Tsukuba Economics Working Papers
No. 2010-008

Term Structure Models Can Predict Interest Rate Volatility. But How?

by

Hideyuki Takamizawa

November 2010

This version: April 2011

UNIVERSITY OF TSUKUBA
Department of Economics
1-1-1 Tennodai
Tsukuba, Ibaraki 305-8571
JAPAN

Term Structure Models Can Predict Interest Rate Volatility. But How?

Hideyuki Takamizawa

Graduate School of Humanities and Social Sciences,
University of Tsukuba
Tsukuba Ibaraki 305-8571, JAPAN
E-mail: takamiza[at]social.tsukuba.ac.jp

April 17, 2011

Abstract

This paper attempts to predict the volatility of interest rates through dynamic term structure models. For this attempt, the models are improved, based on the three-factor Gaussian model, to have level-dependent volatilities supported by data. The empirical results show that the predictive power of the proposed models is higher than that of the affine models. Compared with time-series models, it is low for the four-week forecasting horizon but can be comparable for middle to long term rates by extending the horizon up to 32 weeks. The combination of these two different types of forecasts can lead to higher predictive power.

Keywords: Term structure, Stochastic volatility, Realized volatility, Approximation of conditional moments.

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 paper attempts to predict the volatility by utilizing such information through dynamic term structure models. For this attempt, the models are improved without sacrificing the goodness-of-fit to the yield curve or the predictive power to the level of interest rates. This improvement will contribute to extending the versatility of the term structure models.

The idea of relating interest rate volatility to the yield curve is not new. Brown and Schaefer (1994), Christiansen and Lund (2005), 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 relation, however, no longer seems to be a decisive feature for relatively recent data. Figure 1 shows the time series of interest rates and realized volatilities (annualized standard deviations) constructed from U.S. dollar LIBOR and swap rates over 1991–2009: the details of the construction of the realized volatility measure are provided in Section 2. We notice that the sharp rise in the volatility, observed around 2001–03 and 2008–09, is actually accompanied by the fall in the level of interest rates. 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 (AB) (2010) test affine spanning conditions that the yield variance, both *ex ante* and *ex post*, can be expressed by a linear combination of yields if affine term structure models are true, and reject these conditions. Collin-Dufresne, Goldstein, and Jones (CDGJ) (2009), and Jacobs and Karoui (JK) (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.

Nevertheless, it may be too early to draw the conclusion. Some important issues are left unaddressed. The first issue is the predictability of the volatility over long forecasting horizons: AB (2010) considered up to one-month horizons and CDGJ (2009) considered the one-week horizon. Longer horizons are worth considering, however. In predicting the level of yields such as in the context of the efficient hypothesis of the term structure (EHT), the horizons longer than a month are often considered; see, e.g., Bekaert and Hodrick (2001), Campbell and Shiller (1991), Duffee (2002), and Fama and Bliss (1987). Then, if we match the horizon of the volatility forecast with that of the level forecast typically considered, the predictability of the risk-return relation implied by the yield curve can be explored. The second issue is the use of non-affine term structure models for extracting the volatility. The affine models used by CDGJ (2009), and JK (2009) have restrictions that are necessary for obtaining a closed-form expression of bond prices and yields. Removing such restrictions may improve the model's ability of capturing the volatility.

To address the first issue, we perform the volatility forecasting regression with particular attention to long horizons and out-of-sample performance. We use daily data to construct weekly data on a realized yield variance measure over 4-, 8-, 16-, and 32-week horizons. It is then regressed on a linear combination of the current yields. The possibility of a nonlinear relation between them is also considered in a relatively simple way. We make nonlinear transformations of the dependent variable, specifically the square-root and logarithmic transformations, while leaving the independent variables linear. The results are that the in-sample predictive power of the yield curve is not as bad as it is reported: it improves the longer the forecasting horizon. On the other hand, the results of the out-of-sample predictive power of the yield curve are mixed. We prepare two approaches for the out-of-sample analysis. In the first approach where the parameter values of the regression models are fixed at the in-sample estimates throughout the out-of-sample period, the predictive power deteriorates. In the second approach where the parameters are estimated every time the prediction is made, the predictive power still remains high. Taken together, at least some portion of volatility information seems to be contained in the yield curve, but a simple linear combination of yields fails to produce a long-run predictive relation. This motivates us to use dynamic term structure models for predicting the volatility.

To address the second issue, we attempt to develop term structure models without sacrificing the goodness-of-fit to the cross-section of yields or the predictive power to

the level of yields. Such models may naturally be outside a class of models having a closed-form expression of bond prices and yields, which will be a major obstacle to this research. We overcome this obstacle by relying on an analytical approximation proposed by Shoji (2002), and Takamizawa and Shoji (2009). This method approximates a vector of conditional moments as the solution to an ordinary differential equation (ODE). Since the price of a zero-coupon bond is derived as the conditional expectation of the stochastic discount factor, this method can be directly applied. The accuracy of the approximation, which is carefully investigated also in this paper, seems to be maintained when realistic values of parameter and state vectors are provided.

Now, what specification is appropriate for the volatility prediction without sacrificing the desirable properties of the existing models? Clues to the answer can be found in the existing models as well as in the results of the volatility forecasting regression. It is well known that the three-factor Gaussian model can capture many important features of the actual data, such as the failure of the EHT and a hump shape of unconditional volatilities of changes in logarithmic yields; see Dai and Singleton (2003). An obvious drawback of this model is constant volatility despite the fact that time-varying volatility is one of the decisive features of the data. We then overcome this difficulty by making the volatility in this model level-dependent, leading to a model in which all factors potentially contribute to time-varying volatility.

To introduce the level-dependent volatility most effectively, we carefully choose a baseline specification of the Gaussian model, which has numerous specifications depending on factor rotation. One candidate is a classical specification, which is composed of the instantaneous risk-free rate, a stochastic central tendency factor, and an additional factor; see, e.g., Andersen and Lund (1997b), Balduzzi, Das, and Foresi (1998), Balduzzi, Das, Foresi, and Sundaram (1996), and Bikbov and Chernov (2011). The former two factors are expected to be highly correlated with short-term and long-term yields, respectively, as is indeed the case shown later. Therefore, we can take advantage of the results of the volatility forecasting regression that a combination of yields is more or less useful for predicting the volatility.

Another care in the introduction of the level-dependent volatility is to keep the instantaneous covariance matrix of the factors positive definite. One simple way to meet this requirement is to model eigenvalues of the covariance matrix as functions taking positive

values. We propose two models for the time-varying eigenvalues: one given by quadratic functions and the other by exponential functions. Apart from keeping the volatility positive, the positive definiteness of the covariance matrix has a significant merit. Since by construction the inverse of the covariance matrix is well defied, the market prices of risks can be modeled as flexibly as those for the original Gaussian model, which is the so-called essentially affine specification; see Duffee (2002). Therefore, many of the advantages of the Gaussian model are expected to be inherited into the proposed models, which is indeed the case as shown later.

In comparing the volatility forecasting performance of the proposed models, we consider both affine term structure models and time-series models. The latter include the GARCH(1,1) model, which is estimated from lower frequency data. In addition, a variant of the AR model employed by AB (2010) and the mixed data sampling (MIDAS) approach developed by Ghysels, Santa-Clara, and Valkanov (2005, 2006) are also considered, both of which are estimated from higher frequency data. The results of the performance comparison are summarized as follows. First, compared with the affine models, the proposed models exhibit a superior performance in most of the cases. Second, compared with the time-series models, the proposed models are outperformed for the four-week forecasting horizon. By extending the horizon up to 32 weeks, while the high performance of the time-series models continues at short maturities, the relative performance of the proposed models increases. In fact, there are cases at middle to long maturities in which the predictive power of the proposed models is comparable, or even superior, to that of the time-series models in both the in-sample and out-of-sample periods. Third, by combining the forecast of the proposed models with that of the high-frequency regression model, the predictive power can further increase at middle to long maturities. We also investigate whether the improvement of the volatility prediction is attributable to the imposition of non-arbitrage.

In spite of the improvement, the proposed models have a number of drawbacks. Apart from the computational complexity due to the lack of closed-form expressions for bond and derivative prices, we particularly point out two drawbacks. First, since the proposed models are based on the Gaussian model, negative interest rates occur with positive probability. Second, the proposed models do not accommodate factors that affect derivative prices but not underlying bond prices, in spite of the earlier work pointing out the exis-

tence and significance of such unspanned factors; see, e.g., Collin-Dufresne and Goldstein (2002), Han (2007), Heidari and Wu (2003, 2009), Jarrow, Li, and Zhao (2007), and Li and Zhao (2006). We also discuss how to overcome these difficulties.

Section 2 explains the data and how to construct a realized volatility measure. Section 3 performs the volatility forecasting regression. Section 4 introduces our models, and explains pricing and estimation methods. Section 5 examines the volatility forecasting performance of the proposed models. Section 6 provides concluding remarks. Appendices collect supplementary analyses as well as technical arguments regarding the derivation and accuracy of the approximation method.

2 Data and realized variance measure

2.1 Dataset

We use data on U.S. dollar LIBOR and swap rates, covering the period from January 4, 1991 to May 27, 2009. There are mainly two reasons for choosing this dataset. First, we can make the purpose of predicting the volatility through term structure models more challenging. As mentioned in Introduction and exhibited in Figure 1, parallel movements in the level and volatility of a particular yield disappear in the recent data. Second, we can focus on the 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.

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.

2.2 Realized measure

A realized measure of the conditional variance of changes in yields is constructed from the daily data. We generate weekly, Wednesday, realized variance, where it is assumed that a week begins with Thursday and ends with Wednesday. We choose Wednesday because we later estimate term structure models using weekly, Wednesday, data. We have 960 weekly observations of the realized variance in total, among which first 640 observations (two-

thirds of the sample), ending on April 9, 2003, are used for the in-sample estimation and the remaining 320 observations are used for the out-of-sample prediction. This division is aimed at incorporating information on the lowest range of interest rates into model estimation as well as reserving sufficient out-of-sample observations. Other ratios, three-to-one and four-to-one, are also tried, and the differences in the results are not large enough to change the conclusions of this paper.

A realized measure of the one-week ahead conditional variance of a τ-maturity yield is computed as

$$RV_{t,t+\Delta,\tau} = \sum_{i=1}^{m_{t+\Delta}} \left(y_{t+\frac{\Delta}{m_{t+\Delta}}i,\tau} - y_{t+\frac{\Delta}{m_{t+\Delta}}(i-1),\tau}\right)^2 , \tag{1}$$

where

- $y_{t,\tau}$: a zero-coupon bond yield at time t with τ years to maturity.
- Δ: a week interval set to $1/52$.
- m_t: 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

$$RV_{t,t+h\Delta,\tau} = \sum_{j=1}^{h} RV_{t,t+j\Delta,\tau} \ . \tag{2}$$

The annualized realized variance is obtained by dividing $RV_{t,t+h\Delta,\tau}$ by $h\Delta$. We consider $h = \{4, 8, 16, 32\}$. The forecasting horizons of $h > 4$ are longer than previously considered.

The realized measure based on the daily data is crude. Discretization error arising from approximating the integral of the quadratic variation may be large. This crude measure, however, is more or less unavoidable because of the limited availability of intra-daily data on interest rates; see CDGJ (2009), and JK (2009), where daily data are also used to construct realized measures. Another concern is longer forecasting horizons. In equations (1) and (2), the variation in the unpredictable component alone is reflected: the covariation between the predictable and unpredictable components and the variation in the predictable component are omitted, which are of smaller order than the variation in the unpredictable component; see, e.g., Andersen, Bollerslev, Diebold, and Labys (2003). The omission can be justified for a short horizon, say day or week, but it may not for a long horizon considered in this paper. In particular, if the mean reversion, which is captured by the predictable component, is crucial, the realized measure based on equations (1) and

(2) will overestimate the true conditional variance since the mean reversion reduces the yield variation over a finite horizon.

In Appendix A, we verify the impact of omitting the mean reversion on the realized measure by applying to each yield the Cox, Ingersoll, and Ross (CIR) (1985) model, which has a closed-form expression of the conditional variance for a finite horizon. Specifically, letting β be a parameter in the drift controlling the speed of mean reversion, we compare the conditional variance with β set at a point estimate from the actual data to that with β set at zero. The summary of the results is as follows. The longer the maturity, the speed of mean reversion tends to increase, so does the resulting overestimation of the realized measure. However, since the estimate of β is not significantly different from zero, as is often the case for interest rate data, the actual impact of regarding β to be zero is not as serious as it appears.

3 Volatility Forecasting Regression

The purpose of this section is to examine if the current yield curve contains useful information for predicting the volatility of changes in yields. We perform the volatility forecasting regression with particular attention to the predictability for longer forecasting horizons and out-of-sample.

3.1 Setup

We regress the h-week ahead realized measure computed over time t and $t+h\Delta$ on linear combinations of yields observed at time t:

$$\text{Y-PART:} \qquad f\left(\frac{RV_{t,t+h\Delta,\tau}}{h\Delta}\right) = a_{h,\tau} + b'_{h,\tau} Y_t^p + u^p_{t+h\Delta,\tau} \ , \tag{3}$$

where

$$Y_t^p = (y_{t,0.5} \quad y_{t,2} \quad y_{t,10})' \ ,$$

and

$$\text{Y-ALL:} \qquad f\left(\frac{RV_{t,t+h\Delta,\tau}}{h\Delta}\right) = a_{h,\tau} + b'_{h,\tau} Y_t^a + u^a_{t+h\Delta,\tau} \ , \tag{4}$$

where

$$Y_t^a = (y_{t,0.5} \quad y_{t,1} \quad y_{t,2} \quad y_{t,3} \quad y_{t,5} \quad y_{t,10})' \ .$$

By considering a nonlinear function, f, the possibility of a nonlinear relation between the realized measure and current yields can be addressed in a simple way. We actually consider $f(x) = \sqrt{x}$ and $f(x) = \ln x$, as well as $f(x) = x$ as in the earlier studies taking account of an implication of the affine models. To conserve space, we do not report all of the results with different f. But not surprisingly, the econometric specification is more adequate on the standard deviation basis or the log variance basis than it is on the variance basis. Below, we limit our attention to the standard deviation basis, i.e., $f(x) = \sqrt{x}$, which allows for a more intuitive interpretation of the results.

For comparison, we also consider the following regression models based on the time series of the realized measure:

$$\text{AR-RV:} \qquad f\left(\frac{RV_{t,t+h\Delta,\tau}}{h\Delta}\right) = a_{h,\tau} + b_{h,\tau} f\left(\frac{RV_{t-h\Delta,t,\tau}}{h\Delta}\right) + u^{ar}_{t+h\Delta,\tau}\,, \tag{5}$$

$$\text{HAR-RV:} \qquad f\left(\frac{RV_{t,t+h\Delta,\tau}}{h\Delta}\right) = a_{h,\tau} + \sum_{i=\{4,8,16,32\}} b_{h,\tau,i} f\left(\frac{RV_{t-i\Delta,t,\tau}}{i\Delta}\right) + u^{har}_{t+h\Delta,\tau} \tag{6}$$

The HAR-RV model is employed by AB (2010) and shown to have a superior performance. It is noted that the data of the first 32 weeks are not used for the regression of the HAR-RV model. To make the comparison equal, these data are not used for the regression of the other models.

3.2 In-sample results

We estimate equations (3)–(6) using the in-sample data. The in-sample forecasting performance is evaluated based on the adjusted R-squared coefficient, $\bar{R}^2$. Table 1 presents the results.[1] First, for the four-week forecasting horizon, $h = 4$, the $\bar{R}^2$ coefficients are high at short maturities but decrease with maturity. The result that the volatility of long-term yields is more difficult to explain than the volatility of short-term yields is consistent with the previous results; see AB(2010), and CDGJ (2009). Also of note is a small difference in the forecasting performance between the yield-based and RV-based regression models, although the HAR-RV model exhibits a slightly superior performance at both short and long maturities. Second, the increase in h has different impacts on the $\bar{R}^2$ coefficients between the yield-based and RV-based models. It is generally favorable for the former

[1] Extending the in sample to the full sample does not materially change the results on the relative ranking of the models.

but not for the latter, which is most evident at long maturities. At short maturities, the forecasting performance remains high for both types of models.

3.3 Out-of-sample results

High performance in-sample does not necessarily lead to high performance out-of-sample. We perform an out-of-sample analysis in two approaches. In the first approach, the parameter values of the regression models are fixed at the in-sample estimates throughout the out-of-sample period. This is aimed at examining whether the regression models can produce a long-run predictive relation. In the second approach, the parameters are estimated every time the prediction is made in a rolling window fashion, where the size of the data for the estimation is fixed at the same as the in-sample data. This is aimed at examining whether the current yield curve contains useful information for the volatility prediction. In both approaches, the out-of-sample forecasting performance is evaluated based on the root mean squared error (RMSE).

Table 2 presents the RMSEs expressed in basis points (bps, 1bp = 0.01%) in the fixed-parameter approach. Generally, the RV-based regression models, especially HAR-RV, have smaller RMSEs than the yield-based regression models. This is most evident at $\tau \leq 2$: the differences in the RMSE between the Y-ALL and HAR-RV models reach 12.4 bps for $h = 32$ at $\tau = 1$. The performance gap tends to narrow the longer the maturity and forecasting horizon. Nevertheless, the performance deterioration of the yield-based models is obvious, taking much higher $\bar{R}^2$ coefficients of the in-sample regression into consideration.[2] The results indicate overfitting of the yield-based models to the in-sample data. Once the parameter values are fixed, the models cannot respond to changes in the forecasting environment.

Table 3 presents the RMSEs in the varying-parameter approach. For $h = 4$, the HAR-RV model again exhibits the best performance, followed by the Y-ALL model. The RMSEs for the Y-PART model are the largest, however, the differences in the RMSE between the

[2] The high $\bar{R}^2$ coefficients of the yield-based models for long horizons seem to be caused by the high persistence of both the regressor and regressand. Interest rates on the right-hand side are known to be persistent. The realized volatility on the left-hand side becomes more persistent by extending the horizon since more overlapping daily observations are used in successive weekly observations of the realized volatility. For example, for $h = 32$, $RV_{t,t+32\Delta}$ and $RV_{t+\Delta,t+33\Delta}$ have overlapping observations over 31 weeks.

Y-PART and HAR-RV models do not exceed 5 bps. The extension of the forecasting horizon is again favorable for the yield-based models. Although the high performance of the HAR-RV model continues at $\tau \leq 2$, the performance of the yield-based models becomes comparable or superior at $\tau = \{3, 5, 10\}$.

Taken both the in-sample and out-of-sample results together, the current yield curve seems to fit the volatility realized in the future. But a simple linear combination of yields fails to produce a long-run predictive relation, which after all seems difficult to obtain without information on dynamic properties of the data. The time-varying parameter approach possibly allows a simple linear combination of yields to incorporate such information, however, it lacks consistency. This motivates us to use dynamic term structure models for predicting the volatility.

4 Model

The existing term structure models are known to have difficulties in matching both time-series and cross-sectional properties of the data. Therefore, the purpose of predicting the volatility through term structure models, while maintaining the cross-sectional fit, is originally difficult to achieve, and some breakthrough is required. In this section, we first review the earlier work attempting to capture the volatility in order to see why this purpose is difficult and how our solution is different. We then propose appropriate models. Since the models do not have a closed-form expression of bond prices, we rely on an approximation method.

4.1 A review of the volatility prediction using term structure models

A standard practice to extract volatility factors from the yield curve is to equate a term structure model with a particular set of observed yields and to solve this equation for latent factors including the volatility factors. In this inversion method, affine term structure models are frequently used since they provide a one-to-one correspondence between observed and latent variables in closed-form. To be more specific on the affine models, we use a notation introduced by Dai and Singleton (2000): $A_m(n)$ denotes an n-factor affine term structure model in which m factors affect the instantaneous covariance matrix of the n factors. To model stochastic volatility, $m = 1$ seems to be preferred as long as $n \leq 4$.

This is because the dynamics of the m factors are restricted to be in a non-negative region, which, however, restricts the correlation structure of the factors and the market prices of risks at the same time. In short, the more m, the more restrictions are necessary, which deteriorates the model performance.

Nevertheless, the combination of the inversion method and the $A_1(n)$ model does not work well. CDGJ (2009) and JK (2009) report that the volatility implied by the $A_1(3)$ model does not behave consistently with typical volatility measures in time series. This inconsistency indicates that it is difficult for the volatility factor to explain both the cross-section of yield levels and the time-series of yield changes. It is an outcome of the volatility factor playing as a cross-sectional factor, rather than a time-series factor. The increase in n or m does not seem to be a solution, as long as the inversion method is employed with the affine models.

A solution taken by the earlier work for reducing the tension on the volatility factor and avoiding the resulting inconsistency is to abandon the inversion method. Bikbov and Chernov (2011) use the extended Kalman filter to obtain the time series of the volatility factor. Thompson (2008) uses an estimation method where the volatility factor is integrated out from the likelihood function. A more extreme solution is to develop a term structure model where the volatility factors do not appear in the cross-section of bond prices (but do appear in the cross-section of derivative prices). Collin-Dufresne and Goldstein (2002) propose the so-called unspanned stochastic volatility (USV) model, which is a special case of the $A_m(n)$ model by placing certain parameter constraints.

Our solution is different from the previous ones. We maintain the inversion method, which has advantages of facilitating the factor identification as well as maintaining the cross-sectional fit with a relatively small number of factors. Since the combination of the inversion method and the existing models does not work well, we naturally abandon the existing models. Our view is that the tension on the volatility arises because some factors are presumed to be volatility factors. If instead such prespecification is not made, the tension and the resulting inconsistency will be naturally avoided. Our solution is thus to abandon the prespecification of any factor as the volatility factor. As a result, all factors are supposed to explain the cross-section of yields as the $A_0(n)$ model. This does not seem to be a bad idea, given a number of advantages of the $A_0(n)$ model reported by many studies. To overcome a major drawback of the $A_0(n)$ model that the volatility is constant,

we make the volatility of each factor level-dependent.

4.2 A baseline $A_0(3)$ model

We build our models based on the $A_0(3)$ model, and hence begin with its specification. Among numerous specifications of the $A_0(3)$ model depending on factor rotations, we adopt a classical one characterized by the mean-reverting behavior of the instantaneous risk-free rate; see, e.g., Andersen and Lund (1997b), Balduzzi et al. (1996, 1998), and Bikbov and Cherenov (2005). Let $X_t = (r_t\ \theta_t\ \epsilon_t)'$ be a three-dimensional state vector, and the risk-neutral distribution of the instantaneous change in X_t is given by

$$dX_t \sim N\left[\begin{pmatrix} \kappa_1(\theta_t - r_t) + \epsilon_t \\ \kappa_2(\bar{\theta} - \theta_t) \\ -\kappa_3\epsilon_t \end{pmatrix} dt\ ,\quad \Sigma\ dt\right]\ , \tag{7}$$

where

$$\Sigma = \begin{pmatrix} \sigma_{11} & \sigma_{12} & \sigma_{13} \\ \sigma_{12} & \sigma_{22} & \sigma_{23} \\ \sigma_{13} & \sigma_{23} & \sigma_{33} \end{pmatrix}\ . \tag{8}$$

In this specification, θ_t is presumed to be the central tendency, toward which the instantaneous risk-free rate r_t reverts. ϵ_t is presumed to be a shock that makes the mean-reverting behavior a bit irregular. In reality, however, we have to place the following constraint *a priori*, as pointed out by CDGJ (2008) and Joslin (2006), for each process to be interpreted as above: $0 < \kappa_2 < \kappa_1 < \kappa_3$. That is, θ_t also mean-reverts to some constant $\bar{\theta}$, but its speed is slower than that of r_t reverting to θ_t. The speed of mean reversion of ϵ_t is assume to be the fastest, aiming at making the shock temporary.

This classical specification is beneficial in introducing the level-dependent volatility in the next step. We can expect r_t and θ_t to be highly correlated with short-term and long-term yields, respectively, as is indeed the case *ex post*. In other words, the dynamics of both ends of the yield curve can be more naturally described. Because at least two of the factors are directly related to some yields and because a combination of yields has some information useful for the volatility prediction as seen in Section 3, the introduction of the level-dependent volatility becomes more effective within this classical specification.

To obtain the physical distribution of the instantaneous change in X_t for the purposes of estimation and prediction, the market prices of risks, $\Lambda(X_t)$, need to be specified.

Following the essentially affine specification proposed by Duffee (2002),

$$\Lambda(X_t) = \Sigma^{-1}(\Lambda^0 + \Lambda^1 X_t) \ . \tag{9}$$

Then,

$$dX_t \sim N\left[\left\{\begin{pmatrix} \kappa_1(\theta_t - r_t) + \epsilon_t \\ \kappa_2(\bar{\theta} - \theta_t) \\ -\kappa_3\epsilon_t \end{pmatrix} + \begin{pmatrix} \lambda_1^0 \\ \lambda_2^0 \\ \lambda_3^0 \end{pmatrix} + \begin{pmatrix} \lambda_{11}^1 & \lambda_{12}^1 & \lambda_{13}^1 \\ \lambda_{21}^1 & \lambda_{22}^1 & \lambda_{23}^1 \\ \lambda_{31}^1 & \lambda_{32}^1 & \lambda_{33}^1 \end{pmatrix} \begin{pmatrix} r_t \\ \theta_t \\ \epsilon_t \end{pmatrix}\right\} dt \ , \ \Sigma \ dt\right] . \tag{10}$$

4.3 Proposed models

Using $A_0(3)$ as a baseline model, we make the instantaneous covariance matrix of changes in X_t time-varying. But, first of all, it must be positive definite. We wish to meet this requirement without heavy parameterization. One possible way, similar in spirit to Fan, Gupta, and Ritchken (2003), Han (2007), Jarrow et al. (2007), and Longstaff, Santa-Clara, and Schwartz (2001), is to take the spectral decomposition of the covariance matrix, and then to specify the eigenvalues as functions of the state vector while holding the corresponding eigenvectors fixed. Specifically, the risk-neutral distribution of the instantaneous change in X_t is given by

$$dX_t \sim N\left[\begin{pmatrix} \kappa_1(\theta_t - r_t) + \epsilon_t \\ \kappa_2(\bar{\theta} - \theta_t) \\ -\kappa_3\epsilon_t \end{pmatrix} dt \ , \ \Sigma_t \ dt\right] \ , \tag{11}$$

where Σ_t is decomposed as

$$\Sigma_t = P L_t P' \ . \tag{12}$$

L_t is a diagonal matrix consisting of the eigenvalues, which are functions of the state vector:

$$L_t = \begin{pmatrix} L_1(X_t) & 0 & 0 \\ 0 & L_2(X_t) & 0 \\ 0 & 0 & L_3(X_t) \end{pmatrix} . \tag{13}$$

P is an orthogonal matrix having the corresponding eigenvectors of unit length in its columns. By the conditions of the orthogonality and unit length, the number of free

parameters in P is actually three. We then parameterize P using three rotation matrices:

$$P = \begin{pmatrix} 1 & 0 & 0 \\ 0 & \cos\varphi_3^P & -\sin\varphi_3^P \\ 0 & \sin\varphi_3^P & \cos\varphi_3^P \end{pmatrix} \begin{pmatrix} \cos\varphi_2^P & 0 & -\sin\varphi_2^P \\ 0 & 1 & 0 \\ \sin\varphi_2^P & 0 & \cos\varphi_2^P \end{pmatrix} \begin{pmatrix} \cos\varphi_1^P & -\sin\varphi_1^P & 0 \\ \sin\varphi_1^P & \cos\varphi_1^P & 0 \\ 0 & 0 & 1 \end{pmatrix} . \quad (14)$$

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 $\cos\varphi_i^P = \sqrt{1 - \sin^2\varphi_i^P}$.

For Σ_t to be positive definite, all eigenvalues must be positive, i.e., $L_i(X_t) > 0$ for any X_t $(i = 1, 2, 3)$. Taking this into consideration, we propose two models for the time-varying eigenvalues.

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

$$L_i(X_t) = c_i + X_t'\Gamma^i X_t \quad (i = 1, 2, 3) , \quad (15)$$

where Γ^i is either a positive definite matrix if $c_i \geq 0$ or a nonnegative definite matrix if $c_i > 0$. In the estimation, we impose the latter restriction on Γ^i and c_i as this can lead to a more parsimonious specification. For example, $\Gamma^i = \mathbf{0}$ 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:

$$\Gamma^i = Q^i M^i Q^{i\prime} \quad (i = 1, 2, 3) , \quad (16)$$

where

$$M^i = \begin{pmatrix} m_1^i & 0 & 0 \\ 0 & m_2^i & 0 \\ 0 & 0 & m_3^i \end{pmatrix} , \qquad \text{with} \quad 0 \leq m_1^i \leq m_2^i \leq m_3^i , \quad (17)$$

and

$$Q^i = \begin{pmatrix} 1 & 0 & 0 \\ 0 & \cos\varphi_3^{Q^i} & -\sin\varphi_3^{Q^i} \\ 0 & \sin\varphi_3^{Q^i} & \cos\varphi_3^{Q^i} \end{pmatrix} \begin{pmatrix} \cos\varphi_2^{Q^i} & 0 & -\sin\varphi_2^{Q^i} \\ 0 & 1 & 0 \\ \sin\varphi_2^{Q^i} & 0 & \cos\varphi_2^{Q^i} \end{pmatrix} \begin{pmatrix} \cos\varphi_1^{Q^i} & -\sin\varphi_1^{Q^i} & 0 \\ \sin\varphi_1^{Q^i} & \cos\varphi_1^{Q^i} & 0 \\ 0 & 0 & 1 \end{pmatrix} , \quad (18)$$

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 all 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

$$L_i(X_t) = \exp\left\{c_i + \gamma^{i\prime} X_t\right\} \quad (i = 1, 2, 3) \ . \tag{19}$$

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

Since Σ_t is designed to be positive definite, the inverse matrix, Σ_t^{-1}, is well defined. This is extremely important for modeling the market prices of risks. That is, they can be specified in the same way as those for the $A_0(3)$ model, i.e., equation (9) with Σ replaced by Σ_t. Likewise, the physical distribution of the instantaneous change in X_t is also obtained by substituting Σ_t for Σ in (10). Therefore, the entire drift parameters can virtually be replaced with new ones in changing the risk-neutral measure to the physical measure as in the $A_0(3)$ model. Consequently, we expect that the proposed models inherit a number of desirable properties from the $A_0(3)$ model, which is verified in the next section.

At the same time, there are also a number of drawbacks in the proposed models. Most critically, the models do not accommodate such factors as affecting derivative prices but not underlying bond prices. We discuss this issue and present a possible remedy in Section 5.1. Another drawback is that the models generate negative interest rates with positive probability as the original $A_0(3)$ model. This problem, however, can actually be avoided. As originally proposed by Aït-Sahalia (1996), r_t, starting from $r_0 > 0$, does not reach zero in finite time by adding a reciprocal term to the drift of r_t, i.e., by considering $a_{-1}/r_t + \kappa_1(\theta_t - r_t) + \epsilon_t$ with $a_{-1} > 0$. If a sufficiently small value is given to a_{-1}, say 10^{-8}, the impact of the reciprocal term is negligible in the observed range of the data. Therefore, while the empirical results shown below are based on the models without the reciprocal term, virtually identical results will be obtained using the models with the reciprocal term.

4.4 Bond price and its approximation

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

$$P(X_t, \tau) = E_t^Q\left[\exp\left\{-\int_t^{t+\tau} r_u du\right\}\right] \ , \tag{20}$$

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

The proposed models do not have a closed-form expression of $P(X_t, \tau)$ due to the volatility specifications. To make the estimation of the models and the extraction of the latent state variables feasible, we approximate $P(X_t, \tau)$ by relying on a method proposed by Shoji (2002), and Takamizawa and Shoji (2009). This method approximates a vector of conditional moments as a solution to the ODE. Since the zero-coupon bond price is given as the conditional expectation, this method can be directly applied. We actually use the third-order approximation. The outline and application of this method are presented in Appendix B, and the accuracy of the approximation is checked in Appendix C. In short, the method seems to work at least for pricing bonds with maturities of up to ten years when realistic values of parameter and state vectors are provided.

4.5 Estimation method

We estimate the models with the quasi-maximum likelihood method, which is one of the standard methods for estimating term structure models; see, e.g., CDGJ (2008), Chen and Scott (1993), Duffee (2002), and Pearson and Sun (1994). The zero-yields with maturities of 0.5, 1, 2, 3, 5, and 10 years are used. Among them, those with maturities of 0.5, 2, and 10 years are assumed to be explained exactly by the models in order to back out the latent state variables. Specifically, let Θ be a parameter vector to be estimated, and let $\Upsilon(X_t; \Theta)$ be a three-dimensional vector consisting of the approximate yield functions:

$$\Upsilon(X_t; \Theta) = \left(\tilde{Y}(X_t, 0.5; \Theta) \quad \tilde{Y}(X_t, 2; \Theta) \quad \tilde{Y}(X_t, 10; \Theta) \right)' , \tag{21}$$

where $\tilde{Y}$ is for clarifying the approximation. Then, X_t is obtained by solving $Y_t^p = \Upsilon(X_t; \Theta)$. This is done numerically, however, only a few iterations are sufficient if a good initial value of X_t is given: it is actually the value of X_t implied by the $A_0(3)$ model.

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

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

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

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

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

The first equality is from the changes of variables from Y_t^p to X_t, by which the Jacobian term appears, and from Y_t^e to U_t. The second equality is from the independence between U_t and X_s for any s. For both the SV-Q and SV-E models, the transition density, f_T, has no analytical expression for finite Δ. We approximate it by the multivariate normal density function, which might be justified by a relatively short interval, $\Delta = 1/52$. The conditional first and second moments to be substituted are computed with the same approximation method. It is noted that for the SV-Q model, these moments can be computed exactly since the drift terms are linear and instantaneous (co)variances are quadratic in the state variables. The Jacobian term is already computed in backing out the latent state variables. On the other hand, f_C is the multivariate normal density function from (22). The objective function for estimating Θ is then

$$\sum_t \left\{ \ln f_T(X_t | X_{t-\Delta}; \Theta) + \ln f_C(U_t; \Theta) - \ln \left| \frac{d\Upsilon(X_t;\Theta)}{dX_t'} \right| \right\} . \tag{24}$$

5 Main Empirical Analysis

Parameter estimates and time series of implied state variables are first reported. We then examine whether the proposed models can reproduce some important properties of the actual data: the failure of the EHT and a hump shape of unconditional volatilities of changes in logarithmic yields. Furthermore, we examine the predictive power of the proposed models to the level of yields. After confirming that the proposed models have at least a comparable performance to the $A_0(3)$ model in these dimensions, we finally proceed to examining whether the predictive power to the volatility is improved.

5.1 Parameter estimates and implied state variables

We first report the parameter estimates for the baseline $A_0(3)$ model in Table 4. It is noted that when the full parameters are estimated, some parameters, particularly those of the market prices of risks, are insignificant. We set such parameters to zero, and re-estimate

the model until all t-values exceed 1.5 in absolute value. As a result, the instantaneous covariance between θ_t and ϵ_t is zero: $\sigma_{23} = 0$. Also, θ_t alone contributes to the variation in the market prices of risks: the first and third columns of Λ^1 are all zeros. This result is consistent with Joslin, Singleton, and Zhu (2011, Table 5), where the constraint of $\text{rank}(\Lambda^1) = 1$ is not rejected. The speed of mean reversion, captured by the estimate of κ_i, is by far the lowest for θ_t. It is similar between r_t and ϵ_t, suggesting that the three factors are sufficient for describing the cross-section of yields. This is also suggested by a relatively small standard deviation of the measurement errors, 6.1 bps.

Figure 2 exhibits the time series of the state variables implied by the $A_0(3)$ model. As is indeed anticipated, the dynamics of r_t and θ_t are very closely tracked by those of the six-month and ten-year yields, respectively. It is so even in the out-of-sample period, implying that there is no structural break between the samples. ϵ_t ranges between -0.07 and 0.04, and looks more volatile. This is confirmed by the estimate of σ_{33} shown in Table 4, which is much larger than the estimates of σ_{11} and σ_{22}. In sum, the $A_0(3)$ model is shown to be appropriately specified in that the *ex post* dynamics of the state variables are consistent with the *ex ante* roles.

Tables 5 and 6 present the parameter estimates for the SV-Q and SV-E models, respectively. Similar to the estimation of the $A_0(3)$ model, insignificant parameters are set to zero in the final estimation.[3] It is also noted that only the same parameters of $\Lambda(X_t)$ as those for the $A_0(3)$ model are estimated, which is aimed at distinguishing the models solely by the specification of the covariance matrix. We notice that the estimates of the risk-neutral drift are very similar across the three models. For both the SV-Q and SV-E models, the estimate of $\sin\varphi_2^P$ alone is significant in the eigenvector matrix P. As a result, the instantaneous covariance matrix for these models is

$$\Sigma_t = \begin{pmatrix} L_1(X_t)\cos^2\varphi_2^P + L_3(X_t)\sin^2\varphi_2^P & 0 & \{L_1(X_t) - L_3(X_t)\}\sin\varphi_2^P\cos\varphi_2^P \\ 0 & L_2(X_t) & 0 \\ \{L_1(X_t) - L_3(X_t)\}\sin\varphi_2^P\cos\varphi_2^P & 0 & L_1(X_t)\sin^2\varphi_2^P + L_3(X_t)\cos^2\varphi_2^P \end{pmatrix}. \quad (25)$$

It turns out that $L_2(X_t)$ is the instantaneous variance of θ_t. In addition, because $\sin\varphi_2^P < \cos\varphi_2^P$, $L_1(X_t)$ is more closely related to the instantaneous variance of r_t, whereas $L_3(X_t)$ is to the instantaneous variance of ϵ_t. Among the instantaneous covariances, only that

[3] As exceptions, for the SV-Q model, c_1 is set at a very small number, 10^{-8}, due to the constraint of $c_1 > 0$, and $\sin\varphi_3^{Q^2} = -\sin\varphi_2^{Q^2}$.

between r_t and ϵ_t is significant.

The estimates of ς, the standard deviation of the measurement errors, for the SV-Q and SV-E models are 6.1 bps, which are little changed from that for the $A_0(3)$ model. The result reconfirms that as far as the description of the cross-section of yields, the $A_0(3)$ model has already done a sufficient job. Time-varying volatility does not seem to be effective in this dimension. It is, however, in the time-series dimension of the data. Indeed, compared with the $A_0(3)$ model, the log-likelihood value for the SV-Q model is increased by 152 with additional 10 parameters and that for the SV-E model is increased by 95 with additional 7 parameters.

Figure 3 exhibits the time series of the state variables implied by the SV-Q (left panels) and SV-E (right panels) models. For each state variable, two implied series are visually indistinguishable. Furthermore, they are almost identical to that implied by the $A_0(3)$ model shown in Figure 2. The result also indicates that the specification of the instantaneous covariance matrix little affects the behavior of the state variables extracted from the cross-section of yields, given the identical specification of the risk-neutral drift. In other words, it is the risk-neutral drift that matters with the identification of the state variables.

But if this is indeed the case, the following question naturally arises: can a state variable be identified from the cross-section of yields that appears in the instantaneous covariance matrix but not in the risk-neutral drift? If not, this state variable would virtually act as an unspanned factor. Indeed, a similar argument is made by Joslin (2010) using affine term structure models. [4] In addition, in CDGJ (2009, p.51), the sufficient conditions to obtain the $A_1(4)$ USV model from the $A_1(4)$ model are presented, one of which indicates that the coefficient of a volatility factor in the risk-neutral drift term of a conditionally Gaussian factor should be very small. In fact, their Table 2A shows that the estimate of this coefficient for the $A_1(4)$ USV model is -0.260, which is in sharp contrast

[4] Specifically, the convexity arising from the Jensen's inequality can be defined as

$$CV(X_t, \tau) \equiv \frac{1}{\tau} \int_t^{t+\tau} E_t^Q[r(X_u)]du - Y(X_t, \tau) \ .$$

But Joslin (2010) showed that both the level and variation of $CV(X_t, \tau)$ are small even at $\tau = 10$. This indicates that $Y(X_t, \tau)$ is mostly determined by $\int_t^{t+\tau} E_t^Q[r(X_u)]du/\tau$ and that this relation is stable over time. Then, if the volatility factors are not included in $E_t^Q[r(X_u)]$, they cannot have an impact on $Y(X_t, \tau)$ large enough to be backed out from $Y(X_t, \tau)$ with some precision.

with the corresponding estimate of -809.5 for the $A_1(4)$ model obtained by the inversion method. Other parameters in the risk-neutral drift terms do not exhibit such a dramatic change. The result indicates that by this restriction alone the volatility factor becomes nearly, if not completely, unspanned. A similar result is reported by Thompson (2008, Tables 3 and 4). Taken together, the introduction of a nearly unspanned factor does not seem as difficult as previously thought: simply exclude a factor from the risk-neutral drift terms.

5.2 Ability of the models to replicate real properties of the data

We have seen many similarities between the $A_0(3)$ and proposed models. We then expect that the advantages of the $A_0(3)$ model are inherited into the SV-Q and SV-E models. We verify this expectation. Specifically, we examine whether these models can replicate the failure of the EHT and the hump shape of unconditional volatilities of changes in logarithmic yields.

5.2.1 Failure of the EHT

Following Campbell and Shiller (1991) but using a slightly different notation, we consider the following regression model:

$$y_{t+\tau_1,\tau_2-\tau_1} - y_{t,\tau_2} = \phi^0_{\tau_2} + \phi^1_{\tau_2}\frac{\tau_1}{\tau_2-\tau_1}(y_{t,\tau_2} - y_{t,\tau_1}) + u_{t+\tau_1}\ , \tag{26}$$

where τ_1 and τ_2 are measured in years and τ_2/τ_1 is integer. Since we use weekly data, we set $\tau_1 = \Delta\ (= 1/52)$. But, this requires the one-week yield. Instead of generating the one-week yield from the original LIBOR and swap curves by extrapolation, we use as a proxy the six-month yield. This simplification might be justified because our purpose is to examine whether the proposed models can replicate an empirical pattern rather than to test the EHT rigorously. Also, we replace $y_{t+\Delta,\tau_2-\Delta}$ on the left-hand side of equation (26) with $y_{t+\Delta,\tau_2}$ for convenience. We consider $\tau_2 = \{1, 2, 3, 5, 10\}$.

It is know that the actual estimates of $\phi^1_{\tau_2}$ tend to be negative, and more so with increasing τ_2, contrary to the EHT that $\phi^1_{\tau_2} = 1$ holds for all τ_2. We estimate equation (26) using both actual and simulated data. For each model, we simulate 1,000 sets of data with the same sample size as the actual in-sample data. Precisely, the initial realization is drawn randomly from the actual data. The subsequent realizations are generated from

the physical distribution of changes in X_t given by (10): for the SV-Q and SV-E models, Σ is replaced by Σ_t. For the SV-Q and SV-E models, since analytical expressions of the distribution over a finite interval are unavailable, we use a discretized version of the distribution, where dt is replaced by $\Delta t = 1/1,000$. We then obtain 1,000 estimates of $\phi^1_{\tau_2}$, and record the average, and the 5th and 95th percentiles of these estimates.

The left panels of Figure 4 exhibit the actual (the solid line) and model-implied (the marked line) estimates of $\phi^1_{\tau_2}$ together with the 90% confidence band (the dotted line). While the extent of the failure of the EHT is weaker for the simulated data than for the actual data, all three models have a similar pattern, indicating that the proposed models can regenerate the failure at least as well as the $A_0(3)$ model.

5.2.2 Hump shape of unconditional volatilities

The right panels of Figure 4 exhibit the actual (the solid line) and model-implied (the marked line) estimates of unconditional volatilities (annualized standard deviations) of changes in logarithmic yields with the 90% confidence band (the dotted line). Again, all models can regenerate a hump shape that is close to the actual one.

5.3 Predictive power to the level of yields

Duffee (2002) shows that the essentially $A_0(3)$ model having the market prices of risks given by equation (9) has a superior predictive power to the level of yields than the other essentially $A_m(3)$ models with $m > 0$. We examine whether the proposed models have at least a comparable ability in this dimension. To compute the conditional expectation of the h-week ahead τ-maturity yield for the proposed models, $E_t[\tilde{Y}(X_{t+h\Delta}, \tau)]$, we also employ the approximation method of conditional moments: the details of the application are provided in Appendix B and the accuracy analysis is in Appendix C. To match the forecasting horizons with those for the volatility, we also consider $h = \{4, 8, 16, 32\}$.

Table 7 presents the in-sample RMSEs in basis points, where the prediction error is defined as the difference between predicted and realized yields. As a benchmark, we provide the results for random walk (RW), whose forecast is the current yield for any h. The forecasts of the three term structure models are generally better than those of RW, except for $h = \{8, 16\}$ at $\tau = 1$. Among the term structure models, the performance is similar, though the RMSEs for the SV-Q model tend to be slightly larger.

Table 8 presents the out-of-sample RMSEs in basis points. Generally, the three term structure models have smaller RMSEs than RW. At $\tau = 10$, however, RW is comparable for $h \leq 16$ and superior for $h = 32$ to the term structure models. The result indicates the difficulty of predicting the ten-year yield in terms of both the level and volatility. Among the term structure models, while the proposed models are slightly better at short maturities, the performance is generally similar, confirming that the ability of the proposed models to predict the level of yields is maintained.

5.4 Predictive power to the volatility of yields

We consider several competing models. First, we select the $A_0(3)$ and $A_1(3)$ models. Since, given h and τ, the $A_0(3)$ model produces a constant forecast independently from the current state, it serves as a benchmark, as does RW for the level prediction. As for the specification of the $A_1(3)$ model, we select one proposed by CDGJ (2008, p.762). The $A_1(3)$ model is estimated in the same way as the other term structure models.

Second, we consider a time-series model applied to the same weekly data. 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 necessary beat the simplest one in the out-of-sample period; see Hansen and Lunde (2005). The model is fitted to each yield, $y_{t,\tau}$:

$$y_{t+\Delta,\tau} = \alpha + \beta y_{t,\tau} + \sqrt{V_{t+\Delta,\tau}}\, z_{t+\Delta,\tau} \,, \tag{27}$$

$$V_{t+\Delta,\tau} = \omega_0 + \omega_1 V_{t,\tau} z_{t,\tau}^2 + \omega_2 V_{t,\tau} \,, \tag{28}$$

where it is assumed for simplicity that $z_{t,\tau} \sim i.i.d.N(0,1)$. The h-week ahead conditional variance, $\mathrm{var}_t[y_{t+h\Delta,\tau}]$, is obtained by iteration, an explanation of which is provided in Appendix D.

Third, we consider time-series models applied directly to the daily data. We select the HAR-RV model presented in Section 3 and the mixed data sampling (MIDAS) approach developed by Ghysels et al. (2005, 2006). The MIDAS regression in this paper is as follows:

$$\text{MIDAS:} \quad \sqrt{\frac{RV_{t,t+h\Delta,\tau}}{h\Delta}} = a_{h,\tau} + b_{h,\tau}\sqrt{\frac{\sum_{i=1}^{n_t}(y_{t-L\Delta(i-1)/n_t,\tau} - y_{t-L\Delta i/n_t,\tau})^2 w_{h,\tau}(i)}{(L\Delta/n_t)}} + u^m_{t+h\Delta,\tau} \,, \tag{29}$$

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

$$w_{h,\tau}(i) = \frac{\exp\{w_{h,\tau,1}i + w_{h,\tau,2}i^2\}}{\sum_{k=1}^{n_t}\exp\{w_{h,\tau,1}k + w_{h,\tau,2}k^2\}} , \tag{30}$$

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 HAR-RV and MIDAS. The difference between the two is in the weighting structure to past observations.

The predictive power to the volatility is evaluated by the RMSE criterion using both the in-sample and out-of-sample data. The RMSE for the term structure and GARCH models is computed based on the following equations:

$$\sqrt{\frac{RV_{t,t+h\Delta,\tau}}{h\Delta}} = forecast_{t,h,\tau} + u^1_{t+h\Delta,\tau} , \tag{31}$$

$$\sqrt{\frac{RV_{t,t+h\Delta,\tau}}{h\Delta}} = a_{h,\tau} + b_{h,\tau}forecast_{t,h,\tau} + u^2_{t+h\Delta,\tau} , \tag{32}$$

where $forecast_{t,h,\tau}$ is a model forecast at time t for the h-week ahead volatility of the τ-maturity yield. Some explanation is needed as to how to compute $forecast_{t,h,\tau}$ for the proposed models. First, while we assumed that the yields not used for backing out the factors are measured with error in estimating the models, we omit the variance of this error in computing the conditional variance of these yields, as it is a small constant. That is, we assume here that $y_{t,\tau} = \tilde{Y}(X_t, \tau)$ for all τ. Second, since $forecast_{t,h,\tau} = \sqrt{\text{var}_t[\tilde{Y}(X_t, \tau)]}$ has no closed-form expression, we employ the approximation method of conditional moments again; see Appendices B and C for the application and the accuracy, respectively.

For the HAR-RV and MIDAS regressions, the RMSE is computed from equations (6) and (29), respectively.

As in Section 3, the out-of-sample predictive power is evaluated in both the fixed- and varying-parameter approaches. The out-of-sample RMSE in the fixed-parameter approach is computed with the parameter values held fixed at the in-sample estimates throughout the out-of-sample period. The out-of-sample RMSE in the varying-parameter approach is computed as follows. First, for the term structure and GARCH models, the model parameters are fixed at the in-sample estimates. The parameters that vary every time the prediction is made are $(a_{h,\tau},\ b_{h,\tau})$ in equation (32). Second, for the MIDAS regression, $(w_{h,\tau,1},\ w_{h,\tau,2})$ in equation (30) are fixed at the in-sample estimates, while $(a_{h,\tau},\ b_{h,\tau})$ in

equation (29) are re-estimated. Finally, for the HAR-RV regression (6), all parameters are re-estimated as before.

It is noted that the above comparison scheme is challenging to the term structure models. There is an information gap between the term structure and GARCH models. While each yield is used for estimating the GARCH model, only three yields, the six-month, two-year, and ten-year yields, are used for estimating time-series properties of the term structure models. The information gap widens further against the HAR-RV and MIDAS regressions, which are constructed directly from the realized volatility.

5.4.1 In-sample results

To obtain an intuition about the model performance, we first look at Figures 5–8, where the time series of the four-week ahead forecast of the volatility (annualized standard deviation) of the six-month, two-year, and ten-year yields are displayed together with the corresponding realized measure. In Figure 5, the GARCH forecast tracks the realized measure well with the correlation coefficients being around 0.6, though it tends to overestimate the level of the volatility of the six-month yield and underestimate the variation of the volatility of the ten-year yield. In Figure 6, the $A_1(3)$ forecast fluctuates little. Besides, the correlations between the predicted and realized series are low: for the ten-year yield, it is -0.19. The results are consistent with those reported by CDGJ (2008, 2009), and JK (2009). In Figure 7, the SV-Q forecast is apparently better than the $A_1(3)$ forecast with the correlation coefficients being around 0.5, yet it does not truck the realized measure as closely as the GARCH forecast. More specifically, the forecast is occasionally wrong especially during 2002–04, however, it can remarkably capture the sharp rise and fall in the realized measure observed during 2008–09 even though this period is out-of-sample. At $\tau = 10$, while the correlation remains reasonably high, the variability of the model forecast is much smaller than that of the realized measure. Finally, in Figure 8, the SV-E forecast is also improved from the $A_1(3)$ forecast, but it does not track the realized measure as closely as the GARCH forecast, either. Compared with the SV-Q forecast, the correlation at $\tau = 0.5$ is slightly higher but those at $\tau = \{2, 10\}$ are lower. Still, the SV-E forecast at $\tau = 10$ appears to track the trend of the realized measure better than the SV-Q forecast. More generally, the SV-E forecast appears to be more persistent. It produces neither large deviations nor intensive variations. In fact, the serial correlations for the SV-E forecast

are all around 0.99, which are larger than those for the SV-Q forecast, around 0.95.

Having these figures in mind, we compare the in-sample model performance. Table 9 presents the RMSEs in basis points. For the term structure and GARCH models, the name alone indicates the results based on equation (31), whereas the name "+ Reg" indicates the results based on equation (32). Note that there is no $A_0(3)$ + Reg, as $forecast_{t,h,\tau}$ is constant. In reporting the results, we focus on the comparison first without then with the forecasting regression (32).

For the four-week forecasting horizon, $h = 4$, the proposed models predict the volatility better than the affine models, except at $\tau = 10$, where the benchmark $A_0(3)$ model has a smaller RMSE. Compared with the GARCH model, the RMSEs for the proposed models are larger, except for the SV-E model at $\tau = 0.5$. The proposed models are also outperformed by the HAR-RV and MIDAS regressions especially at short maturities $\tau \leq 1$. However, the performance gap narrows with increasing τ. Even with the forecasting regression (32), the RMSEs for the proposed models are larger than those for the HAR-RV and MIDAS regressions. But the differences in the RMSE are reduced to around 1 bp at middle to long maturities. The performance of the GARCH + Reg model is almost indistinguishable from that of the HAR-RV and MIDAS regressions.

By extending the forecasting horizon, the relative performance of the proposed models improves. While at $\tau \leq 1$, the HAR-RV and MIDAS regressions continuously exhibit a high performance, the performance of the proposed models becomes comparable for $h = \{16, 32\}$ at $\tau = \{2, 3, 5\}$. With the forecasting regression, this holds at all τ. In fact, the RMSEs for the SV-Q + Reg model are the smallest for $h = \{8, 16, 32\}$ at $\tau = \{2, 3\}$, and those for the SV-E + Reg model are the smallest for $h = \{16, 32\}$ at $\tau = \{5, 10\}$.

5.4.2 Out-of-sample results

We compare the out-of-sample model performance first in the fixed-parameter approach, where the parameter values are fixed at the in-sample estimates throughout the out-of-sample period. Table 10 presents the RMSEs in basis points. For $h = 4$, the SV-E model outperforms the affine models for all τ, whereas the SV-Q model fails to outperform the $A_1(3)$ model at $\tau \leq 1$. However, the SV-Q model improves with increasing τ. In fact, it has the smallest RMSE at $\tau = 10$ among all models. Compared with the GARCH model, the proposed models have a comparable or superior performance at both ends of the maturity

spectrum but are worse at in-between maturities. They are outperformed by the HAR-RV and MIDAS regressions overwhelmingly at short maturities but become comparable at $\tau = 10$. With the forecasting regression, the performance of the proposed models becomes better at $\tau \leq 1$ but worse at $\tau = \{3, 5, 10\}$ than it is without the forecasting regression. Since the decrease in the RMSE at $\tau \leq 1$ is insufficient, the proposed models with the forecasting regression are outperformed by the HAR-RV and MIDAS regressions at all τ.

By extending the horizon, while the HAR-RV and MIDAS regressions continuously exhibit a superior performance at $\tau \leq 1$, the relative performance of the proposed models improves at middle to long maturities, as is the case for the in-sample prediction. For $h = 8$, the SV-Q model is comparable or superior to the HAR-RV and MIDAS regressions at $\tau = \{3, 5, 10\}$, although it is worse than the $A_1(3)$ model at $\tau \leq 1$. The SV-E model also outperform the high frequency regressions at $\tau = 10$. This pattern is maintained, but somewhat strengthened, for $h = \{16, 32\}$. As a result, the SV-Q model exhibits a comparable performance in both sample periods to the time-series models for $h = \{8, 16, 32\}$ at $\tau = \{3, 5, 10\}$. The forecasting regression again increases the out-of-sample RMSEs for the proposed models at middle to long maturities. But the degree of increase is relatively minor. As a result, the SV-Q + Reg model is comparable in both sample periods to the time-series models for $h = \{16, 32\}$ at $\tau = \{2, 3, 5, 10\}$. Likewise, the SV-E + Reg model is comparable for $h = \{16, 32\}$ at $\tau = 10$.

Next, we compare the out-of-sample model performance in the varying-parameter approach, where the regression parameters are re-estimated every time the prediction is made. Table 11 presents the RMSEs in basis points. For $h = 4$, the SV-Q model outperforms the $A_1(3)$ model at all τ, whereas the SV-E model fails to do so at $\tau \leq 1$. The proposed models are outperformed by the time-series models, however. In contrast with the previous performance criteria, the relative performance of the proposed models does not much improve by extending the horizon. This is partly because the performance of the HAR-RV regression remains high due to the advantage of more varying parameters: it has five varying parameters, not two like the other models. Still, there are cases in which the differences in the RMSE between the SV-Q and HAR-RV models are within 2 bps: for $h = 8$ at $\tau = \{3, 10\}$ and for $h = \{16, 32\}$ at $\tau = 3$. Compared with the GARCH model, the proposed models are outperformed but catch up at $\tau = 10$.

5.4.3 A possible economic reasoning for the results

Why is the predictive power of the time-series models especially high for short forecasting horizons at short maturities and why does the predictive power of the proposed models relatively improve for long forecasting horizons at middle to long maturities?

The short-term yields are closely related to policy interest rates controlled by the central bank. Usually, the policy rates do not change frequently in a short period of time. Even if they do, they rather move in an anticipated direction. Therefore, the short-term yields can be more easily predicted, in terms of both the level and volatility, using information on the past state of the economy. The time-series models are good at exploiting such information, and thus exhibit a superior forecasting performance for short horizons at short maturities.

On the other hand, the role of the expectation about the future state of the economy increases not only in making longer-term forecasts but also in pricing longer-term bonds. By the specifications of the covariance matrix, Σ_t, the volatility forecast of the proposed models depends on the current yield curve, which by nature contains forward-looking information. Thus, an adequately specified term structure model has potential to have a relatively high predictive power for long forecasting horizons at middle to long maturities.

These results suggest that a combination of the two different types of forecasts, one from the time-series models and the other from the term structure models, may further improve the predictive power. This possibility is explored in the next subsection. Also of interest is whether the non-arbitrage constraint, given specifically by equation (20), has some contribution to the predictive power, beyond the specifications of Σ_t. This question is addressed in Section 5.6.

5.5 Combined forecasts

We make a combined forecast based on the HAR-RV model, which has a simple structure and exhibits a superior performance especially at short maturities:

$$\sqrt{\frac{RV_{t,t+h\Delta,\tau}}{h\Delta}} = a_{h,\tau} + \sum_{i=\{4,8,16,32\}} b_{h,\tau,i}\sqrt{\frac{RV_{t-i\Delta,t,\tau}}{i\Delta}} + c_{h,\tau} forecast_{t,h,\tau} + u^c_{t+h\Delta,\tau} \ , \quad (33)$$

where $forecast_{t,h,\tau}$ is taken from the term structure and GARCH models. Then, we examine the sign, magnitude, and significance of the estimate of $c_{h,\tau}$ using the in-sample

data. Furthermore, we examine the extent to which the combined forecast reduces the out-of-sample RMSE relative to the HAR-RV forecast alone.

Table 12 presents the estimates of $c_{h,\tau}$ in equation (33) together with the standard errors in parenthesis computed using the method of Newey and West (1987) with 32 lags. First, for the four-week forecasting horizon, $h = 4$, the slope estimates for the SV-Q forecast lie in a relatively narrow range from 0.36 to 0.47, which are all significantly different from zero but not from 0.5. Those for the SV-E forecast are also all significant with the range of estimates somewhat wider. The GARCH forecast has significant estimates except at $\tau = \{0.5, 2\}$. The result indicates that incremental information of the GARCH forecast is relatively small. This is not surprising given the similarity between the HAR-RV and GARCH models in the way of using time-series information to make forecasts. The slope estimates for the $A_1(3)$ forecast are significant only at $\tau \leq 1$. Though insignificant, the estimates at $\tau = \{5, 10\}$ are negative. A similar pattern is observed for longer horizons, with a notable exception that the estimates for the proposed models at $\tau = 10$ tend to be less significant.

Table 13 presents the % changes in the out-of-sample RMSE in the fixed-parameter approach. Of particular attention is the success of the HAR-RV + SV-Q forecast. This combined forecast reduces the out-of-sample RMSEs at middle to long maturities for all h, confirming an incremental value of the SV-Q forecast. Adding the GARCH forecast also reduces the RMSEs at middle to long maturities, however, the magnitude of reduction is less than that achieved by the SV-Q forecast. On the other hand, adding the SV-E forecast worsens the out-of-sample performance except at $\tau = 10$. Adding the $A_1(3)$ forecast generally deteriorates the performance with the increase in the RMSE reaching 30% for $h = 32$ at $\tau = 1$.

Table 14 presents the % changes in the out-of-sample RMSE in the varying-parameter approach. The overall picture agrees with that in Table 13. Again, the performance of the HAR-RV + SV-Q forecast is remarkable. In all cases except for $h = 32$ at $\tau = 0.5$, it reduces the RMSEs by up to 16%. Adding the SV-E forecast, however, leads to the worst out-of-sample performance in this criterion: the RMSEs increase in all cases.

Taken together, the combined forecast based on the HAR-RV model works especially at middle to long maturities, as expected. But the choice of the other model matters. In our sample, the SV-Q model is an appropriate one. The GARCH model also contributes

to the reduction in the RMSE, but not as much as the SV-Q model.

5.6 Is the non-arbitrage relevant with the volatility prediction?

Recently, Joslin et al. (2011) have proposed a canonical representation of the Gaussian term structure model in which linear combinations of observed yields serve as underlying factors. Since there is no need to back out the factors from the cross-section of yields, the physical drift terms of the factors, which involve the prediction of the level of factors, are free of any cross-sectional constraint including the non-arbitrage. In predicting the level of the other yields not used for making factors, some cross-sectional constraint relating the factors to these yields is required. But Duffee (2011) shows that a simple linear regression is sufficient for constructing this relation, so that the more complicated non-arbitrage condition is unnecessary for the level prediction of both yields used and not used for making factors.

The above studies have shed new light on the relation between the role of non-arbitrage and the level prediction of yields, using the Gaussian term structure model. However, less clear is whether this relation also holds for the prediction of the volatility that changes stochastically. Since the covariance matrix, Σ_t, is invariant to measure changes, it is related to both time-series and cross-sectional properties of the yield curve once the non-arbitrage is imposed, though the relation to the latter may be weak as argued in Section 5.1. Therefore, the estimate of Σ_t with the non-arbitrage might differ from that without the non-arbitrage obtained using time-series properties of the data alone, which would produce the difference in the volatility prediction between with and without the non-arbitrage.[5]

To examine the impact of non-arbitrage on the volatility prediction, we first clarify the role of non-arbitrage in this particular context:

(i) description of the cross-sectional relation between yields,

(ii) estimation of Σ_t,

(iii) identification of factors.

According to these roles, we control the degree of impact in the following three stages. In the first stage, we take (ii) and (iii) as given, and remove (i). Specifically, we use the

[5] This is the case for options data on stocks and currencies; see, e.g., Bakshi, Cao, and Chen (1996), and Bates (1996, 2000).

time-series of the state vector, X_t, extracted through the SV-Q and SV-E models with the parameter values fixed at the estimates presented in Tables 5 and 6, respectively. The cross-sectional relation is described by a linear regression as

$$y_{t,\tau} = \delta_{0,\tau} + \delta'_{1,\tau} X_t + u^{cross}_{t,\tau} \ . \tag{34}$$

The coefficients, $(\delta_{0,\tau}\ \delta'_{1,\tau})$, are estimated by OLS using the in-sample data. These estimates are fixed throughout the out-of-sample period to maintain the consistency with the previous analysis in Section 5.4. But it is noted that even though they are changed every time the prediction is made in the out-of-sample analysis, the results are not changed much, indicating that the cross-sectional relation is stable over the whole sample period. In computing the h-week ahead conditional variance, we omit the variance of $u^{cross}_{t,\tau}$ as in Section 5.4:

$$\mathrm{var}_t[y_{t+h\Delta,\tau}] = \delta'_{1,\tau} \mathrm{var}_t[X_{t+h\Delta}] \delta_{1,\tau} \ , \tag{35}$$

It is noted that although the constant term in equation (34), $\delta_{0,\tau}$, does not appear in equation (35), it may affect the volatility prediction by affecting the estimate of $\delta_{1,\tau}$. But this impact is negligible. The results of the volatility prediction with $\delta_{0,\tau}$ excluded preliminarily from equation (34) little change from those with $\delta_{0,\tau}$ included, which are presented below. Equations (34) and (35) are common also in the next two stages.

In the second stage, we take only (iii) as given, and remove (i) and (ii). The physical distribution of changes in X_t with the same SV-Q and SV-E specifications of Σ_t is estimated using the in-sample data. These estimated parameters are fixed throughout the out-of-sample period.

In the last stage, we remove all of (i)–(iii). Thus, the underlying factors need to be identified first. To avoid arbitrariness, we use the conventional level (lev_t), slope (slo_t), and curvature (cur_t) factors, which are given by

$$\begin{pmatrix} lev_t \\ slo_t \\ cur_t \end{pmatrix} = \begin{pmatrix} 0 & 0 & 1 \\ -1 & 0 & 1 \\ -1 & 2 & -1 \end{pmatrix} \begin{pmatrix} y_{t,0.5} \\ y_{t,2} \\ y_{t,10} \end{pmatrix} \ . \tag{36}$$

The rest of the process is the same as the second stage. It is noted that since X_t is composed of the observed yields with $\tau = \{0.5, 2, 10\}$, there is no residual in equation (34) for these maturities. Also, though the same SV-Q and SV-E specifications are used, significant parameters remaining in Σ_t are different because different factors are used.

Table 15 presents the in-sample RMSEs in basis point, where panels A and B present the results based on equations (31) and (32), respectively. The rows labeled SV-Q and SV-E show the original results, taken from Table 9, with the non-arbitrage imposed. NCi ($i = 1, 2, 3$) indicates "No Constraint", beyond the linear regression, on the cross-sectional relation with the number corresponding to the stage where the non-arbitrage is involved. We limit our attention to $h = 4$ since the results for other forecasting horizons are basically the same: the imposition of non-arbitrage seems irrelevant to the forecasting horizon.

First, in panel A, the RMSEs for NC1 are little changed from the original ones for both the SV-Q and SV-E specifications, indicating that once the factor and parameter values are provided, there is virtually no difference between using the non-arbitrage condition and the linear regression. Also, the result implies that although the SV-Q and SV-E models are outside a class of affine models, the cross-sectional relation between the factors and yields is close to linear. The RMSEs for NC2 are somewhat changed from the original ones, however, the differences do not seem large. The result is attributed to similar estimates of Σ_t between with and without the non-arbitrage, indicating that Σ_t is estimated mostly by time-series properties of the data even though the non-arbitrage is imposed. The impact of non-arbitrage, if any, seems neutral for the volatility forecasting performance: it is favorable for SV-Q but not for SV-E. The RMSEs for NC3 are not much changed although some reduction is observed for the SV-E specification. These results together indicate that the key to the volatility forecasting performance is neither the imposition of non-arbitrage nor the choice of factors associated with this imposition, but the specification of Σ_t.

Next, looking at panel B, we notice that there is no difference at any stage, with a minor exception for NC3 of the SV-Q specification at $\tau = \{0.5, 1\}$, where the RMSEs are a bit increased. Thus, the volatility forecasting regression further decreases the impact of non-arbitrage to a level where it is almost negligible. It is not surprising that these in-sample results are inherited into the out-of-sample results, so they are omitted for brevity. In sum, the non-arbitrage condition has little effect on the volatility prediction, as is the case for the level prediction shown by Duffee (2011).

6 Concluding remarks

Andersen and Benzoni (2010, p.644) pointed out, after rejecting the affine spanning conditions associated with the yield variance, that "further extensions to the term structure

modeling framework are warranted." This paper has shown one such extension. We improved the predictive power of term structure models with respect to the volatility. Based on the three-factor Gaussian model, we made the instantaneous covariance matrix of the factors level-dependent. Specifically, the eigenvalues of this matrix are specified by quadratic (model SV-Q) and exponential (model SV-E) functions of the factors. The proposed models indeed predicted the volatility more accurately than the affine modes. The key to the success is not to identify the volatility factors directly from the yield curve, but to specify the volatility functions with the yield curve factors and estimate them using time-series properties of the data. The non-arbitrage constraint on cross-sectional properties of the data is of little relevance to this success, however.

Compared with the time-series models, GARCH(1,1) and high frequency regressions, the volatility forecasting performance of the proposed models is low for the four-week forecasting horizon especially at short maturities. However, by extending the horizons up to 32 weeks, the relative performance of the proposed models increases. In fact, there are some cases at middle to long maturities in which the performance is comparable or even superior to that of the time-series models in both the in-sample and out-of-sample periods. We further considered combined forecasts, where a forecast of the proposed models was added to the high-frequency regression model as an explanatory variable. The predictive power indeed increased when the forecast of the SV-Q model was included, which, however, did not hold for the SV-E model.

Level-dependent volatility works from a statistical perspective. But less clear is whether it does from an economic perspective. One approach to measure the economic significance of level-dependent volatility, as developed by Fleming, Kirby, and Ostdiek (2001), is to construct bond portfolios based on model-implied first and second moments of returns and to examine their profitability. In this analysis, the predictability of the covariances, which is not a main focus of this paper, has a crucial role. It may be the case that adequate models differ between statistical and economic perspectives. Another question is, as argued in Section 5.1, whether volatility-specific factors that do not appear in the risk-neutral drift terms can further increase the predictive power to the time series of interest rates, or contribute to the description of the cross-section of interest-rate option prices. Addressing these questions is left for future research.

Appendix A: Impact of omitting the mean reversion on the realized volatility

We consider a CIR-type square-root process for each yield, $y_{t,\tau}$:

$$dy_{t,\tau} = (\alpha + \beta y_{t,\tau})dt + \sigma\sqrt{y_{t,\tau}}dW_t \ , \tag{37}$$

where W_t is Brownian motion in the physical measure. The variance of $y_{t+h\Delta,\tau}$ conditioned on $y_{t,\tau}$ is

$$\mathrm{var}_t[y_{t+h\Delta,\tau};\alpha,\beta,\sigma] = \frac{\sigma^2}{\beta}(e^{2\beta h\Delta} - e^{\beta h\Delta})y_{t,\tau} + \frac{\alpha\sigma^2}{2\beta^2}(e^{\beta h\Delta} - 1)^2 \ . \tag{38}$$

By setting $\beta = 0$, it is simplified as

$$\mathrm{var}_t[y_{\tau,t+h\Delta};\sigma] = \sigma^2 h\Delta y_{t,\tau} \ . \tag{39}$$

Then, we compute the time series of % difference between the two conditional standard deviations:

$$100\left(\sqrt{\frac{\mathrm{var}_t[y_{t+h\Delta,\tau};\sigma]}{\mathrm{var}_t[y_{t+h\Delta,\tau};\alpha,\beta,\sigma]}} - 1\right) \ .$$

We evaluate the impact of omitting the mean reversion by mean and standard deviation of the above series. The parameters are estimated using the in-sample data on each yield. The quasi-maximum likelihood method is employed, where the conditional mean and variance are substituted into the normal density function.

In panel A of Table A, the parameter estimates (standard errors) are presented. The speed of mean reversion tends to increase the longer the maturity, as the estimates of β become more negative. However, none of them is significant. In panel B of Table A, the means (standard deviations) of the time series of % difference are presented: in computing $\mathrm{var}_t[y_{t+h\Delta,\tau};\sigma]$, we do not re-estimate σ but simply use the estimates shown in panel A. For $h = 32$, the mean difference reaches nearly 15% at $\tau = 10$. However, due to the insignificant estimates of β, the actual impact of omitting the mean reversion is not as serious as it appears.

Appendix B: An approximation method of conditional moments and its application to the pricing of bonds

B1. 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 stochastic differential equation (SDE):

$$dx_{t,i} = f_i(X_t)dt + \xi_i(X_t)'dW_t \qquad (i = 1, 2)\ , \tag{40}$$

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 (40) 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 ODE.

By integrating equation (40) and taking the conditional expectation,

$$E_s[x_{t,i} - x_{s,i}] = E_s\left[\int_s^t f_i(X_u)du\right]\ . \tag{41}$$

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

$$\begin{aligned} f_i(X_u) = {} & f_i(X_s) \\ & + f_i^{(1,0)}(X_s)(x_{u,1} - x_{s,1}) + f_i^{(0,1)}(X_s)(x_{u,2} - x_{s,2}) + \frac{1}{2}f_i^{(2,0)}(X_s)(x_{u,1} - x_{s,1})^2 \\ & + \frac{1}{2}f_i^{(0,2)}(X_s)(x_{u,2} - x_{s,2})^2 + f_i^{(1,1)}(X_s)(x_{u,1} - x_{s,1})(x_{u,2} - x_{s,2}) + e_i\ , \end{aligned} \tag{42}$$

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 (42) into equation (41) and expressing the resulting equation in a vector form

$$\begin{aligned} E_s[x_{t,i} - x_{s,i}] &= f_i(t-s) \\ &+ \left(f_i^{(1,0)} \quad f_i^{(0,1)} \quad \frac{1}{2} f_i^{(2,0)} \quad \frac{1}{2} f_i^{(0,2)} \quad f_i^{(1,1)} \right) \int_s^t \Psi_{s,u} du + R_i \ , \end{aligned} \tag{43}$$

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,

$$E_s[(x_{t,1} - x_{s,1})^2] = E_s \left[\int_s^t \{2 f_1(X_u)(x_{u,1} - x_{s,1}) + g_{11}(X_u)\} du \right] \ , \tag{44}$$

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 (44) becomes

$$\begin{aligned} &2 f_1(X_u)(x_{u,1} - x_{s,1}) + g_{11}(X_u) \\ &\quad = g_{11}(X_s) + \{2 f_1(X_s) + g_{11}^{(1,0)}(X_s)\}(x_{u,1} - x_{s,1}) + g_{11}^{(0,1)}(X_s)(x_{u,2} - x_{s,2}) \\ &\quad + \{2 f_1^{(1,0)}(X_s) + \frac{1}{2} g_{11}^{(2,0)}(X_s)\}(x_{u,1} - x_{s,1})^2 + \frac{1}{2} g_{11}^{(0,2)}(X_s)(x_{u,2} - x_{s,2})^2 \\ &\quad + \{2 f_1^{(0,1)}(X_s) + g_{11}^{(1,1)}(X_s)\}(x_{u,1} - x_{s,1})(x_{u,2} - x_{s,2}) + e_{11} \ , \end{aligned} \tag{45}$$

where $g^{(k,l)}$ is defined analogously with $f^{(k,l)}$, and e_{11} is a residual term. By substituting equation (45) into equation (44),

$$\begin{aligned} E_s[(x_{t,1} - x_{s,1})^2] &= g_{11}(t-s) \\ &+ \left(2 f_1 + g_{11}^{(1,0)} \quad g_{11}^{(0,1)} \quad 2 f_1^{(1,0)} + \frac{1}{2} g_{11}^{(2,0)} \quad \frac{1}{2} g_{11}^{(0,2)} \quad 2 f_1^{(0,1)} + g_{11}^{(1,1)} \right) \\ &\times \int_s^t \Psi_{s,u} du + R_{11} \ , \end{aligned} \tag{46}$$

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

$$\Psi_{s,t} = A(X_s) \int_s^t \Psi_{s,u} du + b(X_s)(t-s) + R \ , \tag{47}$$

where

$$A = \begin{pmatrix} f_1^{(1,0)} & f_1^{(0,1)} & \frac{1}{2} f_1^{(2,0)} & \frac{1}{2} f_1^{(0,2)} & f_1^{(1,1)} \\ f_2^{(1,0)} & f_2^{(0,1)} & \frac{1}{2} f_2^{(2,0)} & \frac{1}{2} f_2^{(0,2)} & f_2^{(1,1)} \\ 2 f_1 + g_{11}^{(1,0)} & g_{11}^{(0,1)} & 2 f_1^{(1,0)} + \frac{1}{2} g_{11}^{(2,0)} & \frac{1}{2} g_{11}^{(0,2)} & 2 f_1^{(0,1)} + g_{11}^{(1,1)} \\ g_{22}^{(1,0)} & 2 f_2 + g_{22}^{(0,1)} & \frac{1}{2} g_{22}^{(2,0)} & 2 f_2^{(0,1)} + \frac{1}{2} g_{22}^{(0,2)} & 2 f_2^{(1,0)} + g_{22}^{(1,1)} \\ f_2 + g_{12}^{(1,0)} & f_1 + g_{12}^{(0,1)} & f_2^{(1,0)} + \frac{1}{2} g_{12}^{(2,0)} & f_1^{(0,1)} + \frac{1}{2} g_{12}^{(0,2)} & f_1^{(1,0)} + f_2^{(0,1)} + g_{12}^{(1,1)} \end{pmatrix} \ , \tag{48}$$

$$
\begin{aligned}
b &= (f_1 \quad f_2 \quad g_{11} \quad g_{22} \quad g_{12})' \,, \\
R &= (R_1 \quad R_2 \quad R_{11} \quad R_{22} \quad R_{12})' \,.
\end{aligned}
$$

Equation (47) can be solved as

$$
\Psi_{s,t} = \int_s^t e^{A(X_s)(t-u)} b(X_s) du + \hat{R} \,. \tag{48}
$$

If, in addition, A is invertible, we obtain

$$
\Psi_{s,t} = A^{-1}(X_s)\{e^{A(X_s)(t-s)} - I\} b(X_s) + \hat{R} \,. \tag{49}
$$

It is noted that equations (47)–(49) 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 pricing bonds, we consider $n = 3$.

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. Even in such a case, the use of this formula may be beneficial when the derivation of closed-form conditional moments is demanding.

B2. Application of the method

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

$$
z_{s,t} = \exp\left\{ - \int_s^t r(X_u) du \right\} \,, \tag{50}
$$

and the price of a discount bond at time t maturing at time T is equal to the conditional first moment of $z_{t,T}$ under the risk-neutral measure. This (actually $E_t^Q[z_{t,T} - z_{t,t}]$) is computed as one of the elements of the moment vector, $\Psi_t(T)$. Specifically, we first extend a state vector as $\hat{X}_t = (X_t' \; z_{s,t})'$, where X_t is a d-dimensional diffusion process and

$z_{s,t}$ is treated as the $(d+1)$-th process. By the Ito formula,

$$dz_{s,t} = -r(X_t)z_{s,t}dt \ , \qquad z_{s,s} = 1 \ , \tag{51}$$

and we have

$$f_{d+1}(\hat{X}_t) = -r(X_t)z_{s,t} \ , \tag{52}$$

$$g_{i\,d+1}(\hat{X}_t) = 0 \qquad (i = 1, ..., d+1) \ . \tag{53}$$

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

The approximation method is also applied to the computation of conditional first and second moments of a model-implied yield, $E_s[\tilde{Y}(X_t, \tau)]$ and $E_s[\tilde{Y}(X_t, \tau)^2]$. Similar to the case in $z_{s,t}$, we first extend a state vector as $\hat{X}_t = (X_t' \quad \tilde{Y}(X_t, \tau))'$, and then derive the SDE for $\tilde{Y}(X_t, \tau)$:

$$d\tilde{Y}(X_t, \tau) = f_{d+1}(X_t, \tau)dt + \xi_{d+1}'(X_t, \tau)dW_t \ , \tag{54}$$

where

$$f_{d+1}(X_t, \tau) = \frac{\partial \tilde{Y}(X_t, \tau)}{\partial X_t'}\mu_t + \frac{1}{2}\mathrm{tr}\left(\frac{\partial^2 \tilde{Y}(X_t, \tau)}{\partial X_t \partial X_t'}\Sigma_t\right) \ , \tag{55}$$

$$\xi_{d+1}(X_t, \tau) = \Sigma_t^{0.5\prime}\frac{\partial \tilde{Y}(X_t, \tau)}{\partial X_t} \ , \tag{56}$$

and where W_t is d-dimensional Brownian motion, and μ_t and Σ_t are the physical drift vector and the instantaneous covariance matrix of dX_t, respectively. $g_{i\,d+1}$ is obtained by $\xi_i'\xi_{d+1}$ $(i = 1, ..., d+1)$.

In computing $A(\hat{X}_s)$, the derivatives of f_{d+1} and $g_{i\,d+1}$ are required. But f_{d+1} and $g_{i\,d+1}$ already contain the derivatives of $\tilde{Y}(X_t, \tau)$ up to the second order, which complicates the calculation. To avoid the tedious calculation, the derivatives of $\tilde{Y}(X_t, \tau)$ higher than the second order are omitted, while the derivatives of μ_t and Σ_t are taken as many times as necessary. The accuracy of the approximation to $E_t[\tilde{Y}(X_{t+h\Delta}, \tau)^n]$ $(n = 1, 2)$ is also investigated in the next appendix.

Appendix C: Accuracy of the approximation

The purpose of this appendix is to let the cost of using the approximation be known. By construction of the method, the accuracy becomes worse the longer the time interval, $t-s$. Here, the interval is up to ten years for pricing bonds, which raises concern with the application of this method. To check the accuracy of the approximation, we consider two cases with and without a closed-form solution for bond prices. This is aimed at examining how close the approximate solution is to the closed-form and numerical solutions, respectively.

C1. Comparison to the closed-form solution

We treat $A_0(3)$ as the true model. We divide the analysis into three steps according to the degrees of approximation involved. Let Θ_0 be the parameter vector of the $A_0(3)$ model, the elements of which are set at the estimates presented in Table 4 (also re-exhibited in Table C2). The extracted state vector can then be expressed as $X(Y_t^p;\Theta_0)$.

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

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

where $Y(\cdot)$ and $\tilde{Y}(\cdot)$ stand for the closed-form and approximate functions, respectively.

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

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

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

In the last step, we examine the impact of the approximation on the estimation of model parameters as well as on the pricing of bonds and the extraction of state variables. Here, no prior information on the true values of the parameter and state vectors is given. Using the approximation, the parameter vector of the $A_0(3)$ model is first estimated;

denote it as $\tilde{\Theta}_0$. Next, the state vector is backed out; denote it as $\tilde{X}(Y_t^p; \tilde{\Theta}_0)$. Finally, the rest of the yields are computed. Then, we compare:

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

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

Apart from the parameter vector, the key input for these comparisons is Y_t^p. We use the actual data on Y_t^p. To condense the analysis, we pick up nine observations from the entire sample in the following way. First, Y_t^p is transformed to the conventional level (lev_t), slope (slo_t), and curvature (cur_t) factors by equation (36). Then, we choose three dates in which lev_t takes the minimum, median, or maximum value. Likewise, the three dates are chosen for each of the other proxies, leading to nine dates in total. In this way, the accuracy of the approximation is evaluated at not only the usual times but also the unusual times.

Table C1 presents the differences, expressed in basis points, between the approximate and true yields or state variables. Panel A presents the results for the first step comparison, where the true values of the parameter and state vectors are given as input for the approximation. For maturities of up to five years, the approximation errors are negligibly small at all states. Even for the ten-year maturity, the error exceeds 2 bps only at the maximum-spread state. The values of the curvature, on the other hand, little affects the accuracy.

Panel B of Table C1 presents the results for the second-step comparison, where only the true value of the parameter vector is given. A systematic pattern is found in the approximation errors for the state vector. Specifically, both r and θ are undervalued, which is compensated by the overvaluation of ϵ. The difficulty of the approximation arises again at the maximum-spread state. On the other hand, the approximation errors for the remaining yields are small.

Panel C of Table C1 presents the results for the third-step comparison, where no prior information is given. Before discussing the results, we first look at Table C2, where the parameter estimates (standard errors) obtained by the approximation method are presented. Overall, the estimates do not seem to differ much from the corresponding true

values. Except for $\bar{\theta}$, the differences in absolute value are smaller than the standard errors. Now, looking back at panel C of Table C1, we find, nevertheless, that the magnitude of the approximation errors generally increases. This is the reality. Here, an error pattern is less clear, but it is seen that r tends to be overvalued and θ tends to undervalued. Also, the difficulty of the approximation is not limited to the maximum-spread state. In fact, the largest errors in absolute value for r and θ appear at the minimum-curvature state and the minimum-level state, respectively. The remaining yields are accurately computed.

C2. Comparison to the numerical solution

We employ the Monte Carlo (MC) method to evaluate the accuracy. The models we use are SV-Q and SV-E. Let Θ_i $(i = \{Q, E\})$ be the true parameter vectors, the elements of which are set at the estimates presented in Tables 5 and 6, respectively. Further, the state vector is extracted by the approximation method, not by the MC method; denote it as $\tilde{X}(Y_t^p; \Theta_i)$. Then, we compare

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

In the MC simulations, $\{X_s\}_t^{t+\tau}$ is generated from (11) (the risk-neutral distribution), where dt is replaced by Δt. We set $\Delta t = 1/1,000$, an interval corresponding roughly to four observations per day. The number of repetition is set at 10,000 with antithetic variates.

Table C3 presents the differences, expressed in basis points, between the approximate and MC yields. Generally, the error pattern is similar to that for the first-step comparison to the closed-form solution. For maturities of up to five years, the approximation errors are within 1 bp at all states for both models. For the ten-year yield, the approximation error exceeds 2 bps only at the maximum-spread state. However, this comparison does not take the approximation errors in both the parameter and state vectors into consideration. In reality, therefore, the approximation errors for the resulting yields would be larger, as seen in the third-step comparison to the closed-form solution.

C3. Accuracy to conditional first and second moments of a model-implied yield

Based on the SV-Q and SV-E models, we again employ the MC method to evaluate the accuracy at the nine states. In the MC simulations, $\{X_s\}_t^{t+h\Delta}$ is generated from (10) (the physical distribution), where Σ is replaced by Σ_t and dt by Δt $(= 1/1,000)$. For simplicity, once $X_{t+h\Delta}$ is obtained, a model-implied τ-maturity yield (and its squared value) is computed by the approximation method. This procedure is repeated 10,000 times with antithetic variates. Then, we compare

$$E_t[\tilde{Y}(X_{t+h\Delta},\tau)] \qquad v.s. \quad E_t^{ap}[\tilde{Y}(X_{t+h\Delta},\tau)]\ ,$$

$$\sqrt{\mathrm{var}_t[\tilde{Y}(X_{t+h\Delta},\tau)]/h\Delta} \quad v.s. \quad \sqrt{\mathrm{var}_t^{ap}[\tilde{Y}(X_{t+h\Delta},\tau)]/h\Delta}\ ,$$

where the superscript "ap" is for clarifying the approximation. We set $h = 32$ as this is the longest horizon we consider and thus the accuracy is the worst.

Table C4 presents the differences, expressed in basis points, in the conditional mean. The errors in absolute value are all within 1 bp for both models. Table C5 presents the differences, expressed in basis points, in the annualized conditional standard deviation. Although the accuracy to the second moment becomes worse than that to the first moment, the errors in absolute value are all within 2 bps.

It is noted that the more rigorous analysis of the accuracy requires to compare

$$E_t[Y(X_{t+h\Delta},\tau)^n] \qquad v.s. \qquad E_t^{ap}[\tilde{Y}(X_{t+h\Delta},\tau)^n] \qquad (n=1,2)\ .$$

That is, no approximation is involved in the former in computing the τ-maturity yield at time $t + h\Delta$: but this analysis is computationally very demanding. In reality, therefore, the approximation errors would be larger than presented in Tables C4 and C5.

Appendix D: Computation of the conditional variance for the GARCH (1,1) model

The variance of $y_{t+\Delta,\tau}$ conditioned on time t is simply $V_{t+\Delta,\tau}$, which is observed at time t. The variance of $y_{t+k\Delta,\tau}$ $(k = 2, ..., h)$ conditioned on time t is computed iteratively as follows. In equation (27), by substituting $t + (k-1)\Delta$ for t and taking the variance conditioned on time t,

$$\begin{aligned}
\mathrm{var}_t[y_{t+k\Delta,\tau}] &= \mathrm{var}_t[\alpha + \beta y_{t+(k-1)\Delta,\tau} + \sqrt{V_{t+k\Delta,\tau}}z_{t+k\Delta,\tau}] \\
&= \beta^2 \mathrm{var}_t[y_{t+(k-1)\Delta,\tau}] + \mathrm{var}_t[\sqrt{V_{t+k\Delta,\tau}}z_{t+k\Delta,\tau}] \\
&\qquad +2\beta\,\mathrm{cov}_t[y_{t+(k-1)\Delta,\tau},\ \sqrt{V_{t+k\Delta,\tau}}z_{t+k\Delta,\tau}] \\
&= \beta^2 \mathrm{var}_t[y_{t+(k-1)\Delta,\tau}] + E_t[V_{t+k\Delta}] \qquad (k = 2, ..., h)\ . \qquad (57)
\end{aligned}$$

On the other hand, in equation (28), by substituting $t + (k-1)\Delta$ for t and taking the expectation conditioned on time t,

$$\begin{aligned}
E_t[V_{t+k\Delta}] &= E_t[\omega_0 + \omega_1 V_{t+(k-1)\Delta,\tau} z^2_{t+(k-1)\Delta,\tau} + \omega_2 V_{t+(k-1)\Delta,\tau}] \\
&= \omega_0 + (\omega_1 + \omega_2) E_t[V_{t+(k-1)\Delta,\tau}] \qquad (k = 2, ..., h)\ . \qquad (58)
\end{aligned}$$

Then, $\mathrm{var}_t[y_{t+h\Delta,\tau}]$ is obtained by iteratively solving equations (57) and (58) starting from $\mathrm{var}_t[y_{t+\Delta,\tau}] = V_{t+\Delta,\tau}$.

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., T. Bollerslev, F. X. Diebold, and P. Labys, 2003, Modeling and forecasting realized volatility, *Econometrica* 71, 579-625.

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.

Bakshi, G., C. Cao, and Z. Chen, 1997, Empirical performance of alternative option pricing models, *Journal of Finance* 52, 2003-2049.

Balduzzi, P., S. R. Das, and S. Foresi, 1998, The central tendency: A second factor in bond yields, *Review of Economics and Statistics* 80, 62-72.

Balduzzi, P., S. R. Das, S. Foresi, and R. Sundaram, 1996, A simple approach to three-factor affine term structure models, *Journal of Fixed Income* 6, 43-53.

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.

Bates, D. S., 1996, Jumps and stochastic volatility: Exchange rate processes implicit in Deutsche mark options, *Review of Financial Studies* 9, 69-107.

Bates, D. S., 2000, Post-'87 crash fears in the S&P 500 futures option market, *Journal of Econometrics* 94, 181-238.

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.

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

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, 2008, Identification of maximal affine term structure models, *Journal of Finance* 63, 743-795.

Collin-Dufresne, P., R. S. Goldstein, and C. S. Jones, 2009, Can interest rate volatility be extracted from the cross section of bond yields? *Journal of Financial Economics* 94, 47-66.

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., 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.

Duffee, G. R., 2011, Forecasting with the term structure: The role of no-arbitrage restrictions, Working Paper, Johns Hopkins University.

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.

Fleming, J., C. Kirby, and B. Ostdiek, 2001, The economic value of volatility timing, *Journal of Finance* 56, 329-352.

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., 2006, Can unspanned stochastic volatility models explain the cross section of bond volatilities? Working Paper, MIT Sloan School of Management.

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.

Newey, W. K., and K. D. West, 1987, A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix, *Econometrica* 55, 703-708.

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

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.

Maturity	0.5	1	2	3	5	10
$h = 4$						
Y-PART	0.34	0.30	0.17	0.13	0.13	0.07
Y-ALL	0.36	0.33	*0.23*	**0.16**	**0.15**	0.08
AR-RV	*0.39*	*0.35*	0.22	0.13	0.14	**0.12**
HAR-RV	**0.43**	**0.39**	**0.24**	*0.14*	**0.15**	**0.12**
$h = 8$						
Y-PART	0.39	0.35	*0.23*	*0.18*	*0.17*	*0.10*
Y-ALL	*0.42*	*0.40*	**0.29**	**0.21**	**0.21**	**0.11**
AR-RV	0.40	*0.40*	0.21	0.11	0.13	0.07
HAR-RV	**0.45**	**0.42**	*0.23*	0.12	0.14	0.09
$h = 16$						
Y-PART	0.42	0.39	*0.28*	*0.24*	*0.22*	*0.16*
Y-ALL	**0.45**	**0.44**	**0.35**	**0.27**	**0.25**	**0.17**
AR-RV	0.40	0.34	0.15	0.06	0.06	0.00
HAR-RV	*0.44*	*0.40*	0.22	0.13	0.12	0.07
$h = 32$						
Y-PART	*0.40*	*0.40*	*0.31*	*0.31*	*0.27*	*0.28*
Y-ALL	**0.43**	**0.44**	**0.40**	**0.35**	**0.32**	**0.29**
AR-RV	0.33	0.33	0.17	0.11	0.06	0.03
HAR-RV	0.38	0.37	0.20	0.12	0.07	0.03

Table 1: Adjusted R^2 coefficients for the volatility forecasting regression
The regression models are given by equations (3)–(6), where $f(x) = \sqrt{x}$ and $h = \{4, 8, 16, 32\}$. The best and the second best results in each column of each panel are displayed in bold and italic, respectively. The in-sample data from January 4, 1991 to April 9, 2003 are used for the estimation.

Maturity	0.5	1	2	3	5	10
$h = 4$						
Y-PART	45.0	41.9	42.9	41.4	39.9	40.6
Y-ALL	40.4	38.3	42.2	39.8	39.7	39.7
AR-RV	*31.5*	*28.6*	*32.2*	*34.0*	*34.3*	**35.5**
HAR-RV	**30.7**	**27.3**	**30.5**	**32.2**	**32.4**	*35.7*
$h = 8$						
Y-PART	42.4	39.3	39.8	38.6	37.4	37.7
Y-ALL	36.6	34.7	38.0	35.3	35.5	36.0
AR-RV	*30.7*	*25.0*	*29.1*	*32.2*	*32.0*	**35.2**
HAR-RV	**29.3**	**24.4**	**28.1**	**30.6**	**31.2**	*35.3*
$h = 16$						
Y-PART	40.3	37.5	37.7	37.0	36.5	36.3
Y-ALL	35.1	33.8	36.3	33.7	*34.1*	**34.1**
AR-RV	*29.4*	*23.8*	*30.1*	*33.6*	34.4	38.3
HAR-RV	**27.7**	**23.2**	**28.4**	**30.6**	**31.7**	*34.9*
$h = 32$						
Y-PART	37.3	35.1	36.6	36.7	35.4	34.5
Y-ALL	35.1	34.8	35.9	34.4	33.2	33.0
AR-RV	**25.6**	*23.4*	*29.9*	*31.5*	*32.7*	*31.5*
HAR-RV	**25.6**	**22.3**	**28.4**	**30.6**	**32.0**	**31.4**

Table 2: Out-of-sample RMSEs in the fixed-parameter approach
Out-of-sample forecasts are generated from the regression equations (3)–(6), where $f(x) = \sqrt{x}$ and $h = \{4, 8, 16, 32\}$. The 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 of each panel are displayed in bold and italic, respectively.

Maturity	0.5	1	2	3	5	10
$h = 4$						
Y-PART	34.3	32.4	34.8	33.9	32.3	33.5
Y-ALL	*31.4*	30.6	34.0	32.9	*30.7*	**28.8**
AR-RV	32.0	*29.1*	*31.7*	*32.3*	31.8	33.1
HAR-RV	**31.1**	**27.7**	**30.1**	**30.9**	**29.9**	*32.2*
$h = 8$						
Y-PART	32.1	29.3	31.3	30.4	28.4	*29.2*
Y-ALL	**29.3**	27.7	30.6	*29.2*	**26.3**	**23.7**
AR-RV	31.0	*25.3*	*27.9*	29.4	28.2	31.7
HAR-RV	*29.8*	**24.9**	**27.5**	**28.8**	*27.2*	30.0
$h = 16$						
Y-PART	31.2	27.7	29.8	29.0	27.4	*28.0*
Y-ALL	*28.2*	26.0	29.4	**28.3**	**25.9**	**23.3**
AR-RV	29.8	*24.3*	*29.0*	30.6	29.5	33.4
HAR-RV	**27.5**	**23.6**	**27.8**	*28.9*	*26.7*	28.6
$h = 32$						
Y-PART	29.4	25.8	28.0	28.0	27.4	*27.0*
Y-ALL	27.6	24.9	*27.6*	**27.2**	**26.0**	**23.5**
AR-RV	**25.9**	*23.8*	28.5	29.3	27.6	27.3
HAR-RV	**25.9**	**22.2**	**26.3**	*27.5*	**26.0**	*27.0*

Table 3: Out-of-sample RMSEs in the varying-parameter approach
Out-of-sample forecasts are generated from the regression equations (3)–(6), where $f(x) = \sqrt{x}$ and $h = \{4, 8, 16, 32\}$. The parameters are re-estimated in a rolling-window fashion with the sample size fixed at the same as the in-sample data. The best and the second best results in each column of each panel are displayed in bold and italic, respectively. The out-of-sample period is from April 16, 2003 to May 27, 2009.

Parameter / Index	$i=1$		$i=2$		$i=3$	
κ_i	0.947	(0.087)	0.017	(0.002)	0.957	(0.090)
$\bar{\theta}$			0.146	(0.008)		
$\sigma_{1i} \times 10^4$	0.722	(0.030)	0.095	(0.035)	−0.985	(0.115)
$\sigma_{2i} \times 10^4$			1.220	(0.063)	0.000	
$\sigma_{3i} \times 10^4$					8.036	(0.438)
λ_i^0	0.000		0.030	(0.019)	−0.013	(0.008)
λ_{1i}^1	0.000		−0.195	(0.037)	0.000	
λ_{2i}^1	0.000		−0.514	(0.280)	0.000	
λ_{3i}^1	0.000		0.000		0.000	
$\varsigma \times 10^2$			0.061	(0.001)		
LogL			22263			

Table 4: Parameter estimates (standard errors) for the $A_0(3)$ model
The risk-neutral and physical distributions of the instantaneous change in X_t for the $A_0(3)$ model are given by (7) and (10), respectively. ς is the standard deviation of the measurement errors, $u_{t,\tau} = y_{t,\tau} - \tilde{Y}(X_t, \tau)$ with $\tau = \{1, 3, 5\}$. LogL stands for the maximum log-likelihood value. Standard errors in parenthesis are computed by the outer product of the gradient vector of the log-likelihood function. The in-sample data from January 4, 1991 to April 9, 2003 are used for the estimation.

Parameter / Index	$i=1$		$i=2$		$i=3$	
κ_i	1.034	(0.073)	0.019	(0.002)	0.917	(0.063)
$\bar{\theta}$			0.137	(0.007)		
$\sin\varphi_i^P$	0.000		0.126	(0.010)	0.000	
$c_i \times 10^3$	1e−5		0.111	(0.008)	0.149	(0.107)
m_i^1	0.000		0.022	(0.002)	0.260	(0.034)
$\sin\varphi_i^{Q^1}$	0.250	(0.047)	0.631	(0.021)	−0.705	(0.042)
m_i^2	0.000		0.000		0.097	(0.054)
$\sin\varphi_i^{Q^2}$	0.000		0.547	(0.105)	−0.547	
m_i^3	0.000		0.228	(0.061)	2.153	(0.627)
$\sin\varphi_i^{Q^3}$	0.273	(0.134)	0.551	(0.058)	−0.616	(0.087)
λ_i^0	0.000		0.020	(0.020)	−0.011	(0.008)
λ_{1i}^1	0.000		−0.150	(0.031)	0.000	
λ_{2i}^1	0.000		−0.357	(0.301)	0.000	
λ_{3i}^1	0.000		0.000		0.000	
$\varsigma \times 10^2$			0.061	(0.001)		
LogL			22415			

Table 5: Parameter estimates (standard errors) for the SV-Q model
The risk-neutral and physical distributions of the instantaneous change in X_t for the SV-Q model are given by (7) and (10), respectively, with Σ replaced by Σ_t. Σ_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 (14). The i-th diagonal element of L_t is specified 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 (18). ς is the standard deviation of the measurement errors, $u_{t,\tau} = y_{t,\tau} - \tilde{Y}(X_t, \tau)$ with $\tau = \{1, 3, 5\}$. LogL stands for the maximum log-likelihood value. Standard errors in parenthesis are computed by the outer product of the gradient vector of the log-likelihood function. The in-sample data from January 4, 1991 to April 9, 2003 are used for the estimation.

Parameter / Index	$i=1$		$i=2$		$i=3$	
κ_i	0.955	(0.070)	0.023	(0.003)	0.981	(0.078)
$\bar{\theta}$			0.118	(0.006)		
$\sin\varphi_i^P$	0.000		0.123	(0.013)	0.000	
c_i	−12.942	(0.263)	−8.693	(0.214)	−8.574	(0.379)
γ_i^1	10.538	(2.235)	35.110	(3.476)	−17.267	(2.431)
γ_i^2	−8.828	(3.697)	0.000		−10.551	(2.943)
γ_i^3	−13.489	(4.665)	29.079	(4.460)	−9.665	(2.733)
λ_i^0	0.000		0.036	(0.022)	−0.014	(0.008)
λ_{1i}^1	0.000		−0.171	(0.038)	0.000	
λ_{2i}^1	0.000		−0.595	(0.318)	0.000	
λ_{3i}^1	0.000		0.000		0.000	
$\varsigma \times 10^2$			0.061	(0.001)		
LogL			22358			

Table 6: Parameter estimates (standard errors) for the SV-E model
The risk-neutral and physical distributions of the instantaneous change in X_t for the SV-E model are given by (7) and (10), respectively, with Σ replaced by Σ_t. Σ_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 (14). The i-th diagonal element of L_t is specified as $L_i(X_t) = \exp\{c_i + \gamma^{i\prime} X_t\}$ $(i = 1, 2, 3)$. ς is the standard deviation of the measurement errors, $u_{t,\tau} = y_{t,\tau} - \tilde{Y}(X_t, \tau)$ with $\tau = \{1, 3, 5\}$. LogL stands for the maximum log-likelihood value. Standard errors in parenthesis are computed by the outer product of the gradient vector of the log-likelihood function. The in-sample data from January 4, 1991 to April 9, 2003 are used for the estimation.

Maturity	0.5	1	2	3	5	10
$h = 4$						
RW	23.1	26.8	30.3	30.1	28.4	26.1
$A_0(3)$	22.0	28.2	30.1	29.7	28.1	25.7
SV-Q	22.1	28.7	30.2	29.8	28.2	25.8
SV-E	22.0	28.3	30.1	29.7	28.1	25.7
$h = 8$						
RW	37.8	42.9	47.7	46.9	43.6	37.8
$A_0(3)$	35.9	44.0	47.4	46.1	42.5	36.9
SV-Q	36.2	44.8	47.6	46.3	42.7	37.0
SV-E	35.9	44.1	47.4	46.1	42.5	36.9
$h = 16$						
RW	63.4	69.2	74.3	72.1	66.2	55.6
$A_0(3)$	60.7	70.5	73.5	70.0	63.2	52.8
SV-Q	61.4	71.8	74.2	70.7	63.7	53.2
SV-E	60.6	70.6	73.4	69.9	63.1	52.8
$h = 32$						
RW	107.0	109.9	110.8	105.4	96.1	82.4
$A_0(3)$	102.7	109.0	106.1	98.4	86.9	73.3
SV-Q	104.3	111.6	108.3	100.6	89.0	74.8
SV-E	102.4	108.8	105.7	97.8	86.3	72.9

Table 7: In-sample RMSEs for the level prediction

Root mean squared errors (RMSEs) for the in-sample prediction of the h-week ahead yields are presented in basis points with $h = \{4, 8, 16, 32\}$. RW represents random walk, whose forecast is the current yield for any h. The in-sample period is from January 4, 1991 to April 9, 2003.

Maturity	0.5	1	2	3	5	10
$h = 4$						
RW	28.9	28.2	29.4	31.0	32.4	31.5
$A_0(3)$	25.2	27.8	28.2	28.8	30.7	31.5
SV-Q	24.8	27.6	28.2	29.0	30.8	31.4
SV-E	24.9	27.4	28.2	28.9	30.9	31.7
$h = 8$						
RW	44.5	42.7	45.1	46.5	47.3	45.5
$A_0(3)$	35.9	37.2	42.1	42.2	43.7	45.4
SV-Q	35.0	36.9	42.2	42.5	44.0	45.3
SV-E	35.1	36.4	42.0	42.4	44.1	45.9
$h = 16$						
RW	70.3	66.6	70.1	68.4	63.8	58.2
$A_0(3)$	55.3	55.4	63.7	61.2	59.1	59.0
SV-Q	53.4	54.6	64.0	61.6	59.1	58.2
SV-E	53.6	53.9	63.7	61.7	60.1	60.6
$h = 32$						
RW	112.3	104.8	102.6	93.8	81.0	69.4
$A_0(3)$	91.1	85.0	88.8	82.9	78.2	77.2
SV-Q	88.0	83.6	89.4	83.5	77.3	73.8
SV-E	88.0	82.6	89.1	84.6	81.4	81.8

Table 8: Out-of-sample RMSEs for the level prediction
Root mean squared errors (RMSEs) for the out-of-sample prediction of the h-week ahead yields are presented in basis points with $h = \{4, 8, 16, 32\}$. RW represents random walk, whose forecast is the current yield for any h. The out-of-sample period is from April 16, 2003 to May 27, 2009.

Maturity	0.5	1	2	3	5	10
$h = 4$						
$A_0(3)$	39.2	39.8	34.8	34.6	28.9	26.6
$A_1(3)$	36.1	37.0	33.4	34.1	29.4	28.0
SV-Q	33.7	35.0	32.8	33.9	28.3	26.9
SV-E	32.0	33.8	32.8	33.6	28.4	27.8
GARCH	33.1	31.4	31.1	32.3	27.0	25.0
$A_1(3)$ + Reg	29.8	34.1	33.1	33.9	28.7	26.4
SV-Q + Reg	28.7	32.5	31.1	32.4	27.5	26.3
SV-E + Reg	27.8	31.9	31.5	32.3	27.1	25.8
GARCH + Reg	27.1	*29.1*	30.5	31.9	26.6	**24.7**
HAR-RV	**25.6**	*29.1*	*29.8*	*31.8*	*26.4*	24.9
MIDAS	**25.6**	**28.8**	**29.5**	**31.6**	**26.1**	**24.7**
$h = 8$						
$A_0(3)$	37.5	37.4	30.6	29.5	24.2	22.1
$A_1(3)$	34.0	33.9	28.8	28.8	24.9	23.7
SV-Q	32.2	32.0	27.7	27.9	23.4	22.4
SV-E	30.1	30.8	28.0	28.1	23.7	23.2
GARCH	31.6	27.8	26.9	27.5	22.9	*20.9*
$A_1(3)$ + Reg	27.0	30.6	28.7	28.8	24.2	21.9
SV-Q + Reg	25.9	28.7	**26.0**	**26.7**	22.7	21.8
SV-E + Reg	24.7	28.1	26.7	*26.9*	*22.4*	21.2
GARCH + Reg	24.6	**25.5**	26.3	27.1	22.5	*20.9*
HAR-RV	**23.3**	26.1	26.2	27.4	*22.4*	21.0
MIDAS	*23.5*	*25.9*	**26.0**	27.2	**22.1**	**20.8**

Table 9: In-sample RMSEs for the volatility prediction

Maturity	0.5	1	2	3	5	10
$h = 16$						
$A_0(3)$	37.8	36.2	26.9	24.8	20.5	18.6
$A_1(3)$	33.1	31.4	24.9	24.3	21.6	20.7
SV-Q	34.1	31.4	24.1	22.5	19.6	18.7
SV-E	31.1	29.7	24.0	22.9	20.1	19.8
GARCH	32.0	26.6	24.0	23.4	19.9	17.7
$A_1(3)$ + Reg	24.2	27.3	24.7	24.0	20.4	17.5
SV-Q + Reg	23.7	25.5	**21.9**	**21.5**	*19.0*	17.7
SV-E + Reg	22.2	25.2	*22.6*	*21.8*	**18.5**	**17.1**
GARCH + Reg	22.8	**23.9**	23.4	22.9	19.3	17.4
HAR-RV	**21.6**	*24.4*	23.1	23.0	19.1	*17.2*
MIDAS	**21.6**	*24.4*	23.4	23.4	*19.0*	17.3
$h = 32$						
$A_0(3)$	42.2	36.2	22.8	20.0	17.7	17.2
$A_1(3)$	34.3	28.8	21.1	21.3	20.5	20.1
SV-Q	41.8	33.9	21.5	17.9	15.9	15.5
SV-E	37.8	31.1	20.6	18.7	18.2	19.1
GARCH	35.0	24.9	20.0	19.3	17.5	16.4
$A_1(3)$ + Reg	21.4	23.8	20.5	18.8	16.2	*12.5*
SV-Q + Reg	21.7	22.5	**18.3**	**16.8**	*15.0*	12.9
SV-E + Reg	20.6	22.7	*18.8*	*17.0*	**14.7**	**12.4**
GARCH + Reg	20.9	**21.5**	19.7	18.2	15.6	12.9
HAR-RV	*20.2*	22.4	19.8	18.5	15.6	12.7
MIDAS	**20.1**	*22.3*	19.7	18.6	15.7	13.0

Table 9 (continued): In-sample RMSEs for the volatility prediction
Root mean squared errors (RMSEs) for the in-sample prediction of the h-week ahead yield volatilities (annualized standard deviations) are presented in basis points with $h = \{4, 8, 16, 32\}$. For the term structure and GARCH models, the name alone indicates the results based on equation (31), whereas the name "+ Reg" indicates the results based on equation (32). The best and the second best results in each column of each panel are displayed in bold and italic, respectively. The in-sample period is from January 4, 1991 to April 9, 2003.

Maturity	0.5	1	2	3	5	10
$h = 4$						
$A_0(3)$	50.0	42.3	42.9	43.4	44.3	44.3
$A_1(3)$	43.9	37.1	44.2	46.7	50.5	49.9
SV-Q	47.5	42.5	37.7	34.5	32.5	**34.1**
SV-E	39.9	35.9	37.9	37.2	35.7	35.5
GARCH	46.7	34.3	31.5	**31.0**	**30.0**	34.7
$A_1(3)$ + Reg	51.5	47.5	46.8	46.1	45.3	43.6
SV-Q + Reg	36.1	33.8	34.3	34.6	35.9	39.8
SV-E + Reg	39.3	35.7	38.8	39.1	39.3	40.2
GARCH + Reg	34.1	30.7	31.8	32.3	32.5	*34.3*
HAR-RV	*30.7*	*27.3*	*30.5*	*32.2*	32.4	35.7
MIDAS	**30.3**	**26.8**	**30.3**	32.5	*31.8*	34.5
$h = 8$						
$A_0(3)$	48.8	40.8	40.4	41.1	42.3	42.2
$A_1(3)$	42.0	34.6	42.1	45.2	49.4	48.6
SV-Q	46.8	40.5	33.5	30.5	*28.8*	**30.8**
SV-E	39.0	34.1	35.2	34.9	33.5	*32.6*
GARCH	47.2	31.9	**27.7**	**27.9**	**27.6**	33.8
$A_1(3)$ + Reg	49.4	45.8	44.7	44.2	42.9	40.9
SV-Q + Reg	34.6	31.0	30.6	30.9	32.6	37.0
SV-E + Reg	37.9	33.5	36.1	36.5	36.8	37.5
GARCH + Reg	33.6	28.5	28.6	*29.9*	31.1	33.8
HAR-RV	**29.3**	*24.4*	*28.1*	30.6	31.2	35.3
MIDAS	*29.6*	**24.3**	28.4	31.6	30.9	34.0

Table 10: Out-of-sample RMSEs for the volatility prediction in the fixed-parameter approach

Maturity	0.5	1	2	3	5	10
$h = 16$						
$A_0(3)$	49.8	41.2	38.8	39.5	41.3	40.5
$A_1(3)$	40.9	32.9	40.8	44.5	48.8	47.2
SV-Q	50.8	42.4	32.1	*29.4*	**29.1**	**30.7**
SV-E	40.6	34.5	33.9	34.0	33.3	*31.6*
GARCH	49.2	30.9	**27.3**	**28.2**	*29.4*	35.2
$A_1(3)$ + Reg	46.6	44.0	43.0	42.7	41.2	38.6
SV-Q + Reg	35.1	30.6	29.5	29.9	32.4	36.3
SV-E + Reg	37.2	32.4	34.5	34.9	35.5	35.4
GARCH + Reg	33.6	27.0	28.7	30.3	32.5	35.0
HAR-RV	**27.7**	**23.2**	*28.4*	30.6	31.7	34.9
MIDAS	*28.1*	*23.3*	29.0	32.7	32.7	35.3
$h = 32$						
$A_0(3)$	54.6	43.6	37.2	37.6	39.4	36.6
$A_1(3)$	41.4	31.7	39.2	43.4	47.1	42.9
SV-Q	56.7	44.8	30.4	**28.7**	**30.2**	**30.0**
SV-E	46.2	37.5	33.6	34.1	34.0	*30.3*
GARCH	39.4	25.9	**26.5**	*29.1*	32.2	35.1
$A_1(3)$ + Reg	42.0	40.5	40.4	40.1	37.9	34.6
SV-Q + Reg	34.2	29.4	29.2	30.1	32.1	32.8
SV-E + Reg	35.8	31.5	33.7	34.0	33.8	31.6
GARCH + Reg	27.9	22.4	*28.3*	29.8	32.5	33.3
HAR-RV	*25.6*	*22.3*	28.4	30.6	*32.0*	31.4
MIDAS	**25.4**	**22.1**	28.5	31.0	32.7	32.7

Table 10 (continued): Out-of-sample RMSEs for the volatility prediction in the fixed-parameter approach

Root mean squared errors (RMSEs) for the out-of-sample prediction of the h-week ahead yield volatilities (annualized standard deviations) are presented in basis points with $h = \{4, 8, 16, 32\}$. For the term structure and GARCH models, the name alone indicates the results based on equation (31), whereas the name "+ Reg" indicates the results based on equation (32). 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 of each panel are displayed in bold and italic, respectively.

Maturity	0.5	1	2	3	5	10
$h = 4$						
$A_1(3)$ + Reg	38.8	35.4	42.1	43.2	43.4	42.9
SV-Q + Reg	37.1	33.2	34.5	34.1	33.6	34.4
SV-E + Reg	39.8	36.4	38.1	37.7	35.6	36.6
GARCH + Reg	32.3	28.9	31.7	31.4	30.3	32.8
HAR-RV	*31.1*	*27.7*	*30.1*	*30.9*	*29.9*	**32.2**
MIDAS	**30.6**	**27.1**	**29.9**	**30.7**	**29.7**	**32.2**
$h = 8$						
$A_1(3)$ + Reg	36.9	32.9	39.9	41.1	41.2	40.3
SV-Q + Reg	35.9	30.6	30.9	30.5	29.4	*30.2*
SV-E + Reg	38.4	34.3	35.3	35.0	32.7	33.5
GARCH + Reg	31.5	26.5	28.2	**28.3**	27.7	31.3
HAR-RV	**29.8**	*24.9*	**27.5**	*28.8*	**27.2**	**30.0**
MIDAS	**29.8**	**24.6**	*27.6*	29.2	*27.6*	30.7
$h = 16$						
$A_1(3)$ + Reg	35.0	31.2	38.3	39.6	39.9	38.4
SV-Q + Reg	36.3	30.0	29.8	30.0	29.5	*30.6*
SV-E + Reg	37.5	33.1	34.0	33.8	31.6	31.9
GARCH + Reg	31.6	25.7	28.1	**28.4**	*28.6*	32.3
HAR-RV	**27.5**	**23.6**	**27.8**	*28.9*	**26.7**	**28.6**
MIDAS	*28.5*	*23.7*	*28.0*	30.2	28.9	31.8
$h = 32$						
$A_1(3)$ + Reg	33.0	29.6	36.4	37.4	37.5	34.4
SV-Q + Reg	36.0	29.7	28.7	29.2	29.0	29.3
SV-E + Reg	36.4	32.2	33.5	33.5	31.1	29.6
GARCH + Reg	29.3	23.4	**25.7**	**25.7**	*26.3*	30.9
HAR-RV	**25.9**	**22.2**	*26.3*	27.5	**26.0**	**27.0**
MIDAS	*26.2*	*22.4*	26.5	*27.4*	26.7	*29.2*

Table 11: Out-of-sample RMSEs for the volatility prediction in the varying-parameter approach

RMSEs for the out-of-sample prediction of the h-week ahead yield volatilities (annualized standard deviations) are presented in basis points with $h = \{4, 8, 16, 32\}$. For the term structure and GARCH models, the model parameters are fixed at the in-sample estimates and the regression parameters in equation (32) are re-estimated in a rolling-window fashion. For MIDAS, the weighting parameters in equation (30) are fixed at the in-sample estimates and $(a_{h,\tau}, b_{h,\tau})$ in equation (29) are re-estimated. For HAR-RV, all parameters in equation (6) are re-estimated. The best and the second best results in each column of each panel are displayed in bold and italic, respectively. The out-of-sample period is from April 16, 2003 to May 27, 2009.

Maturity	0.5	1	2	3	5	10
$h = 4$						
+ $A_1(3)$	1.036	0.867	0.482	0.324	−0.056	−0.374
	(0.437)	(0.429)	(0.325)	(0.314)	(0.259)	(0.267)
+ SV-Q	0.360	0.427	0.439	0.461	0.466	0.361
	(0.142)	(0.157)	(0.115)	(0.114)	(0.126)	(0.174)
+ SV-E	0.555	0.516	0.358	0.420	0.373	0.279
	(0.129)	(0.204)	(0.137)	(0.128)	(0.120)	(0.116)
+ GARCH	0.156	0.553	0.355	0.528	0.443	0.832
	(0.120)	(0.141)	(0.202)	(0.208)	(0.197)	(0.270)
$h = 8$						
+ $A_1(3)$	1.164	1.066	0.650	0.482	−0.022	−0.418
	(0.527)	(0.535)	(0.414)	(0.377)	(0.313)	(0.320)
+ SV-Q	0.362	0.459	0.518	0.542	0.506	0.329
	(0.159)	(0.198)	(0.129)	(0.126)	(0.146)	(0.213)
+ SV-E	0.615	0.563	0.432	0.482	0.413	0.314
	(0.170)	(0.248)	(0.160)	(0.142)	(0.141)	(0.142)
+ GARCH	0.273	0.718	0.460	0.578	0.422	0.802
	(0.132)	(0.193)	(0.229)	(0.237)	(0.213)	(0.305)

Table 12: Estimates (standard errors) of the slope coefficient on the model forecast in the combined regression

Maturity	0.5	1	2	3	5	10
$h = 16$						
+ $A_1(3)$	1.375	1.390	0.842	0.665	0.020	−0.566
	(0.631)	(0.681)	(0.527)	(0.455)	(0.375)	(0.342)
+ SV-Q	0.302	0.463	0.525	0.563	0.470	0.225
	(0.161)	(0.236)	(0.161)	(0.147)	(0.169)	(0.288)
+ SV-E	0.613	0.580	0.488	0.534	0.446	0.354
	(0.229)	(0.315)	(0.207)	(0.169)	(0.169)	(0.179)
+ GARCH	0.304	0.759	0.444	0.626	0.443	0.909
	(0.170)	(0.290)	(0.302)	(0.291)	(0.246)	(0.541)
$h = 32$						
+ $A_1(3)$	1.635	1.632	0.972	0.840	−0.013	−0.822
	(0.739)	(0.883)	(0.742)	(0.595)	(0.515)	(0.357)
+ SV-Q	0.313	0.513	0.542	0.565	0.501	0.235
	(0.167)	(0.247)	(0.179)	(0.166)	(0.209)	(0.312)
+ SV-E	0.542	0.513	0.481	0.516	0.442	0.343
	(0.249)	(0.350)	(0.232)	(0.198)	(0.201)	(0.207)
+ GARCH	0.289	0.978	0.485	0.661	0.524	0.561
	(0.196)	(0.303)	(0.414)	(0.456)	(0.311)	(0.913)

Table 12 (continued): Estimates (standard errors) of the slope coefficient on the model forecast in the combined regression

The combined forecasting regression is based on the HAR-RV model:

$$\sqrt{\frac{RV_{t,t+h\Delta,\tau}}{h\Delta}} = a_{h,\tau} + \sum_{i=\{4,8,16,32\}} b_{h,\tau,i}\sqrt{\frac{RV_{t-i\Delta,t,\tau}}{i\Delta}} + c_{h,\tau} forecast_{t,h,\tau} + u^c_{t+h\Delta,\tau} \ ,$$

where $forecast_{t,h,\tau}$ is taken from the term structure and GARCH models. The table presents the estimates of $c_{h,\tau}$. The standard errors, computed using the method of Newey and West (1987) with 32 lags, are in parenthesis. The in-sample data from January 4, 1991 to April 9, 2003 are used for the estimation.

Maturity	0.5	1	2	3	5	10
$h = 4$						
+ $A_1(3)$	3.7	5.0	3.3	3.2	−0.4	−1.1
+ SV-Q	−2.4	0.6	−3.7	−7.8	−10.5	−10.6
+ SV-E	0.9	2.3	1.7	2.5	0.1	−4.7
+ GARCH	1.3	6.2	−0.4	−3.6	−2.0	−3.5
$h = 8$						
+ $A_1(3)$	6.0	10.8	7.4	6.3	−0.2	−2.0
+ SV-Q	−2.0	1.3	−5.4	−11.0	−14.3	−11.5
+ SV-E	3.5	5.7	4.0	4.0	0.3	−6.2
+ GARCH	3.2	11.6	−1.9	−5.5	−2.5	−3.6
$h = 16$						
+ $A_1(3)$	12.1	21.4	11.4	9.3	0.2	−2.3
+ SV-Q	2.4	5.0	−5.5	−10.2	−11.6	−7.1
+ SV-E	9.9	10.5	5.7	4.5	0.1	−7.3
+ GARCH	4.6	11.9	−1.4	−4.6	−1.7	−2.5
$h = 32$						
+ $A_1(3)$	25.0	30.7	13.0	10.1	−0.1	0.6
+ SV-Q	8.0	7.3	−6.7	−9.3	−8.4	−5.1
+ SV-E	18.9	14.1	6.5	4.0	−0.5	−6.1
+ GARCH	−1.5	0.9	−1.9	−4.3	−1.1	0.2

Table 13: % changes in the out-of-sample RMSE by combined forecasts in the fixed-parameter approach

A combined forecast is based on the HAR-RV model, where a forecast taken from the term structure and GARCH models is added. The table presents % changes in the RMSE from the HAR-RV forecast alone to the combined forecast. A negative (positive) number indicates that the combined forecast decreases (increases) the RMSE. All parameters are fixed at the in-sample estimates throughout the out-of-sample period from April 16, 2003 to May 27, 2009.

Maturity	0.5	1	2	3	5	10
$h = 4$						
+ $A_1(3)$	0.1	−0.6	−0.2	0.3	1.1	0.0
+ SV-Q	−3.5	−2.7	−3.6	−5.5	−6.4	−9.8
+ SV-E	0.2	0.4	1.0	2.9	5.2	0.8
+ GARCH	1.0	1.1	0.3	−1.8	−1.2	−1.7
$h = 8$						
+ $A_1(3)$	−0.2	−1.4	−0.2	0.7	1.7	−0.5
+ SV-Q	−3.5	−3.5	−5.9	−8.3	−11.0	−15.7
+ SV-E	0.8	1.1	0.8	3.8	7.3	1.8
+ GARCH	1.6	2.0	−1.1	−4.4	−2.2	−2.3
$h = 16$						
+ $A_1(3)$	−0.8	−2.5	−0.5	0.8	1.9	−0.8
+ SV-Q	−0.7	−2.6	−6.1	−6.2	−8.8	−12.1
+ SV-E	1.9	1.7	1.3	5.2	8.4	3.2
+ GARCH	2.9	3.0	−0.2	−3.9	−1.1	−0.9
$h = 32$						
+ $A_1(3)$	−2.3	−4.0	−2.1	−1.9	0.2	−0.8
+ SV-Q	3.4	−0.8	−9.0	−6.4	−6.8	−7.7
+ SV-E	1.1	1.8	2.6	6.6	6.8	2.8
+ GARCH	0.8	3.7	−1.0	−6.0	−2.0	0.1

Table 14: % changes in the out-of-sample RMSE by combined forecasts in the varying-parameter approach

A combined forecast is based on the HAR-RV model, where a forecast taken from the term structure and GARCH models is added. The table presents % changes in the RMSE from the HAR-RV forecast alone to the combined forecast. A negative (positive) number indicates that the combined forecast decreases (increases) the RMSE. For the term structure and GARCH models, the model parameters are fixed at the in-sample estimates and the parameters in the combined HAR-RV regression are re-estimated in a rolling-window fashion. The out-of-sample period is from April 16, 2003 to May 27, 2009.

Maturity	0.5	1	2	3	5	10
Panel A: RMSEs based on eq. (31) (without the forecasting regression)						
SV-Q	33.7	35.0	32.8	33.9	28.3	26.9
NC1	33.7	34.4	32.7	33.7	28.4	26.9
NC2	34.5	35.2	33.7	35.0	30.3	29.1
NC3	35.3	35.0	32.3	33.4	28.2	28.0
SV-E	32.0	33.8	32.8	33.6	28.4	27.8
NC1	32.0	33.3	32.8	33.6	28.6	27.6
NC2	31.7	33.0	31.8	32.6	27.7	26.7
NC3	32.2	33.0	31.6	32.4	27.2	26.7
Panel B: RMSEs based on eq. (32) (with the forecasting regression)						
SV-Q + Reg	28.7	32.5	31.1	32.4	27.5	26.3
NC1	28.7	32.5	31.1	32.4	27.5	26.3
NC2	28.9	32.6	31.0	32.4	27.6	26.4
NC3	29.7	33.4	31.3	32.4	27.3	26.1
SV-E + Reg	27.8	31.9	31.5	32.3	27.1	25.8
NC1	27.8	31.9	31.5	32.3	27.1	25.8
NC2	27.8	31.8	31.5	32.3	27.1	25.8
NC3	27.9	31.9	31.6	32.3	26.9	25.7

Table 15: In-sample RMSEs for the 4-week ahead volatility prediction with and without the non-arbitrage

Root mean squared errors (RMSEs) for the in-sample prediction of the 4-week ahead yield volatilities (annualized standard deviations) are presented in basis points, where panels A and B present the results based on equations (31) and (32), respectively. The rows labeled SV-Q and SV-E show the original results with the non-arbitrage imposed. NCi ($i = 1, 2, 3$) indicates "No Constraint", beyond the linear regression (34), on the cross-sectional relation with the number corresponding to the stage where the non-arbitrage is involved: it is involved with both identification of the state vector X_t and estimation of the covariance matrix Σ_t in stage 1; with only the latter in stage 2; and with none in stage 3, where X_t is given by the conventional level, slope, and curvature factors. The in-sample period is from January 4, 1991 to April 9, 2003.

	$y_{0.5}$	y_1	y_2	y_3	y_5	y_{10}
Panel A: Parameter estimates for $dy_{t,\tau} = (\alpha + \beta y_{t,\tau})dt + \sigma\sqrt{y_{t,\tau}}dW_t$						
$\alpha \times 10^2$	−0.270	−0.278	0.253	0.787	1.503	2.626
	(0.689)	(1.001)	(2.404)	(1.723)	(1.803)	(1.992)
β	−0.043	−0.040	−0.133	−0.215	−0.311	−0.446
	(0.125)	(0.178)	(0.405)	(0.282)	(0.280)	(0.289)
$\sigma \times 10^2$	3.248	3.745	4.229	4.212	4.019	3.774
	(0.091)	(0.105)	(0.119)	(0.118)	(0.113)	(0.106)
Panel B: Means (standard deviations) of % differences between two conditional standard deviations with and without β						
$h = 4$	0.38	0.36	0.67	0.96	1.31	1.80
	(0.07)	(0.07)	(0.04)	(0.09)	(0.11)	(0.13)
$h = 8$	0.76	0.71	1.34	1.92	2.63	3.62
	(0.15)	(0.13)	(0.08)	(0.18)	(0.22)	(0.26)
$h = 16$	1.53	1.43	2.70	3.87	5.29	7.29
	(0.30)	(0.27)	(0.16)	(0.36)	(0.46)	(0.53)
$h = 32$	3.11	2.90	5.45	7.82	10.70	14.75
	(0.62)	(0.57)	(0.33)	(0.74)	(0.96)	(1.12)

Table A: Impact of omitting the mean reversion on the realized volatility
In Panel A, the estimates (standard errors) for the CIR-type square-root process fitted to each yield are presented. In Panel B, the means (standard deviations) for the time series of

$$100\left(\sqrt{\frac{\mathrm{var}_t[y_{t+h\Delta,\tau};\sigma]}{\mathrm{var}_t[y_{t+h\Delta,\tau};\alpha,\beta,\sigma]}}-1\right)$$

are presented, where $\mathrm{var}_t[y_{t+h\Delta,\tau};\alpha,\beta,\sigma]$ and $\mathrm{var}_t[y_{t+h\Delta,\tau};\sigma]$ are presented in equations (38) and (39). The in-sample data from January 4, 1991 to April 9, 2003 are used for the estimation.

Panel A: the first-step comparison							
		$y_{0.5}$	y_1	y_2	y_3	y_5	y_{10}
Minimum	Level	0.00	0.00	0.00	0.00	0.01	0.25
	Slope	0.00	0.01	0.00	0.00	0.00	0.16
	Curvature	0.01	0.02	−0.01	−0.02	0.00	0.39
Median	Level	0.00	0.00	0.00	0.01	0.06	1.20
	Slope	0.00	0.00	0.00	0.01	0.05	0.78
	Curvature	0.00	0.00	0.00	0.00	0.03	0.73
Maximum	Level	0.00	0.00	0.00	0.01	0.06	1.29
	Slope	0.00	0.00	0.00	0.01	0.14	2.20
	Curvature	0.00	−0.01	0.00	0.01	0.04	0.62

Panel B: the second-step comparison							
		r	θ	ϵ	y_1	y_3	y_5
Minimum	Level	−0.09	−0.39	0.80	0.03	−0.06	−0.14
	Slope	−0.06	−0.25	0.52	0.02	−0.04	−0.10
	Curvature	−0.16	−0.62	1.34	0.06	−0.10	−0.23
Median	Level	−0.41	−1.90	3.89	0.13	−0.26	−0.66
	Slope	−0.27	−1.24	2.55	0.08	−0.17	−0.42
	Curvature	−0.25	−1.16	2.37	0.08	−0.16	−0.41
Maximum	Level	−0.44	−2.03	4.16	0.14	−0.28	−0.71
	Slope	−0.75	−3.46	7.09	0.23	−0.47	−1.18
	Curvature	−0.22	−0.99	2.03	0.06	−0.13	−0.34

Table C1: Comparison to the closed-form solution using the $A_0(3)$ model

Panel C: the third-step comparison							
		r	θ	ϵ	y_1	y_3	y_5
Minimum	Level	0.04	−4.98	5.24	0.02	−0.13	−0.39
	Slope	0.71	−1.28	0.34	−0.16	0.18	0.22
	Curvature	2.10	−2.28	−8.07	−0.46	0.52	0.64
Median	Level	0.50	−0.88	−7.17	−0.10	0.06	−0.10
	Slope	−0.37	2.78	−4.07	0.07	−0.04	−0.01
	Curvature	0.03	−1.78	−1.90	0.01	−0.07	−0.25
Maximum	Level	1.26	3.13	−13.49	−0.31	0.40	0.52
	Slope	0.02	−1.91	−8.58	0.03	−0.17	−0.63
	Curvature	−0.78	2.98	−1.53	0.16	−0.13	−0.11

Table C1 (continued): Comparison to the closed-form solution using the $A_0(3)$ model

Approximation errors, defined as the difference between the approximate and true yields or state variables, are presented in bps. The errors are evaluated at nine states taken from the actual data, where the level, slope, and curvature factors take the minimum, median, or maximum value. Panel A presents the results for the first-step comparison, where the true values of both the parameter and state vectors are given as input for the approximation. Panel B presents the results for the second-step comparison, where the true value of the parameter vector alone is given. Panel C presents the results for the third-step comparison, where no prior information is given.

	True	Approximation		Diff.	%Diff.
	Values	Est.	S.E.		
κ_1	0.947	0.966	(0.088)	−0.019	−2.0
κ_2	0.017	0.019	(0.002)	−0.002	−13.6
κ_3	0.957	0.964	(0.089)	−0.006	−0.7
$\bar{\theta}$	0.146	0.135	(0.006)	0.011	7.5
$\sigma_{11} \times 10^4$	0.722	0.730	(0.030)	−0.008	−1.1
$\sigma_{22} \times 10^4$	1.220	1.257	(0.066)	−0.037	−3.0
$\sigma_{33} \times 10^4$	8.036	8.240	(0.441)	−0.204	−2.5
$\sigma_{12} \times 10^4$	0.095	0.099	(0.036)	−0.004	−3.8
$\sigma_{13} \times 10^4$	−0.985	−1.019	(0.115)	0.034	−3.4
λ_2^0	0.030	0.030	(0.019)	0.000	−0.9
λ_3^0	−0.013	−0.014	(0.008)	0.001	−4.9
λ_{12}^1	−0.195	−0.191	(0.037)	−0.004	2.0
λ_{22}^1	−0.514	−0.518	(0.285)	0.004	−0.8
$\varsigma \times 10^2$	0.061	0.061	(0.001)	0.000	0.4

Table C2: Parameter estimates of the $A_0(3)$ model by the approximation method

The true parameter values are taken from the estimates presented in Table 4. Est. and S.E. are the estimates and standard errors, respectively, obtained by the approximation method. Diff. and % Diff. are differences and % differences between the true values and estimates, respectively.

Panel A: Based on the SV-Q model							
		$y_{0.5}$	y_1	y_2	y_3	y_5	y_{10}
Minimum	Level	0.11	0.04	0.12	0.13	−0.03	0.19
	Slope	0.17	0.22	0.35	0.44	0.55	0.79
	Curvature	0.37	0.13	−0.15	−0.11	−0.04	0.99
Median	Level	−0.03	−0.02	0.00	0.03	0.10	1.41
	Slope	−0.18	−0.13	−0.08	−0.02	0.01	0.45
	Curvature	−0.05	−0.05	−0.08	−0.08	−0.16	0.14
Maximum	Level	0.06	0.14	0.23	0.27	0.13	1.34
	Slope	−0.14	−0.14	−0.19	−0.19	−0.08	2.64
	Curvature	−0.23	−0.21	−0.14	−0.04	−0.05	0.83
Panel B: Based on the SV-E model							
		$y_{0.5}$	y_1	y_2	y_3	y_5	y_{10}
Minimum	Level	0.12	0.11	0.09	0.10	0.07	0.46
	Slope	0.19	0.16	0.12	0.07	0.04	0.42
	Curvature	0.39	0.35	0.20	0.10	0.08	1.39
Median	Level	−0.03	−0.05	−0.08	−0.04	−0.03	1.38
	Slope	−0.20	−0.16	−0.09	−0.06	−0.04	0.30
	Curvature	−0.04	−0.04	−0.04	−0.08	−0.09	0.16
Maximum	Level	0.04	0.02	0.00	−0.06	−0.01	0.85
	Slope	−0.18	−0.18	−0.16	−0.12	0.11	2.37
	Curvature	−0.25	−0.20	−0.10	−0.03	0.02	0.41

Table C3: Comparison to the MC solution using the SV-Q and SV-E models
Approximation errors, defined as the difference between the approximate and MC yields, are presented in bps. The errors are evaluated at nine states taken from the actual data, where the level, slope, and curvature factors take the minimum, median, or maximum value. Panels A and B present the results based on the SV-Q and SV-E models, respectively.

Maturity		0.5	1	2	3	5	10
Panel A: Based on the SV-Q model							
	Level	−0.16	−0.16	−0.17	−0.15	−0.13	−0.12
Minimum	Slope	−0.24	−0.31	−0.30	−0.28	−0.23	−0.22
	Curvature	−0.46	−0.61	−0.67	−0.63	−0.64	−0.76
	Level	0.00	−0.06	−0.06	−0.06	−0.05	−0.04
Median	Slope	−0.14	−0.09	−0.08	−0.06	0.01	0.10
	Curvature	−0.06	−0.03	−0.05	−0.02	−0.01	−0.02
	Level	0.44	0.49	0.51	0.45	0.33	0.18
Maximum	Slope	−0.06	−0.12	−0.09	−0.08	−0.06	−0.01
	Curvature	−0.01	−0.07	−0.09	−0.05	0.03	0.22
Panel B: Based on the SV-E model							
	Level	0.06	0.04	0.04	0.02	0.01	−0.07
Minimum	Slope	0.05	0.03	0.01	0.01	−0.01	−0.07
	Curvature	0.07	0.05	0.01	0.05	0.06	−0.38
	Level	0.02	0.02	0.02	0.02	0.01	−0.07
Median	Slope	−0.09	−0.07	−0.04	−0.02	−0.01	0.01
	Curvature	0.02	0.02	0.01	0.02	0.01	−0.05
	Level	0.00	−0.05	−0.07	−0.08	−0.08	−0.22
Maximum	Slope	−0.04	−0.01	−0.02	−0.03	−0.01	0.01
	Curvature	−0.06	−0.04	−0.01	0.00	0.02	0.03

Table C4: Accuracy to $E_t[\tilde{Y}(X_{t+h\Delta}, \tau)]$ for $h = 32$
Approximation errors, defined as the difference between the approximate and MC moments, are presented in bps. The errors are evaluated at nine states taken from the actual data, where the level, slope, and curvature factors take the minimum, median, or maximum value. Panels A and B present the results based on the SV-Q and SV-E models, respectively.

Maturity		0.5	1	2	3	5	10
Panel A: Based on the SV-Q model							
	Level	0.63	0.72	0.81	0.74	0.59	0.41
Minimum	Slope	0.87	1.36	1.53	1.41	1.00	0.49
	Curvature	−0.21	0.11	0.53	0.68	0.50	0.33
	Level	0.56	0.74	0.79	0.75	0.64	0.52
Median	Slope	−0.22	0.09	0.31	0.39	0.50	0.46
	Curvature	0.64	0.73	0.58	0.33	−0.02	−0.31
	Level	−1.52	−1.90	−1.89	−1.53	−0.88	−0.20
Maximum	Slope	−0.77	−1.01	−1.20	−1.26	−1.21	−1.19
	Curvature	0.05	−0.34	−0.79	−0.90	−0.85	−0.82
Panel B: Based on the SV-E model							
	Level	−0.51	−0.62	−0.62	−0.55	−0.43	−0.28
Minimum	Slope	0.37	0.33	0.29	0.27	0.24	0.23
	Curvature	0.98	0.53	0.10	−0.08	−0.23	−1.28
	Level	0.06	−0.06	−0.16	−0.16	−0.12	−0.15
Median	Slope	−0.23	0.02	0.13	0.10	0.01	−0.09
	Curvature	−0.12	−0.15	−0.12	−0.03	0.12	0.22
	Level	1.99	1.76	1.49	1.28	1.07	0.19
Maximum	Slope	0.08	0.01	−0.12	−0.32	−0.66	−1.37
	Curvature	−0.02	−0.17	−0.36	−0.44	−0.52	−0.57

Table C5: Accuracy to $\sqrt{\mathbf{var}_t[\tilde{Y}(X_{t+h\Delta}, \tau)]/h\Delta}$ for $h = 32$
Approximation errors, defined as the difference between the approximate and MC moments, are presented in bps. The errors are evaluated at nine states taken from the actual data, where the level, slope, and curvature factors take the minimum, median, or maximum value. Panels A and B present the results based on the SV-Q and SV-E models, respectively.

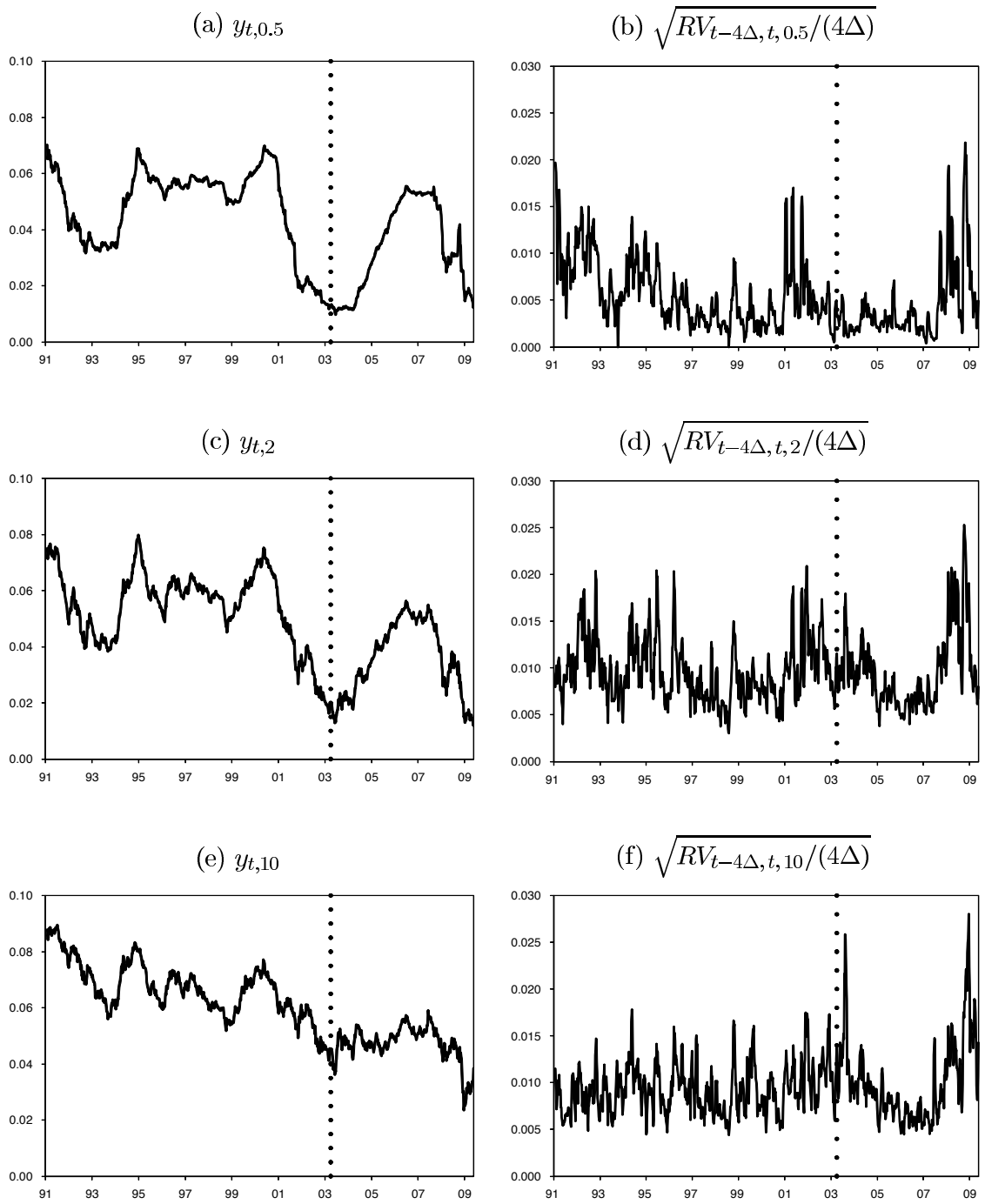

Figure 1: Time series of interest rates and realized volatilities over 1991–2009
On the left panels are the time series of the six-month, two-year, and ten-year yields. On the left panels are the time series of the four-week realized volatility (annualized standard deviation) of the corresponding yields. The vertical dotted line separates the in-sample and out-of-sample periods.

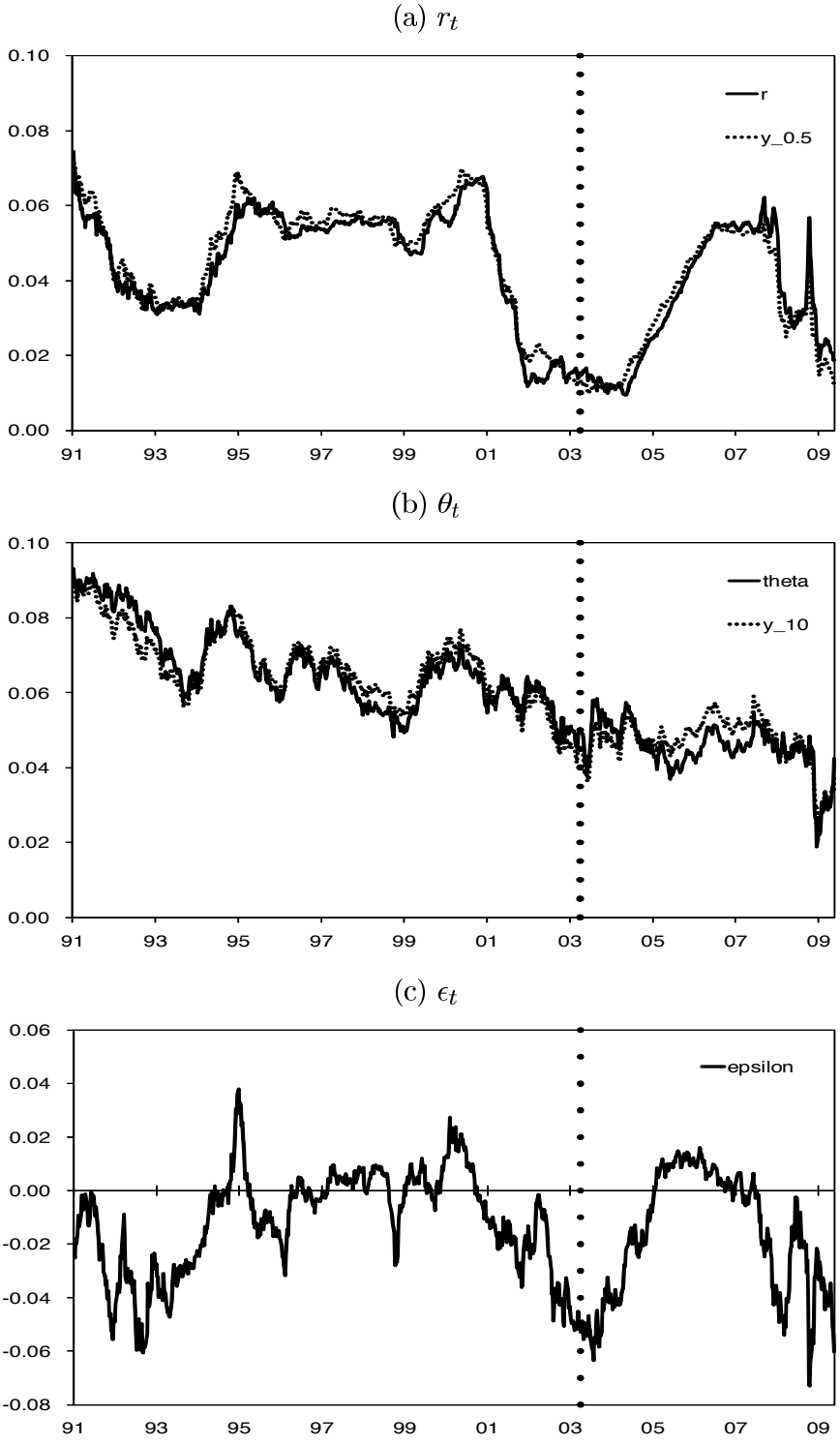

Figure 2: Time series of the state variables implied by the $A_0(3)$ model
In panels (a) and (b), the time series of the six-month and ten-year yields, respectively, are also plotted by the dotted line. The vertical dotted line separates the in-sample and out-of-sample periods.

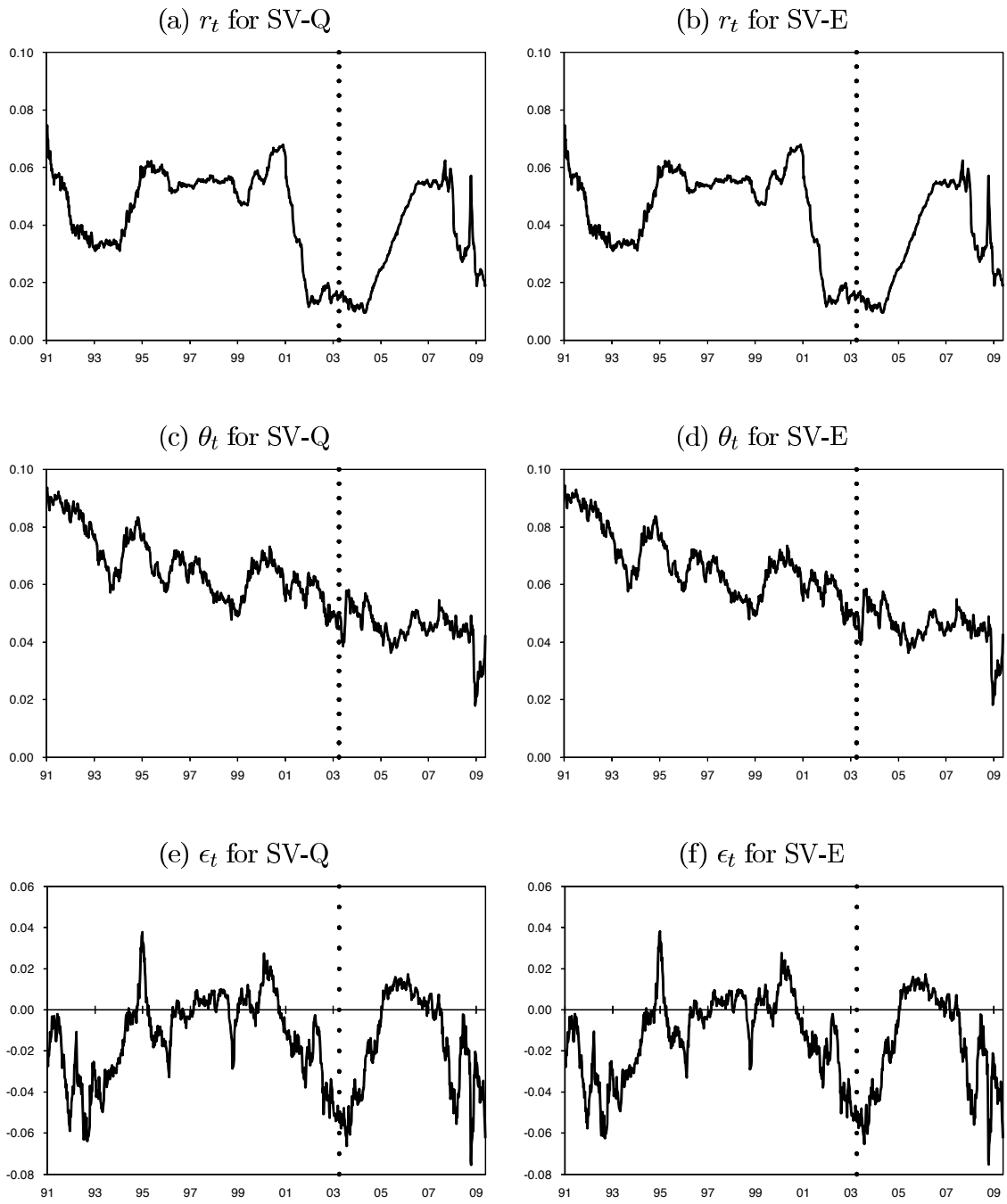

Figure 3: Time series of the state variables implied by the SV-Q and SV-E models

The left (right) panels present the time series for the SV-Q (SV-E) model. The vertical dotted line separates the in-sample and out-of-sample periods.

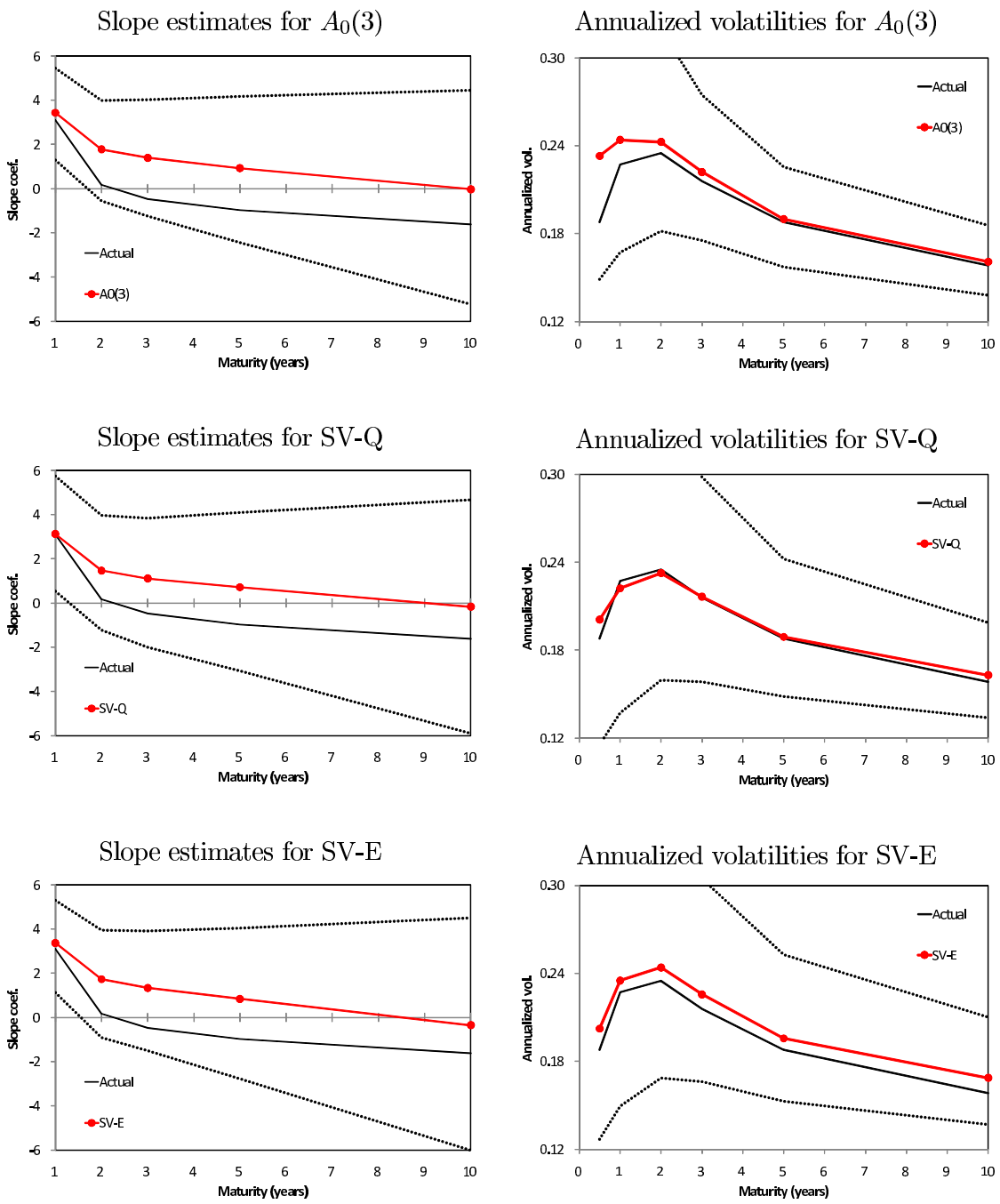

Figure 4: EHT regression coefficients and annualized unconditional volatilities In each panel, the solid line represents the actual estimate. The marked and dotted lines represent the model-implied estimate and 90% confidence band obtained from 1,000 sets of simulated data, respectively. The left panels present the slope estimate of equation (26) with τ_1 replaced by $\Delta(= 1/52)$ and $y_{t+\Delta,\tau_2-\Delta}$ by $y_{t+\Delta,\tau_2}$, where $\tau_2 = \{1, 2, 3, 5, 10\}$. The right panels present annualized standard deviations of changes in logarithmic yields.

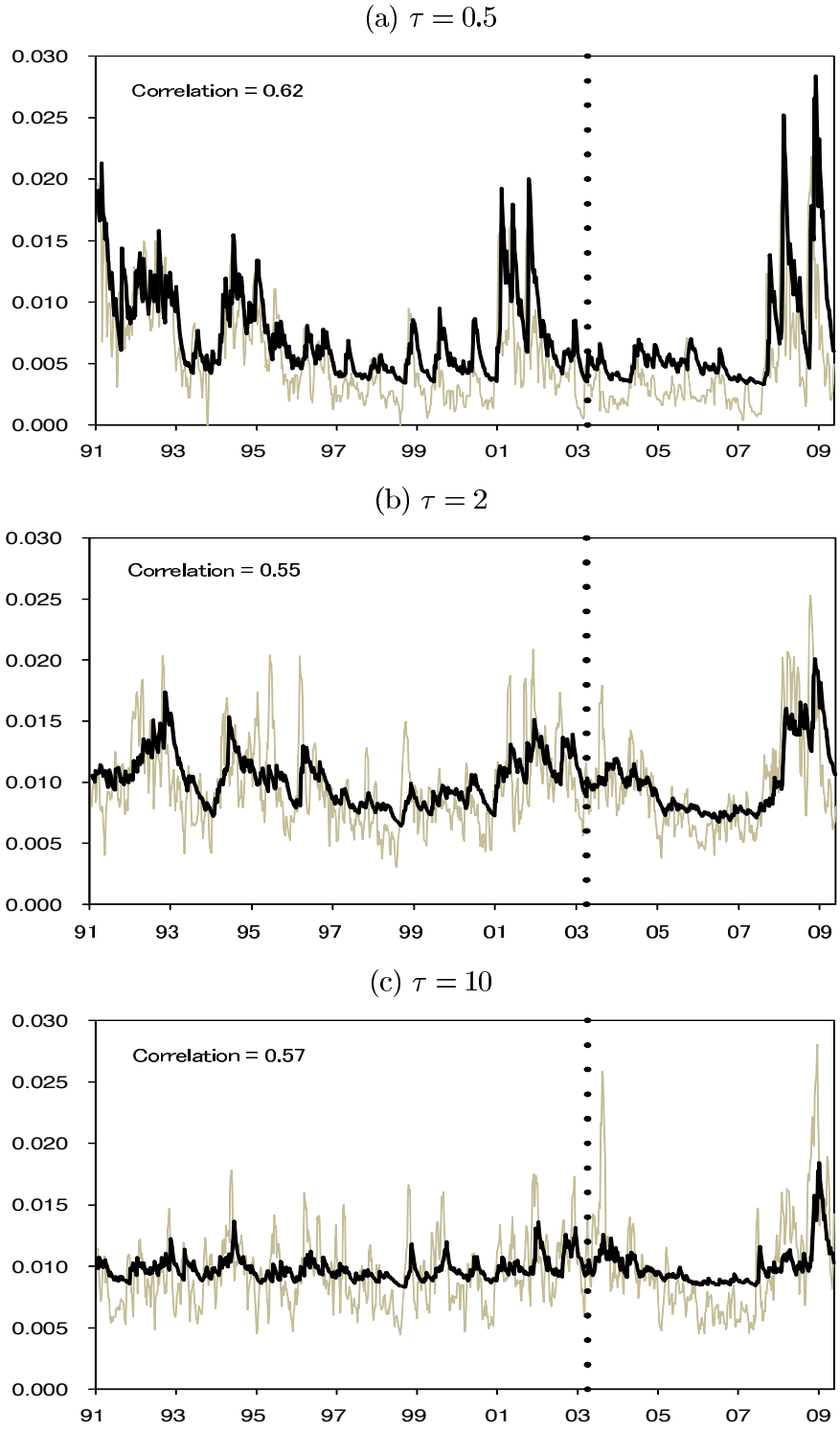

Figure 5: Time series of the four-week ahead volatility forecast by the GARCH (1,1) model

The model forecast (the thick line) and the corresponding realized measure (the thin line) for the six-month, two-year, and ten-year yields are displayed. The vertical dotted line separates the in-sample and out-of-sample periods.

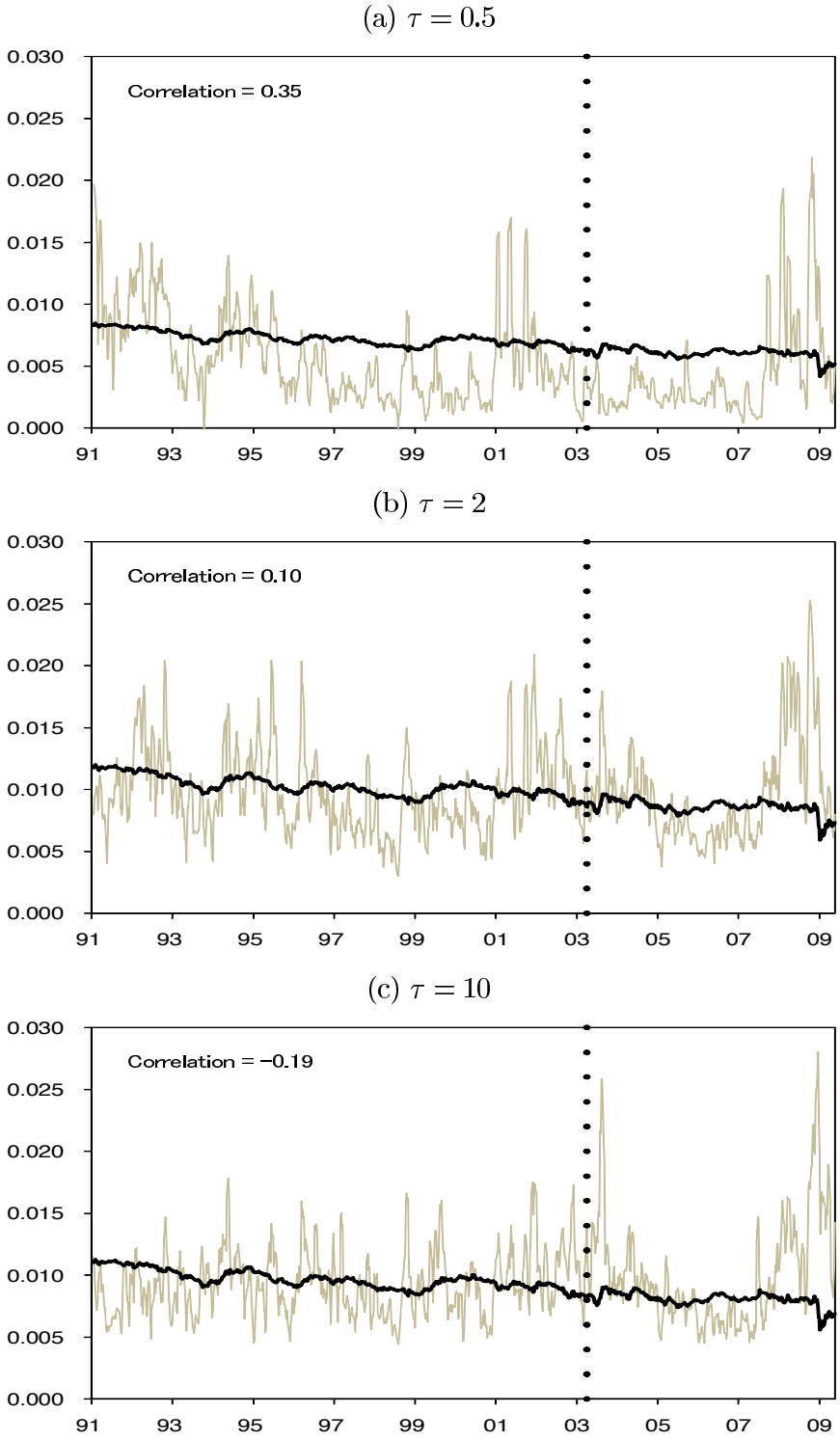

Figure 6: Time series of the four-week ahead volatility forecast by the $A_1(3)$ model

The model forecast (the thick line) and the corresponding realized measure (the thin line) for the six-month, two-year, and ten-year yields are displayed. The vertical dotted line separates the in-sample and out-of-sample periods.

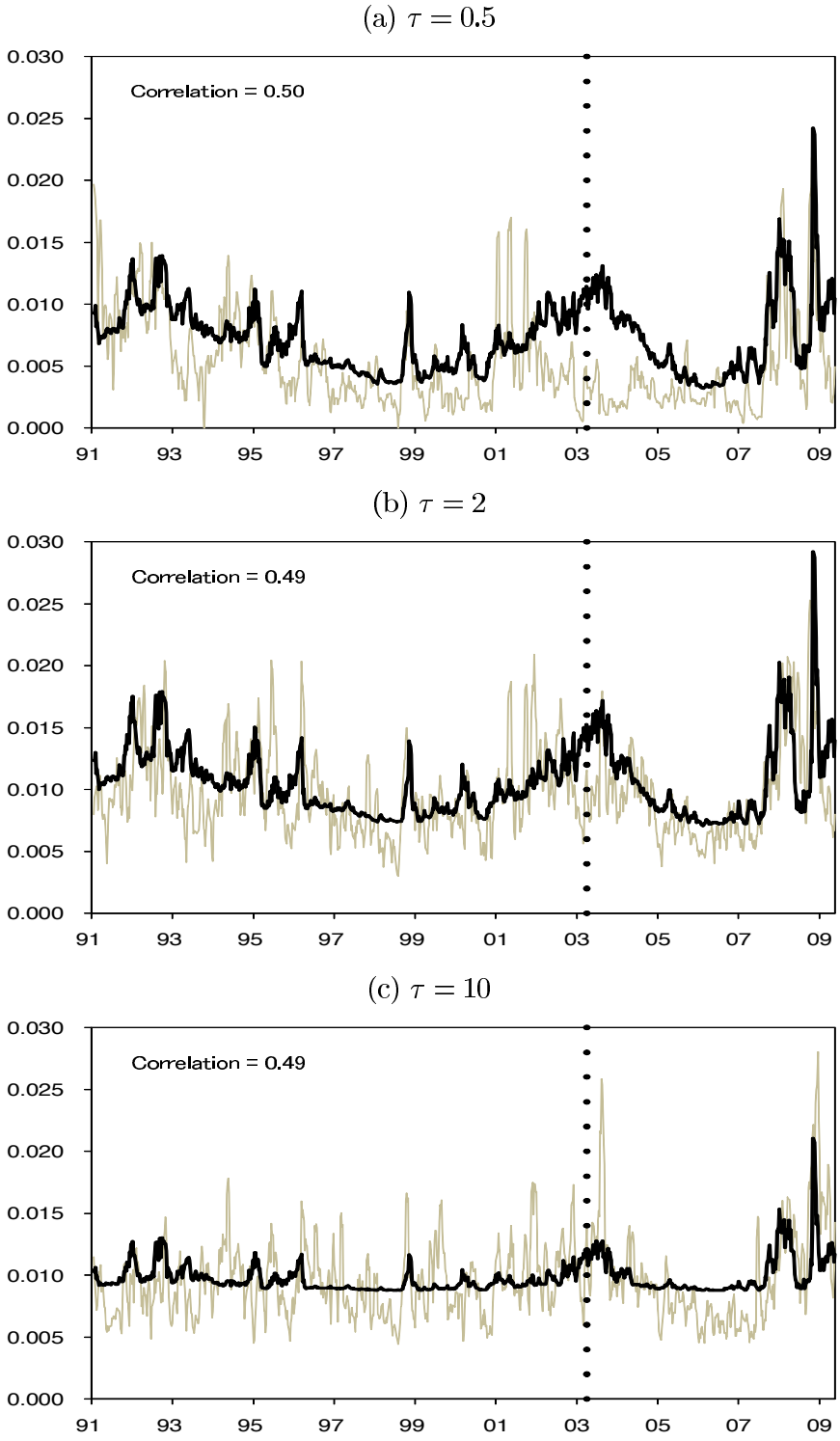

Figure 7: Time series of the four-week ahead volatility forecast by the SV-Q model

The model forecast (the thick line) and the corresponding realized measure (the thin line) for the six-month, two-year, and ten-year yields are displayed. The vertical dotted line separates the in-sample and out-of-sample periods.

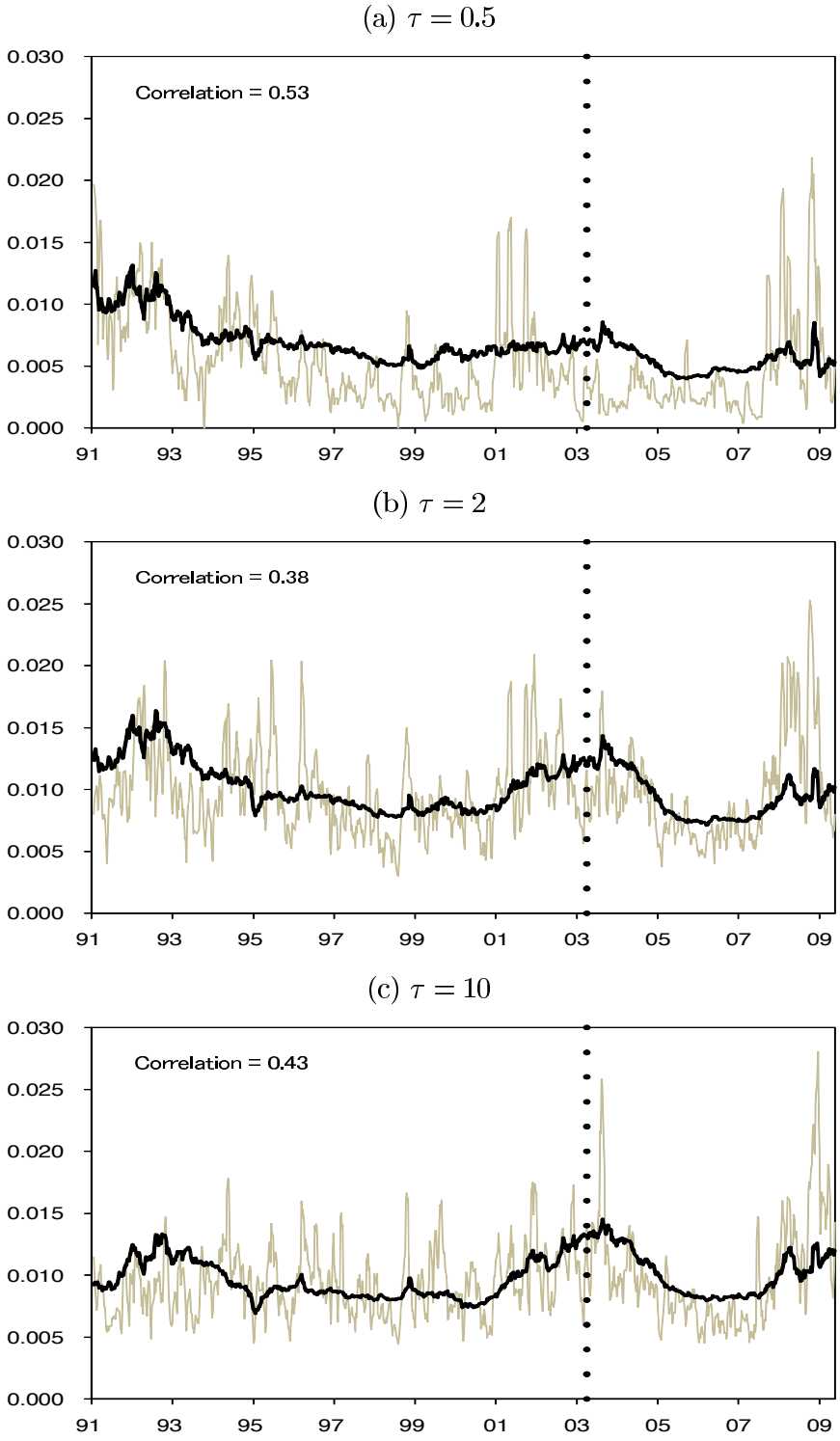

Figure 8: Time series of the four-week ahead volatility forecast by the SV-E model

The model forecast (the thick line) and the corresponding realized measure (the thin line) for the six-month, two-year, and ten-year yields are displayed. The vertical dotted line separates the in-sample and out-of-sample periods.