Forecast Combination of Affine and Quadratic Term Structure Models near the Zero Lower Bound*

Tsz-Kin Chung† Hirokuni Iiboshi‡
Shingo Mukunoki§ Kosuke Oya¶

August 7, 2017

Abstract

In this paper, we study the density forecasts of the affine term structure model (ATSM) and the quadratic term structure model (QTSM) under the zero and negative interest rate policy of Japan, besides considering macro-finance features. As both the two models can be potentially misspecified, we adopt the optimal pooling prediction scheme following Geweke and Amisano (2011). We find that the QTSM provides a more realistic statistical description when bond yields are close to the zero lower bound. The ATSM gives a good fit to the macroeconomic variables and bond yields simultaneously, however, it predicts a large probability of negative interest rates and hence is not appropriate for the forecasting of bond yields. One should use a forecast combination of the two models for the prediction of future bond yields during different time periods. Furthermore, we expand it in order to examine the combination of dataset and models. According to the results, the ATSM with the macro-finance feature defeats for forecasting of short term rate, whereas the QTSM with the financial factors is superior for the long terms rate.

Keywords: Term structure; Density Forecasts; Financial markets and the macroeconomy; Optimal pool; Dynamic prediction pool; Markov-switching mixture; Bayesian estimation

JEL Classifications: C52, E43, E44,

*Acknowledgment: The authors would like to thank Keiichi Tanaka for useful comments and discussion, and appreciate comments from Masaaki Maruyama, Tastuyoshi Matsumae, Daisuke Nakamura, Kosuke Oya, Hisashi Tanizaki, Motostugu Shintani, and participants at the seminar held at Osaka Univ. and Economic and Social Research Institute, Cabinet Office. The views expressed herein are of our own and do not represent those of the organizations the authors belongs to.

†Graduate School of Social Sciences, Tokyo Metropolitan University
‡Graduate School of Social Sciences, Tokyo Metropolitan University
§Graduate School of Economics, Osaka University
¶Graduate School of Economics, Osaka University

1 Introduction

The Gaussian affine term structure model (ATSM) has been a popular choice in the modeling of yield curve given its analytical tractable bond pricing formula as well as the linear dependence of the model-implied bond yields to the underlying factors or state variables (Piazzesi, 2010). The model allows one to summarize the complex movements of bond yields into a small number of factors while imposing the no-arbitrage restrictions among bond yields with different maturities. Traditionally, these factors are regarded as *latent* and are usually related to the first three principal components of bond yields, including the level, slope and curvature factors. From an economic perspective, bond yields should interact closely with the macroeconomy and it is very tempting to relate these factors driving bond yields to various macroeconomic variables such as measures of inflation, real activity and monetary stance. This exercise of linking bond yields to macroeconomic variables, called the macro-finance term structure modeling, allows researchers to explain the movements in bond yields with a richer economic interpretation and potentially improve the prediction of future bond yields by incorporating information beyond the bond market.

There have been large amount of literature that explore the role of macroeconomic variables in the arbitrage-free term structure modeling. Ang and Piazzesi (2003) employ two measures of inflation and real activity and find that these macroeconomic factors explain up to 85% of the time-series variation of bond yields. Diebold et al. (2006) study the dynamic interaction between the macroeconomy and the yield curve. They find that macro variables strongly affect future movements in the yield curve with a feedback from the yield curve to the macroeconomy. Ang et al. (2006) explore the Taylor rule interpretation of a macro-finance model by taking the inflation rate and output gap as the state variables. Li et al. (2012) extend the idea to model a time-varying Taylor rule by incorporating regime-dependent policy response coefficients. Diebold and Rudebusch (2013) provide a succinct summary on the recent development in term structure modeling with macro-finance features.

Despite its popularity in the macro-finance literature, there is one major shortcoming of the Gaussian ATSM: its normal assumption predicts a large probability of negative interest rates. It is noted that another class of ATSM built on the square root process (Cox et al., 1985) is not suitable for macro-finance modeling because the state variables are positive by construction. This is in contrast to the fact that most of the macroeconomic variables can take negative values (e.g., inflation and output gap). This may be problematic for the prediction of future bond yields when interest rates are very close to the zero lower bound or negative interest rates, such as the cases of the Japanese government bond (JGB) yields since 1995 and the US treasury yields after the financial crisis of 2008, plus the recent emergence of negative interest rates in Japan and the euro zone. Against this background, an alternative is the quadratic term structure model (QTSM) as advocated by Ahn et al. (2002) and Leippold and Wu (2002), which allows one to has the extra flexibility to control the range of future interest

rates by setting the short rate equation.

In this paper, we evaluate forecasting performance of factors including macroeconomic variables for term structure of the zero coupon bond yields by building forecast combination of the ATSM and QTSM. To do so, we adopt the pooling prediction of the future bond yields (term structure) of the Gaussian ATSM and QTSM with macro-finance feature in which real GDP gap or inflation incorporates into it as a factor. The contribution is three-fold. Firstly, we compare the empirical performance of the two macro-finance term structure models under the zero interest rate policy in the JGB market. To our knowledge, this is the first study to compare the ATSM and the QTSM under the macro-finance setting across an extended historical period including close-to-ZLB period and negative interest rate period. Secondly, we attempt to derive a better forecast of the term structure by combining the advantages of the two models on hand. Thirdly, we expand the forecast combination of the ATSM and the QTSM to that of six models including three dataset characterized by sorts of factors. And we consider what combination of factors and terms structure model realize best forecast.

We adopt two different novel approaches, with increasing complexity, in modeling the weighting coefficient that pools the bond yield prediction densities of the two models. In particular, the later approaches, say dynamic prediction pool, with time-varying weighting coefficient allows us to investigate the relative goodness in forecasting of the ATSM and QTSM during different sample periods (Del Negro et al., 2016).

Our estimation results show that the QTSM dominates in its forecasting performance when interest rates are close to zero, while the ATSM provides a better fitting of the bond yields and macro factors simultaneously. It is worth to note that the ATSM predicts negative interest rate with almost 40% to 50% of the probability when the JGB yields are close to zero since late 1995. In addition, when moving to consider combination of sorts of factors and sort of terms structure models, our results also show that the ATSM with macro-finance feature is strong for forecasting of short term, while the QTSMs with both spread and real GDP gap are superior for the long terms rate. In other word, macroeconomic variables, i.e., GDP gap and inflation, include useful information on prediction of short term rate in spite of the presence or absence of zero lower bound of interest rate. In contrast, the second moment of the factor, especially curvature, included in the QTSM plays an important role on prediction of long term interest rate.

The paper is structured as follows. Section 2 presents the ATSM and QTSM with factors deciding variations of yields including macro-finance features and discusses the data and estimation procedure. In Section 3, we mention about empirical results of both term structure model. Section 4 reviews forecast combination in terms of the methods of prediction pools. Section 5 reports the estimation results and findings of forecast combination. Section 6 expand to forecast combination of six models in order to consider what factor should be used for forecasting. Section 7 concludes.

2 Macro-Finance Models

2.1 Setup

In this paper, we adopt a discrete time setting for the macro-finance term structure modeling. All the data used in this paper are quarterly and hence we can interpret one period to be one quarter. The key ingredient in the macro-finance term stucture modeling is the linkage between the short-rate r_t and the Gaussian state vector X_t taking values in $\mathbb{R}^M$ as

$$\begin{aligned} r_t &= \phi(X_t), \\ X_{t+1} &= \mu^Q + \Phi^Q X_t + \Sigma\varepsilon_{t+1}, \end{aligned}$$

with $\varepsilon_t \sim N(\mathbf{0}, \mathbf{I}_{M\times M})$, μ^Q is a $M \times 1$ vector and Φ^Q is a $M \times M$ matrix. The notation Q denotes the risk-neutral probability measure. Without much loss of generality, we can specify the market price of risk as

$$\lambda_t = \lambda_0 + \lambda_1 X_t,$$

where λ_0 is a $M \times 1$ vector and λ_1 is a $M \times M$ matrix. Hence, the real-world dynamics of the state vector is given by

$$X_{t+1} = \mu^P + \Phi^P X_t + \Sigma\varepsilon_{t+1},$$

with

$$\mu^Q = \mu^P - \Sigma\lambda_0, \qquad \Phi^Q = \Phi^P - \Sigma\lambda_1,$$

where P denotes the real-world measure (Wright, 2011; Ang et al., 2011). The corresponding pricing kernel has the form

$$\xi_{t+1} = \exp\left(-r_t + \frac{1}{2}\lambda_t^T\lambda_t - \lambda_t^T\varepsilon_{t+1}\right)\xi_t,$$

and the time$-t$ price of a n-period zero-coupon bond can be formulated as

$$P_t^n = \mathbb{E}_t^P\left[\xi_{t+1}P_t^{n-1}\right] = \mathbb{E}_t^Q\left[\exp\left(-\sum_{i=0}^{n-1} r_{t+i}\right)\right].$$

We can also compute the n-period bond yield as

$$y_t^n = -\frac{1}{n}\log P_t^n.$$

Under the ATSM or the QTSM specification of the short rate function $r_t = \phi(X_t)$, it is possible to derive the bond pricing formula in terms of a recursive relationship. In continuous-time modeling, this corresponds to a system of ordinary differential equation that determines the bond prices. We refer the readers to Piazzesi (2010) for continuous-time affine model and Ahn

et al. (2002) for continuous-time quadratic Gaussian model.

2.2 Affine term structure model

The Gaussian ATSM is specified as

$$r_t = \delta_0 + \delta_1^T X_t, \tag{1}$$

i.e., the one-period short rate is a linear function to the selected macroeconomic state variables. As the state variable X_t is Gaussian, there is no guarantee that the short rate is non-negative.

The bond pricing formula follows from Duffie and Kan (1996) as

$$P_t^n = \exp\left(A_n + B_n^T X_t\right), \tag{2}$$

where A_n is a scalar and B_n is a $M \times 1$ vector satisfying the recursive relationship

$$\begin{aligned} A_n &= -\delta_0 + A_{n-1} + B_{n-1}^T \mu^Q + \frac{1}{2} B_{n-1} \Sigma \Sigma^T B_{n-1}^T, \\ B_n^T &= -\delta_1^T + \Phi^Q B_{n-1}^T, \end{aligned} \tag{3}$$

for $n = 1, 2, ..., N$ with $A_1 = -\delta_0$ and $B_1 = -\delta_1$. As a result, the model-implied bond yield is a linear function to the state variable X_t as

$$y_t^n = -\frac{1}{n} \log P_t^n = a_n + b_n^T X_t, \tag{4}$$

by taking $a_n = -A_n/n$ and $b_n = -B_n/n$ as the factor loadings.

Finally, we discuss about what factors are used as state variables X_t. A standard ATSM consists of three fundamental factors of yields such as level, slope and curvature, i.e., $X_t = (f_t, s_t, c_t)$, whose definitions are explained in Section 2.5. However, according to many empirical studies using a principal component approach (e.g., Litterman and Scheinkman, 1991), both factors of the level and the slope must account for more than 90-95% of variations of yields, whereas the explaining power of the curvature is less than 4%. Accordingly, these factors including the third factor are evaluated in Section 6. Instead, we replace the curvature with GDP gap, g_t, as the third factor, i.e., $X_t = (f_t, s_t, g_t)$, and consider the forecast combination of the ATSM and the QTSM based on these factors from the next section to Section 5. In section 6, since we turn to consider what factors should be used, we also introduce the other three factors based on an macro-finance approach. A typical example in the macro-finance ATSM is to choose the base interest rate f_t, output gap g_t and inflation rate π_t as the state variables such that $X_t = (f_t, g_t, \pi_t)$ and

$$r_t = \delta_0 + \delta_{1,1} f_t + \delta_{1,2} g_t + \delta_{1,3} \pi_t, \tag{5}$$

in which $\delta_{1,2}$ and $\delta_{1,3}$ can consist of the coefficients of the policy reaction function under the Taylor rule when we take

$$r_t = \alpha + \beta(\pi_t - \pi^*) + \gamma g_t + (f_t - \pi_t),$$

where π^* denotes inflation target, β is requested to over one to keep an economy stable according to the Taylor principle, and $(f_t - \pi_t)$ represents real rate obtained from physical capital. Accordingly, we have $\delta_0 = \alpha - \beta\pi^*$, $\delta_{1,1} = 1$, $\delta_{1,2} = \gamma$, and $\delta_{1,3} = \beta - 1 > 0$ based on the contemporary theory of monetary policy.

2.3 Quadratic term structure model

For the general QTSM, the short-rate function is specified as

$$r_t = \alpha_0 + \beta_0^T X_t + X_t^T \Psi_0 X_t, \tag{6}$$

i.e., the one-period short rate is a quadratic function to the selected macroeconomic state variables.

The n-period zero coupon bond price can be formulated as

$$P_t^n = \exp\left(A_n + B_n^T X_t + X_t^T C_n X_t\right), \tag{7}$$

where A_n is a scalar, B_n is a $M \times 1$ vector and C_n is a $M \times M$ matrix satisfying the recursive relationship

$$\begin{aligned}
A_n &= -\alpha_0 + A_{n-1} + B_{n-1}^T \mu^Q + \mu^T C_{n-1} \mu^Q - \frac{1}{2} \det\left(\mathbf{I} - 2\Sigma^T C_{n-1} \Sigma\right) \\
&\quad + \frac{1}{2}\left(\Sigma^T B_{n-1} + 2\Sigma^T C_{n-1} \mu^Q\right)^T \left(\mathbf{I} - 2\Sigma^T C_{n-1}\Sigma\right)^{-1} \left(\Sigma^T B_{n-1} + 2\Sigma^T C_{n-1}\mu\right), \\
B_n^T &= -\beta_0^T + B_{n-1}^T \Phi^Q + 2\mu C_{n-1} \Phi^Q \\
&\quad + 2\left(\Sigma^T B_{n-1} + 2\Sigma^T C_{n-1}\mu\right)^T \left(\mathbf{I} - 2\Sigma^T C_{n-1}\Sigma\right)^{-1} \Sigma^T C_{n-1} \Phi^Q, \\
C_n &= -\Psi_0 + \left(\Phi^Q\right)^T C_{n-1} \Phi^Q + 2\left(\Sigma^T C_{n-1} \Phi^Q\right)^T \left(\mathbf{I} - 2\Sigma^T C_{n-1}\Sigma\right)^{-1} \left(\Sigma^T C_{n-1}\Phi^Q\right),
\end{aligned} \tag{8}$$

for $n = 1, 2, ..., N$ with $A_1 = -\alpha_0$, $B_1 = -\beta_0$ and $C_1 = -\Psi_0$. As a result, the model-implied bond yield can be expressed as

$$y_t^n = -\frac{1}{n} \log P_t^n = a_n + b_n^T X_t + X_t^T c_n X_t \tag{9}$$

by taking $a_n = -A_n/n$, $b_n = -B_n/n$ and $c_n = -C_n/n$ as the factor loadings.

2.4 Estimation method

Given the bond pricing formula that relates the model-implied bond yields to the selected macro variables, we can formulate our estimation procedure in terms of a non-linear state-space model as follows:

- **Measurement Equation**

  The measurement equation describes the evolution of the observed bond yields $\hat{y}_t^n$ as

  $$\hat{y}_t^n = a_n + b_n^T X_t + X_t^T c_n X_t + \omega_{n,t}, \tag{10}$$

  with $n = 1, 2, ..., N$ and $\omega_{n,t}$ are the measurement errors which are i.i.d. normals. Moreover, we assume the selected macro variables are observed with measurement errors $\omega_{X,t}$:

  $$\hat{X}_t = X_t + \omega_{X,t}, \tag{11}$$

  where $\hat{X}_t$ is the observed macro variables and $\omega_{X,t}$ are i.i.d. normals.

- **State Equation**

  The state equation is given by the evolution of the latent state vector X_t under the real-world measure P as

  $$X_{t+1} = \mu^P + \Phi^P X_t + \Sigma \varepsilon_{t+1}, \tag{12}$$

  which is a standard VAR(1) system.

Therefore, (10), (11) and (12) together form a non-linearity state space model with 9 observables (6 observed bond yields and 3 macro variables) and 3 latent factors. The Appendix present the Bayesian MCMC method to estimate the model parameters. To this end, it is important to calibrate the size of the measurement errors for macro variables and bond yields. After a number of trial runs, we set the measurement errors to be 2.5 bps for our quarterly data which can be translated to 10 bps for annualized data. For a fair comparison of the two models, we do not constraint the QTSM to produce non-negative yields.

2.5 Data

In this paper, we calculate zero coupon yields from the data for the JGB market during the sample period from 1985:Q1 to 2016:Q1 using the method proposed by Kikuchi and Shintani (2012). To keep the consistency with previous empirical studies, we use the JGB yields of the 2, 4, 12, 20, 32 and 40 quarters, (or 0.5, 1, 3, 5, 8 and 10 years). These time series are depicted in Figure 1 (a).

And data of factors is obtained as follows. As monetary policy stance, the overnight call rate is obtained from the Bank of Japan (BOJ). The real GDP gaps are derived from difference

between real GDP reported in the Cabinet Office, Government of Japan and potential GDP reported in the BOJ. Log first difference of Consumer Price Index (CPI) is adopted as inflation. The spread (or slope) is difference between yields of 2 and 40 quarter. The curvature is derived from discrepancy of two slopes in three points of yields curve, i.e.,

$$c_t = (y_t^{(20)} - y_t^{(2)}) - (y_t^{(40)} - y^{(20)}).$$

These series are drawn in Figure 1 (b). From Section 3 to Section 5, we adopt the dataset of factors consisting of three series, say the policy rate, the spread and the real GDP gap. On the other hand, in Section 6, we expand it to three kinds of the datasets in order to verify which factors perform better prediction of yields. Additional two datasets are (1) fundamental factors of yields curve such as level, slope and curvature the first two of which are the policy rate,and the spread, and (2) macro-finance factors, say the policy rate, the GDP gap and CPI inflation.

To better understand the data and the estimation results, let us take a brief review on the BOJ's monetary policy. The BOJ started to ease the base interest rate (the uncollateralized overnight call rate) in the early 1990s, which is subsequently lowered down to 0.5 percent and 0.25 percent in 1995:Q4 and 1998:Q4 respectively. To further simulate the economy, the BOJ adopted the zero interest rate policy (ZIRP) during the period from 1999:Q1 to 2000:Q3 by keeping the base interest rate effectively to zero. After a short-term recovery in early 2000s, the Japan economy went back to a recession against the background of the internet bubble, which led to the introduction of the quantitative monetary easing policy (QMEP) in order to combat deflationary pressure. Since then, the Japanese base interest rate has been kept very close to the zero lower bound.[1]

In July 2006, the BOJ released the ZIRP because of recovery of economic activity shown as an increase in real GDP gap following a rise of CPI inflation drawn in Panel (b). The BOJ, however, attributed return of the ZIRP to Lehman Brother's collapse occurred in September 2008. In April 2013, the replacement of the governor of the BOJ drastically changed monetary policy to one referred to as quantitative-qualitative easing policy (QQE) in which the operation target was replaced the call rate with monetary base, and the monetary base was quantitatively increased to nearly double amount. In spite of unconventional policy such as the QQE, breakaway from deflation has never been achieved, so that it brought its policy to the negative interest rate policy whose target is set to around -0.1% in Jan 2016.

[Insert Figures 1 around here]

[1]Baba (2006) provides a comprehensive review of the Bank of Japan monetary policy and the JGB market development over the sample period.

3 Density Forcasts of Two Models

3.1 Model estimation

First-of-all, we look at three factors such as policy rate, spread and real GDP gap, and bond yields to investigate the goodness-of-fit of the Gaussian ATSM and QTSM. Table 1 reports the summary statistics of the posterior estimations of the ATSM parameters, while Table 2 reports the corresponding results for the QTSM which imposes a quadratic mapping in between bond yields and macro factors. These estimations are obtained by 30,000 draws of MCMC samplings after discarding the first 10,000 burn-in draws based on the Bayesian methods described in the Appendix. Figures 2 and 3 show the filtered factors and the fitted bond yields of the both models, respectively. Panel (a) shows the ATSM, while Panel (b) shows the QTSM. The dashed black line and the solid red lines represent actual values and fitted values, respectively. In Figure 3, the dashed blue line represents degree of pricing errors (or discrepancy between them), and the shaded grey band represents 90% confidence interval of the distribution. It can be seen that the three factors track quite closely to the actual data and the fitting of the bond yields are reasonably good across maturities.

As can be seen from Figure 2 (a), ATSM is adequate to jointly model the dynamics in bond yields and macro factors. Nevertheless, it is noted that the model-implied bond yields often bleach the zero lower bound and become negative during the sample periods after late 1995 in Figure 3 (a). This generates notable degree of pricing errors when the actual short-term bond yields are effectively zero. Meanwhile, the filtered policy rate and spread factors of the QTSM as shown in Figure 2 (b) also track closely to the actual data, however, there is a small deviation of the GDP gap factor. The latter case indicates that the enforced quadratic mapping in the macro-finance QTSM can be potentially misspecified.

Figure 4 depicts actual and estimated yields curve of the ATSM and the QTSM at specified six points including both of non-zero (1991:Q1, 1996:Q1), zero interest rate policy periods (2001:Q1, 2006:Q1, 2011:Q1) and the negative interest rate policy periods (2016:Q1). These graphs indicate goodness of fit in terms of cross section aspect of time series of term structure of Figure 3. In contrast to the ATSM which produces a large amount of probability of negative interest rates, the model-implied bond yields of QTSM tend to be positive. Hence, the QTSM is able to produce a much better fit to the actual short-term bond yields near the zero lower bound. Figure 4 (b) draws a counterpart of the QTSM for the yields curve of the ATSM in Figure 4 (a). It is noteworthy that both of the ATSM and the QTSM successfully estimate even yields curve of the negative interest rate period as shown in the bottom right graphs of Panels (a) and (b).

[Table 3 around here]

[Insert Figures 2, 3 and 4 around here]

[Insert Table 1 and 2 around here]

To understand how the term structure model predicts the responses of bond yields to shocks in the underlying macro variables (i.e., impulse response), it is important to take a closer look at the estimated factor loadings. Figure 5 reports the factor loadings of the estimated Gaussian ATSM and QTSM. The dashed red line of Panel (a) shows the factor loadings of the estimated Gaussian ATSM using the recursive relationship (3) and the posterior means under Q-measure in Table 1. Meanwhile, the solid blue line shows the factor loadings of linear terms of the estimated QTSM using the recursive relationship (8) and the posterior means under Q-measure in Table2. In Panel (b), the solid blue lines show 6 loadings through the quadratic terms of the QTSM: c_{11}, c_{22}, c_{33},c_{12},c_{13}, c_{23}. [2]

Now, we focus on results of the ATSM. Because the short-term interest rate is taken as one of the state variables, we can impose the initial loading of the spread and output factors: $\delta_{1,2}$ and $\delta_{1,3}$ at Eq.(5), to be zero as in Ang et al. (2011). For the ATSM, we see that the output and inflation loadings: b_2 and b_3, are positive for all maturities, which is consistent with the Taylor rule specification. For example, a positive shock to output gap induces an upward shift up to 10 Quarters, and a steepening of the yield curve, which is consistent with the view that the slope of yield curve is highly related to economic outlook (Diebold and Rudebusch, 2013). As expected, the loading to the policy rate: b_1, is less than one and hence the transmission effect of the short-term interest rate to the long-end of the yield curve is imperfect as can be seen from Figures 3 and 4.

We turn to those of the QTSM. To keep consistency with the setting of ATSM, we set the initial loadings to the output and spread factors: Φ_{22} andΦ_{33}, and other off-diagonal elements in Eq.(6), to zero as shown in Figure 5 (b). Firstly, it is worth to take a look at the diagonal elements of the factor loading c_n, which captures most of the variation in the yield curve. As expected, the loading to the quadratic terms of the inflation and output factors: c_{22} and c_{33}, are positive, indicating that investor demands a higher bond yields when spread and output uncertainty are high as shown in Figure 5(b). Moreover, the QTSM allows a flexible interaction in between different factors through the cross terms (i.e., the off-diagonal elements in the factor loading c_n). Diebold and Rudebusch (2013) note that the negative interaction in between factors are important to model interest rates near the zero lower bound. In our case, the factor loadings for the cross terms of (f_t, g_t) and (g_t, s_t) : c_{13} and c_{23}, are estimated to be negative, which reflect the high flexibility of the QTSM in relating bond yields to the selected macro factors. We argue that it is the negative loadings of the cross terms that generate the off-setting effects such that the QTSM is able to capture the persistent and sticky short-term

[2]Note that the factor loading c_n is symmetric by construction. Accordingly, $c_{12} = c_{21}$, $c_{13} = c_{31}$, and $c_{23} = c_{32}$.

bond yields under the ZIRP, e.g., see the fitting of the 2Q yield in Figure 3 (b).

[Insert Figure 5 around here]

3.2 Density Forecasts of macro factors and bond yields

Before analyzing the optimal prediction pool, we calculate the posterior prediction distributions of the three factors and the JGB yields of the two models as described in Section 2, using 10,000 draws of posterior estimates over the full sample as shown in Table 1 and 2, as well as generating 10,000 draws of shocks ε_t. Here, we briefly describe calculation of predictive densities of the factors and yields. First, we conduct sampling of h-step-ahead factors, X_{t+h}, from the state equation Eq.(12) with draws of posterior estimates and from 1-step-ahead shocks to h-step-ahead shocks. Next, we sample h-step-ahead yields, y^n_{t+h}, from the measurement equation, Eq. (10), using new draws of h-step-ahead factors, X_{t+h}. The sampling is regarded as their predictive densities.

Figure 6 (a) shows the ATSM prediction of the JGB yield curve across 6 maturities for the following forecasting periods: 1988:Q1 – 1998:Q1, while Figure 6 (b) shows the QTSM prediction. The solid black line represents actual values, the solid red line represents the median of posterior prediction distributions and the shaded blue band represents 90% confidence interval of the distribution. In the period of 1992Q4 - 1998Q1, in which monetary policy has not stand under ZIRP yet, the bond yields are quite far away from the zero lower bound as in Figure 6(a). Although the median forecast fits well to the actual data, the ATSM predicts negative bond yields when the forecasting horizon is beyond 4 to 8 quarters. For the in-sample prediction in 1988:Q1 - 1998Q1 in Panel (b), the QTSM produces a less accurate forecast as with the ATSM model as in Panel (a).

Similar to Figure 6, Figure 7 shows the both model prediction for the forecasting periods: 2008:Q1 – 2016:Q1. When the Bank of Japan adopted the ZIRP and the QMEP in 2008:Q4, the prediction of bond yields by the ATSM is even more unrealistic: as the short-term bond yields are close to the zero lower bound, the model predicts with almost half of the probability that the bond yields are negative as in Panel (a). Even for the 20Q (5-year) bond yield, there is a substantial probability of bleaching the zero lower bound when the forecasting horizon is beyond 4 quarters. This reflects that the Gaussian ATSM is very unreliable for the prediction of bond yields when interest rates are close to zero. In contrast, as can be seen from Panel (b), the prediction density of the QTSM is positively skewed because the bond yields are bounded below by zero due to the imposition of non-negative short rate in QTSM. The strength of the QTSM is found to be prominent during the period of 2008:Q1 – 2016:Q1 when the zero lower bound is binding: the prediction produces only positive bond yields even though the short-term interest rate is extremely close to zero as in Panel (b). From the fan chart of the

QTSM predictive density, we can observe that the probability mass near zero is significant even for mediam-term to long-term forecasting horizons. This reflects the stickiness nature of the QTSM which allows one to capture the persistence of the zero interest rate policy (Kim and Singleton, 2012).

[Insert Figures 6 and 7 around here]

4 Methods of Prediction Pool

4.1 Predictive Scores

From a Bayesian perspective, the *marginal likelihood* is commonly used as a criterion of model choice, since it is interpreted as the predictive density of a model obtained by integrating with respect to the prior density of the model parameters Θ. A model with the highest predictive density is thought to be the best model explaining behaviors of observations based on information on the whole data. Let us denote a vector of future observations as y_{t+h}, where h is h-step-ahead forecast, and its history as $Y_t^o = \{y_g, ..., y_t\}$, where $g \leq 1$ is the starting date and the superscript "o" denotes the observed data. The predictive density of a model with respect to the prior of parameters Θ is defined as

$$p^{Prior}\left(y_{t+h}^f - y_{t+h}^o | Y_t^o, \mathcal{M}\right) \equiv \int p\left(y_{t+h}^f - y_{t+h}^o | Y_t^o, \Theta, \Sigma, \mathcal{M}\right) p(\Theta|\mathcal{M}) p(\Sigma) d\Theta d\Sigma,$$

where y_{t+h}^f, and y_{t+h}^o are forecasted and observed values in period $t+h$, respectively, and the difference between them is their forecasting errors ε_{t+h}. Σ is a covariance matrix of the forecasting errors, ε_{t+h}, and $\mathcal{M}$ is a prediction model. $p\left(\varepsilon_{t+h}|Y_t, \Theta, \Sigma, \mathcal{M}\right)$, and $p(\Theta|\mathcal{M})$ denote the likelihood function and the prior density of Θ of a prediction model $\mathcal{M}$, respectively. When we set $h = 1$, then the density is regarded as the marginal likelihood. When by replacing the prior density with the posterior density above predictive density as noted by Geweke (2010), it can be redefined as a *posterior* predictive density, say

$$p^{Post}\left(y_{t+h}^f - y_{t+h}^o \,|\, Y_t^o, \mathcal{M}\right) \equiv \int p\left(y_{t+h}^f - y_{t+h}^o \,|\, Y_t^o, \Theta, \Sigma, \mathcal{M}\right) p(\Theta \,|\, Y_t^o, \mathcal{M}) d\Theta d\Sigma,$$

where $p(\Theta|Y_t, \mathcal{M})$ is the posterior density of Θ conditional on history of observations until period t, Y_t^O, and a model, $\mathcal{M}$. Following Geweke and Amisano (2011), we use the posterior predictive density in order to construct a *predictive score* for evaluating the forecasting performances of a single prediction model and of a convex combination of multiple prediction models with the optimal model weights. We define the predictive score of a model $\mathcal{M}$, $p(y_{t+h}^f; Y_t^o, \mathcal{M})$,

for h-step-ahead forecast as

$$p(y_{t+h}^f;\, Y_t^o,\, \mathcal{M}) \equiv p^{Post}.3\left(y_{t+h}^f - y_{t+h}^o \,|\, Y_t^o,\, \mathcal{M}\right), \tag{13}$$

and regard it as the key element of the following prediction pooling methods.

4.2 Static prediction pool

Geweke and Amisano (2011) propose forecasting combination using the predictive score. Since they assume model weights are constant over sample period, it is referred to as the static prediction pool. Let us redefine $\mathcal{M}$ as the collection of competing multiple models, e.g., $\mathcal{M} = (\mathcal{M}_1, \mathcal{M}_2)$. Given two prediction models, $\mathcal{M}_1$ and $\mathcal{M}_2$, and constant model weights, the predictive score for h-step-ahead forecast can be constructed as the convex combination of predictive scores of competing models,

$$p^{SP}\left(y_{t+h}^f;\, Y_t^o, \mathcal{M}\right) \equiv \lambda\, p\left(y_{t+h}^f;\, Y_t^o, \mathcal{M}_1\right) + (1-\lambda)\; p\left(y_{t+h}^f;\, Y_t^o, \mathcal{M}_2\right), \tag{14}$$

where $\lambda \in (0,1)$ and $1-\lambda$ are constant values indicating model weights in favor of $\mathcal{M}_1$ and $\mathcal{M}_2$, respectively. The optimal prediction pool is then obtained by maximizing the cumulative log predictive score, LPS^{SP}, for the whole prediction periods as

$$LPS^{SP}(\lambda, h) \equiv \sum_{t=1}^{T} \log\left[\lambda p\left(y_{t+h}^f;\, Y_t^o, \mathcal{M}_1\right) + (1-\lambda)\, p\left(y_{t+h}^f;\, Y_t^o, \mathcal{M}_2\right)\right], \tag{15}$$

by choosing $\lambda^* = \arg\max\, LPS^{SP}(\lambda, h)$. In our study, we generate a predictive density of macroeconomic observations based on each of the two DSGE models described in Section 2 from posterior estimations of their model parameters.

4.3 Dynamic prediction pool

We also adopt another pool method with time-varying model weight, say dynamic prediction pool method proposed by Del Negro et al. (2016). In the dynamic prediction pool method, time-varying weights follow continuous values between zero and one by incorporating probit model. The dynamic model consists of the following two equations,

$$\begin{aligned} \lambda_t &= \Phi(x_t), \\ x_t &= (1-\rho)\mu - \rho x_{t-1} + \sqrt{1-\rho^2}\,\sigma\,\varepsilon_t,\;\; x_0 \sim N(\mu, \sigma^2), \end{aligned} \tag{16}$$

where $\lambda_t \in [0,1]$ is a model weight at period t, and x_t is a latent variable indicating an input of a probit transformation and following an AR(1) process. ρ is the autocorrelation coefficient. $\Phi(\cdot)$ is the cumulative density function of the standard normal distribution, the disturbance

term follows $\varepsilon_t \sim N(0,1)$ and x_0 is the initial value of x_t. The autocorrelation coefficient ρ captures how smoothly the weighting coefficient can change over time. The closer ρ is to one, the more slowly the model weights, λ_t , change over time. When $\rho = 1$, the model reduces to the case of static prediction pooling in Geweke and Amisano (2011) by taking $\lambda_t = \lambda$. When $\rho = 0$, it indicates that λ_t is serially-independent and follows a random walk. μ is the mean of the unconditional distribution of the model weights, and σ is the variance of x_t, the large value of which makes the model weights fluctuate drastically. From these equations, we obtain conditional expectations and variances of the latent variables for h-step-ahead forecast, x_{t+h},

$$\begin{aligned} E(x_{t+h}|x_t) &= \rho^h x_t + (1-\rho)\mu \sum_{i=0}^{h-1} \rho^i, \\ Var(x_{t+h}|x_t) &= (1-\rho^2)\sigma^2 \sum_{i=0}^{h-1} \rho^{2i}, \end{aligned} \tag{17}$$

where both conditional values converge to unconditional values, $E(x_{t+h}) = \mu$, and $Var(x_{t+h}) = \sigma^2$, when $h \to \infty$. And coefficient μ and variance σ^2 of the initial value of latent variable are also equivalent to the unconditional values.

This study examines two versions of the above model following Del Negro et al (2014). The one is set as $\mu = 0$ and $\sigma^2 = 1$. $\mu = 0$ indicates that unconditional expectation of model weight is 0.5, $\Phi(0) = 0.5$, since unconditional expectation of weight is assumed to equivalent between both models. And setting $\sigma^2 = 1$ comes from assumption of the latent variable in probit model. Accordingly, we only estimate a coefficient ρ in this version. The second set three parameters freely and estimate them.

We obtain the dynamic prediction pooling of the log predictive score as

$$LPS^{DP}(\lambda_{t+h},\, h) \equiv \sum_{t=1}^{T} \log\left[\lambda_{t+h}\, p\left(y^f_{t+h};\, Y^o_t,\, \mathcal{M}_1\right) + (1-\lambda_{t+h})\, p\left(y^f_{t+h};\, Y^o_t,\, \mathcal{M}_2\right)\right]. \tag{18}$$

We adopt a particle filter for coping with a nonlinear model such as a probit model, and incorporate the nonlinear filtering method into a Bayesian estimation with MCMC procedure, following Del Negro at el. (2016). We set number of particles of the filter as 5,000 and calculate approximate values of log predictive scores defines as Eq. (18). And we conduct 10,000 iterations as the MCMC procedure and discard the first 5,000 draws as burn-in[3].

[3]We code the algorithm of a particle filter following Johanness and Polson (2009). And A joint of the MCMC procedure and the partile filter in our study is also following Andrieu et al. (2010).

4.4 Estimation methodology of pooling methods

In order to estimate and compare the predictive scores of individual prediction models, say two term structure models, and the pooling methods, we adopt a Bayesian approach with the Markov Chain Monte Carlo (MCMC) method for the static and the dynamic prediction pool methods.

Although Waggoner and Zha (2012) simultaneously estimated two macroeconomic models and the pool method, the simultaneous estimation of the model parameters, Θ, under a regime sustaining only for a short period is thought to have only a low level of accuracy. This is because a regime generated in every MCMC iteration of a pooling method is different from that of the previous iteration, and a different regime period expands variations of drawing Θ in the step of MCMC iteration in the term structure model. By adopting a two-step procedure following Geweke and Amisano (2011) and Del Negro et al. (2016), we can avoid generating instability in the model parameters, Θ, estimated based on different regime periods.

The two-step procedure for the dynamic pool is described as below.

Step 1. Make density forecasts of the term structure models.

- The posterior estimates of parameters, $p(\Theta|Y_{t-1}^o, \mathcal{M}_i)$, under the term structure models, M_i, for $i = 1, \cdots, n$, are obtained for the full sample period, using the MCMC method.
- We compute the predictive densities and predictive scores of obserbations, $p(y_{t+h}^f|Y_t^o, \Theta, M)$, from sampling of $p(\Theta|Y_{t-1}^o, \mathcal{M}_i)$ of each DSGE model, $\mathcal{M}_\text{i}$, by Monte Carlo simulation technique.

Step 2. Make the optimal combination of density forecasts.

- We calculate the optimal combination of the log scores of the term structure models obtained in the previous step, using parameters of pooling methods drawn from the Gibbs sampling method with particle filter.

5 Empirical Results

5.1 Prediction Score

In this section, we explore the combination of the ATSM and QTSM in the prediction of bond yields using the optimal prediction pooling as described in the last Section. To begin, it is useful to look at the comparison of the predictive densities of the two individual models based on the log-score criteria. While we have performed the comparison using both one-quarter-ahead ($h = 1$) and four-quarter-ahead forecasts ($h = 4$), we only report the charts of four-quarter-ahead forecasts for exposition purpose. Panel (a) of Figure 8 shows two lines indicating

log scores of forecasting all six yields by each of the two term structure models, calculated from Eq.(13). The scores of the ATSM (dashed red line) dominates those of the QTSM (solid blue line) for the sampling period from 1985:Q1 - 1995:Q4 (before entering zero interest rate policy), while the QTSM dominates the ATSM when the JGB bond yields are close to zero percent rate since 1996Q1. Panel (b) of Figure 8 also shows log scores of forecasting but focusing on single yields from 0.5 years maturity to 10 years maturity. Left hand side of Table 3 represents sum of log scores of both models overall period. In terms of the whole period, forecasting performance of the ATSM is overwhelmingly superior to that of the QTSM, since the former is around 3,660 against 3,584 for the latter.

These results, however, suggest that one can potentially improve the predictive density by combining appropriately the two models which appear to perform better in different sample periods. In other words, they capture different properties of the movements of bond yields and their interaction with the macroeconomy. To fix idea, recall that in Section 4 that we are looking at the log-score function: $LPS(y_t; Y_{t-1}, \text{Pool})$, that is a convex combination of the prediction density at the time-t observation of the ATSM and QTSM as

$$LPS(y_t; Y_{t-1}, \text{Pool}) \equiv \log \left[\lambda_t p \left(y_t^O; Y_{t-1}^O, \Theta_{QTSM} \right) + (1 - \lambda_t) \, p \left(y_t^O; Y_{t-1}^O, \Theta_{ATSM} \right) \right],$$

in which we take λ_t as the weighting assigned to the QTSM while $1 - \lambda_t$ as the weighting assigned to the ATSM, and Θ_{QTSM} and Θ_{ATSM} are the posterior estimates of the QTSM and ATSM parameters over the full sample period, respectively. As noted in Waggoner and Zha (2012), we can take the estimated paramters for both models as given before we pool the models. Then, we compute the prediction scores as

$$p(y_t^O; Y_{t-1}^O, \mathcal{M}_1) \equiv p \left(y_t^O; Y_{t-1}^O, \Theta_{QTSM} \right),$$

$$p(y_t^O; Y_{t-1}^O, \mathcal{M}_2) \equiv p \left(y_t^O; Y_{t-1}^O, \Theta_{ATSM} \right),$$

in order to evaluate the log-score criteria. This two-step procedure significantly reduces the computational burden.

[Table 3 around here]

[Insert Figures 8 around here]

5.2 Static pool

Table 4 reports the posterior estimates of weighting coefficient λ for the static pooling scheme as Eq.(15). This coefficient is estimated with MCMC simulation and obtained from 50,000

MCMC draws after discarding the first 20,000 draws as burn-in. For the four-quarter-ahead forecast, the posterior distribution of λ is skewed to the right, indicating that the parameter restriction of $\lambda \geq 0$ is binding and one should over-weight the ATSM model and under-weight the QTSM. Table 4 (a) shows the posterior mean of the constant weighting coefficient in favor of the QTSM as around 11% for the four-quarter-ahead forecast. However, focusing on the period of the zero interest rate policy after 2000:Q1, the constant model weight on the QTSM goes up to 48% and balances with the counterpart as Panel (b).

We also compute the simulation inefficiency statistics as in Kim et al. (1998) represented in the seventh column of Table 4. According to the statistics, samples of 30,000 draws might be sufficient iterations of simulations to obtain the posterior estimates of model weight λ.

[Table 4 around here]

5.3 Dynamic pool

Table 5 represents posterior estimates of two versions, say flexible one parameter and three parameters, for the dynamic pooling scheme which imposes a smooth transition between the two selected models as described in Eq.(18). And Figure 9 shows the posterior distributions of time varying weight in favor of the QTSM, say, the forecast combination of all six yields of one flexible parameter version in Panel (a) and that of three parameters version. Meanwhile Panel (c) draws those of each of six yields of one flexible parameter version. The solid black line denotes the posterior means, while the blue shaded area represents their 90% credible interval. The red line denotes 50% model weight. The estimation is obtained from 5,000 draws of particle filter with constant autocorrelation coefficient ρ following Del Negro et al (2016).

The sum of log score of the QTSM is defeated by the ATSM as Tabel 3. However, as can be seen from Panel (a), the weighting to QTSM increases after 1995:Q1 and reachs to 50% in 1999:Q1 entering into the ZIRP. And then it keeps to balance by following dominance of the ATSM afterwards in particular after 2011, although the weights of the QTSM model decline as much as 10% before 1995. The case of three parameters version is also similar but swing become wider and reach to 50% in 1996:Q1 earlier than the one parameter version as Panel (b).

From Panel (C), weight of different yields does not seem to depend on different maturities. In other word, for each yield the QTSM is generally defeated by the ATSM before1995. Between 1995 and 2010, performances of the both models countervail each other. After 2010, the QTSM prevail the other.

However, the dispersion of the posterior distribution of λ_t appears to be large with the 90% band fluctuating around 0.1 to 0.9 for the four-quarter-ahead forecasts. This indicates that the

dynamic pooling scheme does not allow us to obtain a clear cut in between QTSM and ATSM.

[Table 5 around here]

[Insert Figure 9 around here]

5.4 Comparison

Lastly, let us compare the performance of different models and pooling schemes in terms of the log score of prediction density. Table 3 summarizes the corresponding cumulative log score performance for the four-quarter-ahead forecasts obtained from Eq.(13), Eq.(15) and Eq.(18). Figure 10 shows the time series of the log score of the three pooling schemes as well as those of the two individual models. As can be seen in Table 3, the dynamic pooling scheme with three parameter version produces the best cumulative log score: this is because it allows one to combine the ATSM and QTSM efficiently by switching from the ATSM before 1996:Q1 to the QTSM after 1996:Q1 as depicted in Figure 9 (b). Interestingly, the static pooling scheme only marginally improve the cumulative log-score performance for the four-quarter-ahead forecast, although it may improve the prediction density in certain sub-sample periods. This suggests that an appropriate pooling scheme is important for one to achieve an overall improvement (in terms of the log-score criteria) in the prediction of future bond yields when models are combined.

[Insert Figure 10 around here]

6 Extension to Dynamic Pool with Six Models

Up to the last section, we cling on to use only one dataset for three observable factors, say the level, the slope and the real GDP gap, and consider the forecast combination of the ATSM and the QTSM. Instead, we expand it to three datasets of the three observable factors, and apply dynamic prediction pool to forecast combination of six models by combining one out of three datasets with one out of the two term structure models. The expansion indicates that we cannot only measure which model performs better forecasting for a certain period than the other, but also measure which combination of dataset and model does for that period. In addition, we evalute model weights of forecasts for single yields as well as those for total of six yields. These empirical results show us what characteristic of different dataset affect different maturities of term structure.

To decide time-varying model weights corresponding to more than three models, instead of using probit formation, Eq.(16), the dynamic pool is built as a new formation based on Dirichlet distribution, $Dir(\bullet)$, as follows.

$$(\lambda_{1t},\, \lambda_{2t}, \cdots,\, \lambda_{mt}) = Dir(\,\exp(s_t)\lambda_{1t-1},\, \exp(s_t)\lambda_{2t-1},\, \cdots,\, \exp(s_t)\lambda_{mt-1}), \tag{19}$$

$$s_t = (1-\rho)\mu_0 + \rho s_{t-1} + \sqrt{1-\rho^2}\,\sigma\,\varepsilon_t,\;\; \varepsilon_t \sim N(0,1),$$

where s_t is time-varying scale parameter which follows AR(1) process and controls variance of time-varying model weights λ_{it} for $i = 1, \cdots, m$. μ_0 is unconditional expected value of scale parameter s_t. As s_t become smaller, then the band of fluctuation of model weight become wider. And the Dirichlet distribution constrains $\Sigma_{i=1}^{m}\lambda_{i,t} = 1$, and we additionally restrict model weights to $\lambda_{it} > 0.01$, since model weights in current period,$\lambda_{i,t-1}$, must become zero if those of previous period,$\lambda_{i,t-1}$ are zero. In our case, six models are combined so that $m = 6$, and the log predictive score of six models is derived as

$$LPS^{DP}(\lambda_{t+h},\, h) \equiv \sum_{t=1}^{T} \log\left[\lambda_{1,t+h}\, p\left(y_{t+h}^{f};\, Y_t^{o},\, \mathcal{M}_1\right) + \cdots + \lambda_{m,\,t+h} p\left(y_{t+h}^{f};\, Y_t^{o},\, \mathcal{M}_m\right)\right]. \tag{20}$$

We conduct the particle filter for estimating optimal prediction in term of the log predictive score, Eq. (20) with time varying model weight, Eq.(19). To do so, we set number of particle as 10,000 particles, and also fix paramters of Eq. (19) as $\mu_0 = 1$, $\rho = 0.75$ and $\sigma = 1$.

The three datasets of the factors adopted in this section are (1) level, slope and curvature, (2) level, slope and real GDP gap, and (3) level, real GDP gap and inflation. We refer to the first dataset, the second one and the third one as 'Standard', 'GDP gap', and 'Macro Finance', respectively.

The dynamic model weights, λ_{it}, of the six models are drawn in Figure 11. Panel (a) of the figure depicts the model weight of forecasting for all six yields. Panel (b) depicts the weight of forecasting for single yield from 0.5 years to ten years in terms of maturities. As you see from Panel (a), the QTSMs with dataset "GDP gap" (the deep blue shade area) and dataset "Standard" (the deep green shade area) tend to prevail forecasting for all six yields after Lehman Brother collapse happened in September 2008. However, before Lehman's failure, the ATSM with dataset "Standard" (the pink shade area) dominates the six models until 1995, say before starting extreme low interest rate policy. And then the ATSM with dataset "GDP gap" (the light red area) took over this position, and the ATSM with dataset "Macro Finance" (the white area) prevail for the next period between 2002 and 2006, i.e., a boom period triggred by subprime loan of the US.

Turning to single yield forecasting cases as Panel (b), we realize that the ATSM with dataset "Macro Finance" (the white area) is relatively strong for forecasting of short term

rate such as 0.5 years and 1 years, while the QTSMs with dataset "GDP gap" (the deep blue shade area) and dataset "Standard" (the deep green shade area) are stronger for the long terms rate such as 8years and 10 years. These graphs indicate that macroeconomic series such as GDP gap and inflation include useful information on forecasting of short term interest rate in spite of the presence or absence of zero lower bound of interest rate. In contrast, the second moment of the factor, especially curvature, included in the QTSM plays an important role on forecasting of long term interest rate.

Finally, we compare the cumulative log score of the dynamic prediction pool for all yields with those of six individual models as Table 6. According to the LHS of Table 6, the highest log scores (3665,6) out of the six individual models is the ATSM with the data set 'standard' version following the ATSM with 'GDP gap', while the lowest (3450.9) is the QTSM with the data set 'Macro-Finance' following the ATSM with the same data set. For all of the sample period including the zero interest rate period, the ATSMs would be a good model for prediction of yields, if we only have to pay attention to selecting data set. The RHS of Table 6 shows the score of dynamic prediction with 6 models is around 3709 which completely dominates those of all six models and the dynamic prediction with 2 models represented in Table 3. These results suggest that combination of data sets as well as of models are another important factor of improving predictive densities.

[Insert Table 6 around here]

[Insert Figure 11 around here]

7 Conclusion

In this paper, we study the optimal prediction pool of the Gaussian ATSM and QTSM with macro-finance features using the JGB data from 1985:Q1 to 2016:Q1 that cover the zero interest rate policy in Japan. Our estimation results show that the QTSM provides a more realistic description of bond yields when the zero lower bound is binding, although the ATSM appears to provide a better fit to bond yields and macroeconomic variables simultaneouly. This suggests that one should combine the two models for the prediction of future bond yields under different market scenarios.

In addition, we expand the forecast combination of the ATSM and the QTSM to that of six models including three dataset characterized by sorts of factors. And we consider what combination of factors and terms structure model realize best forecast. We show that the ATSM with macro-finance feature is strong for forecasting of short term, while the QTSMs with spread and real GDP gap are superior for the long terms rate. In other word, macroeconomic

variables, i.e., GDP gap and inflation, include useful information on prediction of short term rate in spite of the presence or absence of zero lower bound of interest rate. In contrast, the second moment of the factor, especially curvature, included in the QTSM plays an important role on prediction of long term interest rate.

For future research, it is instructive to explore a wider combination of macroeconomic variables, such as unemployment rate, M2 growth and credit-to-GDP ratio. Moreover, it is interesting to repeat the exercise using the US treasury yield data since the financial crisis of 2008, although the history may be limited for a robust statistical identification. A potential remedy is to use macroeconomic variables with higher frequency such as monthly data. An alternative is to use the estimation technique with mixing frequency data such as the one proposed in Camacho and Perez-Quiros (2010).

References

[1] Ahn, D.H., Dittmar, R.F. and Gallant, A.R. (2002) Quadratic Term Structure Models: Theory and Evidence. Review of Financial Studies 15, 243-288.

[2] Albert, J.H. and Chib, S. (1993) Bayes Inference via Gibbs Sampling of Autoregressive Time Series Subject to Markov Mean and Variance Shifts. Journal of Business and Economic Statistics 11 (1), 1-15.

[3] Andreasen, M. and Meldrum, A. (2013) Likelihood Inference in Non-linear Term Structure Models: the importance of the lower bound. Bank of England Working Paper No. 481.

[4] Andersen, L.B.G. and Piterberg, V.V. (2010) Interest Rate Modeling. Volume 2: Term Structure Models.

[5] Ang, A., Boivin, J., Dong, S. and Loo-Kung, R. (2011) Monetary Policy Shifts and the Term Structure. Review of Economic Studies 78, 429-457.

[6] Ang, A., Dong, S. and Piazzesi, M. (2007) No-Arbitrage Taylor Rules. NBER Working Paper No. 13448.

[7] Ang, A. and Piazzesi, M. (2003) A No-Arbitrage Vector Autoregression of Term Structure Dynamics with Macroeconomic and Latent Variables. Journal of Monetary Economics 50 (4), 745-787.

[8] Ang, A., Piazzesi, M. and Wei, M. (2006) What Does the Yield Curve Tell us about GDP Growth? Journal of Econometrics 131, 745-787.

[9] Baba, N. (2006) Financial Market Functioning and Monetary Policy: Japan's Experience. Monetary and Economic Studies 24, 39-71.

[10] Bernanke, B.S., Reinhart, V.R., and Sack, B.P. (2004) Monetary Policy Alternatives at the Zero Bound: An Empirical Assessment. Brookings Papers on Economic Activity 2, 1-100.

[11] Camacho, M. and Perez-Quiros, G. (2010) Introducing the Euro-Sting: Short-term Indicator of Euro Area Growth. Journal of Applied Econometrics 25, 663-694.

[12] Cox, J.C., Ingersoll, J.E., Ross, S. (1985) A Theory of the Term Structure of Interest Rates. Econometrica 53, 385-408.

[13] Del Negro, M., Hasegawa, R.B. and Schorfheide, F. (2016) Dynamic Prediction Pools: an Investigation of Financial Frictions and Forecasting Performance. Journal of Econometrics, Vol.192 (2), p.391-405.

[14] Del Negro, M. and Schorfheide, F. (2010) Bayesian Macroeconometrics. Handbook of Bayesian Econometrics.

[15] Diebold, F.X., Piazzesi, M. and Rudebusch, G.D. (2005) Modeling Bond Yields in Finance and Macroeconomics. American Economic Review 95, 415-420.

[16] Diebold, F.X., Rudebusch, G.D. (2013) Yield Curve Modeling and Forecasting. The Dynamic Nelson-Siegel Approach. Princeton University Press.

[17] Diebold, F.X., Rudebusch, G.D. and Aruoba, B. (2006) The Macroeconomy and the Yield Curve: A Dynamic Latent Factor Approach. Journal of Econometrics 131, 309-338.

[18] Duffie, D., Kan, R. (1996) A Yield-factor Model of Interest Rates. Mathematical Finance 6, 379-406.

[19] Eo, Y. and Kang K.H. (2014) Forecasting the Term Structure of Interest Rates with Potentially Misspecified Models. Working paper.

[20] Geweke J. (2010) Complete and Incomplete Econometric Models. The Econometric and Tinbergen Institutes Lectures, Princeton University Press.

[21] Geweke, J. and Amisano, G. (2011) Optimal Prediction Pools. Journal of Econometrics 164, 130-141.

[22] Geweke, J. and Amisano, G. (2012) Prediction with Misspecified Models. American Economic Review: Papers & Proceedings 2012, 102 (3), 482-486.

[23] Hoeting, J., Madigan, D., Raftey, A. and Volinsky, C. (1999) Bayesian Model Averaging. Statistical Science 14, 382-401.

[24] Kikuchi, Kentaro and Kohei Shintani (2012) "Comparative Analysis of Zero Coupon Yield Curve Estimation Methods Using JGB Price Data", IMES Discussion Paper Series 2012-E-4, Bank of Japan.

[25] Kim, S., Shephard, N. and Chib S. (1998) Stochastic Volatility: Likelihood Inference and Comparison with ARCH Models. Review of Economic Studies 65 (3), 361-393.

[26] Kim, D., Singleton, K.J. (2012) Term Structure Models and the Zero Nound: an Empirical Investigation of Japanese Yields. Journal of Econometrics 170, 32-49.

[27] Li, H, Tao, L and Yu, C. (2013) No-Arbitrage Taylor Rules with Switching Regimes. Management Science 59 (10), 2278-2294.

[28] Leippold, M., Wu, L. (2002) Asset Pricing under the Quadratic Class. Journal of Financial and Quantitative Analysis 37 (2), 271-295.

[29] Johanness, M. and Polson, N. (2009) Particle Filtering, Handbook of Financial Time Series, edited by T.G. Anderson et al. pp.1015-1029. Springer-Verlag Berlin Heidelberg.

[30] Piazzesi, M. (2010) Affine Term Structure Models. Handbook of Financial Econometrics, edited by Y. Ait-Sahalia and L.P. Hansen, pp. 691-766. North Holland, Elsevier.

[31] Rafty, A.E., Madigan, D. and Hoeting, J.A. (1997) Bayesian Model Averaging for Linear Regression Models. Journal of the American Statistical Association 92, 179-191.

[32] Waggoner, D.F. and Zha, T. (2012) Confronting Model Misspecification in Macroeconomics. Journal of Econometrics 171, 167-184.

[33] Wright, J.H. (2011) Term Premia and Inflation Uncertainty: Empirical Evidence from an International Panel Dataset. American Economic Review 101, 1514-1534.

A Appendix

A.1 Bond pricing

For notational convenience, we will take $\mu^Q = \mu$ and $\Phi^Q = \Phi$ as the risk-neutral parameters and all expectations are under the risk neutral measure Q.

A.1.1 ATSM

The n-period zero coupon bond price can be formulated as

$$\begin{aligned} P_t^n &= \mathbb{E}_t\left[e^{-r_t} P_{t+1}^{n-1}\right] \\ &= \mathbb{E}_t\left[\exp\left(-r_t + A_{n-1} + B_{n-1}^T X_{t+1}\right)\right], \end{aligned}$$

where

$$r_t = \delta_0 + \delta_1^T X_t.$$

and X_t follows the VAR dynamics $X_{t+1} = \mu + \Phi X_t + \Sigma\varepsilon_{t+1}$ with $\varepsilon_t \sim N(\mathbf{0}, \mathbf{I})$. We can substitute the expession of X_{t+1} such that

$$\begin{aligned} P_t^n &= \mathbb{E}_t\left[\exp\left(-r_t + A_{n-1} + B_{n-1}^T X_{t+1}\right)\right] \\ &= \exp\left(-r_t + A_{n-1} + \mu + B_{n-1}^T \Phi X_t\right) \\ &\quad \times \mathbb{E}_t\left[\exp\left(B_{n-1}^T \Sigma\varepsilon_{t+1}\right)\right]. \end{aligned}$$

Then, we can make use of the moment generating function of $\varepsilon \sim N(\mathbf{0}, \mathbf{I})$ to compute the expectation as

$$\mathbb{E}_t\left[\exp\left(B_{n-1}^T \Sigma\varepsilon\right)\right] = \exp\left[\frac{1}{2} B_{n-1} \Sigma\Sigma^T B_{n-1}^T\right].$$

by collecting separately the constant terms and linear terms in X_t, we obtain the recursive relationship for ATSM.

A.1.2 QTSM

The n-period zero coupon bond price can be formulated as

$$\begin{aligned} P_t^n &= \mathbb{E}_t\left[e^{-r_t} P_{t+1}^{n-1}\right] \\ &= \mathbb{E}_t\left[\exp\left(-r_t + A_{n-1} + B_{n-1}^T X_{t+1} + X_{t+1}^T C_{n-1} X_{t+1}\right)\right], \end{aligned}$$

where

$$r_t = \alpha_0 + \beta_0^T X_t + X_t^T \Psi_0 X_t,$$

and X_t follows the VAR dynamics $X_{t+1} = \mu + \Phi X_t + \Sigma\varepsilon_{t+1}$ with $\varepsilon_t \sim N(\mathbf{0}, \mathbf{I})$. Similarly, we substitute the expession of X_{t+1} such that

$$(\mu + \Phi X_t + \Sigma\varepsilon_{t+1})^T C_{n-1} (\mu + \Phi X_t + \Sigma\varepsilon_{t+1}) = 2(\mu + \Phi X_t)^T C_{n-1}\Sigma\varepsilon_{t+1},$$

and hence

$$\begin{aligned} P_t^n &= \exp\left(-r_t + A_{n-1} + B_{n-1}^T\mu + \Phi X_t + (\mu + \Phi X_t)^T C_{n-1}(\mu + \Phi X_t)\right) \\ &\quad \times \mathbb{E}_t\left[\exp\left(\Gamma_0^T\varepsilon_{t+1} + \varepsilon_{t+1}^T\Gamma_1\varepsilon_{t+1}\right)\right], \end{aligned}$$

where

$$\Gamma_0^T = B_{n-1}^T\Sigma + 2(\mu + \Phi X_t)^T C_{n-1}\Sigma, \qquad \Gamma_1 = \Sigma^T C_{n-1}\Sigma.$$

In this case, we can make use of the (exponential) quadratic-form expectation for $\varepsilon \sim N(\mathbf{0}, \mathbf{I})$ as

$$\mathbb{E}_t\left[\exp\left(\Gamma_0^T\varepsilon + \varepsilon^T\Gamma_1\varepsilon\right)\right] = \exp\left[-\frac{1}{2}\det(\mathbf{I} - 2\Gamma_1) + \frac{1}{2}\Gamma_0(\mathbf{I} - 2\Gamma_1)^{-1}\Gamma_0\right].$$

See, for example, Chapter 12 in Andersen and Piterberg (2010). Therefore,

$$\begin{aligned} &\mathbb{E}_t\left[\exp\left(\Gamma_0^T\varepsilon_{t+1} + \varepsilon_{t+1}^T\Gamma_1\varepsilon_{t+1}\right)\right] \\ = \;& \exp\left(-\frac{1}{2}\det\left(\mathbf{I} - 2\Sigma^T C_{n-1}\Sigma\right)\right) \\ & \times \exp\left(\left(B_{n-1}^T\Sigma + 2(\mu + \Phi X_t)^T C_{n-1}\Sigma\right)\left(\mathbf{I} - 2\Sigma^T C_{n-1}\Sigma\right)^{-1}\right. \\ & \left.\left(B_{n-1}^T\Sigma + 2(\mu + \Phi X_t)^T C_{n-1}\Sigma\right)^T\right), \end{aligned}$$

collecting separately the constant terms, linear terms in X_t and quadratic terms in X_t, we obtain the recursive relationship for QTSM.

A.2 Bayesian Estimation of Macro-Finance Models

A.2.1 State space formulation

In this subsection, we discuss the Bayesian estimation procedure in more details. First-of-all, it is useful to express more explicitly the state space model in Section 3.4 as follows:

- Measurement equation. Factor loadings a_n, b_n, and c_n are derived from the recursive relationship as described in Section 3.3. The measurement equations for the observable bond yields $\hat{y}_t^n$ and macro variables $\hat{X}_t$ are related to the latent factors X_t as

$$\hat{X}_t = X_t + \omega_{X,t},$$

and

$$\hat{y}_t^n = a_n + b_n^T X_t + X_t^T c_n X_t + \omega_{n,t}.$$

Formally, this can be stacked into one equation and expressed as

$$\underbrace{\begin{bmatrix} \hat{f}_t \\ \hat{g}_t \\ \hat{\pi}_t \\ --- \\ y_t^1 \\ \vdots \\ y_t^n \\ \vdots \\ y_t^N \end{bmatrix}}_{(M+N)\times 1} = \underbrace{\begin{bmatrix} 0 \\ 0 \\ 0 \\ --- \\ a_1 \\ \vdots \\ a_n \\ \vdots \\ a_N \end{bmatrix}}_{(M+N)\times 1} + \underbrace{\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ -- & -- & -- \\ b_{1,1} & b_{2,1} & b_{3,1} \\ \vdots & \vdots & \vdots \\ b_{1,n} & b_{2,n} & b_{3,n} \\ \vdots & \vdots & \vdots \\ b_{1,N} & b_{2,N} & b_{3,N} \end{bmatrix}}_{(M+N)\times M} \underbrace{\begin{bmatrix} f_t \\ g_t \\ \pi_t \end{bmatrix}}_{M\times 1} + \begin{bmatrix} f_t & g_t & \pi_t \end{bmatrix} \odot \underbrace{\begin{bmatrix} \mathbf{0_{M\times M}} \\ \mathbf{0_{M\times M}} \\ \mathbf{0_{M\times M}} \\ --- \\ \mathbf{c_1} \\ \vdots \\ \mathbf{c}_n \\ \vdots \\ \mathbf{c_N} \end{bmatrix}}_{M(M+N)\times M} \odot \begin{bmatrix} f_t \\ g_t \\ \pi_t \end{bmatrix} + \underbrace{\begin{bmatrix} \omega_{r,t} \\ \omega_{y,t} \\ \omega_{\pi,t} \\ -- \\ \omega_{y1,t} \\ \vdots \\ \omega_{yn,t} \\ \vdots \\ \omega_{yN,t} \end{bmatrix}}_{(M+N)\times 1},$$

where $\hat{X}_t = \left(\hat{f}_t,\, \hat{g}_t,\, \hat{\pi}_t\right)$ is the observable state vector of macro variables with measurement errors ω_{it} and $X_t = (f,\, g_t,\, \pi_t)$ is the unobservable state vector. Here, M and N denote the numbers of macro variables and yields respectively. The third term of RHS represents the quadratic multiplication where $\mathbf{0_{M\times M}}$ and $\mathbf{c}_n$ are $M \times M$ matrices. When we set the matrix $\mathbf{c}_n = \mathbf{0_{M\times M}}$, the QTSM reduces to the ATSM and we have a linear state-space model.

- State equation. The state equation with the parameters μ^P and Φ^P is given by

$$X_{t+1} = \mu^P + \Phi^P X_t + \Sigma\varepsilon_{t+1},$$

which can be expressed as

$$\begin{bmatrix} f_{t+1} \\ g_{t+1} \\ \pi_{t+1} \end{bmatrix} = \begin{bmatrix} \mu_1 \\ \mu_2 \\ \mu_3 \end{bmatrix} + \begin{bmatrix} \phi_{11} & \phi_{12} & \phi_{12} \\ \phi_{12} & \phi_{12} & \phi_{12} \\ \phi_{12} & \phi_{12} & \phi_{12} \end{bmatrix} \begin{bmatrix} f_t \\ g_t \\ \pi_t \end{bmatrix} + \begin{bmatrix} \varepsilon_{f,t+1} \\ \varepsilon_{g,t+1} \\ \varepsilon_{\pi,t+1} \end{bmatrix}.$$

The equation is a standard VAR(1) system.

A.2.2 MCMC algorithm

As can be seen from the measurement equation, the state space model is non-linear so that we have adopted *MH within Gibbs* with single-move sampler for unobservable macro variables $X_t = (f_t,\, g_t,\, \pi_t)$, following the Bayesian procedure in Ang et al. (2011).

The algorithm of MCMC based on Ang et al. (2011) is consist of the following five steps.

- **Step 1: Drawing the latent factor** $X_t = (f_t,\, g_t,\, \pi_t)$**.** We adopt the single-move sampler and generate the latent factors using random walk MH with the conditional posterior

density:

$$P(X_t|X_{t-1},\tilde{Y},\Theta) \propto P(X_t|X_{t-1})P(\tilde{Y}_t|\, X_t,\, \Theta)P(X_{t+1}|X_t)$$

where

$$P(X_t|X_{t-1},\Theta) \propto \exp(-\frac{1}{2}(X_t - \mu^P - \Phi^P X_{t-1})^T(\Sigma\Sigma^T)^{-1}(X_t - \mu^P - \Phi^P X_{t-1})\,)$$

and

$$P(\widetilde{Y}_t|X_t,\Theta) \propto \left(-\frac{1}{2}\sum_n \left[\frac{\left(\widetilde{y}_t^n - \left(a_n + b_n^T X_t + X_t^T c_n X_t\right)\right)^2}{\sigma_n^2}\right]\right)$$

where $\tilde{Y}_t$ is observable variables including yields and macro variables and Θ is pamameters. The standard deviation of the random walk MH step is taken to be 0.0001 (i.e., 1 bps).

- **Step 2: Drawing μ^P and Φ^p under the real-world measure P.** We use the Gibbs sampler to sample μ^P and Φ^p with the conditional posterior density

$$P(\mu^P, \Phi^P|\Theta_-,\, X,\, \tilde{Y}) \propto P(X|\mu^P, \Phi^P, \Sigma)P(\mu^P, \Phi^P)$$

where $P(X|\mu^P, \Phi^P, \Sigma)$ is the likelihood function and $P(\mu^P, \Phi^P)$ is the prior (see Del Negro and Schorfheide, 2010).

- **Step 3: Drawing $\Sigma\Sigma'$, the variance of state equation.** We take the inverse Wishart distribution as the prior and sample from the proposal density

$$q(\Sigma\Sigma') = P(X|\mu, \Phi, \Sigma)P(\Sigma\Sigma'),$$

where $P(X|\mu, \Phi, \Sigma)$ and $P(\Sigma\Sigma')$ are the likelihood function and prior, respectively. A proposal draw is then accepted with the probability

$$\alpha = \min\left\{\frac{P(\tilde{Y}|(\Sigma\Sigma')^{m+1}, \Theta_-, X)}{P(\tilde{Y}|(\Sigma\Sigma')^{m}, \Theta_-, X)},\, 1\right\},$$

where $P(\tilde{Y}|(\Sigma\Sigma')^{m+1}, \Theta_-, X)$ is the likelihood function.

- **Step 4: Drawing $\mu^Q$and Φ^Q under the risk-neutral measure Q.** We use the random walk MH algorith and sample $\mu^Q$and Φ^Q from a proposal draw using the random walk process $x^m = x^{m-1} + \varepsilon^m$, where m is iteration and $\varepsilon^m \sim N(0, \sigma^2)$. A proposal draw is then accepted with the probability

$$\alpha = \min\left\{\frac{P(\tilde{Y}|(\mu^Q, \Phi^Q)^{m+1}, \Theta_-, X)}{P(\tilde{Y}|(\mu^Q, \Phi^Q)^{m}, \Theta_-, X)},\, 1\right\},$$

where $P(\tilde{Y}|(\mu^Q, \Phi^Q)^{m+1}, \Theta_{-,}X)$ is the likelihood function or the posterior density as we assume a flat prior as in Ang et al. (2011). The standard deviation of the random walk MH step is taken to be 0.1% of the magnitude of the initial parameters.

- **Step 5: Drawing the variance of measurement error (σ_u).** We take the inverted Gamma distribution as prior with $IG(0, 0.0025^2)$ in order to sample σ_u.

Although we adopt the single-move sampler for the non-linear state space model as in step 1, it is noted that an alternative is to estimate the model using the particle filter: Andreasen et al. (2013) estimate a two-factor QTSM using particle filter with the maximum likelihood estimation. However, we find that one needs to spend an extensive computational time to estimate our 3-factor QTSM when the filter is used along with the Bayesian estimation.

A.2.3 Short rate specification

It is important to note that we do not estimate explicitly the loading coefficients for the ATSM and QTSM. This allows us to avoid identification problem (as our macro factors are observed with errors) and also a more efficient estimation on the model parameters. We follow Ang et al. (2011) to pre-set the initial loading coefficients such that the moments of the bond yields and macro factors are consistent. We take $\delta_0 = 0$ and $\delta_1 = (1, 0, 0)$ for the ATSM and the QTSM, and set Ψ_0 to a diagonal matrix for the QTSM.

A.2.4 Optimal pooling

We describe below the MCMC procedures relate to the three optimal pooling schemes. In the method of static prediction pooling, the random walk Metropolis-Hasting (MH) algorithm is adopted to sample posterior of constant weighting λ. In the Markov-switching prediction pooling, we use a MH within Gibbs algorithm in which the regime s_t at period t is sampled by a single-move sampling as proposed by Albert and Chib (1993), with simultaneously sampled posterior estimates of the weighting $\lambda(s_t)$ under regime s_t. In the dynamic prediction pooling, the particle filter is used following Del Negro et al. (2013), in which the autocorrelation coefficient ρ is set to be a constant as $\rho = 0.9$.

Table 1: Posterior estimates of the model parameters for ATSM. The reported values for the parameters μ and $(\Sigma\Sigma^T)_{ij}$ are multiplied by 10,000.

	Mean	10 Percentile	90-Percentile	Std. Dev.
	VAR(1)-system under P-measure			
Φ_{11}	0.9335	0.9278	0.9392	0.0044
Φ_{12}	-0.0542	-0.0684	-0.04	0.0111
Φ_{13}	0.0102	0.0087	0.0117	0.0011
Φ_{21}	0.0391	0.0344	0.0438	0.0037
Φ_{22}	0.9447	0.9332	0.9563	0.0091
Φ_{23}	-0.0083	-0.0095	-0.0071	0.0009
Φ_{31}	0.0652	0.014	0.1177	0.0406
Φ_{32}	-0.0888	-0.2172	0.0405	0.1017
Φ_{33}	0.939	0.9256	0.9524	0.0105
μ_1	3.0324	2.4853	3.5871	0.4318
μ_2	-0.3067	-0.7539	0.1457	0.352
μ_3	-2.9073	-7.9218	2.0844	3.9352
	VAR(1)-system under Q-measure			
Φ_{11}	0.9668	0.9632	0.9701	0.0027
Φ_{12}	0.0594	0.0555	0.0647	0.0035
Φ_{13}	0.0053	0.0045	0.0062	0.0006
Φ_{21}	0.0515	0.0487	0.0548	0.0024
Φ_{22}	1.0023	0.9931	1.0117	0.0072
Φ_{23}	-0.0107	-0.0114	-0.01	0.0005
Φ_{31}	0.1173	0.1096	0.1257	0.0075
Φ_{32}	0.0495	0.0362	0.0602	0.009
Φ_{33}	0.6082	0.5152	0.6865	0.0604
μ_1	-0.2169	-0.3628	-0.0719	0.1133
μ_2	-0.1703	-0.2062	-0.1386	0.0272
μ_3	-1.9503	-2.3147	-1.5617	0.3034
	Variance Matrix			
$(\Sigma\Sigma^T)_{11}$	0.0006	0.0005	0.0007	0.0001
$(\Sigma\Sigma^T)_{12}$	-0.0002	-0.0003	-0.0002	0
$(\Sigma\Sigma^T)_{13}$	0.0019	0.0015	0.0024	0.0003
$(\Sigma\Sigma^T)_{21}$	-0.0002	-0.0003	-0.0002	0
$(\Sigma\Sigma^T)_{22}$	0.0004	0.0003	0.0005	0.0001
$(\Sigma\Sigma^T)_{23}$	-0.0007	-0.001	-0.0005	0.0002
$(\Sigma\Sigma^T)_{31}$	0.0019	0.0015	0.0024	0.0003
$(\Sigma\Sigma^T)_{32}$	-0.0007	-0.001	-0.0005	0.0002
$(\Sigma\Sigma^T)_{33}$	0.0469	0.0366	0.0587	0.0082

Notes:

1. The first 10,000 draws of MCMC sampling are discarded to guarantee convergence and then the next 20,000 draws are used for calculating the posterior means, the standard deviations (Std. Dev.), as well as the 10 and 90 percentiles.
2. The posterior mean is computed by averaging the MCMC draws.
3. Std. Dev. is computed as the sample standard deviation of the MCMC draws.

Table 2: Posterior estimates of the model parameters for QTSM. The reported values for the parameters μ and $(\Sigma\Sigma^T)_{ij}$ are multiplied by 10,000.

	Mean	10 Percentile	90-Percentile	Std. Dev.
	VAR(1)-system under P-measure			
Φ_{11}	0.9333	0.9296	0.9369	0.0029
Φ_{12}	-0.0553	-0.0641	-0.0464	0.007
Φ_{13}	0.0103	0.0093	0.0112	0.0007
Φ_{21}	0.0386	0.0356	0.0415	0.0023
Φ_{22}	0.9462	0.939	0.9534	0.0057
Φ_{23}	-0.0081	-0.0089	-0.0073	0.0006
Φ_{31}	0.0543	0.023	0.0858	0.0246
Φ_{32}	-0.0621	-0.1411	0.0161	0.0614
Φ_{33}	0.9457	0.9374	0.9539	0.0065
μ_1	3.0869	2.7378	3.4331	0.2722
μ_2	-0.3333	-0.6168	-0.0466	0.2227
μ_3	-2.9874	-6.0233	0.0744	2.3951
	VAR(1)-system under Q-measure			
Φ_{11}	0.9755	0.9725	0.9789	0.0024
Φ_{12}	0.0593	0.0553	0.0632	0.003
Φ_{13}	0.0113	0.0101	0.0123	0.0008
Φ_{21}	0.0306	0.026	0.0343	0.0031
Φ_{22}	1.0066	1.0003	1.0127	0.0049
Φ_{23}	-0.012	-0.0125	-0.0116	0.0003
Φ_{31}	0.1333	0.1032	0.1545	0.0179
Φ_{32}	0.0611	0.0471	0.0773	0.0123
Φ_{33}	0.5369	0.4699	0.6016	0.052
μ_1	-0.2153	-0.3379	-0.0937	0.0886
μ_2	-0.2613	-0.3242	-0.1864	0.0492
μ_3	-2.068	-2.2524	-1.8879	0.1435
	Variance Matrix			
$(\Sigma\Sigma^T)_{11}$	0.0002	0.0002	0.0003	0
$(\Sigma\Sigma^T)_{12}$	-0.0001	-0.0001	-0.0001	0
$(\Sigma\Sigma^T)_{13}$	0.0007	0.0006	0.0009	0.0001
$(\Sigma\Sigma^T)_{21}$	-0.0001	-0.0001	-0.0001	0
$(\Sigma\Sigma^T)_{22}$	0.0002	0.0001	0.0002	0
$(\Sigma\Sigma^T)_{23}$	-0.0003	-0.0004	-0.0003	0.0001
$(\Sigma\Sigma^T)_{31}$	0.0007	0.0006	0.0009	0.0001
$(\Sigma\Sigma^T)_{32}$	-0.0003	-0.0004	-0.0003	0.0001
$(\Sigma\Sigma^T)_{33}$	0.017	0.0134	0.0214	0.0029

Notes:

1. The first 10,000 draws of MCMC sampling are discarded to guarantee convergence and then the next 20,000 draws are used for calculating the posterior means, the standard deviations (Std. Dev.), as well as the 10 and 90 percentiles.
2. The posterior mean is computed by averaging the MCMC draws.
3. Std. Dev. is computed as the sample standard deviation of the MCMC draws.

Table 3: Cumulative log scores

Four-quarter-ahead forecast

Component Models		Model Pooling	
Model	Log Score	Methods	Log Score
ATSM	3659.72	Static	3659.73
QTSM	3584.00	Dynamic (one parameter)	3661.02
		Dynamic (three parameters)	3665.63

Notes:

1. The predictive densities for the ATSM and QTSM are obtained by simulation using the MCMC draws of the posterior model parameters.

2. The cumulative log score is computed as

$$\sum_{t=1}^{T} \log \left[\lambda_t p \left(y_t ; Y_{t-1}, \mathcal{M}_1 \right) + \left(1 - \lambda_t \right) p \left(y_t ; Y_{t-1}, \mathcal{M}_2 \right) \right]$$

   where $\mathcal{M}_1 = \text{QTSM}, \mathcal{M}_2 = \text{ATSM}$.

Table 4: Posterior estimates of the static prediction pool

(a) Full Sample Period (Four-quarter-ahead forecast)

Parameter	Prior	Mean	Std. Dev.	90 % Band	Inefficiency
λ	$G(0.05, 0.1)$	0.102	0.077	[0.008 0.250]	18.18

(b) After 2000:Q1 (Four-quarter-ahead forecast)

Parameter	Prior	Mean	Std. Dev.	90 % Band	Inefficiency
λ	$G(0.05, 0.1)$	0.480	0.225	[0.101 0.846]	85.73

Notes:

1. λ denotes the constant weighting coefficient determined in the following optimal prediction pool:

$$p\left(y_t; Y_{t-1}, \mathcal{M}\right) = \lambda p\left(y_t; Y_{t-1}, \mathcal{M}_1\right) + (1-\lambda)\, p\left(y_t; Y_{t-1}, \mathcal{M}_2\right), \qquad 0 \leq \lambda \leq 1,$$

   where $\mathcal{M}_1 = \text{QTSM}$ $\mathcal{M}_2 = \text{ATSM}$, and $p\left(y_t; Y_{t-1}, \mathcal{M}_1\right)$ denotes log prediction score.

2. The coefficient λ is estimated with MCMC simulation and obtained from 50,000 draws after discarding the first 20,000 draws. And the posterior means, the standard deviations (Std. Dev.), and the percentiles are derived from the sampled draws.

3. The simulation ineffciency statisctic is a useful diagnostic for measuring how well the chain mixes accroding to Kim, Shephard, Chib (1998). The statistic is derived from:

$$\hat{R}_{B_M} = 1 + \frac{2B_M}{B_M - 1} \sum_{i=1}^{B_M} K(\frac{i}{B_M})\hat{\rho}(i),$$

   where $\hat{p}(i)$ is an estimate of the autocorrelation at lag i of MCMC sampler, B_M represents the bandwidth and K the Parzen Kernel.

Table 5: Dynamic Prediction Pool

(a) One flexible parameter version

Parameter	Prior	Mean	Std. Dev.	90 % band		Inefficiency
ρ	$Beta(5,5)$	0.842	0.118	[0.607	0.965]	32.359

(b) Three flexible parameters version

Parameter	Prior	Mean	Std. Dev.	90 % band		Inefficiency
ρ	$Beta(5,5)$	0.643	0.141	[0.486	0.787]	7.93
μ	N(0.5,1)	0.414	0.041	[0.375	0.452]	151.40
σ	IG(1,10)	1.558	0.402	[1.094	1.969]	50.27

Notes:

1. For estimation of Dynamic prediction pool method, we conduct 10,000 MCMC iterations with 5,000 particles, the first 5,000 iterations are discarded.
2. In prior, U, $Beta$, N and IG stand for uniform, beta, normal and inverse gamma distributions, respectively.

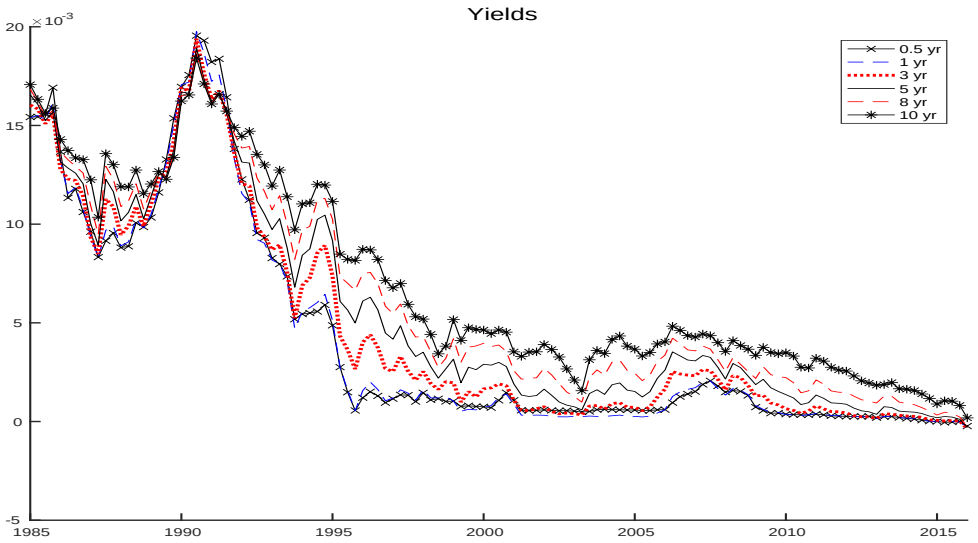
Table 6: Cumulative log scores

Four-quarter-ahead forecast

Component Models		Model Pooling	
Model	Log Score	Methods	Log Score
ATSM (GDP gap)	3659.72		
ATSM (Standard)	**3665.64**		
ATSM (Macro-Finance)	3473.10	Dynamic Pool with 6 models	3709.25
QTSM (GDP gap)	3584.00		
QTSM (Standard)	3553.84		
QTSM (Macro-Finance)	3450.86		

Figure 1: Data

(a) Yields



(b) Factors and Macro Variables

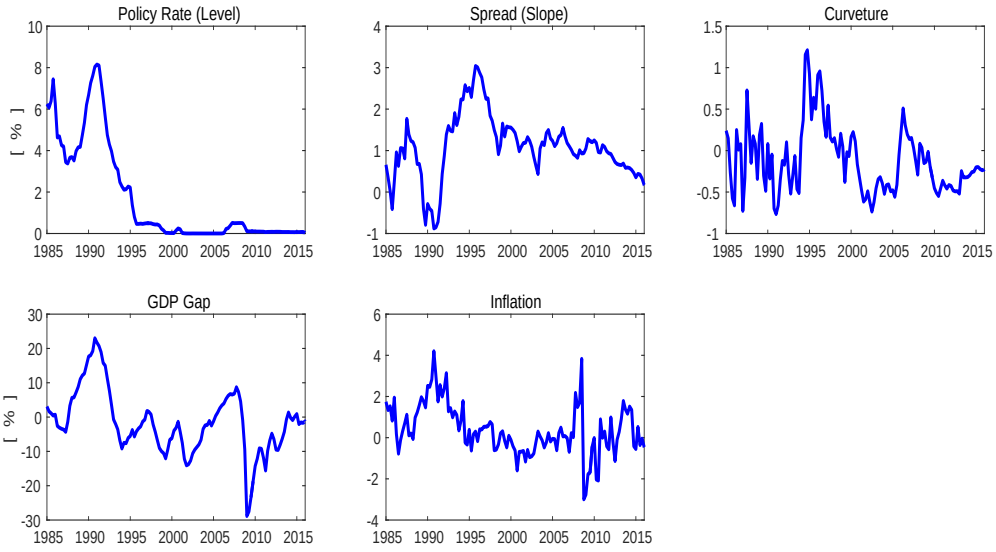
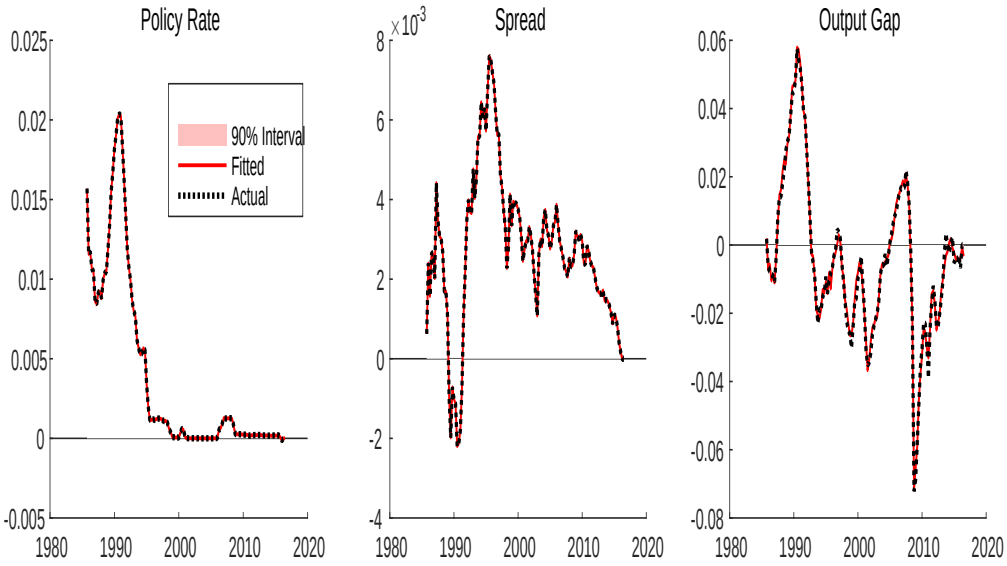
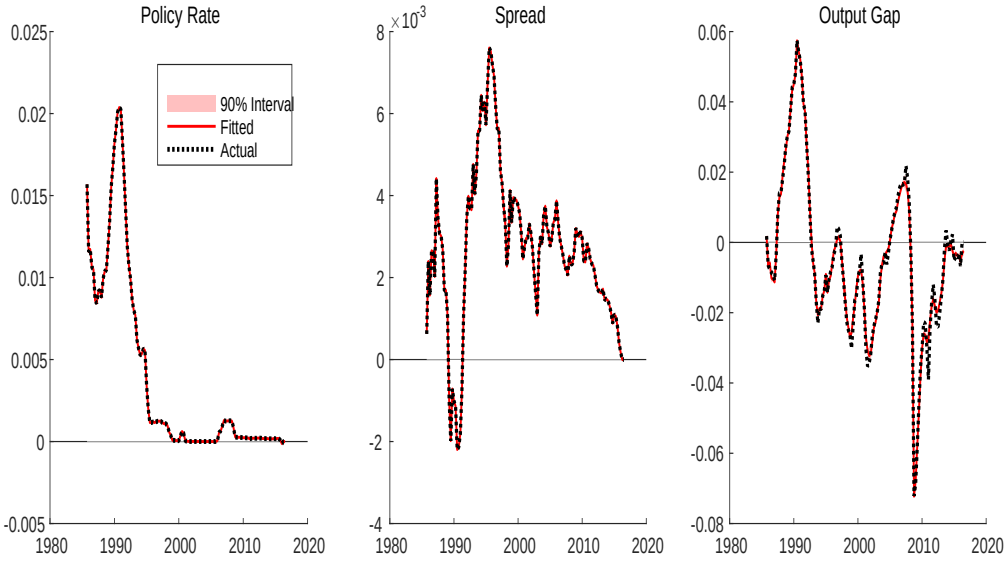


Figure 2: Estimation of Macro Factors

(a) ATSM



(b) QTSM



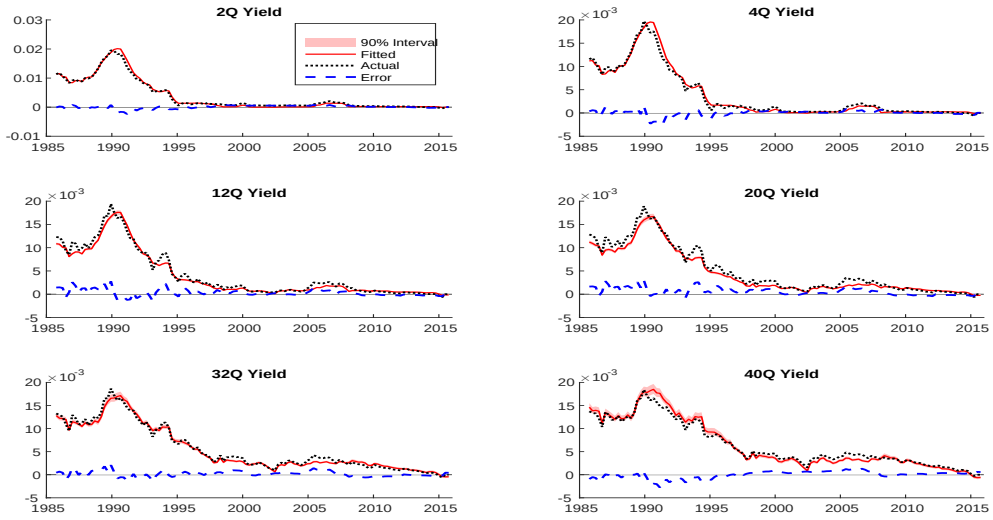
Notes:

1. The solid blue line and the dashed red represent actual values and fitted values respectively, the dashed blue line represents discrepancy between them, and the shaded grey band represents 90% confidence interval of the distribution.

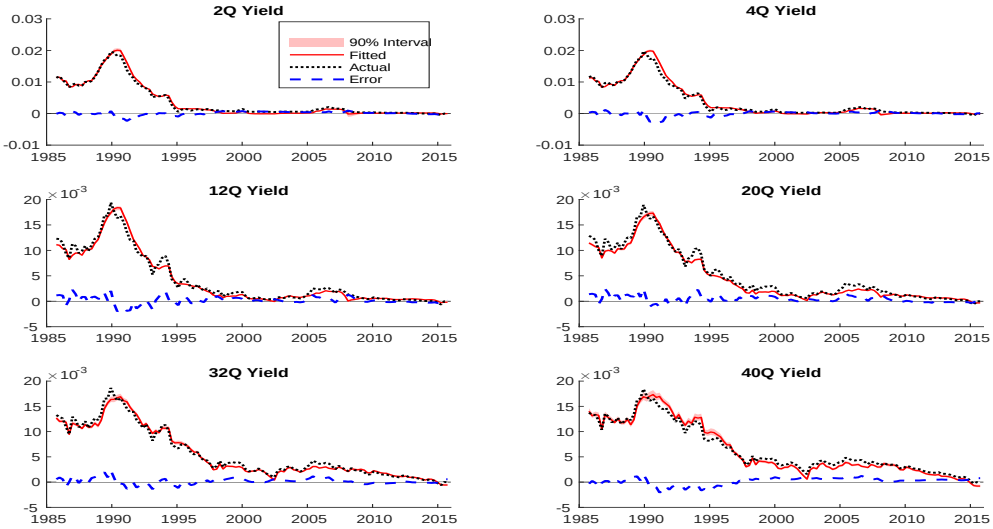
2. These estimations are obtained by 10,000 draws of MCMC samplings after discarding 5000 burn-in draws based on the Bayesian estimation described in the Appendix.

Figure 3: Estimation of Bond Yields

(a) ATSM



(b) QTSM



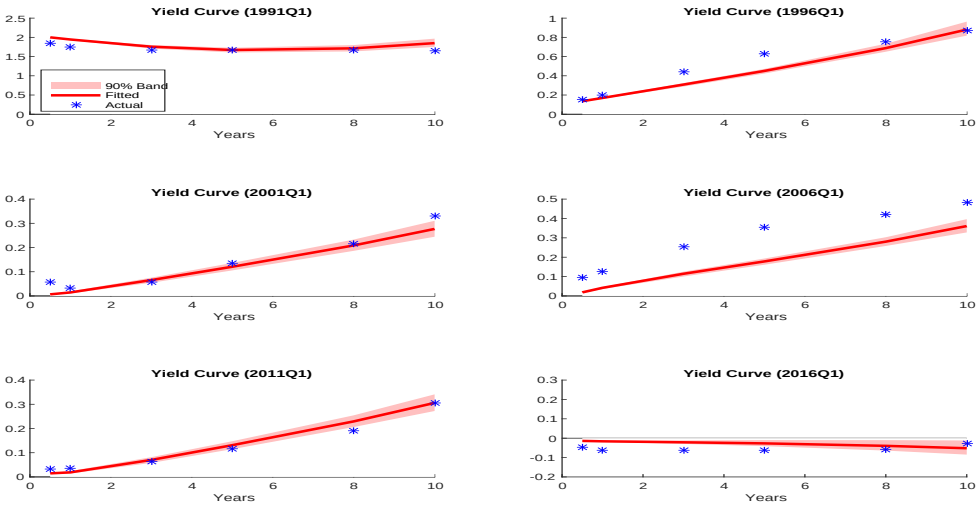
Notes:

1. The solid blue line and the dashed red represent actual values and fitted values respectively, the dashed blue line represents discrepancy between them, and the shaded grey band represents 90% confidence interval of the distribution.

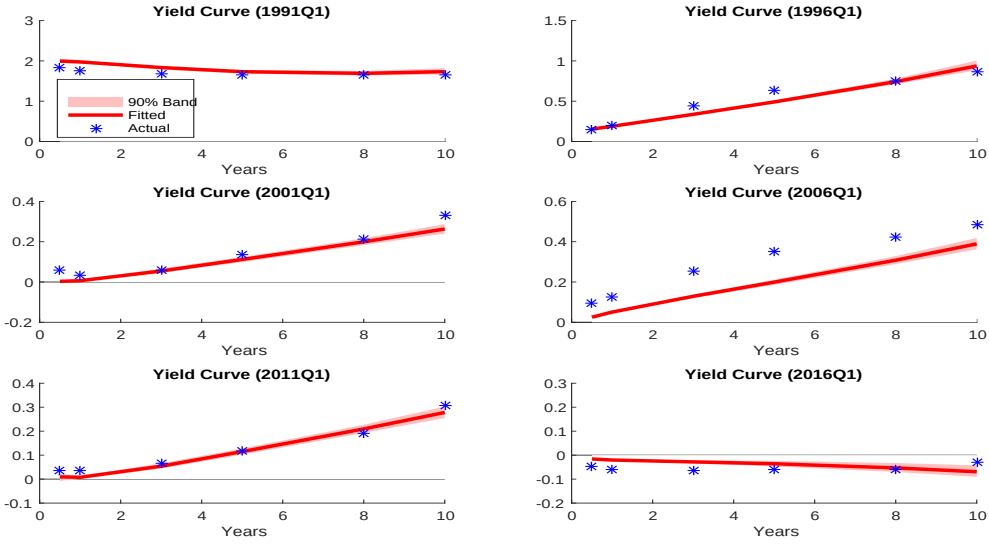
2. These estimations are obtained by 10,000 draws of MCMC samplings after discarding 5000 burn-in draws based on the Bayesian estimation described in the Appendix.

Figure 4: Estimation of Yield Curves

(a) ATSM



(b) QTSM



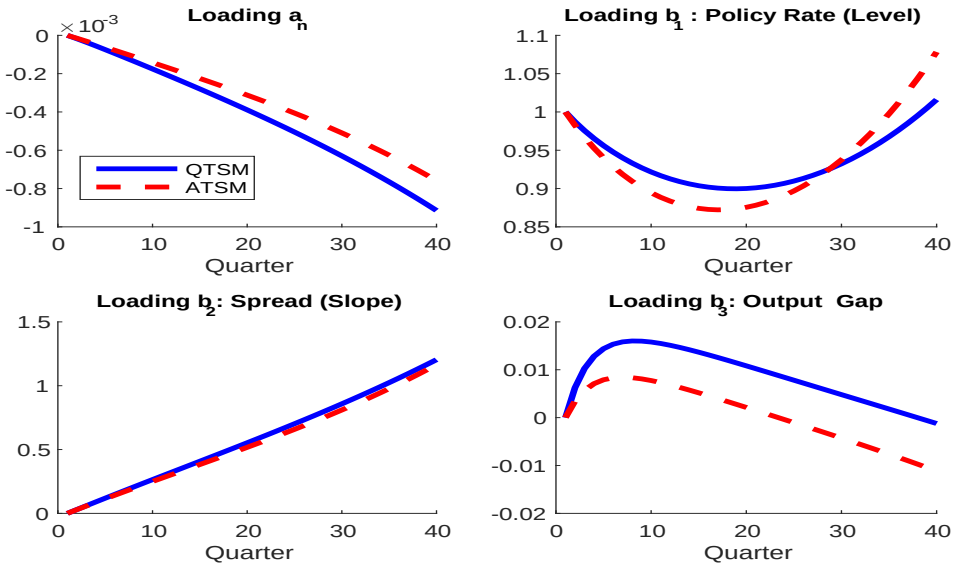
Notes:

1. The solid blue line and the dashed red represent actual values and fitted values respectively, the dashed blue line represents discrepancy between them, and the shaded grey band represents 90% confidence interval of the distribution.

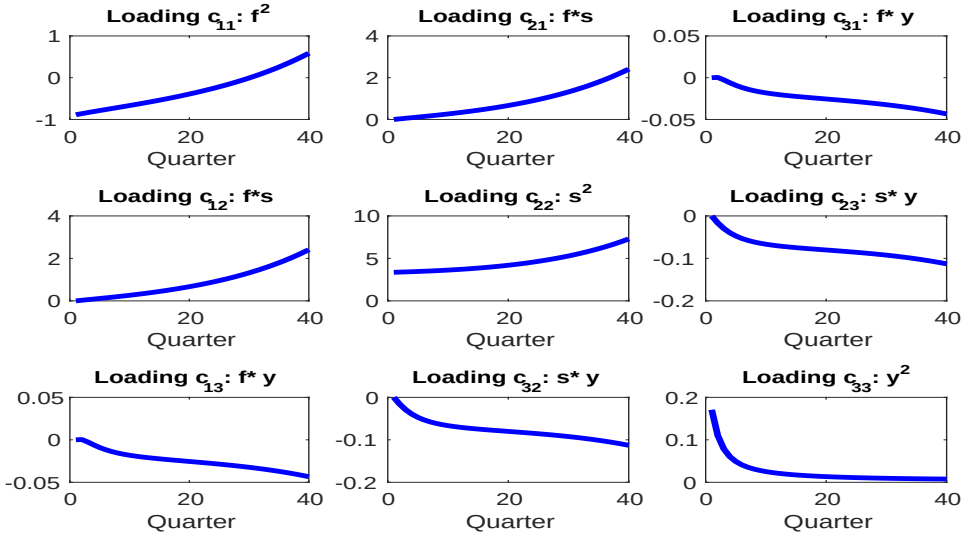
2. These estimations are obtained by 10,000 draws of MCMC samplings after discarding 5000 burn-in draws based on the Bayesian estimation described in the Appendix.

Figure 5: Factor loadings

(a) ATSM and QTSM
Factor loadings a_n and b_n



(b) QTSM
Factor loading c_n

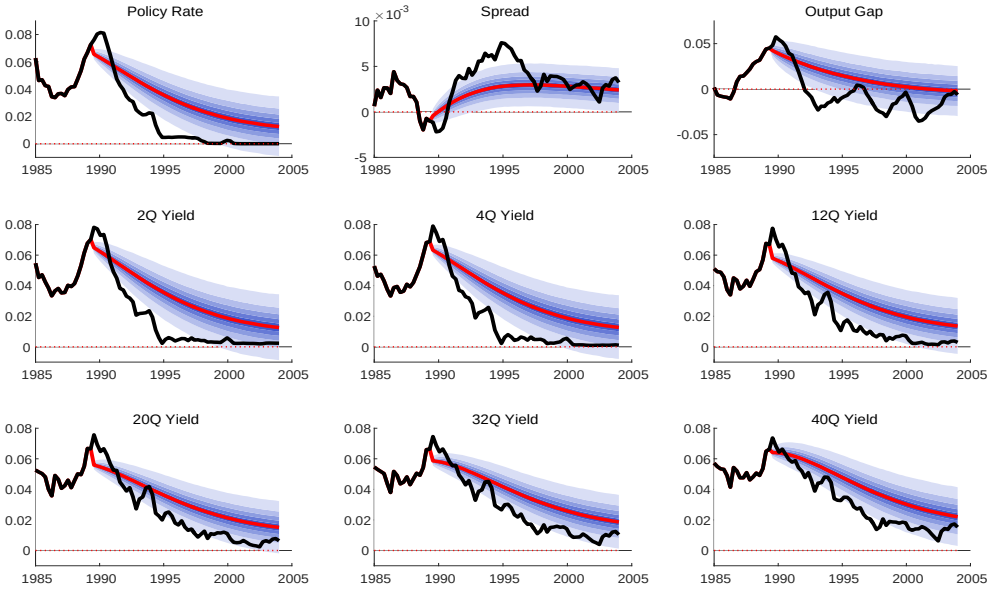


Notes:

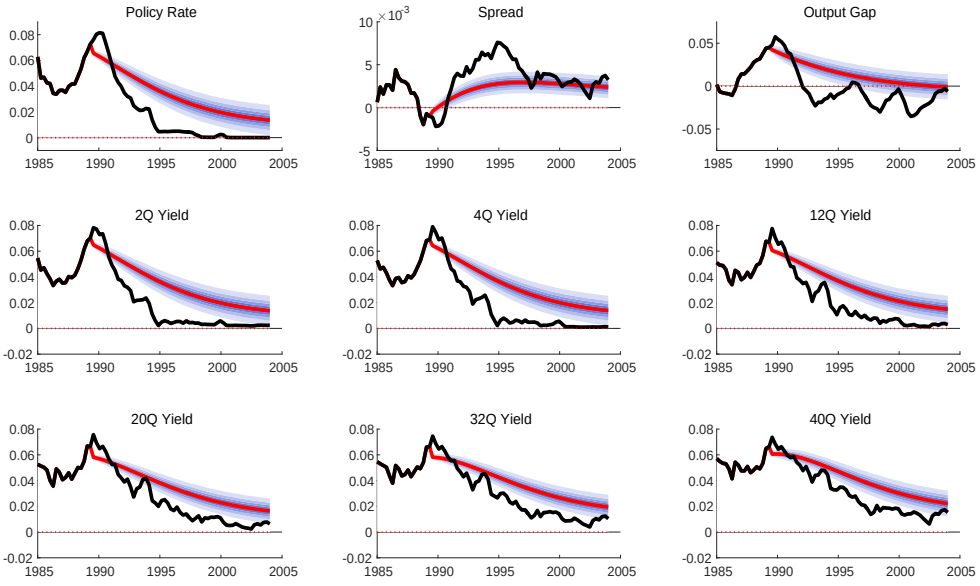
- The factor loadings of the estimated Gaussian ATSM are calculated using the recursive relationship (7) and the posterior means under Q-measure in Table 1.
- The factor loadings of the estimated QTSM are calculated using the recursive relationship (12) and the posterior means under Q-measure in Table 2.

Figure 6: Density Forecasts under Non-ZIRP (as at 1988:Q1)

(a) ATSM



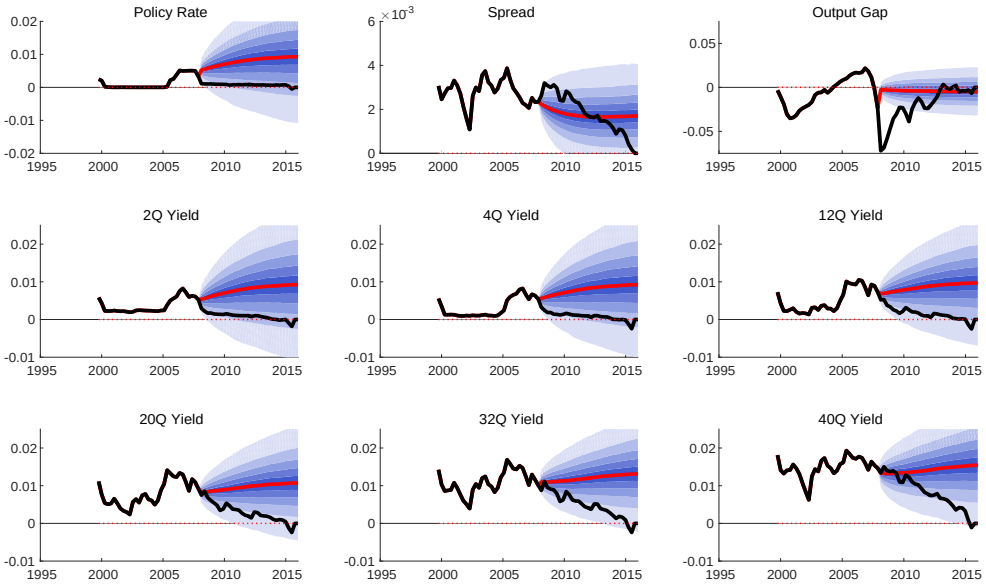
(b) QTSM



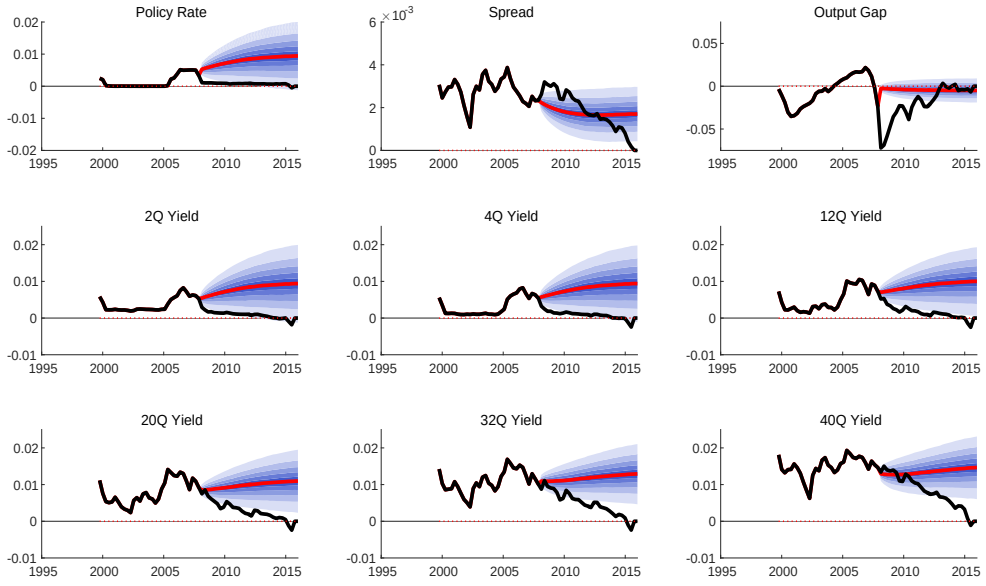
Note: The posterior prediction distributions of the macro factor and the JGB yields of the ATSM models are calculated based on the procedure as described in Section 2.1, using 10,000 draws of posterior estimates over the full sample as shown in Table 1.

Figure 7: Density Forecasts under ZIRP (as at 2008:Q1)

(a) ATSM



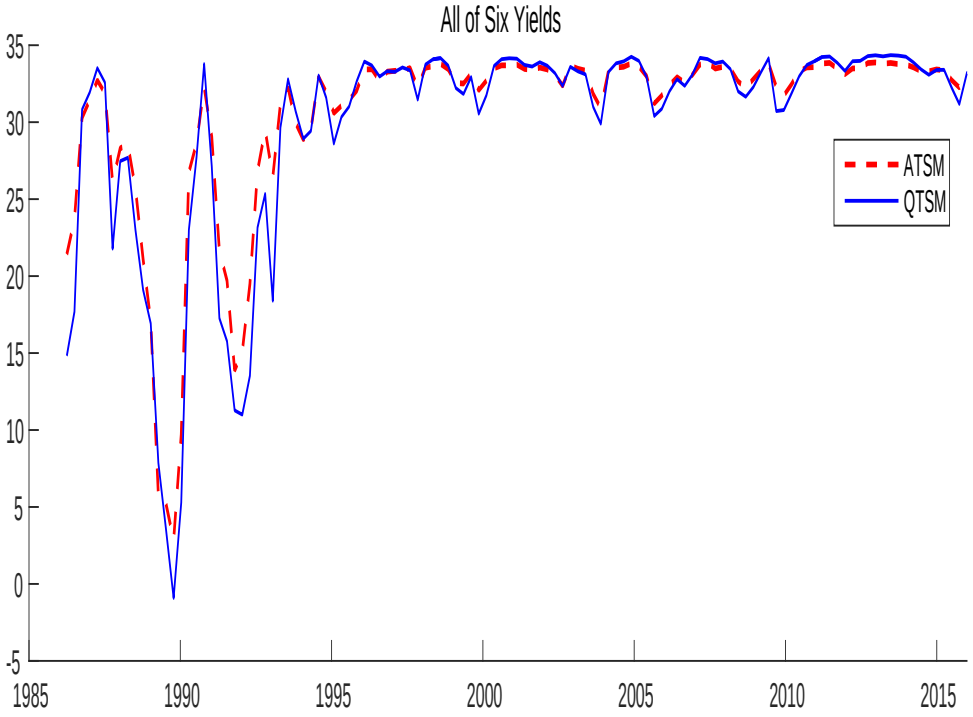
(b) QTSM



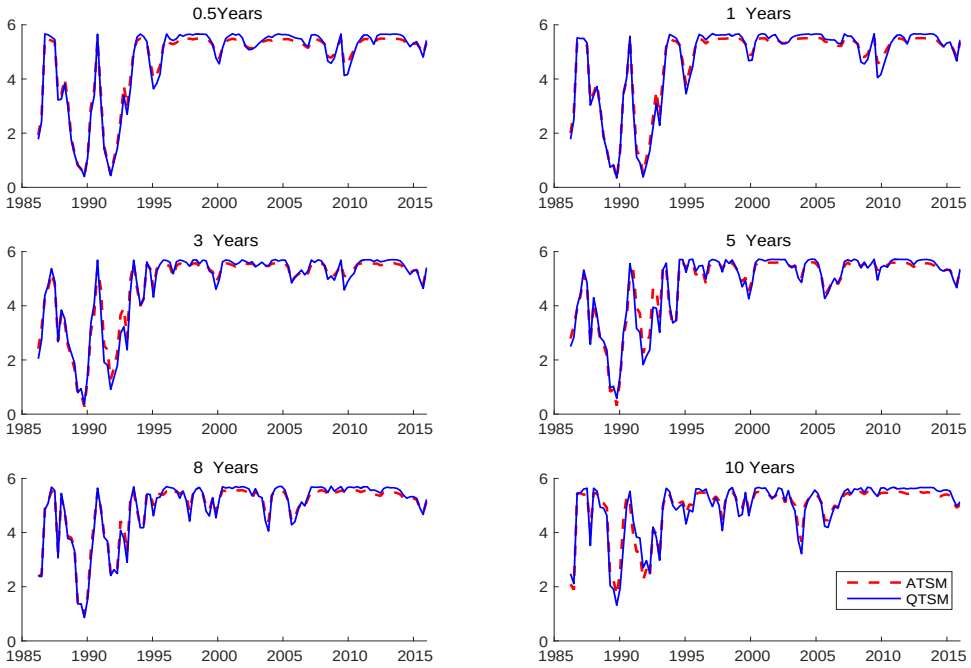
Note: The posterior prediction distributions of the macro factor and the JGB yields of the QTSM models are calculated based on the procedure as described in Section 2.1, using 10,000 draws of posterior estimates over the full sample as shown in Table 2.

Figure 8: Log score comparison of the ATSM and QTSM (based on 4Q-ahead forecast)

(a) All Six Yields



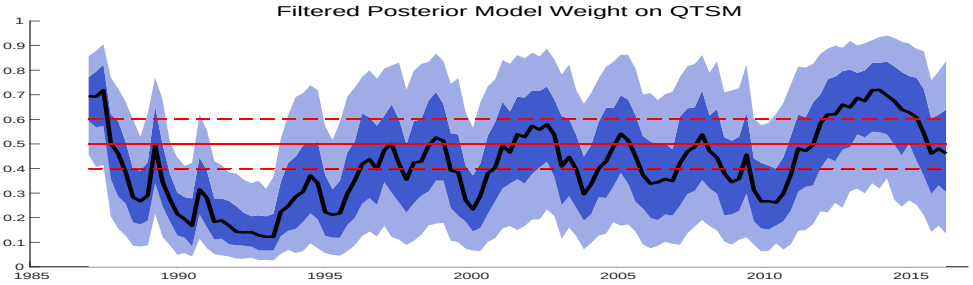
(b) Single Yields



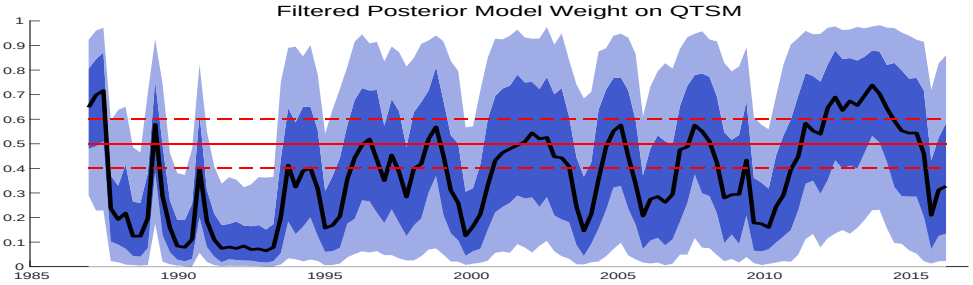
Note: The log score at each period is calculated from log $p(y_t^O; Y_{t-1}^O, M_i)$ for i = 1,2, of the individual model such as the QTSM and the ATSM as explained in Sec 4.3.

Figure 9: Dynamic prediction pool (4Q-ahead forecast)

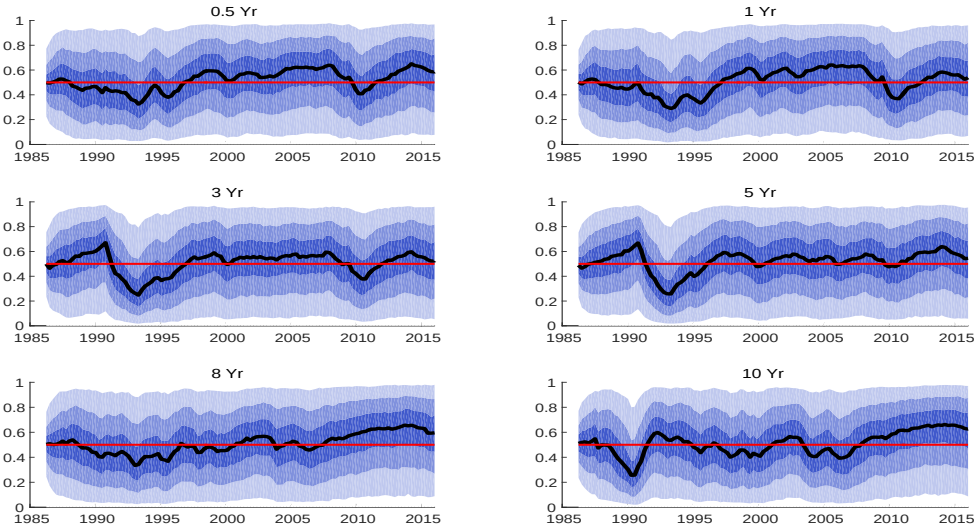
(a) All Six Yields (One Parameter Version)



(b) All Six Yields (Three Parameter Version)



(c) Single Yields (One Parameter Version)

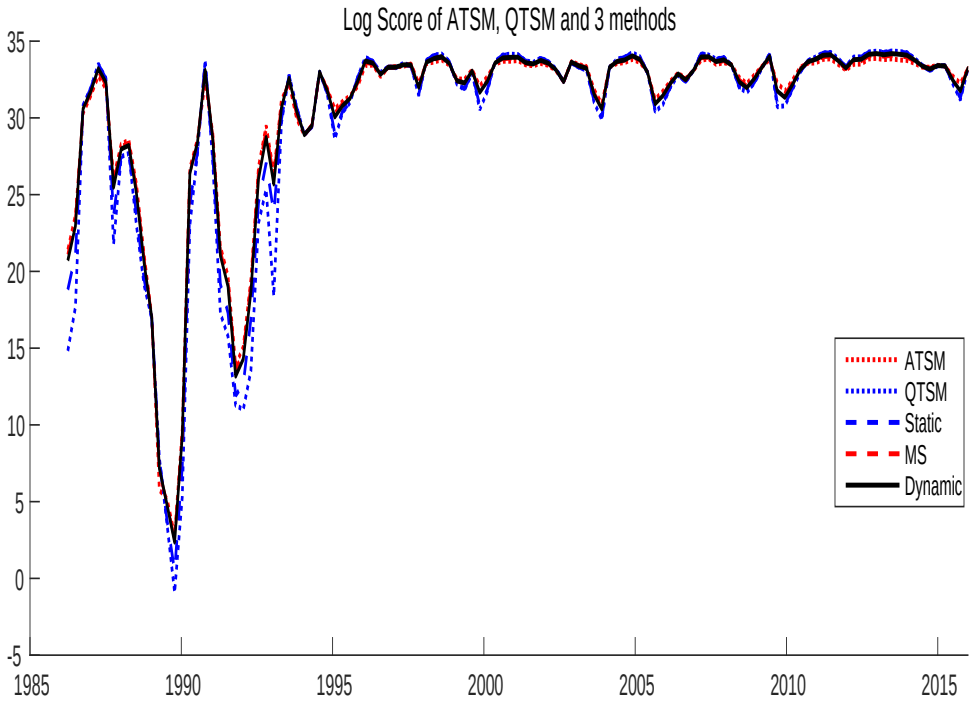


Notes:

1. Dynamic pooling model is calculated from Eq.(3). The time-varying coefficient is obtained from 5,000 draws of particle filter with estimating one parameter: ρ (panel (a)) and three parameters: ρ, μ, σ (panel (b)), following Del Negro et al (2013).

2. The solid black line denotes their posterior means and the blue shaded area represents their 90% confidence interval.

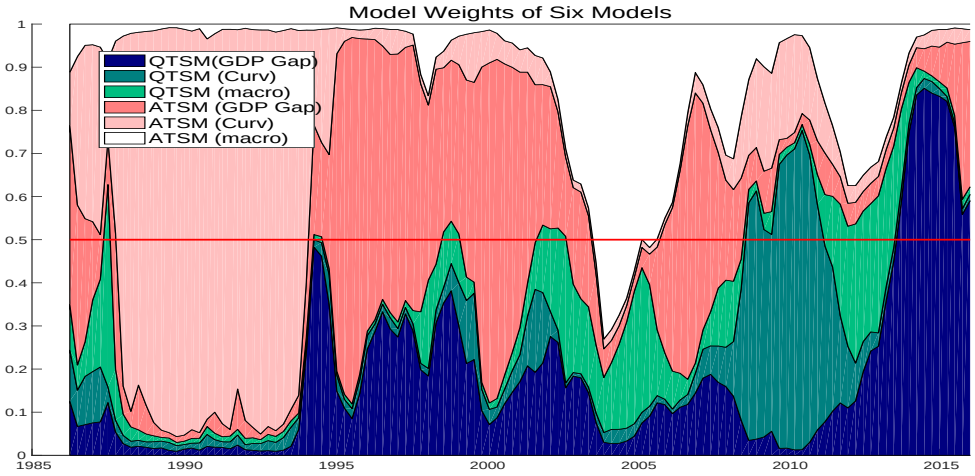
Figure 10: Log scores comparison of all models and pooling schemes (4Q-ahead forecast)



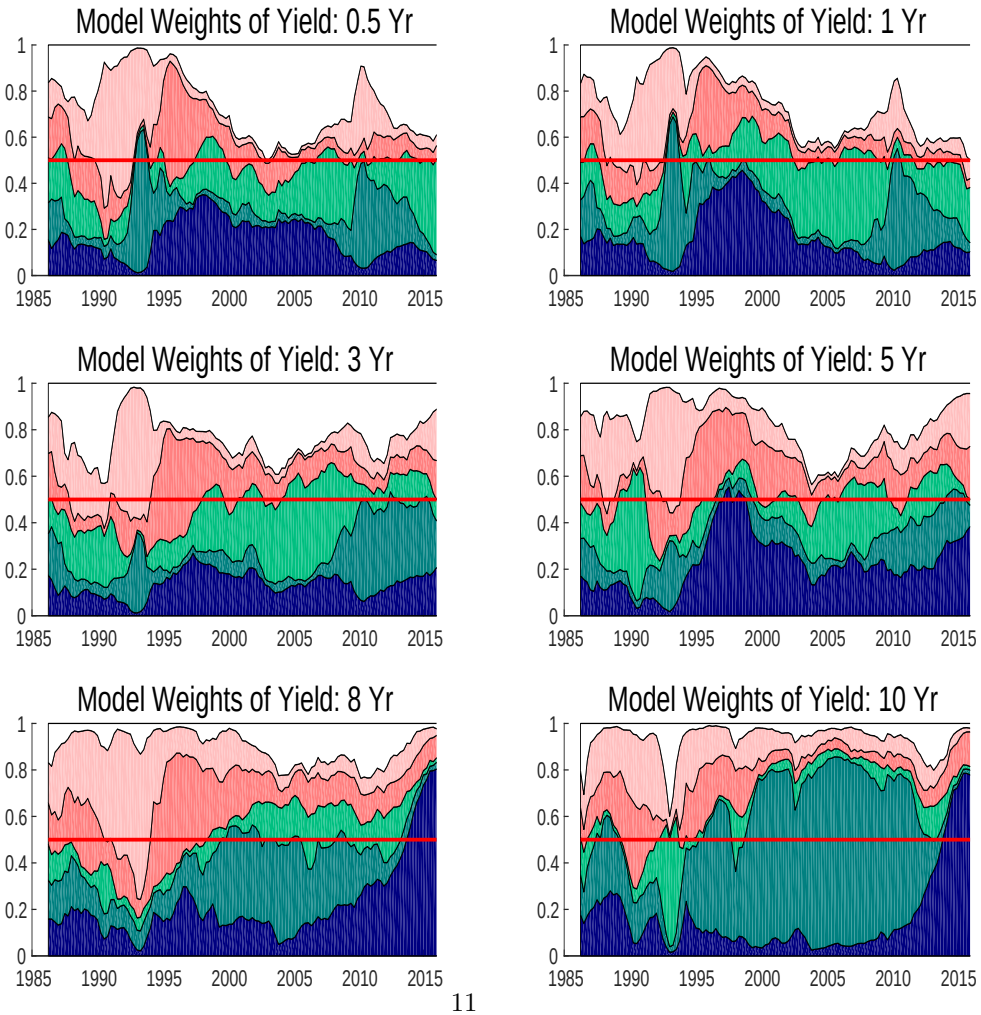
Note: The log scores at each period is calculated from log $p(y_t^O; Y_{t-1}^O, M_i)$ for i = 1,2, of the individual model such as the QTSM and the ATSM as explained in Sec 4.3. The log scores of the three optimal pooling methods are derived from Eq.(1), Eq.(2) and Eq.(3).

Figure 11: Model Weight of Six Models (4Q-ahead forecast)

(a) All Six Yields



(b) Each of Six Yields



Note: