# Forecasting the Term Structure of Interest Rates with Potentially Misspecified Models

Yunjong Eo*
Kyu Ho Kang†

This Version: January 2014
Very Preliminary

**Abstract**

Since Diebold and Li (2006) showed the outstanding performance of a dynamic Nelson-Sieglel model (DNSM) in forecasting the yield curve, the DNSM has been widely used in many macro and finance area. Because of its parsimonious but flexible model specification the Bayesian model-averaging method based on the Bayes factor typically gives a weight of nearly one on the DNSM excluding a standard arbitrage-free affine term structure model (ATSM). Nevertheless, the ATSM has been also commonly used because it provides plenty of economically interpretable outcomes such as term premium and model-implied term structure of real interest rates. Meanwhile, the random-walk (RW) is often used as a benchmark in out-of-sample forecasting comparison. Despite the popularity of these three frameworks, none of them dominates the others across all maturities and forecast horizons. This fact indicates that those models are potentially misspecified. In this paper we investigate whether combining the possibly misspecified models in a linear form suggested by Geweke and Amisano (2011) and Waggoner and Zha (2012) help improve the predictive accuracy. For this we compare out-of-sample prediction performance from the merged models with a constant model weight with those of the three individual prediction models and the merged models with a Markov-switching model weight for eight different maturities and forecast horizons of 1, 3, 6 and 12 months. We find that overall the constant mixture model is most supported. In particular, the constant mixture model consistently forecasts better than the individual prediction models across all maturities and forecast horizons. (JEL G12, C11, F37)

*Keywords*: Model combination, forecasting, Markov switching process, Bayesian MCMC method, Dynamic Nelson-Siegel, Affine term structure model

* *Address for correspondence*: School of Economics, University of Sydney, NSW 2006, Australia. E-mail: Yunjong.Eo@sydney.edu.au

† *Address for correspondence*: Department of Economics, Korea University, Seoul, South Korea, 136-701, *Email*: kyuho@korea.ac.kr

# 1 Introduction

Forecasting the yield curve is extremely important for financial portfolio risk management, monetary policy, business cycle analysis and so forth. In previous literatures three classes of yield curve prediction models have been widely used. One is arbitrage-free affine term structure models, which are a theoretical approach[1]. This approach provides many economically interpretable outcomes such as term premium and term structure of real interest rates. Despite that, this class of models is known to be difficult to estimate because of the nonlinearity and irregular likelihood surface. Another is a purely statistical approach, which is a dynamic version of the Nelson-Siegel model[2]. Since this modeling approach is parsimonious but flexible for fitting the yield curve, overall its forecasting performance is better than the theoretical approach. The other is the random-walk model (RWM), and it is often used as a benchmark in forecasting ability comparison.

Interestingly, beating the RWM is a challenging task although DNSM or ATSM can be better at some particular maturities and forecast horizons. None of the three alternative models uniformly outperforms at all maturities and forecast horizons. For example, Diebold and Li (2006) finds that the three-factor DNSM (DNSM(3))'s 1-month-ahead forecasts outperform those of the RWM at short maturities, but for long-term bond yields the random-walk dominates the DNSM(3). Zantedeschi et al. (2011) confirm that the RWM forecasts better in the short run whereas at three- and six-step-ahead forecast horizons the predictions from their DNSM with time-varying factorloadings are much improved. The forecasts from the ATSM estimated by Moench (2008) are found to be more accurate than those from the RWM only for the 6-month yield. Using an ATSM Carrieroa and Giacomini (2011) produce 1-step-ahead forecasts and find positive prediction gains against the RWM for intermediate and long maturities, not for short maturities.

These mixed results for out-of-sample prediction comparison strongly indicate that

[1]For example, Moench (2008), Christensen, Diebold, and Rudebusch (2011), Chib and Kang (2013), Almeida and Vicenteb (2008), and Carrieroa and Giacomini (2011)

[2]For example, Diebold and Li (2006), De Pooter (2007), and Zantedeschi, Damien, and Polson (2011)

all prediction models are potentially somewhat misspecified. Our goal of this paper is to investigate whether it is possible to improve the out-of-sample prediction performance when all alternative models are potentially misspecified. In a Bayesian context a standard way to consider the model uncertainty is using the Bayesian model-averaging method based on the marginal likelihood computation. However, the Bayesian model averaging typically gives a weight of nearly one on the DNSM excluding the ATSM and RWM.

As an alternative way to consider the model misspecifications we take the pooling method recently suggested by Geweke and Amisano (2011) and Waggoner and Zha (2012). The key idea of their approach is to construct the one-step ahead predictive density as a linear combination of the predictive densities obtained from each of alternative prediction models. In this paper the three individual yield curve prediction models and mixture models of two or three of the prediction models are compared in terms of out-of-sample predictive accuracy. Further, the model weights are specified to be constant or Markov-switching over time. Using these mixture models we forecast the monthly yields with eight different maturities over the forecast horizons of one through twelve months, and conduct model comparison based on the predictive accuracy.

The key finding of our empirical work is that the predictive gains from the pooling method are surprisingly substantial. In particular, the constant mixture model of the three prediction models consistently forecasts better than each of them over all maturities and forecast horizons. This finding implies that all three alternatives are operative over time, and so ATSM and RWM are never negligible in forecasting unlike in the Bayesian model averaging.

The rest of the paper is organized as follows. Section 2 describes our econometric models. Section 3 discusses the Bayesian estimation procedure. Section 4 presents the empirical results. Section 5 concludes. The Appendices provide details for the model derivation and the estimation method.

## 2 Econometric Methodology

In this section we illustrate our statistical method using an example of two prediction models, $\mathcal{M}_1$ and $\mathcal{M}_2$. Let $\Theta_1$ and $\Theta_2$ be the set of parameters in $\mathcal{M}_1$ and $\mathcal{M}_2$, respectively. The $\tau$-period bond yield at time $t$ is denoted by $y_t(\tau)$ and the vector of yields with $N$ different maturities at time $t$ is $y_t = (y_t(\tau_1),\ y_t(\tau_2), .., \ y_t(\tau_N))$. We let $Y_t^o = \{y_1^o, \ldots, y_t^o\}$ denote the observed yield curve data up to time $t$. Then Geweke and Amisano (2011) study predictive densities of the form

$$w_1 \times p(y_t|Y_{t-1}^o, \Theta_1, \mathcal{M}_1) + (1 - w_1) \times p(y_t|Y_{t-1}^o, \Theta_2, \mathcal{M}_2) \tag{2.1}$$

with $w_1 \in [0, 1]$ is the model weight on $\mathcal{M}_1$. Waggoner and Zha (2012) extend Geweke and Amisano (2011)'s approach and allow the model weights to vary over time. Then $w_1$ in equation (2.1) is replaced by $w_{1,s_t} \in [0, 1]$ where $s_t$ takes either 1 or 2 following a first-order two-state Markov process with constant transition probabilities

$$q_{ij} = \Pr\left[s_t = j|s_{t-1} = i\right], \ i, j = 1, 2$$

By doing this they consider the case that the relative importance of each of the prediction models can change over time. The resulting predictive density conditioned on the regime $s_t$ is given by

$$w_{1,s_t} \times p(y_t|Y_{t-1}^o, \Theta_1, \mathcal{M}_1) + (1 - w_{1,s_t}) \times p(y_t|Y_{t-1}^o, \Theta_2, \mathcal{M}_2)$$

On letting the model specific parameters $\Theta = \{\Theta_1,\ \Theta_2\}$, transition probabilities $Q = \{q_{11},\ q_{22}\}$, and the regime-dependent model weight $w = \{w_{1,1},\ w_{1,2}\}$ the likelihood can be constructed as

$$\ln p(Y_T^o|\Theta, P, w) = \sum_{t=1}^{T} \ln p(y_t|Y_{t-1}^o, \Theta, P, w)$$

where the regime $s_t$ is integrated out because it is never observed by econometricians. For more details for likelihood computations refer to Appendix A.

It should be noted that although we follow Geweke and Amisano (2012) and Waggoner and Zha (2012)'s methodological approach our study differs from theirs in several dimensions. First, we concentrate on yield curve forecasting while they forecast macroeconomic variables such as the GDP growth rate and inflation. Second, in our work the model specific parameters, the model weights and the transition probabilities are estimated simultaneously, not sequentially. Third, most importantly, is that both short- and long-term forecasts are produced and used for model comparison whereas they assess the predictive performance of pooled models based on the log predictive score, which is a good measurement of one-step ahead predictive accuracy.

# 3 Estimation

## 3.1 Competing Models

In our pool of models we consider the three classes of predictions models: ATSM(3), DNSM(3), and the random-walk. In a standard affine term structure model the $\tau$-period bond yield at time $t$ is assumed to be affine to a vector of exogenous driving factors $\mathbf{f}_t$

$$y_t(\tau) = a(\tau) + b(\tau)'\mathbf{f}_t$$

The intercept term $a(\tau)$ and the factorloadings $b(\tau)$ are both maturity-specific. These coefficients are endogenously determined by the no-arbitrage condition given certain assumptions about the dynamic evolution of the factors and the stochastic discount factor.

In the DNSM(3) the bond yield is also specified as a linear function of three exogenous latent factors. Unlike in the affine term structure model, the intercept term and the factorloadings are exogenously fixed, so that the factors are interpreted as time-varying level, slope and curvature of the yield curve. Finally, the random-walk model (RWM) is considered and often used as a benchmark in forecasting ability comparison. The details about the model specification of the three prediction models can be found in Appendix B.

Diebold and Li (2006) show that overall DNSM(3) produces better forecast accuracy in out-of-sample prediction compared to Duffie (2002)'s best essentially affine model although the RWM forecasts better at short forecast horizons. Then the Bayesian model-averaging method based on the Bayes factor would yield nearly one weight on the DNSM(3) excluding the ATSM(3) and the random-walk. Nevertheless, all those prediction models have been commonly used for forecasting the term structure of interest rates, and none of them consistently outperforms at all maturities and forecast horizons. One potential reason is that the alternative models are somewhat misspecified.

Given the potential model misspecification of the alternative models we investigate whether combining the multiple models in a linear form helps improve the predictive accuracy. Table 1 presents twelve competing models with various combinations. Basically, we consider the individual models. Also the linear combinations of two and three of the alternatives are used for prediction. The model weights can be either constant or time-varying. For the constant mixture case, the weights are to be estimated or equally given. For instance, the NS-AF-RW is the constant mixture of the DNSM(3), ATSM(3) and the random-walk whereas the *MS*-NS-AF-RW is the mixture model in which the model weights vary over time according to the Markov process. In *Equal*-NS-AF-RW each of the model weights is fixed at 1/3.

| | DNSM(3) | ATSM(3) | random-walk |
|---|---|---|---|
| *Single* | | | |
| DNSM | Yes | - | - |
| ATSM | - | Yes | - |
| RWM | - | - | Yes |
| *Constant mixture* | | | |
| NS-AF | Yes | Yes | - |
| NS-RW | Yes | - | Yes |
| AF-RW | - | Yes | Yes |
| NS-AF-RW | Yes | Yes | Yes |
| *Equal*-NS-AF-RW | Yes | Yes | Yes |
| *Markov-switching mixture* | | | |
| *MS*-NS-AF | Yes | Yes | - |
| *MS*-NS-RW | Yes | - | Yes |
| *MS*-AF-RW | - | Yes | Yes |
| *MS*-NS-AF-RW | Yes | Yes | Yes |

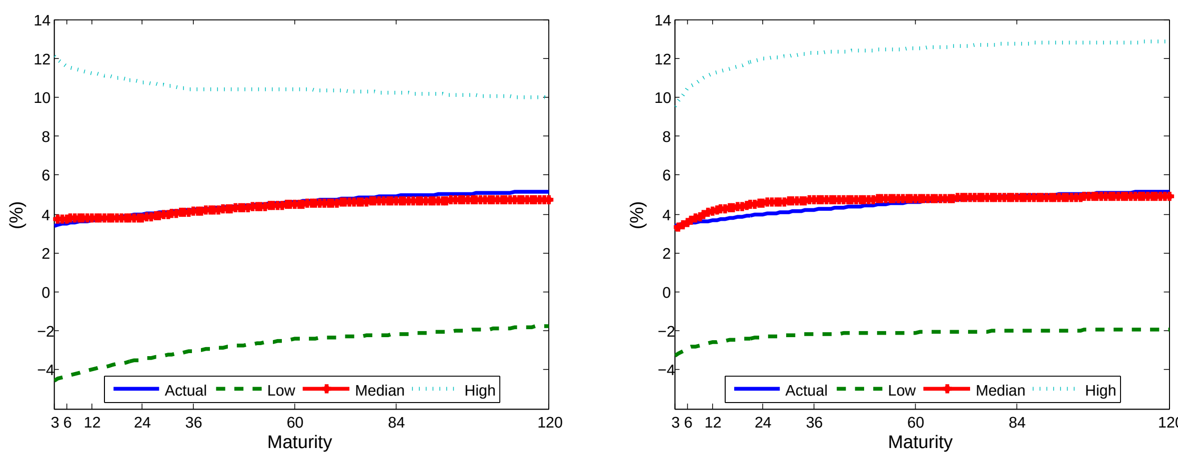
**Table 1: Model combinations**

## 3.2 Posterior Simulation

In the Bayesian context our regime switching mixture model is the joint posterior distribution of the regime indicators ($\mathbf{S} = \{s_t\}_{t=1}^{T}$), continuous latent variables ($\mathbf{X} = \{\mathbf{x}_t\}_{t=1,2,..,T}$ and $\mathbf{F} = \{\mathbf{f}_t\}_{t=1,2,..,T}$) and the model parameters ($\boldsymbol{\psi}= \{\Theta, P, w\}$). It has the form

$$\pi\left(\boldsymbol{\psi}, \mathbf{X}, \mathbf{F}, \mathbf{S}|\mathbf{Y}\right) \propto f\left(\mathbf{Y}|\boldsymbol{\psi}, \mathbf{X}, \mathbf{F}, \mathbf{S}\right) \times f\left(\mathbf{X}, \mathbf{F}|\boldsymbol{\psi}\right) \times p\left(\mathbf{S}|\boldsymbol{\psi}\right) \times \pi(\boldsymbol{\psi}) \qquad (3.1)$$

where $\pi\left(\boldsymbol{\psi}\right)$ is the prior density of the parameters, $p\left(\mathbf{S}|\boldsymbol{\psi}\right)$ is the prior density function for regime-indicators given the parameters and it is specified as the discrete two-state Markov switching process, $f\left(\mathbf{X}, \mathbf{F}|\mathbf{S},\boldsymbol{\psi}\right)$ is the prior density of the factors and $f\left(\mathbf{Y}|\boldsymbol{\psi}, \mathbf{X}, \mathbf{F}, \mathbf{S}\right)$ is the joint density of the $\mathbf{Y} = \{y_t\}_{t=1}^{T}$. Our prior which we give in the paper is set up to reflect the apriori belief that the yield curve is gently upward sloping and concave on average. We arrive at this prior by prior simulation technique, sampling parameters from the assumed prior, then sampling the data given the parameters, and then repeating this process many times. This mildly upward sloping and concave yield curve prior tends to smooth out the many local modes of the likelihood surface as Chib and Ergashev (2009) show. Table 2 reports our prior and Figure 1 plots the resulting prior-implied unconditional distribution of the yield curve.




(a) DNSM(3)

(b) ATSM(3)

**Figure 1: Prior-implied yield curve**

For regime identification, we impose a restriction that the weight on the model $\mathcal{M}_1$

should be higher in regime 1 than in regime 2

$$1 > w_{1,s_t=1} > 0.5 > w_{1,s_t=2} > 0$$

For the ATSM(3), factors are identified by the difference in the persistence. These identification restrictions are imposed through the prior specification. Because the joint posterior distribution in equation 3.1 is not analytically tractable, we rely on a MCMC simulation method and sample the parameters and the states recursively from the joint posterior distribution as follows:

*Algorithm 1: MCMC sampling*

- Step 1: Sample $\boldsymbol{\psi}= \{\Theta_1, \Theta_2, P, w\}|\mathbf{Y}, \mathbf{S}$ using the tailored randomized blocking Metropolis-Hastings algorithm (Chib and Ergashev (2009))
- Step 2: Sample the discrete states $\mathbf{S}|\mathbf{Y},\boldsymbol{\psi}$ based on the multi-move method (Chib (1998))
- Step 3: Calculate $\mathbf{X}(\mathbf{F})$ given the most recent values of $\Theta_1(\Theta_2)$
- Step 4: Simulate the predictive density for the yields given $(\boldsymbol{\psi},\mathbf{X}, \mathbf{F})$

For each posterior draw $(s_T, \mathbf{x}_T, \mathbf{f}_T, \boldsymbol{\psi})$ and forecast horizon of $h = 1, 2, .., H,$, we can also simulate the posterior predictive density of the bond yields.

*Algorithm 2: Posterior predictive simulation*

- Step 1: Sample the factors $(\mathbf{x}_{T+h}, \mathbf{f}_{T+h})$
- Step 2: Given the factors, sample the yields $y_{1,T+h}(y_{2,T+h})$ from $\mathcal{M}_1(\mathcal{M}_2)$
- Step 3: Sample the regimes, $s_{T+h}$ conditioned on $(s_{T+h-1}, Q)$
- Step 4: Given the $s_{T+h}$ compute $y_{T+h}$ as

$$y_{T+h} = w_{1,s_{T+h}} \times y_{1,T+h} + (1 - w_{1,s_{T+h}}) \times y_{2,T+h}$$

- Step 5: Retain $y_{T+h}$ as a posterior predictive draw

Following Zantedeschi et al. (2011) and Chib and Kang (2013) we evaluate the predictive accuracy of the forecasts in terms of the posterior predictive criterion (PPC) of Gelfand and Ghosh (1998). For a given model $\mathcal{M}$ and the observations up to time $T$, the PPC for $\mathsf{h}$-step ahead posterior predictive density of $\tau-$period bond yield is computed as

$$\mathsf{PPC}_T(\tau, \mathsf{h}) = \mathsf{D}_T(\tau, \mathsf{h}) + \mathsf{W}_T(\tau, \mathsf{h})$$

where

$$\mathsf{D}_T(\tau, \mathsf{h}) = \mathsf{Var}\left(y_{T+h}(\tau)|\mathbf{Y}, \mathcal{M}\right)$$

and

$$\mathsf{W}_T(\tau, \mathsf{h}) = \left[y^o_{T+h}(\tau) - E\left(y_{T+h}(\tau)|\mathbf{Y}, \mathcal{M}\right)\right]^2$$

By definition, smaller values of PPC are preferable. Since the value of the PPC can be different in different out-of-sample periods, we compute the average PPC over the twelve different forecast periods, which is shown below

| in-sample | out-of-sample |
|---|---|
| 1990:M1 - 2010:M12 | 2011:M1 - 2011:M12 |
| 1990:M1 - 2011:M1 | 2011:M2 - 2012:M1 |
| 1990:M1 - 2011:M2 | 2011:M3 - 2012:M2 |
| ⋮ | ⋮ |
| 1990:M1 - 2011:M11 | 2011:M12 - 2012:M11 |

# 4 Results

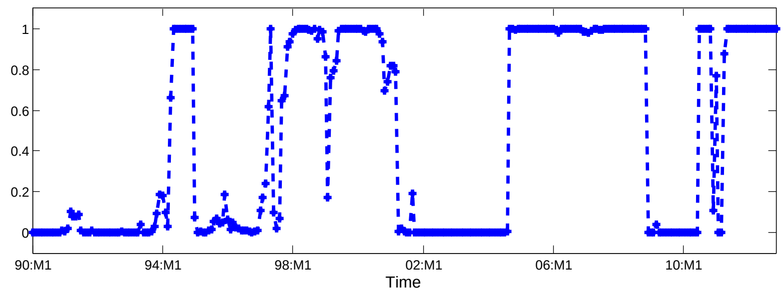
The set of maturities in month is given by $\{3, 6, 12, 24, 36, 60, 84, 120\}$. We denote the parameters in the ATSM(3), DNSM(3) and RWM by $\Theta_1$, $\Theta_2$, and $\Theta_3$, respectively. That is,

$$\Theta_1 = \{\kappa, \phi, V_1, \Lambda_1, \mathbf{D}_1\},\ \Theta_2 = \{\delta, \bar{\boldsymbol{\gamma}}, \boldsymbol{\mu}, \boldsymbol{\Phi}, \mathbf{G}, \beta, \Lambda_2, \mathbf{D}_2\},\ \Theta_3 = \{\mathbf{D}_3\} \tag{4.1}$$

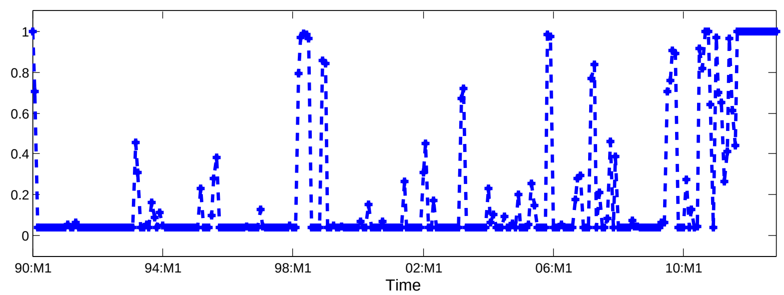
Table 3 reports the results for the PPC averaged over all maturities and forecast horizons of one- through twelve-months. These values capture the predictive accuracy taking into account the cross-sectional and time-series variation of the government bond yields. One can see from the table that the constant mixture of the DNSM, ATSM and RWM (i.e. NS-AF-RW) provides the most accurate forecasts. It implies that all three prediction models are operative, and that the pooling method is better than the Bayesian averaging method (i.e. DNSM(3)). Interestingly, the performance of the *Equal*-NS-AF-RW, which is the constant mixture of the DNSM, ATSM and RWM with equal weights, is remarkable. It produces the second best predictive ability. The importance of the constant mixture model with equal weights has been already emphasized by Geweke and Amisano (2012). In addition, it should be noted that merging perdition models does not guarantee improvement in forecasts. For instance, the AF-RW and *MS*-AF-RW are even worse than the RWM.

As seen from Table 4, for the NS-AF-RW, the estimated weight on the DSNS(3) is largest and the estimated weights on the ATSM(3) and RWM are almost equal. The estimated constant weights from NS-AF-RW are not substantially different from the equal weights, 1/3. For this reason, the *Equal*-NS-AF-RW and NS-AF-RW produce a comparable prediction performance. Table 5 and Figure 2 show that the model weights appear to be regime-specific and change over time. The source of the regime switches seems to be the changes in the factor loadings and conditional correlations (Table 6 and Figure 3). However, the Markov-switching mixture models produce less accurate forecasts because of the inefficiency caused by the future regime forecasting.

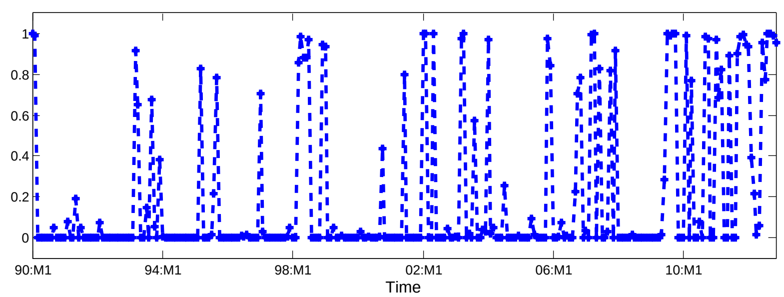
Tables 7 and 8 presents the PPC across maturities and forecast horizons. The PPCs in Table 8 indicate the predictive accuracy of entire yield curve across horizons while the PPCs in Table 7 are the predictive accuracy across maturities averaged over the forecast horizons. One can see that the NS-AF-RW and *Equal*-NS-AF-RW both uniformly forecast better than the RWM across all maturities and forecast horizons. The *Equal*-NS-AF-RW is slightly worse than NS-AF-RW, but its performance is remarkably better than those of the *MS*-NS-AF-RW and individual models. Therefore, beating the constant mixture with equal weights is hard, which is consistent with the finding of Geweke and Amisano (2012). For short-term bond yields the *MS*-NS-AF-RW forecasts best across all forecast horizons as Tables 9 and 10 shows. It appears to be because short term bond yields are more subject to regime shifts than long term bond yields and hence the regime switching specification helps to improve the prediction.



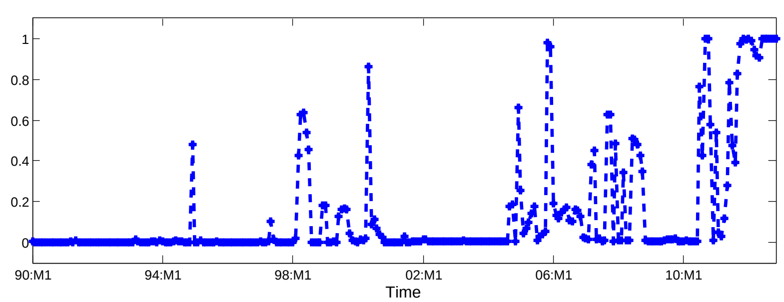

(a) *MS*-NS-AF




(b) *MS*-NS-RW

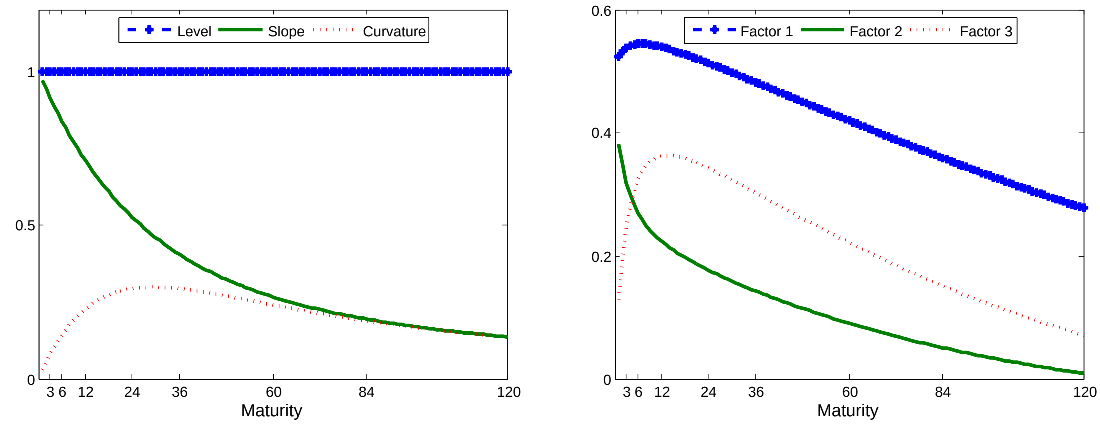



(c) *MS*-AF-RW




(d) *MS*-NS-AF-RW

**Figure 2: Posterior Probability of Regime 2** *These graphs plot the estimates of the probabilities of regime state 2. These graphs are based on 10,000 simulated draws of the posterior simulation.*




(a) Factor Loadings: DNSM(3) (b) Factor Loadings: ATSM(3)

**Figure 3: Factors Loadings: Merged Model with Markov-switching** *These graphs plot the estimates of the factor loadings of the models. These graphs are based on 10,000 simulated draws of the posterior simulation.*

# 5 Concluding Remarks

In this study we attempt to improve the predictive accuracy of government bond yields using the three popular prediction models, which are the three-factor arbitrage-free affine term structure model, the three-factor dynamic Nelson-Siegel model, and the random-walk. All these prediction models are somewhat misspsecified in sense that none of them dominates the others although the Bayesian model averaging would yield the DNSM(3). To mitigate the model misspecification and achieve prediction gains we consider various linear combinations of the three prediction models. According to our out-of-sample prediction comparison, the constant mixture provides the most accurate forecasts. In particular, it outperforms the individual prediction models consistently at all maturities and forecast horizons. It implies that the ATSM(3) and the random-walk as well as the DNSM(3) are still useful in forecasting, and that the pooling method seems to be better than the Bayesian model averaging. More importantly, our findings suggest that for prediction purpose one need to try the pooling method when multiple yield curve prediction models are considered and compared.

We do not argue that our yield curve forecasts obtained from the NS-AF-RW are the best. As Geweke and Amisano (2011) point out, the performance of the pooling method depends on the prediction models contained in the pool. It would be possible to improve the forecasts by including additional prediction models or by changing the model specification of each of the prediction models although it can be computationally more expensive. We leave them as future work.

# Appendix

## A Likelihood

The section summarizes the step by step procedure for the likelihood calculation. Suppose that $p(s_{t-1}|Y^o_{t-1}, \Theta, P, w)$ and $(\Theta, P, w)$ are given and the log likelihood $\ln L$ is initialized at 0. At time 1, $p(s_{t-1}|Y^o_{t-1}, \Theta, P, w)$ is replaced by the unconditional probability of regime $s_t$. For $t = 1, 2, .., T$, the following steps are sequentially repeated.

*Algorithm 3: Likelihood calculation*

- Step 1: The predictive probability of regime $s_t$ $p(s_t = j|Y^o_{t-1}, \Theta, P, w)$ is computed as

$$\begin{aligned} p(s_t &= j|Y^o_{t-1}, \Theta, P, w) \\ &= \sum_{i=1}^{2} Pr[s_t = j|s_{t-1} = i] \times p(s_{t-1} = i|Y^o_{t-1}, \Theta, P, w) \\ &= \sum_{i=1}^{2} q_{ij} \times p(s_{t-1} = i|Y^o_{t-1}, \Theta, P, w) \end{aligned}$$

- Step 2: the predictive model weight on $\mathcal{M}_1$, $W_{1,t}$ is given by

$$\sum_{s_t=1}^{2} w_{1,s_t} \times p(s_t|Y^o_{t-1}, \Theta, P, w)$$

  so the predictive model weight on $\mathcal{M}_2$ is $W_{2,t} = 1 - W_{1,t}$.

- Step 3: we now have the conditional likelihood $p(y_t|Y^o_{t-1}, \Theta, P, w)$ as

$$W_{1,t} \times p(y_t|Y^o_{t-1}, \Theta_1, \mathcal{M}_1) + W_{2,t} \times p(y_t|Y^o_{t-1}, \Theta_2, \mathcal{M}_2)$$

  and $\ln L = \ln L + \log p(y_t|Y^o_{t-1}, \Theta, P, w)$

- Step 4: the updated probability of regime $s_t$ $p(s_t = i|Y_t^o, \Theta, P, w)$ is calculated and retained as

$$
\begin{aligned}
p(s_t &= i|Y_t^o, \Theta, P, w) \\
&= p(s_t = i|Y_{t-1}^o, \Theta, P, w, y_t^o) \\
&= \frac{p(s_t = i, y_t^o|Y_{t-1}^o, \Theta, P, w)}{p(y_t^o|Y_{t-1}^o, \Theta, P, w)} \\
&= \frac{p(y_t|Y_{t-1}^o, \Theta, P, w, s_t = i)p(s_t = i|Y_{t-1}^o, \Theta, P, w)}{p(y_t^o|Y_{t-1}^o, \Theta, P, w)}
\end{aligned}
$$

where the predictive density of $y_t$ given $s_t$ is simply given by

$$
\begin{aligned}
& p(y_t|Y_{t-1}^o, \Theta, P, w, s_t) \\
= \; & w_{k=1,s_t} p(y_t|Y_{t-1}^o, \Theta_1, \mathcal{M}_1) + (1 - w_{k=1,s_t})\, p(y_t|Y_{t-1}^o, \Theta_2, \mathcal{M}_2)
\end{aligned}
$$

# B Alternative Prediction Models in the Pool

## B.1 The Three Factor Dynamic Nelson-Siegel Model: DNSM(3)

We now describe the dynamic Nelson-Siegel model (Diebold and Li (2006)). The vector of yields is statistically modeled by

$$\mathbf{y}_t = \mathbf{\Lambda} \times \mathbf{x}_t + u_t \tag{B.1}$$

where

$$
\begin{aligned}
\mathbf{\Lambda} &= \begin{pmatrix} 1 & \frac{1-e^{\tau_1\lambda}}{\tau_1\lambda} & \frac{1-e^{\tau_1\lambda}}{\tau_1\lambda} - e^{-\tau_1\lambda} \\ 1 & \frac{1-e^{\tau_2\lambda}}{\tau_2\lambda} & \frac{1-e^{\tau_2\lambda}}{\tau_2\lambda} - e^{-\tau_2\lambda} \\ \vdots & \vdots & \vdots \\ 1 & \frac{1-e^{\tau_N\lambda}}{\tau_N\lambda} & \frac{1-e^{\tau_N\lambda}}{\tau_N\lambda} - e^{-\tau_N\lambda} \end{pmatrix} && \text{(B.2)} \\
\mathbf{x}_t &= \begin{pmatrix} \mathbf{x}_t^L & \mathbf{x}_t^S & \mathbf{x}_t^C \end{pmatrix}' && \text{(B.3)} \\
u_t &= \begin{pmatrix} u_t(\tau_1) & u_t(\tau_2) & \cdots & u_t(\tau_N) \end{pmatrix}' && \text{(B.4)}
\end{aligned}
$$

Due to the functional form of the factorloadings $\mathbf{\Lambda}$, the latent dynamic factors, $\mathbf{x}_t^L$, $\mathbf{x}_t^S$ and $\mathbf{x}_t^C$ are usually interpreted as level, slope and curvature factors, respectively. The vector of the dynamic factors $\mathbf{x}_t$ is also assumed to follow the first-order stationary vector autoregressive process.

$$\mathbf{x}_t = \kappa + \phi\,(\mathbf{x}_{t-1} - \kappa) + \varepsilon_t \tag{B.5}$$

where

$$\begin{bmatrix} \varepsilon_t \\ u_t \end{bmatrix} \sim \text{i.i.d.}\mathcal{N}\left(0_{(N+3)\times 1}, \begin{pmatrix} V_1\Lambda_1 V_1 & 0_{N\times 3} \\ 0_{3\times N} & \mathbf{D}_1 \end{pmatrix}\right) \tag{B.6}$$

$$V_1 = diag(l_1, l_2, l_3), \tag{B.7}$$

$$\Lambda_1 = \begin{pmatrix} 1 & \rho_{12,1} & \rho_{13,1} \\ \rho_{12,1} & 1 & \rho_{23,1} \\ \rho_{13,1} & \rho_{23,1} & 1 \end{pmatrix} \text{ and } \Omega_1 = V_1\Lambda_1 V_1 \tag{B.8}$$

We let $\mathcal{N}(.,.)$ denote the multivariate normal distribution. The coefficient $\lambda$, referred to the shape parameter, determines the exponential decay rate of the factor loadings. $\Lambda_1$ is the conditional correlation matrix among the factors. Finally, for identification $V_1$ and $\mathbf{D}_1$ are assumed to be both diagonal matrices with positive diagonal elements. For computational convenience we follow Bansal and Zhou (2002) and Chib and Kang (2013), and assume that three basis bonds (the three-month, two-year, and ten-year) are priced exactly by the model.

## B.2 The Three Factor Gaussian Affine Term Structure Model: ATSM(3)

Let $P_t(\tau)$ denote the price of the bond at time $t$ that matures in period $(t+\tau)$. Following Duffie and Kan (1996), we assume that $P_t(\tau)$ is an exponential affine function of the vector of three-dimensional factors $\mathbf{f}_t$ taking the form

$$P_t(\tau) = \exp(-\tau y_t(\tau)) \tag{B.9}$$

where $y_t(\tau)$ is the continuously compounded yield given by

$$y_t(\tau) = -\frac{\log P_t(\tau)}{\tau} = a(\tau) + b(\tau)'\mathbf{f}_t$$

and $a(\tau)$ is a scalar and $b(\tau)$ is a $3 \times 1$ vector, both depending on $\tau$. In order to impose the no-arbitrage condition

$$P_t(\tau) = \mathbb{E}[M_{t,t+1}P_{t+1}(\tau - 1)|\mathbf{f}_t]$$

given the stochastic discount factor (SDF), $M_{t,t+1}$, we solve risk-neutral pricing equation for these coefficients. To do this, we should specify the factor process and the stochastic discount factor (SDF). The distribution of $\mathbf{f}_t$, conditioned on $\mathbf{f}_{t-1}$, is determined by a Gaussian mean-reverting first-order autoregression

$$\mathbf{f}_t = \mathbf{G}\mathbf{f}_{t-1} + \boldsymbol{\eta}_t, \quad \boldsymbol{\eta}_t \sim \text{i.i.d.}\mathcal{N}(\mathbf{0}_{3\times 1}, \boldsymbol{\Omega}_2) \tag{B.10}$$

where

$$V_2 = \mathbf{I}_3, \tag{B.11}$$

$$\Lambda_2 = \begin{pmatrix} 1 & \rho_{12,2} & \rho_{13,2} \\ \rho_{12,2} & 1 & \rho_{23,2} \\ \rho_{13,2} & \rho_{23,2} & 1 \end{pmatrix} \text{ and } \mathbf{\Omega}_2 = V_2 \Lambda_2 V_2 \tag{B.12}$$

$\mathbf{G}$ is a $3 \times 3$ matrix, and $\Lambda_2$ is the conditional correlation matrix. In the sequel, we will express $\boldsymbol{\eta}_t$ in terms of a vector of i.i.d. standard normal variables $\boldsymbol{\omega}_t$ as $\boldsymbol{\eta}_t = \mathbf{L}\boldsymbol{\omega}_t$ where $\mathbf{L}$ is the lower-triangular Cholesky decomposition of $\mathbf{\Omega}_2$

We complete our modeling by assuming that the SDF $M_{t,t+1}$ that converts a time $(t+1)$ payoff into a payoff at time $t$ is given by

$$M_{t,t+1} = \exp\left(-r_t - \frac{1}{2}\boldsymbol{\gamma}_t'\boldsymbol{\gamma}_t - \boldsymbol{\gamma}_t'\boldsymbol{\omega}_{t+1}\right) \tag{B.13}$$

where $r_t$ is the short-rate, $\boldsymbol{\gamma}_t$ is the vector of time-varying market prices of factor risks and $\boldsymbol{\omega}_{t+1}$ is the i.i.d. vector of factor shocks at time $t+1$. We suppose that the short rate and the market price of factor risk are both affine in the factors and of the form

$$r_t = \delta + \boldsymbol{\beta}'\mathbf{f}_t, \tag{B.14}$$

$$\boldsymbol{\gamma}_t = \bar{\boldsymbol{\gamma}} + \Phi\mathbf{f}_t \tag{B.15}$$

, respectively. We find the expressions for the latter functions by the method of undetermined coefficients. Incorporating the assumptions for the factor and SDF process into the risk-neutral pricing formula yields the following recursive system for the unknown functions

$$\begin{aligned} a(\tau) &= \delta/\tau + a(\tau-1) - b(\tau-1)'\mathbf{L}\bar{\boldsymbol{\gamma}} - \frac{\tau}{2}b(\tau-1)'\mathbf{L}\mathbf{L}'b(\tau-1) \\ b(\tau) &= \boldsymbol{\beta}/\tau + (\mathbf{G} - \mathbf{L}\Phi)'b(\tau-1) \end{aligned} \tag{B.16}$$

where $\tau$ runs over the positive integers. These recursions are initialized by setting $a(0) = 0$ and $b(0) = 0_{3\times 1}$.

As defined above, $\mathbf{y}_t$ denotes the vector of yields with different maturities. Also let $\mathbf{a}$ and $\mathbf{b}$ be the corresponding intercept and factor loadings for $\mathbf{y}_t$ obtained from the recursive equations in (B.16). Then for estimation purpose, we follow Chib and Ergashev (2009) and assume that all of the yields are measured with pricing error. The resulting measurement equation has the form

$$\mathbf{y}_t = \mathbf{a} + \mathbf{b}\mathbf{f}_t + \boldsymbol{e}_t, \ \ \boldsymbol{e}_t \sim \mathrm{iid}\mathcal{N}(0, \mathbf{D}_2) \tag{B.17}$$

where $\mathbf{D}_2$ is a diagonal matrix. The transition equation is given by the equation (B.10), which completes the state-space representation.

## B.3 Random-walk Model (RWM)

The third perdition model contained in our pool is the random-walk.

$$\mathbf{y}_t = \mathbf{y}_{t-1} + \boldsymbol{\nu}_t,\ \ \boldsymbol{\nu}_t \sim \mathsf{iid}\mathcal{N}(0, \mathbf{D}_3) \tag{B.18}$$

where $\mathbf{D}_3$ is a $N \times N$ diagonal matrix.

# References


Almeida, C. and Vicenteb, J. (2008), "The role of no-arbitrage on forecasting: Lessons from a parametric term structure model," *Journal of Banking and Finance*, 32(12), 2695–2705.

Bansal, R. and Zhou, H. (2002), "Term structure of interest rates with regime shifts," *Journal of Finance*, 57(5), 463–473.

Carrieroa, A. and Giacomini, R. (2011), "How useful are no-arbitrage restrictions for forecasting the term structure of interest rates?" *Journal of Econometrics*, 164(1), 21–34.

Chib, S. (1998), "Estimation and comparison of multiple change-point models," *Journal of Econometrics*, 86, 221–241.

Chib, S. and Ergashev, B. (2009), "Analysis of multi-factor affine yield curve Models," *Journal of the American Statistical Association*, 104, 1324–1337.

Chib, S. and Kang, K. H. (2013), "Change Points in Affine Arbitrage-free Term Structure Models," *Journal of Financial Econometrics*, 11(2), 302–334.

Christensen, J. H. E., Diebold, F. X., and Rudebusch, G. D. (2011), "The affine arbitrage-free class of Nelson-Siegel term structure models," *Journal of Econometrics*, 164, 4–20.

De Pooter, M. (2007), "Examining the Nelson-Siegel Class of Term Structure Models," *Tinbergen Institute Discussion Paper*.

Diebold, F. X. and Li, C. L. (2006), “Forecasting the term structure of government bond yields,” *Journal of Econometrics*, 130, 337–364.

Duffie, G. (2002), “Term premia and interest rate forecasts in affine models,” *Journal of Finance*, 57, 405–43.

Duffie, G. and Kan, R. (1996), “A yield-factor model of interest rates,” *Mathematical Finance*, 6, 379–406.

Gelfand, A. E. and Ghosh, S. K. (1998), “Model choice: A minimum posterior predictive loss approach,” *Biometrika*, 85, 1–11.

Geweke, J. and Amisano, G. (2011), “Optimal prediction pools,” *Journal of Econometrics*, 164(1), 130–141.

— (2012), “Prediction with Misspecified Models,” *American Economic Review*, 102(3), 482–486.

Moench, E. (2008), “Forecasting the yield curve in a data-rich environment: A no-arbitrage factor-augmented VAR approach,” *Journal of Econometrics*, 146, 26–43.

Waggoner, D. and Zha, T. (2012), “Confronting model misspecification in macroeconomics,” *Journal of Econometrics*, 171(2), 167–184.

Zantedeschi, D., Damien, P., and Polson, N. G. (2011), “Predictive Macro-Finance With Dynamic Partition Models,” *Journal of the American Statistical Association*, 106(494), 427–439.

| Parameter | Density | Mean | S.D. | Range |
|---|---|---|---|---|
| $\kappa_1$ | Normal | 2.00 | 5.00 | $(-\infty, +\infty)$ |
| $\kappa_i$ $(i = 2, 3)$ | Uniform | -1.00 | 5.00 | $(-\infty, +\infty)$ |
| $\phi_i$ $(i = 1, 2, 3)$ | Beta | 0.80 | 0.05 | $(0, 1)$ |
| $\rho_{ij,1}$ $(i \neq j,\ i, j = 1, 2, 3)$ | Uniform | 0.00 | 0.58 | $(-1, 1)$ |
| $l_i$ $(i = 1, 2, 3)$ | Uniform | 1.00 | 0.58 | $(0, 2)$ |
| $50 \times \sigma_{\tau_i^*,1}$ $(i = 1, 2, .., 5)$ | Inverse gamma | 2.00 | 0.30 | $(0, +\infty)$ |

(a) DNSM(3)

| Parameter | Density | Mean | S.D. | Range |
|---|---|---|---|---|
| $\delta$ | Normal | 3.00 | 5.00 | $(-\infty, +\infty)$ |
| $\bar{\boldsymbol{\gamma}}_i$ $(i = 1, 2, 3)$ | Normal | -2.00 | 1.00 | $(-\infty, +\infty)$ |
| $\boldsymbol{\Phi}_i$ $(i = 1, 2, 3)$ | Normal | 0.00 | 1.00 | $(-\infty, +\infty)$ |
| $\mathbf{G}_{ij}$ $(i, j = 1, 2, 3)$ | Uniform | 0.00 | 0.58 | $(-1, 1)$ |
| $\beta_i$ $(i = 1, 2, 3)$ | Normal | 0.50 | 0.30 | $(-\infty, +\infty)$ |
| $\rho_{ij,2}$ $(i \neq j,\ i, j = 1, 2, 3)$ | Uniform | 0.00 | 0.58 | $(-1, 1)$ |
| $50 \times \sigma_{\tau_i^*,2}$ $(i = 1, 2, .., 5)$ | Inverse gamma | 2.00 | 0.30 | $(0, +\infty)$ |

(b) ATSM(3)

| | | | | |
|---|---|---|---|---|
| $50 \times \sigma_{\tau_i^*,3}$ $(i = 1, 2, .., 8)$ | Inverse gamma | 2.00 | 0.30 | $(0, +\infty)$ |

(c) Random-walk

| | | | | |
|---|---|---|---|---|
| $w_{i,s_t}$ $(i = 1, 2,\ s_t = 1, 2)$ | Uniform | 0.50 | 0.29 | $(0, 1)$ |

(d) Model weights

| | | | | |
|---|---|---|---|---|
| $q_{ii}$ $(i = 1, 2)$ | Beta | 0.90 | 0.05 | $(0, 1)$ |

(e) Transition probability

**Table 2: Prior** *This table presents the prior distribution for the parameters. $\tau_i^{*\prime}s$ are the maturities of the bonds priced with errors*

| | PPC |
|---|---|
| *Single* | |
| DNSM | 0.997 |
| ATSM | 1.416 |
| RWM | 0.938 |
| *Constant weights* | |
| NS-AF | 0.995 |
| NS-RW | 0.893 |
| AF-RW | 1.319 |
| NS-AF-RW | 0.772 |
| *Fixed*-NS-AF-RW | **0.766** |
| *Markov-switching weights* | |
| *MS*-NS-AF | 1.087 |
| *MS*-NS-RW | 0.875 |
| *MS*-AF-RW | 1.250 |
| *MS*-NS-AF-RW | 0.877 |

**Table 3: PPC averaged over all maturities and forecast horizons** *This table presents the PPCs averaged over all maturities and forecast horizons. The values of the PPCs are based on 10,000 simulated posterior predictive draws of the bond yields.*

| | DNSM(3) | ATSM(3) | Random-walk |
|---|---|---|---|
| NS-AF | 0.442 | 0.558 | |
| | (0.037) | (0.041) | |
| NS-RW | 0.732 | | 0.268 |
| | (0.032) | | (0.012) |
| AF-RW | | 0.225 | 0.775 |
| | | (0.029) | (0.058) |
| NS-AF-RW | 0.441 | 0.305 | 0.253 |
| | (0.036) | (0.100) | (0.100) |
| Equal-NS-AF-RW | 0.333 | 0.333 | 0.333 |

**Table 4: Constant model weights** *The posterior standard errors are in the parentheses.*

| | DNSM(3) | ATSM(3) | Random-walk | | DNSM(3) | ATSM(3) | Random-walk |
|---|---|---|---|---|---|---|---|
| | | $s_t = 1$ | | | | $s_t = 2$ | |
| NS-AF | 0.964 | 0.036 | | NS-AF | 0.037 | 0.963 | |
| | (0.026) | (0.001) | | | (0.028) | (0.122) | |
| NS-RW | 0.868 | | 0.132 | NS-RW | 0.151 | | 0.849 |
| | (0.051) | | (0.010) | | (0.110) | | (0.221) |
| AF-RW | | 0.766 | 0.234 | AF-RW | | 0.028 | 0.972 |
| | | (0.140) | (0.089) | | | (0.023) | (0.184) |
| NS-AF-RW | 0.741 | 0.127 | 0.133 | NS-AF-RW | 0.123 | 0.425 | 0.452 |
| | (0.023) | (0.014) | (0.015) | | (0.019) | (0.181) | (0.110) |

**Table 5: Markov switching model weights** *The posterior standard errors are in the parentheses.*

| | 3m | 6m | 12m | 24m | 36m | 60m | 84m | 120m |
|---|---|---|---|---|---|---|---|---|
| 3m | 1.00 | | | | | | | |
| 6m | 0.89 | 1.00 | | | | | | |
| 12m | 0.68 | 0.81 | 1.00 | | | | | |
| 24m | 0.55 | 0.77 | 0.88 | 1.00 | | | | |
| 36m | 0.47 | 0.70 | 0.84 | 0.98 | 1.00 | | | |
| 60m | 0.39 | 0.62 | 0.77 | 0.93 | 0.94 | 1.00 | | |
| 84m | 0.35 | 0.58 | 0.74 | 0.90 | 0.92 | 0.94 | 1.00 | |
| 120m | 0.32 | 0.53 | 0.69 | 0.86 | 0.89 | 0.94 | 0.98 | 1.00 |

(a) DNSM(3)

| | 3m | 6m | 12m | 24m | 36m | 60m | 84m | 120m |
|---|---|---|---|---|---|---|---|---|
| 3m | 1.00 | | | | | | | |
| 6m | 0.93 | 1.00 | | | | | | |
| 12m | 0.87 | 0.88 | 1.00 | | | | | |
| 24m | 0.89 | 0.92 | 0.94 | 1.00 | | | | |
| 36m | 0.88 | 0.90 | 0.92 | 0.99 | 1.00 | | | |
| 60m | 0.85 | 0.86 | 0.88 | 0.95 | 0.96 | 1.00 | | |
| 84m | 0.81 | 0.82 | 0.83 | 0.91 | 0.92 | 0.93 | 1.00 | |
| 120m | 0.81 | 0.80 | 0.80 | 0.88 | 0.91 | 0.93 | 0.95 | 1.00 |

(b) ATSM(3)

**Table 6: Model-dependent conditional correlation** *These graphs are based on 50,000 simulated draws of the posterior simulation.*

| | 3m | 6m | 12m | 24m | 36m | 60m | 84m | 120m |
|---|---|---|---|---|---|---|---|---|
| *Single* | | | | | | | | |
| DNSM | 0.983 | 0.952 | 0.959 | 1.036 | 1.087 | 1.054 | 0.989 | 0.913 |
| ATSM | 1.658 | 1.626 | 1.564 | 1.458 | 1.403 | 1.325 | 1.217 | 1.077 |
| RWM | 0.933 | 0.917 | 0.920 | 0.960 | 0.988 | 0.968 | 0.932 | 0.882 |
| *Constant weights* | | | | | | | | |
| NS-AF | 1.134 | 1.124 | 1.066 | 1.014 | 0.989 | 0.937 | 0.886 | 0.809 |
| NS-RW | 0.890 | 0.863 | 0.862 | 0.917 | 0.957 | 0.937 | 0.890 | 0.829 |
| AF-RW | 1.304 | 1.404 | 1.435 | 1.400 | 1.382 | 1.329 | 1.228 | 1.069 |
| NS-AF-RW | 0.821 | 0.803 | **0.773** | 0.774 | 0.785 | 0.772 | 0.747 | 0.703 |
| *Fixed*-NS-AF-RW | 0.818 | 0.802 | 0.778 | **0.768** | **0.775** | **0.761** | **0.736** | **0.691** |
| *Markov-switching weights* | | | | | | | | |
| *MS*-DN-AF | 1.224 | 1.143 | 1.040 | 0.967 | 1.015 | 1.102 | 1.115 | 1.092 |
| *MS*-NS-RW | 0.876 | 0.855 | 0.849 | 0.897 | 0.934 | 0.914 | 0.867 | 0.809 |
| *MS*-AF-RW | 1.318 | 1.350 | 1.339 | 1.289 | 1.272 | 1.237 | 1.160 | 1.033 |
| *MS*-NS-AF-RW | **0.798** | **0.794** | 0.831 | 0.916 | 0.963 | 0.949 | 0.909 | 0.857 |

**Table 7: PPC($\tau$): PPC averaged over all forecast horizons** *This table presents the PPCs averaged over all forecast horizons. The values of the PPCs are based on 10,000 simulated posterior predictive draws of the bond yields.*

| forecast horizon | 1-month ahead | 3-month ahead | 9-month ahead | 12-month ahead |
|---|---|---|---|---|
| *Single* | | | | |
| DNSM | 0.375 | 0.670 | 0.988 | 1.489 |
| ATSM | 0.535 | 0.962 | 1.424 | 2.078 |
| RWM | 0.351 | 0.647 | 0.946 | 1.366 |
| *Constant weights* | | | | |
| NS-AF | 0.341 | 0.655 | 1.010 | 1.486 |
| NS-RW | 0.340 | 0.613 | 0.896 | 1.312 |
| AF-RW | 0.389 | 0.769 | 1.282 | 2.169 |
| NS-AF-RW | **0.282** | 0.528 | 0.792 | 1.117 |
| NS-AF-RW-Fixed | 0.285 | **0.528** | **0.786** | **1.106** |
| *Markov-switching weights* | | | | |
| *MS*-DN-AF | 0.387 | 0.700 | 1.078 | 1.649 |
| *MS*-NS-RW | 0.328 | 0.601 | 0.880 | 1.282 |
| *MS*-AF-RW | 0.401 | 0.779 | 1.238 | 1.981 |
| *MS*-NS-AF-RW | 0.331 | 0.600 | 0.879 | 1.290 |

**Table 8: PPC(h): PPC averaged over all maturities** *This table presents the PPCs averaged over all maturities. The values of the PPCs are based on 10,000 simulated posterior predictive draws of the bond yields.*

| | 3m | 6m | 12m | 24m | 36m | 60m | 84m | 120m |
|---|---|---|---|---|---|---|---|---|
| *Single* | | | | | | | | |
| DNSM | 1.073 | 0.995 | 1.126 | 0.986 | 1.141 | 1.119 | 1.011 | 0.999 |
| ATSM | 2.635 | 2.462 | 2.027 | 1.378 | 1.245 | 1.145 | 1.085 | 0.902 |
| *Constant mixture* | | | | | | | | |
| NS-AF | 1.020 | 1.163 | 1.010 | 0.762 | 0.886 | 0.895 | 0.895 | 0.869 |
| NS-RW | 0.882 | 0.926 | 0.955 | 0.887 | 1.115 | 1.140 | 1.046 | 0.932 |
| AF-RW | 1.643 | 1.608 | 1.368 | 0.955 | 0.926 | 0.929 | 0.908 | 0.788 |
| NS-AF-RW | 0.656 | 0.755 | **0.752** | **0.666** | **0.796** | **0.817** | **0.806** | **0.773** |
| *Fixed*-NS-AF-RW | 0.674 | 0.777 | 0.768 | 0.673 | 0.789 | 0.811 | 0.815 | 0.785 |
| *Markov-switching mixture* | | | | | | | | |
| *MS*-DN-AF | 1.629 | 1.656 | 1.452 | 1.108 | 1.135 | 1.079 | 1.033 | 0.936 |
| *MS*-NS-RW | 0.792 | 0.840 | 0.850 | 0.784 | 0.969 | 1.002 | 0.946 | 0.863 |
| *MS*-AF-RW | 1.844 | 1.817 | 1.523 | 1.034 | 0.974 | 0.979 | 0.935 | 0.808 |
| *MS*-NS-AF-RW | **0.511** | **0.645** | 0.814 | 0.843 | 1.077 | 1.107 | 1.036 | 0.956 |

(a) 1-month ahead

| | 3m | 6m | 12m | 24m | 36m | 60m | 84m | 120m |
|---|---|---|---|---|---|---|---|---|
| *Single* | | | | | | | | |
| DNSM | 1.095 | 1.176 | 1.223 | 1.032 | 1.027 | 1.034 | 0.958 | 0.924 |
| ATSM | 2.584 | 2.444 | 2.085 | 1.482 | 1.253 | 1.085 | 0.995 | 0.887 |
| *Constant mixture* | | | | | | | | |
| NS-AF | 1.272 | 1.282 | 1.077 | 0.791 | 0.818 | 0.876 | 0.874 | 0.892 |
| NS-RW | 0.886 | 0.884 | 0.899 | 0.933 | 1.074 | 1.127 | 1.061 | 0.968 |
| AF-RW | 1.696 | 1.642 | 1.425 | 1.031 | 0.904 | 0.830 | 0.795 | **0.755** |
| NS-AF-RW | 0.710 | 0.755 | 0.719 | **0.666** | **0.745** | 0.803 | **0.793** | 0.779 |
| *Fixed*-NS-AF-RW | 0.739 | 0.789 | 0.744 | 0.683 | 0.749 | **0.801** | 0.795 | 0.785 |
| *Markov-switching mixture* | | | | | | | | |
| *MS*-DN-AF | 1.793 | 1.723 | 1.485 | 1.172 | 1.135 | 1.088 | 1.035 | 0.985 |
| *MS*-NS-RW | 0.835 | 0.847 | 0.864 | 0.862 | 0.970 | 1.025 | 0.976 | 0.899 |
| *MS*-AF-RW | 1.955 | 1.915 | 1.656 | 1.181 | 1.018 | 0.924 | 0.878 | 0.815 |
| *MS*-NS-AF-RW | **0.506** | **0.555** | **0.716** | 0.856 | 1.023 | 1.084 | 1.036 | 0.974 |

(b) 3-month ahead

**Table 9: PPC($\tau$,h)** *These graphs are based on 10,000 simulated draws of the posterior simulation.*

| | 3m | 6m | 12m | 24m | 36m | 60m | 84m | 120m |
|---|---|---|---|---|---|---|---|---|
| *Single* | | | | | | | | |
| DNSM | 1.618 | 1.718 | 1.631 | 1.267 | 1.113 | 0.967 | 0.910 | 0.848 |
| ATSM | 2.877 | 2.943 | 2.638 | 1.982 | 1.670 | 1.235 | 1.062 | 0.934 |
| *Constant mixture* | | | | | | | | |
| NS-AF | 1.563 | 1.535 | 1.230 | 0.823 | 0.789 | 0.828 | 0.860 | 0.900 |
| NS-RW | 0.905 | 0.924 | 0.933 | 0.988 | 1.114 | 1.135 | 1.086 | 1.020 |
| AF-RW | 1.785 | 1.827 | 1.604 | 1.158 | 0.987 | 0.836 | 0.811 | **0.805** |
| NS-AF-RW | 0.955 | 0.962 | 0.857 | 0.724 | **0.740** | **0.783** | **0.806** | 0.822 |
| *Fixed*-NS-AF-RW | 0.878 | 0.903 | 0.818 | **0.723** | 0.775 | 0.812 | 0.827 | 0.841 |
| *Markov-switching mixture* | | | | | | | | |
| *MS*-DN-AF | 1.979 | 1.909 | 1.621 | 1.256 | 1.202 | 1.122 | 1.075 | 1.036 |
| *MS*-NS-RW | 0.861 | 0.884 | 0.880 | 0.910 | 1.022 | 1.054 | 1.018 | 0.963 |
| *MS*-AF-RW | 2.155 | 2.207 | 1.917 | 1.366 | 1.144 | 0.950 | 0.903 | 0.867 |
| *MS*-NS-AF-RW | **0.502** | **0.565** | **0.699** | 0.872 | 1.033 | 1.073 | 1.041 | 0.996 |

(c) 6-month ahead

| | 3m | 6m | 12m | 24m | 36m | 60m | 84m | 120m |
|---|---|---|---|---|---|---|---|---|
| *Single* | | | | | | | | |
| DNSM | 0.927 | 0.983 | 0.988 | 0.934 | 0.985 | 1.125 | 1.137 | 1.087 |
| ATSM | 2.877 | 2.943 | 2.638 | 1.982 | 1.670 | 1.235 | 1.062 | 0.934 |
| *Constant mixture* | | | | | | | | |
| NS-AF | 2.033 | 1.971 | 1.599 | 1.025 | 0.829 | **0.770** | 0.814 | 0.865 |
| NS-RW | 0.885 | 0.902 | 0.923 | 0.976 | 1.103 | 1.137 | 1.108 | 1.035 |
| AF-RW | 2.168 | 2.281 | 2.073 | 1.539 | 1.272 | 0.942 | 0.843 | **0.805** |
| NS-AF-RW | 1.102 | 1.091 | 0.957 | **0.756** | **0.754** | 0.780 | **0.808** | 0.821 |
| *Fixed*-NS-AF-RW | 1.140 | 1.131 | 0.989 | 0.779 | 0.767 | 0.786 | 0.812 | 0.824 |
| *Markov-switching mixture* | | | | | | | | |
| *MS*-DN-AF | 2.319 | 2.251 | 1.942 | 1.454 | 1.309 | 1.156 | 1.110 | 1.068 |
| *MS*-NS-RW | 0.868 | 0.875 | 0.875 | 0.895 | 1.009 | 1.061 | 1.045 | 0.982 |
| *MS*-AF-RW | 2.629 | 2.721 | 2.441 | 1.808 | 1.503 | 1.117 | 0.993 | 0.914 |
| *MS*-NS-AF-RW | **0.525** | **0.576** | **0.695** | 0.858 | 1.020 | 1.084 | 1.069 | 1.020 |

(d) 12-month ahead

**Table 10: PPC($\tau$,h)** *These graphs are based on 10,000 simulated draws of the posterior simulation.*