Prediction of Term Structure with Potentially Misspecified Macro-Finance Models near the Zero Lower Bound

Chung, Tsz-Kin and Iiboshi, Hirokuni

Tokyo Metropolitan University, Economic and Social Research Institute, Cabinet Office

January 2015

Prediction of Term Structure with Potentially Misspecified Macro-Finance Models near the Zero Lower Bound*

Tsz-Kin Chung†
Tokyo Metropolitan University

Hirokuni Iiboshi‡
Tokyo Metropolitan University
Economic and Social Research Institute, Cabinet Office

January 2015

Abstract

In this paper, we study the forecasting performances of the affine term structure model (ATSM) and the quadratic term structure model (QTSM) with macro-finance features under the zero interest rate policy of Japan. As both the two models can be potentially misspecified, we adopt the optimal pooling prediction scheme following the recent work by 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. The Markov-switching prediction pool dominates individual models as well as the static and dynamic pools. Our results suggest that both of the ATSM and QTSM macro-finance models are potentially misspecified and one should use a combination of the two models for the prediction of future bond yields during different time periods. Our analysis sheds light on the macro-finance modeling using US data amid the Federal Reserve's zero interest rate policy since December 2008.

Keywords: Term structure; Forecasting; 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.

†E-mail: btkchung@gmail.com (Correspondence Author)

‡E-mail: iiboshi@tmu.ac.jp

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 a number of papers 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: it does not constraint the interest rate to be non-negative.[1] This may be problematic for the prediction of future bond yields when interest rates are very close to the zero lower bound, such as the cases of the Japanese government bond (JGB) yields since 1995 and the US treasury yields after the financial crisis of 2008. 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 naturally accommodates non-negative interest rates. Indeed, the quadratic models have been widely adopted by market participants in the pricing and hedging of interest rate derivatives given its nice analytical tractability and the guarantee of

[1] It is noted that another class of ATSM built on the square root process (Cox et al., 1985) is not suitable for macro-finance modelings 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).

non-negative model-implied interest rates (Piterberg, 2005; Kijima et al., 2009; Piterberg and Andersen, 2011). However, there have been very few formal empirical studies on the QTSM in particular its performance under the zero interest rate policy. Until recently, Kim and Singleton (2012) and Andreasen and Meldrum (2013) demonstrate the strength of the QTSM model in the statistical description of the yield curve data for the JGB yields and the US treasury yields, respectively.

In this paper, we study the pooling prediction of the future bond yields (term structure) of the Gaussian ATSM and QTSM with macro-finance features. The contribution is two-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. Secondly, we attempt to derive a better forecast of the term structure by combining the advantages of the two models on hand. The idea of optimal prediction pool is pioneered by Geweke and Amisano (2011, 2012) in which potentially misspecified models are pooled together in order to improve the prediction density using a log score criteria. Under the context of macro-finance modeling, the two models are potentially misspecified in different aspects:

1. *The Gaussian ATSM is naturally linked to the macroeconomic variables in a linear fashion and empirical studies have shown that it provides a good-fit to the macroeconomic variables and bond yields simultaneously. However, the model does not constraint non-negative interest rates which can be problematic when the zero lower bound is binding.*

2. *The QTSM naturally precludes negative interest rates. However, it is not clear whether the enforced non-linear mapping of the bond yields to the macroeconomic variables would provide a good fit to the data.*

We adopt three 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 two later approaches with time-varying weighting coefficient allows us to investigate the relative goodness in forecasting of the ATSM and QTSM during different sample periods (Waggoner and Zha, 2012; Del Negro et al., 2013). A related paper to ours is Eo and Kang (2014) who consider the model combination of the dynamic Nelson-Siegel model (DNSM) and ATSM using latent factors.[2] We differ from Eo and Kang (2014) in focusing on the macro-finance modeling and the combination of two class of models in which the prediction densities are substantially different when interest rates are near the zero lower bound.

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 simultaneouly. 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

[2]The DNSM can be described as a sub-class of the Gaussian ATSM with certain parameter restrictions.

since late 1995. This indicates the importance to take into account the zero lower bound when interest rates are low. As both the ATSM and QTSM with macro-finance features can be potentially misspecified, it is recommended that one should use a combination of the two models in the prediction of future bond yields. Although this paper focuses on the JGB data, the empirical results shed light on the future research on macro-finance modeling using the US data given the Federal Reserve's zero interest rate policy since December 2008.

The paper is structured as follows. Section 2 reviews the methods of prediction pools in the recent literature. Section 3 presents the ATSM and QTSM with macro-finance features and disucsses the data and estimation procedure. Section 4 reports the estimation results and findings. Section 5 concludes.

2 Methods of Prediction Pooling

2.1 Motivation

From a Bayesian perspective, the *marginal likelihood* is commonly used as a criteria of model choice because it is interpreted as the prediction distribution after integrating the prior density with respect to the model parameters. Let us denote y_t a vector time series, and its history as $Y_{t-1}^O = \{y_h, ..., y_{t-1}\}$ where $h \leq 1$ is a starting date and the superscript "O" denotes the observed data. The marginal likelihood is given by

$$p^{Prior}\left(y_t|Y_{t-1}^O, \mathcal{M}\right) = \int p\left(y_t|Y_{t-1}^O, \Theta, \mathcal{M}\right) p(\Theta|\mathcal{M}) d\Theta,$$

where $\mathcal{M}$ is a prediction model, $p\left(y_t|Y_{t-1}, \Theta, \mathcal{M}\right)$ and $p(\Theta|\mathcal{M})$ denote the likelihood function and the prior density of parameters Θ on a specified model $\mathcal{M}$, respectively. If we regard the marginal likelihood as a *prior* prediction distribution as noted in Geweke (2010), we can propose a *posterior* prediction distribution by replacing the prior density $p(\Theta|\mathcal{M})$ using the posterior density $p(\Theta|Y_{t-1}, \mathcal{M})$ as

$$p^{Post}\left(y_t|Y_{t-1}^O, \mathcal{M}\right) = \int p\left(y_t|Y_{t-1}^O, \Theta, \mathcal{M}\right) p(\Theta|Y_{t-1}^O, \mathcal{M}) d\Theta,$$

where $p(\Theta|Y_{t-1}, \mathcal{M})$ is the posterior density of parameters conditional on a specified model $\mathcal{M}$. Following Geweke and Amisano (2011), we use the posterior prediction distribution to construct a *log prediction scoring rule* in order to evaluate the forecasting performances using individual models as well as the calculation of the optimal weights when two models are combined for forecasting. We set up the prediction score of observation y_t^O at period t from the posterior prediction density as

$$p(y_t^O;\, Y_t^O,\, \mathcal{M}) = p^{Post}\left(y_t^O|Y_{t-1}^O, \mathcal{M}\right),$$

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

Recently, statistical approaches to model combination for forecasting and decision making have been paid attenstions, and one of them is Bayesian model averaging (BMA) proposed by Raftery et al. (1997) and Hoeting et al. (1999), in which a predictive distribution for future observation y_t is obtained by averaging a set of competing models and given as

$$p^{BMA}\left(y_t \,|\, Y_{t-1}\right) = \sum_{i=1}^{m} p(y_t^O;\, Y_t^O,\, \mathcal{M}_i)\, p\left(\mathcal{M}_i \,|\, Y_{t-1},\, S_{t-1}, \Theta\right), \text{ for } i = 1, \cdots, m,$$

where $p\left(\mathcal{M}_i \,|\, Y_{t-1},\, S_{t-1}, \Theta\right)$ is a posterior model probability (or a posterior model weights). The key of the BMA is that the posterior model probability is derived from $\frac{p^{Prior}\left(y_t|Y_{t-1}^O,\mathcal{M}_i\right)}{\sum_{i=1}^{m} p^{Prior}\left(y_t|Y_{t-1}^O,\mathcal{M}_i\right)}$ using the marginal likeloohood. However, this value makes one problem such that the posterior probability of one of competing models is often set as a extreme value such as almost 100% and those of others almost 0%. It indicates that forecating of BMA is likely to be same as that of one model selected from the marginal likeloohood, and that it is not preferable to use the marginal likeloohood as the weighting coefficient of forecasting future observations. Instead, we will consider what model weights or functions should be used as the posterior model weights to improve performance of forecasting empirically in the optimal prediction pooling. The aim of this study is to examine various model choices and combinations in terms of the macro-finance term structure modeling using new model weighting criteria. To this end, we will conduct three proposed approaches of prediction pool which we explain as follows.

2.2 Static prediction pooling

Firstly, we discuss the static pooling based on a constant weighting as in Geweke and Amisano (2011). Given two prediction models $\mathcal{M}_1$ and $\mathcal{M}_2$, the pool of the prediction density can be constructed as the convex combination

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

with $\mathcal{M} = (\mathcal{M}_1, \mathcal{M}_2)$ is the collection of models and $p\left(y_t^O; Y_{t-1}^O, \mathcal{M}_i\right) = p\left(y_t^O; Y_{t-1}^O, \hat{\Theta}_i\right)$ is the prediction density with the posterior estimates $\hat{\Theta}_i$ of the model $\mathcal{M}_i$. The optimal prediction pooling is then obtained by maximizing the cumulative log prediction score, LPS^{SP}, as

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

by choosing $\lambda^* = \arg\max LPS^{SP}(\lambda)$. An important assumption as noted in Geweke and Amisano (2011) is that the two candidate prediction models have to be substantially different in terms of the functional form of their predictive densities (i.e., non-nested models). In our case, the Gaussian ATSM generates a prediction density of bond yield close to a normal

distribution, while the QTSM generates an asymmetric prediction density which is bounded explicitly from zero (Kim and Singleton, 2012).

2.3 Markov-switching prediction pooling

Waggoner and Zha (2012) extend the static prediction pool by allowing the weighting coefficient λ_t to be dependent on a regime variable s_t following a Markov chain as

$$\lambda_t = \lambda(s_t) = \begin{cases} \lambda_1, & s_t = 1 \\ \lambda_2, & s_t = 2 \end{cases},$$

in which the transition matrix Q for s_t is given by

$$Q = \begin{bmatrix} q_{11} & q_{12} \\ q_{21} & q_{22} \end{bmatrix},$$

where $q_{ij} = \Pr(s_t = j | s_{t-1} = i)$ with $q_{11} + q_{12} = 1$ and $q_{21} + q_{22} = 1$. Conditional on the state s_t, the pool of the prediction density can be expressed as

$$p\left(y_t^O;\, Y_{t-1}^O, \mathcal{M},\, s_t\right) = \lambda(s_t)\, p\left(y_t^O; Y_{t-1}^O, \mathcal{M}_1\right) + (1 - \lambda(s_t))\, p\left(y_t^O; Y_{t-1}^O, \mathcal{M}_2\right),$$

with $\mathcal{M} = (\mathcal{M}_1, \mathcal{M}_2)$. Hence, integrating out the unobservable regime s_t, we have the pooled prediction density as

$$p^{MS}\left(y_t^O; Y_{t-1}^O, \mathcal{M}\right) = \sum_{s_t=1}^{2} p\left(y_t^O; Y_{t-1}^O, \mathcal{M},\, s_t\right) p(s_t | s_{t-1})\, p\left(s_{t-1} | Y_{t-1}^O,\, \mathcal{M}\right).$$

where $p(s_t | s_{t-1})$ is the transition probability of the Markov chain. Using the Bayes theorem, posterior conditional density of s_{t-1}, $p\left(s_{t-1} | Y_{t-1}^O,\, \mathcal{M}\right)$, can be shown to be equal to $p(y_{t-1} | s_{t-1}, Y_{t-2,}^O \mathcal{M}) \times p(s_{t-1} | s_{t-2}) \times p\left(s_{t-2} | Y_{t-2}^O, \mathcal{M}\right)$, where $p(y_{t-1}^O | s_{t-1}, Y_{t-2,}^O \mathcal{M})$ is a likelihood function of y_{t-1}^O given s_{t-1} and Y_{t-2}^O. It also indicates that $p\left(s_{t-1} | Y_{t-1}^O,\, \mathcal{M}\right)$ is recursively obtained from the density of previous period's regime, $p\left(s_{t-1} | Y_{t-1}^O,\, \mathcal{M}\right)$, for $t = 1 \cdots T$.

From the above equation, we have the log prediction score of Markov-switching prediction pooling with two regimes as

$$LPS^{MS}(\lambda_1, \lambda_2) \equiv \sum_{t=1}^{T} \log\, p^{MS}\left(y_t^O; Y_{t-1}^O, \mathcal{M}\right). \tag{2}$$

An advantage of using the Markov-switching modeling for the weighting coefficient is that we can identify the relative importance of the models during different sample periods. Waggoner and Zha (2012) show that the DSGE model plays an important role relative to a BVAR model only in the late 1970s and the early 1980s. It is important to note that the we do not assume

regime-switching in the interest rate dynamics under the bond pricing model. Hence, the regime only reflects the particular times of history in which the model dominates in terms of its prediction ability.

2.4 Dynamic prediction pooling

Finally, we adopt the dynamic prediction pooling scheme as proposed in Del Negro et al. (2013). The idea is to generate a smooth time-varying weighting coefficient, $\lambda_t \in [0,1]$, based on a probit transformation of a AR(1) process of a latent variable x_t with the autocorrelation coefficient ρ as

$$\begin{aligned} \lambda_t &= N(x_t), \\ x_t &= \rho x_{t-1} + \sqrt{1-\rho^2}\varepsilon_t, \end{aligned}$$

where $N(\cdot)$ is the cumulative density function of standard normal distribution, the disturbance term follows $\varepsilon_t \sim N(0,1)$ and the initial value takes $x_0 \sim N(0,1)$. The autocorrelation coefficient ρ captures how smooth the weighting coefficient can 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 are serially-independent and follows a random walk. Then, we have the log score of dynamic prediction pooling as

$$LPS^{DP}(\lambda_t) \equiv \sum_{t=1}^{T} \log\left[\lambda_t\, p\left(y_t^O; Y_{t-1}^O, \mathcal{M}_1\right) + (1-\lambda_t)\, p\left(y_t^O; Y_{t-1}^O, \mathcal{M}_2\right)\right]. \tag{3}$$

Del Negro et al (2013) fixed the parameter ρ as 0.9 or 0.75 and estimated the time-varying weighting coefficient λ_t of nonlinear model with probit transformation using particle filtering. We refer the readers to Johannes and Polson (2009) for a survey on the application of particle filtering in financial econometrics.

2.5 Estimation strategy of prediction pooling

To estimate and compare the prediction pools using the above three methods, we adopt the Bayesian approach with Monte-Carlo Markov Chain (MCMC) for the static and Markov-switching approaches, and resort to the particle filtering for the dynamic prediciton pool (with a fixed parameters ρ). We adopt a two-step procedure as noted in Waggoner and Zha (2012) in which the posterior parameters are saved from the model estimation procedure and used for forecasting and computation of the prediction density/score. In the first step, posterior estimates of parameters, $p(\Theta|Y_{t-1}^O, \mathcal{M})$, under the prediction models, $\mathcal{M} = (\mathcal{M}_1, \mathcal{M}_2)$, are obtained using the MCMC method based on the full sample period in order to obtain the prediction scores $p\left(y_t^O; Y_{t-1}^O, \mathcal{M}\right)$ of the three methods. For each set of posterior estimate of the prediction

macro-finance models $\mathcal{M}$, we then compute the forecast and prediction density by simulation technique in order to make use of the entire joint posterior distribution. In the second step, we estimate the optimal combination using the log scores from the individual prediction models (ATSM and QTSM) obtained in the previous step, based on the aforementioned three prediction pooling methods, i.e., Eq.(1), Eq(2) and Eq.(3). The details on the MCMC procedure of the prediction macro-finance models are provided in the Appendix.

3 Macro-Finance Models

3.1 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.

3.2 Affine term structure model

The Gaussian ATSM is specified as

$$r_t = \delta_0 + \delta_1^T X_t, \tag{4}$$

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. 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.

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

$$P_t^n = \exp\left(A_n + B_n^T X_t\right), \tag{6}$$

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

$$\begin{aligned} A_n &= -\delta_0 + A_{n-1} + B_{n-1}^T \mu^Q + \frac{1}{2} B_{n-1} \Sigma\Sigma^T B_{n-1}^T, \\ B_n^T &= -\delta_1^T + \Phi^Q B_{n-1}^T, \end{aligned} \tag{7}$$

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

$$y_t^n = -\frac{1}{n} \log P_t^n = a_n + b_n^T X_t, \tag{8}$$

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

3.3 Quadratic term structure model

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

$$r_t = \alpha_0 + \beta_0^T X_t + X_t^T \Psi_0 X_t, \tag{9}$$

i.e., the one-period short rate is a quadratic function to the selected macroeconomic state variables. To ensure non-negative interest rate, it is common to set $\alpha_0 = 0$ and $\beta_0 = 0_M$ as

$$r_t = X_t^T \Psi_0 X_t, \tag{10}$$

such that the non-negativity of r_t is implied by the assumption that Ψ_0 is positive-semidefinite (Kim and Singleton, 2012).

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

$$P_t^n = \exp\left(A_n + B_n^T X_t + X_t^T C_n X_t\right), \tag{11}$$

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

$$\begin{aligned}
A_n &= -\alpha_0 + A_{n-1} + B_{n-1}^T \mu^Q + \mu^T C_{n-1} \mu^Q - \frac{1}{2} \det\left(\mathbf{I} - 2\Sigma^T C_{n-1} \Sigma\right) \\
&\quad + \frac{1}{2} \left(\Sigma^T B_{n-1} + 2\Sigma^T C_{n-1} \mu^Q\right)^T \left(\mathbf{I} - 2\Sigma^T C_{n-1} \Sigma\right)^{-1} \left(\Sigma^T B_{n-1} + 2\Sigma^T C_{n-1} \mu\right), \\
B_n^T &= -\beta_0^T + B_{n-1}^T \Phi^Q + 2\mu C_{n-1} \Phi^Q \\
&\quad + 2\left(\Sigma^T B_{n-1} + 2\Sigma^T C_{n-1} \mu\right)^T \left(\mathbf{I} - 2\Sigma^T C_{n-1} \Sigma\right)^{-1} \Sigma^T C_{n-1} \Phi^Q, \\
C_n &= -\Psi_0 + \left(\Phi^Q\right)^T C_{n-1} \Phi^Q + 2\left(\Sigma^T C_{n-1} \Phi^Q\right)^T \left(\mathbf{I} - 2\Sigma^T C_{n-1} \Sigma\right)^{-1} \left(\Sigma^T C_{n-1} \Phi^Q\right),
\end{aligned} \tag{12}$$

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

$$y_t^n = -\frac{1}{n} \log P_t^n = a_n + b_n^T X_t + X_t^T c_n X_t \tag{13}$$

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

when we set the initial loadings as $\alpha_0 = 0$ and $\beta_0 = 0_M$, the bond yield has a constant term loading and a linear term loading as a_n and b_n respectively.

3.4 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

  $$\hat{y}_t^n = a_n + b_n^T X_t + X_t^T c_n X_t + \omega_{n,t}, \tag{14}$$

  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 $\omega_{X,t}$:

  $$\hat{X}_t = X_t + \omega_{X,t}, \tag{15}$$

  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

  $$X_{t+1} = \mu^P + \Phi^P X_t + \Sigma\varepsilon_{t+1}, \tag{16}$$

  which is a standard VAR(1) system.

Therefore, (14), (15) and (16) 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.

3.5 Data

In this paper, we use the data in Wright (2011) for the JGB market during the sample period from 1990Q1 to 2008Q3. The JGB yield curve data is obtained from Datastream and the author's calculation based on the Svensson interpolation methodology. Following Diebold et al. (2006), we construct the macro-finance model with variables that represent the monetary policy instrument, level of real economic activity and inflation rate. The marcoeconomic variables are:

1. Measure of monetary policy (f_t): based on the Bank of Japan's uncollateralized overnight call rate.

2. Measure of real activity (g_t): based on the exponentially weighted moving average of quarterly GDP growth.

3. Measure of inflation (π_t): based on the exponentially weighted moving average of quarterly inflation.

As noted in Diebold et al. (2006), these selected marco variables $X_t = (f_t, y_t, \pi_t)$ are widely taken to be a set of fundamentals that capture the macroeconomic dynamics. A similar set of marcoeconomic variables have been employed in Ang and Piazzesi (2003) and Bernanke et al. (2005). The overnight call rate is obtained from the Bank of Japan's website, while the last two macroeconomic variables are obtained from the dataset of Wright (2011). To keep the consistency with previous empirical studies, we use the JGB yields of the 1, 4, 8, 12, 16 and 20 quarters.

To better understand the data and the estimation results, let us take a brief review on the Bank of Japan's monetary policy since the 1990s. The Bank of Japan 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 1995Q4 and 1998Q4 respectively. To further simulate the economy, the Bank of Japan adopted the zero interest rate policy (ZIRP) during the period from 1999Q1 to 2000Q3 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. Baba (2006) provides a comprehensive review of the Bank of Japan monetary policy and the JGB market development over the sample period.

4 Results

4.1 Model estimation

First-of-all, we look at the filtered macro factors 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 Figure 1 (a) and (b) show the filtered macro factors and the fitted bond yields of the model. These estimations are obtained by 10,000 draws of MCMC samplings after discarding the first 5000 burn-in draws based on the Bayesian methods described in the Appendix. 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. It can

be seen that macro factors track quite closely to the actual data and the fitting of the bond yields are reasonably good across maturities. This demonstrates that 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. This generates notable degree of pricing errors when the actual short-term bond yields are effectively zero. Figure 1 (c) depicts actual and estimated yields curve at specified four points including both of non-zero and zero interest rate policy periods. This graph indicates goodness of fit in terms of cross section aspect of time series of term structure of panel (b). Table 2 and Figure 2 report the corresponding results for the QTSM which imposes a quadratic mapping in between bond yields and macro factors. The filtered output and inflation factors as shown in Figure 2 (a) and (b) also track closely to the actual data, however, there is a substantial deviation of the monetary stance factor. The latter case indicates that the enforced quadratic mapping in the macro-finance QTSM can be potentially misspecified. In contrast to the ATSM, the model-implied bond yields of QTSM are guaranteed 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 2 (c) draws a counterpart of the QTSM for the yields curve of the ATSM in Figure 1 (c).

[Insert Figures 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 3 reports the factor loadings of the estimated Gaussian ATSM using the recursive relationship (7) and the posterior means under Q-measure in Table 1. Because the short-term interest rate is taken as one of the state variables, we can impose the initial loading of the inflation 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 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 monetary stance: 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 Figure 3.

Figure 4 reports the factor loadings of the estimated QTSM using the recursive relationship (12) and the posterior means under Q-measure in Table2. In contrast to the ATSM which only has 4 factor loadings, the QTSM provides in total 10 loading combinations to the three macroeconomic factors, including 4 loadings through the linear terms: a, b_1, b_2, b_3, and 6 load-

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	90 Percentile	10-Percentile	Std. Dev.
	VAR(1)-system under P-measure			
Φ_{11}	0.8672	0.9196	0.8128	0.0428
Φ_{12}	0.0438	0.1318	-0.0457	0.0696
Φ_{13}	0.1014	0.2287	-0.0246	0.1000
Φ_{21}	0.1210	0.2150	0.0326	0.0716
Φ_{22}	0.5443	0.7098	0.3652	0.1333
Φ_{23}	-0.1947	0.0048	-0.3901	0.1593
Φ_{31}	0.0397	0.1141	-0.0312	0.0572
Φ_{32}	0.0565	0.2055	-0.1206	0.1271
Φ_{33}	0.5287	0.6922	0.3699	0.1257
μ_1	-0.6343	2.6977	-3.9661	2.5966
μ_2	12.1554	19.1896	5.6553	5.2071
μ_3	1.0896	8.0413	-4.4578	4.9450
	VAR(1)-system under Q-measure			
Φ_{11}	0.9383	0.9674	0.9080	0.0230
Φ_{12}	0.0691	0.0833	0.0538	0.0107
Φ_{13}	0.0138	0.0182	0.0093	0.0033
Φ_{21}	0.1336	0.1685	0.1105	0.0213
Φ_{22}	0.8502	0.9220	0.7832	0.0521
Φ_{23}	-0.0329	0.1774	-0.2265	0.1671
Φ_{31}	0.1347	0.1586	0.1132	0.0173
Φ_{32}	0.0825	0.1017	0.0642	0.0146
Φ_{33}	0.4517	0.5255	0.3747	0.0570
μ_1	-0.3371	-0.0775	-0.5338	0.1782
μ_2	8.7984	9.8244	7.6819	0.8208
μ_3	-2.2497	-1.7340	-2.9599	0.4368
	Variance Matrix			
$(\Sigma\Sigma^T)_{11}$	0.0155	0.0201	0.0114	0.0034
$(\Sigma\Sigma^T)_{12}$	-0.0051	-0.0001	-0.0106	0.0041
$(\Sigma\Sigma^T)_{13}$	-0.0018	0.0031	-0.0073	0.0041
$(\Sigma\Sigma^T)_{21}$	-0.0051	-0.0001	-0.0106	0.0041
$(\Sigma\Sigma^T)_{22}$	0.0370	0.0476	0.0272	0.0082
$(\Sigma\Sigma^T)_{23}$	-0.0001	0.0073	-0.0078	0.0060
$(\Sigma\Sigma^T)_{31}$	-0.0018	0.0031	-0.0073	0.0041
$(\Sigma\Sigma^T)_{32}$	-0.0001	0.0073	-0.0078	0.0060
$(\Sigma\Sigma^T)_{33}$	0.0296	0.0442	0.0145	0.0116

Notes:

1. The first 5,000 draws of MCMC sampling are discarded to guarantee convergence and then the next 10,000 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	90 Percentile	10-Percentile	Std. Dev.
	VAR(1)-system under P-measure			
Φ_{11}	0.8131	0.8934	0.7320	0.0640
Φ_{12}	0.0568	0.1696	-0.0544	0.0872
Φ_{13}	0.2270	0.3758	0.0794	0.1173
Φ_{21}	0.0169	0.1213	-0.0911	0.0827
Φ_{22}	0.6470	0.7770	0.5121	0.1035
Φ_{23}	0.0068	0.2000	-0.1789	0.1462
Φ_{31}	0.0670	0.1666	-0.0133	0.0719
Φ_{32}	0.0590	0.1544	-0.0387	0.0774
Φ_{33}	0.7496	0.9090	0.5348	0.1463
μ_1	2.2411	6.3008	-1.8670	3.2158
μ_2	7.3223	12.0844	2.6005	3.7285
μ_3	-2.4398	0.8460	-6.0533	2.8044
	VAR(1)-system under Q-measure			
Φ_{11}	0.9942	0.9993	0.9892	0.0039
Φ_{12}	0.0392	0.0484	0.0333	0.0057
Φ_{13}	0.0187	0.0215	0.0162	0.0021
Φ_{21}	0.0951	0.1243	0.0718	0.0203
Φ_{22}	0.7553	0.8453	0.6581	0.0732
Φ_{23}	-0.2717	-0.2504	-0.2947	0.0165
Φ_{31}	0.1105	0.1302	0.0922	0.0132
Φ_{32}	0.0671	0.0952	0.0479	0.0166
Φ_{33}	0.4003	0.4472	0.3493	0.0372
μ_1	-1.2593	-1.0521	-1.3488	0.1056
μ_2	7.4830	7.9915	6.9087	0.4519
μ_3	-2.4335	-2.1773	-2.7327	0.2461
	Variance Matrix			
$(\Sigma\Sigma^T)_{11}$	0.0310	0.0407	0.0225	0.0072
$(\Sigma\Sigma^T)_{12}$	0.0027	0.0110	-0.0055	0.0065
$(\Sigma\Sigma^T)_{13}$	-0.0043	0.0011	-0.0101	0.0044
$(\Sigma\Sigma^T)_{21}$	0.0027	0.0110	-0.0055	0.0065
$(\Sigma\Sigma^T)_{22}$	0.0393	0.0553	0.0268	0.0116
$(\Sigma\Sigma^T)_{23}$	-0.0070	-0.0002	-0.0141	0.0054
$(\Sigma\Sigma^T)_{31}$	-0.0043	0.0011	-0.0101	0.0044
$(\Sigma\Sigma^T)_{32}$	-0.0070	-0.0002	-0.0141	0.0054
$(\Sigma\Sigma^T)_{33}$	0.0201	0.0349	0.0097	0.0097

Notes:

1. The first 5,000 draws of MCMC sampling are discarded to guarantee convergence and then the next 10,000 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.

ings through the quadratic terms : c_{11}, c_{22}, c_{33}, c_{12}, c_{13}, c_{23}.[3] To keep consistency with the setting of ATSM, we set the initial loadings to the output and inflation factors: Φ_{22} and Φ_{33}, and other off-diagonal elements in Eq.(9), to zero as shown in Figure 4(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 inflation and output uncertainty are high as shown in Figure 4(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, π_t) and (g_t, π_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 bond yields under the ZIRP, e.g., see the fitting of the 1Q yield in Figure 2(b).

[Insert Figures 3 and 4 around here]

4.2 Prediction of macro factors and bond yields

Before analysing the optimal prediction pool, we calculate the posterior prediction distributions of the macro factors and the JGB yields of the two models as described in Section 2.1, using 10,000 draws of posterior estimates over the full sample as shown in Table 1 and 2. Figure 5 shows the ATSM prediction of the macro factors and the JGB yield curve across 6 maturities for the following two forecasting periods: (i) 1992Q4 - 1998Q1 and (ii) 2003Q4 - 2008Q3. 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 5(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. When the Bank of Japan adopted the ZIRP and the QMEP in 2003Q4, 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 Figure 5(b). Even for the 5-year bond yield, there is a substantial probability of bleaching the

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

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.

Figure 6 shows the QTSM prediction of the macro factors and the bond yields during the corresponding two forecasting periods. For the in-sample prediction in 1992Q4 - 1998Q1, the QTSM produces a less accurate forecast as with the ATSM model as in Figure 6(a). Moreover, the prediction density 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 2003Q4 - 2008Q3 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 Figure 6(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 5 and 6 around here]

4.3 Prediction pool

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 Section 2, following Geweke and Amisano (2011), Waggoner and Zha (2012) and Del Negro et al. (2013). 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 and four-quarter-ahead forecasts, we only report the charts of four-quarter-ahead forecasts for exposition purpose. As shown in Figure 7, the ATSM (dashed red line) dominates the QTSM (solid blue line) for the sampling period from 1990Q1 - 1995Q4 while the QTSM dominates the ATSM when the JGB bond yields are close to zero since 1996Q1. This suggets that one can potentially improve the predictive density by combining appropriately the two models which appear to perform better in different sample periods, i.e., they capture different properties of the movements of bond yields and their interaction with the macroeconomy. To fix idea, recall that in Section 2 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.

[Insert Figures 7 around here]

4.3.1 Static pooling

Figure 8 reports the posterior density and corresponding trace plot of the weighting coefficient λ for the static pooling scheme as Eq.(1). This coefficient is estimated with MCMC simulation and obtained from 10,000 MCMC draws after discarding the first 5000 draws as burn-in. For the four-quarter-ahead forecast, the posterior distribution of λ is skewed to the left, indicating that the parameter restriction of $\lambda \leq 1$ is binding and one should over-weight the QTSM model and under-weight the ATSM. Table 3 reports the posterior mean of the pooling coefficient as $\lambda = 0.8627$ for the four-quarter-ahead forecast. We also compute the simulation inefficiency statistics as in Kim et al. (1998) and show that MCMC draws of the parameter λ are efficient.

[insert Figure 8 around here]

4.3.2 Markov-switching (MS) pooling

Figure 9 reports the estimation for the Markov-switching pooling scheme, Eq.(2), in which the extra model parameters include the transition matrix of the Markov chain and the corresponding weighting coefficients λ_1 and λ_2 under the two regimes $s_t = 1$ and $s_t = 2$. Figure 9(a) shows the posterior estimates of probability of the regime variable $s_t = 2$ conditional on data and prediction models (left-hand side) and the time-varying weight λ_t calculated from

Table 3: Posterior estimates of the static prediction pool

(a) One-quarter-ahead forecast

Parameter	Mean	90 Percentile	10 Percentile	Std. Dev.	Inefficiency
λ	0.2982	0.5466	0.0615	0.1480	79.035

(b) Four-quarter-ahead forecast

Parameter	Mean	90 Percentile	10 Percentile	Std. Dev.	Inefficiency
λ	0.8627	0.9886	0.6631	0.1014	81.189

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 10,000 draws after discarding the first 5000 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.

s_t (right-hand side), respectively.[4] In both graphs, the solid black and red lines denote their posterior means and medians, respectively, while the blue shaded area represents their 90% credible interval. The probabilities of regime 2 conditional on Y_t and $\mathcal{M}$, $Prob(s_t = 2 \,|\, Y_t, \mathcal{M})$, as shown in the left graph of panel (a), are equivalent to the posterior distribution of regime variable $s_t = 2$ and they are made from MCMC draws using a smoothing method for a Markov-switching model proposed by Albert and Chib (1993), whereas the right graph shows that the posterior density of the MCMC draws of λ_1 and λ_2 are concentrated around 0 and 1 respectively, which means that we can clearly distinguish the two regimes for model combination. Figure 9(b) shows the histograms and traces of the MCMC draws of the both paramters λ_1 and λ_2. Again, these posterior estimates are also obtained from MCMC 10,000 draws after discarding the first 5000 draws as burn-in. As shown in Table 4, the posterior means of the weighting coefficients are $\lambda_1 = 0.1305$ and $\lambda_2 = 0.9451$ for four-quarter-ahead forecast. This indicates that $s_t = 1$ corresponds to the regime in which the ATSM dominates while $s_t = 2$ corresponds to regime in which the QTSM dominates as the right panel of Figure 9(a). The left panel of Figure 9(a) shows that the probability of regime 2 is close to one since 1996 as the JGB yields continued to move towards the zero lower bound amid the Bank of Japan's interest rate cuts. In particular, it is interesting to look closer to the posterior distribution of λ_1: as the model switches gradually from the ATSM to QTSM in the early 1990s, the dispersion of λ_1 is large, indicating that both the two models are useful in explaining the bond yields: as the Bank of Japan adopts the ZIRP from 1999Q1 to 2000Q3, the distribution of λ_1 is concentrated around 1, indicating that the QTSM model captures much better the joint dynamics of bond yields and the macroeconomy under the ZIRP.

[Insert Figure 9 around here]

4.3.3 Dynamic pooling

Figure 10 shows the estimation for the dynamic pooling scheme which imposes a smooth transition in between the two selected models as described in Eq.(3). The solid black line denotes the posterior means of the time varying weighting coefficients while the blue shaded area represents their 90% credible interval. The estimation is obtained from 5000 draws of particle filter with constant autocorrelation coefficient ρ fixed as 0.9 following Del Negro et al (2013). Similar to the Markov-switching pooling, the weighting to QTSM increases after 1995Q1 and keep dominating the ATSM afterwards. However, the dispersion of the posterior distribution

[4] The draws of the time-varying weight λ_t are conducted based on the following equation. $\lambda_t = \lambda_1 \times Prob(s_t = 1) + \lambda_2 \times Prob(s_t = 2)$.

Table 4: Posterior estimates of the Markov-switching prediction pool

(a) One-quarter-ahead forecast

Parameter	Mean	90 Percentile	10 Percentile	Std. Dev.	Inefficiency
λ_1	0.0899	0.2670	0.0039	0.0831	22.930
λ_2	0.9014	0.9951	0.7142	0.0906	44.189
q_{11}	0.9167	0.9774	0.8403	0.0564	50.122
q_{22}	0.9259	0.9776	0.8623	0.0476	53.547

(b) Four-quarter-ahead forecast

Parameter	Mean	90 Percentile	10 Percentile	Std. Dev.	Inefficiency
λ_1	0.1305	0.3733	0.0070	0.1167	22.341
λ_2	0.9451	0.9970	0.8395	0.0507	42.194
q_{11}	0.8922	0.9706	0.7911	0.0743	44.401
q_{22}	0.9772	0.9976	0.9467	0.0240	48.601

Notes:

1. λ_1 and λ_2 denote the regime switching weighting coefficient determined by regime variables s_t in the following optimal prediction pool:

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

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

$$\lambda_t = \lambda\left(s_t\right) = \begin{cases} \lambda_1, & s_t = 1 \\ \lambda_2, & s_t = 2 \end{cases},$$

2. The coefficient λ is estimated with MCMC simulation and obtained from 10,000 draws after discarding the first 5000 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.

of λ_t appears to be large with the median fluctuating around 0.3 to 0.8 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. The Markov-switching type pooling maybe more effective in capturing the abrupt change in bond yield movements as the monetary policy switches from time to time.

[Insert Figure 10 around here]

4.3.4 Comparison

Lastly, let us compare the performance of different models and pooling schemes in terms of the log score of prediction density. Table 5 summarizes the corresponding cumulative log score performance for the one-quarter-ahead and four-quarter-ahead forecasts obtained from Eq.(1), Eq.(2) and Eq.(3). Figure 11 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 5, the Markov-switching (MS) pooling scheme 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 1995 to the QTSM after 1995 as depicted in Figure 11. 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 11 around here]

5 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 1990Q1 to 2008Q3 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.

Table 5: Cumulative log scores

(a) One-quarter-ahead forecast

Component Models		Model Pooling	
Model	Log Score	Methods	Log Score
ATSM	2006.85	Static	2008.38
QTSM	1987.93	Markov Switching	2018.13
		Dynamic	2013.82

(b) Four-quarter-ahead forecast

Component Models		Model Pooling	
Model	Log Score	Methods	Log Score
ATSM	1909.20	Static	1922.78
QTSM	1923.03	Markov Switching	1930.92
		Dynamic	1925.87

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

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. (2013) Dynamic Prediction Pools: an Investigation of Financial Frictions and Forecasting Performance. Working Paper.

[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] 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.

[25] 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.

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

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

[28] 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.

[29] 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.

[30] 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.

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

[32] 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

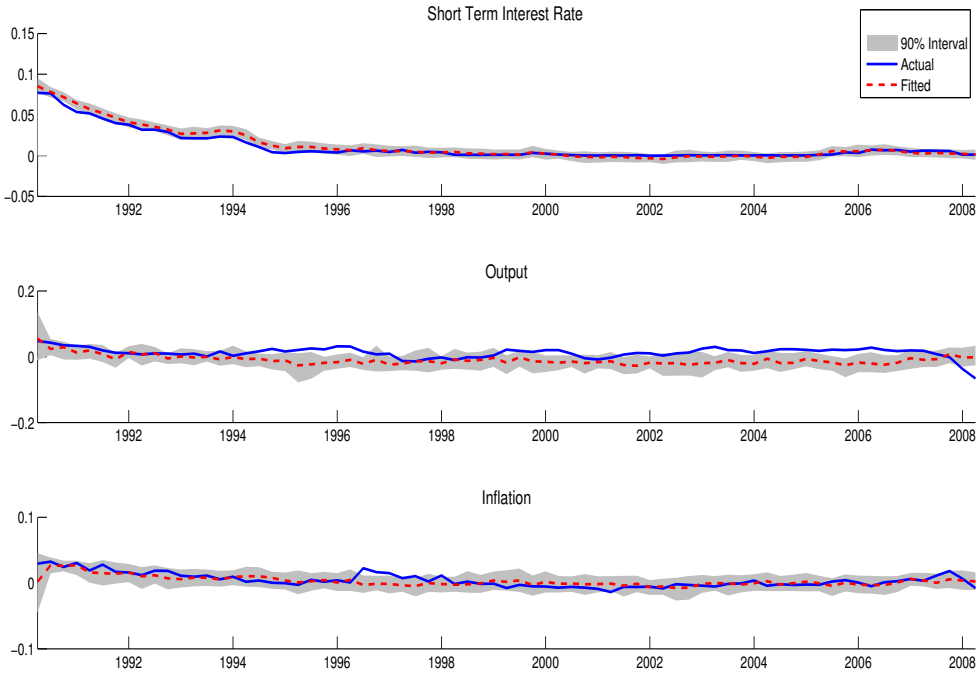
It is important to note that we do not estimate explicitly the loading coefficients for the ATSM and QTSM. This allows us to avoid identification problem (as our macro factors are observed with errors) and also a more efficient estimation on the model parameters. We follow Ang et al. (2011) to pre-set 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 take $\Psi_0 = \text{diag}(50, 0, 0)$ for the QTSM which is obtained by running a preliminary OLS regression of the average yield against the short rate.

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 $\rho = 0.9$.

Figure 1: Filtered macro factors and fitted bond yields by ATSM

(a) Macro factors



(b) Bond yields

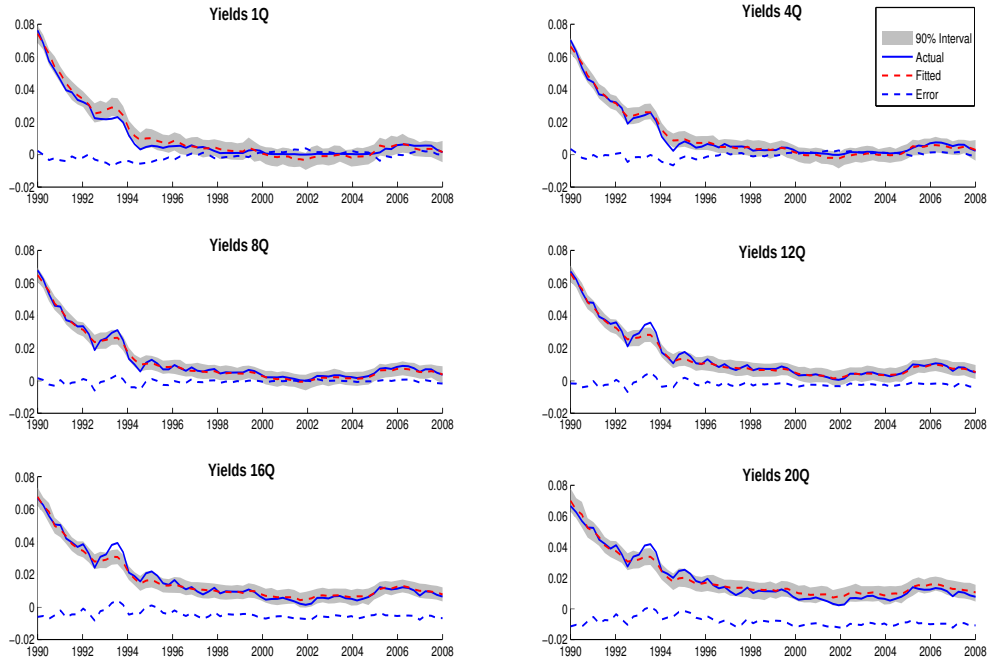
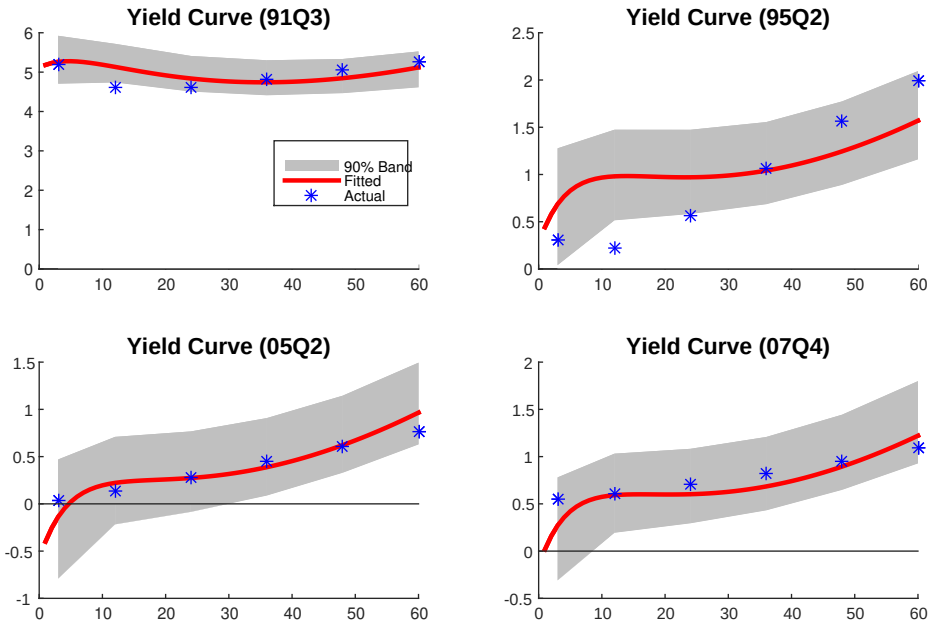


Figure 1: (cont'd)

(c) Yield Curve



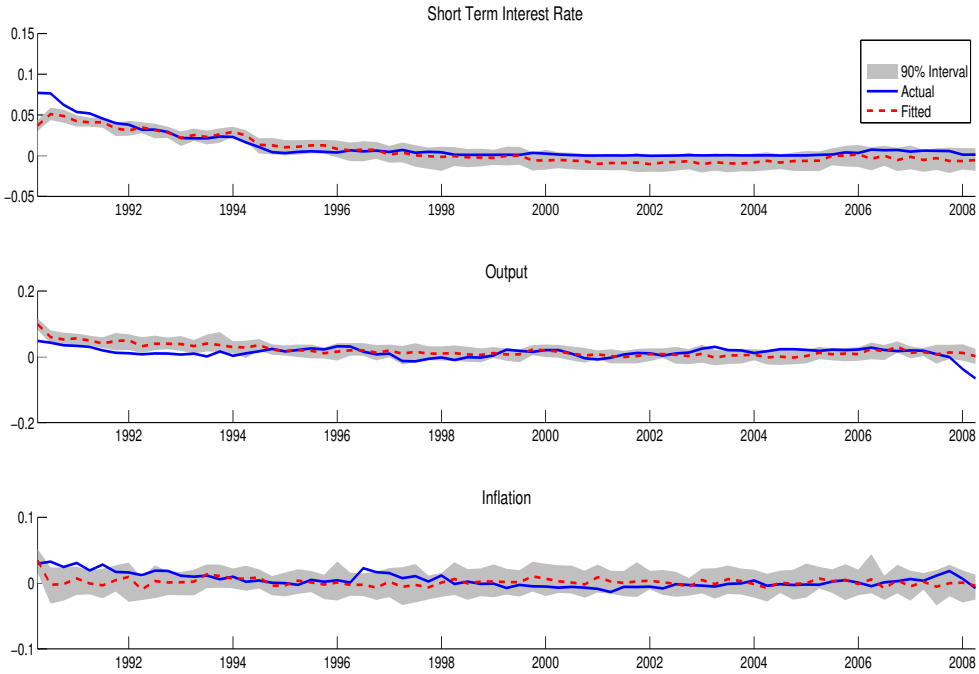
Notes:

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 2: Filtered macro factors and fitted bond yields by QTSM

(a) Macro factors



(b) Bond yields

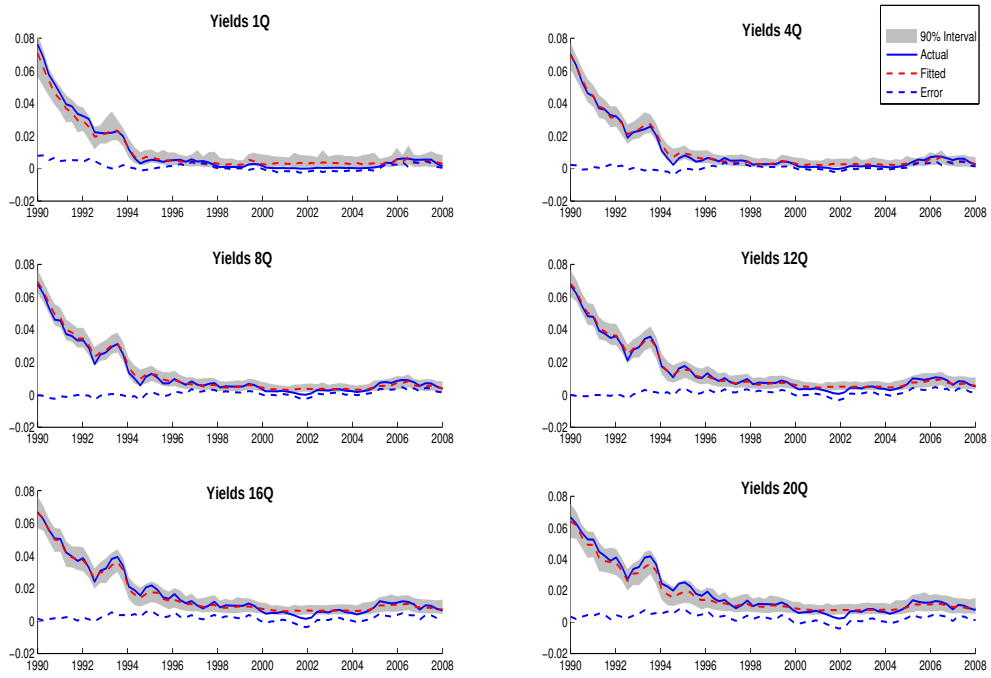
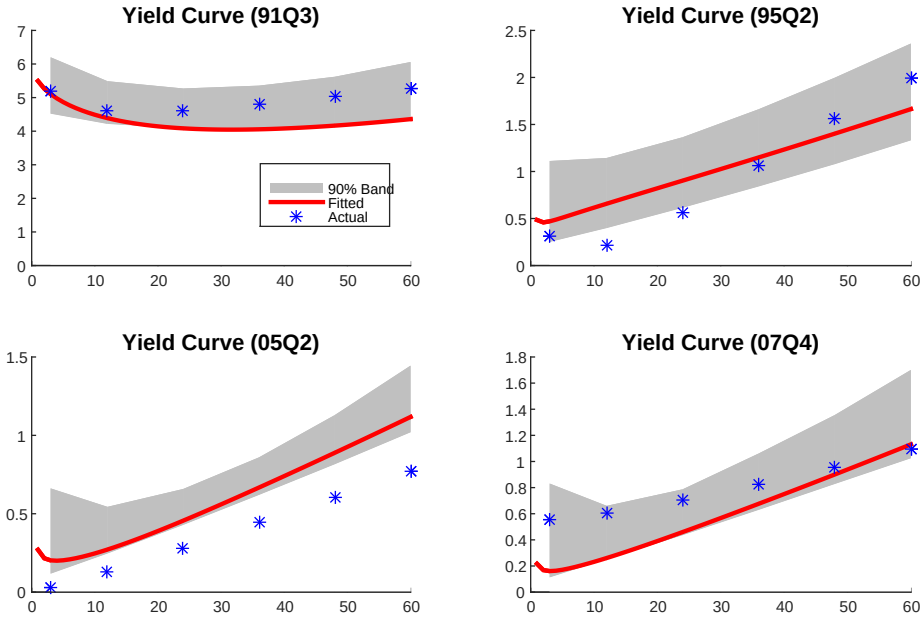


Figure 2: (cont'd)

(c) Yield Curve

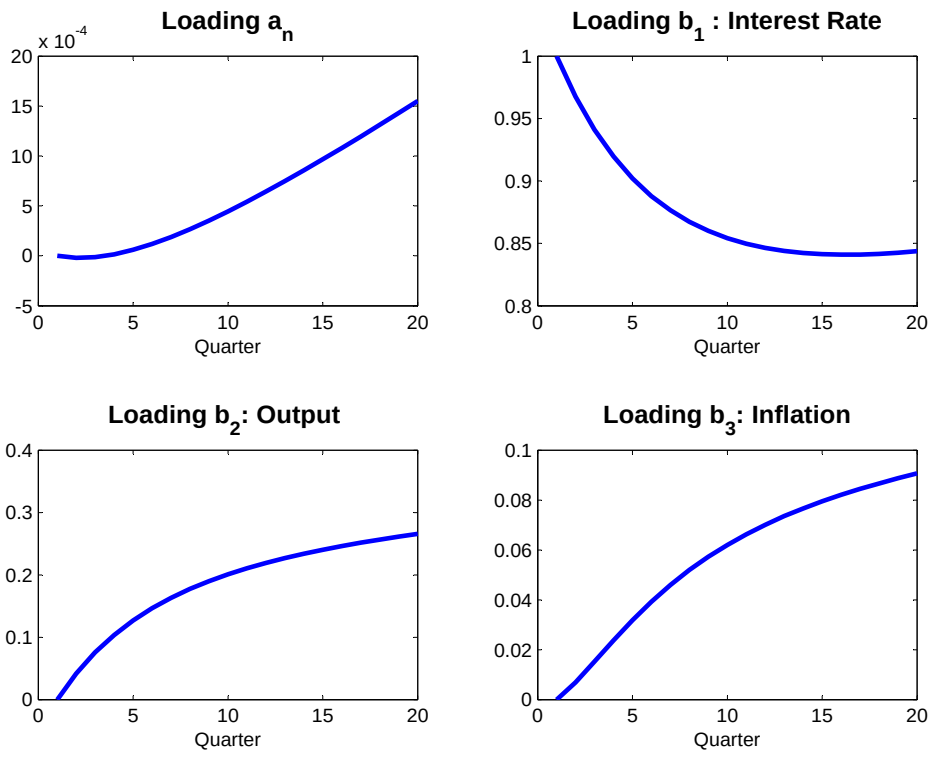


Notes:

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.

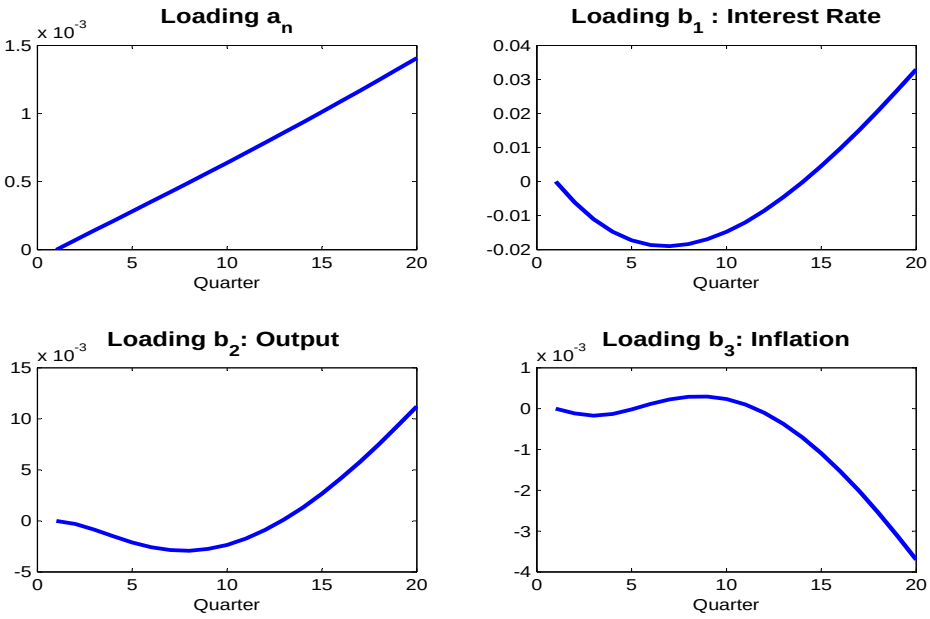
Figure 3: Posterior Means of Factor loadings a_n and b_n for ATSM



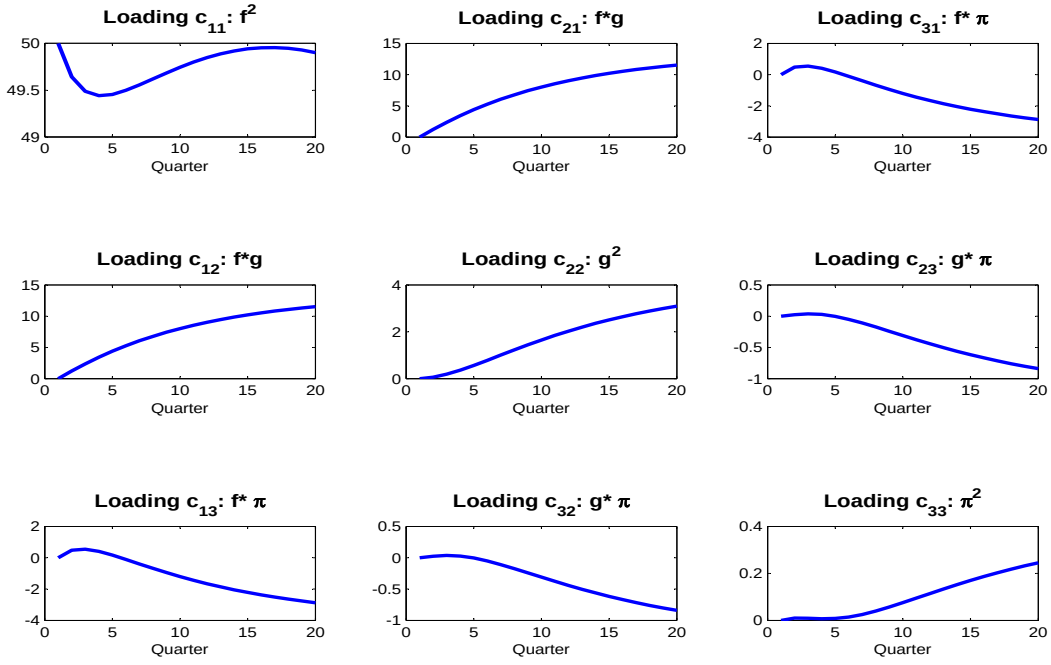
Notes: The factor loadings of the estimated Gaussian ATSM are calculated using the recursive relationship (7) and the posterior means under Q-measure in Table 1.

Figure 4: Posterior Means of Factor loadings a_n, b_n and c_n for QTSM

(a) Factor loadings a_n and b_n



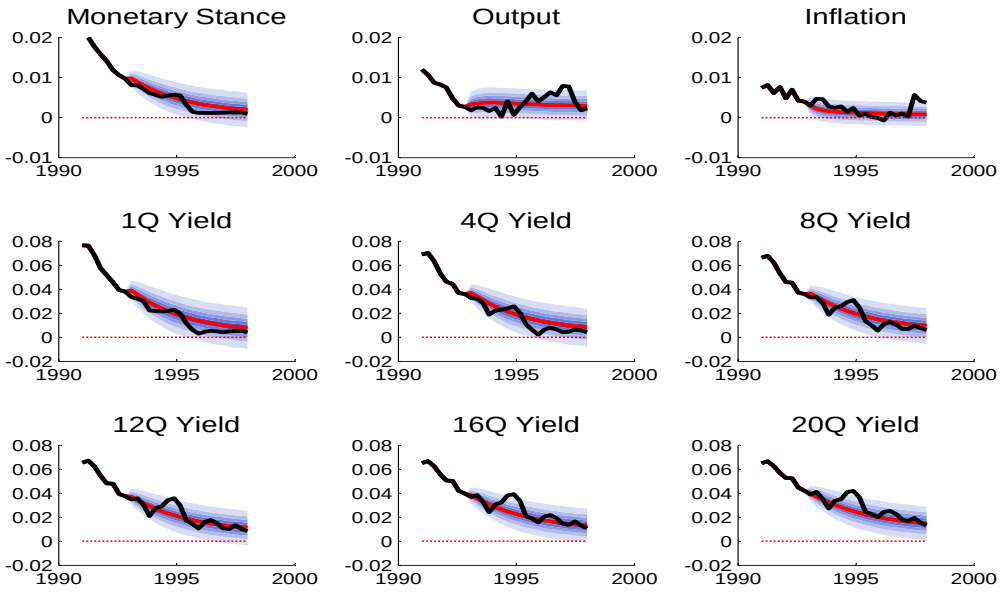
(b) Factor loading c_n



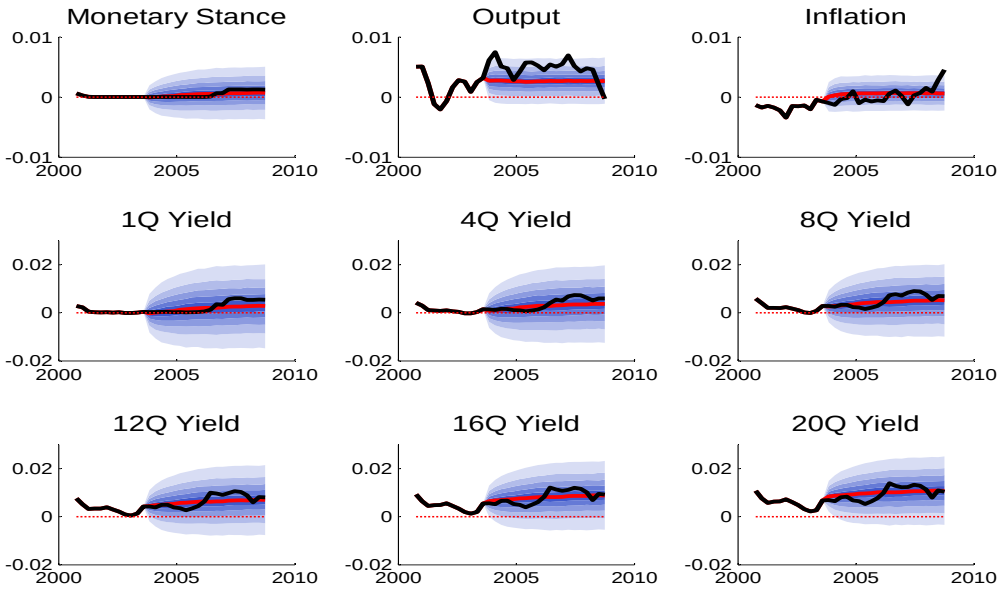
Note: The factor loadings of the estimated QTSM are calculated using the recursive relationship (12) and the posterior means under Q-measure in Table 2.

Figure 5: Forecasting of bond yields by ATSM

(a) Periods: 1992Q4 - 1998Q1 (under non-ZIRP)



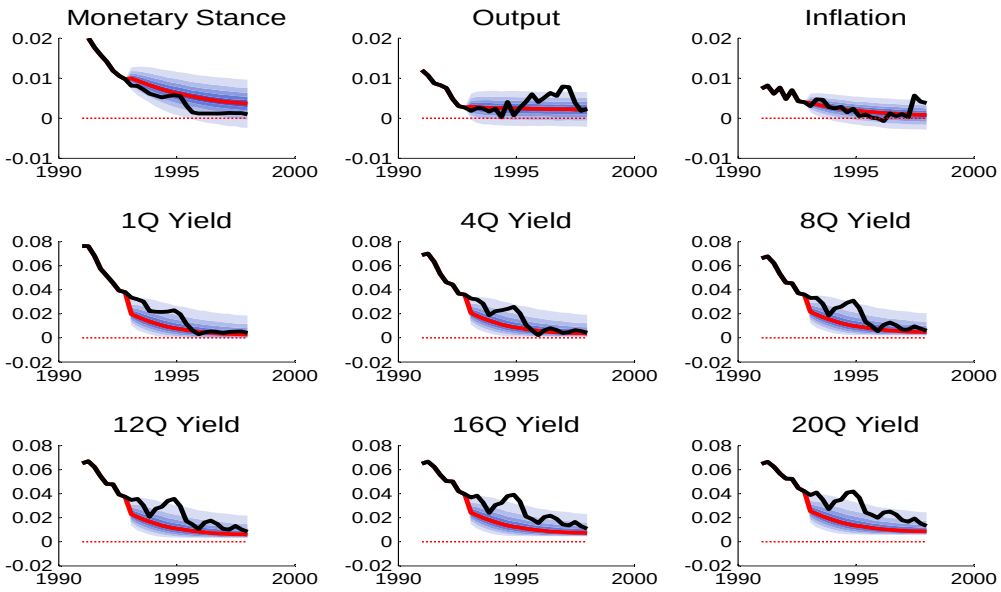
(b) Periods: 2004Q2 - 2009Q1 (under ZIRP)



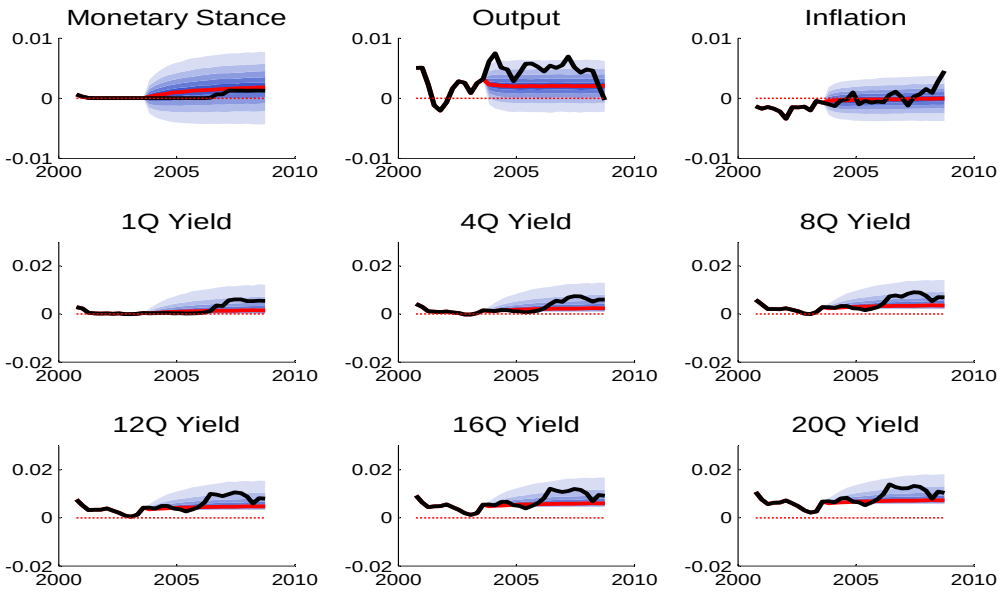
Note: The posterior prediction distributions of the macro factor and the JGB yields of the ATSM models are calculated based on the procedure as described in Section 2.1, using 10,000 draws of posterior estimates over the full sample as shown in Table 1.

Figure 6: Forecasting of bond yields by QTSM

(a) periods: 1992Q4 - 1998Q1 (under non-ZIRP)

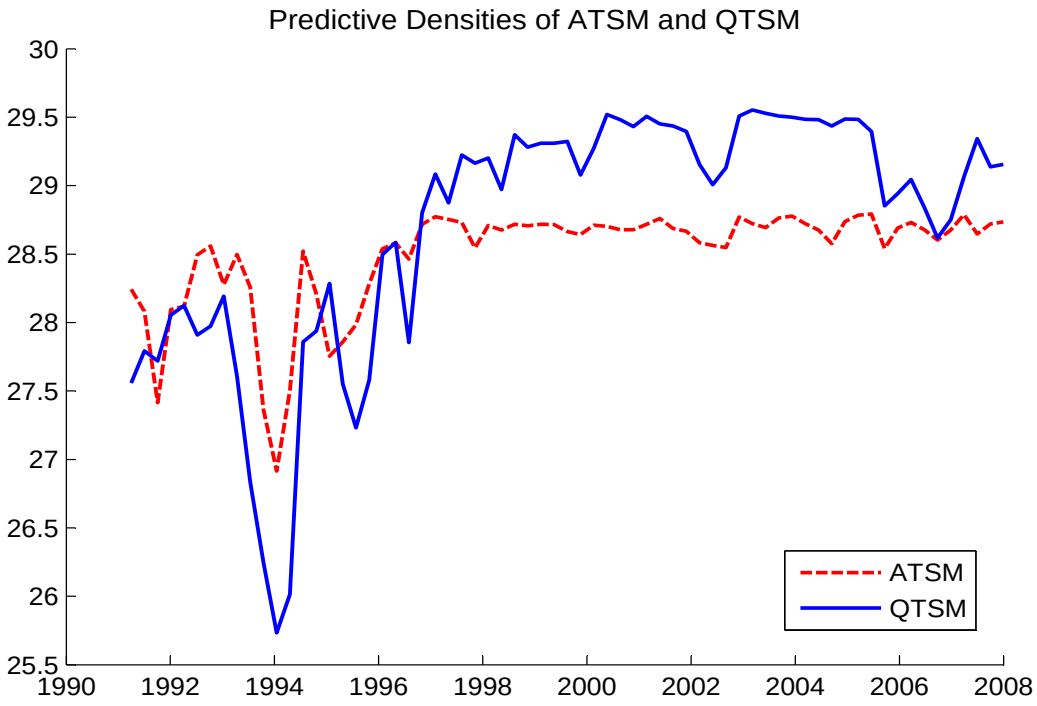


(b) periods: 2004Q2 - 2009Q1 (under ZIRP)



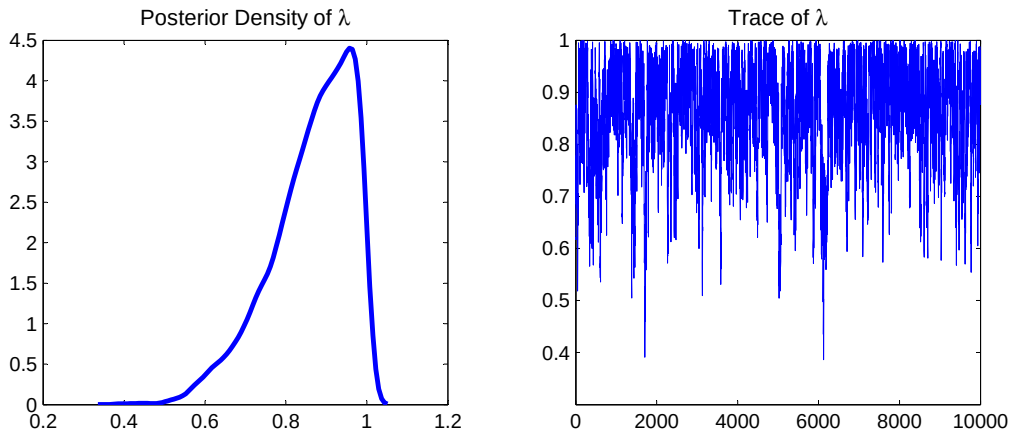
Note: The posterior prediction distributions of the macro factor and the JGB yields of the QTSM models are calculated based on the procedure as described in Section 2.1, using 10,000 draws of posterior estimates over the full sample as shown in Table 2.

Figure 7: Log score comparison of the ATSM and QTSM (based on 4Q-ahead forecast)



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.

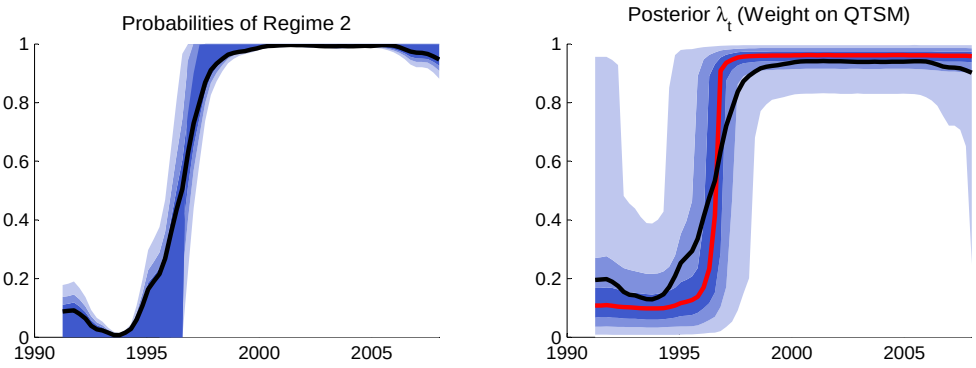
Figure 8: Static prediction pool (4Q-ahead forecast)



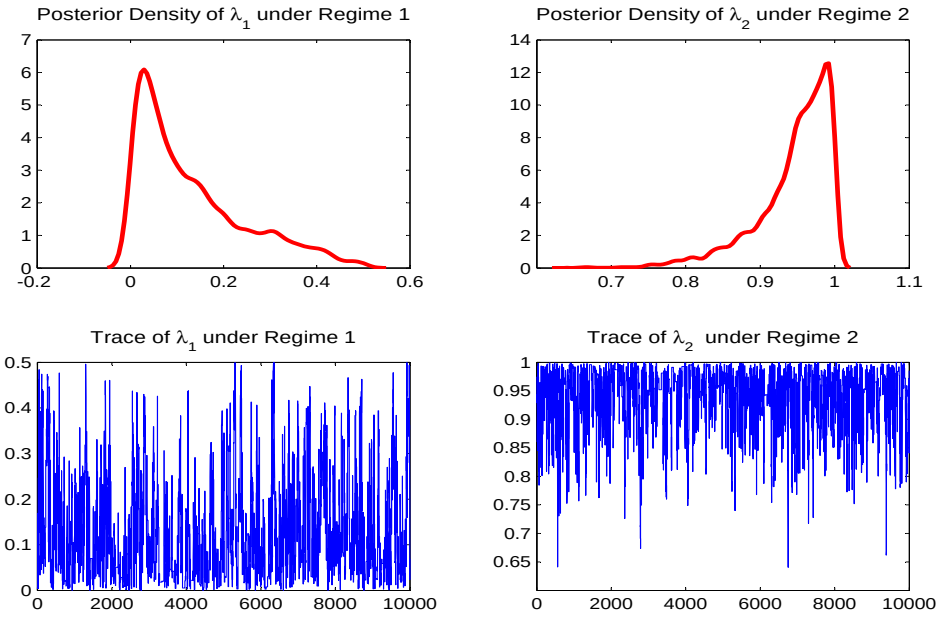
Note: Static pooling model is calculated from Eq.(1). The weighting coefficient on QTSM, λ, is estimated with MCMC simulation and obtained from 10,000 draws after discarding the first 5000 burn-in draws.

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

(a) Probability of regime 2 (QTSM) and the corresponding posterior density of λ_2.



(b) Posterior Distributions of Parameters

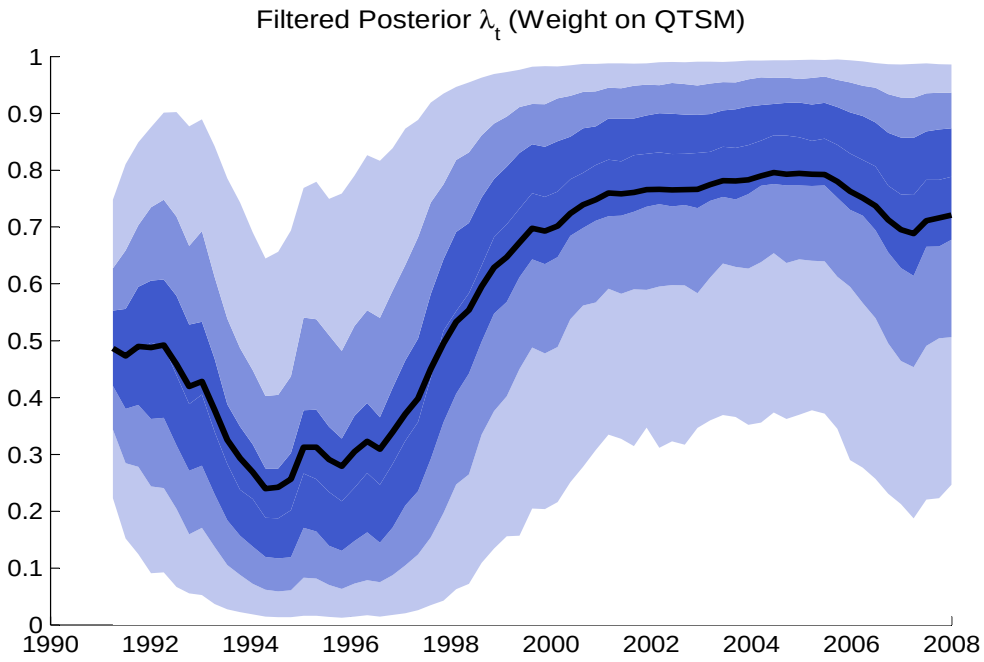


Notes:

1. Markov-switching pooling model is calculated from Eq.(2). The weighting coefficients on QTSM, λ_i, are estimated with MCMC simulation and obtained from 10,000 draws after discarding the first 5000 burn-in draws.

2. The solid black and red lines denote their posterior means and medians, respectively, and the blue shaded area represents their 90% confidence interval.

Figure 10: Dynamic prediction pool (4Q-ahead forecast)

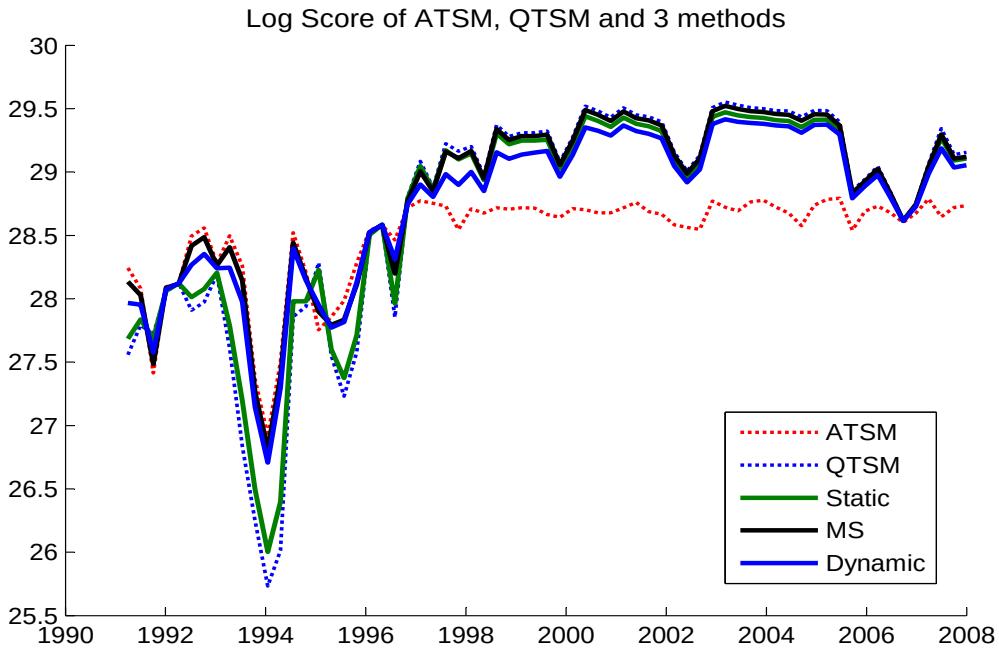


Notes:

1. Dynamic pooling model is calculated from Eq.(3). The time-varying coefficient is obtained from 5000 draws of particle filter with constant autocorrelation coefficient ρ fixed as 0.9 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 11: 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).