

THE UNIVERSITY OF
SYDNEY

Economics Working Paper Series

2016 - 05

Forecasting the Term Structure of
Interest Rates with Potentially
Misspecified Models

Yunjong Eo & Kyu Ho Kang

March 2016

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

YUNJONG EO[†]
KYU HO KANG[‡]

March 2016

Abstract

This paper assesses the predictive gains of the pooling method in yield curve prediction. We consider three individual yield curve prediction models: the dynamic Nelson-Siegel model (DNS) and the arbitrage-free Nelson-Siegel model in addition to the random walk (RW) model as a benchmark. 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. We investigate whether combining the possibly misspecified models in a linear form helps improve the predictive accuracy. To do this, we evaluate the out-of-sample forecasts of the pooled models in comparison with the individual models. In terms of density prediction, the pooled model of the DNS and RW models consistently outperforms those individual models regardless of maturities and forecast horizons. Our findings strongly suggest that one needs to try the pooling method rather than choosing one of the alternative models.

JEL classification: G12, C11, F37

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

*We thank all seminar participants at the Texas A&M University, Society for Nonlinear Dynamics and Econometrics Conference, Recent Developments in Financial Econometrics and Applications, Reserve Bank of New Zealand, University of California, Riverside, EFaB@Bayes 250 Workshop, Hitotsubashi University, 2013 Econometric Society Australasian Meeting, and the Bank of Korea. All remaining errors are our own.

[†]*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:* kyuhok@korea.ac.kr

1 Introduction

Forecasting the yield curve is extremely important for bond portfolio risk management, monetary policy, business cycle analysis, and so forth. In the previous literature, three classes of yield curve prediction models have been widely used. One is arbitrage-free affine term structure models, which take a theoretical bond pricing approach¹. This approach provides many economically interpretable outcomes such as term premium and term structure of real interest rates. Despite its flexibility and microfoundation, 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 (DNS) model². 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 (RW), and it is often used as a benchmark in forecasting ability comparison.

Interestingly, beating the RW is a challenging task although the DNS or the arbitrage free term structure model 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) find that the three-factor DNS outperforms the RW at 1-month-ahead horizon for short maturities, but for long-term bond yields the RW dominates the DNS. Zantedeschi et al. (2011) confirm that the RW forecasts better in the short run whereas at three- and six-step-ahead forecast horizons the predictions from their DNS with time-varying factorloadings are much improved. Moench (2008) finds that the arbitrage free affine term structure model forecasts better than the RW at he six-month maturity only. Using an affine term structure model, Carrieroa and Giacomini (2011) produce one-step-ahead forecasts and find positive prediction gains against the RW at intermediate and long maturities, not at short maturities.

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

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

These mixed results for out-of-sample prediction comparison strongly indicate that all the 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 DNS excluding other models.

As an alternative way to consider the model misspecifications we take the pooling method recently suggested by Geweke and Amisano (2012), 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 pooled models of two or three of the individual models are compared in terms of out-of-sample predictive accuracy. When considering the pooling approach, the weights are equally given, to be estimated as constant parameters, or follow the first-order Markov-switching process. Using these pooled models we forecast the monthly yields with eight different maturities over the forecast horizons of one through twelve months, and conduct model comparison analysis using the root mean squared error and the posterior predictive criterion.

According to our empirical results based on 108 out-of-sample periods, the pooled model of the DNS and RW with equal model weights dominates all individual models across all maturities and forecast horizons in terms of the bond yield density forecasting. Further, the predictive gains from using the pooling method are remarkable. In contrast, in the point forecasting, we do not find a strong evidence that the pooling method can help improve the predictive accuracy. Our findings strongly indicate that for density prediction the pooling method is worth trying, so one should examine the pooled models of the individual models in addition to choosing one of them.

The rest of the paper is organized as follows. Section 2 briefly describes our pooling methods, and Section 3 specifies all competing models including the three individual prediction models and the pooled models. In Section 4 we lay out our Bayesian Markov

chain Monte Carlo (MCMC) algorithm for estimation. Sections 5 and 6 provide the empirical results and discuss their implications. Finally, Section 7 concludes the paper.

2 Pooling Method

In this section we illustrate the pooling method using an example of two prediction models, \mathcal{M}_1 and \mathcal{M}_2 . Let Θ_1 and Θ_2 be the set of parameters in \mathcal{M}_1 and \mathcal{M}_2 , respectively. The set of maturities is $\{\tau_i\}_{i=1}^N$, the τ -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

$$\mathbf{y}_t = (y_t(\tau_1), y_t(\tau_2), \dots, y_t(\tau_N))'.$$

We let $Y_t = \{y_i\}_{i=1}^t$ denote the observed yield curve data up to time t . Then Geweke and Amisano (2011) study predictive densities of the form

$$w_1 \times p(\mathbf{y}_t | Y_{t-1}, \Theta_1, \mathcal{M}_1) + (1 - w_1) \times p(\mathbf{y}_t | Y_{t-1}, \Theta_2, \mathcal{M}_2) \quad (2.1)$$

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. They replace w_1 in equation (2.1) 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

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

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

$$w_{1,s_t} \times p(\mathbf{y}_t | Y_{t-1}, \Theta_1, \mathcal{M}_1) + (1 - w_{1,s_t}) \times p(\mathbf{y}_t | Y_{t-1}, \Theta_2, \mathcal{M}_2)$$

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

$$\log p(Y_T | \Theta, Q, w) = \sum_{t=1}^T \log p(\mathbf{y}_t | Y_{t-1}, \Theta, Q, w)$$

where the regime s_t is integrated out because it is never observed by econometricians. For more details for the likelihood computation, refer to 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, 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 Models

In this section we briefly discuss the three individual yield curve prediction models which have been commonly used for yield curve prediction. Then, we introduce our pooled models as alternative prediction models.

3.1 Individual Yield Curve Prediction Models

3.1.1 Dynamic Nelson-Siegel Model

We begin by describing the three-factor dynamic Nelson-Siegel model (DNS, Diebold and Li (2006)). In the DNS model, the bond yields are specified as a linear function of the vector of three exogenous latent factors \mathbf{x}_t

$$\mathbf{y}_t | \mathbf{x}_t, \boldsymbol{\Sigma}_{NS} \sim \mathcal{N}(\boldsymbol{\Lambda} \mathbf{x}_t, \boldsymbol{\Sigma}_{NS}) \quad (3.1)$$

where $\mathcal{N}(\cdot, \cdot)$ denotes the multivariate normal distribution, the measurement error variances $\boldsymbol{\Sigma}_{NS}$ is a diagonal matrix, λ is a decay parameter,

$$\boldsymbol{\Lambda} = \begin{pmatrix} 1 & \frac{1-e^{-\tau_1\lambda}}{\tau_1\lambda} & \frac{1-e^{-\tau_1\lambda}}{\tau_1\lambda} - e^{-\tau_1\lambda} \\ 1 & \frac{1-e^{-\tau_2\lambda}}{\tau_2\lambda} & \frac{1-e^{-\tau_2\lambda}}{\tau_2\lambda} - e^{-\tau_2\lambda} \\ \vdots & \vdots & \vdots \\ 1 & \frac{1-e^{-\tau_N\lambda}}{\tau_N\lambda} & \frac{1-e^{-\tau_N\lambda}}{\tau_N\lambda} - e^{-\tau_N\lambda} \end{pmatrix}, \quad (3.2)$$

$$\text{and } \mathbf{x}_t = \left(\mathbf{x}_t^L \quad \mathbf{x}_t^S \quad \mathbf{x}_t^C \right)' . \quad (3.3)$$

The vector of the dynamic factors \mathbf{x}_t is assumed to follow the first-order stationary vector autoregressive (VAR) process,

$$\mathbf{x}_t | \mathbf{x}_{t-1}, \kappa, \phi, \Omega_{NS} \sim \mathcal{N}(\kappa + \phi \mathbf{x}_{t-1}, \Omega_{NS}) . \quad (3.4)$$

For stationarity, the absolute of all eigen values of $\phi : 3 \times 3$ is constrained to be less than 1, and \mathbf{x}_0 is assumed to be generated from the unconditional distribution of \mathbf{x}_t . 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 identified and interpreted as level, slope, and curvature effects, respectively. The coefficient λ , referred to the decay parameter, determines the exponential decay rate of the factor loadings and it is fixed at 0.0607 like in Diebold and Li (2006).

Then, the set of the parameters to be estimated in the DNS model is

$$\Theta_{NS} = \{ \kappa, \phi, \Omega_{NS} = V_{NS} \Gamma_{NS} V_{NS}, \Sigma_{NS} \}$$

where $V_{NS} : 3 \times 3$ and $\Gamma_{NS} : 3 \times 3$ are the factor shock volatility and correlation matrix, respectively.

Finally, as the equations (3.1) and (3.4) are a standard state-space representation, the resulting conditional density of \mathbf{y}_t at each time point

$$p(\mathbf{y}_t | Y_{t-1}, \Theta_{NS}, \mathcal{M}_{NS}) \quad (3.5)$$

can be easily obtained by the usual Kalman filtering procedure.

3.1.2 Arbitrage-free Nelson-Siegel Model

Bond Prices Arbitrage-free Nelson-Siegel Model (AFNS) model is a theoretical bond pricing approach based on a partial equilibrium while the DNS model is a purely statistical approach. Satisfying the arbitrage-free condition, the bond prices are endogenously determined by economic agents who know the model parameters.

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

$$P_t(\tau) = \exp(-\tau y_t(\tau)) \quad (3.6)$$

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 3×1 vector, both depending on τ . 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 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 (SDF). The distribution of \mathbf{f}_t , conditioned on \mathbf{f}_{t-1} , is determined by a Gaussian mean-reverting first-order autoregression

$$\mathbf{f}_t = G\mathbf{f}_{t-1} + \boldsymbol{\eta}_t, \quad \boldsymbol{\eta}_t \sim \mathcal{N}(\mathbf{0}, \Omega_{AF}) \quad (3.7)$$

where $G : 3 \times 3$ is VAR coefficients. 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 Ω_{AF} .

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

$$M_{t,t+1} = \exp\left(-r_t - \frac{1}{2}\boldsymbol{\gamma}'_t\boldsymbol{\gamma}_t - \boldsymbol{\gamma}'_t\boldsymbol{\omega}_{t+1}\right) \quad (3.8)$$

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

$$r_t = \delta + \boldsymbol{\beta}'\mathbf{f}_t, \quad (3.9)$$

$$\boldsymbol{\gamma}_t = \bar{\boldsymbol{\gamma}} + \boldsymbol{\Phi}\mathbf{f}_t \quad (3.10)$$

, respectively.

Given the assumptions above, we find the solutions for $a(\tau)$ and $b(\tau)$ in terms of the structural parameters by using 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

$$\begin{aligned} a(\tau) &= \delta/\tau + a(\tau - 1) - b(\tau - 1)' \mathbf{L} \bar{\gamma} - \frac{\tau}{2} b(\tau - 1)' \Omega_{AF} b(\tau - 1) \\ b(\tau) &= \beta/\tau + (G - \mathbf{L}\Phi)' b(\tau - 1) \end{aligned} \quad (3.11)$$

where $G^Q = G - \mathbf{L}\Phi$ and τ runs over the positive integers. These recursions are initialized by setting $a(0) = 0$ and $b(0) = \mathbf{0}_{3 \times 1}$.

Econometric Model Now we express the AFNS model into an econometric model for estimation. First of all, we let \mathbf{a} and \mathbf{b} be the corresponding intercept and factor loadings for \mathbf{y}_t obtained from the recursive equations in (3.11).

$$\begin{aligned} \mathbf{a} &= (a(\tau_1) \ a(\tau_2) \ \cdots \ a(\tau_N))' : N \times 1 \\ \mathbf{b} &= (b(\tau_1) \ b(\tau_2) \ \cdots \ b(\tau_N))' : N \times 3 \end{aligned} \quad (3.12)$$

For computational convenience, we follow Bansal and Zhou (2002) and Chib and Kang (2013), and assume that three basis bonds (the three-month, three-year, and ten-year) are observed without errors. These three maturities are the first, fifth, and eighth maturities in our data set. This means that there is a one-to-one mapping between the three latent factors and basis yields such as

$$\begin{aligned} \mathbf{y}_t^B &= \mathbf{a}_B + \mathbf{b}_B \mathbf{f}_t \\ \text{or } \mathbf{f}_t &= (\mathbf{b}_B)^{-1} \times (\mathbf{y}_t^B - \mathbf{a}_B) \end{aligned}$$

where

$$\begin{aligned} \mathbf{y}_t^B &= (y(\tau_1) \ y(\tau_5) \ y(\tau_8))', \\ \mathbf{a}_B &= (a(\tau_1) \ a(\tau_5) \ a(\tau_8))' : 3 \times 1, \\ \text{and } \mathbf{b}_B &= (b(\tau_1) \ b(\tau_5) \ b(\tau_8))' : 3 \times 3. \end{aligned}$$

Let \mathbf{a}_{NB} and \mathbf{b}_{NB} denote the intercept term and factorloadings corresponding to the non-basis yields. The non-basis yields, denoted by \mathbf{y}_t^{NB} , are observed with errors,

$$\mathbf{y}_t^{NB} | \mathbf{a}_{NB}, \mathbf{b}_{NB}, \mathbf{f}_t \sim \mathcal{N}(\mathbf{a}_{NB} + \mathbf{b}_{NB} \mathbf{f}_t, \boldsymbol{\Sigma}_{AF}).$$

where $\boldsymbol{\Sigma}_{AF} : 5 \times 5$ is a diagonal matrix.

Identifying Restrictions For factor identification we impose two restrictions. First of all, the matrix \mathbf{G}^Q has the form

$$\mathbf{G}^Q = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \exp(-g^Q) & g^Q \exp(-g^Q) \\ 0 & 0 & \exp(-g^Q) \end{bmatrix}. \quad (3.13)$$

Secondly, the vector $\boldsymbol{\beta}$ is constrained to be

$$\boldsymbol{\beta} = (1, 1, 0,)'.$$

As proved by Niu and Zeng (2012), due to these restrictions, \mathbf{b} in equation (3.12) reduces to exactly the form of the dynamic Nelson-Siegel factorloadings $\boldsymbol{\Lambda}$. Therefore, the factors \mathbf{f}_t are also identified as the level, slope, and curvature effects as in the DNS model.

Unlike in the DNS model, however, the structural parameters in the AFNS model determining the intercept term, factorloadings, factor persistence, and factor volatilities are jointly estimated. Meanwhile, in the DNS model, the intercept term and factorloadings are fixed, and Θ_{NS} consists of the parameters in the factor process and measurement error variances.

Suppose that $V_{AF} : 3 \times 3$ is the factor shock volatility and $\Gamma_{AF} : 3 \times 3$ is the factor shock correlation matrix. In addition, following Dai, Singleton, and Yang (2007) we fix δ at the sample mean of the short rate because the short rate is highly persistent and δ tends to be estimated very inefficiently. After all, the set of parameters in the AFNS model to be estimated is

$$\Theta_{AF} = \{G, g^Q, \Omega_{AF} = V_{AF} \Gamma_{AF} V_{AF}, \boldsymbol{\Sigma}_{AF}\}.$$

The resulting conditional density of \mathbf{y}_t is obtained as

$$p(\mathbf{y}_t | Y_{t-1}, \Theta_{AF}, \mathcal{M}_{AF}) = p(\mathbf{y}_t^{NB} | \mathbf{y}_t^B, \Theta_{AF}, \mathcal{M}_{AF}) \times p(\mathbf{y}_t^B | \mathbf{y}_{t-1}^B, \Theta_{AF}, \mathcal{M}_{AF}) \quad (3.14)$$

$$= \mathcal{N}(\mathbf{y}_t^{NB} | \mathbf{a}_{NB} + \mathbf{b}_{NB} \mathbf{f}_t, \boldsymbol{\Sigma}_{AF}) \times \mathcal{N}(\mathbf{f}_t | G \mathbf{f}_{t-1}, \boldsymbol{\Omega}_{AF}) \times |\mathbf{b}_B^{-1}|$$

where $\mathbf{f}_t = (\mathbf{b}_B)^{-1} \times (\mathbf{y}_t^B - \mathbf{a}_B)$, and $\mathcal{N}(x|m, V)$ denotes the multivariate normal density of x with mean m and variance-covariance V .

3.1.3 Random-walk Model

The third individual prediction model contained in our pool is the random-walk,

$$\mathbf{y}_t | \mathbf{y}_{t-1}, \boldsymbol{\Sigma}_{RW} \sim \mathcal{N}(\mathbf{y}_{t-1}, \boldsymbol{\Sigma}_{RW}) \quad (3.15)$$

where $\boldsymbol{\Theta}_{RW} = \boldsymbol{\Sigma}_{RW}$ is an $N \times N$ diagonal matrix. The conditional density of \mathbf{y}_t is simply

$$p(\mathbf{y}_t | Y_{t-1}, \boldsymbol{\Theta}_{RW}, \mathcal{M}_{RW}) = \mathcal{N}(\mathbf{y}_t | \mathbf{y}_{t-1}, \boldsymbol{\Sigma}_{RW}). \quad (3.16)$$

As demonstrated by Altavilla, Giacomini, and Ragusa (2014) and Diebold and Li (2006), outperforming the random walk in terms of out-of-sample yield curve forecasting is quite challenging, and it is often used as a benchmark in prediction ability comparison. For this reason, we include the random-walk in our pool.

3.2 Pooled Models

Diebold and Li (2006) show that overall DNS produces better forecast accuracy in out-of-sample prediction compared to Duffie (2002)'s best essentially affine model although the RW forecasts better at short forecast horizons. Then the Bayesian model-averaging method based on the Bayes factor would yield nearly one weight on the DNS model excluding the AFNS and RW models. 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 15 competing pooled 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

Table 1: List of model combinations

	DNS	AFNS	RW
<i>Single</i>			
DNS	○	×	×
AFNS	×	○	×
RW	×	×	○
<i>Constant weight</i>			
NS-AF-Const	○	○	×
NS-RW-Const	○	×	○
AF-RW-Const	×	○	○
NS-AF-RW-Const	○	○	○
NS-AF-Equal	○	○	×
NS-RW-Equal	○	×	○
AF-RW-Equal	×	○	○
NS-AF-RW-Equal	○	○	○
<i>Markov-switching weight</i>			
NS-AF-MS	○	○	×
NS-RW-MS	○	×	○
AF-RW-MS	×	○	○
NS-AF-RW-MS	○	○	○

Note: ○ indicates the inclusion of the individual model into the pool.

constant or time-varying. For the pooled models with constant weights, the weights are to be estimated or equally given. For instance, the NS-AF-RW-Const is the pooled model of the DNS, AFNS and the random-walk with a constant weight. The NS-AF-RW-MS is the pooled model with Markov regime-switching weights, in which the model weights vary over time according to the Markov process. Meanwhile, in NS-AF-RW-Equal each of the model weights is fixed at 1/3.

4 Posterior Simulation

This section discusses the posterior sampling scheme for the most general model among the competing models, NS-AF-RW-MS model. The other models can be estimated as a special case of NS-AF-RW-MS model. In the Bayesian context our pooled model with Markov regime-switching weights is the joint prior distribution of the yield curves

($\mathbf{Y} = \{\mathbf{y}_t\}_{t=1}^T$), the regime indicators ($\mathbf{S} = \{s_t\}_{t=1}^T$), continuous latent variables ($\mathbf{X} = \{\mathbf{x}_t\}_{t=1,2,\dots,T}$ and $\mathbf{F} = \{\mathbf{f}_t\}_{t=1,2,\dots,T}$), and the model parameters ($\boldsymbol{\psi} = \{\Theta_{NS}, \Theta_{AF}, \Theta_{RW}, Q, w\}$). Given the joint density, our objective is to simulated the posterior distribution of $(\boldsymbol{\psi}, \mathbf{X}, \mathbf{F}, \mathbf{S})$ conditioned on the observed yield curves \mathbf{Y} . Its density has the form

$$\pi(\boldsymbol{\psi}, \mathbf{X}, \mathbf{F}, \mathbf{S} | \mathbf{Y}) \propto f(\mathbf{Y} | \boldsymbol{\psi}, \mathbf{X}, \mathbf{F}, \mathbf{S}) \times f(\mathbf{X}, \mathbf{F} | \boldsymbol{\psi}) \times p(\mathbf{S} | \boldsymbol{\psi}) \times \pi(\boldsymbol{\psi}) \quad (4.1)$$

where $\pi(\boldsymbol{\psi})$ is the prior density of the parameters, $p(\mathbf{S} | \boldsymbol{\psi})$ is the prior density function for regime-indicators given the parameters and it is specified as the discrete two-state Markov switching process. $f(\mathbf{X}, \mathbf{F} | \mathbf{S}, \boldsymbol{\psi})$ is the prior density of the factors and $f(\mathbf{Y} | \boldsymbol{\psi}, \mathbf{X}, \mathbf{F}, \mathbf{S})$ is the joint conditional density of the observed data.

4.1 Prior

Our prior for $(\Theta_{NS}, \Theta_{AF})$ 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. Particulary, the prior of Θ_{AF} should be chosen very carefully because the bond yields are highly nonlinear to the parameters and the likelihood surface of the AFNS tends to be irregular. The irregularity of the posterior surface can be more serious or mitigated depending on the choice of the prior. We use the prior in the work of Chib and Kang (2016). They arrive at the 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 until the mean of the resulting prior-implied unconditional distribution of yield curve is mildly upward sloping and concave. As Chib and Ergashev (2009) and Chib and Kang (2016) show, this simulation-based prior can smooth out the many local modes of the likelihood surface.

In addition, we assume that ϕ and G are diagonal since this restriction does not only helps improve the predictive accuracy, but also reduces the computational burden (Christensen et al. (2011)). Table 2 summarizes our prior.

For regime identification, we impose a restriction that the weight on the DNS model \mathcal{M}_{NS} , denoted by w_{NS,s_t} , should be higher in regime 2 than in regime 1

$$0 < w_{NS,s_t=1} < w_{NS,s_t=2} < 1 \text{ and } 0 < w_{NS,s_t} + w_{AF,,s_t} < 1$$

Table 2: Prior

Parameter	Density	Mean	S.D.
$1200 \times \kappa$	Normal	(0.2, -0.1, -0.1)'	(0.01, 0.01, 0.01)'
$\text{diag}(\phi)$	Normal	(0.9, 0.9, 0.9)'	(0.05, 0.05, 0.05)'
$1200 \times \text{diag}(V_{NS})$	Inverse Gamma	1.000	0.200
$\rho_{NS,ij}$ ($i \neq j, i, j = 1, 2, 3$)	Uniform	0.000	0.580
$1.4 \times 10^4 \times \text{diag}(\Sigma_{NS})$	Inverse Gamma	2.000	0.300

(a) DNS

Parameter	Density	Mean	S.D.
$\bar{\gamma}$	Normal	(-0.15, -0.07, 0)'	(0.01, 0.01, 0.01)'
\mathbf{g}^Q	Normal	0.067	0.031
$\text{diag}(G)$	Normal	(0.9, 0.9, 0.9)'	(0.05, 0.05, 0.05)'
$10^4 \times \text{diag}(V_{AF})$	Inverse Gamma	(2, 2.5, 5)'	(0.25, 0.25, 0.25)'
$\rho_{AF,ij}$ ($i \neq j, i, j = 1, 2, 3$)	Uniform	0.000	0.580
$1.4 \times 10^4 \times \text{diag}(\Sigma_{AF})$	Inverse gamma	2.000	0.300

(b) AFNS

$1.4 \times 10^4 \times \text{diag}(\Sigma_{RW})$	Inverse gamma	2.000	0.300
---	---------------	-------	-------

(c) Random-walk

w_{i,s_t} ($s_t = 1, 2$)	Uniform	0.500	0.29
------------------------------	---------	-------	------

(d) Model weights

q_{ii} ($i = 1, 2$)	Beta	0.900	0.05
-------------------------	------	-------	------

(e) Transition probability

Note: $\rho_{NS,ij}$ and $\rho_{AF,ij}$ are the (i, j) elements of Γ_{NS} and Γ_{AF} , respectively.

where w_{AF,s_t} is the weight on the AFNS model. For the *MS*-AF-RW model, the restriction is replaced by

$$0 < w_{AF,s_t=1} < w_{AF,s_t=2} < 1.$$

All the restrictions including factor identification and regime identification restrictions are imposed through the prior.

4.2 MCMC Sampling

Because the joint posterior distribution in equation (4.1) is not analytically tractable, we rely on a MCMC simulation method and sample the parameters, factors, regimes, and predictive yield curves recursively from the joint posterior distribution as follows:

Algorithm 1: MCMC sampling

- Step 1: Sample $\Theta_{NS}, \Theta_{AF}, \Theta_{RW}, w | \mathbf{Y}, Q$
 - Step 1(a): Sample $\Theta_{NS}, \Theta_{AF}, \Theta_{RW}, w | \mathbf{Y}, Q$
 - Step 1(b): Sample $Q | \mathbf{Y}, \mathbf{S}, \Theta_{NS}, \Theta_{AF}, \Theta_{RW}, w$
- Step 2: Sample $\mathbf{S} | \mathbf{Y}, \psi$
- Step 3: Sample $\mathbf{X} | \mathbf{Y}, \psi$ and $\mathbf{F} | \mathbf{Y}, \psi$
- Step 4: Sample $\{\mathbf{y}_{T+h}\}_{h=1}^H | \mathbf{Y}, \mathbf{X}, \mathbf{F}, \mathbf{S}, \psi$

More details about MCMC sampling in each step can be found in B .

5 Predictive Accuracy Measures

We evaluate the predictive accuracy using two measures. One is the posterior predictive criterion (PPC) of Gelfand and Ghosh (1998), which is often used for density forecast evaluation. The other measure is the root mean squared error (RMSE), which is a popular measure of point forecast accuracy.

5.0.1 Posterior Predictive Criterion

We follow Zantedeschi et al. (2011) and Chib and Kang (2013), and evaluate the predictive accuracy of the density forecasts by using the posterior predictive criterion (PPC) of Gelfand and Ghosh (1998). Given each pooled model and in-sample data \mathbf{Y} , the PPC for h -month-ahead posterior predictive density of τ -period bond yield, $\text{PPC}_{\mathbf{Y}}(\tau, h)$ is computed as

$$\text{PPC}_{\mathbf{Y}}(\tau, h) = D_{\mathbf{Y}}(\tau, h) + W_{\mathbf{Y}}(\tau, h)$$

where

$$D_{\mathbf{Y}}(\tau, h) = \text{Var}(y_{T+h}(\tau)|\mathbf{Y})$$

and

$$W_{\mathbf{Y}}(\tau, h) = [y_{T+h}^o(\tau) - \mathbb{E}(y_{T+h}(\tau)|\mathbf{Y})]^2$$

The term $D_{\mathbf{Y}}(\tau, h)$ is the posterior variance of the predictive yield, which is large when the model has too many restrictions or redundant parameters. The term $W_{\mathbf{Y}}(\tau, h)$ is the squared errors and evaluates the predictive goodness-of-fit. Let T_H denote the number of the out-of-sample datasets. Then, the $\text{PPC}(\tau, h)$ is obtained as the average of $\text{PPC}_{\mathbf{Y}}(\tau, h)$ over the 108(= T_H) out-of-samples.

$$\text{PPC}(\tau, h) = \frac{1}{T_H} \sum_{\mathbf{Y}} \text{PPC}_{\mathbf{Y}}(\tau, h)$$

By definition, smaller values of PPC and RMSE are preferable.

5.0.2 Root Mean Squared Error

Suppose that $y_{T+h}^o(\tau)$ is the realized τ -month bond yield, and $\mathbb{E}(y_{T+h}(\tau)|\mathbf{Y})$ is the posterior mean of $y_{T+h}(\tau)$. Then, RMSE of the h-month-ahead bond yield with τ -month to maturity, denoted by $\text{RMSE}(\tau, h)$, is given by

$$\text{RMSE}(\tau, h) = \sqrt{\frac{1}{T_H} \sum_{\mathbf{Y}} [y_{T+h}^o(\tau) - \mathbb{E}(y_{T+h}(\tau)|\mathbf{Y})]^2}.$$

6 Empirical Results

In this section, we evaluate and compare the out-of-sample forecasting performance of the pooled models. In particular, we concentrate on the predictive gain of pooling the individual yield curve prediction models.

Our data comprise monthly yields on U.S. government bonds ranging February 1994 to December 2013. The set of maturities in month is $\{3, 6, 12, 24, 36, 60, 84, 120\}$. The data are obtained from the Federal Reserve Bank of St. Louis economic data.

For model comparison, we calculate PPC and RMSE values from rolling window estimation. The window size is 120 months and the forecast horizon is one- through

Table 3: Rolling Windows: in-sample and out-of-sample

	in-sample	out-of-sample
1	Feb. 1994 - Jan. 2004	Feb. 2004 - Jan. 2005
2	Mar. 1994 - Feb. 2004	Mar. 2004 - Feb. 2005
3	April 1994 - Mar. 2004	April 2004 - Mar. 2005
	⋮	⋮
54	Aug. 1997 - July 2008	Aug. 2008 - July 2009
55	Sept. 1997 - Aug. 2008	Sept. 2008 - Aug. 2009
	⋮	⋮
107	Dec. 2002 - Nov. 2012	Dec. 2012 - Nov. 2013
108	Jan. 2003 - Dec. 2012	Jan. 2013 - Dec. 2013

twelve-months. The first out-of-sample period is February 2004 to January 2005 and the corresponding in-sample period is February 1994 to January 2004. We simulate the predictive yield curves and compute the squared errors and $PPC_{\mathbf{Y}}$. Then, we move forward the in-sample and out-of-sample by one-month, and compute the squared errors and $PPC_{\mathbf{Y}}$. This procedure is repeated 108 times where the last out-of-sample period is January 2013 to December 2013. After all, we obtain 108 squared errors and $PPC_{\mathbf{Y}}$ s for each pair of maturity and forecast horizon. Using them we are able to compute and compare the PPCs and RMSEs of the pooled models with those of the individual models. The pairs of the in-sample and out-of-sample are given in Table 3.

6.1 PPC Comparison

Table 4 presents the best models based on the PPC. For easier reference, we use shorter model specification indices: ‘ N ’ stands for the DNS, ‘ A ’ for the AFNS, ‘ R ’ for the RW. The subscript ‘ E ’ stands for the equal weights, ‘ C ’ for the constant weights, and ‘ M ’ for the Markov-switching weights. For example, ‘ NR_E ’ in Table 4 indicates the pooled model of DNS and RW with equal weight model (i.e., NS-RW-*Equal*).

Table 4 shows that all the best models are one of the pooled models. In particular, the equal weight NS-RW model, NS-RW-*Equal* outperforms the other models for the maturities of three months through three years regardless of the forecast horizon. For the seven-year bond yield, the NS-RW-*Equal* forecasts best at all the forecast horizons

except for the one-month-ahead forecast. Further, pooling all individual models with equal weights is found to help improve the predictive accuracy of five- and ten-year bond yields.

There are three important implications from the PPC comparison. First, as shown in Tables 5 and 6, the predictive gain from the pooling method is remarkable. Especially, the NS-RW-*Equal* consistently outperforms the three individual models across all maturities and forecast horizons. Meanwhile, the Bayesian model averaging tends to pick one of the three individual models. The selected model is usually either DNS or RW model, but their density forecasting performance is found to be worse than that of the pooled model of the DNS and RW. This does not imply that pooling individual models guarantees better density forecasts than choosing one of the individual models. As the Tables 5 and 6 demonstrate, many pooled models produce less accurate forecasts than does the DNS or RW model. As the more models are included in the pool, the more parameters are to be estimated and prediction can be inefficient since the information contained in the data is fixed.

Second, although the RW model is not the best prediction model for density forecasting, all the best model combinations include the RW model in the pool. Therefore, the RW plays an essential role in improving the density forecasting as well as the point forecasting. Third, the theoretical model, AFNS can help produce more accurate density forecasts of long-term bond yields, and thus this model should not be excluded from the pool. This finding is very meaningful in sense that the AFNS has been less popular than the DNS and RW models in yield curve prediction.

It is interesting that the best pooled models have equal model weights. Estimating the weights which are either constant or follow the Markov-process can improve in-sample-fit, but at the same time, it causes inefficiency. This inefficiency has been already pointed in the literature, for instance, Smith and Wallis (2009), Stock and Watson (2004), and Winkler and Clemen (1992).

Table 4: Best Models: PPC

	3m	6m	12m	24m	36m	60m	84m	120m
1-month-ahead	NR_E	NR_E	NR_E	NR_E	NR_E	AR_E	AR_E	AR_E
2-month-ahead	NR_E	NR_E	NR_E	NR_E	NR_E	AR_E	NR_E	AR_E
3-month-ahead	NR_E	NR_E	NR_E	NR_E	NR_E	AR_E	NR_E	NR_E
4-month-ahead	NR_E	NR_E	NR_E	NR_E	NR_E	NAR_E	NR_E	NR_E
5-month-ahead	NR_E	NR_E	NR_E	NR_E	NR_E	NAR_E	NR_E	NAR_E
6-month-ahead	NR_E	NR_E	NR_E	NR_E	NR_E	NAR_E	NR_E	NAR_E
7-month-ahead	NR_E	NR_E	NR_E	NR_E	NR_E	NAR_E	NR_E	NAR_E
8-month-ahead	NR_E	NR_E	NR_E	NR_E	NR_E	NAR_E	NR_E	NAR_E
9-month-ahead	NR_E	NR_E	NR_E	NR_E	NR_E	NAR_E	NR_E	NAR_E
10-month-ahead	NR_E	NR_E	NR_E	NR_E	NR_E	NAR_E	NR_E	NAR_E
11-month-ahead	NR_E	NR_E	NR_E	NR_E	NR_E	NAR_E	NR_E	NAR_E
12-month-ahead	NR_E	NR_E	NR_E	NR_E	NR_E	NAR_E	NR_E	NAR_E

Note: NR_E , AR_E , and NAR_E indicate NS-RW-Equal, AF-RW-Equal, and NS-AF-RW-Equal, respectively.

6.2 RMSE Comparison

We move to the point forecasting evaluation of the pooling method. For each pair of maturity and forecast horizon, we calculate $RMSE(\tau, h)$, and choose the best prediction models producing the smallest RMSE value among the competing models. Table 7 presents the best models for each maturity across forecast horizons. Tables 8 and 9 reports that the RMSEs for all the maturities at 1, 3, 6, and 12-month-ahead horizons to show the relative performance of the alternative models. The RMSEs are normalized by that of the RW model for each maturity. Bold entries indicate the best prediction models for each pair of maturity and forecast horizon.

Two important findings emerge from the tables. First, in terms of the RMSE comparison the RW model outperforms the other models in 73 cases out of 96 cases (12 forecast horizons times 8 maturities) in contrast to the results from the PPCs. Particularly, the RW produces the most accurate point forecasts of 3, 24, 36, 60, and 84 month bond yields at all forecast horizon.

Second, the pooled model, NS-RW-MS forecasts best for six-month bond yield at long forecast horizons. In addition, the NS-AF-Const is superior in short-term forecasting of

10-year bond yield. As shown in Tables 8 and 9, however, the differences between the RMSEs of the two pooled models and the RW model are not substantial. Therefore, in terms of point forecasts the predictive gain from pooling various yield curve models does not seem to be remarkable.

6.3 Further Discussion

6.3.1 Robustness to the Out-of-Sample Selection

For a robustness check, we evaluate the pooling method based on the PPC and RMSE values over the last 54 forecast periods, starting in September 2008 to August 2009. Table 10 reports the best models which achieve the smallest PPC and RMSE values for each pair of maturity and forecasting horizon. The results are quite consistent with those from the 108 forecast periods.

6.3.2 Model Weights and Posterior Probability of Regimes

Finally, we discuss the estimates of the model weights using the full sample of February 1994 to December 2013. These estimates can be a measure of the relative importance of each individual model in the pooled models. Table 11 presents the estimated weights for the constant weight model. It is found that the DNS model has a higher weight in general, and that the RW model has a lower weight during the full sample period.

However, the models weights do not seem to be constant over time. Table 12 presents the estimated Markov-switching weights, and Figure 1 plots the posterior probabilities of the regime 2. First of all, it can be seen that the model weights are strongly regime-dependent, and the regime changes are drastic over time. As a result, the regimes seem to be well-identified.

More importantly, the role of the RW model in the pooled models became more important after the recent crisis. For instance, in the case of the NS-AF-RW-*MS*, the weights on the DNS and RW models in regime 1 are 0.219 and 0.429, respectively. Meanwhile, the weights are 0.752 and 0.133 in regime 2. As the figure shows, the posterior probability of regime 2 has been persistent, but decreased drastically around 2008. Hence, the predictive ability of the DNS model was superior before the recent

financial crisis, but the RW model was dominant after the crisis. The bigger role of the RW model during the crisis can be attributed to the fact that the RW enables to capture a structural change in the U.S. yield curve dynamics. Table 12 and Figure 1 confirm that this is consistent for the NS-RW-*MS* and AF-RW-*MS* models.

7 Conclusion

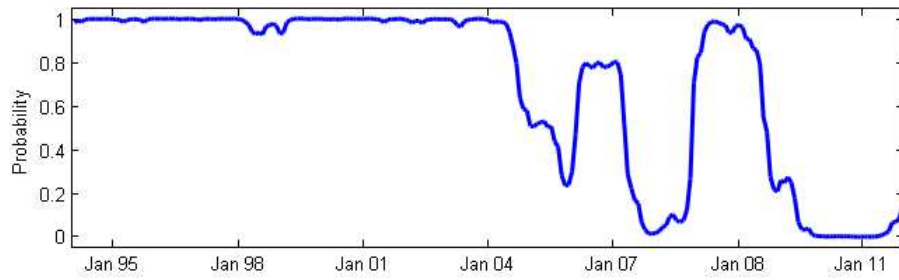
In this paper we examine whether the pooling method can help improve the accuracy of yield forecasts. The pool of models includes the dynamic Nelson-Siegel model, the arbitrage-free Nelson-Siegel model, and the random walk model. We combine these potentially misspecified models in a linear form suggested by Geweke and Amisano (2012), Geweke and Amisano (2011) and Waggoner and Zha (2012). We evaluate the out-of-sample forecasts of the pooled models based on the PPC and RMSE. The empirical results of the PPC comparison demonstrate that the pooled model of the DNS and RW with equal weights produces accurate density forecasts dominating the three individual models at all maturities and forecast horizons. Meanwhile, the predictive gain from the pooling method does not seem to be substantial in the point forecasts.

Our empirical finding implies that despite the higher computational cost, the pooling method is worth trying when multiple prediction models compete. However, we do not argue that our pooled model is the best model because possibly a different pooled model of different individual models can show a better forecasting performance. Maybe, one may include the DNS and AFNS models with macroeconomic factors or regime-switching yield curve models in the pool. In addition, it would be interesting to examine various approaches to averaging bond yield forecasts and compare their performance like in Clark and McCracken (2010). We leave them as future work.

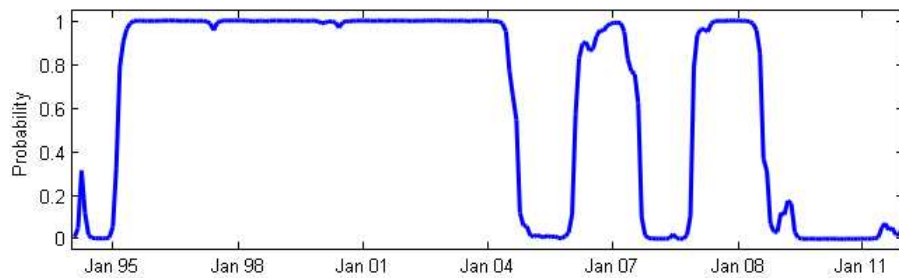
References

Almeida, C. and Vicente, 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.

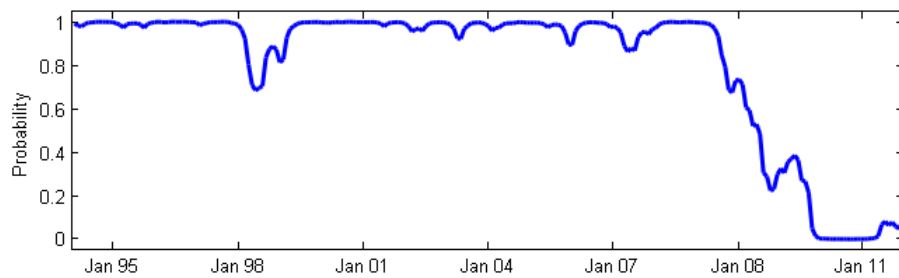
Figure 1: Posterior probabilities of regime 2



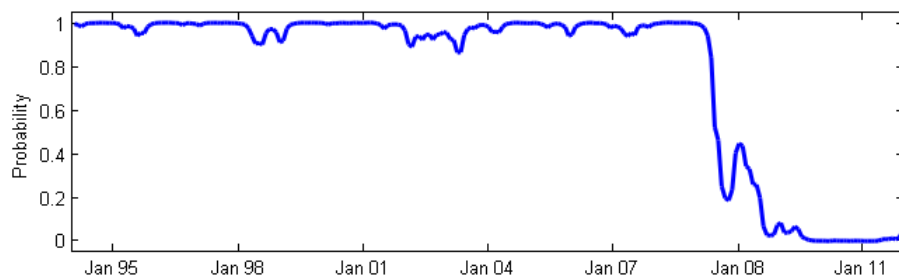
(a) NS-AF-RW-MS



(b) NS-AF-MS



(c) NS-RW-MS



(d) AF-RW-MS

Note: The solid line is the posterior probability of regime 2.

- Altavilla, C., Giacomini, R., and Ragusa, G. (2014), “Anchoring the Yield Curve using Survey Expectations?” *European Central Bank Working Paper 1632*, 1–28.
- Bansal, R. and Zhou, H. (2002), “Term structure of interest rates with regime shifts,” *Journal of Finance*, 57(5), 463–473.
- Carriero, 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.
- Carter, C. and Kohn, R. (1994), “On Gibbs sampling for state space models,” *Biometrika*, 81, 541–553.
- 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.
- (2016), “An Efficient Posterior Sampling in Gaussian Affine Term Structure Models,” *Manuscript*, 1–28.
- Chib, S. and Ramamurthy, S. (2010), “Tailored Randomized-block MCMC Methods with Application to DSGE Models,” *Journal of Econometrics*, 155, 19–38.
- 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.
- Clark, T. E. and McCracken, M. W. (2010), “Averaging forecasts from VARs with uncertain instabilities,” *Journal of Applied Econometrics*, 25(1), 5–29.
- Dai, Q., Singleton, K. J., and Yang, W. (2007), “Regime shifts in a dynamic term structure model of U.S. treasury bond yields,” *Review of Financial Studies*, 20, 1669–1706.
- 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.
- Niu, L. and Zeng, G. (2012), “The Discrete-Time Framework of Arbitrage-Free Nelson-Siegel Class of Term Structure Models,” *manuscript*, 1–68.
- 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.

A Likelihood

This section presents the step by step procedure for the log likelihood calculation. Suppose that $\boldsymbol{\psi}$ is the set of all model parameters and the log likelihood $\log L$ is initialized at 0. At time 1, $p(s_{t-1}|Y_{t-1}, \boldsymbol{\psi})$ is given at the unconditional probability of regime s_t . For $t = 1, 2, \dots, T$, the following steps are sequentially repeated.

Algorithm 2: Log likelihood calculation

- Step 1: The predictive probability of regime s_t , $p(s_t = j|Y_{t-1}, \boldsymbol{\psi})$ is computed as

$$\begin{aligned} p(s_t = j|Y_{t-1}, \boldsymbol{\psi}) &= \sum_{i=1}^2 \Pr[s_t = j|s_{t-1} = i, \boldsymbol{\psi}] \times \Pr(s_{t-1} = i|Y_{t-1}, \boldsymbol{\psi}) \\ &= \sum_{i=1}^2 q_{ij} \times p(s_{t-1} = i|Y_{t-1}, \boldsymbol{\psi}), \quad j = 1, 2 \end{aligned}$$

- Step 2: The predictive model weight on \mathcal{M}_1 , w_{1,s_t} is given by

$$W_{1,t} = \sum_{s_t=1}^2 w_{1,s_t} \times p(s_t|Y_{t-1}^o, \boldsymbol{\psi})$$

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 density $p(\mathbf{y}_t|Y_{t-1}, \boldsymbol{\psi})$ as

$$W_{1,t} \times p(\mathbf{y}_t|Y_{t-1}, \Theta_1, \mathcal{M}_1) + W_{2,t} \times p(\mathbf{y}_t|Y_{t-1}, \Theta_2, \mathcal{M}_2),$$

and $\log L = \log L + \log p(\mathbf{y}_t|Y_{t-1}, \boldsymbol{\psi})$

- Step 4: The updated probability of regime s_t , $p(s_t = i|Y_t, \boldsymbol{\psi})$ is calculated and retained as

$$\begin{aligned} p(s_t = i|Y_t, \boldsymbol{\psi}) &= \frac{p(s_t = i|Y_{t-1}, \boldsymbol{\psi}, \mathbf{y}_t)}{p(\mathbf{y}_t|Y_{t-1}, \boldsymbol{\psi})} \\ &= \frac{p(\mathbf{y}_t|Y_{t-1}, \boldsymbol{\psi}, s_t = i)p(s_t = i|Y_{t-1}, \boldsymbol{\psi})}{p(\mathbf{y}_t|Y_{t-1}, \boldsymbol{\psi})} \text{ for } i = 1, 2 \end{aligned}$$

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

$$\begin{aligned} &p(y_t|Y_{t-1}, \boldsymbol{\psi}, s_t) \\ &= w_{1,s_t} \times p(\mathbf{y}_t|Y_{t-1}, \Theta_1, \mathcal{M}_1) + (1 - w_{1,s_t}) \times p(\mathbf{y}_t|Y_{t-1}, \Theta_2, \mathcal{M}_2) \end{aligned}$$

- Step 5: $t = t + 1$ and go to Step 1 if $t \leq T$

B Details of MCMC Sampling

We now discuss the details of the each MCMC step. The burn-in is 1,000 and the MCMC simulation size beyond the burn-in is 10,000.

B.1 Parameters Sampling

First of all, $\Theta = (\Theta_{NS}, \Theta_{AF}, \Theta_{RW}, w)$ are simulated by the tailored randomized blocking Metropolis-Hastings algorithm (TaRB-MH, Chib and Ramamurthy (2010)). We note that the posterior density of Θ is proportional to the product of the likelihood and the prior as

$$\pi(\Theta|\mathbf{Y}, Q) \propto f(\mathbf{Y}|\boldsymbol{\psi}) \times \pi(\Theta). \quad (\text{B.1})$$

We simulate $\Theta|\mathbf{Y}, Q$ rather than $\Theta|\mathbf{Y}, \mathbf{S}, \mathbf{X}, Q$ by integrating out the regimes and latent factors because the former is more efficient than the latter. The likelihood computation is illustrated in 'A.

In every MCMC iteration, we apply the TaRB-MH method and sample Θ given (\mathbf{Y}, Q) . This algorithm is particularly useful when the posterior density is high-dimensional and its surface is possibly irregular. For the technical details, refer to Chib and Ramamurthy (2010) or Chib and Kang (2013).

Next, since the transition probability Q is independent of (\mathbf{Y}, Θ) given the regimes \mathbf{S} , it is sampled from

$$Q|\mathbf{S}.$$

Further, its prior is conjugate, and the transition probabilities are sampled from a beta distribution.

B.2 Regime Sampling

The time-series of the regimes \mathbf{S} is simulated in one block by the multi-move method (Chib (1998)). This method consists of two stages. The first stage is to calculate the filtered probabilities, $\Pr(s_t|Y_t, \boldsymbol{\psi})$ as

$$\begin{aligned} \Pr(s_t = j|Y_t, \boldsymbol{\psi}) &= \frac{\sum_{i=1}^2 p(\mathbf{y}_t|Y_{t-1}, s_t = j, \boldsymbol{\psi}) \times \Pr(s_t = j|s_{t-1} = i, \boldsymbol{\psi})}{\sum_{j=1}^2 \left[\sum_{i=1}^2 p(\mathbf{y}_t|Y_{t-1}, s_t = j, \boldsymbol{\psi}) \times \Pr(s_t = j|s_{t-1} = i, \boldsymbol{\psi}) \right]} \\ &= \frac{\sum_{i=1}^2 p(\mathbf{y}_t|Y_{t-1}, s_t = j, \boldsymbol{\psi}) \times q_{ij}}{\sum_{j=1}^2 \left[\sum_{i=1}^2 p(\mathbf{y}_t|Y_{t-1}, s_t = j, \boldsymbol{\psi}) \times q_{ij} \right]} \end{aligned}$$

for $j = 1, 2$. Then, the conditional density of \mathbf{y}_t $p(\mathbf{y}_t|Y_{t-1}, s_t, \boldsymbol{\psi})$ is computed as the linear combination of the model-specific conditional densities of \mathbf{y}_t : $p(\mathbf{y}_t|Y_{t-1}, \Theta_{NS}, \mathcal{M}_{NS})$, $p(\mathbf{y}_t|Y_{t-1}, \Theta_{AF}, \mathcal{M}_{AF})$, and $p(\mathbf{y}_t|Y_{t-1}, \Theta_{RW}, \mathcal{M}_{RW})$. That is,

$$\begin{aligned} & p(\mathbf{y}_t|Y_{t-1}, s_t, \boldsymbol{\psi}) \\ = & w_{NS,s_t} \times p(\mathbf{y}_t|Y_{t-1}, \Theta_{NS}, \mathcal{M}_{NS}) + w_{AF,s_t} \times p(\mathbf{y}_t|Y_{t-1}, \Theta_{AF}, \mathcal{M}_{AF}) \\ & + (1 - w_{NS,s_t} - w_{AF,s_t}) \times p(\mathbf{y}_t|Y_{t-1}, \Theta_{RW}, \mathcal{M}_{RW}) \end{aligned}$$

Those model-specific conditional densities are already given in equations (3.5), (3.14), and (3.16), respectively.

In the second stage, $\{s_t\}_{t=1}^T$ is sampled through a backward recursion. The regime at time T , s_T is first drawn with the filtered probability $\Pr(s_T|Y_T, \boldsymbol{\psi})$. Then, conditioned on s_{t+1} one can compute $\Pr(s_t|Y_t, s_{t+1}, \boldsymbol{\psi})$ using the filtered probabilities as the following:

$$\begin{aligned} & \Pr(s_t = i|Y_t, s_{t+1}, \boldsymbol{\psi}) \tag{B.2} \\ & = \frac{q_{ij} \times \Pr[s_t = i|Y_t, \boldsymbol{\psi}]}{\sum_{i=1}^2 q_{ij} \times \Pr[q_t = i|Y_t, \boldsymbol{\psi}]}, \quad i = 1, 2 \end{aligned}$$

Now given s_{t+1} , s_t is sampled with the probability $\Pr(s_t|Y_t, s_{t+1}, \boldsymbol{\psi})$ for $t = T - 1, T - 2, \dots, 1$, which completes the regime sampling.

B.3 Factor Sampling

The latent factors \mathbf{X} in the DNS model are sampled independently of $(\Theta_{AF}, \Theta_{RW}, w, Q)$. Given (Θ_{NS}, Y) , \mathbf{X} is typically simulated by the Carter and Kohn approach (Carter and Kohn (1994)). Meanwhile, one can exactly compute the factors $\mathbf{F} = \{\mathbf{f}_t\}_{t=1}^T$ in the AFNS model as

$$\mathbf{f}_t = (\mathbf{b}_B)^{-1} \times (\mathbf{y}_t^B - \mathbf{a}_B)$$

by using the basis yields and the model parameters.

B.4 Predictive Yield Curve Sampling

Each MCMC cycle is completed by sampling the posterior predictive draws given $(\mathbf{Y}, \mathbf{X}, \mathbf{F}, \mathbf{S}, \boldsymbol{\psi})$. For each posterior draw $(\mathbf{y}_T, \mathbf{x}_T, \mathbf{f}_T, s_T, \boldsymbol{\psi})$ and forecast horizon of $h = 1, 2, \dots, H$, we first simulate the predictive draws of the factors $\{\mathbf{x}_{T+h}, \mathbf{f}_{T+h}\}_{h=1}^H$. Given the factors and parameters, the model-specific predictive bond yields are generated within the individual model specifications. Next, the predictive regime $\{s_{T+h}\}_{h=1}^H$ is sampled by the Markov-switching process, and it determines the model weights for each forecast horizon. Finally, the predictive yield curve is computed as the linear combination of

the model-specific predictive yield curves, and retained as the predictive draws. The following algorithm summarizes the predictive yield curve simulation.

Algorithm 3: Posterior predictive distribution simulation

- Step 1: Sample the factors $\{\mathbf{x}_{T+h}\}_{h=1}^H | \mathbf{X}, \Theta_{NS}$ and $\{\mathbf{f}_{T+h}\}_{h=1}^H | \mathbf{F}, \Theta_{AF}$
- Step 2: Sample the model-specific predictive yields

$$\{\mathbf{y}_{NS,T+h}\}_{h=1}^H | \mathbf{X}, \Theta_{NS}, \quad \{\mathbf{y}_{AF,T+h}\}_{h=1}^H | \mathbf{F}, \Theta_{AF}, \quad \text{and} \quad \{\mathbf{y}_{RW,T+h}\}_{h=1}^H | \mathbf{Y}, \Theta_{RW}$$

- Step 3: Sample the predictive regimes, $\{s_{T+h}\}_{h=1}^H | \mathbf{S}, Q$
- Step 4: For $h = 1, 2, \dots, H$, compute the predictive yield curve as

$$\begin{aligned} \mathbf{y}_{T+h} &= w_{NS,s_{T+h}} \times \mathbf{y}_{NS,T+h} + w_{AF,s_{T+h}} \times \mathbf{y}_{AF,T+h} \\ &\quad + (\mathbf{1} - w_{NS,s_{T+h}} - w_{AF,s_{T+h}}) \times \mathbf{y}_{RW,T+h} \end{aligned}$$

- Step 5: Retain $\{\mathbf{y}_{T+h}\}_{h=1}^H$ as a posterior predictive draw

Table 5: 1- and 3-month ahead density forecasts evaluations: PPC

	3m	6m	12m	24m	36m	60m	84m	120m
DNS	1.112	1.209	1.124	1.102	1.320	1.408	1.267	1.032
AFNS	4.596	4.583	3.627	2.200	1.777	1.469	1.216	0.795
RW	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
NS-AF - <i>Equal</i>	1.790	1.671	1.300	1.043	1.158	1.035	1.140	0.767
NS-AF - <i>Const</i>	1.116	1.102	0.956	0.937	1.081	1.027	1.051	0.846
NS-AF - <i>MS</i>	1.941	1.980	1.655	1.310	1.272	1.116	1.146	0.931
NS-RW - <i>Equal</i>	0.749	0.708	0.653	0.677	0.748	0.779	0.752	0.676
NS-RW - <i>Const</i>	0.957	0.983	0.931	0.952	1.088	1.128	1.042	0.882
NS-RW - <i>MS</i>	0.887	0.907	0.850	0.870	0.978	0.999	0.945	0.833
AF-RW - <i>Equal</i>	1.655	1.615	1.319	0.953	0.853	0.736	0.750	0.620
AF-RW - <i>Const</i>	3.095	3.047	2.438	1.586	1.345	1.118	1.048	0.749
AF-RW - <i>MS</i>	2.719	2.578	1.990	1.310	1.089	0.934	0.901	0.670
NS-AF-RW - <i>Equal</i>	1.413	1.329	1.116	0.982	1.018	0.904	1.009	0.791
NS-AF-RW - <i>Const</i>	1.141	1.109	1.023	1.049	1.197	1.168	1.193	0.940
NS-AF-RW - <i>MS</i>	1.676	1.730	1.570	1.372	1.402	1.296	1.401	1.163

(a) 1-month-ahead

	3m	6m	12m	24m	36m	60m	84m	120m
DNS	1.113	1.047	1.024	1.074	1.152	1.175	1.106	1.000
AFNS	3.964	3.657	3.004	2.043	1.657	1.170	1.138	0.839
RW	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
NS-AF - <i>Equal</i>	1.680	1.501	1.259	1.077	1.082	0.966	1.023	0.818
NS-AF - <i>Const</i>	1.099	1.005	0.937	0.949	0.990	0.915	0.929	0.824
NS-AF - <i>MS</i>	1.685	1.558	1.354	1.166	1.098	0.924	0.947	0.837
NS-RW - <i>Equal</i>	0.806	0.769	0.743	0.733	0.747	0.741	0.717	0.661
NS-RW - <i>Const</i>	0.988	0.950	0.937	0.974	1.022	1.029	0.978	0.887
NS-RW - <i>MS</i>	0.910	0.882	0.864	0.887	0.924	0.927	0.891	0.826
AF-RW - <i>Equal</i>	1.558	1.474	1.282	1.002	0.897	0.724	0.767	0.663
AF-RW - <i>Const</i>	2.753	2.567	2.163	1.577	1.346	1.006	1.052	0.832
AF-RW - <i>MS</i>	2.453	2.229	1.807	1.320	1.114	0.859	0.922	0.760
NS-AF-RW - <i>Equal</i>	1.211	1.126	1.007	0.882	0.853	0.733	0.802	0.688
NS-AF-RW - <i>Const</i>	0.992	0.929	0.894	0.909	0.943	0.894	0.901	0.790
NS-AF-RW - <i>MS</i>	1.210	1.158	1.074	1.000	0.991	0.894	0.939	0.833

(b) 3-month-ahead

Note: Bold entries indicate the best density forecasting performance for each maturity and forecast horizon.

Table 6: 6- and 12-month ahead density forecasts evaluations: PPC

	3m	6m	12m	24m	36m	60m	84m	120m
DNS	1.103	1.045	1.027	1.047	1.096	1.077	1.023	0.937
AFNS	3.261	3.072	2.666	1.969	1.646	1.107	1.123	0.855
RW	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
NS-AF - <i>Equal</i>	1.473	1.378	1.233	1.079	1.061	0.920	0.969	0.805
NS-AF - <i>Const</i>	1.085	1.028	0.981	0.957	0.968	0.871	0.883	0.799
NS-AF - <i>MS</i>	1.469	1.399	1.283	1.134	1.064	0.874	0.895	0.798
NS-RW - <i>Equal</i>	0.847	0.834	0.805	0.747	0.739	0.712	0.693	0.644
NS-RW - <i>Const</i>	0.998	0.975	0.968	0.966	0.988	0.963	0.922	0.839
NS-RW - <i>MS</i>	0.932	0.909	0.893	0.881	0.900	0.882	0.856	0.795
AF-RW - <i>Equal</i>	1.397	1.359	1.237	0.998	0.894	0.703	0.746	0.649
AF-RW - <i>Const</i>	2.337	2.230	1.984	1.543	1.340	0.956	1.023	0.826
AF-RW - <i>MS</i>	2.082	1.945	1.672	1.296	1.117	0.825	0.904	0.758
NS-AF-RW - <i>Equal</i>	1.068	1.039	0.969	0.848	0.800	0.660	0.706	0.619
NS-AF-RW - <i>Const</i>	0.939	0.914	0.900	0.884	0.890	0.809	0.811	0.729
NS-AF-RW - <i>MS</i>	1.050	1.028	0.987	0.926	0.907	0.799	0.822	0.738

(a) 6-month-ahead

	3m	6m	12m	24m	36m	60m	84m	120m
DNS	1.089	1.045	1.037	1.064	1.100	1.065	0.995	0.929
AFNS	2.468	2.394	2.229	1.873	1.660	1.161	1.152	0.894
RW	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
NS-AF - <i>Equal</i>	1.203	1.169	1.120	1.058	1.045	0.907	0.915	0.788
NS-AF - <i>Const</i>	1.046	1.011	0.994	0.988	0.991	0.890	0.867	0.804
NS-AF - <i>MS</i>	1.225	1.196	1.161	1.107	1.070	0.902	0.894	0.822
NS-RW - <i>Equal</i>	0.884	0.870	0.844	0.769	0.733	0.663	0.618	0.572
NS-RW - <i>Const</i>	0.989	0.970	0.968	0.965	0.973	0.925	0.861	0.792
NS-RW - <i>MS</i>	0.940	0.918	0.913	0.900	0.903	0.859	0.808	0.751
AF-RW - <i>Equal</i>	1.195	1.181	1.125	0.966	0.869	0.659	0.664	0.568
AF-RW - <i>Const</i>	1.786	1.748	1.654	1.428	1.288	0.933	0.948	0.764
AF-RW - <i>MS</i>	1.635	1.574	1.448	1.231	1.090	0.788	0.828	0.690
NS-AF-RW - <i>Equal</i>	0.941	0.942	0.926	0.854	0.797	0.635	0.636	0.556
NS-AF-RW - <i>Const</i>	0.932	0.917	0.921	0.914	0.909	0.813	0.782	0.710
NS-AF-RW - <i>MS</i>	0.966	0.958	0.958	0.931	0.907	0.788	0.773	0.701

(b) 12-month-ahead

Note: Bold entries indicate the best density forecasting performance for each maturity and forecast horizon.

Table 7: Best Models: RMSE

	3m	6m	12m	24m	36m	60m	84m	120m
1-month-ahead	<i>R</i>	<i>NR_E</i>	<i>NR_M</i>	<i>R</i>	<i>R</i>	<i>R</i>	<i>R</i>	<i>NR_M</i>
2-month-ahead	<i>R</i>	<i>N</i>	<i>N</i>	<i>R</i>	<i>R</i>	<i>R</i>	<i>R</i>	<i>NA_M</i>
3-month-ahead	<i>R</i>	<i>N</i>	<i>N</i>	<i>R</i>	<i>R</i>	<i>R</i>	<i>R</i>	<i>NA_C</i>
4-month-ahead	<i>R</i>	<i>N</i>	<i>N</i>	<i>R</i>	<i>R</i>	<i>R</i>	<i>R</i>	<i>NA_C</i>
5-month-ahead	<i>R</i>	<i>N</i>	<i>N</i>	<i>R</i>	<i>R</i>	<i>R</i>	<i>R</i>	<i>NA_C</i>
6-month-ahead	<i>R</i>	<i>N</i>	<i>N</i>	<i>R</i>	<i>R</i>	<i>R</i>	<i>R</i>	<i>NA_M</i>
7-month-ahead	<i>R</i>	<i>R</i>	<i>N</i>	<i>R</i>	<i>R</i>	<i>R</i>	<i>R</i>	<i>R</i>
8-month-ahead	<i>R</i>	<i>R</i>	<i>R</i>	<i>R</i>	<i>R</i>	<i>R</i>	<i>R</i>	<i>R</i>
9-month-ahead	<i>R</i>	<i>NR_M</i>	<i>R</i>	<i>R</i>	<i>R</i>	<i>R</i>	<i>R</i>	<i>R</i>
10-month-ahead	<i>R</i>	<i>NR_M</i>	<i>R</i>	<i>R</i>	<i>R</i>	<i>R</i>	<i>R</i>	<i>R</i>
11-month-ahead	<i>R</i>	<i>NR_M</i>	<i>R</i>	<i>R</i>	<i>R</i>	<i>R</i>	<i>R</i>	<i>R</i>
12-month-ahead	<i>R</i>	<i>NR_M</i>	<i>R</i>	<i>R</i>	<i>R</i>	<i>R</i>	<i>R</i>	<i>R</i>

Note: *N*, *R*, *NR_E*, *NR_M*, *NA_C*, and *NA_M* indicate DNS, RW, NS-RW-*Equal*, NS-RW-*MS*, NS-AF-*Const*, and NS-AF-*MS*, respectively.

Table 8: 1- and 3-month-ahead point forecasts evaluations: RMSE

	3m	6m	12m	24m	36m	60m	84m	120m
DNS	1.036	1.007	0.905	1.043	1.232	1.326	1.206	1.006
AFNS	1.074	1.079	0.983	1.112	1.227	1.101	1.272	1.047
RW	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
NS-AF - <i>Equal</i>	1.178	1.124	1.010	1.155	1.342	1.294	1.385	1.079
NS-AF - <i>Const</i>	1.090	1.040	0.925	1.058	1.207	1.163	1.196	1.005
NS-AF - <i>MS</i>	1.257	1.273	1.173	1.230	1.279	1.145	1.253	1.079
NS-RW - <i>Equal</i>	1.027	0.980	0.927	1.035	1.098	1.133	1.086	0.998
NS-RW - <i>Const</i>	1.039	1.001	0.913	1.052	1.177	1.234	1.147	1.001
NS-RW - <i>MS</i>	1.021	0.990	0.904	1.028	1.126	1.162	1.101	0.998
AF-RW - <i>Equal</i>	1.083	1.081	1.029	1.098	1.125	1.036	1.145	1.040
AF-RW - <i>Const</i>	1.154	1.158	1.079	1.174	1.244	1.112	1.279	1.093
AF-RW - <i>MS</i>	1.138	1.106	1.008	1.142	1.173	1.075	1.222	1.063
NS-AF-RW - <i>Equal</i>	1.274	1.236	1.166	1.271	1.338	1.252	1.354	1.163
NS-AF-RW - <i>Const</i>	1.275	1.239	1.164	1.276	1.388	1.368	1.374	1.157
NS-AF-RW - <i>MS</i>	1.450	1.468	1.400	1.435	1.481	1.416	1.474	1.278

(a) 1-month-ahead

	3m	6m	12m	24m	36m	60m	84m	120m
DNS	1.045	0.974	0.942	1.057	1.131	1.186	1.128	1.017
AFNS	1.150	1.125	1.092	1.180	1.223	1.059	1.251	1.108
RW	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
NS-AF - <i>Equal</i>	1.218	1.144	1.085	1.176	1.252	1.235	1.299	1.136
NS-AF - <i>Const</i>	1.103	1.023	0.976	1.064	1.114	1.079	1.100	0.993
NS-AF - <i>MS</i>	1.203	1.136	1.073	1.132	1.143	1.048	1.104	1.008
NS-RW - <i>Equal</i>	1.036	0.998	0.986	1.050	1.070	1.092	1.063	0.998
NS-RW - <i>Const</i>	1.054	1.007	0.987	1.085	1.134	1.174	1.129	1.035
NS-RW - <i>MS</i>	1.022	0.985	0.962	1.045	1.084	1.115	1.081	1.012
AF-RW - <i>Equal</i>	1.106	1.093	1.079	1.130	1.140	1.067	1.160	1.087
AF-RW - <i>Const</i>	1.215	1.194	1.174	1.245	1.268	1.141	1.301	1.181
AF-RW - <i>MS</i>	1.168	1.128	1.085	1.170	1.180	1.090	1.235	1.141
NS-AF-RW - <i>Equal</i>	1.155	1.109	1.082	1.142	1.161	1.108	1.185	1.088
NS-AF-RW - <i>Const</i>	1.139	1.077	1.046	1.123	1.159	1.146	1.153	1.043
NS-AF-RW - <i>MS</i>	1.183	1.135	1.100	1.159	1.179	1.142	1.179	1.078

(b) 3-month-ahead

Note: Bold entries indicate the best point forecasting performance for each maturity and forecast horizon.

Table 9: 6- and 12-month ahead point forecasts evaluations: RMSE

	3m	6m	12m	24m	36m	60m	84m	120m
DNS	1.039	0.997	0.986	1.062	1.127	1.153	1.104	1.009
AFNS	1.125	1.127	1.131	1.211	1.265	1.084	1.277	1.159
RW	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
NS-AF - <i>Equal</i>	1.169	1.139	1.123	1.193	1.261	1.225	1.286	1.158
NS-AF - <i>Const</i>	1.091	1.052	1.031	1.085	1.123	1.074	1.091	1.001
NS-AF - <i>MS</i>	1.154	1.120	1.095	1.133	1.144	1.037	1.078	0.988
NS-RW - <i>Equal</i>	1.029	1.017	1.016	1.055	1.081	1.088	1.069	1.014
NS-RW - <i>Const</i>	1.044	1.023	1.025	1.097	1.144	1.164	1.127	1.044
NS-RW - <i>MS</i>	1.018	0.998	0.995	1.054	1.095	1.114	1.089	1.025
AF-RW - <i>Equal</i>	1.076	1.081	1.088	1.133	1.156	1.074	1.162	1.098
AF-RW - <i>Const</i>	1.160	1.163	1.177	1.255	1.297	1.154	1.310	1.210
AF-RW - <i>MS</i>	1.128	1.115	1.105	1.178	1.206	1.099	1.244	1.164
NS-AF-RW - <i>Equal</i>	1.079	1.068	1.066	1.110	1.130	1.056	1.119	1.044
NS-AF-RW - <i>Const</i>	1.075	1.050	1.045	1.099	1.131	1.092	1.097	1.013
NS-AF-RW - <i>MS</i>	1.080	1.062	1.060	1.113	1.140	1.088	1.113	1.032

(a) 6-month-ahead

	3m	6m	12m	24m	36m	60m	84m	120m
DNS	1.031	1.008	1.017	1.107	1.190	1.249	1.200	1.127
AFNS	1.070	1.084	1.119	1.244	1.347	1.242	1.426	1.334
RW	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
NS-AF - <i>Equal</i>	1.081	1.074	1.090	1.189	1.285	1.305	1.358	1.272
NS-AF - <i>Const</i>	1.065	1.044	1.049	1.120	1.181	1.169	1.167	1.102
NS-AF - <i>MS</i>	1.084	1.071	1.079	1.148	1.198	1.144	1.166	1.100
NS-RW - <i>Equal</i>	1.019	1.011	1.022	1.068	1.109	1.129	1.099	1.054
NS-RW - <i>Const</i>	1.024	1.013	1.030	1.112	1.183	1.232	1.188	1.122
NS-RW - <i>MS</i>	1.004	0.992	1.006	1.078	1.142	1.185	1.149	1.091
AF-RW - <i>Equal</i>	1.026	1.032	1.054	1.120	1.170	1.117	1.190	1.134
AF-RW - <i>Const</i>	1.066	1.078	1.116	1.230	1.318	1.241	1.378	1.298
AF-RW - <i>MS</i>	1.059	1.059	1.073	1.162	1.221	1.145	1.287	1.226
NS-AF-RW - <i>Equal</i>	1.018	1.022	1.045	1.114	1.158	1.108	1.146	1.082
NS-AF-RW - <i>Const</i>	1.042	1.031	1.048	1.123	1.180	1.175	1.161	1.090
NS-AF-RW - <i>MS</i>	1.030	1.024	1.047	1.123	1.175	1.155	1.162	1.095

(b) 12-month-ahead

Note: Bold entries indicate the best point forecasting performance for each maturity and forecast horizon.

Table 10: Best Models: Robustness

	3m	6m	12m	24m	36m	60m	84m	120m
1-month-ahead	NR_E	NR_E	NR_E	NR_E	NR_E	AR_E	NR_E	AR_E
2-month-ahead	NR_E	NR_E	NR_E	NR_E	NR_E	AR_E	NR_E	NR_E
3-month-ahead	NR_E	NR_E	NR_E	NR_E	NR_E	AR_E	NR_E	NR_E
4-month-ahead	NR_E	NR_E	NR_E	NR_E	NR_E	NR_E	NR_E	NR_E
5-month-ahead	NR_E	NR_E	NR_E	NR_E	NR_E	NAR_E	NR_E	NR_E
6-month-ahead	NR_E	NR_E	NR_E	NR_E	NR_E	NAR_E	NR_E	NAR_E
7-month-ahead	NR_E	NR_E	NR_E	NR_E	NR_E	NAR_E	NR_E	NAR_E
8-month-ahead	NR_E	NR_E	NR_E	NR_E	NR_E	NAR_E	NR_E	NAR_E
9-month-ahead	NR_E	NR_E	NR_E	NR_E	NR_E	NAR_E	NR_E	NAR_E
10-month-ahead	NR_E	NR_E	NR_E	NR_E	NR_E	NAR_E	NR_E	NAR_E
11-month-ahead	NR_E	NR_E	NR_E	NR_E	NR_E	NR_E	NR_E	NAR_E
12-month-ahead	NR_E	NR_E	NR_E	NR_E	NR_E	NR_E	NR_E	NAR_E

(a) PPC

	3m	6m	12m	24m	36m	60m	84m	120m
1-month-ahead	R	N	NR_M	R	R	R	R	NR_E
2-month-ahead	R	N	N	R	R	R	R	NA_M
3-month-ahead	R	N	N	R	R	R	R	NR_E
4-month-ahead	R	R	R	R	R	R	R	R
5-month-ahead	R	R	R	R	R	R	R	NA_M
6-month-ahead	R	R	R	R	R	R	R	NA_M
7-month-ahead	R	R	R	R	R	R	R	NA_M
8-month-ahead	R	R	R	R	R	R	R	R
9-month-ahead	R	R	R	R	R	R	R	R
10-month-ahead	R	R	R	R	R	R	R	R
11-month-ahead	R	R	R	R	R	R	R	R
12-month-ahead	R	R	R	R	R	R	R	R

(b) RMSE

Note: N , R , NR_M , NR_M , NA_C , NA_M , AR_E , and NAR_E indicate DNS, RW, NS-RW-Equal, NS-RW-MS, NS-AF-Const, NS-AF-MS, AF-RW-Equal, and NS-AF-RW-Equal, respectively. These models are chosen based on the recent 54 out-of-sample predictions.

Table 11: Constant model weight estimates

	DNS	AFNS	RW
NS-AF- <i>Const</i>	0.678 (0.034)	0.322 (0.034)	-
NS-RW- <i>Const</i>	0.749 (0.029)	-	0.251 (0.029)
AF-RW- <i>Const</i>	-	0.702 (0.033)	0.298 (0.033)
NS-AF-RW- <i>Const</i>	0.593 (0.034)	0.181 (0.027)	0.226 (0.028)

Note: Standard errors are in parentheses.

Table 12: Markov-switching model weight estimates

	Regime 1 ($s_t = 1$)			Regime 2 ($s_t = 2$)		
	DNS	AFNS	RW	DNS	AFNS	RW
NS-AF- <i>MS</i>	0.216 (0.048)	0.784 (0.048)	-	0.883 (0.015)	0.117 (0.015)	-
NS-RW- <i>MS</i>	0.339 (0.080)	-	0.661 (0.080)	0.858 (0.027)	-	0.142 (0.027)
AF-RW- <i>MS</i>	-	0.276 (0.068)	0.724 (0.068)	-	0.834 (0.030)	0.166 (0.030)
NS-AF-RW- <i>MS</i>	0.219 (0.051)	0.352 (0.060)	0.429 (0.062)	0.752 (0.027)	0.115 (0.014)	0.133 (0.023)

Note: Standard errors are in parentheses.