

Impact of No-arbitrage on Interest Rate Dynamics

Hideyuki Takamizawa *

This draft: May 17, 2015

(First draft: January 2013)

Abstract

By imposing no-arbitrage condition (NAC), the volatility of changes in interest rates is linked to the cross-section of interest rates. Due to this link, the cross-section may have impact on estimation and prediction of volatility using interest rate data. Furthermore, the volatility may have impact on identification of latent factors from the cross-section. In this study, these impacts arising from the NAC are examined, and found to be minor or mitigated without much difficulty. It follows that the resulting dynamics of interest rates do not differ much between estimation with and without imposing the NAC.

Keywords: Term structure, Interest rate, Volatility, Non-affine model, Approximation.

JEL codes: C58, E43, G12, G17.

*Graduate School of Commerce and Management, Hitotsubashi University, Kunitachi Tokyo 186-8601, JAPAN. This draft was written while I was a visiting scholar at Kellogg School of Management, Northwestern University. I am grateful to Torben Andersen, Viktor Todorov, and seminar participants at Osaka University, the University of Tokyo, and YNU-Nanzan finance workshop for comments and discussions. E-mail: Hi.Takamizawa@r.hit-u.ac.jp

1 Introduction

The imposition of no-arbitrage condition (NAC) derives a theoretical relationship in the cross-section of interest rates, which in turn may constitute constraints for modeling and estimating interest rate dynamics. In particular, since it relates the volatility of changes in interest rates to the cross-section of interest rates,¹ the behavior of volatility estimated from both time-series and cross-sectional properties of interest rate data may differ from that estimated from time-series properties alone. Furthermore, if the volatility is driven partially by the same latent factors driving interest rates, the interpretation of these factors identified from the cross-section of interest rates may not be straightforward: It would be difficult to immediately answer whether they are interest-rate specific or volatility specific.

This study investigates the impact of the NAC on the identification of latent factors and the estimation and prediction of volatility using interest rate data. To clarify the research objectives, the following questions, resulting from the imposition of the NAC, are raised:

- Q1.** Does the modeling of volatility practically influence the identification of latent factors from the cross-section of interest rates?
- Q2.** Does the cross-section of interest rates practically influence the estimation and prediction of volatility?

The answers to these questions have important implications for the estimation of the dynamics of interest rates and their volatilities. If the answer to Q1 is no, the identification of factors and the modeling of volatility can be conducted separately. Specifically, it would be possible to first identify latent factors, and then consider realistic volatility models that accommodate time-varying and persistent features. In addition, if the answer to Q2 is no, the behavior of volatility can be estimated using time-series properties of the data alone, without explicitly taking account of cross-sectional properties of the data constrained by the NAC. Taken together, the resulting dynamics of interest rates do not differ much between the estimation with and without imposing the NAC. Meanwhile, the latter estimation allows for flexible modeling and yet is computationally less demanding.

¹This holds true for model parameters as well as latent factors. For example, even with the existence of volatility factors that cannot be extracted from the cross-section of interest rates, the so-called unspanned volatility (USV) factors, parameters regarding the second moments of the other factors are still related to the cross-section of interest rates; see, e.g., Colline-Dufresne and Goldstein (2002, p.1705).

Many studies without considering the NAC seem to take it as granted that the answers to Q1 and Q2 are “no.” In these studies, interest rate factors are first obtained from observed variables by combining key interest rates or applying the principal component analysis. Or, as implemented by Diebold and Li (2006), they are obtained by fitting a descriptive model such as the one proposed by Nelson and Siegel (NS) (1987) to the yield curve at each point in time; see also Almeida and Vicente (2008) adopting a similar approach. In any case, cross-sectional constraints due to no-arbitrage are not always explicitly imposed in the factor identification process. Next, the dynamics of the factors, potentially augmented with time-varying volatility, are estimated using time-series (pooled) data. Again, cross-sectional constraints due to no-arbitrage are not typically imposed in the model estimation process. This convention also holds for the estimation of the dynamics of interest rate volatility using high-frequency data; see, e.g., Andersen and Benzoni (2010).

The literature does not appear to be of little concern to Q1 and Q2, however. In fact, a more recent study by Christensen et al. (2009, 2011) develops the dynamic NS model on which the NAC is imposed. Nevertheless, a thorough investigation of these questions has yet to be conducted.²

Theoretically, the NAC links conditional second moments of the factors to the cross-section of interest rates. Practically, however, the strength of the link depends on how a model is specified and estimated. This means that in order to examine a practical impact of the NAC, some *specific* setting regarding models and estimation methods is unavoidable. Then, a setting should be such that findings and implications obtained there are generalizable to some extent. Since this study raises Q1 and Q2 to examine the validity of the aforementioned approaches to modeling and estimating interest rate dynamics, and these approaches normally stand on the ground that the answers to Q1 and Q2 are no, a setting that potentially increases the likelihood of answering “yes” seems appropriate. Specifically, such an *unfavorable* setting consists of

- (i) model with the volatility that is tightly linked to the cross-section of interest rates;
- (ii) estimation method that closely fits a model to the cross-section of interest rates.

An example of (i) is a level-dependent volatility model, where the factor covariance matrix is driven by the same factors driving the yield curve. Then, the existing link between volatility and cross-section of interest rates is expected to be further strengthened by imposing the NAC.

²Coroneo et al. (2011) examine how arbitrage-free the NS model is. But their focus is on neither factor identification nor volatility prediction, which would potentially be affected by imposing the NAC.

One way to develop such models is to extend existing models. For example, suppose the Gaussian term structure model with constant covariance matrix. Then, an extended version of the Gaussian model is obtained by making the covariance matrix depend on all factors with the positive definiteness maintained. By comparing identified factors between the Gaussian and extended models, therefore, the impact of volatility modeling on factor identification will be uncovered. Furthermore, by estimating these models with and without imposing the NAC, the impact of cross-sectional constraints on volatility estimation and prediction will also be understood. The same extension and analysis are also possible based on the other affine term structure models with stochastic volatility.

As for (ii), the quasi-maximum likelihood method employed by, e.g., Chen and Scott (1993) is appropriate. In the estimation, it is assumed that some interest rates are observed without measurement error. Then, by equating model-implied interest rates (functions of latent factors) to these observed counterparts at each point in time, the latent factors can be obtained by inversion. Since this inversion method forcefully fits the volatility to the cross-section of interest rates when the NAC is imposed, it allows us to highlight the resulting impact of the NAC.

Suppose instead other settings where a model with USV factors that are of little relevance to the cross-section of interest rates is estimated with the assumption that all interest rates are measured with error. Then, it is obvious that the current setting with (i) and (ii) is more prone to being influenced by cross-sectional constraints due to no-arbitrage than others. Nevertheless, the answers to Q1 and Q2 found in this study are “virtually no.”

In comparison with the earlier work, Duffee (2011), and Joslin, Singleton, and Zhu (JSZ) (2011) discuss, using the Gaussian term structure model (a particular model), whether cross-sectional constraints due to no-arbitrage influence the estimation of the physical drift. However, they do not explore whether cross-sectional constraints influence the estimation of the covariance matrix, which is not a trivial issue as the covariance matrix cannot be separated between the physical and risk-neutral probability measures. Joslin (2010) argues, based on the Gaussian model, that the (constant) factor covariance matrix has a minor impact on the cross-section of interest rates. However, he does not explore the impact of the covariance matrix on both cross-section and factor identification when the covariance matrix depends on the same factors driving interest rates.

Duffee (2011) and JSZ (2011) demonstrate that estimation of conditional means of interest rates can be made independently of the NAC.³ This study extends this insight by showing that

³In contrast, Almeida and Vicente (2008), using term structure models that are not exactly the same as the

cross-sectional constraints due to no-arbitrage have a limited impact on both identification of factors and estimation of conditional volatilities of interest rates. Taken together, the imposition of the NAC is not restrictive for modeling and estimating the dynamics of interest rates and their volatilities. Even if it is restrictive for a model, it can be mitigated owing to the introduction of nearly USV factors, which this study shows is possible without much difficulty. After all, standard approaches in time-series analysis are not contradictory to the absence of arbitrage.

It is important to note, however, that this study does not imply that the imposition of the NAC is without merit. It is useful, and in fact essential, for estimating market prices of risk, and more generally, investors' risk preferences and future prospects implicit in interest rate data. What is emphasized here is a virtually innocuous role of no-arbitrage in estimating interest rate dynamics using interest rate data.⁴

Section 2 introduces models based on the Gaussian term structure model. Section 3 explains data and realized volatility measure. Sections 4 and 5 address Q1 and Q2, respectively. Section 6 conducts robustness checks based on the affine models with stochastic volatility. Section 7 concludes. Technical arguments on the computation of no-arbitrage bond prices, when they are unavailable in closed-form, are collected in Appendices.

2 Models

As term structure models appropriate for the research objectives, the Gaussian term structure model is first selected with a number of reasons in Section 2.1 and then extended by making the factor covariance matrix depend on all factors in Section 2.2. Since such extended models do not have closed-form expressions for no-arbitrage bond prices, a method of computing them is explained in Section 2.3.

2.1 The baseline Gaussian model

The Gaussian term structure model is reported to have a number of desirable properties: It can explain observed properties of interest rate data such as failure of the efficient hypothesis of the term structure and hump shape of unconditional volatility of changes in logarithmic yields (Dai and Singleton, 2003); provide accurate prediction of the level of interest rates (Duffee, 2002);

affine models, find that the models' predictive ability to the level of interest rates improves by imposing the NAC. However, whether the same is true for volatility prediction is left unanswered.

⁴This study is silent on whether the same is true for option data, which is left for future research.

and generate interest rate factors whose statistical properties anticipated before estimation are consistent with those realized after estimation (Collin-Dufresne, Goldstein, and Jones (CDGJ), 2009). Additionally, the use of the Gaussian model enables us to obtain implications regarding the impact of the NAC that are useful for the popular dynamics NS model employed by, e.g., Diebold and Li (2006), Diebold et al. (2006, 2008), and Koopman et al. (2010). This is because as shown by Christensen et al. (2009, 2011), the Gaussian model nests the no-arbitrage version of the NS model.

There are numerous specifications of the Gaussian model depending on factor rotations. Among them, this study adopts a conventional one characterized by the mean-reversion of the instantaneous interest rate; see, e.g., Andersen and Lund (1997b), Balduzzi et al. (1996, 1998), and Bikbov and Chernov (2011). The mean-reversion specification, where the roles of the factors are anticipated to some extent, is convenient because model adequacy is easily checked by examining the consistency between the ex-ante and ex-post roles.

Following Litterman and Scheinkman (1991) and the subsequent studies, this study assumes three factors. Let $X_t = (r_t \ \theta_t \ \epsilon_t)'$ be a three-dimensional state vector, and the risk-neutral distribution of instantaneous changes in X_t is given by

$$dX_t \sim N \left[\begin{pmatrix} \kappa_1(\theta_t - r_t) + \epsilon_t \\ \kappa_2(\bar{\theta} - \theta_t) \\ -\kappa_3\epsilon_t \end{pmatrix} dt, \quad \Sigma dt \right], \quad (1)$$

with the (i, j) -th element of Σ denoted as σ_{ij} . In this specification, θ_t is assumed to be the central tendency, toward which the instantaneous interest rate r_t reverts. ϵ_t is assumed to be a shock that makes the mean-reverting behavior a bit irregular. In reality, however, the following constraint has to be placed, as pointed out by CDGJ (2008), for each process to be interpreted as above: $0 < \kappa_2 < \kappa_1 < \kappa_3$. That is, θ_t also mean-reverts to a 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 for the shock to be temporary.

For the purpose of estimation and prediction, the physical distribution of instantaneous changes in X_t also needs to be specified. By assuming the market price of risk that leads to the essentially affine model proposed by Duffee (2002), it can be expressed as

$$dX_t \sim N \left[\left\{ \begin{pmatrix} K_{r,0} \\ K_{\theta,0} \\ K_{\epsilon,0} \end{pmatrix} + \begin{pmatrix} K_{r,1} & K_{r,2} & K_{r,3} \\ K_{\theta,1} & K_{\theta,2} & K_{\theta,3} \\ K_{\epsilon,1} & K_{\epsilon,2} & K_{\epsilon,3} \end{pmatrix} \begin{pmatrix} r_t \\ \theta_t \\ \epsilon_t \end{pmatrix} \right\} dt, \quad \Sigma dt \right], \quad (2)$$

where all parameters in the physical drift are different from those in the risk-neutral drift.

For robustness check, this study also considers an alternative specification of the Gaussian model proposed by CDGJ (2008, equation (29)) and its extended version presented below, and finds that the answers to Q1 and Q2 remain unchanged from those based on the mean-reversion specification (the results are available upon request). However, this study does not consider an alternative specification in which r_t is given by the sum of arbitrary state variables as $r_t = X_{1,t} + X_{2,t} + X_{3,t}$ because this equation unnecessarily imposes an affine structure on the non-affine extended version. Besides, the discussion on factor identification would be easier if the roles of the factors were preliminarily anchored than if they were unknown.

2.2 Extended models

The Gaussian model is extended in a way that the factor covariance matrix depends on all factors with the positive definiteness maintained. The extension to the level dependence is the key to creating an unfavorable setting for conventional approaches without considering the NAC. Namely, if the answers to Q1 and Q2 were no for models with level-dependent volatility that is tightly linked to the cross-section of interest rates, the same would be true for other models with USV factors or GARCH volatilities that are of little relevance to the cross-section of interest rates.

Many previous studies incorporate the level dependence into the dynamics of interest rates; see, e.g., Ait-Sahalia (1996), Andersen and Lund (1997a, b), Bali (2000), Ball and Torous (1999), Brennan and Schwartz (1979), Brenner et al. (1996), Chan et al. (1992), Durham (2003), and Gallant and Tauchen (1998). In what follows, level-dependent specifications are explored that are more complicated than those considered in the literature. However, the purpose is not to propose models. Rather, it is to strengthen the link between volatility and cross-section of interest rates as much as possible within admissible specifications for the factor dynamics.

Based on the Gaussian model, the risk-neutral distribution of instantaneous changes in X_t is given by

$$dX_t \sim N \left[\begin{pmatrix} \kappa_1(\theta_t - r_t) + \epsilon_t \\ \kappa_2(\bar{\theta} - \theta_t) \\ -\kappa_3\epsilon_t \end{pmatrix} dt, \Sigma_t dt \right]. \quad (3)$$

The instantaneous covariance matrix, Σ_t , is now time-dependent. On the other hand, the risk-neutral drift vector is the same as in the original Gaussian model. This same specification allows

us to control for potential influences of the risk-neutral drift vector and hence focus exclusively on potential influences of the covariance matrix on factor identification and volatility prediction.

To satisfy both level dependence and positive definiteness of Σ_t , it is convenient to first decompose Σ_t and then specify its components. The following two approaches to the decomposition are considered. The first approach follows Bollerslev (1990) and Engle (2002) among others, and decomposes Σ_t as

$$\Sigma_t = H_t R_t H_t, \quad (4)$$

where H_t is a diagonal matrix and R_t is a correlation matrix.

The second approach uses the spectral decomposition, which is similar in spirit to Fan et al. (2003), Han (2007), Jarrow et al. (2007), Longstaff et al. (2001), and Pérignon and Villa (2006):

$$\Sigma_t = P L_t P', \quad (5)$$

where L_t is a diagonal matrix consisting of the eigenvalues, which are time-varying, and P is an orthonormal matrix consisting of the corresponding eigenvectors, which are fixed as in the previous studies. In each decomposition, level-dependent specifications are introduced.

Specifications based on the decomposition using the correlation matrix :

The diagonal matrix H_t in equation (4) is given by

$$H_t = \begin{pmatrix} h_1(X_t) & 0 & 0 \\ 0 & h_2(X_t) & 0 \\ 0 & 0 & h_3(X_t) \end{pmatrix}. \quad (6)$$

Then, this study specifies $h_i(X_t)$ simply by a linear function of X_t as ⁵

$$h_i(X_t) = \beta_0^i + \beta^{i'} X_t \quad (i = 1, 2, 3). \quad (7)$$

The correlation matrix R_t is given by

$$R_t = \begin{pmatrix} 1 & \rho_{12}(X_t) & \rho_{13}(X_t) \\ \rho_{12}(X_t) & 1 & \rho_{23}(X_t) \\ \rho_{13}(X_t) & \rho_{23}(X_t) & 1 \end{pmatrix}. \quad (8)$$

⁵This study also considers an alternative specification for $h_i(X_t)$ by following Andersen and Lund (1997b): $h_1(X_t) = \sigma_1 r_t^\gamma \exp\{\frac{1}{2}\epsilon_t\}$; $h_2(X_t) = \sigma_2 \sqrt{\theta_t}$; $h_3(X_t) = \sigma_3$. The answers to Q1 and Q2 remain unchanged from those obtained by the linear specification in equation (7). The results for this alternative specification are available upon request.

The following two specifications for $\rho_{ij}(X_t)$ are considered. The first specification is given by

$$\rho_{ij}(X_t) = \rho_{ij} \quad (i, j = 1, 2, 3, \quad i < j), \quad (9)$$

where $|\rho_{ij}| \leq 1$. The second specification is given by

$$\rho_{ij}(X_t) = \delta_{ij}c(X_t) \quad (i, j = 1, 2, 3, \quad i < j), \quad (10)$$

where $0 \leq \delta_{ij} \leq 1$ and

$$c(X_t) = \frac{2}{1 + \exp\{\gamma_0 + \gamma'X_t\}} - 1. \quad (11)$$

Note that $|c(X_t)| \leq 1$. Further constraints on $\rho_{ij}(X_t)$ in equation (10) are required for R_t to be positive definite, which are specified in estimating the model in Section 4. For convenience, the first model with Quadratic variances and Constant correlations is labeled as QC, and the second model with Quadratic variances and State-dependent correlations as QS.

Specifications based on the spectral decomposition :

The diagonal matrix L_t in equation (5) is given by

$$L_t = \begin{pmatrix} l_1(X_t) & 0 & 0 \\ 0 & l_2(X_t) & 0 \\ 0 & 0 & l_3(X_t) \end{pmatrix}. \quad (12)$$

The orthonormal matrix P can be expressed as

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

where the parameters to be estimated are $\sin \varphi_i^P$ ($i = 1, 2, 3$). For identification, $\varphi_i^P \in [-\pi/2, \pi/2]$ is imposed so that $\cos \varphi_i^P = \sqrt{1 - \sin^2 \varphi_i^P}$. For Σ_t to be positive definite, all eigenvalues in equation (12) must be positive, i.e., $l_i(X_t) > 0$ ($i = 1, 2, 3$) for any X_t . Then, this study specifies $l_i(X_t)$ simply by a quadratic function of X_t as ⁶

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

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

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

⁶This study also considers an alternative specification given by $l_i(X_t) = \exp\{c_0^i + c^{i'} X_t\}$. The answers to Q1 and Q2 remain unchanged from those obtained by the quadratic specification in equation (14). The results for the exponential specification are available upon request.

where

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

and

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

with $\varphi_j^i \in [-\pi/2, \pi/2]$. It is noted that $\sin \varphi_j^i$ cannot be identified for some m_j^i . For example, when $m_j^i = 0$ for all j , $\sin \varphi_j^i$ cannot be identified for all j . In such cases, $\sin \varphi_j^i = 0$ is placed. This model with Quadratic Eigenvalues is labeled as QE.

Since for all extended models Σ_t is designed to be positive definite and there is no sign constraint on the factors, the market price of risk can be specified as flexibly as in the original Gaussian model. It then follows that the physical distribution of instantaneous changes in X_t can also be expressed as (2) with Σ replaced by Σ_t .

2.3 An approximation to no-arbitrage zero-coupon bond prices

To practically examine the impact of the NAC, this study actually estimates the models. For the estimation (including factor identification) to be feasible, closed-form expressions for no-arbitrage bond prices are needed.

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

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

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

$P(X_t, \tau)$, however, has no closed-form under the extended models. It is therefore approximated by relying on a method proposed by Takamizawa and Shoji (2009): The method approximates a vector of conditional moments as the solution to a system of ordinary differential equations. Since the zero-coupon bond price is given as the conditional expectation, the method can be directly applied. The outline of the method is provided in Appendix A, and the accuracy of the approximation is checked in Appendix B. In brief, the accuracy is maintained at least for

maturities ranging up to ten years when reasonable values of parameter and state vectors are provided.

3 Data and realized volatility measure

Weekly (Wednesday) data on U.S. dollar LIBOR and swap rates are used, which cover the period from January 4, 1991 to May 27, 2009 with 961 observations in total. The data are divided into in-sample portion for estimation and out-of-sample portion for prediction. The out-of-sample analysis is conducted to detect the possibility that the impact of the NAC is small in-sample but not out-of-sample. The in-sample period is until April 9, 2003 with 641 observations. This partition is intended to incorporate information on the lowest range of interest rates into model estimation as well as to reserve sufficient out-of-sample observations.

The LIBOR rates with maturities of 6 and 12 months, and swap rates with maturities of 2, 3, 4, 5, 7, and 10 years are used to obtain zero-coupon bond yields on a continuously compounded basis by a bootstrap method with linear interpolation to discount functions. The maturities of the zero yields used for the analysis are 0.5, 1, 2, 3, 5, and 10 years. Following the previous studies, e.g., CDGJ (2008) and Duffee (2002), this study assumes that the yields with maturities of 0.5, 2, and 10 years are measured without error to extract latent factors and that the rest of the yields are measured with error. There may be a concern that this choice is arbitrary. But it is actually convenient because by selecting the yields that are measured without error from each of the short, medium, and long segments of the yield curve, one can measure also without error proxies for the level, slope, and curvature of the yield curve by combining these yields. More importantly from the research objectives, this no-measurement-error assumption is another key to potentially raising the likelihood of answering yes to Q1 and Q2.

For volatility prediction, a realized series is constructed from daily changes in zero yields. A realized measure of the one-week ahead conditional variance of a τ -year yield is computed as

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

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 $m_t = 5$).

The realized measure is generated every Wednesday. The total number of observations is 960, among which the first 640 observations belong to the in-sample.

A realized measure of the h -week ahead conditional variance is computed as

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

The annualized variance is obtained by dividing $RV_{t,t+h\Delta,\tau}$ by $h\Delta$. $RV_{t,t+h\Delta,\tau}$ is also generated weekly, which means that some overlapping daily observations are used in successive observations of $RV_{t,t+h\Delta,\tau}$ with $h > 1$. This study uses $h = 4$, aiming at balancing between ensuring a number of observations for the calculation and avoiding too much overlapping. But other realistic values of h do not qualitatively or quantitatively change the results presented below.

We are fully aware that the realized volatility measure constructed with daily data is crude. But the purpose of this study is not to rank competing models based on predictive accuracy, for which the measurement of the target variable seems to be crucial. Rather, it is to examine within the same model whether the difference in volatility prediction arises between with and without the NAC, regardless of whether the prediction is accurate or not. Therefore, as long as the same conditions are maintained, the use of the crude measure does not seem to be a serious concern.

4 Impact of volatility modeling on factor identification

The purpose of this section is to address Q1: Does the modeling of volatility practically influence the identification of latent factors from the cross-section of interest rates when the NAC is imposed? To achieve the purpose, latent factors extracted from the cross-section of interest rates are compared between the Gaussian model with constant covariance matrix and the extended models with level-dependent covariance matrix. Section 4.1 explains estimation method, and Sections 4.2 and 4.3 report the results regarding factor identification and parameter estimation, respectively.

4.1 Estimation method

The models with the NAC are estimated by the quasi-maximum likelihood method, which is one of the standard methods employed by, e.g., Chen and Scott (1993), Duffee (2002), and Pearson and Sun (1994). First of all, parameter vectors of the term structure models are defined in the

following, which will be useful in comparing parameter estimates between with and without the NAC.

- Θ_{μ^Q} : a parameter vector for the risk-neutral drift vector;
- Θ_{μ^P} : a parameter vector for the physical drift vector;
- Θ_{Σ} : a parameter vector for the covariance matrix.

Then, define

$$\Theta = \begin{pmatrix} \Theta_{\mu^Q} \\ \Theta_{\mu^P} \\ \Theta_{\Sigma} \end{pmatrix}, \quad \Theta_Q = \begin{pmatrix} \Theta_{\mu^Q} \\ \Theta_{\Sigma} \end{pmatrix}, \quad \Theta_P = \begin{pmatrix} \Theta_{\mu^P} \\ \Theta_{\Sigma} \end{pmatrix}. \quad (21)$$

Next, let $Y_t^P = (y_{t,0.5} \ y_{t,2} \ y_{t,10})'$ be a vector of yields measured without error, and $\Upsilon(X_t; \Theta_Q)$ be a vector of model-implied yields with the corresponding maturities:

$$\Upsilon(X_t; \Theta_Q) = (Y(X_t, 0.5; \Theta_Q) \ Y(X_t, 2; \Theta_Q) \ Y(X_t, 10; \Theta_Q))'. \quad (22)$$

Then, X_t is extracted by solving $Y_t^P = \Upsilon(X_t; \Theta_Q)$ for X_t . This extraction is performed numerically for the extended models, 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 Gaussian model.

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

$$U_t \sim i.i.d.N(\mathbf{0}, \varsigma^2 \mathbf{I}). \quad (23)$$

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

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

$$\begin{aligned} f(Y_t^P, Y_t^e | Y_{t-\Delta}^P; \Theta, \varsigma^2) &= f(X_t, U_t | X_{t-\Delta}; \Theta, \varsigma^2) \left| \frac{d\Upsilon(X_t; \Theta_Q)}{dX_t'} \right|^{-1} \\ &= f_T(X_t | X_{t-\Delta}; \Theta) f_C(U_t | X_t; \Theta_Q, \varsigma^2) \left| \frac{d\Upsilon(X_t; \Theta_Q)}{dX_t'} \right|^{-1}. \end{aligned} \quad (24)$$

The first equality is from 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 decomposition of the joint density into the marginal (f_T) and conditional (f_C) components with X_t Markovian.

For the extended models, the transition density, f_T , has no analytical expression for finite Δ . It is then approximated 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 method used for pricing bonds. It is noted that for all models except the QS model, these moments can be computed exactly since the drift vector is linear in X_t and the instantaneous covariance matrix is at most quadratic in X_t . Still, the formula given in equation (49) (without the residual term) is useful for this computation. The Jacobian term is computed in the process of extracting X_t . On the other hand, f_C is the multivariate normal density function from (23). The objective function for estimating the model parameters is then

$$\sum_t \left\{ \ln f_T(X_t|X_{t-\Delta}; \Theta) + \ln f_C(U_t|X_t; \Theta_Q, \varsigma^2) - \ln \left| \frac{dY(X_t; \Theta_Q)}{dX'_t} \right| \right\}. \quad (25)$$

The parameters are estimated using the in-sample data. These estimates are held fixed throughout the out-of-sample period. This means that the models are not re-estimated each time the out-of-sample prediction is made. This approach does not seem to be a serious concern because, as will be discussed in Section 4.2, the model parameters appear to be stable between the two sample periods.

4.2 Identified factors

Before reporting estimated parameters in the next subsection, identified factors are first reported to address Q1. Figure 1 exhibits the time series of r_t extracted through the Gaussian, QC, QS, and QE models over the whole period. It is seen that these plots are hardly visually distinguishable. In fact, sample correlations over the whole period across the four models are all very close to one, regardless of the level or first difference of the data. In Panel (a), the time-series of the six-month yield is also displayed. As anticipated, it tracks the time-series of r_t with sample correlation over the whole period 0.98.

Analogous time-series for θ_t and ϵ_t are displayed in Figures 2 and 3, respectively, showing that they are also hardly distinguishable among the four models with sample correlations all close to one. The results for θ_t and ϵ_t may be a bit more surprising than that for r_t . This is because while r_t is unambiguously the instantaneous interest rate and can be determined as the starting point of the yield curve independently of models considered, this is not the case for θ_t and ϵ_t , which realize model-dependently.

In Panel (a) of Figure 2, the time-series of the ten-year yield is also displayed. As also anticipated, it tracks the time-series of θ_t with sample correlation over the whole period 0.96. Similarly, Panel (a) of Figure 3 displays the time-series of an observed variable that turns out to track well the time-series of ϵ_t . It is actually a curvature factor known as the third factor driving

the yield curve, which is simply measured here by $cur_t = 2y_{t,2} - (y_{t,0.5} + y_{t,10})$. It is noted that these similarities between the state and observed variables continue in the out-of-sample period, suggesting that there is no structural break or regime shift in the model parameters despite experiencing severe market turmoil in the out-of-sample period.

These figures imply that the covariance matrix does not have a significant impact on the identification of latent factors. It then follows that it is the risk-neutral drift vector that dominantly determines factor identification. The identified factors are similar across the four models because the specification of the risk-neutral drift vector is identical with similar parameter values, as will be discussed in Section 4.3.

An important implication can then be obtained from this finding: A factor is difficult to be identified from the cross-section of interest rates if it does not drive the risk-neutral drift of interest rate factors. Put simply, if there is a factor that appears in the covariance matrix but not in the risk-neutral drift vector, it will virtually act as an unspanned factor. Indeed, a similar argument is made by Joslin (2010) using the Gaussian term structure model. Besides, 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 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 to the corresponding estimate of -809.5 for the $A_1(4)$ model obtained by the inversion method. Other parameters in the risk-neutral drift do not exhibit such a dramatic change. The result suggests that by this constraint 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 into no-arbitrage term structure models is not as difficult as is thought: Simply exclude a factor from the risk-neutral drift vector.

4.3 Estimated parameters

The estimation results are detailed, which will be helpful in discussing the impact of the NAC on volatility estimation and prediction in the next section. Since estimated parameters of particular interest are those in the covariance matrix, those in the physical drift vector are omitted from each of the following tables but collected altogether in Table 5. It is noted here that to control for potential influences of the physical drift, the drift parameters to be estimated are exactly the same across all models.

4.3.1 Results for the Gaussian model

In columns labeled “with NAC” of Table 1, the estimates for the baseline Gaussian model are presented with standard errors in parenthesis: The standard errors are computed by the outer product of the gradient of the log-likelihood function. When the full parameters are estimated, some parameters including those in the physical drift are insignificant. Such parameters are set to zero, and the remaining parameters are re-estimated. The parameters of the other models are treated in a similar way unless otherwise noted. Since this study also performs out-of-sample analysis, simple models are not necessarily evaluated unfavorably. By this simplification, $\sigma_{23} = 0$ is placed, indicating that the instantaneous covariance between θ_t and ϵ_t is zero.

The speed of mean reversion of θ_t under the risk-neutral measure is by far the lowest, as implied by the estimate of κ_2 , 0.017. The estimates of κ_1 and κ_3 are similar around 0.95, implying that an additional (fourth) factor is unnecessary for describing the cross-section of interest rates. The high goodness-of-fit to the cross-section is also evidenced by a small standard deviation of the measurement errors: The estimate of $\varsigma \times 10^4$ is 6.15 basis points (bps).

4.3.2 Results for the QC model

Columns labeled “with NAC” of Table 2 show that the estimates in the risk-neutral drift for the QC model are not very different from those for the Gaussian model. Additionally, the estimate of $\varsigma \times 10^4$ (6.13 bps) remains almost the same. These similarities between the QC and Gaussian models indicate that the level-dependent covariance matrix of this form does not much affect the parameters in the risk-neutral drift nor improve the goodness-of-fit to the cross-section of interest rates. In fact, looking at the values of the log-likelihood and its components, the cross-sectional component, which is denoted as LogL_C and corresponds to the second term of (25), is increased only by 6 to 11478 from that for the Gaussian model. In contrast, the time-series component, which is denoted as LogL_T and corresponds to the first term of (25), is indeed increased by 67 to 9417. Consequently, the total log-likelihood value denoted as LogL is increased by 93 with additional 9 parameters. It is the descriptive power for time-series properties of the data that is benefited from the level-dependent covariance matrix.

Furthermore, some interesting results emerge from the estimates in $h_i(X_t)$ ($i = 1, 2, 3$). First, the estimates of β_1^2 and β_2^2 in $h_2(X_t)$ are both insignificant, suggesting that neither the short-term nor long-term yield drives the instantaneous variance of θ_t .⁷ Since θ_t is highly correlated

⁷To highlight this result and differences in estimates between with and without the NAC, the insignificant

with the ten-year yield as seen in Figure 2, $h_2(X_t)$ also governs the variability in long-term yields. Then, the result is consistent with those by the previous studies pointing out that the volatility of long-term yields is especially difficult to be captured by the level of interest rates; see, e.g., Andersen and Benzoni (2010), CDGJ (2009), and Jacobs and Karoui (2009). In contrast, the estimated coefficients on r_t and θ_t in $h_1(X_t)$ and $h_3(X_t)$ are all significant. Specifically, the estimates of β_1^1 and β_2^1 in $h_1(X_t)$ are both positive, implying that a positive shift in the yield curve increases the volatility of r_t and thus of short-term yields. In $h_3(X_t)$, while the estimate of β_2^3 is positive, that of β_1^3 is negative. The result suggests that when the slope of the yield curve steepens, the volatility of ϵ_t and thus of the curvature factor increases.

Second, the estimated coefficients on ϵ_t , β_3^i ($i = 1, 2, 3$), are all significant and negative. These negative coefficients seem to be inconsistent with the theory documented in, e.g., Christiansen and Lund (2005), and Ilmanen (1999). Specifically, as seen in Figure 3, ϵ_t moves closely to the curvature factor of the yield curve measured by $2y_{t,2} - (y_{t,0.5} + y_{t,10})$, which in turn can be interpreted as a butterfly spread. Since this butterfly spread has a negative convexity, the spread position loses money when a large shift in the yield curve occurs regardless of the direction of the shift. To compensate the loss induced by high volatility, the butterfly spread or the curvature factor should move parallel with the volatility. That is, the theory suggests that the sign of β_3^i is positive. The result that the estimates of β_3^i are all negative implies that at any segment of the yield curve the volatility increases with decrease in the curvature factor. Further examination of the empirical relationship between shape and volatility of the yield curve seems interesting and is left for future research.

4.3.3 Results for the QS model

Before estimation, the following constraints are placed based on the previous results to keep the model simple: $\beta_1^2 = \beta_2^2 = \beta_0^3 = 0$ and $\rho_{23}(X_t) = 0$ or equivalently $\delta_{23} = 0$. The latter constraint (zero instantaneous correlation between θ_t and ϵ_t) simplifies a further constraint required for R_t to be positive definite. Namely, it is $\rho_{12}^2(X_t) + \rho_{13}^2(X_t) < 1$ for any X_t , or by substituting equation (10) into this inequality $c^2(X_t)(\delta_{12}^2 + \delta_{13}^2) < 1$. Since $c^2(X_t) \leq 1$, it follows that $\delta_{12}^2 + \delta_{13}^2 < 1$. To further simplify the constraint, however, $\delta_{13} = \sqrt{1 - \delta_{12}^2}$ is actually placed: With this constraint, R_t becomes semi-positive definite in a special case of $c^2(X_t) = 1$.

parameters are left unconstrained. But the answers to Q1 and Q2 do not change if $\beta_1^2 = \beta_2^2 = 0$ is placed consistently with the other insignificant parameters. For the same reason, the insignificant parameters in the correlation matrix R_t for the QS model, presented in Table 3, are left unconstrained.

In columns labeled “with NAC” of Table 3, the estimates for the QS model are presented. First of all, it is found that none of the estimates associated with time-varying correlations, γ_i ($i = 1, 2, 3$), are significant. Consequently, the goodness-of-fit to neither cross-sectional nor time-series properties of the data is improved from the QC model: Actually, due to the constraint of $\beta_1^2 = \beta_2^2 = \beta_3^2 = 0$, the values of LogL and LogL_T are slightly decreased. Therefore, the overall picture of the estimation results remains unchanged from that for the QC model. Here, it is additionally implied that time-varying correlations are more difficult to be captured by the level of interest rates than are time-varying volatilities, though this difficulty may be partly due to the limited specification for the positive definite correlation matrix.⁸

4.3.4 Results for the QE model

There are sign constraints on the parameters of the QE model to make the covariance matrix positive definite. Some parameters reach the boundary values or violate the sign constraints. Such parameters are fixed in the following and then the remaining parameters are re-estimated: $c_0^i = 10^{-8}$ (as $c_0^i > 0$); $m_j^i = 0$ (as $m_j^i \geq 0$).

In columns labeled “with NAC” of Table 4, the estimates for the QE model are presented. Again, the overall picture remains unchanged from the previous models. In particular, the estimates in the risk-neutral drift are very similar to those for the Gaussian model. Also, it is reconfirmed that the goodness-of-fit to cross-sectional properties of the data is not much improved, as evidenced by the estimate of $\varsigma \times 10^4$ (6.13 bps) and the value of LogL_C that is increased only by 9 from that for the Gaussian model. In contrast, the goodness-of-fit to time-series properties of the data exhibits a significant improvement with the value of LogL_T increased to 9472. Consequently, the value of LogL is increased by 151 to 22416 from that for the Gaussian model with additional 11 parameters.

Among the parameters in the eigenvector matrix P , only $\sin \varphi_2^P$ is estimated significantly,

⁸To capture time-varying correlations as well as volatilities, it seems promising to use a Wishert process; see, e.g., Buraschi et al. (2008), and Gourieroux and Sufana (2011). Cieslak and Povala (2011) apply a term structure model where the covariance matrix of interest rate factors is driven by a 2-by-2 Wishert process to the joint data on interest rates and realized (co)variance measures, and show that the proposed model has a surprisingly good performance to the latter data without reducing the goodness-of-fit to the former data.

leading to Σ_t of the following form:

$$\Sigma_t = \begin{pmatrix} l_1(X_t) \cos^2 \varphi_2^P + l_3(X_t) \sin^2 \varphi_2^P & 0 & \{l_1(X_t) - l_3(X_t)\} \sin \varphi_2^P \cos \varphi_2^P \\ 0 & l_2(X_t) & 0 \\ \{l_1(X_t) - l_3(X_t)\} \sin \varphi_2^P \cos \varphi_2^P & 0 & l_1(X_t) \sin^2 \varphi_2^P + l_3(X_t) \cos^2 \varphi_2^P \end{pmatrix}. \quad (26)$$

It turns out that $l_2(X_t)$ is the instantaneous variance of θ_t . In addition, since the estimate indicates $\sin \varphi_2^P < \cos \varphi_2^P$, $l_1(X_t)$ is more closely related to the instantaneous variance of r_t , whereas $l_3(X_t)$ to that of ϵ_t . Among the instantaneous covariances, only that between r_t and ϵ_t is significant. These interpretations will be helpful for understanding the difference in volatility prediction between with and without the NAC in the next section.

5 Impact of cross-sectional constraints on volatility estimation and prediction

The purpose of this section is to address Q2: Does the cross-section of interest rates practically influence the estimation and prediction of interest rate volatility when the NAC is imposed? To achieve the purpose, the Gaussian and extended models are used also here, and estimated with and without imposing the NAC while the other conditions are held constant. Section 5.1 explains a setting for the analysis, and Sections 5.2 and 5.3 report the results regarding parameter estimation and volatility prediction, respectively.

5.1 Empirical setting

5.1.1 Estimation

The analysis bases the no-arbitrage models that have already been estimated in Section 4, and proceeds from a point of view of removing, rather than imposing, the NAC. This seemingly reverse approach makes it easier to control for comparative conditions. The key to this control is to use the same state vector X_t between with and without the NAC. A simple way to guarantee the same X_t is to treat X_t as *observed* by overlooking the fact that X_t is extracted through the no-arbitrage models. The reasons for taking this reverse approach are further addressed in Section 5.1.3.

It is useful to clarify from the context of this study that the NAC is involved in the following two properties:

- (a) time-series behavior of interest rate factors and their volatilities;

(b) cross-sectional relation between the factors and observed interest rates.

Then, when the NAC is removed, properties (a) and (b) need to be determined in other ways. It follows that the estimation procedure without the NAC consists of two parts for determining these properties separately.

To determine property (a), Θ_P defined in equation (21) is estimated. Specifically, given the in-sample data on X_t , it is estimated by maximizing

$$\sum_t \ln f_T(X_t|X_{t-\Delta}; \Theta_P). \quad (27)$$

Then, the estimate of Θ_P obtained by maximizing (27) is compared with that obtained by maximizing (25) to examine the impact of the NAC on parameter estimation. From now on, the estimates of Θ_P are distinguished between with and without the NAC by denoting Θ_P^{TC} and Θ_P^T , respectively: “TC” indicates that the parameter vector is estimated by maximizing (25) and hence using both time-series and cross-sectional dimensions of the data, whereas “T” by maximizing (27) and hence using only time-series dimension of the data.

To determine property (b), the simplest relation is considered:

$$y_{t,\tau} = \alpha_{0,\tau} + \alpha'_{1,\tau} X_t. \quad (28)$$

The linear function is intended to raise the likelihood of answering yes to Q2 as it may overstate the impact of the NAC for the extended models that are not exactly the affine models. However, even if a significant impact is detected, there is still room for considering nonlinear functions to mitigate it. The parameters in equation (28) are estimated for each τ by OLS using the in-sample data. For estimation, measurement errors are introduced in all yields.

5.1.2 Prediction

Model-implied h -week ahead conditional variances of a τ -year yield are given by

$$\text{with NAC: } \text{var}_t[y_{t+h\Delta,\tau}] = \text{var}_t[Y(X_{t+h\Delta}, \tau; \Theta_Q); \Theta_P^{TC}], \quad (29)$$

$$\text{without NAC: } \text{var}_t[y_{t+h\Delta,\tau}] = \alpha'_{1,\tau} \text{var}_t[X_{t+h\Delta}; \Theta_P^T] \alpha_{1,\tau}, \quad (30)$$

where $\text{var}_t[\cdot]$ stands for the conditional variance under the physical measure.

The following notes are in order regarding the computation of equations (29) and (30). First, while measurement errors are introduced in some or all yields for estimation, their variances are omitted for prediction, as the purpose is to compare predicted values generated by the models.

Second, since the conditional variance in (29) has no closed-form for the extended models, it is computed with the same method used for pricing bonds. The application and accuracy of the method are provided in Appendices A and B, respectively. It is noted that the conditional variance in (30) can be computed exactly for all models except the QS model by the same reason as noted in Section 4.1. Third, although the constant term in equation (28), $\alpha_{0,\tau}$, does not appear in (30), it may affect volatility prediction through the estimate of $\alpha_{1,\tau}$. But this effect is negligible. The results of volatility prediction with $\alpha_{0,\tau}$ excluded preliminarily from (28) remain unchanged from those presented below with $\alpha_{0,\tau}$ included.

The magnitude of the difference in forecasts between with and without the NAC is evaluated from both statistical and economic perspectives. From a statistical perspective a mean squared error (MSE) criterion is applied to variance forecasts, whereas a root mean squared error (RMSE) criterion is applied to annualized standard-deviation forecasts from an economic perspective.

Prediction errors for variance are computed according to the degree of involvement of the NAC in properties (a) and (b):

$$e_{1,t,h,\tau} = RV_{t,t+h\Delta,\tau} - \text{var}_t[Y(X_{t+h\Delta}, \tau; \Theta_Q); \Theta_P^{TC}], \quad (31)$$

$$e_{2,t,h,\tau} = RV_{t,t+h\Delta,\tau} - \alpha'_{1,\tau} \text{var}_t[X_{t+h\Delta}; \Theta_P^{TC}] \alpha_{1,\tau}, \quad (32)$$

$$e_{3,t,h,\tau} = RV_{t,t+h\Delta,\tau} - \alpha'_{1,\tau} \text{var}_t[X_{t+h\Delta}; \Theta_P^T] \alpha_{1,\tau}. \quad (33)$$

e_1 and e_3 are prediction errors with and without the NAC, respectively. Our primary interest is in the difference between these two errors. On the other hand, e_2 is prediction error with partial involvement of the NAC, where Θ_P^{TC} is used as the parameter vector for the physical distribution while the simple linear function given by equation (28) is used as the cross-sectional relation.

Using the error in the intermediate step, the MSE ratio of primary interest can be decomposed as

$$\frac{\text{MSE}_3}{\text{MSE}_1} = \frac{\text{MSE}_2}{\text{MSE}_1} \times \frac{\text{MSE}_3}{\text{MSE}_2}, \quad (34)$$

where MSE_i is calculated from e_i . If the ratio on the left-hand side of equation (34), the 3-1 ratio for short, significantly deviates from one, this deviation can be regarded as evidence of the resulting impact of the NAC. Further, if the 3-1 ratio is above (below) one, the imposition of the NAC is (un)favorable for volatility prediction. The first ratio on the right-hand side of (34), the 2-1 ratio, asks if the prediction differs by changing the cross-sectional relation between with and without the NAC while the value of the physical parameter vector is held fixed at

Θ_P^{TC} . The second ratio on the right-hand side, the 3-2 ratio, asks if the prediction differs by changing the value of the physical parameter vector between with and without the NAC while the cross-sectional relation is held fixed by the linear function. This decomposition uncovers which has a dominant effect, cross-section function or time-series estimation.

5.1.3 Discussion

As explained above, the analysis is conducted following the reverse approach, starting from imposing the NAC and ending in removing it. A more natural approach, on the other hand, bases the results without the NAC and then compares them with those obtained by imposing the NAC. The reasons for not taking this natural approach are detailed as follows.

The natural approach starts from providing an observed vector, say X_t^o , consisting of interest rates or their combinations. The physical distribution of instantaneous changes in X_t^o is assumed to follow the Gaussian model or its extended version, and estimated using time-series data on X_t^o . The question then is whether the estimated physical distribution significantly changes if the NAC is imposed. The estimation with the NAC is equivalent to estimating no-arbitrage term structure models. To model the no-arbitrage term structure, the risk-neutral distribution of instantaneous changes in X_t (an unobserved state vector) is required. This distribution can be derived, after providing an appropriate model for market price of risk, by applying the Ito formula to $X_t = g^{-1}(X_t^o)$, where g is a function derived as the solution to the pricing equation under the NAC. Note that in this approach, the same X_t^o is used between with and without the NAC, allowing for examining whether the estimated physical distribution of dX_t^o significantly changes by imposing the NAC. The analogous logic in the reverse approach is that using the same X_t between with and without the NAC allows for examining whether the estimated physical distribution of dX_t significantly changes by removing the NAC.

The above natural approach is applicable to the Gaussian model, which indeed is demonstrated by JSZ (2011). But the application to the extended models is almost infeasible. This is because the implied risk-neutral distribution of dX_t becomes very complicated due to the level dependence of the covariance matrix and a potentially nonlinear relationship between X_t and X_t^o . This complexity makes it extremely difficult to derive analytical expressions of no-arbitrage bond prices even approximately. We believe that it is more constructive to try to understand the impact of the NAC by adopting the feasible reverse approach.

5.2 Comparison of parameter estimation

Parameter estimates in the covariance matrix are compared between with and without the NAC in Tables 1–4. Columns labeled “without NAC” present the estimates obtained by maximizing (27). The estimates in the physical drift vector are collected altogether in Table 5, and the estimates in the cross-section function (28) are not reported for saving space.

Table 1 shows that in the Gaussian model there is no parameter that exhibits a significant change by removing the NAC. Besides, the value of LogL without the NAC is almost the same as that of LogL_T with the NAC. These results imply that cross-sectional constraints due to no-arbitrage have little impact on parameters associated with the volatility.

Tables 2 and 3 also show that a significant impact of the NAC is detected for neither the QC nor QS model. Especially, the estimates in $h_1(X_t)$ and $h_3(X_t)$ are stable to the presence or absence of the NAC. Additionally, the coefficients on r_t and θ_t in $h_2(X_t)$ remain insignificant, confirming that the volatility of long-term yields is difficult to be explained by the level of interest rates. Still, the point estimates in $h_2(X_t)$ are increased in absolute value, though not statistically significantly.

Table 4 presents the results for the QE model, which are generally similar to those for the previous models but contain a notable difference between with and without the NAC. Specifically, the estimate of m_3^2 in $l_2(X_t)$, which is originally 0.058 but not statistically different from zero, is increased to 0.140 and turns significant without the NAC. The result suggests that fitting a model to both time-series and cross-sectional properties of the data tends to reduce the variability in θ_t and thus in long-term yields. But in spite of this change, the overall fit to time-series properties of the data is not changed much. Even by removing the NAC, the value of LogL is increased only by 5 from that of LogL_T with the NAC. In other words, when the NAC is imposed, the loss in the log-likelihood value of the time-series dimension is at most 0.05%. This is an intuitive number representing the cost of no-arbitrage constraints.

5.3 Comparison of volatility prediction

5.3.1 Statistical perspective

Table 6 presents the ratios of the MSEs of four-week-ahead ($h = 4$) variance forecasts generated by the extended models: Those for the Gaussian model are omitted as they are almost identical to one for all cases considered. The last column labeled “Ave.” displays the ratio between average MSEs with each average taken over the maturity spectrum. * and ** indicate that the

Diebold and Mariano (DM) (1995) test rejects the null hypothesis of equal predictive accuracy between e_i and e_j ($i, j = 1, 2, 3$) given in equations (31)–(33) at the 5% and 1% significance levels, respectively. The standard error of the test statistic is computed by the Newey and West (1987) method with 8 lags. As noted by Diebold (2015), the DM test is originally designed to compare forecasts, but not models, where forecast errors are treated as given or primitive variables. Since this study compares forecasts within the same model, the application of the DM test is suitable.

In Panel A of Table 6, the results from the in-sample are presented. The values of the 3-1 ratio at the average are close to one for all models. The same is true for the values of the component ratios at the average, although the value of the 2-1 ratio for the QE model (0.992) indicates a statistically significant deviation from one. Also at $\tau = 0.5, 2$, the 2-1 ratio for the QE model deviates from one statistically significantly though the magnitude of deviation is small. The result indicates that the difference in the cross-sectional relation can produce a small but distinct difference in volatility prediction. Furthermore, the 3-1 ratio for the QE model at $\tau = 1, 2$ is statistically below one, which is attributed to both the 2-1 and 3-2 ratios with the latter having a slightly stronger impact. In contrast, for the QC and QS models, there is no case at respective maturities as well as at the average in which the difference is statistically significant.

In Panel B of Table 6, the results from the out-of-sample are presented. In contrast to the in-sample results, it is apparent that the values of the 3-1 ratio tend to be below one and that the number of cases in which the difference is statistically significant is increased. Specifically, for the QC model, the value of the 3-1 ratio at the average is 0.949 and the values of the 3-1 ratio at respective maturities range from 0.901 ($\tau = 5$) to 0.997 ($\tau = 0.5$). They are all below one, indicating the possibility that cross-sectional constraints are binding out-of-sample though they are not in-sample. A similar pattern is observed for the QS model. For both models, the 3-2 ratio contributes more than the 2-1 ratio, implying that the difference in time-series estimation has a stronger impact on volatility prediction than the difference in cross-section function.

For the QE model, on the other hand, the 3-1 ratio at neither the average nor respective maturities is statistically different from one. This holds true even at $\tau = 10$, where the 3-1 ratio reaches 0.824. Behind the result of not detecting a statistical difference is, aside from the smaller sample size, large variation in forecast errors generated by the QE model, $e_{i,t,4,10}$ ($i = 1, 2, 3$): These errors are further discussed in Section 5.3.3. Behind the result of a seemingly large

deviation (0.824) is the difference in the estimates of m_3^2 presented in Table 4. As discussed in Section 5.2, m_3^2 determines the magnitude of the volatility of θ_t and thus of long-term yields. By removing the NAC, the point estimate of m_3^2 becomes more than double, from 0.06 to 0.14. This change is actually desirable for capturing large variation in the ten-year yield volatility observed in the out-of-sample period, as will be seen in Figure 5. This explanation is also supported by the 3-2 ratio measuring the time-series-related difference, which is further below one, 0.791.

To summarize from a statistical point of view, significant differences in forecasts between with and without the NAC can arise. The next step then is to evaluate the magnitude of these differences from an economic point of view. This evaluation is also informative for the case in which the difference is not statistically significant but appears large. When the impact of the NAC is decomposed into cross-section-related and time-series-related components, the latter measured by the 3-2 ratio is generally stronger. But the cross-section-related difference measured by the 2-1 ratio, though it is small in magnitude, can be statistically significant. Still, it could be made insignificant by considering more flexible functions than the linear one given by equation (28).

5.3.2 Economic perspective

To see the differences in forecasts from an economic point of view, the four-week-ahead ($h = 4$) annualized standard deviation is predicted, instead of the variance, and the predictive accuracy is evaluated by the RMSE criterion. Specifically, the RMSEs are computed from the following prediction errors:

$$\tilde{e}_{1,t,h,\tau} = \sqrt{\frac{RV_{t,t+h\Delta,\tau}}{h\Delta}} - \sqrt{\frac{\text{var}_t[Y(X_{t+h\Delta}, \tau; \Theta_Q); \Theta_P^{TC}]}{h\Delta}}, \quad (35)$$

$$\tilde{e}_{3,t,h,\tau} = \sqrt{\frac{RV_{t,t+h\Delta,\tau}}{h\Delta}} - \sqrt{\frac{\alpha'_{1,\tau} \text{var}_t[X_{t+h\Delta}; \Theta_P^T] \alpha_{1,\tau}}{h\Delta}}. \quad (36)$$

Table 7 presents the RMSEs expressed in bps, where RMSE_i is calculated from \tilde{e}_i . From Panel A displaying the in-sample results, it is found that the difference is at most -1.1 bps (QE, $\tau = 10$). This negative difference, however, does not indicate that the imposition of the NAC is restrictive. Panel B shows that the differences are somewhat larger and generally positive in the out-of-sample period. Still, the differences at the average are well within 1 bp for all models. Besides, the largest difference for each model does not seem to be economically significant: 1.7 (QC, $\tau = 5$), 1.6 (QS, $\tau = 5$), and 2.0 (QE, $\tau = 10$) bps. As presented in Panel B of Table 6, the 3-1 ratio of the variance forecast MSEs falls to as low as 0.824 (QE, $\tau = 10$). This seemingly

large deviation from one is in fact equivalent to a difference of 2 bps in terms of the annualized standard deviation.

5.3.3 Further investigation from time-series perspective

The difference in RMSE of at most 2 bps does not seem large. However, it may still be concerned that the differences become occasionally large although they are small on average. To see the differences in forecasts at each point in time, Figures 4 and 5 display time-series of four-week-ahead ($h = 4$) forecasts of annualized standard deviation of a τ -year yield ($\tau = 0.5, 2, 10$) generated by the QC and QE models, respectively: The bold and dotted lines correspond to the forecasts with and without the NAC, respectively, and the thin line to the realized value. The analogous figures for the QS model are omitted as they are similar to those for the QC model and hence for saving space. For both models, there is no visual difference in forecast series between with and without the NAC at $\tau = 0.5, 2$. At $\tau = 10$, however, some differences can be observed. Specifically, the forecasts without the NAC are higher, especially in high volatility periods, which is attributed to the differences in point estimates in $h_2(X_t)$ for the QC model and in $l_2(X_t)$ for the QE model, as discussed in Section 5.2.

If this magnitude of difference in forecasts observed at some points in time is unacceptable, it may be recommended to introduce (nearly) USV factors. In this particular case, they are to be introduced into $h_2(X_t)$ and $l_2(X_t)$. Then, even though some parameters in these functions were adjusted to cross-sectional properties of the data by imposing the NAC, the goodness-of-fit to time-series properties of the data would not be reduced owing to the nearly USV factors. And this study has argued that their introduction is achieved without much difficulty. That is, the time-series-related difference can also be reduced, as is the case for the cross-section-related difference.

Taken together, even if a large impact of the NAC on volatility prediction is detected, there are ways to mitigate it by considering flexible functions (nonlinear regression equations) for the cross-sectional relation or by introducing nearly USV factors for the time-series estimation, or by both. Normally, models in time-series analysis are sufficiently flexible and embed nearly unspanned factors such as GARCH volatilities. Therefore, even if the NAC were imposed on such models, its impact would not be large.

6 Robustness checks using affine models with stochastic volatility

We have seen using the Gaussian and extended models that the imposition of the NAC does not much affect the identification of latent factors or the estimation and prediction of interest rate volatility. Also, we have discussed that even if there is a significant impact, it can be mitigated without much difficulty. This section examines the robustness of these findings by changing the baseline models to affine models with stochastic volatility. Sections 6.1 and 6.2 present specifications for an affine stochastic volatility model and its extended version, respectively. Sections 6.3 through 6.5 present the results regarding factor identification, parameter estimation, and volatility prediction, respectively. Section 6.6 discusses and summarizes the results of this section.

6.1 The baseline affine model with stochastic volatility

Following the notation developed by Dai and Singleton (2000), let $A_m(3)$ ($0 \leq m \leq 3$) denote an affine term structure model that has 3 factors in total with m factors driving the factor covariance matrix: $m = 0$ corresponds to the Gaussian model. Below, the results with $m = 1$ are presented, however, the results with $m = 2$ (available upon request) do not qualitatively or quantitatively differ from those with $m = 1$. On the other hand, $m = 3$ is not considered as the dynamics of r_t are directly specified here. More specifically, the remaining two factors cannot drive the volatility of changes in r_t , and hence they are difficult to be interpreted as stochastic volatility factors of the yield curve.

Among numerous specifications for the $A_1(3)$ model, this study adopts the one proposed by CDGJ (2008, equations (33) and (34)). Let $X_t = (r_t \ \mu_t \ V_t)'$ be a three-dimensional state vector, and the risk-neutral distribution of instantaneous changes in X_t is given by

$$dX_t \sim N \left[\begin{pmatrix} \mu_t \\ \kappa_{\mu,0} + \kappa_{\mu,1}r_t + \kappa_{\mu,2}\mu_t + V_t \\ \kappa_{V,0} + \kappa_{V,3}V_t \end{pmatrix} dt, \ \Sigma_t dt \right], \quad (37)$$

where

$$\Sigma_t = \begin{pmatrix} s_{11} & s_{12} & s_{13} \\ s_{12} & s_{22} & s_{23} \\ s_{13} & s_{23} & s_{33} \end{pmatrix} V_t + \begin{pmatrix} c_{11} & c_{12} & 0 \\ c_{12} & c_{22} & 0 \\ 0 & 0 & 0 \end{pmatrix}. \quad (38)$$

In this specification, r_t is also the instantaneous interest rate, whose risk-neutral drift is μ_t . V_t is a factor driving both conditional first and second moments of X_t . It is noted that V_t is normalized in the risk-neutral drift of μ_t , where the coefficient on V_t is set to one. This normalization is useful for interpreting extracted V_t , as will be seen in the following subsections.

The market price of risk is specified with the assumption that V_t does not reach zero in finite time under both the risk-neutral and physical probability measures; see Cheridito et al. (2007). Then, the constant term in the drift of V_t can be separated between these measures. The resulting physical distribution of instantaneous changes in X_t is given by

$$dX_t \sim N \left[\left\{ \begin{pmatrix} K_{r,0} \\ K_{\mu,0} \\ K_{V,0} \end{pmatrix} + \begin{pmatrix} K_{r,1} & K_{r,2} & K_{r,3} \\ K_{\mu,1} & K_{\mu,2} & K_{\mu,3} \\ 0 & 0 & K_{V,3} \end{pmatrix} \begin{pmatrix} r_t \\ \mu_t \\ V_t \end{pmatrix} \right\} dt, \quad \Sigma_t dt \right]. \quad (39)$$

V_t does not reach zero in finite time under both measures if $2\kappa_{V,0} \geq s_{33}^2$ and $2K_{V,0} \geq s_{33}^2$. These inequalities are satisfied by the actual estimates presented in Tables 8 and 11 despite the fact that they are not explicitly imposed on the objective function for estimation.

6.2 Extension of the $A_1(3)$ model

The $A_1(3)$ model is extended in the same way as the Gaussian term structure model. Specifically, the extension is achieved by making the covariance matrix of the $A_1(3)$ model depend on all factors, where the dependence structure is exactly the same as that of the QC or QE model specified in Section 2.2. The extension to the QS model is not considered here based on the previous results showing similarities between the QC and QS models. It is important to note that as is the previous extension, the specifications for both the risk-neutral and physical drift vectors are the same between the $A_1(3)$ and extended models in order to control for potential influences of the drift vectors and thus focus exclusively on potential influences of the covariance matrix on factor identification and volatility prediction.

6.3 Identified factors

One of the major findings of this study is that the volatility is of little relevance to factor identification. This is so even using the models with the volatility that is originally linked to the cross-section of interest rates and using the inversion method with the assumption that some interest rates are measured without error.

This finding is confirmed in Figure 6 displaying the time-series of identified factors: The left panels present the results for the $A_1(3)$ model together with the time series of observed variables that track those of the factors, and the right panels present the results for the QC and QE models together. Again, for each factor the three plots are visually indistinguishable. Furthermore, it is seen from the left panels that the identified factors are well tracked by (a combination of) observed interest rates. In particular, V_t and the ten-year yield (a proxy of the level factor) are highly correlated with sample correlation over the whole period 0.96, which is in line with the results by the previous studies. Furthermore, μ_t exhibits a moderately high correlation with the curvature factor, 0.70.

These results imply that all factors in the $A_1(3)$ model, including V_t , are interest-rate specific, when identified from the cross-section of interest rates with the inversion method. This implication, which has already been addressed in CDGJ (2009) based on the affine models, seems to hold more generally, regardless of the specification in the factor covariance matrix. The interpretation and implication of the results are further discussed in Section 6.6.

6.4 Estimated parameters

Among a maximum of fourteen identifiable parameters associated with the risk-neutral distribution for the $A_1(3)$ model, ten are estimated significantly and displayed in Table 8: As before, the estimates in the physical drift vector are collected altogether in Table 11. First, between with and without the NAC, there is no parameter that exhibits a statistically significant change. However, of note is that the estimate of s_{33} is slightly increased by removing the NAC so that V_t becomes slightly more volatile. Meanwhile, V_t is realized as the most persistent factor in the risk-neutral world as evidenced by the estimate of $\kappa_{V,3}$, -0.006 . Second, compared with the Gaussian model, the goodness-of-fit to the cross-section of interest rates remains unchanged. Specifically, the estimate of $\varsigma \times 10^4$ is 6.15 bps and the value of LogL_C is 11472, both of which are almost the same as those for the Gaussian model. The result is in line with Bikbov and Chernov (2011), who document that these two affine models are hardly distinguishable in terms of describing the cross-section of interest rates.

Tables 9 and 10 present the estimates for the QC and QE models, respectively. First of all, looking at columns labeled “with NAC”, the estimates in the risk-neutral drift for both models are basically similar to those for the original $A_1(3)$ model, as is the case for the Gaussian and extended models. Also, it is confirmed by comparing the values of the LogL and its components

between the $A_1(3)$ and extended models that the extension of the covariance matrix does not contribute to the cross-sectional fit but it does to the time-series fit.

Looking more precisely at the estimates for the QC model in Table 9, the level of V_t is found to be irreverent to the volatility of V_t as β_3^3 in $h_3(X_t)$ is set to zero in both cases with and without the NAC. Since V_t is highly correlated with the ten-year yield as seen in Figure 6, the result indicates that long-term yields have little predictive power for their own volatilities. On the other hand, r_t , which also represents an interest rate level at the other end of the yield curve, is found to be reverent to the volatilities of all factors. In particular, because the estimates of β_1^i ($i = 1, 2, 3$) are negative, the decrease in r_t increases the volatility, which is consistent with recent movements in interest rates.

As for the differences in parameter estimates between with and without the NAC, the results from the $A_1(3)$ -based extension are basically the same as those from the Gaussian-based extension. Namely, while the parameters associated with the volatilities of r_t and μ_t are stable, those associated with the volatility of the persistent factor, V_t , are changeable. For example, Table 9 shows that the estimate of β_1^3 for the QC model changes from -0.027 to -0.062 by removing the NAC. In spite of this decrease, however, the volatility of V_t is actually higher without the NAC than with the NAC, which is due to the increase in β_0^3 from 0.011 to 0.013 . From Table 10, the volatility of V_t for the QE model is most affected by $l_3(X_t)$ as the estimate of $\sin \varphi_i^P$ indicates $\sin \varphi_i^P < \cos \varphi_i^P$ for all i . Then, consistent with the result from the Gaussian-based extension, the volatility of V_t is higher without the NAC than with the NAC as m_3^3 is increased from 0.11 to 0.16 .

6.5 Volatility Prediction

Table 12 presents the ratios of the MSEs of four-week-ahead ($h = 4$) variance forecasts. In Panel A displaying the in-sample results, neither the $A_1(3)$ nor QC model exhibits a statistically significant deviation from one at the average or respective maturities. In contrast, the QE model generates forecasts that are significantly different between with and without the NAC at the average and medium to long maturities. But these significant ratios are all above one, indicating that the imposition of the NAC is not necessarily restrictive for volatility prediction.

In Panel B displaying the out-of-sample results, the opposite pattern emerges. Significant differences in forecasts are detected for the $A_1(3)$ and QC models, whereas none for the QE model. Specifically, for the $A_1(3)$ model, the values of the 3-1 ratio at the average and $\tau = 2, 5, 10$

are significantly below one, which is mainly attributed to the 3-2 ratio. However, the 2-1 ratio measuring the cross-section-related difference also exhibits a significant deviation at $\tau \geq 2$. Since for the $A_1(3)$ model, the cross-sectional relation between factors and yields is linear in both cases with and without the NAC, the cross-section-related difference arises as the difference in parameter values: structural model parameters with the NAC versus regression parameters without the NAC. The result again suggests that even a tiny difference in the cross-sectional relation can generate a statistically significant difference in volatility prediction.

To understand the magnitude of the 3-1 ratio from an economic point of view, Table 13 presents the RMSEs of four-week-ahead ($h = 4$) annualized standard deviation forecasts in bps. The differences in forecasts between with and without the NAC are all within 2 bps in absolute value for both in- and out-of-sample.

To see the differences in forecasts at each point in time, Figures 7–9 display time-series of four-week-ahead ($h = 4$) forecasts of annualized standard deviation of a τ -year yield ($\tau = 0.5, 2, 10$) generated by the $A_1(3)$, QC, and QE models, respectively. The basic pattern remains unchanged from that for the Gaussian-based QC and QE models shown in Figures 4 and 5. Namely, for all three models, the differences in forecasts between with and without the NAC are hardly visible at $\tau = 0.5, 2$. But they are occasionally seen at $\tau = 10$, which is attributed to the differences in parameter estimates for the persistent factor (here V_t) as discussed in Section 6.4.

6.6 Discussion

In the $A_1(3)$ model, V_t is realized as the most persistent factor, which in turn is the least relevant factor for volatility. Indeed, looking at Figure 7 displaying the time-series of volatility forecasts generated by this model, it is apparent that the forecast series do not vary largely nor track the realized series closely. But it may not be fair to judge by this phenomenon that the $A_1(3)$ model is misspecified. As noted in Section 6.3, all factors, when identified from the cross-section of interest rates with the inversion method, are interest-rate specific. In this respect, all models, except those accommodating (nearly) USV factors, are more or less misspecified unless linear or nonlinear combinations of interest rates are capable of capturing volatility dynamics. Since the $A_1(3)$ model has a factor covariance matrix that is slightly more restrictive for the recent data on U.S. interest rates than its extended version, implied behavior of the volatility does not match with the realized counterpart as displayed in Figure 7. However, the fact remains that the $A_1(3)$ model and its extended version attempt to capture the volatility with the level of

interest rates.

If one wishes V_t in the $A_1(3)$ model to behave like standard volatility measures in any sample period, the inversion method that forcefully fits the volatility to the cross-section of interest rates should be avoided. Instead, filtering-based estimation techniques such as the (extended) Kalman filter are suitable, where all interest rates are assumed to be measured with error; see, e.g., Chen and Scott (2003), de Jong (2000), Duan and Simonato (1999), and Duffee and Stanton (2012). If the change of estimation methods alone is not sufficient, it may be necessary to change model structure and incorporate (nearly) USV factors as in Collin-Dufresne and Goldstein (2002). It may be the case, however, that the parameter constraints required for obtaining the USV factors are not supported by the data; see e.g., Bikbov and Chernov (2009), and Thompson (2008). Then, an alternative approach is not to impose these constraints but to apply an unrestricted model to the joint data on interest rates and volatility measures constructed from options and/or high-frequency observations as in Bikbov and Chernov (2009, 2011), and Cieslak and Povala (2011).

Although the previous studies have already developed these innovative approaches and obtained a number of insights, it seems worth discussing and confirming them from a perspective of level-dependent volatility that is the opposite of USV.

7 Concluding Remarks

The imposition of no-arbitrage condition (NAC) produces a constraint that the volatility is linked to the cross-section of interest rates. This study has examined using interest rate data whether this constraint is binding from two perspectives: The first is that of the identification of latent factors and the second is that of the prediction of interest rate volatility. Taken into consideration the importance of the earlier study which reveals that the imposition of the NAC can be made irrelevant to the prediction of the level of interest rates (Duffee, 2011, JSZ, 2011), the examination from the above two perspectives seems worth advancing.

To examine a practical impact of the NAC, this study considers a particular setting regarding models and estimation method with care. The models have factor covariance matrix driven by common factors driving interest rates. The estimation method relies on the assumption that some interest rates are measured without error. These considerations are motivated to match the models closely to the cross-section of interest rates and thus highlight the impact of cross-sectional constraints due to no-arbitrage. Nevertheless, the impact is found to be minor on both

factor identification and volatility prediction. Specifically, the factors identified from the cross-section of interest rates do not much differ, after controlling for the risk-neutral drift, across models with different specifications in the factor covariance matrix. It is evident that factor identification is almost dominantly determined by the risk-neutral drift and that the covariance matrix is of little relevance to it. It then follows that it is not difficult to obtain nearly unspanned factors: Simply add a factor to the covariance matrix of existing factors but not to the risk-neutral drift vector. In addition, parameter estimates in the factor covariance matrix do not differ much between with and without the NAC, though the imposition of the NAC tends to slightly reduce the volatility of a persistent factor affecting the long end of the yield curve. Consequently, volatility forecasts do not substantially differ between with and without the NAC on average. Even if there is an unacceptable difference at some points in time, it can be reduced by introducing nearly USV factors into no-arbitrage term structure models, which is achieved without much difficulty as noted above.

These findings will have important implications for time-series analysis on interest rate dynamics. Conventional approaches in time-series analysis are to first identify interest rate factors independently of volatility modeling and then estimate the dynamics of the factors and their volatilities independently of cross-sectional constraints. This study confirms that the results produced by these approaches do not substantially differ from those that would be obtained by explicitly imposing the NAC if a model is sufficiently flexible. This condition is not actually binding as many models in time-series analysis originally embed unspanned factors such as GARCH volatilities.

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

A1. Outline of the method

The method is originally developed by Shoji (2002) and applied to the pricing of bonds by Takamizawa and Shoji (2009). The method generally allows for the computation of up to n -th conditional moments, if they exist, for a d -dimensional diffusion process. To ease the explanation, we limit our attention to the case of $(n, d) = (2, 2)$, i.e., the conditional first and second moments of a two-dimensional diffusion process. As seen below, n can be considered as the order of approximation.

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

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

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 changes in X_t conditioned on time $s < t$:

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

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

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

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

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

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

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

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

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,

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

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

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

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),

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

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

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

where

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

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

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

Equation (47) can be solved as

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

If, in addition, A is invertible, we obtain

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

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} ($i, j = 1, \dots, d$) are taken to compute the elements of $A(X_s)$. Omitting the residual vector, R or \hat{R} , leads to the approximation formula. According to Shoji (2002), both R and \hat{R} have order of $O((t-s)^{(n+3)/2})$. Thus, n can be considered as the order of approximation. In pricing bonds, we consider $n = 3$.

It is also noted that R contains conditional expectations of derivatives of f_i higher than the first order and derivatives of g_{ij} higher than the second order. Then, if f_i and g_{ij} are linear and quadratic in X_s , respectively, there is no residual term. In other words, the conditional moments computed by the formula are exact. Even in this case, the use of this formula may be beneficial when the derivation of closed-form conditional moments is demanding.

A2. Application to bond prices

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

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

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,

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

and we have

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

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

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

A3. Application to yield moments

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

$$dY(X_t, \tau) = f_{d+1}(X_t, \tau)dt + \xi'_{d+1}(X_t, \tau)dW_t, \quad (54)$$

where

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

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

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 $\xi_i' \xi_{d+1}$ ($i = 1, \dots, d+1$).

In computing $A(\hat{X}_t)$, 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 $Y(X_t, \tau)$ up to the second order, which complicates the calculation. To avoid the tedious calculation, the derivatives of $Y(X_t, \tau)$ 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 $E_t[Y(X_{t+h\Delta}, \tau)^n]$ ($n = 1, 2$) is also investigated in Appendix B.

Appendix B: 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 concerns 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, aiming at examining how close the approximate solution is to the closed-form and numerical solutions, respectively.

B1. Comparison with the closed-form solution for bond prices

We treat the Gaussian model as the true model. We divide the analysis into three steps according to the degree of approximation involved. Let Θ_0 be the parameter vector of the the Gaussian model, the elements of which are set at the estimates presented in columns of Table 1 labeled “with NAC”. 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

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

where Y and \tilde{Y} stand for the closed-form and approximate functions, respectively: Note that in Appendix B, the tilde symbol is used for clarifying the fact that the approximation method is involved.

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 method, the state vector is first extracted, which is denoted as $\tilde{X}(Y_t^p; \Theta_0)$, and the rest of the yields are computed. Then, we compare:

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

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

In the last step, we examine the 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 is given regarding the true value of the parameter or state vector. Instead, using the approximation method, the parameter vector of the Gaussian model is first estimated; denote it as $\tilde{\Theta}_0$. Next, the state vector is extracted; denote it as $\tilde{X}(Y_t^p; \tilde{\Theta}_0)$. Finally, the rest of the yields are computed. Then, we compare:

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

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

Apart from the parameter vector, the key input for these comparisons is Y_t^p . We use the actual data on Y_t^p . To condense the analysis, we pick up nine observations from the entire sample as follows. First, Y_t^p is transformed to the conventional level (lev_t), slope (slo_t), and curvature (cur_t) factors by

$$\begin{pmatrix} lev_t \\ slo_t \\ cur_t \end{pmatrix} = \begin{pmatrix} 0 & 0 & 1 \\ -1 & 0 & 1 \\ -1 & 2 & -1 \end{pmatrix} \begin{pmatrix} y_{t,0.5} \\ y_{t,2} \\ y_{t,10} \end{pmatrix}. \quad (57)$$

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 both typical and atypical states.

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

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

Panel C presents the results for the third-step comparison, where no prior information is given. Compared with Panel B, the magnitude of the approximation errors for the state variables

is generally larger, even though parameter estimates do not much differ between the approximate and closed-form solutions (this estimation result is not presented in any table but available upon request). This is the reality. Here, an error pattern is less clear, but it is seen that r tends to be overvalued and θ and ϵ tend to be undervalued. Also, the difficulty of the approximation method is not limited at the maximum-slope 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 as in the second-step comparison.

B2. Comparison with the numerical solution for bond prices

We employ the Monte Carlo (MC) method to evaluate the accuracy. As models that do not have closed-form expressions for no-arbitrage bond prices, we use the Gaussian-based QS and QE models. Let Θ be the true parameter vector, the elements of which are set at the estimates presented in Table 3 for the QS model and Table 4 for the QE model. Further, the state vector is extracted by the approximation method, but not the MC method with which the extraction is computationally very demanding. The extracted state vector is denoted as $\tilde{X}(Y_t^p; \Theta)$. Then, we compare

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

In the MC simulations, $\{X_s\}_t^{t+\tau}$ is generated from (3) (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 B2 presents the differences in yields between the approximate and MC solutions in bps. Generally, the error pattern is similar to that for the first-step comparison with the closed-form solution. For maturities 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 is the largest at the maximum-slope state but still around 2 bps.

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

B3. Accuracy to yield volatility

We continuously use the Gaussian-based QS and QE models and the MC method to evaluate the accuracy of the approximation to yield volatility at the same nine states. In the MC simulations, $\{X_s\}_t^{t+h\Delta}$ is generated from (2) (the physical distribution), where dt is replaced by $\Delta t = 1/1,000$ and Σ by Σ_t . For simplicity, once $X_{t+h\Delta}$ is obtained, a model-implied τ -year yield (and its squared value) is computed by the approximation method, but not the MC method with which the computation is very demanding. This approach seems to be justified by the results presented in Table B2. This procedure is repeated 10,000 times with antithetic variates. Then, we compare

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

where $\text{var}_t^{mc}[\cdot]$ and $\text{var}_t^{ap}[\cdot]$ stand for the conditional variance computed by the MC and approximation methods, respectively. We set $h = 4$, but other values of h up to $h = 52$ do not qualitatively or quantitatively change the results presented below.

Table B3 presents the differences in annualized standard deviations between the approximate and MC solutions in bps. All approximation errors are around or below 1 bp in absolute value.

It is noted that a more rigorous investigation of the accuracy requires for comparing

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

That is, no approximation is involved in the former in computing the τ -year yield at time $t+h\Delta$. In reality, therefore, the approximation errors would be larger than presented in Table B3, even though the differences between $Y(X_{t+h\Delta}, \tau)$ and $\tilde{Y}(X_{t+h\Delta}, \tau)$ at $\tau \leq 10$ are actually small as presented in Table B2.

References

- Aït-Sahalia, Y., 1996, Testing continuous-time models of the spot interest rate, *Review of Financial Studies* 9, 385-426.
- Almeida, C., and J. Vicente, 2008, The role of no-arbitrage on forecasting: Lessons from a parametric term structure model, *Journal of Banking Finance* 32, 2695-2705.
- Andersen, T. G., and L. Benzoni, 2010, Do bonds span volatility risk in the U.S. Treasury market? A specification test for affine term structure models, *Journal of Finance* 65, 603-653.
- Andersen, T. G., and J. Lund, 1997a, Estimating continuous-time stochastic volatility models of the short-term interest rate, *Journal of Econometrics* 77, 343-377.
- Andersen, T. G., and J. Lund, 1997b, Stochastic volatility and mean drift in the short rate diffusion: Sources of steepness, level and curvature in the yield curve, Working paper, Northwestern University Kellogg School of Management.
- 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.
- Bali, T. G., 2000, Testing the empirical performance of stochastic volatility models of the short-term interest rate, *Journal of Financial and Quantitative Analysis* 35, 191-215.
- Ball, C. A., and W. N. Torous, 1999, The stochastic volatility of short-term interest rates: Some international evidence, *Journal of Finance* 54, 2339-2359.
- Bikbov, R., and M. Chernov, 2009, Unspanned stochastic volatility in affine models: Evidence from Eurodollar futures and options, *Management Science* 55, 1292-1305.
- Bikbov, R., and M. Chernov, 2011, Yield curve and volatility: Lessons from Eurodollar futures and options, *Journal of Financial Econometrics* 9, 66-105.
- Bollerslev, T., 1990, Modelling the coherence in short-run nominal exchange rates: A multivariate generalized arch model, *Review of Economics and Statistics* 72, 498-505.

- Brennan, M. J., and E. S. Schwartz, 1979, A continuous time approach to the pricing of bonds, *Journal of Banking Finance* 3, 133-155.
- Brenner, R. J., R. H. Harjes, and K. F. Kroner, 1996, Another look at models of the short-term interest rate, *Journal of Financial and Quantitative Analysis* 31, 85-107.
- Buraschi, A., A. Cieslak, and F. Trojani, 2008, Correlation risk and the term structure of interest rates, Working paper, Imperial College London Business School.
- 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.
- Chen, R., and L. Scott, 2003, Multi-factor CIR models of the term structure: Estimates and tests from a state-space model using a Kalman filter, *Journal of Real Estate Finance and Economics* 27, 143-172.
- Cheridito, P., D. Filipovic, and R. L. Kimmel, 2007, Market price of risk specifications for affine models: Theory and evidence, *Journal of Financial Economics* 83, 123-170.
- Christensen, J. H., F. X. Diebold, and G. D. Rudebusch, 2009, An arbitrage-free generalized Nelson-Siegel term structure model, *Econometrics Journal* 12, C33-C64.
- Christensen, J. H., F. X. Diebold, and G. D. Rudebusch, 2011, The affine arbitrage-free class of Nelson-Siegel term structure models, *Journal of Econometrics* 164, 1-20.
- Christiansen, C., and J. Lund, 2005, Revisiting the shape of the yield curve: The effect of interest rate volatility, Working paper, University of Aarhus.
- Cieslak, A., and P. Povala, 2011, Information in the term structure of yield curve volatility, Working paper, Northwestern University Kellogg School of Management.
- Collin-Dufresne, P., and R. S. Goldstein, 2002, Do bonds span the fixed income markets? Theory and evidence for unspanned stochastic volatility, *Journal of Finance* 57, 1685-1730.
- Collin-Dufresne, P., R. S. Goldstein, and C. S. Jones, 2008, Identification of maximal affine term structure models, *Journal of Finance* 63, 743-795.

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

Coroneo, L., K. Nyholm, and R. Vidova-Koleva, 2011, How arbitrage-free is the Nelson-Siegel model? *Journal of Empirical Finance* 18, 393-407.

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.

de Jong, F., 2000, Time series and cross-section information in affine term-structure models, *Journal of Business and Economic Statistics* 18, 300-314.

Diebold, F. X., 2015, Comparing predictive accuracy, twenty years later: A personal perspective on the use and abuse of Diebold-Mariano tests, *Journal of Business and Economic Statistics* 33, 1-9.

Diebold, F. X., and C. Li, 2006, Forecasting the term structure of government bond yields, *Journal of Econometrics* 130, 337-364.

Diebold, F. X., C. Li, and V. Z. Yue, 2008, Global yield curve dynamics and interactions: A dynamic Nelson-Siegel approach, *Journal of Econometrics* 146, 351-363.

Diebold, F. X., and R. S. Mariano, 1995, Comparing predictive accuracy, *Journal of Business and Economic Statistics* 13, 253-263.

Diebold, F. X., G. D. Rudebusch, and S. B. Aruoba, 2006, The macroeconomy and the yield curve: A dynamic latent factor approach, *Journal of Econometrics* 131, 309-338.

Duan, J., and J. Simonato, 1999, Estimating and testing exponential-affine term structure models by Kalman filter, *Review of Quantitative Finance and Accounting* 13, 111-135.

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.

- Duffee, G. R., and R. H. Stanton, 2012, Estimation of dynamic term structure models, *Quarterly Journal of Finance* 2(2).
- 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.
- Engle, R. F., 2002, Dynamic conditional correlation: A simple class of multivariate generalized autoregressive conditional heteroskedasticity models, *Journal of Business and Economic Statistics* 20, 339-350.
- Fan, R., A. Gupta, and P. Ritchken, 2003, Hedging in the possible presence of unspanned stochastic volatility: Evidence from swaption markets, *Journal of Finance* 58, 2219-2248.
- 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.
- Gourieroux, C., and R. Sufana, 2011, Discrete time Wishart term structure models, *Journal of Economic Dynamics and Control* 35, 815-824.
- Han, B., 2007, Stochastic volatilities and correlations of bond yields, *Journal of Finance* 62, 1491-1524.
- Ilmanen, A., 1999, Convexity bias and the yield curve, in N. Jegadeesh, and B. Tuckman, *Advanced Fixed-Income Valuation Tools*, Wiley, New York, 25-57.
- Jacobs, K., and L. Karoui, 2009, Conditional volatility in affine term-structure models: Evidence from Treasury and swap markets, *Journal of Financial Economics* 91, 288-318.
- Jarrow, R. A., H. Li, and F. Zhao, 2007, Interest rate caps smile too! But can the LIBOR market models capture the smile? *Journal of Finance* 62, 345-382.
- Joslin, S., 2010, Pricing and hedging volatility risk in fixed income markets, working paper, USC Marshall School of Business.
- 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.
- Koopman, S. J., M. I. P. Mallee, and M. Van der Wel, 2010, Analyzing the term structure of interest rates using the dynamic Nelson-Siegel model with time-varying parameters, *Journal of Business and Economic Statistics* 28, 329-343.

Litterman, R., and J. Scheinkman, 1991, Common factors affecting bond returns, *Journal of Fixed Income* 1, 54-61.

Longstaff, F. A., P. Santa-Clara, and E. S. Schwartz, 2001, The relative valuation of caps and swaptions: Theory and empirical evidence, *Journal of Finance* 56, 2067-2109.

Nelson, C. R., and A. F. Siegel, 1987, Parsimonious modeling of yield curves, *Journal of Business* 60, 473-489.

Newey, W., and K. 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.

Pérignon, C., and C. Villa, 2006, Sources of time variation in the covariance matrix of interest rates, *Journal of Business* 79, 1535-1549.

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.

	with NAC		without NAC	
Parameters in the risk-neutral drift vector				
κ_1	0.946	(0.088)		
κ_2	0.017	(0.002)		
κ_3	0.957	(0.090)		
$\bar{\theta}$	0.147	(0.008)		
Parameters in the covariance matrix				
$\sigma_{11} \times 10^4$	0.725	(0.033)	0.724	(0.032)
$\sigma_{22} \times 10^4$	1.218	(0.064)	1.307	(0.062)
$\sigma_{33} \times 10^4$	8.100	(0.449)	7.883	(0.344)
$\sigma_{12} \times 10^4$	0.097	(0.035)	0.103	(0.036)
$\sigma_{13} \times 10^4$	-1.005	(0.120)	-0.983	(0.093)
$\sigma_{23} \times 10^4$	0.000		0.000	
$\varsigma \times 10^4$	6.149	(0.092)		
LogL	22265		9351	
LogL _T	9350			
LogL _C	11472			

Table 1: Parameter estimates (standard errors) for the Gaussian model

The risk neutral and physical distributions of instantaneous changes in $X_t = (r_t \theta_t \epsilon_t)'$ are given by (1) and (2), respectively. To obtain the estimates without NAC, the same X_t as extracted by imposing the NAC is used and its physical distribution is estimated using time-series properties of the data alone. ς is the standard deviation of measurement errors. LogL is the log-likelihood value, which is decomposed into time-series (LogL_T), cross-sectional (LogL_C), and Jacobian (not displayed) parts, for the no-arbitrage case. In-sample data from January 4, 1991 to April 9, 2003 are used for the estimation (641 observations).

	with NAC		without NAC	
Parameters in the risk-neutral drift vector				
κ_1	0.938	(0.091)		
κ_2	0.021	(0.003)		
κ_3	0.968	(0.096)		
$\bar{\theta}$	0.124	(0.007)		
Parameters in $h_1(X_t)$				
β_0^1	-0.006	(0.002)	-0.006	(0.002)
β_1^1	0.041	(0.013)	0.042	(0.009)
β_2^1	0.169	(0.021)	0.169	(0.019)
β_3^1	-0.086	(0.012)	-0.087	(0.011)
Parameters in $h_2(X_t)$				
β_0^2	0.013	(0.002)	0.015	(0.002)
β_1^2	-0.026	(0.025)	-0.038	(0.027)
β_2^2	-0.014	(0.030)	-0.030	(0.031)
β_3^2	-0.042	(0.018)	-0.050	(0.019)
Parameters in $h_3(X_t)$				
β_0^3	0.007	(0.006)	0.007	(0.006)
β_1^3	-0.123	(0.071)	-0.120	(0.068)
β_2^3	0.378	(0.073)	0.365	(0.058)
β_3^3	-0.151	(0.043)	-0.147	(0.041)
Parameters in R				
ρ_{12}	0.081	(0.043)	0.075	(0.040)
ρ_{13}	-0.420	(0.043)	-0.421	(0.036)
ρ_{23}	0.000		0.000	
$\varsigma \times 10^4$	6.131	(0.091)		
LogL	22358		9418	
LogL _T	9417			
LogL _C	11478			

Table 2: Parameter estimates (standard errors) for the QC model

The risk neutral and physical distributions of instantaneous changes in $X_t = (r_t \theta_t \epsilon_t)'$ are given by (3) and (2) (where Σ is replaced by Σ_t), respectively. To obtain the estimates without NAC, the same X_t as extracted by imposing the NAC is used and its physical distribution is estimated using time-series properties of the data alone. The covariance matrix Σ_t is decomposed as $\Sigma_t = H_t R H_t$ with H_t being the diagonal matrix and R the constant correlation matrix. The i -th diagonal element of H_t is specified as $h_i(X_t) = \beta_0^i + \beta^{i'} X_t$ ($i = 1, 2, 3$). ς is the standard deviation of measurement errors. LogL is the log-likelihood value, which is decomposed into time-series (LogL_T), cross-sectional (LogL_C), and Jacobian (not displayed) parts, for the no-arbitrage case. In-sample data from January 4, 1991 to April 9, 2003 are used for the estimation (641 observations).

	with NAC		without NAC	
Parameters in the risk-neutral drift vector				
κ_1	0.921	(0.074)		
κ_2	0.018	(0.002)		
κ_3	0.998	(0.088)		
θ	0.135	(0.007)		
Parameters in $h_1(X_t)$				
β_0^1	-0.007	(0.002)	-0.007	(0.002)
β_1^1	0.038	(0.016)	0.039	(0.014)
β_2^1	0.180	(0.024)	0.181	(0.023)
β_3^1	-0.088	(0.013)	-0.088	(0.013)
Parameters in $h_2(X_t)$				
β_0^2	0.010	(0.000)	0.011	(0.000)
β_1^2	0.000		0.000	
β_2^2	0.000		0.000	
β_3^2	-0.057	(0.014)	-0.067	(0.015)
Parameters in $h_3(X_t)$				
β_0^3	0.000		0.000	
β_1^3	-0.092	(0.056)	-0.089	(0.037)
β_2^3	0.457	(0.051)	0.450	(0.033)
β_3^3	-0.182	(0.032)	-0.176	(0.029)
Parameters in R_t				
γ_0	0.681	(0.552)	0.651	(0.540)
γ_1	-2.587	(7.534)	-2.488	(7.419)
γ_2	5.874	(7.996)	6.125	(7.848)
γ_3	0.559	(5.197)	0.440	(5.064)
δ_{12}	-0.184	(0.088)	-0.187	(0.089)
$\varsigma \times 10^4$	6.129	(0.092)		
LogL	22357		9417	
LogL _T	9416			
LogL _C	11478			

Table 3: Parameter estimates (standard errors) for the QS model

The risk neutral and physical distributions of instantaneous changes in $X_t = (r_t \theta_t \epsilon_t)'$ are given by (3) and (2) (where Σ is replaced by Σ_t), respectively. To obtain the estimates without NAC, the same X_t as extracted by imposing the NAC is used and its physical distribution is estimated using time-series properties of the data alone. The covariance matrix Σ_t is decomposed as $\Sigma_t = H_t R_t H_t$ with H_t being the diagonal matrix and R_t the correlation matrix. The i -th diagonal element of H_t is specified as $h_i(X_t) = \beta_0^i + \beta^i X_t$ ($i = 1, 2, 3$). The off-diagonal elements of R_t are given by $\rho_{12}(X_t) = \delta_{12} c(X_t)$, $\rho_{13}(X_t) = (1 - \delta_{12}^2)^{0.5} c(X_t)$, and $\rho_{23}(X_t) = 0$, where $c(X_t) = \frac{2}{1 + \exp\{\gamma_0 + \gamma' X_t\}} - 1$. ς is the standard deviation of measurement errors. LogL is the log-likelihood value, which is decomposed into time-series (LogL_T), cross-sectional (LogL_C), and Jacobian (not displayed) parts, for the no-arbitrage case. In-sample data from January 4, 1991 to April 9, 2003 are used for the estimation.

	with NAC		without NAC	
Parameters in the risk-neutral drift vector				
κ_1	0.967	(0.077)		
κ_2	0.016	(0.002)		
κ_3	0.975	(0.081)		
$\bar{\theta}$	0.146	(0.009)		
Parameters in P				
$\sin \varphi_1^P$	0.000		0.000	
$\sin \varphi_2^P$	0.130	(0.011)	0.132	(0.010)
$\sin \varphi_3^P$	0.000		0.000	
Parameters in $l_1(X_t)$				
$c_0^1 \times 10^4$	1e-4		1e-4	
m_1^1	0.000		0.000	
m_2^1	0.024	(0.003)	0.024	(0.002)
m_3^1	0.231	(0.036)	0.237	(0.021)
$\sin \varphi_1^1$	0.289	(0.049)	0.286	(0.034)
$\sin \varphi_2^1$	0.616	(0.025)	0.619	(0.020)
$\sin \varphi_3^1$	-0.674	(0.049)	-0.679	(0.037)
Parameters in $l_2(X_t)$				
$c_0^2 \times 10^4$	0.990	(0.066)	1.044	(0.072)
m_1^2	0.000		0.000	
m_2^2	0.000		0.000	
m_3^2	0.058	(0.042)	0.140	(0.067)
$\sin \varphi_1^2$	0.000		0.000	
$\sin \varphi_2^2$	0.413	(0.170)	0.467	(0.097)
$\sin \varphi_3^2$	-0.351	(0.193)	-0.428	(0.099)
Parameters in $l_3(X_t)$				
$c_0^3 \times 10^4$	1e-4		1e-4	
m_1^3	0.043	(0.008)	0.040	(0.009)
m_2^3	0.413	(0.132)	0.346	(0.095)
m_3^3	2.109	(0.633)	2.393	(0.626)
$\sin \varphi_1^3$	0.470	(0.092)	0.391	(0.105)
$\sin \varphi_2^3$	0.586	(0.069)	0.628	(0.045)
$\sin \varphi_3^3$	-0.651	(0.107)	-0.715	(0.075)
$\varsigma \times 10^4$	6.130	(0.090)		
LogL	22416		9477	
LogL _T	9472			
LogL _C	11481			

Table 4: Parameter estimates (standard errors) for the QE model

The covariance matrix Σ_t is decomposed as $\Sigma_t = PL_tP'$ with L_t being the diagonal eigenvalue matrix and P the orthonormal eigenvector matrix parameterized in equation (13). The i -th diagonal element of L_t is specified as $l_i(X_t) = c_0^i + X_t'\Gamma^i X_t$ ($i = 1, 2, 3$), where $c_0^i > 0$ and Γ^i is a non-negative definite matrix. Γ^i is also parameterized based on the spectral decomposition, $\Gamma^i = Q^i M^i Q^{i'}$, 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 orthonormal eigenvector matrix parameterized in equation (17). In-sample data from January 4, 1991 to April 9, 2003 are used for the estimation.

	Gaussian model				QC model			
	with NAC		without NAC		with NAC		without NAC	
$K_{r,0}$	0.028	(0.015)	0.028	(0.014)	-0.002	(0.019)	-0.003	(0.018)
$K_{r,1}$	-1.160	(0.246)	-1.156	(0.219)	-0.813	(0.234)	-0.803	(0.211)
$K_{r,2}$	0.513	(0.230)	0.511	(0.203)	0.704	(0.273)	0.707	(0.249)
$K_{r,3}$	1.164	(0.149)	1.160	(0.145)	0.983	(0.167)	0.977	(0.162)
$K_{\theta,0}$	0.035	(0.019)	0.038	(0.020)	0.036	(0.021)	0.036	(0.022)
$K_{\theta,1}$	0.000		0.000		0.000		0.000	
$K_{\theta,2}$	-0.573	(0.282)	-0.610	(0.301)	-0.585	(0.307)	-0.589	(0.321)
$K_{\theta,3}$	0.000		0.000		0.000		0.000	
$K_{\epsilon,0}$	-0.077	(0.038)	-0.075	(0.037)	-0.052	(0.040)	-0.054	(0.039)
$K_{\epsilon,1}$	1.232	(0.724)	1.187	(0.697)	0.754	(0.759)	0.776	(0.731)
$K_{\epsilon,2}$	0.000		0.000		0.000		0.000	
$K_{\epsilon,3}$	-1.512	(0.525)	-1.461	(0.485)	-1.240	(0.542)	-1.264	(0.513)

	QS model				QE model			
	with NAC		without NAC		with NAC		without NAC	
$K_{r,0}$	-0.002	(0.019)	-0.002	(0.018)	0.010	(0.016)	0.010	(0.015)
$K_{r,1}$	-0.811	(0.231)	-0.784	(0.209)	-1.053	(0.297)	-1.071	(0.281)
$K_{r,2}$	0.692	(0.266)	0.674	(0.251)	0.736	(0.229)	0.739	(0.210)
$K_{r,3}$	0.982	(0.167)	0.971	(0.163)	1.177	(0.216)	1.175	(0.214)
$K_{\theta,0}$	0.039	(0.022)	0.040	(0.023)	0.035	(0.020)	0.036	(0.023)
$K_{\theta,1}$	0.000		0.000		0.000		0.000	
$K_{\theta,2}$	-0.631	(0.318)	-0.634	(0.339)	-0.567	(0.292)	-0.573	(0.337)
$K_{\theta,3}$	0.000		0.000		0.000		0.000	
$K_{\epsilon,0}$	-0.058	(0.038)	-0.058	(0.037)	-0.058	(0.045)	-0.059	(0.044)
$K_{\epsilon,1}$	0.869	(0.730)	0.862	(0.699)	0.930	(0.840)	0.954	(0.822)
$K_{\epsilon,2}$	0.000		0.000		0.000		0.000	
$K_{\epsilon,3}$	-1.293	(0.544)	-1.318	(0.507)	-1.141	(0.696)	-1.090	(0.657)

Table 5: Parameter estimates (standard errors) in the physical drift for the Gaussian and extended models

The specification of the physical drift is given by (2) for all models. The same parameters are estimated between the Gaussian and extended models in order to control for potential influences from the physical drift. In-sample data from January 4, 1991 to April 9, 2003 are used for the estimation (641 observations).

Maturity	0.5	1	2	3	5	10	Ave.
Panel A: In-sample (1991/1 – 2003/4)							
QC model							
MSE ₃ /MSE ₁	1.002	1.002	0.998	0.996	1.001	1.005	1.000
MSE ₂ /MSE ₁	1.000	0.996	1.000	1.000	1.004	0.999	1.000
MSE ₃ /MSE ₂	1.002	1.006	0.998	0.996	0.997	1.006	1.000
QS model							
MSE ₃ /MSE ₁	1.000	0.999	1.000	1.001	1.019	1.016	1.005
MSE ₂ /MSE ₁	1.000	0.998	1.000	1.000	1.006	0.998	1.000
MSE ₃ /MSE ₂	1.000	1.001	1.000	1.001	1.013	1.018	1.005
QE model							
MSE ₃ /MSE ₁	0.996	0.964*	0.965**	0.991	1.053	1.066	0.999
MSE ₂ /MSE ₁	1.000*	0.984	0.993*	0.992	0.994	0.994	0.992*
MSE ₃ /MSE ₂	0.996	0.980	0.972*	0.999	1.059	1.073	1.007
Panel B: Out-of-sample (2003/4 – 2009/5)							
QC model							
MSE ₃ /MSE ₁	0.997	0.995*	0.973	0.960	0.901*	0.925	0.949
MSE ₂ /MSE ₁	1.000*	0.998	1.000	1.008	0.988*	1.004	1.000
MSE ₃ /MSE ₂	0.997	0.996	0.972	0.952	0.912*	0.921	0.949
QS model							
MSE ₃ /MSE ₁	0.997	0.998	0.985*	0.979*	0.930**	0.941*	0.965*
MSE ₂ /MSE ₁	1.000	1.004	1.000*	1.009*	0.986**	1.003	1.000
MSE ₃ /MSE ₂	0.997	0.994	0.985*	0.971*	0.943**	0.939*	0.965*
QE model							
MSE ₃ /MSE ₁	1.013	0.965	1.058	1.073	0.957	0.824	0.983
MSE ₂ /MSE ₁	0.999*	0.943**	0.982	0.982	0.995	1.042	0.991
MSE ₃ /MSE ₂	1.013	1.023	1.078	1.093	0.962	0.791	0.992

Table 6: MSE Ratios for variance prediction

Ratios of mean squared errors (MSEs) of four-week-ahead ($h = 4$) variance forecasts generated by the QC, QS, and QE models are presented. MSE _{i} ($i = 1, 2, 3$) are computed from the following prediction errors:

$$\begin{aligned}
e_{1,t,h,\tau} &= RV_{t,t+h\Delta,\tau} - \text{var}_t[Y(X_{t+h\Delta}, \tau; \Theta_Q); \Theta_P^{TC}], \\
e_{2,t,h,\tau} &= RV_{t,t+h\Delta,\tau} - \alpha'_{1,\tau} \text{var}_t[X_{t+h\Delta}; \Theta_P^{TC}] \alpha_{1,\tau}, \\
e_{3,t,h,\tau} &= RV_{t,t+h\Delta,\tau} - \alpha'_{1,\tau} \text{var}_t[X_{t+h\Delta}; \Theta_P^T] \alpha_{1,\tau},
\end{aligned}$$

where Θ_P^{TC} and Θ_P^T are parameter vectors for the physical distribution estimated with and without the NAC, respectively. Then, MSE₃/MSE₁ measures the total impact of the NAC on variance prediction, whereas MSE₂/MSE₁ and MSE₃/MSE₂ measure cross-section-related and time-series-related components of the impact, respectively. * and ** indicate that the Diebold-Mariano test rejects the null hypothesis of equal predictive accuracy between e_i and e_j at the 5% and 1% significance levels, respectively, with standard errors computed using the Newey-West method with 8 lags. The last column displays the ratios of average MSEs with each average taken over the maturity spectrum. In-sample period is from January 4, 1991 to April 9, 2003, and the out-of-sample period is from April 16, 2003 to May 27, 2009.

Maturity	0.5	1	2	3	5	10	Ave.
Panel A: In-sample (1991/1 – 2003/4)							
QC model							
RMSE ₁	32.9	33.5	31.8	32.6	27.5	25.9	30.7
RMSE ₃	32.9	33.2	31.8	32.5	27.8	26.3	30.8
difference	0.0	0.3	0.0	0.0	−0.4	−0.3	−0.1
QS model							
RMSE ₁	32.7	33.3	32.2	33.1	28.3	26.6	31.0
RMSE ₃	32.7	32.9	32.2	33.1	28.7	27.0	31.1
difference	0.0	0.4	0.0	−0.1	−0.5	−0.4	−0.1
QE model							
RMSE ₁	34.7	34.6	32.7	33.9	28.5	27.1	31.9
RMSE ₃	34.7	34.0	32.3	33.9	29.4	28.2	32.1
difference	0.0	0.6	0.4	0.0	−0.9	−1.1	−0.2
Panel B: Out-of-sample (2003/4 – 2009/5)							
QC model							
RMSE ₁	40.0	35.5	37.6	37.6	37.6	37.6	37.7
RMSE ₃	40.1	35.4	37.3	37.0	36.0	36.9	37.1
difference	−0.1	0.1	0.4	0.6	1.7	0.7	0.6
QS model							
RMSE ₁	39.0	34.2	37.8	38.0	38.5	37.6	37.5
RMSE ₃	39.0	34.0	37.4	37.5	36.8	36.4	36.8
difference	0.1	0.2	0.4	0.5	1.6	1.2	0.7
QE model							
RMSE ₁	48.3	42.8	37.7	34.3	32.4	33.6	38.2
RMSE ₃	48.4	41.9	38.2	35.1	31.9	31.6	37.9
difference	−0.1	0.9	−0.5	−0.8	0.5	2.0	0.3

Table 7: RMSEs for volatility prediction

Root mean squared errors (RMSEs) of four-week-ahead ($h = 4$) annualized standard deviation forecasts generated by the QC, QS, and QE models are presented in basis points. RMSE _{i} ($i = 1, 3$) are computed from the following prediction errors:

$$\begin{aligned}\tilde{e}_{1,t,h,\tau} &= \sqrt{RV_{t,t+h\Delta,\tau}/h\Delta} - \sqrt{\text{var}_t[Y(X_{t+h\Delta}, \tau; \Theta_Q); \Theta_P^{TC}]/h\Delta}, \\ \tilde{e}_{3,t,h,\tau} &= \sqrt{RV_{t,t+h\Delta,\tau}/h\Delta} - \sqrt{\alpha'_{1,\tau} \text{var}_t[X_{t+h\Delta}; \Theta_P^T] \alpha_{1,\tau}/h\Delta},\end{aligned}$$

where Θ_P^{TC} and Θ_P^T are parameter vectors for the physical distribution estimated with and without the NAC, respectively. The last column displays the average RMSEs with the average taken over the maturity spectrum. In-sample period is from January 4, 1991 to April 9, 2003, and the out-of-sample period is from April 16, 2003 to May 27, 2009.

	with NAC		without NAC	
Parameters in the risk-neutral drift vector				
$\kappa_{\mu,0}$	0.000			
$\kappa_{\mu,1}$	-0.882	(0.026)		
$\kappa_{\mu,2}$	-1.845	(0.053)		
$\kappa_{V,0} \times 10^2$	0.156	(0.014)		
$\kappa_{V,3}$	-0.006	(0.002)		
Parameters in the covariance matrix				
$s_{11} \times 10^2$	0.112	(0.005)	0.112	(0.005)
$s_{22} \times 10^2$	1.776	(0.100)	1.736	(0.077)
$s_{33} \times 10^2$	0.155	(0.009)	0.167	(0.008)
$s_{12} \times 10^2$	-0.239	(0.021)	-0.235	(0.019)
$s_{13} \times 10^2$	0.012	(0.005)	0.013	(0.005)
$s_{23} \times 10^2$	0.158	(0.019)	0.163	(0.020)
c_{11}	0.000		0.000	
c_{22}	0.000		0.000	
c_{12}	0.000		0.000	
$\varsigma \times 10^4$	6.150	(0.097)		
LogL	22283		9491	
LogL _T	9490			
LogL _C	11472			

Table 8: Parameter estimates (standard errors) for the $A_1(3)$ model

The risk neutral and physical distributions of instantaneous changes in $X_t = (r_t \ \mu_t \ V_t)'$ are given by (37) and (39), respectively, where Σ_t is given by (38). The rest of the legend is the same as in Table 1.

	with NAC		without NAC	
Parameters in the risk-neutral drift vector				
$\kappa_{\mu,0}$	0.000			
$\kappa_{\mu,1}$	-0.879	(0.026)		
$\kappa_{\mu,2}$	-1.789	(0.052)		
$\kappa_{V,0} \times 10^2$	0.191	(0.016)		
$\kappa_{V,3}$	-0.013	(0.002)		
Parameters in $h_1(X_t)$				
β_0^1	-0.007	(0.001)	-0.007	(0.001)
β_1^1	-0.044	(0.008)	-0.045	(0.008)
β_2^1	-0.089	(0.011)	-0.090	(0.011)
β_3^1	0.291	(0.031)	0.292	(0.029)
Parameters in $h_2(X_t)$				
β_0^2	0.000		0.000	
β_1^2	-0.309	(0.049)	-0.304	(0.046)
β_2^2	-0.184	(0.038)	-0.181	(0.036)
β_3^2	0.790	(0.045)	0.777	(0.043)
Parameters in $h_3(X_t)$				
β_0^3	0.011	(0.001)	0.013	(0.001)
β_1^3	-0.027	(0.012)	-0.062	(0.014)
β_2^3	-0.040	(0.013)	-0.038	(0.013)
β_3^3	0.000		0.000	
Parameters in R				
ρ_{12}	-0.532	(0.037)	-0.532	(0.034)
ρ_{13}	0.098	(0.043)	0.106	(0.043)
ρ_{23}	0.299	(0.033)	0.291	(0.034)
$\varsigma \times 10^4$	6.127	(0.097)		
LogL	22360		9587	
LogL _T	9584			
LogL _C	11479			

Table 9: Parameter estimates (standard errors) for the $A_1(3)$ -based QC model

The risk neutral and physical distributions of instantaneous changes in $X_t = (r_t \mu_t V_t)'$ are given by (37) and (39), respectively, where Σ_t is decomposed as $\Sigma_t = H_t R H_t$. The rest of the legend is the same as in Table 2.

	with NAC		without NAC	
Parameters in the risk-neutral drift vector				
$\kappa_{\mu,0}$	0.000			
$\kappa_{\mu,1}$	-0.892	(0.027)		
$\kappa_{\mu,2}$	-1.843	(0.056)		
$\kappa_{V,0} \times 10^2$	0.152	(0.014)		
$\kappa_{V,3}$	-0.006	(0.002)		
Parameters in P				
$\sin \varphi_1^P$	0.143	(0.011)	0.141	(0.011)
$\sin \varphi_2^P$	-0.363	(0.042)	-0.342	(0.031)
$\sin \varphi_3^P$	0.050	(0.016)	0.065	(0.016)
Parameters in $l_1(X_t)$				
$c_0^1 \times 10^4$	1e-4		1e-4	
m_1^1	0.000		0.000	
m_2^1	0.000		0.000	
m_3^1	0.020	(0.003)	0.021	(0.002)
$\sin \varphi_1^1$	0.000		0.000	
$\sin \varphi_2^1$	0.230	(0.038)	0.243	(0.034)
$\sin \varphi_3^1$	0.324	(0.035)	0.331	(0.030)
Parameters in $l_2(X_t)$				
$c_0^2 \times 10^4$	1e-4		1e-4	(0.072)
m_1^2	0.059	(0.025)	0.067	(0.021)
m_2^2	0.705	(0.256)	0.734	(0.260)
m_3^2	1.595	(0.429)	1.602	(0.425)
$\sin \varphi_1^2$	0.462	(0.086)	0.481	(0.089)
$\sin \varphi_2^2$	0.456	(0.160)	0.487	(0.164)
$\sin \varphi_3^2$	0.684	(0.196)	0.649	(0.218)
Parameters in $l_3(X_t)$				
$c_0^3 \times 10^4$	1e-4		1e-4	
m_1^3	0.000		0.000	
m_2^3	0.032	(0.009)	0.032	(0.009)
m_3^3	0.106	(0.025)	0.160	(0.034)
$\sin \varphi_1^3$	0.289	(0.098)	0.265	(0.095)
$\sin \varphi_2^3$	0.149	(0.083)	0.183	(0.075)
$\sin \varphi_3^3$	0.871	(0.064)	0.895	(0.041)
$\varsigma \times 10^4$	6.136	(0.097)		
LogL	22356		9538	
LogL _T	9536			
LogL _C	11476			

Table 10: Parameter estimates (standard errors) for the $A_1(3)$ -based QE model

The risk neutral and physical distributions of instantaneous changes in $X_t = (r_t \mu_t V_t)'$ are given by (37) and (39), respectively, where Σ_t is decomposed as $\Sigma_t = PL_tP'$. The rest of the legend is the same as in Table 4

	$A_1(3)$ model				QC model			
	with NAC		without NAC		with NAC		without NAC	
$K_{r,0}$	0.000		0.000		0.000		0.000	
$K_{r,1}$	0.000		0.000		0.000		0.000	
$K_{r,2}$	1.077	(0.154)	1.071	(0.146)	0.980	(0.148)	0.987	(0.138)
$K_{r,3}$	-0.220	(0.050)	-0.219	(0.047)	-0.191	(0.057)	-0.191	(0.054)
$K_{\mu,0}$	0.000		0.000		0.000		0.000	
$K_{\mu,1}$	0.000		0.000		0.000		0.000	
$K_{\mu,2}$	-2.416	(0.566)	-2.357	(0.548)	-1.976	(0.538)	-1.986	(0.516)
$K_{\mu,3}$	0.300	(0.174)	0.291	(0.166)	0.198	(0.180)	0.195	(0.172)
$K_{V,0}$	0.035	(0.016)	0.037	(0.016)	0.035	(0.017)	0.035	(0.018)
$K_{V,3}$	-0.633	(0.274)	-0.664	(0.290)	-0.617	(0.278)	-0.622	(0.288)

	QE model			
	with NAC		without NAC	
$K_{r,0}$	0.000		0.000	
$K_{r,1}$	0.000		0.000	
$K_{r,2}$	0.994	(0.149)	0.999	(0.139)
$K_{r,3}$	-0.193	(0.058)	-0.194	(0.054)
$K_{\mu,0}$	0.000		0.000	
$K_{\mu,1}$	0.000		0.000	
$K_{\mu,2}$	-2.013	(0.524)	-1.976	(0.501)
$K_{\mu,3}$	0.199	(0.175)	0.194	(0.167)
$K_{V,0}$	0.033	(0.017)	0.033	(0.017)
$K_{V,3}$	-0.599	(0.269)	-0.599	(0.277)

Table 11: Parameter estimates (standard errors) in the physical drift for the $A_1(3)$ and extended models

The specification of the physical drift is given by (39) for all models. The same parameters are estimated between the $A_1(3)$ and extended models in order to control for potential influences from the physical drift. In-sample data from January 4, 1991 to April 9, 2003 are used for the estimation (641 observations).

Maturity	0.5	1	2	3	5	10	Ave.
Panel A: In-sample (1991/1 – 2003/4)							
<i>A</i> ₁ (3) model							
MSE ₃ /MSE ₁	1.000	1.001	1.000	1.000	1.013	1.014	1.004
MSE ₂ /MSE ₁	1.000	1.001	1.000	1.000	1.003	1.000	1.001
MSE ₃ /MSE ₂	1.000	1.000	1.000	1.000	1.010	1.014	1.003
QC model							
MSE ₃ /MSE ₁	0.998	0.995	0.993	0.993	0.990	0.992	0.993
MSE ₂ /MSE ₁	1.000	0.996	1.000	0.999	1.007	1.000	1.000
MSE ₃ /MSE ₂	0.998	0.999	0.993	0.994	0.983	0.992	0.993
QE model							
MSE ₃ /MSE ₁	1.005	1.004	1.016	1.020*	1.121**	1.140**	1.053**
MSE ₂ /MSE ₁	1.000	0.994	0.998	0.989	1.016*	0.979	0.995
MSE ₃ /MSE ₂	1.005	1.010*	1.018	1.031*	1.103**	1.164**	1.058**
Panel B: Out-of-sample (2003/4 – 2009/5)							
<i>A</i> ₁ (3) model							
MSE ₃ /MSE ₁	1.000	1.002	0.998*	1.000	0.976**	0.979**	0.989**
MSE ₂ /MSE ₁	1.000	1.001	1.000*	1.006**	0.992**	1.000**	1.000
MSE ₃ /MSE ₂	1.000	1.000	0.998*	0.994**	0.984**	0.979**	0.990**
QC model							
MSE ₃ /MSE ₁	1.000	1.007	1.001	0.998	0.936**	0.943	0.975*
MSE ₂ /MSE ₁	1.000	1.004	1.000**	1.012**	0.987**	1.003*	1.001
MSE ₃ /MSE ₂	1.000	1.003	1.001	0.987*	0.949**	0.940	0.975*
QE model							
MSE ₃ /MSE ₁	0.980	1.008	1.037	1.039	1.031	1.015	1.025
MSE ₂ /MSE ₁	1.000	0.982	0.995	0.987	0.983	0.997	0.990
MSE ₃ /MSE ₂	0.980	1.027	1.042	1.053	1.049	1.019	1.035

Table 12: MSE Ratios for variance prediction generated by the $A_1(3)$ and extended models

The same note applies as in Table 6.

Maturity	0.5	1	2	3	5	10	Ave.
Panel A: In-sample (1991/1 – 2003/4)							
<i>A</i> ₁ (3) model							
RMSE ₁	36.9	36.7	33.4	34.4	30.3	28.6	33.4
RMSE ₃	36.9	36.4	33.4	34.4	30.8	29.1	33.5
difference	0.0	0.3	0.0	0.0	−0.5	−0.5	−0.1
QC model							
RMSE ₁	32.2	32.7	32.5	33.4	28.3	26.1	30.9
RMSE ₃	32.1	32.3	32.3	33.2	28.3	26.2	30.7
difference	0.1	0.4	0.2	0.2	0.0	−0.1	0.2
QE model							
RMSE ₁	33.0	32.8	32.2	35.1	32.4	33.4	33.2
RMSE ₃	33.1	32.6	32.5	35.4	33.8	35.0	33.7
difference	0.0	0.2	−0.2	−0.3	−1.4	−1.7	−0.6
Panel B: Out-of-sample (2003/4 – 2009/5)							
<i>A</i> ₁ (3) model							
RMSE ₁	44.0	37.1	44.2	46.8	50.5	49.9	45.4
RMSE ₃	44.0	37.0	44.1	46.8	49.6	49.1	45.1
difference	0.0	0.2	0.0	0.0	0.9	0.8	0.3
QC model							
RMSE ₁	39.1	35.4	40.5	40.3	39.9	38.2	38.9
RMSE ₃	39.1	35.5	40.6	40.2	38.2	37.2	38.5
difference	0.0	−0.1	0.0	0.0	1.7	1.0	0.4
QE model							
RMSE ₁	38.5	33.5	36.1	35.9	35.7	34.2	35.6
RMSE ₃	38.8	33.6	36.6	36.1	35.1	33.7	35.6
difference	−0.2	−0.1	−0.4	−0.3	0.6	0.5	0.0

Table 13: RMSEs for volatility prediction generated by the $A_1(3)$ and extended models

The same note applies as in Table 7.

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

Panel C: the third-step comparison		r	θ	ϵ	y_1	y_3	y_5
Minimum	Level	0.06	-4.95	5.13	0.02	-0.13	-0.38
	Slope	0.72	-1.26	0.21	-0.16	0.18	0.22
	Curvature	2.11	-2.27	-8.23	-0.47	0.53	0.65
Median	Level	0.51	-0.86	-7.15	-0.10	0.06	-0.10
	Slope	-0.37	2.81	-4.02	0.07	-0.04	-0.01
	Curvature	0.03	-1.76	-1.90	0.01	-0.07	-0.25
Maximum	Level	1.26	3.15	-13.47	-0.31	0.40	0.53
	Slope	0.02	-1.89	-8.47	0.03	-0.17	-0.63
	Curvature	-0.79	3.01	-1.47	0.16	-0.13	-0.11

Table B1: Comparison with the closed-form solution using the Gaussian model

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

Maturity		0.5	1	2	3	5	10
Panel A: QS model							
Minimum	Level	0.12	0.10	0.09	0.10	0.08	0.29
	Slope	0.21	0.18	0.13	0.14	0.11	0.45
	Curvature	0.39	0.32	0.14	0.07	0.03	0.39
Median	Level	-0.04	-0.05	-0.07	-0.02	0.04	0.78
	Slope	-0.18	-0.17	-0.15	-0.08	0.02	0.61
	Curvature	-0.04	-0.06	-0.09	-0.12	-0.15	0.31
Maximum	Level	0.02	0.03	0.06	0.02	-0.05	1.20
	Slope	-0.18	-0.17	-0.06	0.05	0.13	1.81
	Curvature	-0.27	-0.25	-0.23	-0.28	-0.24	0.45
Panel B: QE model							
Minimum	Level	0.10	0.08	0.09	0.18	0.31	0.42
	Slope	0.19	0.23	-0.09	-0.14	-0.11	0.19
	Curvature	0.37	0.20	-0.26	-0.82	-1.45	-1.07
Median	Level	-0.03	-0.03	-0.03	-0.04	0.10	1.54
	Slope	-0.21	-0.14	-0.14	-0.11	-0.06	0.39
	Curvature	-0.04	-0.04	-0.07	-0.10	-0.10	0.57
Maximum	Level	0.03	0.04	0.04	0.01	-0.27	0.80
	Slope	-0.17	-0.11	-0.05	-0.05	0.04	2.02
	Curvature	-0.27	-0.34	-0.29	-0.25	-0.26	-0.02

Table B2: Comparison with the MC solution using the Gaussian-based QS and QE models

Approximation errors, defined as the differences in yields between the approximate and MC solutions, are presented in bps. The errors are evaluated at nine states taken from the actual data from January 4, 1991 to May 27, 2009, where the level, slope, and curvature factors take the minimum, median, or maximum value. Panels A and B present the results for the Gaussian-based QS and QE models with the parameter values given in Tables 3 and 4, respectively.

Maturity		0.5	1	2	3	5	10
Panel A: QS model							
Minimum	Level	0.08	0.09	0.09	0.08	0.07	0.05
	Slope	0.35	0.22	0.10	0.00	-0.13	-0.22
	Curvature	-0.31	-0.34	-0.26	-0.13	0.08	0.25
Median	Level	0.65	0.71	0.71	0.63	0.49	0.36
	Slope	-0.35	-0.78	-0.95	-0.87	-0.68	-0.50
	Curvature	0.44	0.60	0.52	0.33	0.07	-0.14
Maximum	Level	0.21	-0.53	-0.71	-0.45	0.08	0.60
	Slope	-0.36	-0.63	-0.81	-0.87	-0.90	-0.91
	Curvature	-0.29	-0.36	-0.45	-0.46	-0.43	-0.40
Panel B: QE model							
Minimum	Level	-0.54	-0.75	-1.01	-1.10	-1.14	-1.16
	Slope	0.15	-0.18	-0.52	-0.61	-0.59	-0.54
	Curvature	1.12	0.57	0.06	-0.02	0.09	0.37
Median	Level	0.07	0.02	0.09	0.29	0.63	0.93
	Slope	0.23	0.22	0.21	0.25	0.33	0.41
	Curvature	-0.13	0.12	0.29	0.32	0.32	0.30
Maximum	Level	0.55	0.78	0.89	0.77	0.50	0.22
	Slope	0.38	-0.08	-0.40	-0.35	-0.07	0.27
	Curvature	-0.16	-0.57	-0.83	-0.68	-0.27	0.15

Table B3: Accuracy to 4-week annualized standard deviation using the Gaussian-based QS and QE models

Approximation errors, defined as the differences in annualized standard deviations between the approximate and MC solutions, are presented in bps. The errors are evaluated at nine states taken from the actual data from January 4, 1991 to May 27, 2009, where the level, slope, and curvature factors take the minimum, median, or maximum value. Panels A and B present the results for the Gaussian-based QS and QE models with the parameter values given in Tables 3 and 4, respectively.

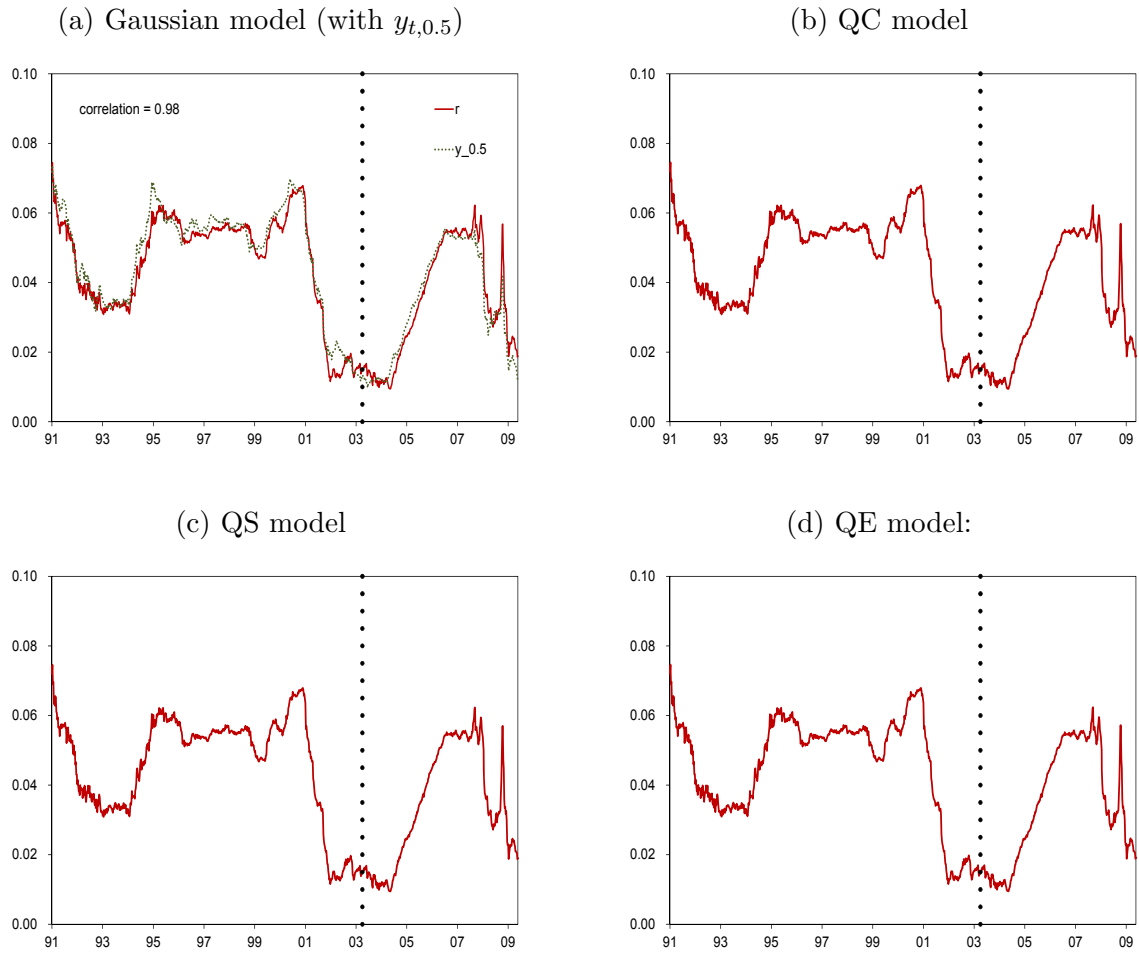


Figure 1: Time series of r_t

In panel (a), time-series of the six-month yield is also displayed. The vertical dotted line separates the in-sample and out-of-sample periods.

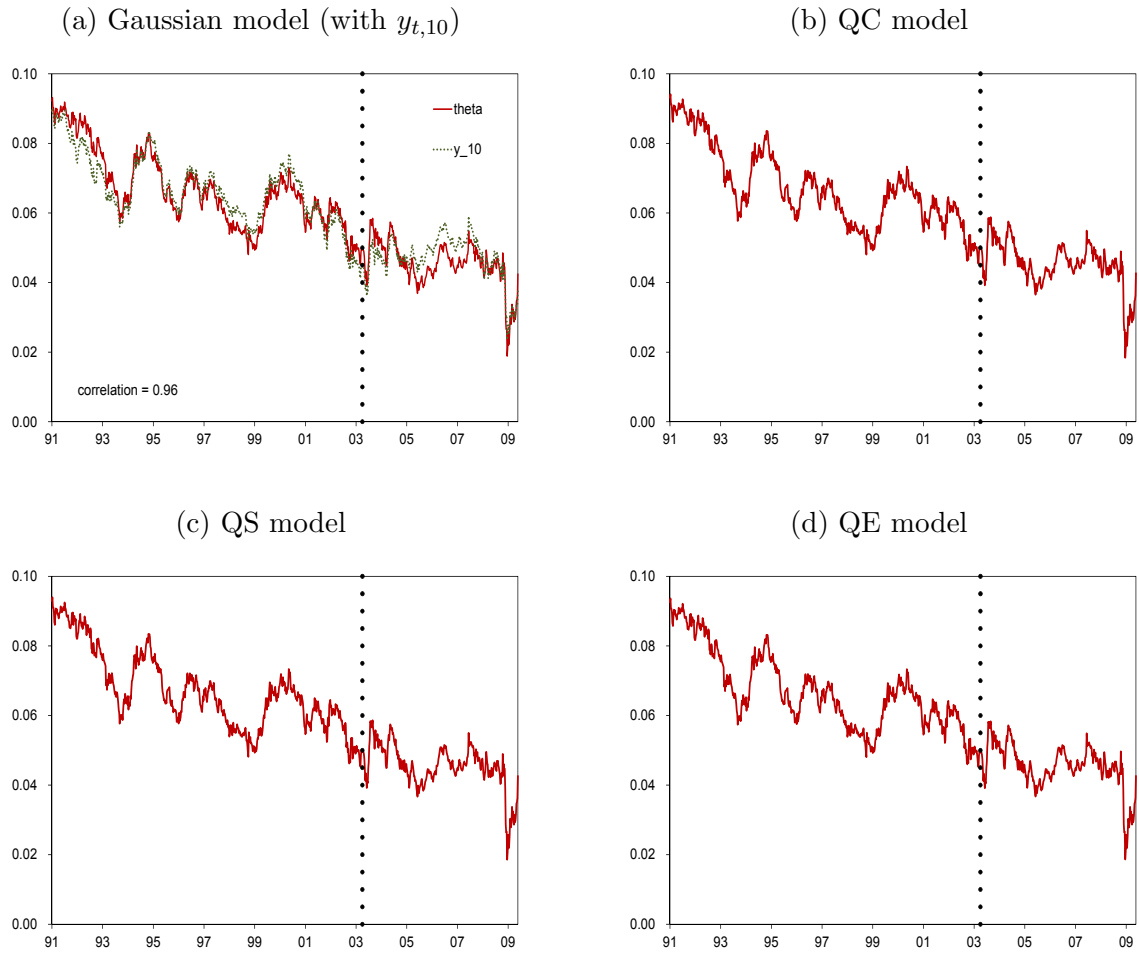


Figure 2: Time series of θ_t

In panel (a), time-series of the ten-year yield is also displayed. The vertical dotted line separates the in-sample and out-of-sample periods.

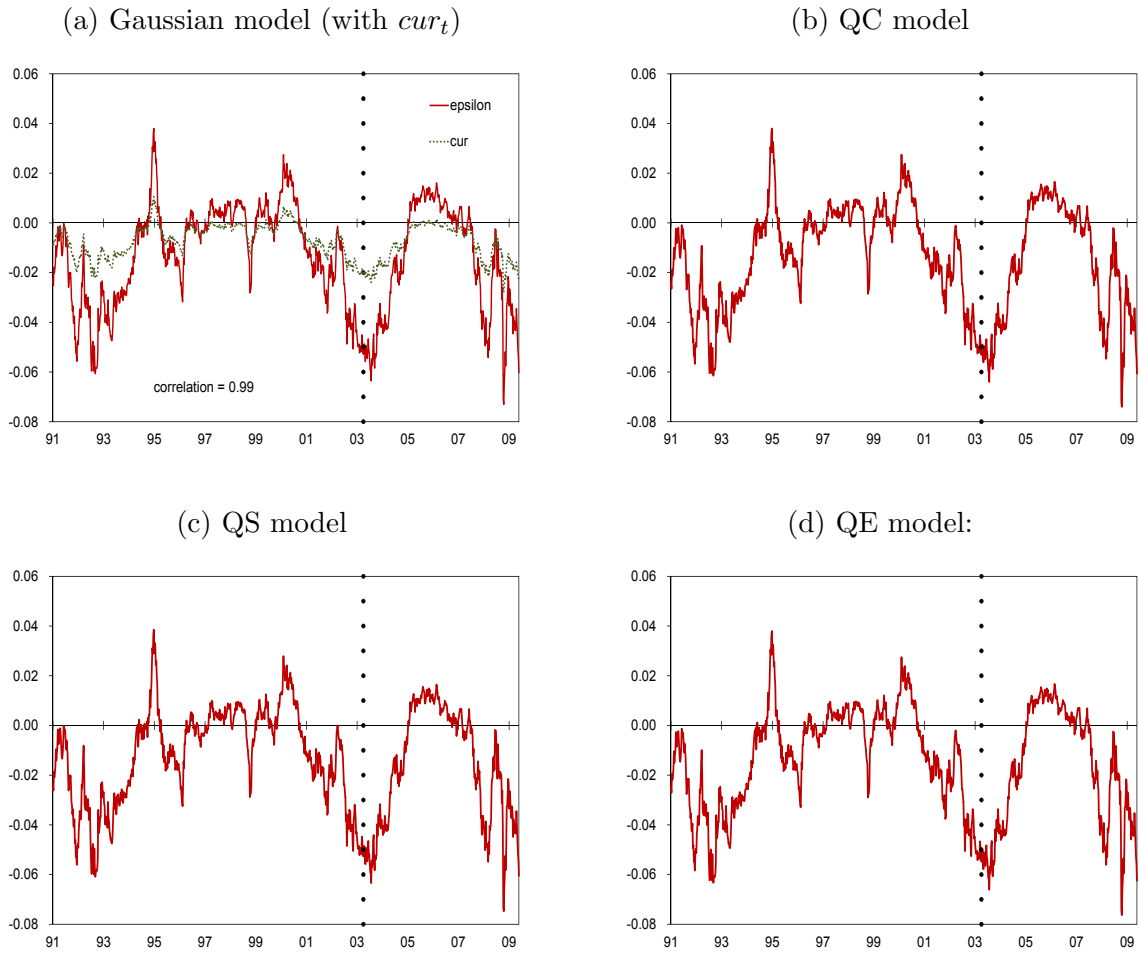
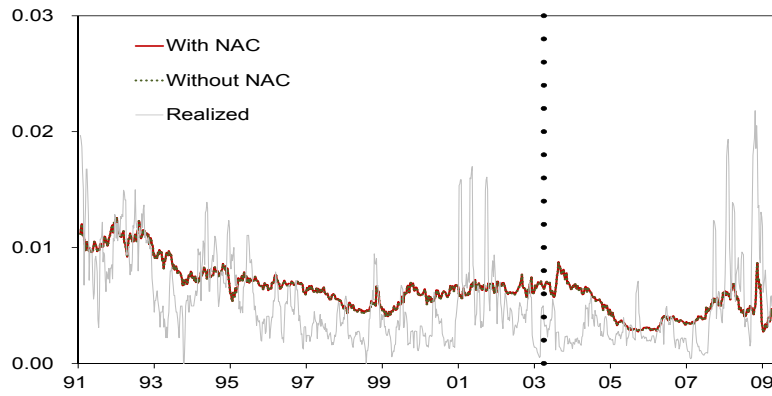


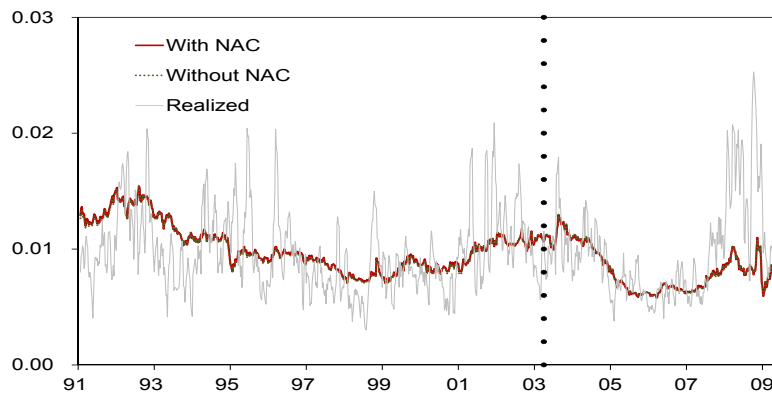
Figure 3: Time series of ϵ_t

In panel (a), time-series of a proxy for the curvature factor given by $cur_t = 2y_{t,2} - (y_{t,0.5} + y_{t,10})$ is also displayed. The vertical dotted line separates the in-sample and out-of-sample periods.

(a) $\tau = 0.5$



(b) $\tau = 2$



(c) $\tau = 10$

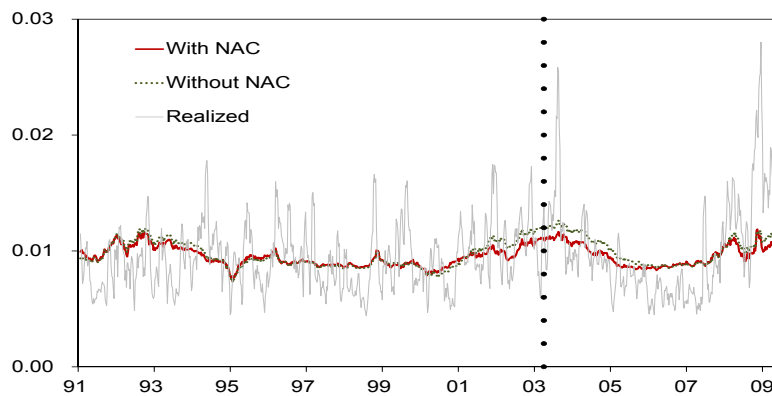
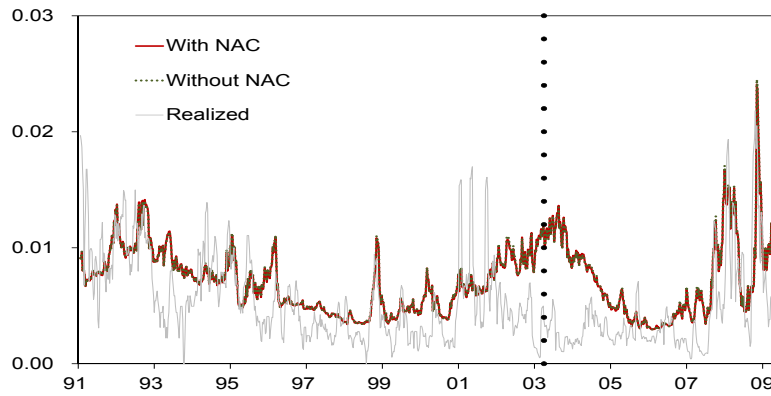


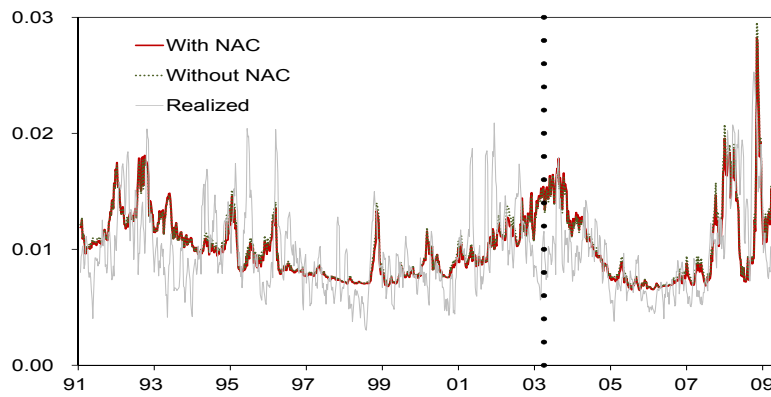
Figure 4: Time series of 4-week-ahead forecasts of τ -year yield volatility generated by the QC model

The bold and dotted lines correspond to forecasts with and without NAC, and the thin line to the realized value. The vertical dotted line separates the in-sample and out-of-sample periods.

(a) $\tau = 0.5$



(b) $\tau = 2$



(c) $\tau = 10$

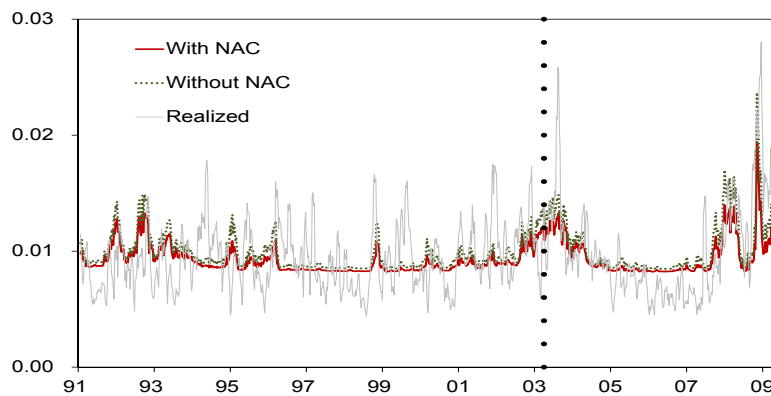


Figure 5: Time series of 4-week-ahead forecasts of τ -year yield volatility generated by the QE model

The bold and dotted lines correspond to forecasts with and without NAC, and the thin line to the realized value. The vertical dotted line separates the in-sample and out-of-sample periods.

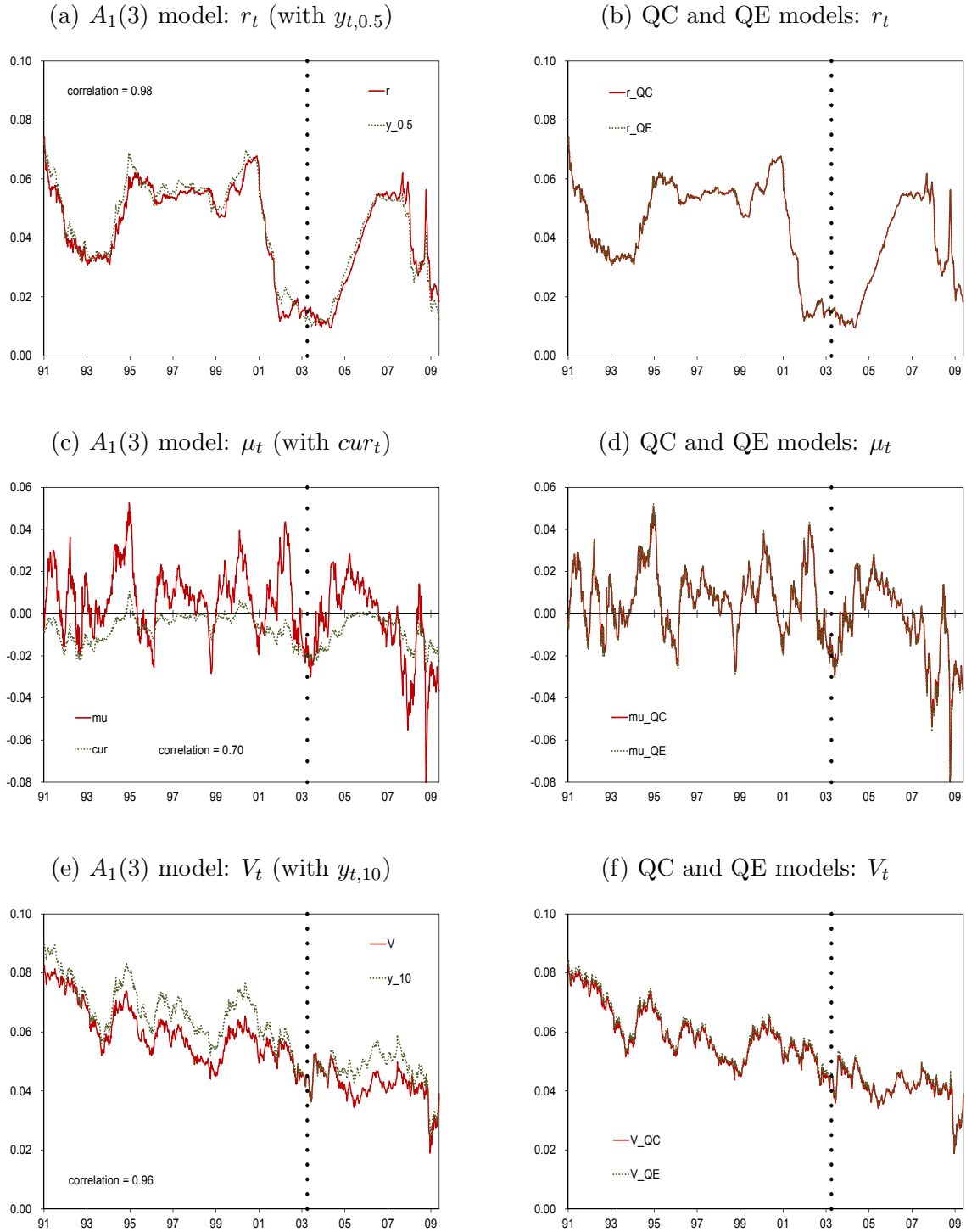
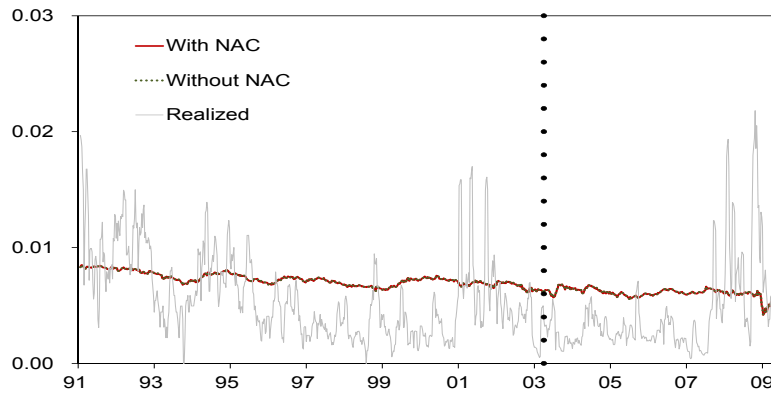
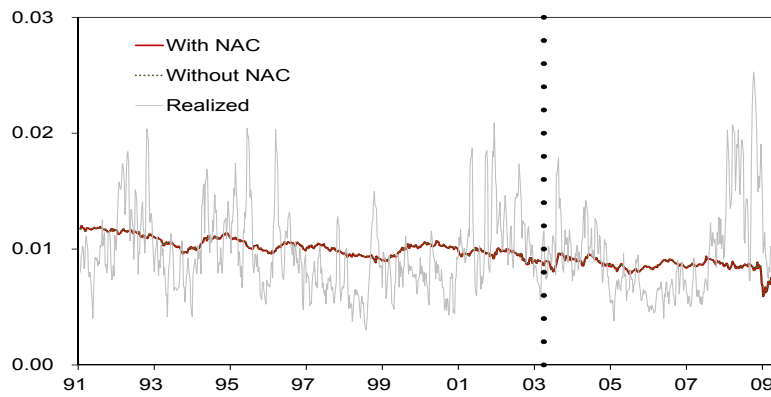


Figure 6: Time series of factors extracted through the $A_1(3)$ and extended models
The left panels present the time-series of r_t , μ_t , and V_t extracted through the $A_1(3)$ model together with the time-series of observed variables that track those of the factors. The corresponding factor time-series for the QC and QE models are displayed together in the right panels. The vertical dotted line separates the in-sample and out-of-sample periods.

(a) $\tau = 0.5$



(b) $\tau = 2$



(c) $\tau = 10$

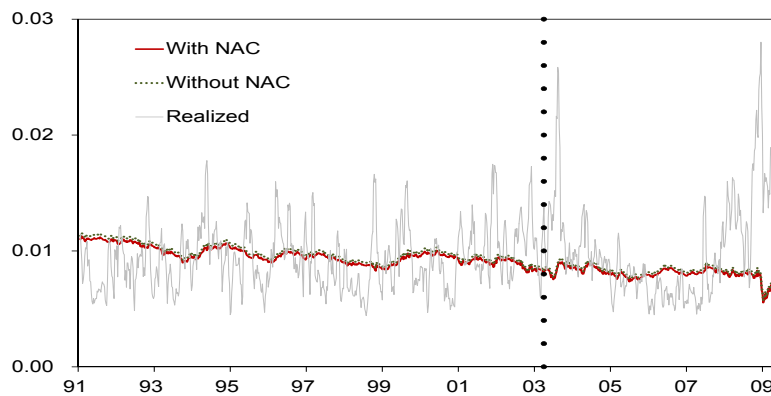
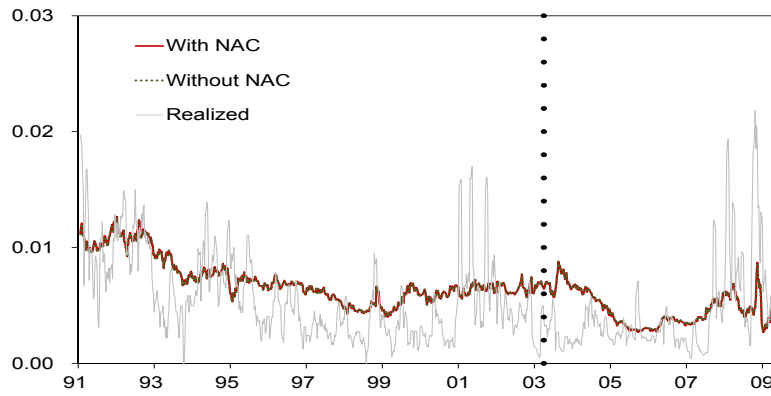
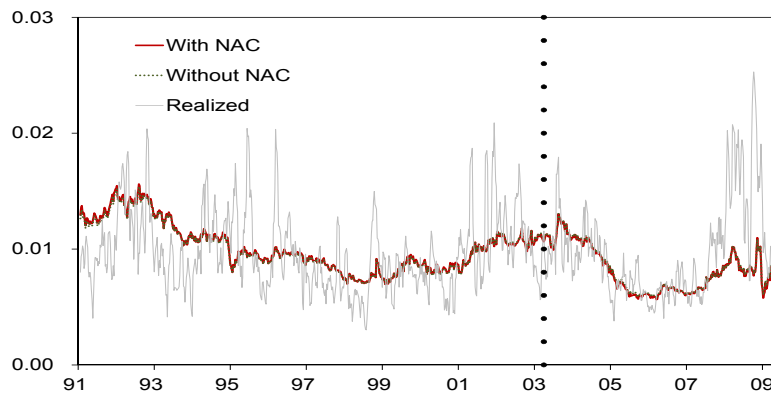


Figure 7: Time series of 4-week-ahead forecasts of τ -year yield volatility generated by the $A_1(3)$ model

(a) $\tau = 0.5$



(b) $\tau = 2$



(c) $\tau = 10$

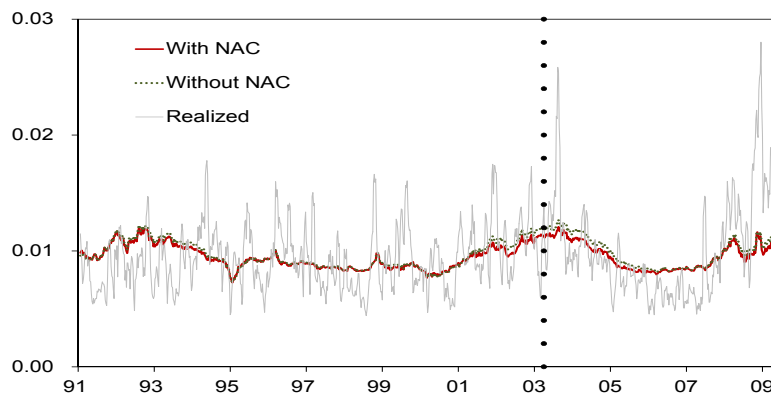
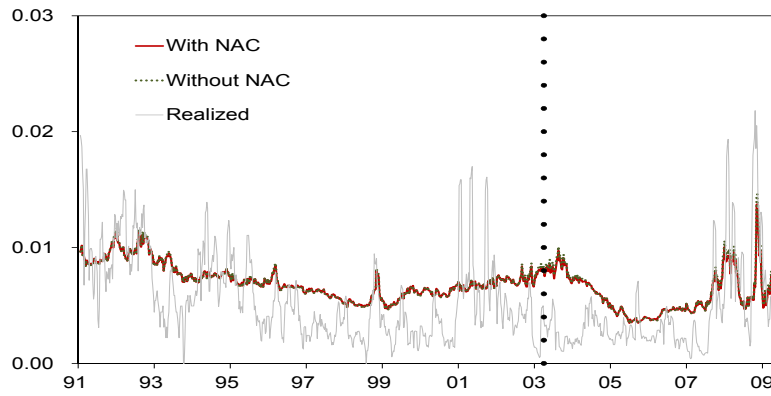
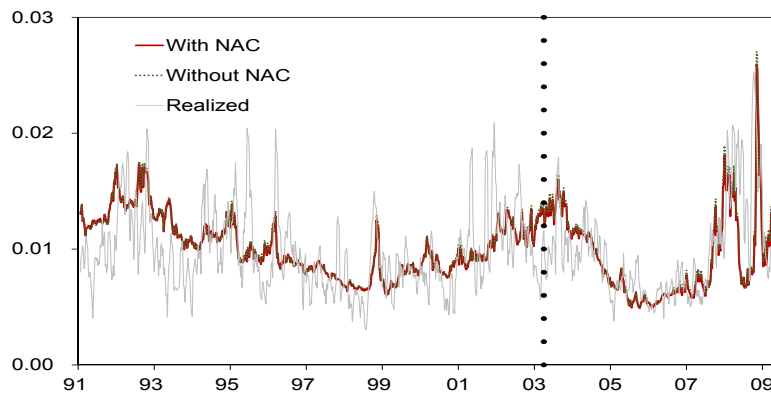


Figure 8: Time series of 4-week-ahead forecasts of τ -year yield volatility generated by the $A_1(3)$ -based QC model

(a) $\tau = 0.5$



(b) $\tau = 2$



(c) $\tau = 10$

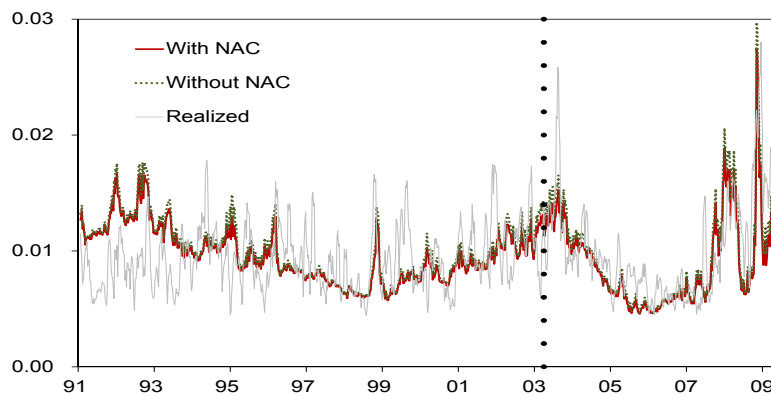


Figure 9: Time series of 4-week-ahead forecasts of τ -year yield volatility generated by the $A_1(3)$ -based QE model