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: August 6, 2013

(First draft: February 2012)

Abstract

This study attempts to predict the volatility of yield curve factors using information on the cross-section of yields. Since a linear relationship between variances and levels of yields is reported to be inconsistent with recent data, we focus on nonlinear relationships. Further, rather than employing regression models, we employ dynamic models, where information on the yield curve is used for specifying factor covariance matrix as nonlinear functions of yield curve factors. Through such dynamic models, we uncover both usefulness and limitations of information content of the yield curve with respect to volatility prediction.

Keywords: Term structure, Level-dependence, Realized volatility, 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. This study attempts to predict the volatility by utilizing such information. This is done by modeling the dynamics of yield curve factors as diffusion processes, aiming at exploring appropriate underlying models for bonds and their derivatives.

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 (annualized standard deviation) of the first principle component (PC) constructed from U.S. dollar LIBOR and swap rates over 1991–2009: The details of these data are provided in Section 3. Note that the first PC is interpretable as a level factor of the yield curve. We notice that the sharp rise in the volatility of changes in the first PC, observed around 2001–03 and 2008–09, is actually accompanied by the fall in the level of the first PC.

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. Consequently, it is implied that the relationship between the volatility and the curvature of the yield curve up to ten years is not supported since 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 in time series. Due in part to these difficulties, the recent literature seems to prefer the Gaussian term structure model, a family of affine models with constant volatility, to the other affine models with stochastic volatility.

Is the yield curve really of little relevance to the volatility? There seems to be a room for further discussion. 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. For this purpose, we employ dynamic models, rather than regression models, aiming at incorporating both time-series and cross-sectional properties of interest-rate data into estimation and prediction. This approach will shed new light on adequate underlying models of the term structure of interest rates that have ability of predicting the volatility of yields while maintaining the goodness-of-fit to the cross-section of yields.

A clue to finding such dynamic models is to rely on existing models successful in many ways. The Gaussian model is a candidate with capability of describing and predicting the level of yields; see, e.g., Dai and Singleton (2003), Duffee (2002), and Joslin, Singleton, and Zhu (2011). Besides, since all factors in the Gaussian model are naturally interpreted as yield-curve specific, the misidentification of factors such that ex-ante and ex-post roles are different can be avoided. An obvious drawback of the model, on the other hand, is constant volatility despite mounting evidence supporting time-varying volatility.

We then overcome this drawback by making the covariance matrix of the factors level-dependent, leading to a model in which all factors potentially contribute to time-varying volatility. This is how the information on the cross-section of yields is used: It is used not for directly identifying volatility factors, but for specifying volatility functions of the yield curve factors. The level-dependent covariance matrix, however, cannot be introduced without caution. It must be positive definite. One simple way to meet this requirement is to model eigenvalues of the covariance matrix as functions taking positive values. Specif-

ically, we propose two models for the time-varying eigenvalues: one given by quadratic functions and the other by exponential functions. Naturally, the relationship between variances and levels of yields is nonlinear.

Our approach of examining information content of the yield curve with respect to the volatility is different from the earlier work. Bikbov and Chernov (2011), and Thompson (2008) change estimation methods while using the affine models. We change models themselves. Compared with dynamic models for the yield curve employed by, e.g., Christiansen (2005), and Pérignon and Smith (2007), the current models are more parsimonious in that GARCH or regime-switching effect is not taken into consideration. Rather, we broaden interpretation of the level effect and explore its adequacy, which we think deserves research attention for the following reasons. First, in developing dynamic models to term structure models by imposing no arbitrage conditions, less complicated models have advantages of lower computational cost. Second, since the Gaussian model becomes increasingly popular in the empirical term structure literature, it may be of interest to examine the degree to which volatility prediction improves by extending the Gaussian model while holding the number of factors fixed. Third, even though unspanned volatility factors as proposed by Collin-Dufresne and Goldstein (2002) may be necessary for fully capturing the behavior of volatility, it is important to know in advance where to introduce them, as increasing the number of factors is not costless. To ensure the effective introduction, it seems necessary to clarify what yield curve factors alone can do. The contribution of this study is to present both usefulness and limitations of information content of the yield curve with respect to volatility prediction through dynamic models.

Section 2 proposes models. Section 3 explains the data and how to construct a realized volatility measure. Section 4 examines the volatility forecasting performance of the proposed models. Section 5 provides concluding remarks. Appendices collect technical arguments including explanation of an approximation method of conditional moments used for estimation.

2   Model

The research objective is to examine whether information on the cross-section of yields is useful for predicting the volatility of yields. To achieve this objective, we use dynamic models and try nonlinear relationships between the yield curve and volatility. In Section

2.1, models appropriate for the objective are proposed. Strengths and weaknesses of the proposed models are then discussed in Section 2.2.

2.1 Specification

In modeling the dynamics of the yield curve, we consider three factors by following a tradition in the empirical term structure literature. Let X_t be a three dimensional state vector, the elements of which are all supposed to drive the yield curve. Then, the instantaneous change in X_t is assumed to follow

$$dX_t \sim N\left[(K_0 + K_1 X_t)dt, \ \Sigma_t dt\right] \ . \tag{1}$$

We incorporate information on the yield curve into the specification of the instantaneous covariance matrix, Σ_t. Specifically, we make Σ_t dependent on X_t. But such level-dependent specifications cannot be introduced without caution because Σ_t must be positive definite. To meet this condition, one simple approach is to take the spectral decomposition of Σ_t, and then to specify the eigenvalues as functions of X_t that take positive values for any X_t while holding the corresponding eigenvectors fixed:

$$\Sigma_t = P L_t P' \ , \tag{2}$$

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. The idea of the spectral decomposition to capture time-varying second moments is not new: It is employed by, e.g., Fan, Gupta, and Ritchken (2003), Han (2007), Jarrow, Li, and Zhao (2007), and Longstaff, Santa-Clara, and Schwartz (2001), and supported empirically by Pérignon and Villa (2006). Here, the idea is combined with level-dependent specifications.

Before specifying eigenvalue functions, we first parameterize P in equation (2). By the conditions of orthogonality and unit length, the number of free parameters in P is actually three. Then, it is parameterized using three rotation matrices:

$$P = \begin{pmatrix} 1 & 0 & 0 \\ 0 & \cos\varphi_3^P & -\sin\varphi_3^P \\ 0 & \sin\varphi_3^P & \cos\varphi_3^P \end{pmatrix} \begin{pmatrix} \cos\varphi_2^P & 0 & -\sin\varphi_2^P \\ 0 & 1 & 0 \\ \sin\varphi_2^P & 0 & \cos\varphi_2^P \end{pmatrix} \begin{pmatrix} \cos\varphi_1^P & -\sin\varphi_1^P & 0 \\ \sin\varphi_1^P & \cos\varphi_1^P & 0 \\ 0 & 0 & 1 \end{pmatrix} . \tag{3}$$

This rotation is called the yaw, pitch, and roll rotation. The parameters to be estimated are $\sin\varphi_i^P$ $(i = 1, 2, 3)$. For identification, we restrict $\varphi_i^P \in [-\pi/2, \pi/2]$, so that $\cos\varphi_i^P = \sqrt{1 - \sin^2\varphi_i^P}$.

L_t in equation (2) is expressed by the following form:

$$L_t = \begin{pmatrix} L_1(X_t) & 0 & 0 \\ 0 & L_2(X_t) & 0 \\ 0 & 0 & L_3(X_t) \end{pmatrix}. \tag{4}$$

To have $L_i(X_t) > 0$ $(i = 1, 2, 3)$ for any X_t, we propose two models.

The first model, abbreviated as SV-Q (Stochastic Volatility with Quadratic specification), specifies $L_i(X_t)$ as

$$L_i(X_t) = c_i + X_t'\Gamma^i X_t \quad (i = 1, 2, 3), \tag{5}$$

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, we impose the latter restriction 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:

$$\Gamma^i = Q^i M^i Q^{i\prime} \quad (i = 1, 2, 3), \tag{6}$$

where

$$M^i = \begin{pmatrix} m_1^i & 0 & 0 \\ 0 & m_2^i & 0 \\ 0 & 0 & m_3^i \end{pmatrix}, \qquad \text{with} \quad 0 \leq m_1^i \leq m_2^i \leq m_3^i, \tag{7}$$

and

$$Q^i = \begin{pmatrix} 1 & 0 & 0 \\ 0 & \cos\varphi_3^{Q^i} & -\sin\varphi_3^{Q^i} \\ 0 & \sin\varphi_3^{Q^i} & \cos\varphi_3^{Q^i} \end{pmatrix} \begin{pmatrix} \cos\varphi_2^{Q^i} & 0 & -\sin\varphi_2^{Q^i} \\ 0 & 1 & 0 \\ \sin\varphi_2^{Q^i} & 0 & \cos\varphi_2^{Q^i} \end{pmatrix} \begin{pmatrix} \cos\varphi_1^{Q^i} & -\sin\varphi_1^{Q^i} & 0 \\ \sin\varphi_1^{Q^i} & \cos\varphi_1^{Q^i} & 0 \\ 0 & 0 & 1 \end{pmatrix}, \tag{8}$$

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, we set $\sin\varphi_j^{Q^i} = 0$.

The second model, abbreviated as SV-E (Stochastic Volatility with Exponential specification), specifies $L_i(X_t)$ as

$$L_i(X_t) = \exp\left\{s_{i0} + s_i' X_t\right\} \quad (i = 1, 2, 3). \tag{9}$$

No parameter restriction is required for the SV-E model. This 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).

At first sight, the proposed models may appear odd. However, they can be thought of as a family of models with level-dependent volatility considered by many studies. The level dependence is more involved here to achieve the research objective, and the more involved specification is feasible due to the spectral decomposition of the covariance matrix, which is also employed by previous studies.

2.2 Strengths and weaknesses of the proposed models, and possible remedies for the weaknesses

Time-varying covariance matrix must be positive definite in the first place. The affine models satisfy this requirement by placing constraints on factor processes driving this matrix. These constraints, however, often result in reduction of the goodness-of-fit to the cross-section of yields or misidentification of volatility factors. On the other hand, the proposed models satisfy this requirement by placing constraints on eigenvalue functions. There is no sign constraint on factor processes here, as is the case for the original Gaussian model. Therefore, the modeling of the market prices of risk, which is required for deriving no-arbitrage bond prices, can also be made as flexibly as the Gaussian model. It is noted that this flexibility is one of the major strengths of the Gaussian model especially in terms of predicting the level of yields. In short, the proposed models inherit empirical strengths of the Gaussian model.

At the same time, the proposed models have obvious weaknesses. We mention three that seem to be particularly crucial in using them as underlying models for the term structure of interest rates. First, there is no closed-form expression for no-arbitrage bond prices. Second, negative interest rates cannot be avoided as for the original Gaussian model. Third, there is no unspanned factor that affects derivative prices but not bond prices in spite of the earlier work pointing out the existence and significance of such factors; see, e.g., Collin-Dufresne and Goldstein (2002), Han (2007), Heidari and Wu (2003, 2009), Jarrow et al. (2007), and Li and Zhao (2006).

There are possible remedies for these weaknesses, however. A remedy for the lack of closed-form expression is to rely on analytical approximations of no-arbitrage bond prices.

Takamizawa and Shoji (2009) propose one such method, the accuracy of which is sufficient as long as reasonable parameter and state variable values are used. The second weakness of generating negative interest rates can virtually be avoided. As originally proposed by Aït-Sahalia (1996), the key is to add to the risk-neutral drift of the instantaneous risk-free rate process, r_t, a term such that it increases sufficiently rapidly as r_t approaches zero. If the coefficient of this term is set to a very small number, the impact of this term is negligible at, e.g., $r_t = 0.0001$, a level lower than the minimum in our sample. Hence, estimation and prediction results using real data do not change between with and without it. At least from an empirical point of view, we can work as if it is added though it is actually not. Resolving the third weakness of not accommodating unspanned factors may not be as difficult as it appears. As suggested by Joslin (2010) using affine models and Takamizawa (2011) using non-affine models, a factor that is nearly, if not completely, unspanned can be introduced by adding it to the covariance matrix but not to the risk-neutral drift vector.

3   Data and realized variance measure

3.1   Dataset

We use 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 (641 observations), and the out-of-sample data for prediction contain 320 observations. This division allows for incorporating information on the lowest range of interest rates into model estimation as well as reserving sufficient out-of-sample observations. Other divisions are also tried, and the differences in the results are not large enough to change the conclusions of this study.

Reasons for selecting this dataset are as follows. First, we can make more challenging the purpose of predicting the volatility using information on the cross-section of yields. As shown in Figure 1, a simple level-volatility relationship disappears in the recent data.

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 the history of U.S. interest rates.

3.2 Factor identification

As the yield curve factors, we choose the first three PCs, which are interpretable as level, slope, and curvature factors of the yield curve, respectively. 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 over the entire period to make the PCs identical between the in-sample and out-of-sample periods.

This choice of factors has advantages relative to others. First, it allows for focusing on the volatilities of PCs without paying much attention to the covariances between PCs. Second, it allows for a more intuitive understanding of the results. Specifically, since P in equation (2) is expected to be nearly the identity matrix, as is empirically the case shown in Section 4.1, $L_i(X_t)$ corresponds to the instantaneous variance of the i-th PC, $x_{t,i}$. Still, a nonlinear relationship between the yield curve and the volatility can be detected based on the PCs. Since the PCs are obtained by linear combinations of yields, if the volatility of PCs is nonlinear in the level of PCs, the volatility of yields is also nonlinear in the level of yields.

3.3 Realized measure

Using daily data, we construct a realized measure of the conditional variance of changes in PCs. We first transform the daily series of zero yields into those of PCs. This is done using the same rotation matrix as calculated with the weekly data to make the PCs identical between the weekly and daily data. Then, a realized measure of the one-week ahead conditional variance of $x_{t,i}$ is computed as

$$RV_{t,t+\Delta,i} = \sum_{k=1}^{m_{t+\Delta}} (x_{t+\frac{\Delta}{m_{t+\Delta}}k,i} - x_{t+\frac{\Delta}{m_{t+\Delta}}(k-1),i})^2 \, , \tag{10}$$

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

$$RV_{t,t+h\Delta,i} = \sum_{j=1}^{h} RV_{t,t+j\Delta,i} \ . \tag{11}$$

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, i.e., $h = \{4, 8, 16, 32\}$, which are longer than those considered in the earlier work. This is aimed at matching the horizon of the volatility forecast with that of the level forecast including the test for the efficient hypothesis of the term structure, which is usually longer than a month; see, e.g., Bekaert and Hodrick (2001), Campbell and Shiller (1991), and Fama and Bliss (1987).

4 Empirical Analysis

4.1 In-sample estimation

The proposed models are estimated by the quasi-maximum likelihood method, where the conditional first and second moments of the factors are substituted into the multivariate normal density function. This method might be justified by a relatively short interval, $\Delta = 1/52$. The conditional moments to be substituted are computed using a method proposed by Shoji (2002), which approximates a vector of conditional moments as the solution to an ordinary differential equation: Appendix A provides a brief explanation of this method. Note that the conditional moments for the SV-Q model can be computed exactly, as the drift vector is linear and the covariance matrix is quadratic in X_t, respectively.

Fully-parameterized models are first estimated, but estimates of some parameters are not significant. We set such parameters to zero, and re-estimate the models to obtain t-values all exceeding 1.5 in absolute value. In addition, parameters in the SV-Q model that do not satisfy the sign constraints are fixed as follows: $c_i = 10^{-8}$ and $m_i^j = 0$.

Table 1 presents parameter estimates of the covariance matrix, Σ_t. Since our primary interest is in the volatility, parameter estimates of the drift vector are omitted for saving space. First, we report the estimates for the SV-Q model provided in Panel A. All parameters in P, $\sin \varphi_i^P$ $(i = 1, 2, 3)$, are not significant at the first round of estimation as expected, and thus they are set to zero, i.e., $P = I$, at the final round. Hence, $L_i(X_t)$ can be regarded as the instantaneous variance of the i-th PC. The constant terms, c_i $(i = 1, 2, 3)$, and the smallest eigenvalues in M^i, m_1^i $(i = 1, 2, 3)$, are all fixed to make Σ_t positive definite. The second largest eigenvalue in M^1, m_2^1, is also fixed at zero. Conse-

quently, $\sin \varphi_1^{Q^1}$ and $\sin \varphi_2^{Q^1}$ are unidentified, so they are set to zero. Hence, $L_1(X_t)$, the instantaneous variance of the first PC, has a simpler driving force than do $L_2(X_t)$ and $L_3(X_t)$.

Panel B of Table 1 presents the parameter estimates for the SV-E model. As with the SV-Q model, $\sin \varphi_i^P$ $(i = 1, 2, 3)$ in P are all set to zero. A notable feature is that none of the coefficients of the first PC, s_{i1} $(i = 1, 2, 3)$, is estimated significantly, so that they are set to zero at the final round of estimation. That is, the most persistent level factor is the least relevant to the volatility of yields, which is consistent with the earlier work and Figure 1 of this study showing that the level factor alone is difficult to capture the behavior of volatility. Also, the result implies a limitation of affine models: A prespecified volatility factor, when identified from the cross-section of yields, often has the highest correlation with the level factor.

4.2 Setup for volatility prediction

To make a more intuitive interpretation of the results, we predict the standard deviation, not the variance. The predictive power of the models is evaluated by the root mean squared error (RMSE) criterion using both in-sample and out-of-sample data. The RMSEs are computed from the residuals of the following equations:

$$\sqrt{\frac{RV_{t,t+h\Delta,i}}{h\Delta}} \quad = \quad \sqrt{\frac{\mathrm{var}_t[x_{t+h\Delta,i}]}{h\Delta}} + u_{t+h\Delta,i} , \tag{12}$$

$$\sqrt{\frac{RV_{t,t+h\Delta,i}}{h\Delta}} \quad = \quad a_{h,i} + b_{h,i}\sqrt{\frac{\mathrm{var}_t[x_{t+h\Delta,i}]}{h\Delta}} + u_{t+h\Delta,i}^{reg} , \tag{13}$$

where $\mathrm{var}_t[\cdot]$ stands for conditional variance.

The out-of-sample predictive power is evaluated in two approaches. In the first approach, both the model parameters in (1) and the regression parameters in (13) are held fixed at the in-sample estimates throughout the out-of-sample period to examine whether the models can produce a long-run predictive relation. In the second approach, a part of the parameters are estimated every time the prediction is made in a rolling window fashion with the sample size fixed at the same as the in-sample data. More specifically, the model parameters in (1) are held fixed at the in-sample estimates whereas the regression parameters in (13) are re-estimated. This treatment is consistent with the idea behind model estimation that the structural parameters are stable overtime. This idea will be supported by looking at Figures 2 and 3, the details of which are provided in Section 4.3.

It is noted that $\text{var}_t[X_{t+h\Delta}]$ has no analytical expression for the SV-E model, and thus is computed by the same approximation method as used for the estimation. But here, the interval is extended up to 32Δ, which raises concern on the accuracy of the approximation. In Appendix A, it is shown that the accuracy is maintained given reasonable parameter and state variable values.

To evaluate the predictive power of the proposed models, we prepare several competing models. First, we select the Gaussian model, which is given by (1) with Σ_t replaced by Σ, a constant positive definite matrix. Since, given h, the Gaussian model produces a constant forecast independently of the current state, it serves as a benchmark, as does Random Walk for predicting the level of yields. The RMSEs are computed based on equation (12) alone.

Second, we select the GARCH(1,1) model. Although there are a number of variants of the GARCH model, it is selected because more complicated models do not necessarily beat the simplest one in out-of-sample tests; see Hansen and Lunde (2005). The model is fitted to weekly data on each $x_{t,i}$:

$$x_{t+\Delta,i} = \alpha_i + \beta_i x_{t,i} + \sqrt{V_{t+\Delta,i}}\, z_{t+\Delta,i}\,, \tag{14}$$

$$V_{t+\Delta,i} = \omega_i + \phi_i V_{t,i} z_{t,i}^2 + \rho_i V_{t,i}\,, \tag{15}$$

where it is assumed for simplicity that $z_{t,i} \sim i.i.d.N(0,1)$. It is noted that since we work with the PCs, the estimation with each, but not joint, series seems to be justified.

The h-week ahead conditional variance can be computed by iteration, an explanation of which is provided in Appendix B. The RMSEs are computed based on both equations (12) and (13). More specifically, the out-of-sample RMSEs in the varying parameter approach are computed in a similar way to those for the SV-Q and SV-E models, where the model parameters in equations (14) and (15) are fixed at the in-sample estimates whereas the regression parameters in equation (13) are re-estimated.

Third, we consider forecasting regressions directly applied to the realized volatility series. We select the HAR-RV model employed by Andersen and Benzoni (2010) and the mixed data sampling (MIDAS) approach developed by Ghysels, Santa-Clara, and Valkanov (2005, 2006). In this research, the HAR-RV model is given by

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

The MIDAS regression is given by

$$\text{MIDAS:} \qquad \sqrt{\frac{RV_{t,t+h\Delta,i}}{h\Delta}} = a_{h,i}$$

$$+ b_{h,i}\sqrt{\frac{\sum_{j=1}^{n_t}(x_{t-L\Delta(j-1)/n_t,i} - x_{t-L\Delta j/n_t,i})^2 \, w_{h,i}(j)}{(L\Delta/n_t)}} + u_{t+h\Delta,i}^m \,, \quad (17)$$

where following Ghysels et al. (2005, equation (3)), the weighting function, $w_{h,i}(j)$, is given by

$$w_{h,i}(j) = \frac{\exp\{w_{h,i,1}j + w_{h,i,2}j^2\}}{\sum_{k=1}^{n_t}\exp\{w_{h,i,1}k + w_{h,i,2}k^2\}} \,, \qquad (18)$$

and n_t stands for the number of daily observations over the past L weeks: Usually $n_t = L \times 5$. We set $L = 32$ to make the amount of information equal between the HAR-RV and MIDAS regressions: The difference between the two is in the weighting structure to past observations. It is noted that the data of the first 32 weeks are not used in equations (16) and (17). To make competitive conditions equal, these data are not used in equations (12) and (13). To compute the out-of-sample RMSEs in the varying parameter approach by the MIDAS regression, $(w_{h,i,1}, \; w_{h,i,2})$ in equation (18) are fixed at the in-sample estimates, while $(a_{h,i}, \; b_{h,i})$ in equation (17) are re-estimated. For the HAR-RV regression (16), all parameters are re-estimated.

It is noted that volatility prediction is not the only purpose for the proposed models: They can be a basis for term structure models that are expected to explain both time-series and cross-sectional properties of the data. Nevertheless, we consider the GARCH model as a competitor, in contrast to Jacobs and Karoui (2009) where it is treated as a model generating the true volatility. Furthermore, an information gap exists against the HAR-RV and MIDAS regressions which are constructed directly from the realized volatility. By considering such a challenging setup, we uncover both possibilities and limitations of dynamic models composed solely of yield curve factors from a point of view of volatility prediction.

4.3 In-sample results

To obtain an intuition about the model performance, we first look at Figures 2–4, which display the time series of the four-week-ahead forecast of the volatility (annualized standard deviation) by the GARCH, SV-Q, and SV-E models, respectively, together with the corresponding realized series. First, comparing Figure 3 with Figure 2, we notice that the SV-Q model cannot predict the volatility of PC1 (the first PC) as accurately as the

GARCH model: The sample correlation over 1991–2009 is 0.35, which is smaller than that for the GARCH model, 0.57. Still, the trend of the realized series appears to be tracked. On the other hand, the SV-Q model generates forecasts of PC2 (the second PC) that are more highly correlated with the realized series than does the GARCH model. Moreover, it seems successful in capturing the intensive variation of the realized series during 2008–09, even though this period is out-of-sample. For PC3 (the third PC), the SV-Q model is again outperformed by the GARCH model judged by the sample correlation, however, it still appears to work in the out-of-sample period.

Next, looking at Figure 4, we notice that while the forecast series of PC1 generated by the SV-E model is similar to that generated by the SV-Q model, those of PC2 and PC3 are quite different. In particular, the SV-E model fails to generate sufficient variation relative to the realized series. A reason for these differences is that an exponential function can be well approximated by a linear function in a narrow range where the eigenvalues of the covariance matrix of changes in yields are involved and that the linear specification is not sufficient for capturing the intensive variation in the realized volatility.

Having these figures in mind, we compare the in-sample model performance. Table 2 presents the RMSEs, which are multiplied by 10^4 and thus interpretable in units of basis points (1 bp = 0.0001). For the proposed and GARCH models, the name alone indicates the results based on equation (12), whereas the name "+ Reg" indicates the results based on equation (13). In reporting the results, we focus on the comparison first without then with the forecasting regression.

First, compared with the RMSEs for the benchmark Gaussian model, those for the proposed models are smaller in most of the cases: Exceptions are PC1 at $h = 32$ for the SV-Q model and PC3 at $h = \{16, 32\}$ for the SV-E model. Between the proposed models, the SV-E model is better at predicting the volatility of PC1, whereas the SV-Q model is better at predicting the volatility of PC2 and PC3. Second, compared with the GARCH model, the performance of the proposed models is similar with respect to PC2. Specifically, the RMSEs for the SV-Q model range from 33.1 bps at $h = 4$ to 25.2 bps at $h = 32$, which are all smaller than the corresponding RMSEs for the GARCH model. For PC1 and PC3, however, the GARCH model forecasts better than the proposed models. Third, compared with the HAR-RV and MIDAS regressions, the proposed models are clearly outperformed. But for PC1, the performance gap tends to narrow with h. Fourth,

the forecasting regression (13) is helpful for the proposed models. In particular, for PC1, the SV-E + Reg model exhibits the best performance at longer horizons, $h = \{16, 32\}$, followed by the SV-Q + Reg model. For PC3, the differences in the RMSE between the SV-Q/E + Reg models and HAR-RV/MIDAS regressions shrink to around 1 bp or less by extending the horizon to $h \geq 8$. For PC2, however, the performance gap against the HAR-RV/MIDAS regressions still remains even after introducing the forecasting regression (13).

4.4   Out-of-sample results

Table 3 presents the out-of-sample RMSEs in basis points in the fixed-parameter approach, where all parameter values are fixed at the in-sample estimates throughout the out-of-sample period. First, compared with the RMSEs for the benchmark Gaussian model, those for the proposed models are smaller except PC3 for the SV-E model. Between the proposed models, the SV-Q model is better at predicting the volatility of PC2 and PC3 while both exhibit a similar performance for PC1.

Second, compared with the GARCH model, the proposed models are outperformed largely with respect to PC1. For example, the differences in the RMSE between the SV-E and GARCH models reach 18 bps at $h = 8$. In fact, the GARCH model is the best at predicting the out-of-sample volatility of PC1 at $h \geq 8$, showing its robust performance. Looking at PC2, the SV-Q model is comparable to the GARCH model with the former (latter) exhibiting a superior performance at shorter (longer) horizons. The SV-E model, on the other hand, is outperformed by the GARCH model with the gap being around 4 bps for all h. Throughout the sample period, therefore, the level-dependent covariance matrix with quadratic specification is at least as effective as the GARCH(1,1) specification in predicting the volatility of PC2. Also for PC3, only the SV-Q model is comparable to, or slightly worse than, the GARCH model.

Third, introducing the forecasting regression (13) to the proposed models does not necessarily improve the out-of-sample performance when the regression parameters are held fixed. Specifically, the RMSEs of PC1 for the SV-Q/E + Reg models are 83–84 bps at $h = 4$, which increase from those without the forecasting regression, 79–80 bps. Such differences disappear at $h = 32$, however. For PC2 and PC3, the forecasting regression is generally favorable for the SV-E model while it is not for the SV-Q model.

Fourth, the proposed models, in both cases with and without the forecasting regression, are outperformed by the HAR-RV/MIDAS regressions. By extending the forecasting horizon, the performance gap tends to narrow for PC1 while it does not for PC2 or PC3.

Table 4 presents the out-of-sample RMSEs in basis points in the varying-parameter approach, where the regression parameters are re-estimated every time the prediction is made. First, for both PC1 and PC3, the proposed models are clearly outperformed. Between the proposed models, the SV-Q (SV-E) model is slightly better at predicting the volatility of PC1 (PC3). The GARCH model exhibits a comparable performance to the HAR-RV/MIDAS regressions for PC1 at $h = 32$. Second, for PC2, the SV-Q model is comparable to the GARCH model while the SV-E model is not. Compared with the HAR-RV/MIDAS regressions, the SV-Q model is slightly worse at $h \leq 8$ with the gap being around 2 bps, but it reaches more than 5 bps at $h = 32$.

5   Concluding remarks

We have predicted the volatility of yield curve factors with the purpose of examining information content of the yield curve. The information is used for specifying the instantaneous covariance matrix of yield curve factors. Specifically, the eigenvalues are modeled by quadratic (model SV-Q) and exponential (model SV-E) functions of the factors. We find that the quadratic specification has potential to capture the intensive variation in the realized volatility. Furthermore, it exhibits a comparable, or even superior, performance to the GARCH(1,1) specification in predicting the volatility of the second principle component of the yield curve that is interpretable as a slope factor. That is, the yield curve can be considered to have predictive power for the volatility of the slope factor. Against forecasting regressions constructed directly from the realized volatility series, however, a performance gap inevitably exists.

At the same time, the empirical results also uncover limitations of the role of the yield curve in predicting the volatility of the first and third principle components. Information on the yield curve alone is not sufficient, and therefore, it is for the prediction of their volatility that the introduction of unspanned volatility factors is effective. It is then of interest to examine how unspanned volatility factors introduced for capturing time series properties are linked to those extracted from the cross-section of option prices, which is left for future research.

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, we limit our attention to 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:

$$dx_{t,i} = f_i(X_t)dt + \xi_i(X_t)'dW_t \qquad (i = 1, 2) , \tag{19}$$

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 (19) 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 (19) and taking the conditional expectation,

$$E_s[x_{t,i} - x_{s,i}] = E_s \left[\int_s^t f_i(X_u)du \right] . \tag{20}$$

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

$$\begin{aligned}
f_i(X_u) = {} & f_i(X_s) \\
& + f_i^{(1,0)}(X_s)(x_{u,1} - x_{s,1}) + f_i^{(0,1)}(X_s)(x_{u,2} - x_{s,2}) + \frac{1}{2} f_i^{(2,0)}(X_s)(x_{u,1} - x_{s,1})^2 \\
& + \frac{1}{2} f_i^{(0,2)}(X_s)(x_{u,2} - x_{s,2})^2 + f_i^{(1,1)}(X_s)(x_{u,1} - x_{s,1})(x_{u,2} - x_{s,2}) + e_i , \tag{21}
\end{aligned}$$

17

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 (21) into equation (20) and expressing the resulting equation in a vector form

$$E_s[x_{t,i} - x_{s,i}] = f_i(t-s)$$
$$+ \left(f_i^{(1,0)} \quad f_i^{(0,1)} \quad \frac{1}{2} f_i^{(2,0)} \quad \frac{1}{2} f_i^{(0,2)} \quad f_i^{(1,1)} \right) \int_s^t \Psi_{s,u} du + R_i , \qquad (22)$$

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,

$$E_s[(x_{t,1} - x_{s,1})^2] = E_s \left[\int_s^t \{2f_1(X_u)(x_{u,1} - x_{s,1}) + g_{11}(X_u)\} du \right] , \qquad (23)$$

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 (23) becomes

$$2f_1(X_u)(x_{u,1} - x_{s,1}) + g_{11}(X_u)$$
$$= g_{11}(X_s) + \{2f_1(X_s) + g_{11}^{(1,0)}(X_s)\}(x_{u,1} - x_{s,1}) + g_{11}^{(0,1)}(X_s)(x_{u,2} - x_{s,2})$$
$$+ \{2f_1^{(1,0)}(X_s) + \frac{1}{2} g_{11}^{(2,0)}(X_s)\}(x_{u,1} - x_{s,1})^2 + \frac{1}{2} g_{11}^{(0,2)}(X_s)(x_{u,2} - x_{s,2})^2$$
$$+ \{2f_1^{(0,1)}(X_s) + g_{11}^{(1,1)}(X_s)\}(x_{u,1} - x_{s,1})(x_{u,2} - x_{s,2}) + e_{11} , \qquad (24)$$

where $g^{(k,l)}$ is defined analogously with $f^{(k,l)}$, and e_{11} is a residual term. By substituting equation (24) into equation (23),

$$E_s[(x_{t,1} - x_{s,1})^2] = g_{11}(t-s)$$
$$+ \left(2f_1 + g_{11}^{(1,0)} \quad g_{11}^{(0,1)} \quad 2f_1^{(1,0)} + \frac{1}{2} g_{11}^{(2,0)} \quad \frac{1}{2} g_{11}^{(0,2)} \quad 2f_1^{(0,1)} + g_{11}^{(1,1)} \right)$$
$$\times \int_s^t \Psi_{s,u} du + R_{11} , \qquad (25)$$

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

$$\Psi_{s,t} = A(X_s) \int_s^t \Psi_{s,u} du + b(X_s)(t-s) + R , \qquad (26)$$

where

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

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

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

Equation (26) can be solved as

$$\Psi_{s,t} = \int_s^t e^{A(X_s)(t-u)} b(X_s) du + \hat{R} \,. \tag{27}$$

If, in addition, A is invertible, we obtain

$$\Psi_{s,t} = A^{-1}(X_s)\{e^{A(X_s)(t-s)} - I\}b(X_s) + \hat{R} \,. \tag{28}$$

It is noted that equations (26)–(28) 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} $(i, j = 1, ..., d)$ 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 of the proposed models, we consider $n = 2$.

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 SV-Q 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 SV-E model

Under the SV-E model, the conditional standard deviation calculated from equation (28) contains approximation error. We check the accuracy of the approximation to $\sqrt{\mathrm{var}_t[X_{t+h\Delta}]/h\Delta}$ by the Monte Carlo (MC) method. The parameter values for the SV-E model are given in Panel B of Table 1. The starting values for the MC simulations, at which the accuracy is evaluated, are selected from the actual data. Specifically, we pick

up three dates from the entire 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 (1) 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 10,000 with antithetic variates.

Panels A and B of Table A present forecasts of the 32-week ahead volatility (annualized standard deviation in units of basis points) 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 seems to be of little concern in the current setup.

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

The variance of $x_{t+\Delta,i}$ $(i = 1, 2, 3)$ conditioned on time t is simply $V_{t+\Delta,i}$, which is observed at time t. The variance of $x_{t+k\Delta,i}$ $(k = 2, ..., h)$ conditioned on time t is computed iteratively as follows. In equation (14), by substituting $t + (k-1)\Delta$ for t and taking the variance conditioned on time t,

$$
\begin{aligned}
\mathrm{var}_t[x_{t+k\Delta,i}] &= \mathrm{var}_t[\alpha_i + \beta_i x_{t+(k-1)\Delta,i} + \sqrt{V_{t+k\Delta,i}}z_{t+k\Delta,i}] \\
&= \beta_i^2 \mathrm{var}_t[x_{t+(k-1)\Delta,i}] + \mathrm{var}_t[\sqrt{V_{t+k\Delta,i}}z_{t+k\Delta,i}] \\
&\quad + 2\beta_i \, \mathrm{cov}_t[x_{t+(k-1)\Delta,i}, \quad \sqrt{V_{t+k\Delta,i}}z_{t+k\Delta,i}] \\
&= \beta_i^2 \mathrm{var}_t[x_{t+(k-1)\Delta,i}] + E_t[V_{t+k\Delta,i}] \qquad (k = 2, ..., h) \, . \qquad (29)
\end{aligned}
$$

On the other hand, in equation (15), by substituting $t + (k-1)\Delta$ for t and taking the expectation conditioned on time t,

$$
\begin{aligned}
E_t[V_{t+k\Delta,i}] &= E_t[\omega_i + \phi_i V_{t+(k-1)\Delta,i} z_{t+(k-1)\Delta,i}^2 + \rho_i V_{t+(k-1)\Delta,i}] \\
&= \omega_i + (\phi_i + \rho_i)E_t[V_{t+(k-1)\Delta,i}] \qquad (k = 2, ..., h) \, . \qquad (30)
\end{aligned}
$$

Then, $\mathrm{var}_t[x_{t+h\Delta,i}]$ is obtained by iteratively solving equations (29) and (30) starting from $\mathrm{var}_t[x_{t+\Delta,i}] = V_{t+\Delta,i}$.

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.

Bekaert, G., and R. J. Hodrick, 2001, Expectations hypotheses tests, *Journal of Finance* 56, 1357-1394.

Bikbov, R., and M. Chernov, 2011, Yield curve and volatility: Lessons from Eurodollar futures and options, *Journal of Financial Econometrics* 9, 66-105.

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.

Campbell, J. Y., and R. J. Shiller, 1991, Yield spreads and interest rate movements: A bird's eye view, *Review of Economic Studies* 58, 495-514.

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., 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.

Dai, Q., and K. J. Singleton, 2003, Term structure dynamics in theory and reality, *Review of Financial Studies* 16, 631-678.

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.

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

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

Fama E. F., and R. R. Bliss, 1987, The information in long-maturity forward rates, *American Economic Review* 77, 680-692.

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.

Joslin, S., K. J. Singleton, and H. Zhu, 2011, A new perspective on Gaussian dynamic term structure models, *Review of Financial Studies* 24, 926-970.

Li, H., and F. Zhao, 2006, Unspanned stochastic volatility: Evidence from hedging interest rate derivatives, *Journal of Finance* 61, 341-378.

Litterman, R., J. Scheinkman, and L. Weiss, 1991, Volatility and the yield curve, *Journal of Fixed Income* 1, 49-53.

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.

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., 2010, Term Structure Models Can Predict Interest Rate Volatility. But How?, Tsukuba Economics Working Papers 2010-008, University of Tsukuba.

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: Estimates for the SV-Q 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: Estimates for the SV-E 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 1: Parameter estimates (standard errors) of $L_i(X_t)$ for the SV-Q and SV-E models

The instantaneous covariance matrix of changes in the first three PCs, X_t, is decomposed as $\Sigma_t = PL_tP'$, where L_t is the diagonal eigenvalue matrix and P is the orthogonal eigenvector matrix parameterized in equation (4): Actually, $P = I$ in the final estimation for both models. The SV-Q model specifies the i-th diagonal element of L_t as $L_i(X_t) = c_i + X_t'\Gamma^i X_t$ $(i = 1, 2, 3)$, where $c_i > 0$ and Γ^i is a non-negative definite matrix. Γ^i is also parameterized based on the spectral decomposition as $\Gamma^i = Q^i M^i Q^{i\prime}$ $(i = 1, 2, 3)$, where M^i is the diagonal eigenvalue matrix with its elements satisfying $0 \leq m_1^i \leq m_2^i \leq m_3^i$, and Q^i is the orthogonal eigenvector matrix parameterized in equation (8). The SV-E model specifies the i-th diagonal element of L_t as $L_i(X_t) = \exp\{s_{i0} + s_i'X_t\}$ $(i = 1, 2, 3)$. The in-sample data from January 4, 1991 to April 9, 2003 are used for the estimation.

Model/Horizon	h = 4			h = 8			h = 16			h = 32		
	PC1	PC2	PC3	PC1	PC2	PC3	PC1	PC2	PC3	PC1	PC2	PC3
Gaussian	68.6	36.3	27.2	59.4	33.8	27.6	51.4	31.3	29.6	42.3	28.1	33.1
SV-Q	67.2	33.1	26.3	57.8	30.4	26.5	49.9	28.1	28.4	42.9	25.2	31.7
SV-E	64.2	34.3	26.7	54.5	31.7	27.3	46.5	29.2	29.7	39.5	26.2	33.5
GARCH	61.8	33.9	24.8	53.2	31.3	24.5	46.9	29.0	25.3	38.4	26.3	27.9
SV-Q + Reg	58.9	32.2	21.2	49.6	29.1	18.7	*42.1*	26.6	16.2	*34.8*	24.0	14.0
SV-E + Reg	58.4	33.3	21.5	49.1	30.0	19.0	**41.5**	27.1	*16.0*	**34.5**	24.5	13.7
GARCH + Reg	56.4	33.5	20.9	*48.2*	30.7	18.9	42.7	28.0	16.3	36.0	25.0	13.5
HAR-RV	*56.2*	**28.5**	**19.1**	48.7	**25.6**	**17.7**	43.1	**22.9**	**15.3**	36.9	**20.5**	**12.9**
MIDAS	**55.5**	*28.8*	*19.4*	**48.0**	*26.2*	*18.0*	43.5	*23.8*	16.2	36.9	*20.7*	*13.0*

Table 2: In-sample RMSEs for volatility prediction

RMSEs for in-sample prediction of the h-week ahead yield volatility (annualized standard deviation) are presented in basis points with $h = \{4, 8, 16, 32\}$. For the proposed and GARCH models, the name alone indicates the results based on equation (14), whereas the name "+ Reg" indicates the results based on equation (15). The best and the second best results in each column are displayed in bold and italic, respectively. The in-sample period is from January 4, 1991 to April 9, 2003.

Model/Horizon	h = 4			h = 8			h = 16			h = 32		
	PC1	PC2	PC3	PC1	PC2	PC3	PC1	PC2	PC3	PC1	PC2	PC3
Gaussian	87.5	38.6	22.9	82.0	36.3	22.7	78.5	34.3	23.7	74.3	31.6	25.1
SV-Q	80.0	31.0	18.8	74.9	28.7	17.1	72.3	29.0	18.0	70.0	30.1	20.8
SV-E	79.3	36.3	23.5	74.5	34.6	23.9	72.3	33.0	25.5	71.2	30.2	26.8
GARCH	63.5	32.1	17.0	**56.5**	30.5	16.5	**56.4**	28.8	17.7	**55.4**	26.2	20.4
SV-Q + Reg	83.8	31.9	24.3	78.2	30.4	23.8	74.5	30.6	24.1	70.4	31.1	24.5
SV-E + Reg	83.4	35.5	21.3	78.1	33.7	20.3	74.7	32.4	20.0	71.1	31.0	20.4
GARCH + Reg	65.2	34.0	22.3	59.9	32.7	21.6	61.2	31.3	21.8	61.4	29.1	19.0
HAR-RV	*62.8*	*28.4*	**16.1**	58.2	**27.2**	**15.5**	59.3	**25.4**	**15.5**	*60.9*	**24.7**	**15.6**
MIDAS	**62.2**	**28.2**	16.4	58.1	27.5	16.7	57.2	26.6	17.4	61.1	*24.8*	*16.1*

Table 3: Out-of-sample RMSEs for volatility prediction in the fixed-parameter approach

RMSEs for out-of-sample prediction of the h-week ahead yield volatility (annualized standard deviation) are presented in basis points with $h = \{4, 8, 16, 32\}$. For the proposed and GARCH models, the name alone indicates the results based on equation (14), whereas the name "+ Reg" indicates the results based on equation (15). All parameters are fixed at the in-sample estimates throughout the out-of-sample period from April 16, 2003 to May 27, 2009. The best and the second best results in each column are displayed in bold and italic, respectively.

| Model/Horizon	h = 4			h = 8			h = 16			h = 32		
	PC1	PC2	PC3	PC1	PC2	PC3	PC1	PC2	PC3	PC1	PC2	PC3
SV-Q + Reg	78.6	31.0	21.4	72.4	29.5	20.6	69.2	30.2	21.0	66.7	31.7	22.4
SV-E + Reg	80.3	36.7	21.4	74.8	35.1	20.4	71.9	33.9	20.2	69.8	32.9	20.8
GARCH + Reg	64.0	32.1	18.4	57.6	31.0	18.1	58.1	30.3	19.0	**52.1**	28.9	18.2
HAR-RV	*60.8*	*29.1*	*16.1*	**54.5**	**28.0**	**15.6**	**54.0**	**26.5**	**15.7**	*52.2*	**26.1**	**15.6**
MIDAS	**60.4**	**28.8**	**15.8**	*55.1*	*28.2*	*16.1*	*55.9*	*27.8*	*17.1*	52.3	*26.4*	*15.7*

Table 4: Out-of-sample RMSEs for volatility prediction in the varying-parameter approach

RMSEs for out-of-sample prediction of the h-week ahead yield volatility (annualized standard deviation) are presented in basis points with $h = \{4, 8, 16, 32\}$. For the proposed and GARCH models, the model parameters are fixed at the in-sample estimates and the regression parameters in equation (15) are re-estimated in a rolling-window fashion. For MIDAS, the weighting parameters in equation (18) are fixed at the in-sample estimates and $(a_{h,\tau}, b_{h,\tau})$ in equation (17) are re-estimated. For HAR-RV, all parameters in equation (16) are re-estimated. The best and the second best results in each column are displayed in bold and italic, respectively. The out-of-sample period is from April 16, 2003 to May 27, 2009.

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

Table A: Accuracy to $\sqrt{\mathbf{var}_t[X_{t+h\Delta}]/h\Delta}$ with $h = 32$ for the SV-E model

Panels A and B present forecasts of the 32-week ahead volatility (annualized standard deviation in basis points) for the SV-E 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 SV-Q model can be computed exactly.

(a) Level

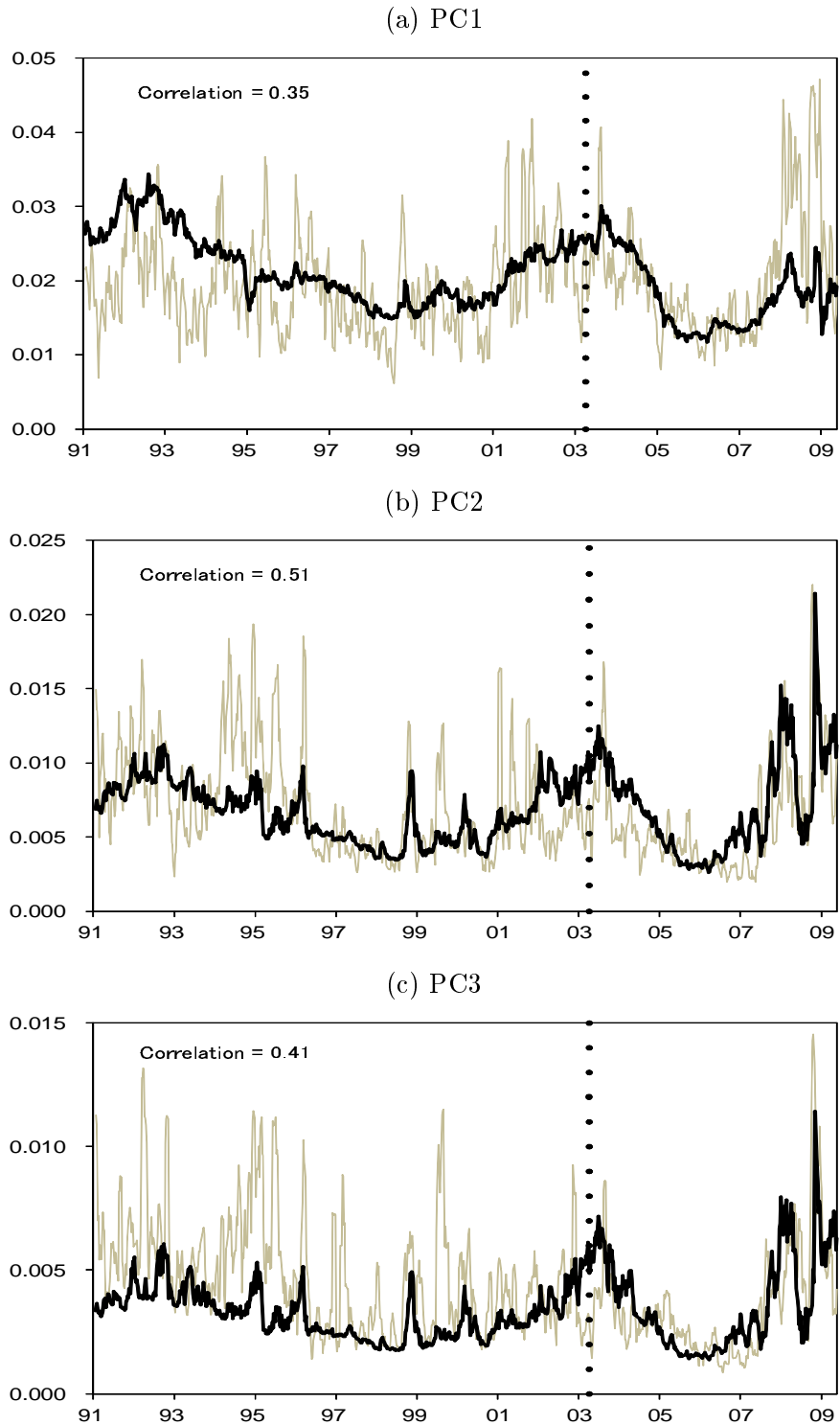(b) Realized Standard Deviation
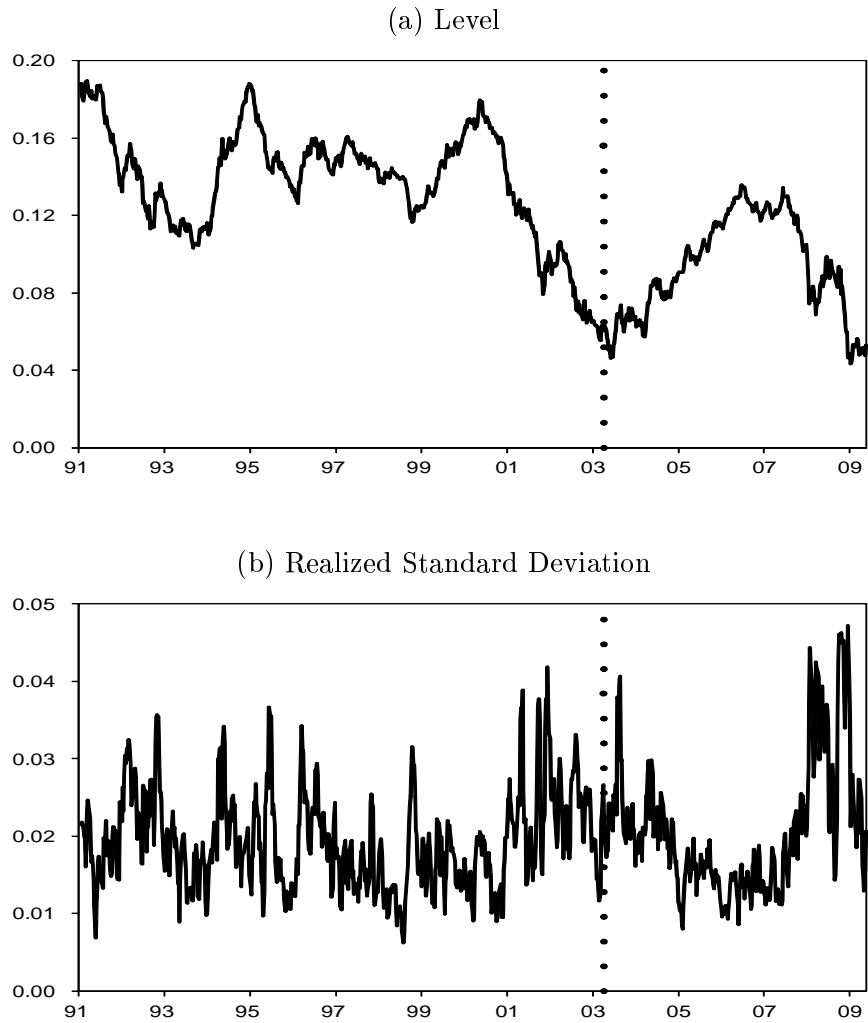
Figure 1: Time-series of the level and realized volatility of PC1 over 1991–2009
Panels (a) and (b) present the time-series of the level and realized volatility (annualized standard deviation) of the first principle component (PC1). The vertical dotted line separates the in-sample and out-of-sample periods.
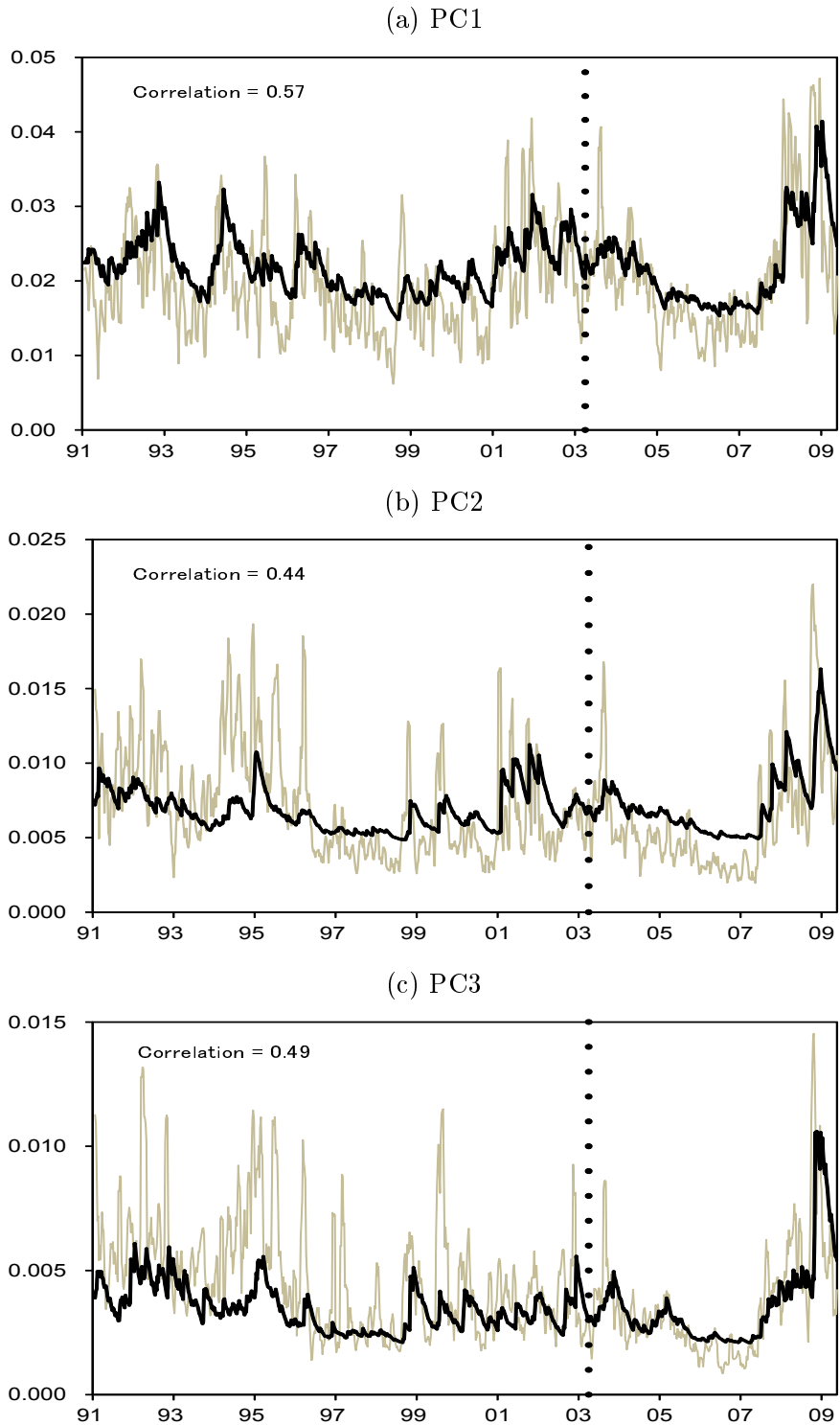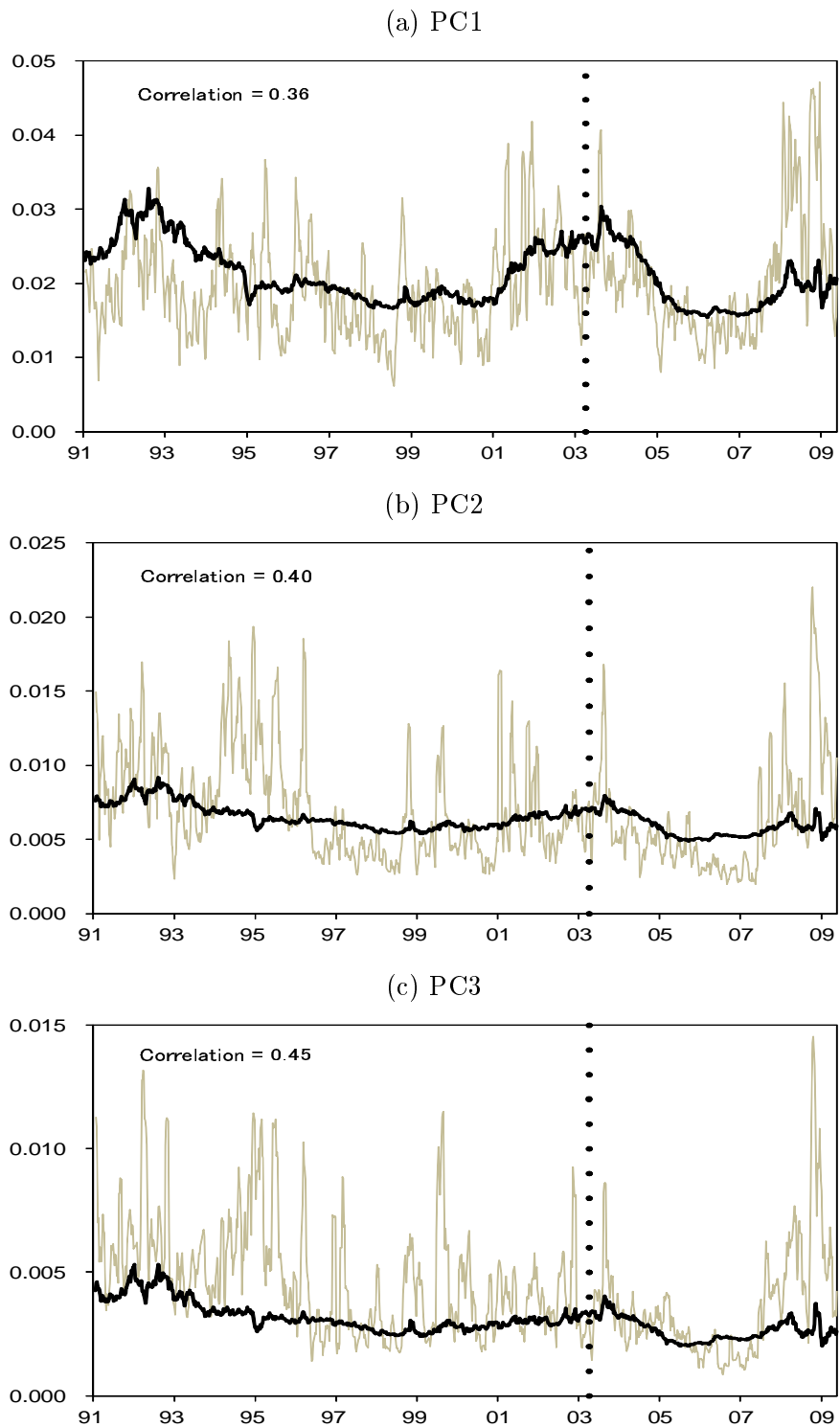
Figure 2: Time series of the four-week ahead volatility forecast by the GARCH (1,1) model

The model forecast (the thick line) and the corresponding realized measure (annualized standard deviation, the thin line) are displayed, with the vertical dotted line separating the in-sample and out-of-sample periods.

(a) PC1

(b) PC2

(c) PC3

Figure 3: Time series of the four-week ahead volatility forecast by the SV-Q model

The model forecast (the thick line) and the corresponding realized measure (annualized standard deviation, the thin line) are displayed, with the vertical dotted line separating the in-sample and out-of-sample periods.

Figure 4: Time series of the four-week ahead volatility forecast by the SV-E model

The model forecast (the thick line) and the corresponding realized measure (annualized standard deviation, the thin line) are displayed, with the vertical dotted line separating the in-sample and out-of-sample periods.