

Predicting Interest Rate Volatility Using Information on the Yield Curve

Hideyuki Takamizawa

Graduate School of Commerce and Management,
Hitotsubashi University
Kunitachi Tokyo 186-8601, JAPAN
E-mail: hi.takamizawa[at]r.hit-u.ac.jp

This draft: May 8, 2015
(First draft: February 2012)

Abstract

This study examines whether information on the yield curve is useful for predicting volatility of the yield curve. The information is used within dynamic models by specifying the covariance matrix of changes in yield factors as nonlinear functions of the factors. Using such models, it is found that the information (1) is useful for predicting volatility of the slope factor, achieving the accuracy comparable to the GARCH model; (2) has incremental value for predicting volatility of the curvature factor when combined with a volatility-specific factor; (3) does not much improve prediction of volatility of the level factor once the volatility-specific factor is introduced.

Keywords: Yield curve, Volatility, Level-dependence, Approximation of conditional moments.

JEL codes: C58, E43, G12, G17.

1 Introduction

It does not seem unreasonable to think that the current yield curve contains some information on the volatility of changes in interest rates. In making bond portfolios or managing interest rate risks, investors will take account of conditional second moments of bond returns or yield changes. The resulting shape of the yield curve will then reflect investors' views toward the volatility. The purpose of this study is to examine whether information on the yield curve is useful for predicting the volatility of the yield curve.

The idea of relating interest rate volatility to the yield curve is not new. Brown and Schaefer (1994), Christiansen and Lund (2005), Joslin (2010), Litterman, Scheinkman, and Weiss (1991), and Phoa (1997) relate the volatility to the curvature, or convexity, of the yield curve. Time-series studies using long historical data on U.S. interest rates find a relation between the volatility and the level of a particular yield, especially the short-term rate, such that high volatility is accompanied by high level; see, e.g., Andersen and Lund (1997a), Ball and Torous (1999), Chan, Karolyi, Longstaff, and Sanders (1992), Durham (2003), and Gallant and Tauchen (1998).

This simple level-volatility relationship, however, no longer seems to be a decisive feature for relatively recent data. Figure 1 shows the time series of the level and realized volatility of the first principal component (PC) constructed from U.S. dollar LIBOR and swap rates over 1991–2009: The details of these data are provided in Section 2. Note that the first PC (PC1) is interpreted as a level factor of the yield curve. It is observed that sharp rise in the volatility of PC1 around 2001–03 and 2008–09 is actually accompanied by the fall in the level of PC1.

It is, therefore, not surprising that more recent studies using these data are skeptical about the possibility of extracting volatility information from the yield curve. Andersen and Benzoni (2010) test affine spanning conditions that yield variances, both *ex ante* and *ex post*, can be expressed by some linear combinations of yield levels if affine term structure models are true, and reject these conditions. A direct implication of this result is that the relationship between the volatility and the curvature of the yield curve is not supported by the data because the curvature is normally measured by a linear combination of yields. Collin-Dufresne, Goldstein, and Jones (2009), and Jacobs and Karoui (2009) report that yield variances extracted from the cross-section of yields through affine term structure models do not behave similarly to typical variance measures computed from time series

data.

It is too early to conclude, however, that the yield curve is of little relevance to the volatility. Although information on the cross-section of yields alone may not be rich enough to identify volatility-specific factors, it may still be useful if it is combined with information on the time-series of yields. Furthermore, nonlinear relationships between variances and levels of yields may exist even though a linear relationship as implied by the affine models is not supported.

This study explores these possibilities that are not fully studied by the earlier work. To combine information on the cross-section of yields with information on the time-series of yields, dynamic models of yield factors, rather than regression models, are employed. Then, nonlinear relationships between variances and levels of yields are incorporated into the dynamic models. Specifically, the covariance matrix of changes in yield factors is specified as nonlinear functions of the factors themselves. This is how information on the yield curve is used: It is used for specification, but not for extraction, of the volatility.

As such, the approach of examining information content of the yield curve with respect to the volatility is different from that in the earlier work. Bikbov and Chernov (2011), and Thompson (2008) use no-arbitrage affine models with particular attention to whether model-implied behavior of the volatility changes by modifying estimation methods or adding options data. This study uses both affine and non-affine models and estimates them using only time-series dimension of interest rate data with particular attention to whether information on the yield curve is useful for predicting the volatility. Also, the dynamic models of the yield curve are different from those in the earlier work. Christiansen (2005) embeds the GARCH volatility. Christiansen (2004), and Pérignon and Smith (2007) incorporate regime switching into volatility modeling. This study considers a volatility-specific factor that has a similar role to the GARCH volatility, but not regime switching. Instead, the models explored here are characterized by more flexible level-dependent specifications of the volatility. Using such models, this study uncovers both usefulness and limitations of information content of the yield curve with respect to the volatility and contributes to the literature by promoting understanding of appropriate underlying models that are capable of predicting the volatility without sacrificing the goodness-of-fit to the cross-section of no-arbitrage bond prices.

Section 2 explains the data and realized measures for yield factors and their volatil-

ities. Section 3 provides level-dependent volatility models that are appropriate for the research objective together with several competing models. Sections 4 and 5 examine predictive accuracy of the level-dependent volatility models before and after introducing the volatility-specific factor, respectively. Section 6 provides concluding remarks. Appendices collect technical arguments including an approximation method used for estimation and prediction.

2 Data and realized measures

2.1 Dataset

This study uses data on U.S. dollar LIBOR with maturities of 6 and 12 months and swap rates with maturities of 2, 3, 4, 5, 7, and 10 years. The sample period is from January 4, 1991 to May 27, 2009. The LIBOR and swap rates are transformed to zero-coupon bond yields on a continuously compounded basis using a bootstrap method with linear interpolation applied to discount functions. The maturities of the zero yields used for the analysis are 0.5, 1, 2, 3, 5, and 10 years.

Weekly data consist of Wednesday observations. The in-sample data for estimation cover up to April 9, 2003 with 641 observations, and the out-of-sample data for prediction contain 320 observations. This division allows for incorporating information on the lowest range of yields into model estimation as well as reserving sufficient out-of-sample observations. But it is noted that other divisions do not materially change the results presented below.

This dataset is selected by the following reasons. First, we can make more challenging the purpose of predicting the volatility using information on the yield curve because a simple level-volatility relationship disappears in the recent data as seen in Figure 1. Furthermore, it is also anticipated from Figure 1 that the out-of-sample behavior of the yield level and volatility makes prediction even more difficult. Second, we can focus on volatility prediction without introducing an additional complexity of regime switching. As documented by Dai, Singleton, and Yang (2007), this sample period can be regarded as a single regime when viewed from a long history of U.S. interest rates.

2.2 Realized factors

Since the primary interest of this study is in volatility prediction using information on the yield curve, but not in factor identification, observed variables are used as proxies for yield factors. First of all, the number of factors is assumed to be three, following Litterman and Scheinkman (1991) and the subsequent studies. Then, the first three PCs are selected as proxies for yield factors, which can be interpreted as the level, slope, and curvature factors of the yield curve. A rotation matrix to obtain the PCs is calculated from the covariance matrix of changes in yields, which is estimated using the in-sample weekly data. This rotation matrix is fixed and used in the out-of-sample period to avoid the case in which the interpretation of the PCs is different between the in- and out-of-sample periods.

The first three PCs are actually convenient proxies relative to the others. They allow us to focus on the volatilities of the PCs without paying much attention to the covariances between the PCs. Additionally, they enhance the interpretation of the results regarding estimation and prediction as will be seen in Sections 4 and 5. Still, a nonlinear relationship between variances and levels of yields can be detected based on the PCs. Since the PCs are obtained by linear combinations of yields, if the volatility of the PCs is nonlinear in the level of the PCs, the volatility of yields is also nonlinear in the level of yields.

2.3 Realized volatility

Using daily data, a realized measure of the conditional variance of changes in the PCs is constructed. The daily series of zero yields is first transformed into those of the PCs. This is done using the same rotation matrix calculated with the weekly data to avoid the case in which the interpretation of the PCs is different between the weekly and daily data.

Denote the i -th PC or PC i at time t by $x_{t,i}$, and a realized measure of the one-week ahead conditional variance of $x_{t,i}$ is computed as

$$RV_{t,t+\Delta,i} = \sum_{k=1}^{m_{t+\Delta}} \left(x_{t+\frac{\Delta}{m_{t+\Delta}}k,i} - x_{t+\frac{\Delta}{m_{t+\Delta}}(k-1),i} \right)^2 \quad (i = 1, 2, 3), \quad (1)$$

where Δ is a week interval set to $1/52$, and m_t is the number of observations during a week ending at time t (usually $m_t = 5$). A realized measure of the h -week ahead conditional variance is computed as

$$RV_{t,t+h\Delta,i} = \sum_{j=1}^h RV_{t+(j-1)\Delta,t+j\Delta,i}. \quad (2)$$

The annualized realized variance is obtained by dividing $RV_{t,t+h\Delta,i}$ by $h\Delta$.

Forecasting horizons are set to 4, 8, 16, and 32 weeks. We start with the 4-week horizon to reduce potential impacts of noise or measurement error and extend the horizon up to 32 weeks to examine the effect of longer horizons on predictive accuracy. These realized variances are generated weekly, which means that there are overlapping observations on daily squared changes in the PCs between successive observations of the realized variances. For example, for $h = 4$, daily observations over three weeks overlap between t and $t + \Delta$ observations of the four-week realized variance.

3 Models

The purpose of this study is to examine whether the current yield curve contains useful information for predicting the volatility of the yield curve. To achieve this purpose, this study uses dynamic models that allow for combining information on the cross-section and time-series of yields and considers nonlinear relationships between variances and levels of yields within these models. Section 3.1 proposes such appropriate models and Section 3.2 provides several competing models.

3.1 Level-dependent volatility models

Let X_t be a vector consisting of the first three PCs, and the instantaneous change in X_t is assumed to follow

$$dX_t \sim N [(K_0 + K_1 X_t)dt, \Sigma_t dt] . \quad (3)$$

The instantaneous mean, or the drift, is simply specified as a linear function of X_t because the primary interest of this study is in volatility prediction, but not in level prediction. Additionally, this simple specification is convenient for comparing the proposed models with the affine models. Specifically, given the same specification of the drift, the difference in volatility prediction, if any, can be attributed solely to the difference in volatility specification. In other words, potential impacts of the drift on volatility prediction can be controlled.

Information on the yield curve is embedded into the covariance matrix, Σ_t , by specifying it as functions of X_t . But it should be considered that Σ_t is positive definite. Then, one simple approach to guarantee the positive definiteness is to take the spectral decomposition of Σ_t and model the eigenvalues as functions of X_t such that they take positive

values for arbitrary X_t . Specifically,

$$\Sigma_t = PL_tP' , \quad (4)$$

where L_t is a diagonal matrix consisting of the eigenvalues and P is an orthogonal matrix having the corresponding eigenvectors of unit length in its columns. Note that L_t is time-varying whereas P is fixed. The previous studies employing this approach also assume fixed P ; see, e.g., Fan, Gupta, and Ritchken (2003), Han (2007), Jarrow, Li, and Zhao (2007), Longstaff, Santa-Clara, and Schwartz (2001), and Pérignon and Villa (2006).

P can be parameterized as,

$$P = \begin{pmatrix} 1 & 0 & 0 \\ 0 & \cos \varphi_3^P & -\sin \varphi_3^P \\ 0 & \sin \varphi_3^P & \cos \varphi_3^P \end{pmatrix} \begin{pmatrix} \cos \varphi_2^P & 0 & -\sin \varphi_2^P \\ 0 & 1 & 0 \\ \sin \varphi_2^P & 0 & \cos \varphi_2^P \end{pmatrix} \begin{pmatrix} \cos \varphi_1^P & -\sin \varphi_1^P & 0 \\ \sin \varphi_1^P & \cos \varphi_1^P & 0 \\ 0 & 0 & 1 \end{pmatrix} . \quad (5)$$

Free parameters in P are actually three: $\sin \varphi_i^P$ ($i = 1, 2, 3$). This is because though P is a three-by-three matrix, there are six constraints arising from the fact that each column of P is of unit length and orthogonal to the others. For identification, $\varphi_i^P \in [-\pi/2, \pi/2]$ is placed, so that $\cos \varphi_i^P = \sqrt{1 - \sin^2 \varphi_i^P}$.

The diagonal elements of L_t in equation (4) are specified as functions of X_t , which are denoted as $L_i(X_t)$ ($i = 1, 2, 3$). To guarantee $L_i(X_t) > 0$ for arbitrary X_t , this study proposes the following two models.

3.1.1 SVQ model

The first model labeled as SVQ (Stochastic Volatility with Quadratic specification) specifies $L_i(X_t)$ as

$$L_i(X_t) = c_i + X_t' \Gamma^i X_t \quad (i = 1, 2, 3) , \quad (6)$$

where Γ^i is either a positive definite matrix with $c_i \geq 0$ or a nonnegative definite matrix with $c_i > 0$. In the estimation, the latter restriction is placed on Γ^i and c_i as this can lead to a more parsimonious specification: $\Gamma^i = \mathbf{0}$ is possible as long as the data support. Similar to Σ_t , the non-negative definite matrix Γ^i is parameterized based on the spectral decomposition:

$$\Gamma^i = Q^i M^i Q^{i'} , \quad (7)$$

where

$$M^i = \begin{pmatrix} m_1^i & 0 & 0 \\ 0 & m_2^i & 0 \\ 0 & 0 & m_3^i \end{pmatrix} \quad \text{with} \quad 0 \leq m_1^i \leq m_2^i \leq m_3^i, \quad (8)$$

and

$$Q^i = \begin{pmatrix} 1 & 0 & 0 \\ 0 & \cos \varphi_3^{Q^i} & -\sin \varphi_3^{Q^i} \\ 0 & \sin \varphi_3^{Q^i} & \cos \varphi_3^{Q^i} \end{pmatrix} \begin{pmatrix} \cos \varphi_2^{Q^i} & 0 & -\sin \varphi_2^{Q^i} \\ 0 & 1 & 0 \\ \sin \varphi_2^{Q^i} & 0 & \cos \varphi_2^{Q^i} \end{pmatrix} \begin{pmatrix} \cos \varphi_1^{Q^i} & -\sin \varphi_1^{Q^i} & 0 \\ \sin \varphi_1^{Q^i} & \cos \varphi_1^{Q^i} & 0 \\ 0 & 0 & 1 \end{pmatrix}, \quad (9)$$

with $\varphi_j^{Q^i} \in [-\pi/2, \pi/2]$ ($j = 1, 2, 3$). It is noted that $\sin \varphi_j^{Q^i}$ cannot be identified for some m_j^i . For example, when $m_j^i = 0$ for all j , $\sin \varphi_j^{Q^i}$ cannot be identified for any j . In such cases, $\sin \varphi_j^{Q^i} = 0$ is placed. In the estimation, some parameters reach the boundary values or violate the sign constraints. Such parameters are fixed in the following and then the remaining parameters are re-estimated: $c_i = 10^{-8}$ for $c_i > 0$ and $m_j^i = 0$ for $m_j^i \geq 0$.

3.1.2 SVE model

The second model labeled as SVE (Stochastic Volatility with Exponential specification) specifies $L_i(X_t)$ as

$$L_i(X_t) = \exp \{s_{i0} + s'_i X_t\} \quad (i = 1, 2, 3). \quad (10)$$

No parameter restriction is required for the SVE model. The exponential specification that naturally avoids negative volatility is popular in time series analysis; see, e.g., Andersen and Lund (1997a, b), Ball and Torous (1999), and Gallant and Tauchen (1998).

3.1.3 Discussion on the level-dependent volatility models proposed here

The SVQ/E models are just a few examples of level-dependent volatility models. Naturally, they are restrictive. The first source of the restriction is the eigenvectors that are fixed: They may also vary over time. This restriction may not be severe, however, as long as the PCs, which are unconditionally uncorrelated with each other, are used as proxies for the factors. In fact, as will be seen in Section 4.1, none of the parameters in P are statistically significant, which means that the PCs are also nearly conditionally uncorrelated. The second source of the restriction is the eigenvalues that are quadratic or exponential functions of X_t : They may be more general functions of X_t taking positive values. Nevertheless, the degree of level-dependence in the SVQ/E models is more

deepened than it is in the previous models where the volatility depends on one or at most two variables. Therefore, bearing in mind that the proposed models are approximations of more general level-dependent volatility models, this study utilizes them for examining information content of the yield curve with respect to the volatility.

Furthermore, when the proposed models are used as underlying models of the term structure of interest rates by providing market price of risk and imposing no-arbitrage conditions, there are both strengths and weaknesses. An obvious weakness is that no-arbitrage bond prices cannot be derived as closed-form functions of X_t . But this weakness may be overcome by relying on analytical approximation. An example of such approximation is proposed by Takamizawa and Shoji (2009). Another weakness is the lack of unspanned volatility factors that are irrelevant to the cross-section of yields. The necessity of such factors for describing joint data on bonds and options is emphasized by many studies; see, e.g., Collin-Dufresne and Goldstein (2002), Han (2007), Heidari and Wu (2003, 2009), Jarrow et al. (2007), and Li and Zhao (2006). The introduction of the unspanned factors into the level-dependent volatility models is straightforward. Therefore, this study also considers extended models having both level-dependent and unspanned volatilities in Section 5, aiming at examining relative importance of the level-dependent volatility over the unspanned volatility.

On the other hand, the proposed models have strength of accommodating the following two properties: (i) the covariance matrix is time-varying; (ii) all factors can change signs over time. The affine models cannot accommodate both: The Gaussian model satisfies property (ii) but not property (i), whereas the affine models with stochastic volatility satisfy property (i) but not property (ii). The advantage of satisfying property (ii) is demonstrated by Duffee (2002). Specifically, property (ii) allows for flexible modeling of the market price of risk, which in turn allows for flexible modeling of the physical dynamics of the factors. Then, Duffee (2002) shows that the Gaussian model having this flexibility predicts the level of yields more accurately than any other affine model. This finding suggests that since the proposed models can be regarded as an extension of the Gaussian model (in a way where the yield factors also drive the covariance matrix), they are expected to inherit the predictive power for the level of yields and at the same time improve the predictive power for the volatility of yields.

3.2 Competing models

3.2.1 Gaussian or $A_0(3)$ model

The instantaneous change in X_t for the Gaussian model is given by (3) with Σ_t replaced by Σ , a constant positive definite matrix. Since given a forecasting horizon h , the Gaussian model produces a constant forecast independently of X_t , it serves as a benchmark as does Random Walk in predicting the level of yields.

This study denotes the Gaussian model as $A_0(3)$, which is a conventional notation developed by Dai and Singleton (2000). In general, $A_m(n)$ stands for an affine model in terms of both physical and risk-neutral dynamics of the factors, where n is the total number of factors, among which m factors drive the covariance matrix of the n factors and jointly follow the square-root processes taking non-negative values. Although this study considers only the physical dynamics, this notation is used for convenience.

3.2.2 $A_1(3)$ model

An affine model with stochastic volatility, $A_1(3)$, is adopted, which strikes the balance between fitting time-series and cross-sectional properties of interest rate data as documented by Dai and Singleton (2000). Besides, since the PCs are used as proxies for the factors, the square-root process is applicable only to PC1, but not to PC2 or PC3 as they potentially change signs over time.

The instantaneous change in X_t for the $A_1(3)$ model is also given by (3), where the i -th diagonal element of Σ_t , denoted as $\Sigma_{t,i}$, is given by

$$\Sigma_{t,1} = \sigma_1 x_{t,1}, \quad (11)$$

$$\Sigma_{t,i} = \sigma_i x_{t,1} + c_i \quad (i = 2, 3), \quad (12)$$

and the off-diagonal elements of Σ_t are assumed to be zero.¹ It may be concerned that $x_{t,1}$ or the level factor also plays a role of driving the volatility. However, even if the volatility factor in the $A_1(3)$ model is treated as a latent process and backed out from the cross-section of yields, it is often the case that the resulting process is very persistent

¹The $A_1(3)$ model with non-diagonal covariance matrix is also estimated. The results (not shown in this draft but available upon request) are that most of the off-diagonal parameters are statistically insignificant and hence that volatility prediction is little affected, which are not surprising as the PCs are used as proxies for the factors.

and highly correlated with the level factor; see, e.g., Jacobs and Karoui (2009, Table 7). Furthermore, the earlier work also uses an observed yield for modeling level-dependent volatility of the yield curve; see, e.g., Pérignon and Smith (2007).

The covariance matrix of the $A_1(3)$ model is also level-dependent, however, the degree of level-dependence is different between the $A_1(3)$ and SVQ/E models. The motivation of adopting the $A_1(3)$ model is to examine whether the following two properties, embedded in the SVQ/E models but not in the $A_1(3)$ model, are jointly helpful for predicting the volatility: One is a nonlinear relationship between variances and levels and the other is the involvement of all factors in driving the volatility.

3.2.3 CEV model

A model with constant elasticity of volatility (CEV) is a representative model with level-dependent volatility. Some affine models can also be regarded as special cases of the CEV class of models. Examples of the CEV class include one-factor models proposed by Cox, Ingersoll, and Ross (1985), Chan et al. (1992), and Aït-Sahalia (1996), and multi-factor models proposed by Longstaff and Schwartz (1992), Brenner, Harjes, and Kroner (1996), and Andersen and Lund (1997a, b). A CEV model considered in this study is such that it nests the $A_1(3)$ model provided in equations (11) and (12). Specifically, the instantaneous change in X_t is given by (3), where the diagonal covariance matrix is specified as

$$\Sigma_{t,1} = \sigma_1 x_{t,1}^{\gamma_1}, \quad (13)$$

$$\Sigma_{t,i} = \sigma_i x_{t,1}^{\gamma_i} + c_i \quad (i = 2, 3). \quad (14)$$

Note that if the value of γ_i in equation (14) is statistically insignificant, it is difficult to identify both σ_i and c_i . In such case, $\sigma_i = \gamma_i = 0$ is placed to retain numerical stability.

The relationship between variances and levels is nonlinear in the CEV model, however, only the level factor $x_{t,1}$ contributes to driving the volatility. The motivation of selecting the CEV model is to examine whether the multivariate level-dependence, which is embedded in the SVQ/E models but not in the CEV model, has additional value for predicting the volatility.

3.2.4 GARCH(1,1) model

Although there are a number of variants of the GARCH model, the GARCH(1,1) model is selected because more complicated models do not necessarily beat the simplest one in

out-of-sample tests; see Hansen and Lunde (2005). In fact, as will be seen in Section 4.2, the model has a remarkably high out-of-sample predictive power for the volatility of PC1. The model is fitted to weekly data on each $x_{t,i}$ and hence specified as

$$x_{t+\Delta,i} = \alpha_i + \beta_i x_{t,i} + \sqrt{h_{t+\Delta,i}} z_{t+\Delta,i}, \quad (15)$$

$$h_{t+\Delta,i} = \omega_i + \phi_i h_{t,i} z_{t,i}^2 + \rho_i h_{t,i}, \quad (16)$$

where it is assumed for simplicity that $z_{t,i} \sim i.i.d.N(0, 1)$ and is independent of $z_{t,j}$ ($j \neq i$). It is noted that since the PCs are used as proxies for the factors, the estimation with each, but not joint, series does not seem unfavorable for the GARCH model.

Since the GARCH model is a well-established volatility model and estimated with the same data used for the other models provided above, it serves as an alternative benchmark. Specifically, predictive accuracy of each model is statistically tested against that of the GARCH model. It is noted that this comparison framework contrasts to that in Jacobs and Karoui (2009), where the EGARCH model is treated as a *true*, but not competing, model. As such, the comparison here is carried out under more severe conditions.

3.2.5 HAR-RV regression

To set an *absolute* standard for volatility prediction, the HAR-RV regression employed by Andersen and Benzoni (2010) is considered, which is directly fitted to the realized volatility series. In this study, it is specified as

$$\text{HAR-RV: } \sqrt{\frac{RV_{t,t+h\Delta,i}}{h\Delta}} = a_{h,i} + \sum_{j=\{4,8,16,32\}} b_{h,i,j} \sqrt{\frac{RV_{t-j\Delta,t,i}}{j\Delta}} + u_{t+h\Delta,i}^{har}. \quad (17)$$

Since the HAR-RV regression is constructed on the object of prediction, it will provide a conventional reference about upper limit of predictive accuracy in the current setting.²

²This study also considers the mixed data sampling (MIDAS) approach developed by Ghysels, Santa-Clara, and Valkanov (2005, 2006), which is directly fitted to the realized volatility series to generate volatility forecasts. Not surprisingly, therefore, predictive accuracy is very similar to that of the HAR-RV regression and therefore not reported in this draft for saving space.

4 Empirical analysis without a volatility-specific factor

4.1 Estimation

The quasi-maximum likelihood (QML) method is used for estimating the models. The log-likelihood function to be maximized is

$$\sum_t \ln f_T(X_t|X_{t-\Delta}; \Theta), \quad (18)$$

where f_T stands for transition density and Θ for parameter vector. For the CEV and SVQ/E models, since f_T does not have a closed-form for finite observation frequency Δ , it is approximated by the multivariate normal density function. This approximation is also applied to the $A_1(3)$ model for convenience. Since the weekly data with $\Delta = 1/52$ are used for the estimation, this approximation seems acceptable. The conditional first and second moments of the factors, which are substituted into f_T , are computed using a method proposed by Shoji (2002): A brief explanation of this method is provided in Appendix A. Note that the conditional moments for the affine and SVQ models can be computed exactly, as the drift vector is linear and the covariance matrix is at most quadratic in X_t , respectively. Even in this case, equation (43) (without the residual term) in Appendix A is useful for computing the conditional moments.

Some insignificant parameters at the first round of estimation are set to zero and the remaining parameters are re-estimated, which is aimed at keeping the models simple to enhance the interpretation of estimation and prediction results. Since out-of-sample tests are conducted, such simplification does not lead to an unfavorable evaluation of the predictive power of the models.

Table 1 presents the estimation results for the $A_1(3)$ and CEV models. Since the primary interest of this study is in volatility prediction, only the parameter estimates in the covariance matrix are reported and those in the drift are summarized in Table 9. Panel A of Table 1 shows that in the $A_1(3)$ model, the volatility of PC1 exhibits level-dependence but the volatilities of PC2 and PC3 do not because σ_2 and σ_3 are insignificant at the first round of estimation and hence set to zero. These results are basically inherited into the CEV model presented in Panel B of Table 1. In particular, although γ_i ($i = 2, 3$) are free parameters, they are not statistically significant and thus set to zero.³ The volatility of PC1

³The result is in line with Christiansen (2005) but not with Pérignon and Smith (2007). These studies use the data starting in 1970, including periods in which yield volatility tends to be high when yield level

also depends on $x_{t,1}$ in the CEV model, however, the estimate of γ_1 sharply distinguishes CEV from $A_1(3)$: It is -0.618 and statistically different from one assumed in $A_1(3)$. The negative estimate of γ_1 indicates that the volatility of PC1 is higher the lower the level of PC1. This inverse relation is consistent with Figure 1 and also reported by Collin-Dufresne et al. (2009, Table 6), and Jacobs and Karoui (2009, Table 7) using relatively recent data on U.S. interest rates. Mathematically, however, γ_1 must not be negative. This is because $x_{t,1}$, starting from an arbitrary initial value $x_{0,1} > 0$, will become negative with positive probability, so the diffusion term, $x_{t,1}^{\gamma_1}$, cannot stay real. Nevertheless, this mathematical inconsistency may be overcome by considering nonlinear drift as proposed by Aït-Sahalia (1996). Specifically, if there is a term in the drift of $x_{t,1}$ such that it increases sufficiently rapidly as $x_{t,1}$ approaches zero, then $x_{t,1}$ does not reach zero in finite time even though the volatility of $x_{t,1}$ becomes also very large. Put simply, the positive drift more than offsets the volatility near the boundary of zero, which prevents $x_{t,1}$ from going below zero. If in addition this term is sufficiently small in an observed range of $x_{t,1}$, it is innocuous to actual estimation.⁴ In this way, since the deficiency of the CEV model with negative γ_1 may virtually be corrected, it is continuously used as a competing model for volatility prediction.

Table 2 presents the estimation results for the SVQ/E models. In both models, none of the parameters in P , $\sin \varphi_i^P$ ($i = 1, 2, 3$), are estimated significantly, leading to $P = I$ (the identity matrix). Consequently, $L_i(X_t)$ can be interpreted as the instantaneous variance of PC i ($i = 1, 2, 3$). Panel A of Table 2 shows that many parameters in the SVQ model are fixed due mainly to sign constraints. Nevertheless, the volatilities of all factors are time-varying. In fact, the volatilities of PC2 and PC3 have more complicated level-dependence than the volatility of PC1, where $m_1^1 = m_2^1 = 0$ is placed. The result highlights the importance of nonlinear and multivariate specifications in modeling level-dependent volatility for recent interest rate data. The estimation results of the SVQ model are further discussed in Section 4.3 in relation to the predictive power for the volatility

is high. Pérignon and Smith (2007) use the three-month yield for driving the volatility of the yield curve, and detect the level-dependence for all factors. Christiansen (2005) uses the slope factor for driving the volatility of the slope factor, and finds little evidence of the level-dependence for the slope factor.

⁴A simple formulation of the above arguments is as follows. Suppose the drift of $x_{t,1}$ is of the form, $\mu(x_{t,1}) = a + bx_{t,1} + cx_{t,1}^d$ with $c > 0$. Then, if $d < \gamma_1 - 1$, $x_{t,1}$ does not reach zero in finite time; see Aït-Sahalia (1996, Appendix). Additionally, if c is fixed at a sufficiently small number, the last term in $\mu(x_{t,1})$ is negligible in an observed range of $x_{t,1}$, which is actually $[0.043, 0.194]$ in the dataset used here.

evaluated in Section 4.2.

The estimation results for the SVE model, presented in Panel B of Table 2, are more intuitive because of the simpler form of the eigenvalue functions. Most notably, none of the coefficients on $x_{t,1}$, s_{i1} ($i = 1, 2, 3$), are statistically significant, indicating that among the yield factors, the level factor is the least relevant to the volatility. It then follows that the $A_1(3)$ model, either in the particular form considered in this study or in a general form where the volatility factor is treated as a latent process, has difficulty in predicting the volatility: As noted in Section 3.2.2, a volatility factor, when backed out from the cross-section of yields, is often highly correlated with the level factor.

4.2 Prediction

4.2.1 Setting

To enhance the interpretation of the results, the object of prediction is annualized standard deviation, but not variance. The predictive accuracy is evaluated by root mean squared prediction error (RMSPE) using both in- and out-of-sample data. The prediction errors are the residuals of the following equations (except the HAR-RV regression given in equation (17)):

$$\sqrt{\frac{RV_{t,t+h\Delta,i}}{h\Delta}} = \sqrt{\frac{\text{var}_t[x_{t+h\Delta,i}]}{h\Delta}} + u_{t+h\Delta,i}, \quad (19)$$

$$\sqrt{\frac{RV_{t,t+h\Delta,i}}{h\Delta}} = a_{h,i} + b_{h,i} \sqrt{\frac{\text{var}_t[x_{t+h\Delta,i}]}{h\Delta}} + u_{t+h\Delta,i}^{reg}, \quad (20)$$

where $\text{var}_t[\cdot]$ stands for model-implied conditional variance. Equation (19) is aimed at examining pure predictive power of the models whereas equation (20) is at examining to what degree the prediction improves by the help of the forecasting regression. $(a_{h,i}, b_{h,i})$ in equation (20) are estimated by OLS. It is noted that in the HAR-RV regression the first 32 weeks' observations are not used for estimating the parameters (they are used for forming the explanatory variables on the right-hand side of equation (17)). To make competitive conditions equal, these observations are not used in equation (19) or (20).

Since $\text{var}_t[x_{t+h\Delta,i}]$ is not available in closed-form for the CEV and SVE models, it is computed approximately with the same method used for estimating the models. But here, since the time interval for the prediction is up to 32 times longer than that for the estimation, there may be a concern about the accuracy of the approximation. In Appendix

A, the accuracy is checked and confirmed to be high as long as reasonable parameter and state variable values are provided. $\text{var}_t[x_{t+h\Delta,i}]$ for the GARCH model is computed by iteration, an explanation of which is provided in Appendix B.

To statistically compare predictive accuracy, the Diebold and Mariano (DM) (1995) test is implemented as in Colline-Dufresne et al. (2009). Two notes are in order. First, as noted in Section 3.2.4, the GARCH model is the fixed benchmark against which each model is tested. Second, as noted in Section 2.3, since overlapping observations are used for computing realized variances with $h > 1$, the standard error of the test statistic is computed by the method of Newey and West (1987) with lag length equal to h .

4.2.2 Predicted time-series of volatility

To obtain intuition about predictive accuracy of the models, Figures 2–4 display the time series of four-week ahead forecasts of volatilities (annualized standard deviations) of the first three PCs produced by the four models: GARCH, $A_1(3)$, SVQ, SVE. In each graph, the time-series of the corresponding realized value is also plotted with the thin line. It is noted that in generating out-of-sample forecasts, the model parameters are held fixed at the in-sample estimates.

First, Figure 2 shows that the GARCH model is the most successful among the four models in forecasting the volatility of PC1. The sample correlation calculated from the whole sample between the forecast and realized series is 0.57. The forecast series produced by the SVQ/E models look similar and reasonably track the trend of the realized series with the sample correlations around 0.35. In contrast, the $A_1(3)$ model fails to even capture the trend. The forecast series is negatively correlated with the realized series, -0.37 . The result is not surprising, however, given similar graphs obtained by the previous studies; see Colline-Dufresne et al. (2009, Figure 1), Jacobs and Karoui (2009, Figures 1 and 2), and Thompson (2008, Figure 3).

Next, Figure 3 shows that the SVQ model performs the best in predicting the volatility of PC2 with the sample correlation 0.51. It is also capable of generating the forecast series that varies intensively in accordance with the realized series. This is particularly evident in the out-of-sample period, implying in turn that the model parameters are stable between the in- and out-of-sample periods. The GARCH model follows with the sample correlation 0.44. The SVE model exhibits a reasonable correlation 0.41 compared with

the GARCH model, however, it fails to produce sufficient variation in forecasts. The forecast series of the $A_1(3)$ model looks constant over time. But actually, it slightly varies and is negatively correlated with the realized series. The slight variation comes from the fact that the instantaneous variance of PC1 is time-varying and that PC1 and PC2 are conditionally correlated through the drift; see Table 9 for the estimates of the drift parameters. Therefore, even though the instantaneous variance of PC2 is constant, the conditional variance of PC2 over the four-week horizon or longer is not.

Finally, Figure 4 shows that the GARCH model returns to the best performer in predicting the volatility of PC3 with the sample correlation 0.49, followed by the SVE model with the sample correlation 0.45. The forecast series of the SVQ model looks similar to that of the GARCH model for many periods, however, the sample correlation is somewhat lower, 0.41. The graph for the $A_1(3)$ model is similar to that presented in Figure 3 for the same reason mentioned above.

As seen in Figures 2–4, while the forecast series of the volatility of PC1 are similar between the SVQ and SVE models, those of the volatilities of PC2 and PC3 are not. The results can be explained as follows. First, for the SVE model, all forecast series are generally less volatile because an exponential function can be well approximated by a linear function in a narrow range where the eigenvalues move and because a linear function is difficult to produce intensive variation. Second, for the SVQ model, the forecast series of the volatility of PC1 exhibits variation as moderate as that for the SVE model because many parameters are set to zero in $L_1(X_t)$. Conversely, the level-dependence of $L_2(X_t)$ and $L_3(X_t)$ is more complex than that of $L_1(X_t)$ and this complexity is the key to generating intensive variation in forecasts. This key point is further discussed in Section 4.3.

4.2.3 Prediction results without forecasting regression

Table 3 presents the RMSPEs without the forecasting regression, where the prediction error is the residual in equation (19). The RMSPEs are multiplied by 10^4 and thus interpreted in units of basis points (bps, 1 bp = 0.0001). The smallest and the second smallest numbers in each column are displayed in bold and italic, respectively. * and ** indicate that the DM test rejects the null hypothesis that predictive accuracy between a model in each row and the GARCH model is equal at the 5% and 1% significance levels, respectively. In making the out-of-sample prediction, the parameter values of the models

are held fixed at the in-sample estimates. To make competitive conditions equal, the parameter values of the HAR-RV regression are also held fixed throughout the out-of-sample period.

Overall, the HAR-RV regression has the smallest RMSPEs for all cases in-sample and all but three cases out-of-sample. Specifically, the superiority of the HAR-RV regression is most evident in predicting the volatility of PC2, the accuracy of which is significantly higher than that of the GARCH model (and most of the other models) both in- and out-of-sample except at $h = 32$. Also, the predictive accuracy for the volatility of PC3 is remarkably high in-sample. In contrast, regarding prediction of the volatility of PC1, the superiority of the HAR-RV regression is not relatively high. The result suggests that it is originally difficult to predict the volatility of the level factor or long-term yields that are highly correlated with the level factor, and is consistent with the previous results; Andersen and Benzoni (2010), Jacobs and Karoui (2009). Instead, the GARCH model works well in this dimension, exhibiting the smallest out-of-sample RMSPEs except at $h = 4$.

More precisely, in Panel A of Table 3 displaying the results for PC1, it is noticed that the linear and univariate level-dependence is worse than constant. The differences in RMSPE between the $A_1(3)$ and $A_0(3)$ models are around 13 bps in-sample and range from 14 to 18 bps out-of-sample. Second, a slight modification of the $A_1(3)$ model dramatically improves the predictive accuracy. The CEV model with free elasticity parameter γ_1 has smaller in-sample RMSPEs than the $A_1(3)$ model with $\gamma_1 = 1$ and the $A_0(3)$ model with $\gamma_1 = 0$ at all horizons. In the out-of-sample period, however, while the CEV model still exhibits better predictive accuracy than the $A_1(3)$ model, it is slightly behind the $A_0(3)$ model. Third, the nonlinear and multivariate level-dependence further improves the predictive accuracy. Although the SVQ/E models do not clearly outperform the CEV model in-sample, they do out-of-sample with the differences in RMSPE ranging from 6 to 10 bps. The result suggests that the nonlinear and multivariate level-dependence is more robust to changes in predictive environments. Fourth, the SVQ/E models are outperformed by the GARCH model. Specifically, between the SVE and GARCH models, the in-sample differences are at most 2.4 bps at $h = 4$ and not statistically significant, however, the out-of-sample differences reach up to 18 bps at $h = 8$. Nevertheless, the out-of-sample differences are not statistically significant except at $h = 4$ because of the

small sample size and of very volatile forecast errors as evidenced in Figure 2.

In Panel B of Table 3 displaying the results for PC2, the RMSPEs for the $A_0(3)$, $A_1(3)$, and CEV models are almost the same, resulting from the constant instantaneous variance of PC2 (they are not exactly the same, however, for the same reason mentioned in Section 4.2.2). The SVQ/E models continuously outperform these models. For example, the differences in RMSPE between the CEV and SVQ models are around 3 bps at all horizons in-sample and reach 7.7 bps at $h = 8$ out-of-sample. Though not statistically significant, the SVQ model even outperforms the GARCH model at all horizons in-sample and at $h = 4, 8$ out-of-sample. The SVE model is also comparable to the GARCH model in-sample but slightly worse than it out-of-sample. These results demonstrate that information on the yield curve is useful for predicting the volatility of PC2 when nonlinear and multivariate level-dependent specifications are considered.

Conversely, Panel C of Table 3 shows that the nonlinear and multivariate level-dependence has difficulty in predicting the volatility of PC3. Compared with the in-sample RMSPEs for the $A_0(3)$ model, those for the SVE model little differ whereas those for the SVQ model are only marginally smaller by around 1 bp. They both are statistically significantly larger than those for the GARCH model, which in turn are statistically significantly larger than those for the HAR-RV regression. The difficulty of the nonlinear and multivariate level-dependence is somewhat mitigated in the out-of-sample period. The SVQ model is better than the $A_0(3)$ model with the differences in RMSPE ranging from 4 to 6 bps and even comparable to the GARCH model.

4.2.4 *Prediction results with forecasting regression*

As expected, the models fitted to the weekly data do not in general predict the volatility as accurately as the HAR-RV regression. Of interest here is whether this is still the case after introducing the forecasting regression. Also of interest is whether the advantage of the nonlinear and multivariate level-dependence over the linear or nonlinear univariate level-dependence remains to hold.

Table 4 presents the RMSPEs with the forecasting regression, where the prediction error is the residual in equation (20). Before reporting the results, the following notes are in order. First, in making the out-of-sample prediction, the parameter values of the models are held fixed at the in-sample estimates as in Section 4.2.3 whereas the parameter values

of the forecasting regression in (20) and of the HAR-RV regression in (17) are re-estimated every time the prediction is made in a rolling window fashion. Second, the RMSPEs for the $A_0(3)$ model are not presented here as the forecasts given h are constant over time. Third, the in-sample RMSPEs are no larger than those in Table 3 by construction of the forecasting regression whereas the out-of-sample RMSPEs are not guaranteed to be smaller. Finally, the in-sample RMSPEs for the HAR-RV regression are the same as those in Table 3, but the results of the statistical significance are different, because predictive accuracy is compared with that of “GARCH + forecasting regression.”

Overall, the differences in RMSPE are narrowed. Indeed, the number of cases in which the difference is statistically significant is decreased from that in Table 3. Nevertheless, the forecasting regression is not effective enough to change the performance ranking. Therefore, only the key results are reported below with particular attention to the performance of the SVQ/E models. First, the SVQ/E models are comparable even to the HAR-RV regression in predicting the volatility of PC1 in-sample but largely outperformed out-of-sample by both the GARCH model and the HAR-RV regression. Second, the SVQ/E models generally produce smaller RMSPEs than the $A_1(3)$ and CEV models, indicating that the advantage of the nonlinear and multivariate level-dependence remains to hold though it is somewhat reduced compared to prior to introducing the forecasting regression. Third, the SVQ model is confirmed to have a robust performance of predicting the volatility of PC2, which is comparable to, or slightly better than, the GARCH model.

4.3 Further evidence on the performance of the SVQ model

This subsection addresses the following two questions: (i) why is the quadratic multivariate specification effective for predicting the volatility of PC2 and relatively effective for predicting the volatility of PC3 compared with the univariate specification, but not for predicting the volatility of PC1; (ii) does information on the past yield curve, in addition to the current one, improve volatility prediction within the quadratic multivariate specification?

To obtain the answer to question (i), we rewrite the instantaneous variances of PC1–3 for the SVQ model given in equations (6)–(9) and substitute the parameter values presented in Panel A of Table 2. First, let $Q^i = (q_1^i \ q_2^i \ q_3^i)$ in equation (9), and the i -th

instantaneous variance can be rewritten as

$$L_i(X_t) = c_i + \sum_{j=1}^3 m_j^i (q_j^{i'} X_t)^2 \quad (i = 1, 2, 3). \quad (21)$$

Then, by substituting the parameter values,⁵

$$L_1 = 10^{-8} + 1.598 \times (-0.295 \times \text{PC2} + 0.956 \times \text{PC3})^2, \quad (22)$$

$$L_2 = 10^{-8} + 0.036 \times (0.201 \times \text{PC1} + 0.964 \times \text{PC2} - 0.176 \times \text{PC3})^2 \\ + 2.933 \times (0.129 \times \text{PC1} + 0.152 \times \text{PC2} + 0.980 \times \text{PC3})^2, \quad (23)$$

$$L_3 = 10^{-8} + 0.006 \times (0.257 \times \text{PC1} + 0.956 \times \text{PC2} - 0.141 \times \text{PC3})^2 \\ + 0.990 \times (0.128 \times \text{PC1} + 0.111 \times \text{PC2} + 0.985 \times \text{PC3})^2. \quad (24)$$

It is seen from equations (22)–(24) that except L_1 where m_1^1 and m_2^1 are set to zero as they are not significantly estimated, the instantaneous variances are decomposed into two quadratic terms with each term consisting of a linear combination of the PCs and hence the current yields. To examine behavior of these components, Figure 5 plots L_i ($i = 1, 2, 3$), where the values of the components of L_i are added up to the value of L_i . A common feature found in L_2 and L_3 is that these two components have different roles. The first term, $m_2(q_2' X_t)^2$, captures persistent movement of the variance whereas the second term, $m_3(q_3' X_t)^2$, captures transitory movement. In fact, it is the second term that generates intensive behavior of the variance in the out-of-sample period. To further examine differences in these terms, Table 5 presents sample statistics of these terms over the whole period. Indeed, compared to the first term, the second term in L_2 and L_3 has a large standard deviation relative to the mean and autocorrelations decaying quickly.

Originally, having different components is the key feature of the GARCH(1,1) model, where the own lag captures persistent movement and the lagged squared innovation of underlying yield process captures transitory movement. The success of the quadratic multivariate specification seems to lie in accommodating this key feature. It is not surprising that persistent component of the volatility is accommodated because persistent yields are used for driving the volatility. A nontrivial result is that transitory component of the

⁵ Q^i ($i = 1, 2, 3$) are realized as follows:

$$Q^1 = \begin{pmatrix} 1.000 & 0.000 & 0.000 \\ 0.000 & 0.956 & -0.295 \\ 0.000 & 0.295 & 0.956 \end{pmatrix}, \quad Q^2 = \begin{pmatrix} 0.971 & 0.201 & 0.129 \\ -0.219 & 0.964 & 0.152 \\ -0.094 & -0.176 & 0.980 \end{pmatrix}, \quad Q^3 = \begin{pmatrix} 0.958 & 0.257 & 0.128 \\ -0.271 & 0.956 & 0.111 \\ -0.094 & -0.141 & 0.985 \end{pmatrix}.$$

volatility can be replicated, at least partially, by linearly combining and then squaring the current yields.

This line of explanation is consistent with why the quadratic multivariate specification does not work well for predicting the volatility of PC1. There is only one component in L_1 as seen in equation (22). From Figure 5 and Table 5, it can capture persistent movement but not transitory movement. The difficulty in predicting the volatility of PC1 is further discussed in Section 5.2 after introducing a volatility specific factor.

If squared changes between the current and past yields embedded in the GARCH model are more or less replicated by squared combinations of the current yields, information on the past yields may not be very helpful for predicting the volatility. Then, question (ii) naturally arises: does information on the past yield curve, in addition to the current one, improve volatility prediction within the quadratic multivariate specification? To obtain the answer to question (ii) more directly, it is convenient to conduct regression analysis, where the realized annual standard deviation is the object of prediction as is the case for the HAR-RV regression. The following regression models are considered, all of which are based on the quadratic multivariate specification.

$$(M1) \quad \sqrt{\frac{RV_{t,t+h\Delta,i}}{h\Delta}} = c_{h,i} + \sqrt{X_t' \Gamma^{h,i} X_t} + e_{t+h\Delta,i}, \quad (25)$$

$$(M2) \quad \sqrt{\frac{RV_{t,t+h\Delta,i}}{h\Delta}} = c_{h,i} + \sqrt{(X_t - a_{h,i} \cdot X_{t-\Delta})' \Gamma^{h,i} (X_t - a_{h,i} \cdot X_{t-\Delta})} + e_{t+h\Delta,i} \quad (26)$$

$$(M3) \quad \sqrt{\frac{RV_{t,t+h\Delta,i}}{h\Delta}} = c_{h,i} + \sqrt{X_t' \Gamma^{h,i} X_t} + \sqrt{X_{t-\Delta}' \Omega^{h,i} X_{t-\Delta}} + e_{t+h\Delta,i}, \quad (27)$$

$$(M4) \quad \sqrt{\frac{RV_{t,t+h\Delta,i}}{h\Delta}} = c_{h,i} + \sum_{j=0}^K b_{h,i,j} \sqrt{X_{t-j\Delta}' \Gamma^{h,i} X_{t-j\Delta}} + e_{t+h\Delta,i}, \quad (28)$$

where $c_{h,i} > 0$, $\Gamma^{h,i}$ and $\Omega^{h,i}$ are non-negative definite matrices, and $b_{h,i,j} \geq 0$.

Several notes are in order. First, the regression models are estimated with each forecasting horizon (index h) and PC (index i) using the in-sample data. Second, in making the out-of-sample prediction, the parameter values are fixed at the in-sample estimates. Third, M1 in equation (25) serves as a baseline model, which uses information on the current yields alone for volatility prediction. Fourth, the other models in equations (26)–(28) are specified such that they nest M1. Specifically, M2 leads to M1 by setting $a_{h,i} = \mathbf{0}$: Note that $a_{h,i} \cdot X_{t-\Delta}$ stands for element-wise product. The lagged level of the PCs, multiplied by $a_{h,i}$, enters into M2 in the form of the first difference. M3 leads to M1 by setting $\Omega^{h,i} = \mathbf{0}$. The lagged level enters into M3 as an additional quadratic term. M4 leads to

M1 by setting $b_{h,i,0} = 1$ and $b_{h,i,j} = 0$ ($j = 1, \dots, K$). More lagged levels are included in M4 in a parsimonious way where $\Gamma^{h,i}$ is fixed independently of lags (index j). To pin down the level of $\Gamma^{h,i}$, $b_{h,i,0} = 1$ is placed in the estimation. The results presented below are those with $K = 1$, however, other values of K do not change the results.⁶

Table 6 presents the RMSPEs expressed in bps. The DM test is also conducted under the null hypothesis that predictive accuracy is equal between M1 and M_i ($i = 2, 3, 4$). The answer to question (ii) derived from Table 6 is that including the lagged level of the PCs does not lead to improving volatility prediction within the quadratic multivariate specification. While the in-sample RMSPEs for M_i ($i = 2, 3, 4$) are no larger than those for M1 by construction, there are generally no noticeable differences. Similarly, the out-of-sample RMSPEs do not vary much across the models. The exception is M3 in predicting the volatility of PC2. However, the statistically better in-sample performance of M3 at $h = 4, 8, 16$ is accompanied by the statistically worse out-of-sample performance, suggesting that M3 is overfitted to the in-sample data.

An additional important finding can be obtained by comparing the results for M1 with those for the SVQ model presented in Table 3. By directly regressing the realized volatility on the quadratic form of the PCs at each forecasting horizon, higher predictive accuracy is indeed achieved for the volatility of PC1: M1 has smaller RMSPEs than SVQ by around 11 bps in-sample and 5–8 bps out-of-sample. This is also the case for the in-sample prediction of the volatility of PC3. However, the regression does not much improve volatility prediction for PC2 or PC3 in the out-of-sample period. The result provides additional evidence that high accuracy can be maintained even by the diffusion model of the PCs estimated with weekly data.

5 Empirical analysis with a volatility-specific factor

We have seen so far that the nonlinear and multivariate level-dependence is more appropriate for predicting the volatility than the linear or nonlinear univariate level-dependence. But it may be concerned that the level-dependence for the $A_1(3)$ and CEV models is too restrictive although such restriction is inevitable as long as the PCs are used as proxies for

⁶This fact may not be surprising because the yields are persistent and hence there is not much incremental information in lagged yields. Also, the result is similar to that for the GARCH volatility, where more lagged terms do not necessarily improve both in- and out-of-sample predictions.

the factors. To address this concern, an additional factor that drives the volatility of the yield curve is introduced into the $A_1(3)$ model. It is also introduced into the SVQ model, which is aimed at examining whether the nonlinear and multivariate level-dependence is still useful.

5.1 Extended models

Let V_t be a latent process affecting the volatilities of all factors considered. This study does not consider multiple volatility factors because according to the principal component analysis applied to the covariance matrix of changes in the three realized volatility series used in this study, the first principal component can explain more than 96% of the variation regardless of the time horizon h in equation (2). V_t is first introduced into the $A_1(3)$ model. Prior to the introduction, $x_{t,1}$ is excluded from the instantaneous variance of PC1 because the linear dependence has an adverse effect on volatility prediction as seen in Table 3. Then, the $A_1(3)$ model is extended to the $A_1(4)$ model, where only V_t drives the covariance matrix. Precisely, the instantaneous change in $(X_t' V_t)'$ for the $A_1(4)$ model is specified as

$$d \begin{pmatrix} X_t \\ V_t \end{pmatrix} \sim N \left[\left\{ \begin{pmatrix} K_0 \\ k_v \theta_v \end{pmatrix} + \begin{pmatrix} K_1 & \mathbf{0} \\ \mathbf{0}' & -k_v \end{pmatrix} \begin{pmatrix} X_t \\ V_t \end{pmatrix} \right\} dt, \Sigma_t dt \right], \quad (29)$$

where $\mathbf{0}$ stands for the three dimensional vector of zeros and where Σ_t is a diagonal covariance matrix with the diagonal elements specified as

$$\Sigma_{t,i} = \sigma_i V_t + c_i \quad (i = 1, 2, 3), \quad (30)$$

$$\Sigma_{t,4} = \sigma_4 V_t. \quad (31)$$

It is noted that Σ_t is diagonal and the drift of X_t does not depend on V_t . This is because we wish to break the interaction between X_t and V_t so that V_t specializes in predicting the volatility as much as possible. In other words, we let it play like an almost unspanned factor in the sense that it is of little relevance to the yield factors. Consequently, this specification will work toward reducing the effectiveness of level-dependence and hence be useful for preventing spurious detection of it. Given the purpose of this study, this benefit will exceed the cost of missing important features such as a leverage effect or an ARCH-in-mean like effect.

V_t is also introduced into the SVQ model in both multiplicative (SVQ-M) and additive (SVQ-A) forms, which is similar in spirit to Brenner et al. (1996) who consider multiplicative and additive combinations of the GARCH and level-dependent volatilities. Since the covariance matrix of the SVQ model is decomposed as $\Sigma_t = PL_tP'$ and P can be regarded as the identity matrix when the PCs are used as proxies for the factors, it holds that $\Sigma_t = L_t$. Then, $\Sigma_{t,i}$ is used in this section, instead of $L_i(X_t)$, for denoting the instantaneous variance of PC i . The instantaneous change in $(X_t' V_t)'$ for the SVQ-M/-A models is specified as in (29), where the diagonal elements of Σ_t are specified as

$$\text{(SVQ-M)} \quad \Sigma_{t,i} = X_t' \Gamma^i X_t V_t + c_i \quad (i = 1, 2, 3), \quad (32)$$

$$\text{(SVQ-A)} \quad \Sigma_{t,i} = \sigma_i V_t + X_t' \Gamma^i X_t \quad (i = 1, 2, 3), \quad (33)$$

and $\Sigma_{t,4}$ is the same as in the $A_1(4)$ model.

It is noted that this study does not explore appropriate specifications for the dynamics of V_t because the purpose is to compare the SVQ-M/-A models with the $A_1(4)$ model and examine whether the nonlinear and multivariate level-dependence is still useful given the presence of the volatility-specific factor. To achieve this purpose, it is necessary to use the same specification for the dynamics of V_t and control for potential impacts associated with V_t . It is also noted that the same extension is not applied to the SVE model as it does not perform equally well to the SVQ model as seen in Section 4.2.

5.2 Estimation and prediction results

The QML method is used for estimating the models. Since V_t is treated as a latent process, it needs to be integrated out of the likelihood function. The likelihood function is computed following Thompson (2008).⁷ For identification of the parameters in (29), $\theta_v = 1$ is placed as in Gallant and Tauchen (1998). This means that given $\kappa_v > 0$ the unconditional mean of V_t is set to one.

Table 7 presents the estimation results for the $A_1(4)$, SVQ-M and SVQ-A models. As before, only the parameter estimates in the covariance matrix are reported and those in the drift are summarized in Table 9. In Panel A, the estimates of σ_i ($i = 1, 2, 3$) in the $A_1(4)$ model are all significant, indicating the importance of introducing the volatility-specific factor for capturing time-varying features of the volatility. In Panel B, the significantly

⁷We first discretize the values of V_t with the number of grid points, d , set to 21 and sum the likelihood function over the points. We also tried $d = 51$ and the results are essentially the same as those with $d = 21$.

estimated parameters in the SVQ-M model are the same as those for the original SVQ model presented in Table 2, indicating that the significance of the level-dependence remains unchanged in the multiplicative form. In Panel C, the level-dependence is also confirmed in the SVQ-A model, however, the number of significant parameters is reduced. Most importantly, $m_2^2 = m_2^3 = 0$ is placed. The simpler level-dependence is supported because the level-dependent term in the SVQ-A model is in place of the constant term in the $A_1(4)$ model and because V_t can generate sufficient variation in forecast as seen below.

Figure 6 presents the time-series of four-week ahead forecasts of volatilities (annualized standard deviations) of the first three PCs generated by the $A_1(4)$ model (the left panel) and the SVQ-A model (the right panel). It is clear from the left panel that owing to V_t the forecast series vary sufficiently and well track the realized series. It is also seen from the right panel that the level-dependent term generates additional variation in forecasts, which appears more desirable particularly in the out-of-sample period.

Table 8 presents the RMSPEs without the forecasting regression.⁸ For comparison, the RMSPEs for the GARCH model and the HAR-RV regression presented in Table 3 are add to the last two rows in each panel. The DM tests for equal predictive accuracy are conducted based on the GARCH model also here. The results are reported to answer whether the introduction of the volatility-specific factor improves the prediction through the comparison of the $A_1(4)$ and $A_1(3)$ models and whether the nonlinear and multivariate level-dependence is still of value through the comparison of the SVQ-M/-A and $A_1(4)$ models.

In Panel A of Table 8 displaying the results for PC1, it is found that the volatility-specific factor significantly improves the prediction both in- and out-of-sample. Specifically, the RMSPEs for the $A_1(4)$ model are uniformly smaller than those for the $A_1(3)$ model by up to 20.8 bps at $h = 16$ in-sample and by up to 38.6 bps at $h = 8$ out-of-sample. In fact, the introduction of V_t is effective enough to make the $A_1(4)$ model comparable to the GARCH model in-sample though it is still not enough out-of-sample. Also, V_t improves predictive accuracy when introduced into the SVQ model. Specifically, the RMSPEs for the SVQ-A model are smaller than those for the original SVQ model by 5–8 bps in-sample and 5–13 bps out-of-sample. Next, the nonlinear and multivariate level-dependence does not seem to add value to predicting the volatility of PC1 once V_t is introduced. Between

⁸The RMSPEs with the forecasting regression are not reported because the points to note are essentially the same as those from Table 8, as is the case between Tables 3 and 4.

the SVQ-A and $A_1(4)$ models, while the former (latter) performs better in-sample (out-of-sample), the differences in RMSPE are minor. The SVQ-M model is outperformed by the $A_1(4)$ model both in- and out-of-sample.

In Panel B of Table 8 displaying the results for PC2, it is noticed that V_t does not necessarily improve the in-sample prediction. For example, the in-sample RMSPE at $h = 32$ for the $A_1(4)$ model is 30.5 bps, which is larger than 27.8 bps for the $A_1(3)$ model with the constant instantaneous variance of PC2. Similarly, the in-sample RMSPEs for the SVQ-A model range from 29–33 bps, which are larger than the corresponding RMSPEs for the SVQ model with the gap increasing with h . The result, however, may be indirect evidence of the advantage of the level-dependence over V_t . In contrast to the in-sample prediction, V_t does improve the out-of-sample prediction. Specifically, the differences in the out-of-sample RMSPE between the $A_1(3)$ and $A_1(4)$ models reach up to 8.4 bps at $h = 4$. Next, it is also noticed that the nonlinear and multivariate level-dependence has incremental value over V_t especially out-of-sample. Specifically, while the differences in the in-sample RMSPE between the $A_1(4)$ and SVQ-A models are at most 1.3 bps at $h = 32$, those in the out-of-sample RMSPE are 3–4 bps. In fact, the SVQ-A model is the best performer out-of-sample.

Finally, the results for PC3 shown in Panel C of Table 8 have a similar pattern to the results for PC2. The introduction of V_t is effective out-of-sample but not in-sample, and the nonlinear and multivariate level-dependence contributes to additionally reducing the RMSPEs. Still, the combination of these two properties is not sufficient to resolve the difficulty of the diffusion model in predicting the volatility of PC3 by taking the in-sample performance of the HAR-RV regression into consideration.

Taken together, among the dynamic models considered in this study, the SVQ-A model seems desirable, which is capable of predicting the volatility of PC1 equally well to the $A_1(4)$ model and better at predicting the volatilities of PC2 and PC3 than the $A_1(4)$ model. There are even cases in which the SVQ-A model is comparable to the HAR-RV regression that is fitted directly to the realized volatility series. The result is not trivial as the SVQ-A model is supposed to capture various features of interest rate data, not limited to the volatility.

Nevertheless, the role of the nonlinear and multivariate level dependence is not remarkably well in predicting the volatility of PC1. As noted in Section 4.2.3, it has been

reported by the earlier work that it is difficult to predict the volatility of the level factor or long-term yields. This study has further shown that the originally small predictable component is difficult to be captured by the yield factors even with their nonlinear combinations. One explanation of this difficulty, which is more general than that presented in Section 4.3, is as follows. It is well-known that long-term yields are more affected by the risk premium than are short-term yields that are tied by monetary policy. And, it is found by Duffee (2011) that the risk premium is driven in part by hidden factors that are of little relevance to the current yield curve.

6 Concluding remarks

This study has predicted yield volatility using information on the current yield curve. The information is embedded in dynamic models of yield factors. Specifically, the covariance matrix of changes in the factors is made dependent on the factors with particular attention to nonlinear and multivariate dependence that is not fully studied by the earlier work. To maintain the positive definiteness of the covariance matrix at the same time, mainly two models are considered in which the eigenvalues of the covariance matrix are specified by quadratic (model SVQ) or exponential (model SVE) function. Although these models are approximations of more general level-dependent volatility models, they contribute to uncovering both usefulness and limitations of information content of the yield curve with respect to the volatility.

It is found that the quadratic level-dependent specification is comparable, or even superior, to the GARCH(1,1) specification in predicting the volatility of the second principal component or the slope factor of the yield curve. The finding is robust to, and in fact reinforced by, the introduction of a volatility-specific factor. The quadratic level-dependent specification is also confirmed to be useful for out-of-sample prediction of the volatility of the third principal component or the curvature factor when it is combined with the volatility-specific factor. Regarding prediction of the volatilities of the slope and curvature factors, therefore, information on the yield curve is useful.

In contrast, the nonlinear and multivariate level-dependent specification has a limited predictive power for the volatility of the first principal component or the level factor of the yield curve. Although it is better than the linear or nonlinear univariate level-dependent specification, it is outperformed by the GARCH specification particularly out-of-sample.

Besides, it does not bring further improvement of predictive accuracy once the volatility-specific factor is introduced.

This study has also found that the level factor is the least relevant to the volatility of the yield curve, which will make challenging the development of no-arbitrage term structure models that can explain both time-series and cross-sectional dimensions of interest rate data without conflict. But the above findings on information content of the yield curve will contribute to overcoming this challenge.

Acknowledgement

We are grateful to the anonymous referee for comments, which enabled us to significantly improve this manuscript.

Appendix A: An approximation method of conditional moments

A1. Outline of the method

The method is originally developed by Shoji (2002) and applied to the pricing of bonds by Takamizawa and Shoji (2009). 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, this study narrows focus on the case of $(n, d) = (2, 2)$, i.e., the conditional first and second moments of a two-dimensional diffusion process. As seen below, n can be considered as the order of approximation.

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

$$dx_{t,i} = f_i(X_t)dt + \xi_i(X_t)'dW_t \quad (i = 1, 2), \quad (34)$$

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 (34) to exist for an arbitrary X_0 .

Let $\Psi_{s,t}$ be a vector consisting of the first and second moments of an increment of 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 an ordinary differential equation.

By integrating equation (34) and taking the conditional expectation,

$$E_s[x_{t,i} - x_{s,i}] = E_s \left[\int_s^t f_i(X_u)du \right] . \quad (35)$$

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

$$\begin{aligned} f_i(X_u) &= f_i(X_s) \\ &+ f_i^{(1,0)}(X_s)(x_{u,1} - x_{s,1}) + f_i^{(0,1)}(X_s)(x_{u,2} - x_{s,2}) + \frac{1}{2}f_i^{(2,0)}(X_s)(x_{u,1} - x_{s,1})^2 \\ &+ \frac{1}{2}f_i^{(0,2)}(X_s)(x_{u,2} - x_{s,2})^2 + f_i^{(1,1)}(X_s)(x_{u,1} - x_{s,1})(x_{u,2} - x_{s,2}) + e_i , \end{aligned} \quad (36)$$

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 (36) into equation (35) and expressing the resulting equation in a vector form

$$E_s[x_{t,i} - x_{s,i}] = f_i(t - s) + \begin{pmatrix} f_i^{(1,0)} & f_i^{(0,1)} & \frac{1}{2}f_i^{(2,0)} & \frac{1}{2}f_i^{(0,2)} & f_i^{(1,1)} \end{pmatrix} \int_s^t \Psi_{s,u} du + R_i, \quad (37)$$

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,

$$E_s[(x_{t,1} - x_{s,1})^2] = E_s \left[\int_s^t \{2f_1(X_u)(x_{u,1} - x_{s,1}) + g_{11}(X_u)\} du \right], \quad (38)$$

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 (38) becomes

$$\begin{aligned} & 2f_1(X_u)(x_{u,1} - x_{s,1}) + g_{11}(X_u) \\ &= g_{11}(X_s) + \{2f_1(X_s) + g_{11}^{(1,0)}(X_s)\}(x_{u,1} - x_{s,1}) + g_{11}^{(0,1)}(X_s)(x_{u,2} - x_{s,2}) \\ &+ \{2f_1^{(1,0)}(X_s) + \frac{1}{2}g_{11}^{(2,0)}(X_s)\}(x_{u,1} - x_{s,1})^2 + \frac{1}{2}g_{11}^{(0,2)}(X_s)(x_{u,2} - x_{s,2})^2 \\ &+ \{2f_1^{(0,1)}(X_s) + g_{11}^{(1,1)}(X_s)\}(x_{u,1} - x_{s,1})(x_{u,2} - x_{s,2}) + e_{11}, \end{aligned} \quad (39)$$

where $g^{(k,l)}$ is defined analogously with $f^{(k,l)}$, and e_{11} is a residual term. By substituting equation (39) into equation (38),

$$\begin{aligned} E_s[(x_{t,1} - x_{s,1})^2] &= g_{11}(t - s) \\ &+ \begin{pmatrix} 2f_1 + g_{11}^{(1,0)} & g_{11}^{(0,1)} & 2f_1^{(1,0)} + \frac{1}{2}g_{11}^{(2,0)} & \frac{1}{2}g_{11}^{(0,2)} & 2f_1^{(0,1)} + g_{11}^{(1,1)} \end{pmatrix} \\ &\times \int_s^t \Psi_{s,u} du + R_{11}, \end{aligned} \quad (40)$$

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

$$\Psi_{s,t} = A(X_s) \int_s^t \Psi_{s,u} du + b(X_s)(t - s) + R, \quad (41)$$

where

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

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

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

Equation (41) can be solved as

$$\Psi_{s,t} = \int_s^t e^{A(X_s)(t-u)} b(X_s) du + \hat{R}. \quad (42)$$

If, in addition, A is invertible,

$$\Psi_{s,t} = A^{-1}(X_s) \{e^{A(X_s)(t-s)} - I\} b(X_s) + \hat{R}. \quad (43)$$

It is noted that equations (41)–(43) hold for any (n, d) with modification to $A(X_s)$ and $b(X_s)$. In general, $\Psi_{s,t}$ consists of $\binom{n+d}{n} - 1 = (n+d)!/(n!d!) - 1$ elements when up to n -th conditional moments for a d -dimensional diffusion process are computed. Correspondingly, up to n -th derivatives of f_i and g_{ij} ($i, j = 1, \dots, d$) are taken to compute the elements of $A(X_s)$. Omitting the residual vector, R or \hat{R} , leads to the approximation formula. According to Shoji (2002), both R and \hat{R} have order of $O((t-s)^{(n+3)/2})$. Thus, n can be considered as the order of approximation. In computing conditional first and second moments for the SVE model, $n = 2$ is actually considered.

It is also noted that R contains the conditional expectation 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. In other words, the conditional moments computed by the formula are exact. The SVQ model applies to this case. Even in this case, the use of this formula may be beneficial when the derivation of closed-form conditional moments is demanding.

A2. Accuracy to the conditional standard deviation under the SVE model

$\text{var}_t[x_{t+h\Delta, i}]$ for the SVE model is computed based on the first term on the right-hand side of equation (43). Naturally, the resulting value contains approximation error. Then, this appendix checks the accuracy to $\sqrt{\text{var}_t[X_{t+h\Delta}]/h\Delta}$ using the Monte Carlo (MC) method. The parameter values are fixed at the estimates presented in Panel B of Table 2. The starting values for the MC simulations, at which the accuracy is evaluated, are selected from the actual data. Specifically, three dates are selected from the whole sample when PC1 takes the minimum, median, or maximum value. The same is applied to PC2 and

PC3, which produces in total nine sets of observations. The accuracy is thus evaluated at not only usual but also unusual times. The subsequent realizations are generated from (3) with dt replaced by $\Delta/20$, an interval corresponding to 20 observations per week or 4 observations per day. The length of a path is up to 32 weeks. The number of repetition is set at 10,000 with the antithetic variate method.

Panels A and B of Table 10 present 32-week ahead forecasts of volatilities (annualized standard deviations in units of bps) of the first three PCs computed by the approximation and MC methods, respectively. Panel C presents percentage differences between the two methods, which range between -1.1% and 1.4% . The accuracy is maintained in the current setting.

Appendix B: Computation of the conditional variance under the GARCH(1,1) model

The variance of $x_{t+\Delta,i}$ conditioned on time t is simply $h_{t+\Delta,i}$, which is observed at time t . The variance of $x_{t+k\Delta,i}$ ($k = 2, \dots, h$) conditioned on time t is computed iteratively as follows. In equation (15), by substituting $t + (k - 1)\Delta$ for t and taking the variance conditioned on time t ,

$$\begin{aligned}
\text{var}_t[x_{t+k\Delta,i}] &= \text{var}_t[\alpha_i + \beta_i x_{t+(k-1)\Delta,i} + \sqrt{h_{t+k\Delta,i}} z_{t+k\Delta,i}] \\
&= \beta_i^2 \text{var}_t[x_{t+(k-1)\Delta,i}] + \text{var}_t[\sqrt{h_{t+k\Delta,i}} z_{t+k\Delta,i}] \\
&\quad + 2\beta_i \text{cov}_t[x_{t+(k-1)\Delta,i}, \sqrt{h_{t+k\Delta,i}} z_{t+k\Delta,i}] \\
&= \beta_i^2 \text{var}_t[x_{t+(k-1)\Delta,i}] + E_t[h_{t+k\Delta,i}] \quad (k = 2, \dots, h) . \quad (44)
\end{aligned}$$

On the other hand, in equation (16), by substituting $t + (k - 1)\Delta$ for t and taking the expectation conditioned on time t ,

$$\begin{aligned}
E_t[h_{t+k\Delta,i}] &= E_t[\omega_i + \phi_i h_{t+(k-1)\Delta,i} z_{t+(k-1)\Delta,i}^2 + \rho_i h_{t+(k-1)\Delta,i}] \\
&= \omega_i + (\phi_i + \rho_i) E_t[h_{t+(k-1)\Delta,i}] \quad (k = 2, \dots, h) . \quad (45)
\end{aligned}$$

Then, $\text{var}_t[x_{t+h\Delta,i}]$ is obtained by iteratively solving equations (44) and (45) starting from $\text{var}_t[x_{t+\Delta,i}] = h_{t+\Delta,i}$.

References

- Aït-Sahalia, Y., 1996, Testing continuous-time models of the spot interest rate, *Review of Financial Studies* 9, 385-426.
- 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.
- Andersen, T. G., and J. Lund, 1997a, Estimating continuous-time stochastic volatility models of the short-term interest rate, *Journal of Econometrics* 77, 343-377.
- Andersen, T. G., and J. Lund, 1997b, Stochastic volatility and mean drift in the short rate diffusion: Sources of steepness, level and curvature in the yield curve, Working paper.
- Ball, C. A., and W. N. Torous, 1999, The stochastic volatility of short-term interest rates: Some international evidence, *Journal of Finance* 54, 2339-2359.
- Bikbov, R., and M. Chernov, 2011, Yield curve and volatility: Lessons from Eurodollar futures and options, *Journal of Financial Econometrics* 9, 66-105.
- Brenner, R. J., R. H. Harjes, and K. F. Kroner, 1996, Another look at models of the short-term interest rate, *Journal of Financial and Quantitative Analysis* 31, 85-107.
- Brown, R. H., and S. M. Schaefer, 1994, Interest rate volatility and the shape of the term structure, *Philosophical Transactions: Physical Sciences and Engineering* 347, 563-576.
- Chan, K. C., G. A. Karolyi, F. A. Longstaff, and A. B. Sanders, 1992, An empirical comparison of alternative models of the short-term interest rate, *Journal of Finance* 47, 1209-1227.
- Christiansen, C., 2004, Regime switching in the yield curve, *Journal of Futures Markets* 24, 315-336.
- Christiansen, C., 2005, Multivariate term structure models with level and heteroskedasticity effects, *Journal of Banking and Finance* 29, 1037-1057.
- Christiansen, C., and J. Lund, 2005, Revisiting the shape of the yield curve: The effect of interest rate volatility, Working paper, University of Aarhus.
- 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, 2009, Can interest rate volatility be extracted from the cross section of bond yields? *Journal of Financial Economics* 94, 47-66.
- Cox, J. C., J. E. Ingersoll, Jr., and S. A. Ross, 1985, A theory of the term structure of interest rates, *Econometrica* 53, 385-407.
- Dai, Q., and K. J. Singleton, 2000, Specification analysis of affine term structure models, *Journal of Finance* 55, 1943-1978.
- Dai, Q., K. J. Singleton, and W. Yang, 2007, Regime shifts in a dynamic term structure model of U.S. Treasury bond yields, *Review of Financial Studies* 20, 1669-1706.
- 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.
- Duffee, G. R., 2011, Information in (and not in) the term structure, *Review of Financial Studies* 24, 2895-2934.
- Durham, G. B., 2003, Likelihood-based specification analysis of continuous-time models of the short-term interest rate, *Journal of Financial Economics* 70, 463-487.
- 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.
- Gallant, A. R., and G. Tauchen, 1998, Reprojecting partially observed systems with application to interest rate diffusions, *Journal of the American Statistical Association* 93, 10-24.
- Ghysels, E., P. Santa-Clara, and R. Valkanov, 2005, There is a risk-return trade-off after all, *Journal of Financial Economics* 76, 509-548.
- Ghysels, E., P. Santa-Clara, and R. Valkanov, 2006, Predicting volatility: Getting the most out of return data sampled at different frequencies, *Journal of Econometrics* 131, 59-95.
- Han, B., 2007, Stochastic volatilities and correlations of bond yields, *Journal of Finance* 62, 1491-1524.
- Hansen, P. R., and A. Lunde, 2005, A forecast comparison of volatility models: Does anything beat a GARCH(1,1)? *Journal of Applied Econometrics* 20, 873-889.

- Heidari, M., and L. Wu, 2003, Are interest rate derivatives spanned by the term structure of interest rates? *Journal of Fixed Income* 13, 75-86.
- Heidari, M., and L. Wu, 2009, A Joint framework for consistently pricing interest rates and interest rate derivatives, *Journal of Financial and Quantitative Analysis* 44, 517-550.
- 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.
- Joslin, S., 2010, Pricing and hedging volatility in fixed income markets, Working Paper, MIT Sloan School of Management.
- Li, H., and F. Zhao, 2006, Unspanned stochastic volatility: Evidence from hedging interest rate derivatives, *Journal of Finance* 61, 341-378.
- Litterman, R., and J. Scheinkman, 1991, Common factors affecting bond returns, *Journal of Fixed Income* 1, 54-61.
- Litterman, R., J. Scheinkman, and L. Weiss, 1991, Volatility and the yield curve, *Journal of Fixed Income* 1, 49-53.
- Longstaff, F. A., and E. S. Schwartz, 1992, Interest rate volatility and the term structure: A two-factor general equilibrium model, *Journal of Finance* 47, 1259-1282.
- 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.
- Newey, W., and K. West, 1987, A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix, *Econometrica* 55, 703-708.
- Pérignon, C., and D. R. Smith, 2007, Yield-factor volatility models, *Journal of Banking and Finance* 31, 3125-3144.
- Pérignon, C., and C. Villa, 2006, Sources of time variation in the covariance matrix of interest rates, *Journal of Business* 79, 1535-1549.
- Phoa, W., 1997, Can you drive market volatility forecasts from the observed yield curve convexity bias? *Journal of Fixed Income* 7, 43-54.
- Shoji, I., 2002, Approximation of conditional moments of diffusion processes, *International Journal of Computational and Numerical Analysis and Applications* 1, 163-190.

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.

Parameter / Index	$i = 1$	$i = 2$	$i = 3$
Panel A: $A_1(3)$ model			
$c_i \times 10^4$	–	0.456 (0.026)	0.133 (0.008)
$\sigma_i \times 10^2$	0.393 (0.022)	0.000	0.000
Panel B: CEV model			
$c_i \times 10^4$	–	0.456 (0.026)	0.133 (0.008)
$\sigma_i \times 10^2$	0.014 (0.007)	0.000	0.000
γ_i	–0.618 (0.233)	0.000	0.000

Table 1: Parameter estimates (standard errors) in the covariance matrix for the $A_1(3)$ and CEV models

The covariance matrix of changes in the first three PCs, Σ_t , is assumed to be diagonal with the i -th diagonal element $\Sigma_{t,i}$ given as

$$\begin{aligned}
 A_1(3) \quad \Sigma_{t,1} &= \sigma_1 x_{1,t}, & \Sigma_{t,i} &= \sigma_i x_{1,t} + c_i \quad (i = 2, 3), \\
 \text{CEV} \quad \Sigma_{t,1} &= \sigma_1 x_{1,t}^{\gamma_1}, & \Sigma_{t,i} &= \sigma_i x_{1,t}^{\gamma_i} + c_i \quad (i = 2, 3),
 \end{aligned}$$

where $x_{t,1}$ corresponds to PC1. Insignificant parameters at the first round of estimation are set to zero and then the remaining parameters are re-estimated to keep the models parsimonious. In-sample data from January 4, 1991 to April 9, 2003 are used for estimation.

Parameter / Index	$i = 1$		$i = 2$		$i = 3$	
Panel A: SVQ model						
c_i	1e-8		1e-8		1e-8	
m_1^i	0.000		0.000		0.000	
m_2^i	0.000		0.036	(0.009)	0.006	(0.002)
m_3^i	1.598	(0.108)	2.933	(0.569)	0.990	(0.179)
$\sin \varphi_1^{Q^i}$	0.000		-0.202	(0.023)	-0.259	(0.037)
$\sin \varphi_2^{Q^i}$	0.000		-0.129	(0.002)	-0.129	(0.002)
$\sin \varphi_3^{Q^i}$	0.295	(0.034)	-0.154	(0.027)	-0.112	(0.023)
Panel B: SVE model						
$s_{i0} \times 10^{-2}$	-0.088	(0.005)	-0.114	(0.004)	-0.136	(0.004)
$s_{i1} \times 10^{-2}$	0.000		0.000		0.000	
$s_{i2} \times 10^{-2}$	0.296	(0.049)	0.187	(0.051)	0.241	(0.046)
$s_{i3} \times 10^{-2}$	-0.617	(0.245)	-0.731	(0.224)	-1.255	(0.235)

Table 2: Parameter estimates (standard errors) in the covariance matrix for the SVQ and SVE models

The covariance matrix of changes in X_t , the first three PCs, is decomposed as $\Sigma_t = PL_tP'$, where L_t is the diagonal eigenvalue matrix and P is the orthonormal eigenvector matrix given in equation (5). Actually, $P = I$ because none of the parameters in P are significant. The i -th diagonal elements of L_t , $L_i(X_t)$, for the SVQ and SVE models are given as

$$\begin{aligned}
\text{SVQ} \quad L_i(X_t) &= c_i + X_t' \Gamma^i X_t & (i = 1, 2, 3), \\
\text{SVE} \quad L_i(X_t) &= \exp\{s_{i0} + s_i' X_t\} & (i = 1, 2, 3),
\end{aligned}$$

where Γ^i is given in equations (7)–(9). In-sample data from January 4, 1991 to April 9, 2003 are used for estimation.

Horizon, h	In-sample				Out-of-sample			
	4	8	16	32	4	8	16	32
Panel A: PC1								
$A_0(3)$	68.6**	59.4*	51.4	42.3	87.5**	82.0**	78.5*	74.3
$A_1(3)$	81.1**	72.7**	64.9**	55.0*	103.6**	99.7**	95.9**	88.3
CEV	67.0*	57.5	49.2	39.9	89.4**	82.7**	79.2**	77.0*
SVQ	67.2	57.8	49.9	42.9	80.0*	74.9	72.3	70.0
SVE	64.2	54.5	<i>46.5</i>	39.5	79.3*	74.5	72.2	71.2
GARCH	<i>61.8</i>	<i>53.2</i>	46.9	<i>38.4</i>	<i>63.5</i>	56.5	56.4	55.4
HAR-RV	56.2*	48.7	43.1	36.9	62.8	<i>58.2</i>	<i>59.3</i>	<i>60.9</i>
Panel B: PC2								
$A_0(3)$	36.3**	33.8*	31.3	28.1	38.6*	36.3*	34.3*	31.6
$A_1(3)$	36.3**	33.8*	31.2	27.8	38.6*	36.4*	34.4*	31.8
CEV	36.3**	33.8*	31.3	28.1	38.6*	36.3*	34.1*	31.5
SVQ	<i>33.1</i>	<i>30.4</i>	<i>28.1</i>	<i>25.1</i>	<i>31.0</i>	<i>28.6</i>	29.0	30.1
SVE	34.3	31.7	29.2	26.2	36.3	34.6	33.0	30.2
GARCH	33.9	31.3	29.0	26.3	32.1	30.5	<i>28.8</i>	<i>26.2</i>
HAR-RV	28.5**	25.6**	22.9*	20.5	28.4**	27.2**	25.4*	24.7
Panel C: PC3								
$A_0(3)$	27.2**	27.6**	29.6**	33.1**	22.9**	22.7*	23.7	25.1
$A_1(3)$	27.2**	27.6**	29.6**	33.0**	22.9**	22.7*	23.7*	25.3
CEV	27.2**	27.6**	29.6**	33.1**	22.9**	22.7*	23.6	24.9
SVQ	26.3*	26.5*	28.4**	31.7**	18.8	17.1	17.9	20.8
SVE	26.7*	27.3**	29.7**	33.5**	23.5*	23.9*	25.5*	26.8
GARCH	<i>24.7</i>	<i>24.5</i>	<i>25.3</i>	<i>27.9</i>	<i>17.0</i>	<i>16.5</i>	<i>17.7</i>	<i>20.4</i>
HAR-RV	19.1**	17.7**	15.3**	12.9**	16.0	15.5	15.5	15.6

Table 3: RMSPEs without forecasting regression

Root mean squared prediction errors (RMSPEs) for h -week ahead annualized standard deviations of the first three principal components are presented in basis points, where the prediction error is the residual of equation (19) (without the forecasting regression). The smallest and second smallest numbers in each column are displayed in bold and italic, respectively. * and ** indicate that the Diebold and Mariano (DM) (1995) test rejects the null hypothesis of equal predictive accuracy between a model in each row and the GARCH model with the 5% and 1% significance levels, respectively. The in-sample period is from January 4, 1991 to April 9, 2003 and the out-of-sample period is from April 16, 2003 to May 27, 2009. Throughout the out-of-sample period the parameter values are fixed at the in-sample estimates.

Horizon, h	In-sample				Out-of-sample			
	4	8	16	32	4	8	16	32
Panel A: PC1								
$A_1(3)$	61.4**	52.5*	45.1	37.3	79.7**	74.0*	71.7	70.7
CEV	61.6**	52.7*	45.3	37.1	80.5**	74.7*	72.5	71.4
SVQ	58.9	49.6	<i>42.1</i>	<i>34.8</i>	78.6	72.4	69.2	66.6
SVE	58.4	49.1	41.5	34.5	80.3*	74.8	71.9	69.8
GARCH	<i>56.4</i>	48.2	42.7	36.0	<i>64.0</i>	<i>57.6</i>	<i>58.1</i>	52.1
HAR-RV	56.2	<i>48.7</i>	43.1	36.9	60.8	54.5	54.0	<i>52.2</i>
Panel B: PC2								
$A_1(3)$	35.7*	32.7	29.6	26.4	37.5*	35.6*	34.2*	32.4
CEV	35.7*	32.7	29.6	26.4	37.6*	35.8*	34.7*	33.0
SVQ	<i>32.2</i>	<i>29.1</i>	<i>26.6</i>	<i>24.0</i>	<i>31.0</i>	<i>29.5</i>	<i>30.2</i>	31.7
SVE	33.2	30.0	27.1	24.5	36.7	35.1	33.9	32.9
GARCH	33.5	30.7	28.0	25.0	32.1	31.0	30.3	<i>28.9</i>
HAR-RV	28.5**	25.6**	22.9*	20.5	29.1*	28.0*	26.5*	26.1
Panel C: PC3								
$A_1(3)$	22.8*	20.6	18.1	15.4	23.1*	22.3	21.7	20.9
CEV	22.9*	20.7	18.1	15.4	23.4*	22.6*	22.1	21.5
SVQ	21.2	<i>18.7</i>	16.2	14.0	21.4*	20.6	21.0	22.4*
SVE	21.5	19.0	<i>16.0</i>	13.7	21.4	20.4	20.2	20.8
GARCH	<i>20.9</i>	18.9	16.3	<i>13.5</i>	<i>18.4</i>	<i>18.1</i>	<i>19.0</i>	<i>18.2</i>
HAR-RV	19.1**	17.7	15.3	12.9	16.1*	15.6*	15.7*	15.6

Table 4: RMSPEs with forecasting regression

Root mean squared prediction errors (RMSPEs) for h -week ahead annualized standard deviations of the first three principal components are presented in basis points, where the prediction error is the residual of equation (20) (with the forecasting regression). The smallest and second smallest numbers in each column are displayed in bold and italic, respectively. * and ** indicate that the DM test rejects the null hypothesis of equal predictive accuracy between a model in each row and the GARCH model with the 5% and 1% significance levels, respectively. The in-sample period is from January 4, 1991 to April 9, 2003 and the out-of-sample period is from April 16, 2003 to May 27, 2009. Throughout the out-of-sample period the model parameters are fixed at the in-sample estimates whereas the regression parameters in equations (17) and (20) are re-estimated in a rolling window fashion.

		Mean ($\times 10^4$)	S.D. ($\times 10^4$)	AR(1)	AR(4)	AR(8)	AR(16)	AR(32)
$L_1(X_t)$	$m_3^1(q_3^1 X_t)^2$	4.509	2.266	0.989	0.961	0.930	0.886	0.829
$L_2(X_t)$	$m_2^2(q_2^2 X_t)^2$	0.290	0.202	0.993	0.971	0.940	0.888	0.766
	$m_3^2(q_3^2 X_t)^2$	0.256	0.452	0.919	0.712	0.540	0.355	0.329
$L_3(X_t)$	$m_2^3(q_2^3 X_t)^2$	0.065	0.038	0.992	0.969	0.937	0.882	0.767
	$m_3^3(q_3^3 X_t)^2$	0.094	0.160	0.926	0.745	0.577	0.383	0.374

Table 5: Sample statistics of the components of instantaneous variances of PC1–3 for the SVQ model

The instantaneous variances $L_i(X_t)$ ($i = 1, 2, 3$) are decomposed as presented in equations (22)–(24). S.D. is standard deviation and $AR(k)$ is autocorrelation with lag k . The sample period is from January 4, 1991 to May 27, 2009 (the whole sample).

Horizon, h	In-sample				Out-of-sample			
	4	8	16	32	4	8	16	32
Panel A: PC1								
M1	56.6	47.2	39.5	31.4	73.5	66.7	65.5	64.6
M2	56.3	47.0	39.5	31.3	76.3	67.3	64.9	64.7
M3	56.4	47.1	39.4	31.3	72.4*	65.6	65.0	64.2
M4	56.6	47.2	39.5	31.4	73.1	66.5	65.4	64.4
Panel B: PC2								
M1	31.4	28.2	25.7	23.2	30.6	28.8	28.2	27.1
M2	31.2	28.0	25.5	23.2	31.7	30.3	28.9	27.3
M3	30.9*	27.5*	25.1*	22.9	32.6**	31.2*	30.8*	29.1
M4	31.3	28.2	25.7	23.2	31.2	28.8	28.2	27.1
Panel C: PC3								
M1	20.1	17.6	15.0	12.5	21.4	20.3	19.1	18.0
M2	20.0	17.5	14.9	12.5	22.1*	20.8	19.4	18.0
M3	19.8	17.5	15.0	12.5	20.1	19.5	18.7	18.3
M4	20.0	17.5	15.0	12.5	22.1*	20.5	19.2	17.9

Table 6: RMSPEs for regression models based on the quadratic multivariate specification

Root mean squared prediction errors (RMSPEs) for h -week ahead annualized standard deviations of the first three principal components are presented in basis points. * and ** indicate that the DM test rejects the null hypothesis of equal predictive accuracy between M1 presented in equation (25) and M_i ($i = 2, 3, 4$) presented in equations (26)–(28) with the 5% and 1% significance levels, respectively. The in-sample period is from January 4, 1991 to April 9, 2003 and the out-of-sample period is from April 16, 2003 to May 27, 2009. Throughout the out-of-sample period the parameter values are fixed at the in-sample estimates.

Parameter / Index	$i = 1$		$i = 2$		$i = 3$	
Panel A: $A_1(4)$ model						
$c_i \times 10^4$	1.199	(0.417)	0.000		0.000	
$\sigma_i \times 10^4$	2.992	(0.711)	0.346	(0.061)	0.107	(0.019)
Panel B: SVQ-M model						
c_i	1e-8		1e-8		1e-8	
m_1^i	0.000		0.000		0.000	
m_2^i	0.000		0.022	(0.009)	0.005	(0.002)
m_3^i	1.567	(0.228)	2.155	(0.624)	0.758	(0.220)
$\sin \varphi_1^{Q^i}$	0.000		-0.242	(0.043)	-0.288	(0.062)
$\sin \varphi_2^{Q^i}$	0.000		-0.126	(0.003)	-0.126	(0.003)
$\sin \varphi_3^{Q^i}$	0.317	(0.047)	-0.121	(0.031)	-0.090	(0.036)
Panel C: SVQ-A model						
$\sigma_i \times 10^4$	2.118	(0.730)	0.274	(0.073)	0.072	(0.019)
m_1^i	0.000		0.000		0.000	
m_2^i	0.000		0.000		0.000	
m_3^i	0.717	(0.174)	1.025	(0.426)	0.658	(0.194)
$\sin \varphi_1^{Q^i}$	0.000		0.000		0.000	
$\sin \varphi_2^{Q^i}$	0.000		-0.128	(0.004)	-0.129	(0.003)
$\sin \varphi_3^{Q^i}$	0.355	(0.073)	-0.114	(0.037)	-0.083	(0.044)

Table 7: Parameter estimates (standard errors) in the covariance matrix for the $A_1(4)$, SVQ-M, and SVQ-A models

For the $A_1(4)$ model, the covariance matrix Σ_t is assume to be diagonal with the i -th diagonal element $\Sigma_{t,i}$ given as

$$A_1(4) \quad \Sigma_{t,i} = \sigma_i V_t + c_i \quad (i = 1, 2, 3), \quad \Sigma_{t,4} = \sigma_4 V_t .$$

Likewise, the diagonal elements of Σ_t for the SVQ-M and SVQ-A models are specified as

$$\begin{aligned} \text{SVQ-M} \quad \Sigma_{t,i} &= X_t' \Gamma^i X_t V_t + c_i & (i = 1, 2, 3), \\ \text{SVQ-A} \quad \Sigma_{t,i} &= \sigma_i V_t + X_t' \Gamma^i X_t & (i = 1, 2, 3), \end{aligned}$$

and $\Sigma_{t,4}$ is the same as in $A_1(4)$. The estimate (standard error) of σ_4 is as follows: $A_1(4)$: 10.5 (5.33); SVQ-M: 18.8 (11.5); SVQ-A: 12.7 (9.10). In-sample data from January 4, 1991 to April 9, 2003 are used for estimation.

Horizon, h	In-sample				Out-of-sample			
	4	8	16	32	4	8	16	32
Panel A: PC1								
$A_1(4)$	63.5	53.2	44.1	<i>35.9</i>	66.6	61.1	62.7	64.0
SVQ-M	68.0*	57.7	49.4	43.0	74.3	70.6	71.3	72.0
SVQ-A	62.4	<i>51.9</i>	<i>42.7</i>	34.7	67.0	62.1	63.2	65.5
GARCH	<i>61.8</i>	53.2	46.9	38.4	<i>63.5</i>	56.5	56.4	55.4
HAR-RV	56.2*	48.7	43.1	36.9	62.8	<i>58.2</i>	<i>59.3</i>	<i>60.9</i>
Panel B: PC2								
$A_1(4)$	33.7	32.3	31.6*	30.5*	30.2	29.1	29.0	28.3
SVQ-M	<i>32.9</i>	<i>30.9</i>	29.4	26.6	28.7*	<i>26.9*</i>	27.5	27.9
SVQ-A	33.4	31.9	31.1*	29.2	27.4**	25.4**	24.9	<i>24.9</i>
GARCH	33.9	31.3	<i>29.0</i>	<i>26.3</i>	32.1	30.5	28.8	26.2
HAR-RV	28.5**	25.6**	22.9*	20.5	<i>28.4**</i>	27.2	<i>25.4*</i>	24.7
Panel C: PC3								
$A_1(4)$	26.3**	27.0**	29.5**	33.3**	18.1	18.9	21.7	24.8
SVQ-M	25.6	25.7	27.1*	29.9*	17.7	16.1	16.7	<i>18.5</i>
SVQ-A	25.9*	26.2*	28.2**	31.7**	<i>16.8</i>	15.3	<i>16.3</i>	19.4
GARCH	<i>24.7</i>	<i>24.5</i>	<i>25.3</i>	<i>27.9</i>	17.0	16.5	17.7	20.4
HAR-RV	19.1**	17.7**	15.3**	12.9**	16.0	<i>15.5</i>	15.5	15.6

Table 8: RMSPEs without forecasting regression

The same legend as in Table 3 applies.

	$A_0(3)$	$A_1(3)$	CEV	SVQ	SVE
$K_{0,1}$	0.000	0.000	0.000	0.000	0.000
$K_{1,11}$	-0.078 (0.045)	-0.076 (0.048)	-0.081 (0.043)	-0.066 (0.041)	-0.071 (0.041)
$K_{1,12}$	0.000	0.000	0.000	0.000	0.000
$K_{1,13}$	0.000	0.000	0.000	0.000	0.000
$K_{0,2}$	0.000	0.000	0.000	0.000	0.000
$K_{1,21}$	-0.498 (0.120)	-0.498 (0.120)	-0.498 (0.120)	-0.524 (0.138)	-0.539 (0.120)
$K_{1,22}$	-0.682 (0.219)	-0.682 (0.219)	-0.682 (0.219)	-0.569 (0.235)	-0.670 (0.215)
$K_{1,23}$	-3.753 (0.874)	-3.753 (0.874)	-3.752 (0.874)	-3.964 (1.012)	-4.053 (0.879)
$K_{0,3}$	-0.025 (0.008)	-0.025 (0.008)	-0.025 (0.008)	-0.021 (0.008)	-0.026 (0.008)
$K_{1,31}$	-0.145 (0.047)	-0.145 (0.047)	-0.145 (0.047)	-0.129 (0.053)	-0.156 (0.047)
$K_{1,32}$	0.000	0.000	0.000	0.000	0.000
$K_{1,33}$	-2.464 (0.562)	-2.464 (0.562)	-2.464 (0.562)	-2.150 (0.600)	-2.614 (0.604)
	$A_1(4)$	SVQ-M	SVQ-A		
$K_{0,1}$	0.000	0.000	0.000	0.000	
$K_{1,11}$	-0.070 (0.040)	-0.085 (0.038)	-0.067 (0.039)		
$K_{1,12}$	0.000	0.000	0.000		
$K_{1,13}$	0.000	0.000	0.000		
$K_{0,2}$	0.000	0.000	0.000		
$K_{1,21}$	-0.311 (0.109)	-0.368 (0.116)	-0.340 (0.111)		
$K_{1,22}$	-0.408 (0.181)	-0.420 (0.189)	-0.401 (0.179)		
$K_{1,23}$	-2.182 (0.798)	-2.651 (0.859)	-2.398 (0.824)		
$K_{0,3}$	-0.021 (0.007)	-0.017 (0.007)	-0.019 (0.008)		
$K_{1,31}$	-0.106 (0.041)	-0.084 (0.047)	-0.088 (0.051)		
$K_{1,32}$	0.000	0.000	0.000		
$K_{1,33}$	-1.970 (0.517)	-1.605 (0.533)	-1.731 (0.555)		
k_v	4.748 (1.170)	12.93 (3.776)	4.415 (1.257)		

Table 9: Parameter estimates (standard errors) in the drift

Insignificant parameters at the first round of estimation are set to zero and then the remaining parameters are re-estimated to keep the models parsimonious. In-sample data from January 4, 1991 to April 9, 2003 are used for estimation.

States	Panel A: Forecasts by Approx.			Panel B: Forecasts by MC			Panel C: % differences		
	PC1	PC2	PC3	PC1	PC2	PC3	PC1	PC2	PC3
PC1	197.5	56.9	16.1	197.2	57.0	16.1	0.2	-0.3	-0.1
PC2	174.4	59.5	18.3	173.9	59.6	18.5	0.3	-0.1	-1.1
PC3	222.6	77.4	25.2	220.7	77.3	24.9	0.9	0.1	1.4
PC1	180.8	59.1	17.8	181.4	59.3	17.8	-0.3	-0.4	-0.2
PC2	197.1	60.7	18.0	196.4	60.8	18.1	0.3	-0.2	-0.8
PC3	234.5	70.5	21.3	236.2	71.2	21.2	-0.7	-1.0	0.5
PC1	222.6	77.4	25.2	222.7	77.3	24.9	-0.1	0.1	1.3
PC2	283.4	78.3	23.3	281.3	78.6	23.5	0.8	-0.4	-0.5
PC3	185.3	55.6	16.1	187.2	56.2	16.2	-1.0	-1.0	-1.1

Table 10: Accuracy to $\sqrt{\text{var}_t[\hat{X}_{t+h\Delta}]/h\Delta}$ with $h = 32$ for the SVE model

Panels A and B present forecasts of the 32-week ahead volatility (annualized standard deviation in basis points) for the SVE model computed by the approximation and MC methods, respectively. Panel C presents percentage differences between the two methods. The forecasts are made at nine states taken from the entire sample, where PC1-3 take the minimum, median, or maximum value. For example, (Minimum, PC1) corresponds to a set of PCs in a day when PC1 takes the sample minimum. Note that conditional standard deviations for the SVQ model can be computed exactly.

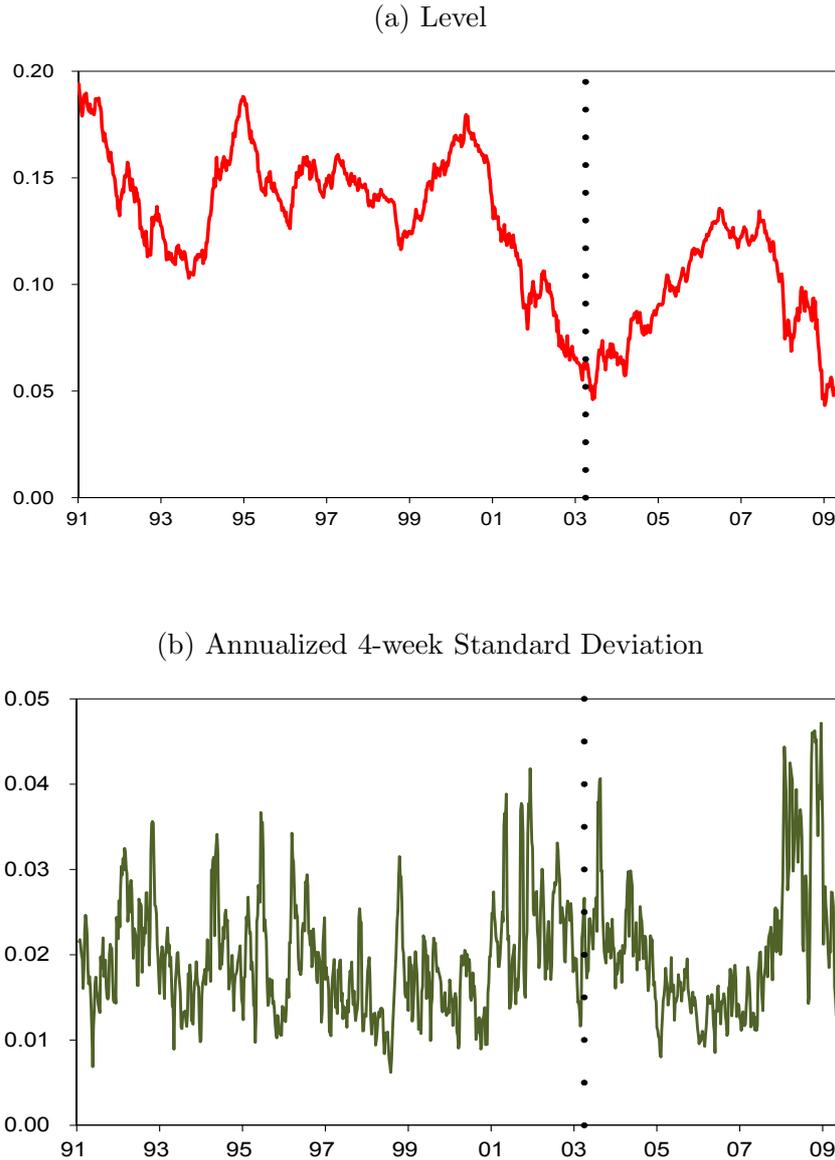


Figure 1: Time-series of the level and realized volatility of PC1 over 1991–2009

Panel (a) presents the time-series of the level of the first principle component (PC1). Panel (b) presents the time-series of the realized volatility (annualized 4-week standard deviation) of PC1, which is calculated by summing squared daily changes in PC1 over the four weeks, dividing the sum by $4/52$, and taking the square-root. The vertical dotted line separates the in-sample and out-of-sample periods.

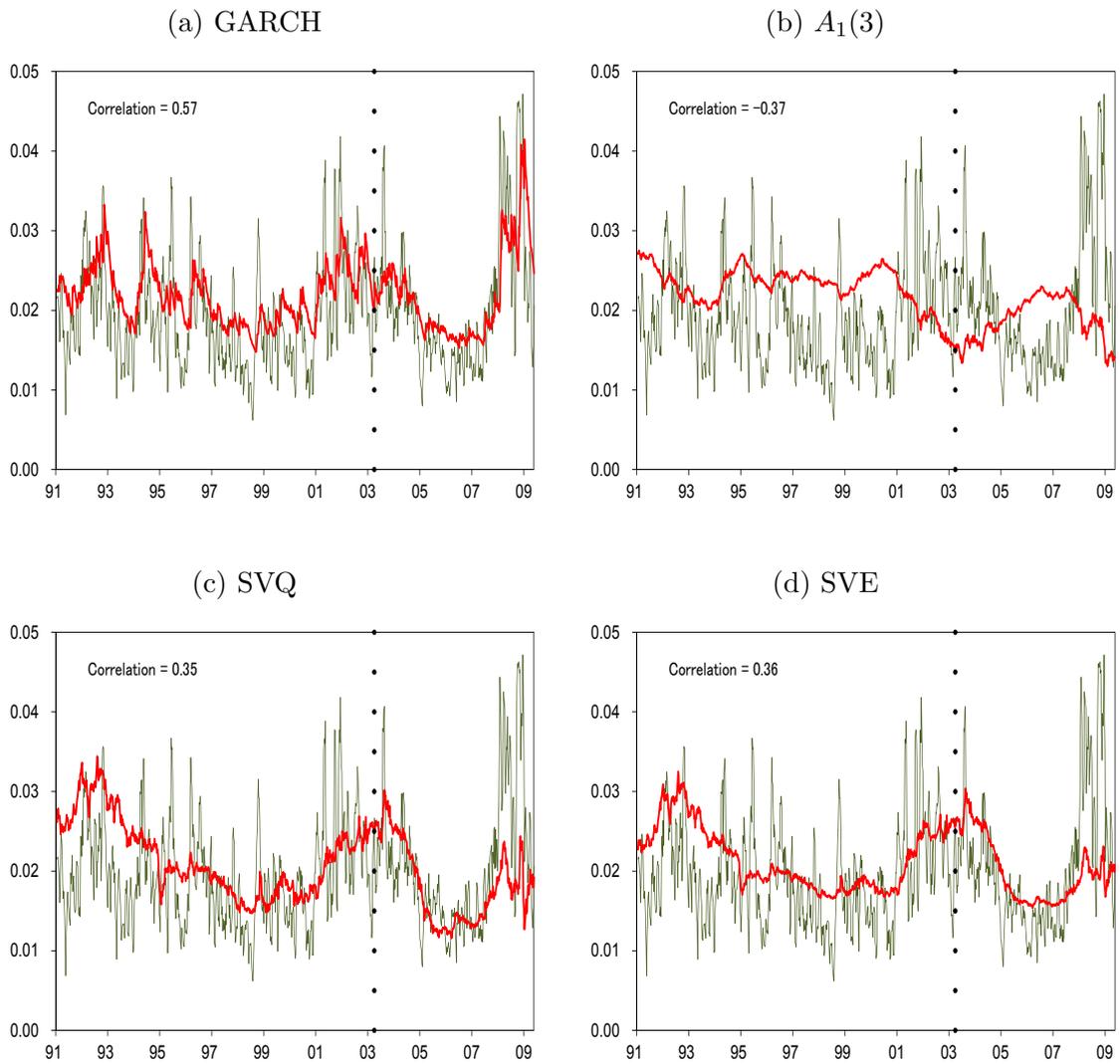


Figure 2: Time series of four-week ahead forecasts of the volatility (annualized standard deviation) of PC1

Model forecasts (thick line) and the corresponding realized values (thin line) are displayed, with the vertical dotted line separating the in-sample and out-of-sample periods.

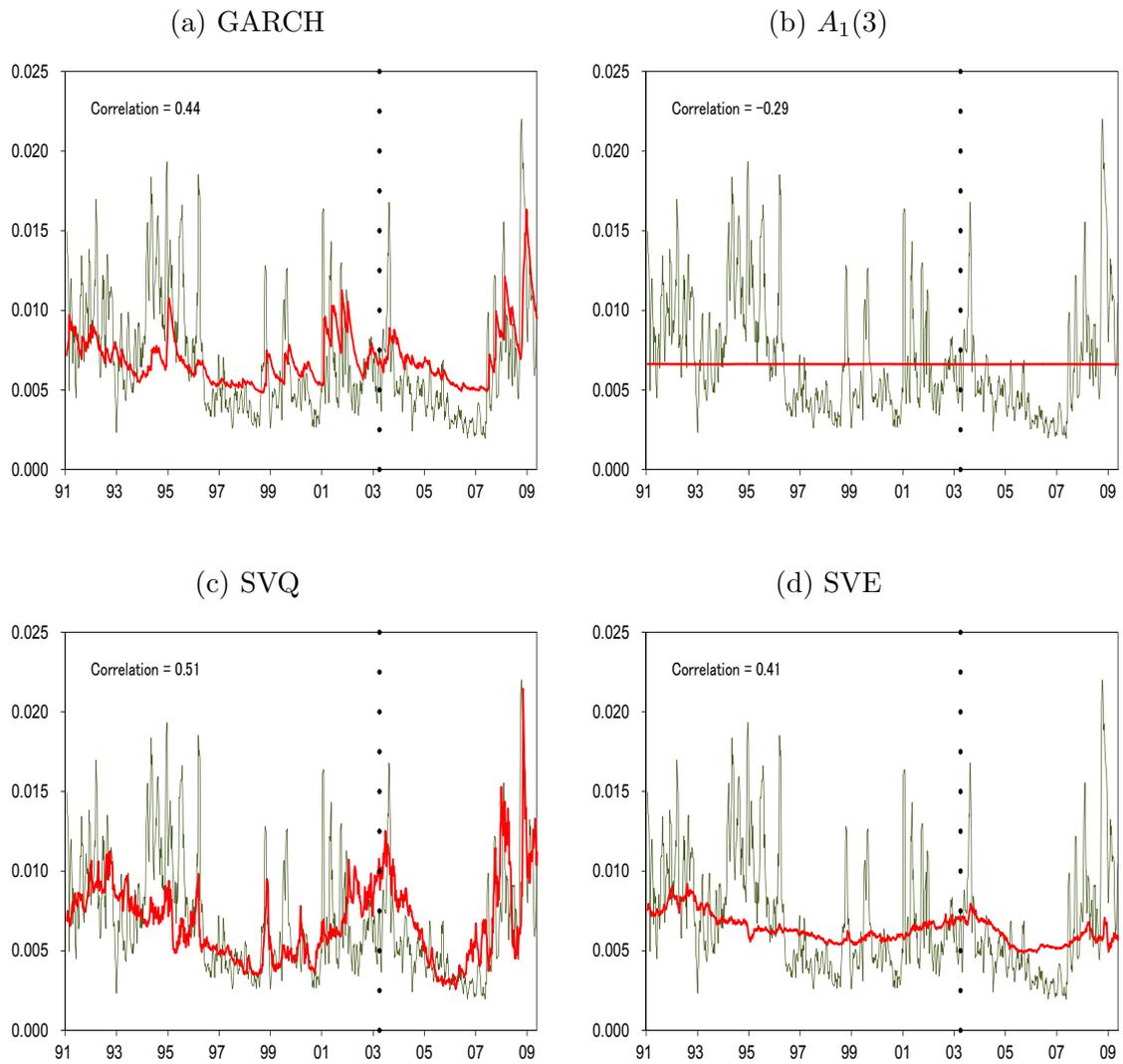


Figure 3: Time series of four-week ahead forecasts of the volatility (annualized standard deviation) of PC2

Model forecasts (thick line) and the corresponding realized values (thin line) are displayed, with the vertical dotted line separating the in-sample and out-of-sample periods.

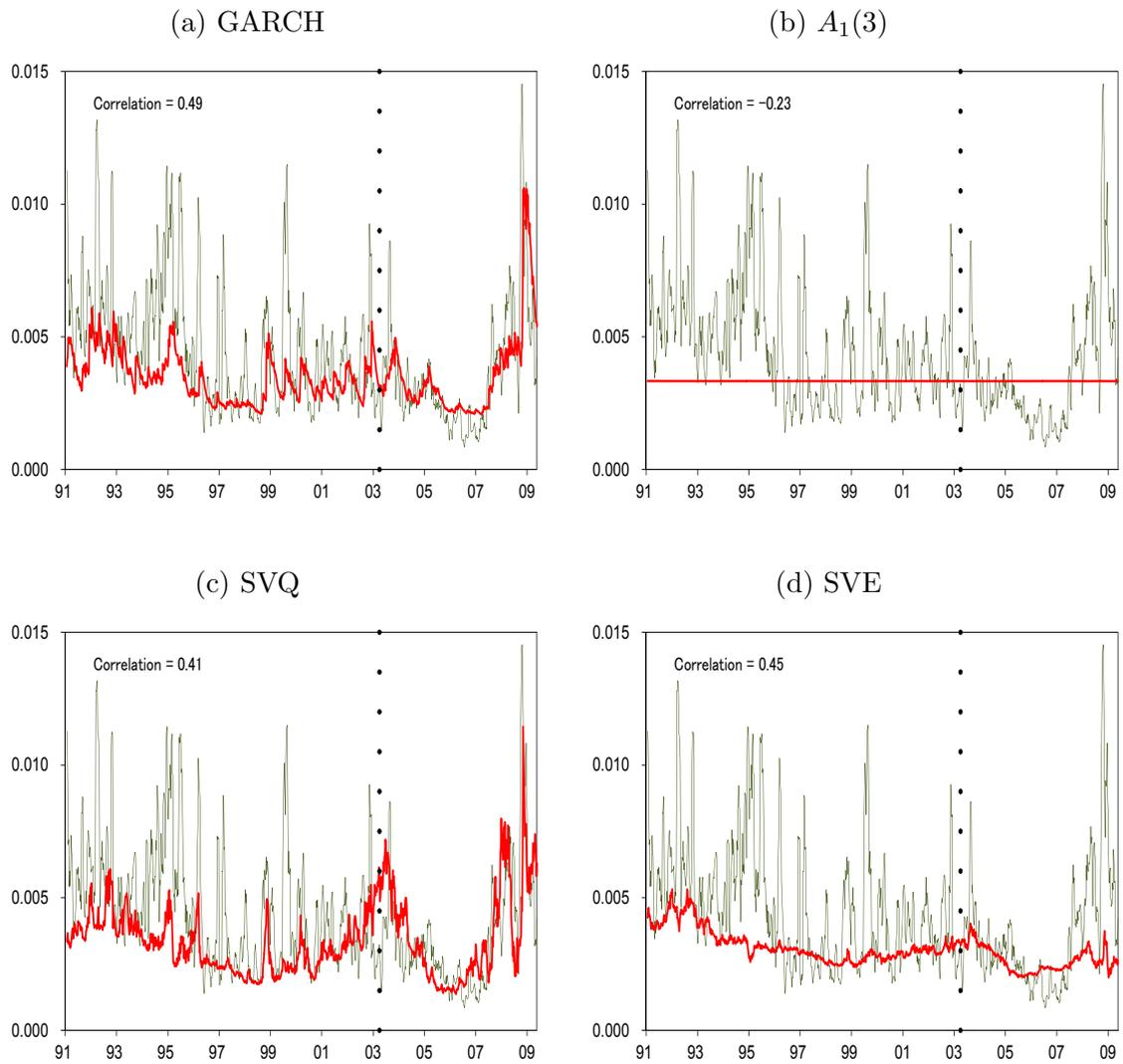


Figure 4: Time series of four-week ahead forecasts of the volatility (annualized standard deviation) of PC3

Model forecasts (thick line) and the corresponding realized values (thin line) are displayed, with the vertical dotted line separating the in-sample and out-of-sample periods.

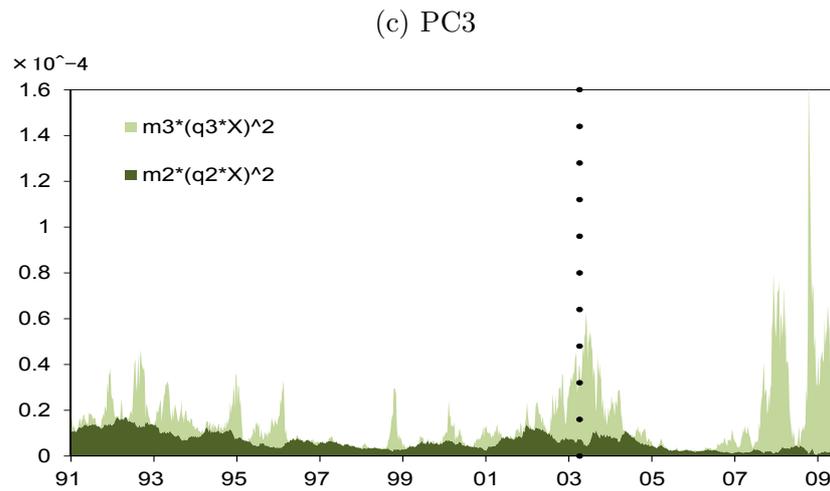
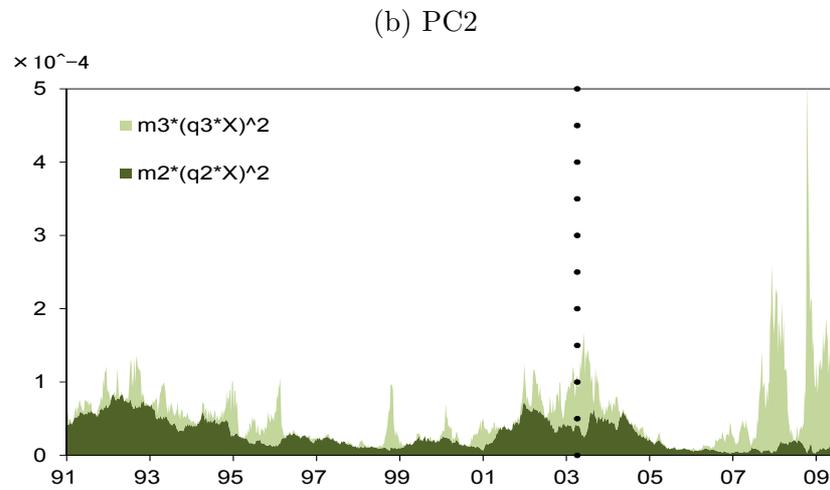
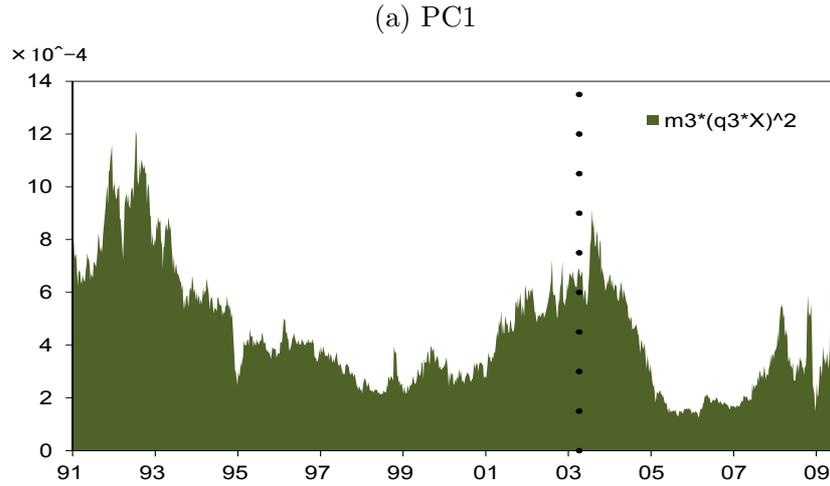


Figure 5: Time-series of instantaneous variances of PC1–3 for the SVQ model
 The instantaneous variances L_i ($i = 1, 2, 3$) are decomposed as presented in equations (22)–(24). Each figure is drawn such that the values of the components of L_i are added up to the value of L_i . The constant term in L_i (10^{-8}) is omitted. The vertical dotted line separates the in-sample and out-of-sample periods.

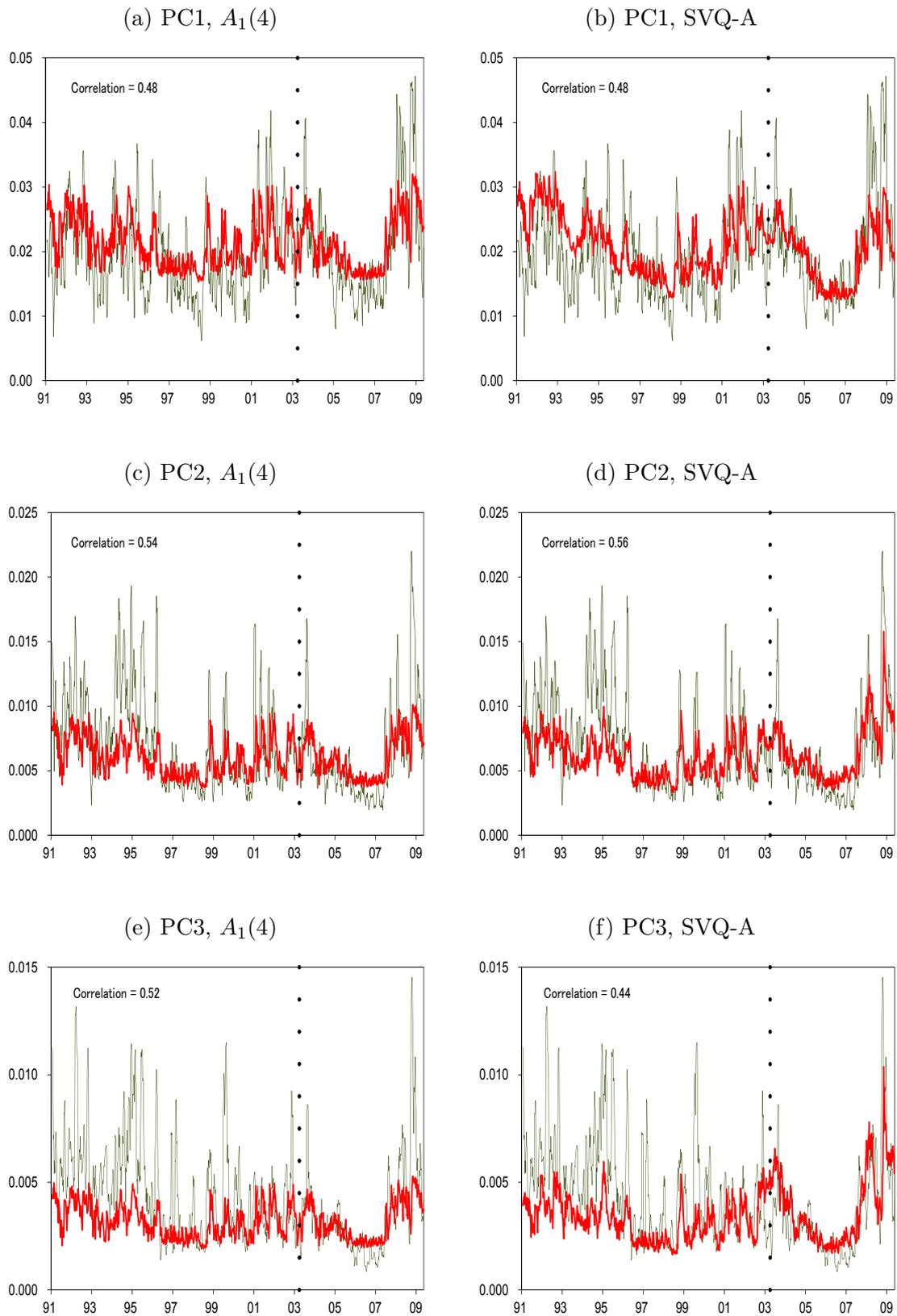


Figure 6: Time series of four-week ahead forecasts of the volatilities (annualized standard deviations) of the PC1–3 generated by the $A_1(4)$ model (the left panel) and the SVQ-A model (the right panel)