Term Structure Dynamics, Macro-Finance Factors and Model Uncertainty

Byrne, Joseph and Cao, Shuo and Korobilis, Dimitris

University of Glasgow

April 2015

Term Structure Dynamics, Macro-Finance Factors and Model Uncertainty*

Joseph P. Byrne†, Shuo Cao‡ and Dimitris Korobilis§

21 April 2015

Abstract

This paper extends the Nelson-Siegel linear factor model by developing a flexible macro-finance framework for modeling and forecasting the term structure of US interest rates. Our approach is robust to parameter uncertainty and structural change, as we consider instabilities in parameters and volatilities, and our model averaging method allows for investors' model uncertainty over time. Our time-varying parameter Nelson-Siegel Dynamic Model Averaging (NS-DMA) predicts yields better than standard benchmarks and successfully captures plausible time-varying term premia in real time. The proposed model has significant in-sample and out-of-sample predictability for excess bond returns, and the predictability is of economic value.

Keywords: Term Structure of Interest Rates; Nelson-Siegel; Dynamic Model Averaging; Bayesian Methods; Term Premia.

JEL Classification Codes: C32; C52; E43; E47; G17.

*An earlier version of this paper was circulated under the title 'Term Structure Dynamics with Many Macroeconomic Fundamentals'. The authors would like to thank Luc Bauwens, Andrea Carriero, Jens Christensen, Pasquale Della Corte, Magnus Dahlquist, Gregory Duffee, Domenico Giannone, Jan Grobovšek, Kris Jacobs, Gary Koop, Rajnish Mehre, Theo Nijman, Bart Taub, Herman van Dijk, Jonathan Wright and participants at the EC^2 Barcelona, Inquire Europe Seminar Stockholm and BMRC-DEMS Conference for helpful discussion and comments.

†School of Management and Languages, Heriot-Watt University, Edinburgh, UK. Email: j.p.byrne@hw.ac.uk

‡Adam Smith Business School, University of Glasgow, Glasgow, UK. Email: s.cao.1@research.gla.ac.uk

§Adam Smith Business School, University of Glasgow, Glasgow, UK. Email: dimitris.korobilis@glasgow.ac.uk

1 Introduction

Modeling the term structure of interest rates using risk factors is a vast and expanding research frontier in financial economics; see Piazzesi (2010), Gürkaynak and Wright (2012), Duffee (2013) and Diebold and Rudebusch (2013) for extensive reviews. A large number of papers have focused on modeling yield dynamics and sought to produce satisfactory forecasting results, such as Nelson and Siegel (1987), Dai and Singleton (2003), Diebold and Li (2006), Christensen, Diebold and Rudebusch (2011) and Dewachter and Iania (2012), among others. A major strand of this yield forecasting literature has been inspired by the seminal contribution of Nelson and Siegel (1987), who extract three linear factors that capture most of the variation in bond yield data. The Nelson and Siegel (1987) (NS) approach has an appealing structure that is parsimonious, flexible, and allows for an easy interpretation of the estimated factors. Diebold and Li (2006) extend the proposed Nelson-Siegel model to a dynamic version, and provide improved predictive power in modeling the yield curve. Joslin, Singleton and Zhu (2011) and Duffee (2013) conclude that, in the absence of restrictions in factor dynamics, forecasts from models which impose no-arbitrage restrictions are equivalent to forecasts from unrestricted, reduced-form econometric models.[1] This observation can generalize to reduced-form estimation with Nelson-Siegel restrictions, where principal component estimates are replaced with NS factors.[2]

In this paper we build upon previous work and propose a dynamic Nelson-Siegel model with several novel features. Firstly, we extend related work by accommodating structural change in our term structure model and incorporating additional financial information. The global financial crisis was an abrupt nonlinear shock that highlighted the importance of financial market for macroeconomic activity and bond yields more generally. Our macro-finance model combines standard Nelson-Siegel factors with macroeconomic and financial factors estimated by a large vector autoregressive (VAR) system with time-varying coefficients and volatility.

[1] See also Joslin, Le and Singleton (2013), who extend the irrelevance proposition of cross-sectional (no-arbitrage) restrictions of Joslin, Singleton and Zhu (2011) to higher order state dynamics.

[2] Nevertheless we test the robustness of core results to the no-arbitrage restrictions.

The time-varying setup is conducted by the application of Bayesian econometric techniques. Building on, and extending, Koop and Korobilis (2013) we develop an efficient Bayesian model that allows us to estimate large systems with many variables.

Secondly, following a large literature we include macro risk factors in our reduced-form specification.[3] The seminal work of Ang and Piazzesi (2003) uses inflation, the output gap and three latent factors to model yields.[4] Other authors consider the dynamics of the term structure augmented with information on exchange rates or survey data; see Anderson, Hammond and Ramezani (2010), Duffee (2014) and Kim and Orphanides (2012). Dewachter and Iania (2012) and Dewachter, Iania and Lyrio (2014) successfully model yield dynamics using standard macro factors plus three additional financial factors: liquidity risk, credit risk and risk premium factors. This innovative approach can be extended to incorporate a more substantial range of macro-finance risk factors with modeling techniques that seek to distill large datasets.

Lastly, following Koop and Korobilis (2012) we employ Dynamic Model Averaging (DMA) methods in order to determine in a data-based way which macro risks are relevant for the yield curve at different points in time. That is, we use DMA in order to choose, at each point in time, between three models: i) one with three Nelson-Siegel (NS) factors only; ii) NS factors plus three key macroeconomic indicators; and iii) NS factors augmented using up to 15 macro and financial factors.[5] DMA allows us to assign probabilities for each of the models at each point in time and thus dynamically implement averaging over time. Model averaging methods have been shown to reduce the total forecast risk associated with using only a single 'best' model; see Avramov (2002), Cremers (2002) and Elliott and Timmermann (2008).

We use our model to empirically examine U.S. term structure dynamics using monthly obser-

[3]See for instance Ang and Piazzesi (2003), Bernanke, Reinhart and Sack (2004), Diebold, Rudebusch and Aruoba (2006), Hördahl, Tristani and Vestin (2006), Bianchi, Mumtaz and Surico (2009), Smith and Taylor (2009), Bekaert, Cho and Moreno (2010) and Joslin, Priebsch and Singleton (2014).

[4]The important roles of macro variables, such as price inflation and indicators of real activity, are particularly emphasized in this paper: the authors show that macro factors can explain up to 85% of the variation in bond yields.

[5]Our third macro-finance model is like a 'kitchen sink' model which fully accounts for, and extends, the point of Dewachter and Iania (2012) and Dewachter, Iania and Lyrio (2014) that financial factors are important for modeling yields, whilst allowing for much more information to be incorporated in the spirit of Ludvigson and Ng (2009).

vations from 1971 to 2013. Our empirical evidence indicates an extended macro-finance model is helpful in modeling yield factor dynamics especially in recession periods. We shed light on the apparent trade-off between incorporating stochastic volatility and fitting the cross-section of yields in affine term structure models.[6] We find that our approach has useful empirical properties in yield forecasting, as it is robust to parameter and model uncertainty as well as potential structural breaks. We compare the forecast performance of our approach to a basic dynamic Nelson-Siegel model and several variations, and show that the gains in predictability is due to the ensemble of salient features – time-varying parameters, stochastic volatility and dynamic model averaging. Our extended macro-finance model forecasts better than the benchmarks, especially at short horizons based upon Diebold and Mariano (1995) tests and predictive likelihood. Using only conditional information, our approach to modeling the yield curve provides us with successful term premium alternatives to full-sample estimates of Kim and Wright (2005), Wright (2011) and Bauer, Rudebusch and Wu (2014), which reveals plausible expectations of investors in real time. Our estimated term premia shows a significant 'flight-to-quality' demand in the global financial crisis, which distinguishes this crisis from the previous recessions. A predictable element estimated from our proposed model has strong in-sample and out-of-sample predictability in terms of future excess bond returns using Clark and West (2007). Moreover, the predictability is of economic value, based upon the methodology of Campbell and Thompson (2008).

This paper is structured as follows. Section 2 describes the estimation method and our framework for modeling bond yield dynamics. Section 3 describes the data, discusses the economic implications of NS factor movements and displays the performance and second moment properties of NS-DMA yield forecasts. Robustness checks with arbitrage-free restrictions are as well present in Section 3. Section 4 evaluates the predictability of NS-DMA for the excess

[6] Anh and Joslin (2013) indicate no-arbitrage affine term structure models with stochastic volatility perform poorly in replicating term premia dynamics in the data, because the no-arbitrage assumption provides strong over-identifying constraints. Creal and Wu (2015) also suggest that in the no-arbitrage framework with constant parameters, the benefit in fitting volatility is at the expense of fitting the cross-section of yields. Our empirical results show that the potential evolution of model parameters needs to be taken into account, so less flexible state dynamics may not be correctly specified to capture the abnormal dynamics of yield factors in recession periods.

bond returns and the economic value of the predictability. Section 4 also shows the implied term premia of NS-DMA has informative economic implications. Section 5 concludes.

2 Methods

2.1 The Nelson-Siegel Restrictions

Following Nelson and Siegel (1987) and Diebold and Li (2006) we assume that three factors summarize most of the information in the term structure of interest rates. Let $y_t(\tau)$ denote yields at maturity τ, then the factor model we use is of the form:[7]

$$y_t(\tau) = L_t^{NS} + \frac{1 - e^{-\tau\lambda^{NS}}}{\tau\lambda^{NS}} S_t^{NS} + \left(\frac{1 - e^{-\tau\lambda^{NS}}}{\tau\lambda^{NS}} - e^{-\tau\lambda^{NS}} \right) C_t^{NS} + \varepsilon_t(\tau), \tag{2.1}$$

where L_t^{NS} is the "Level" factor, S_t^{NS} is the "Slope" factor, C_t^{NS} is the "Curvature" factor and $\varepsilon_t(\tau)$ is the error term. In the formulation above, λ^{NS} is a parameter that controls the shapes of loadings for the NS factors; following Diebold and Li (2006) and Bianchi, Mumtaz and Surico (2009), we set $\lambda^{NS} = 0.0609$. For estimation purposes, we can rewrite the equation (2.1) in the equivalent compact form,

$$y_t(\tau) = \mathbf{B}(\tau) F_t^{NS} + \varepsilon_t(\tau),$$

where $F_t^{NS} = \left[L_t^{NS}, S_t^{NS}, C_t^{NS} \right]'$ is the vector of three NS factors, $\mathbf{B}(\tau)$ is the loading vector and $\varepsilon_t(\tau)$ is the error term.

The Nelson-Siegel restrictions are in fact restrictions on the risk-neutral dynamics. Feunou, Fontaine and Le (2014) show that the NS model is the continuous time limit of their near arbitrage-free class with a unit root under the risk-neutral measure. Joslin, Singleton and Zhu (2011) show that no-arbitrage cross-sectional restrictions cannot improve out-of-sample forecasts. In light of their findings, we specify the cross-sectional loadings with NS restrictions

[7] This is an asymptotically flat approximating function, and Siegel and Nelson (1988) demonstrate that this property is appropriate if forward rates have finite limiting values.

and focus on time-series variation of yield factors, in order to improve the forecast performance.

The NS restrictions also imply a setup of Unspanned Macro Risk, if the time series (physical) dynamics of factors, without imposing any restrictions, are augmented with macro-finance information. In this setup, the macro variables only affect the unobserved NS factors and do not interact directly with the observed yields, so that they are unspanned by the yields. In other words, a 'knife-edge' restriction is imposed on the coefficients of macro variables in the pricing dynamics, while the physical dynamics are left unconstrained, see Joslin, Priebsch and Singleton (2014) for details.

2.2 Yield Factor Dynamics

In our first step, we use a simple ordinary least squares (OLS) to extract three NS factors. We assume these factors are observed without errors, which is a standard assumption in term structure modeling. The interpretation of the Nelson-Siegel factors are of considerable empirical importance. The Level factor L_t^{NS} is identified as the factor that is loaded evenly by the yields of all maturities. The Slope factor S_t^{NS} is equivalent to the spread between short- and a long-term bond yields, and its movements are captured by placing more weights on shorter maturities. The Curvature factor C_t^{NS} captures changes that have their largest impact on medium-term maturities, and therefore medium-term maturities load more heavily on this factor. In particular, using the setting $\lambda^{NS} = 0.0609$, the C_t^{NS} has the largest impact on the bond at 30-month maturity, see Diebold and Li (2006).[8]

An important novel aspect of our methodology is in modeling the factor dynamics in the second step. Following Bianchi, Mumtaz and Surico (2009), the extracted Nelson-Siegel factors augmented with macroeconomic variables follow a time-varying parameter vector autoregression

[8]Further discussion of these factors can be found in Appendix B.

(TVP-VAR) of order p of the form

$$\begin{bmatrix} F_t^{NS} \\ \\ M_t \end{bmatrix} = c_t + B_{1t} \begin{bmatrix} F_{t-1}^{NS} \\ \\ M_{t-1} \end{bmatrix} + \cdots + B_{pt} \begin{bmatrix} F_{t-p}^{NS} \\ \\ M_{t-p} \end{bmatrix} + v_t, \tag{2.2}$$

where c_t are time-varying intercepts, B_{it} are matrices of time-varying autoregressive coefficients for $i = 1, ..., p$, M_t is a vector of macro-finance risk factors, and v_t is the error term. Following Coroneo, Giannone and Modugno (2014) and Joslin, Priebsch and Singleton (2014), we do not impose any restrictions on the above VAR system.

For the purpose of econometric estimation, we work with a more compact form of Eq. (2.2). We can show that the p-lag TVP-VAR can be written as

$$z_t = X_t \beta_t + v_t, \tag{2.3}$$

where $z_t = \left[L_t^{NS}, S_t^{NS}, C_t^{NS}, M_t'\right]'$, M_t is an $q \times 1$ vector of macro-finance factors, $X_t = I_n \otimes \left[z_{t-1}', ..., z_{t-p}'\right]$ for $n = q + 3$, $\beta_t = \left[c_t, vec\left(B_{1t}\right)', \cdots, vec\left(B_{pt}\right)'\right]'$ is a vector summarizing all VAR coefficients, $v_t \sim N\left(0, \Sigma_t\right)$ with Σ_t an $n \times n$ covariance matrix. This regression-type equation is completed by describing the law of motion of the time-varying parameters β_t and Σ_t. For β_t we follow the standard practice in the literature from Bianchi, Mumtaz and Surico (2009) and consider random walk evolution for our VAR coefficients of the form,

$$\beta_{t+1} = \beta_t + \mu_t, \tag{2.4}$$

based upon a prior β_0 discussed below, and $\mu_t \sim N\left(0, Q_t\right)$. Following Koop and Korobilis (2013) we set $Q_t = \left(\Lambda^{-1} - 1\right) cov\left(\beta_{t-1} | \mathcal{D}_{t-1}\right)$ where $\mathcal{D}_{t-1}$ denotes all the available data at time $t - 1$ and scalar $\Lambda \in (0, 1]$ is a 'forgetting factor' discounting older observations.

The covariance matrix Σ_t evolves according to a Wishart matrix discount process (Prado and West (2010)) of the form:

$$\begin{aligned}
\Sigma_t &\sim iW\left(S_t, n_t\right), && (2.5)\\
n_t &= \delta n_{t-1} + 1, && (2.6)\\
n_t S_t &= (n_t - 1) S_{t-1} + f\left(v_t' v_t\right), && (2.7)
\end{aligned}$$

where n_t and S_t are the degrees of freedom and scale matrix, respectively, of the inverse Wishart distribution, δ is a 'decay factor' discounting older observations, and $f\left(v_t' v_t\right)$ is a specific function of the squared residuals of our model and explained in the Appendix A.1.

Therefore, we have specified a VAR with drifting coefficients and stochastic volatility which allows to model structural instabilities and regime changes in the joint dynamics of the NS factors and the macroeconomic and financial factors. When conducting Bayesian inference Markov Chain Monte Carlo for example needs to be employed, which can be computationally demanding especially in a recursive forecasting context. Here we extend the methodology of Koop and Korobilis (2013) and conduct an efficient estimation scheme to provide accurate results while largely speeding up the estimation procedure. We use what is known as a 'forgetting factor' or 'decay factor' to discount the previous information when updating the parameter estimates; detailed information of our empirical methodology can be found in Appendix A.1.

2.3 Model Selection

2.3.1 Uncertainty about Macro-Finance Factors

The previous subsection describes the specification of a single time-varying parameter Nelson-Siegel model. In this paper we argue that the possible set of risk factors, relevant for characterizing the evolution of the yield curve, can change over time. In this case we are faced with multiple models. In that respect we focus on Eq. (2.3) and we work with three different specifications: small, medium, and large. The small-size (NS) model only contains the three yield factors extracted from the Nelson-Siegel model and zero macro variable, i.e. $q = 0$ in Eq.

(2.3). The middle-size (NS + macro) model includes, in addition to the Nelson-Siegel factors, Federal Fund Rate, CPI and Industrial Production, so $q = 3$. The large (NS + macro-finance) model includes $q = 15$ macroeconomic and financial variables.

Having three models $\mathcal{M}^{(i)} = 1, 2, 3$, in our model space, we use the recursive nature of the Kalman filter to chose to forecast with a different model at each point in time. That is, for each t we chose the optimal $\mathcal{M}^{(i)}$ which maximizes the probability/weight

$$\pi_t^{(i)} = f\left(\mathcal{M}_{t-1}^{TRUE} = \mathcal{M}^{(i)} | Y^{t-1}\right)$$

under the regularity conditions $\sum_{i=1}^{K} \pi_t^i = 1$ and $\pi_t^i \in [0, 1]$, and where $\mathcal{M}_{t-1}^{TRUE}$ is the 'true' model at time $t-1$. We estimate these model weights in a recursive manner, in the spirit of the Kalman filtering approach. We follow Koop and Korobilis (2013) and define a linear forgetting prediction step

$$\pi_{t|t-1}^{(i)} = \frac{\left(\pi_{t-1|t-1}^{(i)}\right)^{\alpha}}{\sum_{i=1}^{K}\left[\mu\left(\pi_{t-1|t-1}^{(i)}\right)^{\alpha}\right]} \tag{2.8}$$

and the updating step

$$\pi_{t|t}^{(i)} \propto \pi_{t|t-1}^{(i)} p^{(i)}\left(z_t | z_{t-1}\right). \tag{2.9}$$

where the quantity $p^{(i)}\left(z_t | z_{t-1}\right)$ is the time t predictive likelihood of model i, using information up to time $t-1$. This quantity is readily available from the Kalman filter and it provides an out-of-sample measure of fit for each model which allows us to construct model probabilities. Finally, $0 < \alpha \leq 1$ is a decay factor which allows to discount exponentially past forecasting performance, that is, it allows to give exponentially higher weight to most recent observations; see Koop and Korobilis (2013) for more information. When $\alpha \to 0$ then we have the case of averaging using equal weights for each model, while when $\alpha = 1$ the predictive likelihood of each observation has the same weight which is basically equivalent to recursively implementing static Bayesian Model Averaging. For all other values between $(0, 1)$ Dynamic Model Averaging occurs.

2.3.2 Prior Selection

We define a Minnesota prior for our VAR, which will guarantee some degree of shrinkage that could prevent overfitting of our larger models. This prior is of the form $\beta_0 \sim N\left(\mathbf{0}, V^{MIN}\right)$ where V^{MIN} is a diagonal matrix with element V_i^{MIN} given by

$$V_i^{MIN} = \begin{cases} \gamma/r^2, & \text{for coefficients on lag } r \text{ where } r = 1, ..., p \\ \underline{\alpha}, & \text{for the intercept} \end{cases}, \tag{2.10}$$

where p is the lag length and $\underline{\alpha} = 1$. The prior covariance matrix controls the degree of shrinkage on the VAR coefficients. To be more specific, the larger the prior parameter γ is, the more flexible the estimated coefficients are and, hence, the lower the intensity of shrinkage towards zero. As the degree of the shrinkage can directly affect the forecasting results, we allow for a wide grid for the reasonable candidate values of γ: $[10^{-10}, 10^{-6}, 0.001, 0.005, 0.01, 0.05, 0.1]$. The best prior γ is selected dynamically according to the forecasting accuracy each value in the grid generates. That is, following Koop and Korobilis (2013) we select γ for each of the three models $M^{(i)} = 1, 2, 3$ and for each time period. Details of this Dyanmic Prior Selection (DPS) procedure can also be found in the Appendix A.2.

In this paper we also need to calibrate some other free parameters: the NS factor parameter λ^{NS} in Eq. (2.1), forgetting factors Λ in Eq. (A.3), α in Eq. (A.5), and decay factor δ in Eq. (A.2). We have already mentioned that following Diebold and Li (2006), Bianchi, Mumtaz and Surico (2009) and Van Dijk et al. (2014) we set $\lambda^{NS} = 0.0609$. Regarding the forgetting factors and the decay factor, we may need some more discussion. Intuitively, these parameters control the persistence of previous information. When these parameters are fixed at 1, our time-varying parameter model will become the fixed parameter model. However, as discussed in Koop and Korobilis (2013), too small values may induce sudden changes to outliers, so the state space system is not stable and the results will not be robust. Another reason to calibrate high values for these factors is due to the persistence of bond yields; low values for free parameters will weaken the bond yield predictions. Hence, we choose relatively high values (less than 1) to ensure stability while still allowing for flexibility: The Λ, α and δ are set to 0.99, 0.99 and 0.95, respectively.

3 Data and Results

This study uses the smoothed yields provided from the US Federal Reserve by Gürkaynak, Sack and Wright (2007). We also include 3- and 6-month Treasury Bills (Secondary Market Rate). The empirical analysis focuses on yields with maturities of 3, 6, 12, 24, 36, 48, 60, 72, 84, 96, 108 and 120 months. The key macroeconomic and financial variables that enter our Dynamic Model Averaging model are obtained from St. Louis Federal Reserve Economic Data (FRED). These include inflation, real activity indicators, monetary policy tools, as well as the stock market, exchange rate, house prices and other financial market indicators; the details can be found in Data Appendix. The full sample is from November 1971 to November 2013 and we use end of the month yield data. The 1, 3, 6 and 12 months ahead predictions are produced with a training sample of 38 observations from the start of our sample, up to and including December 1974. We present the yields' descriptive statistics in Table 1. As expected the mean of yields increase with maturity, consistent with the existence of a risk premium for long maturities. Yields have high autocorrelation which declines with lag length and increases with maturity. The short end of the yield curve is more volatile than the long end.

Different numbers of macro-finance variables are selected for the three VARs entering our DMA. As mentioned above, the small-size VAR (NS) does not include any macro or financial variables, but only the Nelson-Siegel factors. The middle-size VAR (NS + macro) includes Federal Fund Rate, inflation and Industrial Production, which are also used in related literature such as Ang and Piazzesi (2003) and Diebold, Rudebusch and Aruoba (2006). The large VAR (NS + macro-finance) includes all 15 macro and financial variables, which should comprehensively include the information the market players are able to acquire.

Table 1: **Descriptive Statistics of Bond Yields**

	Mean	Std. Dev.	Minimum	Maximum	$\hat{\rho}(1)$	$\hat{\rho}(12)$	$\hat{\rho}(30)$
3	5.154	3.341	0.010	16.300	0.987	0.815	0.533
6	5.284	3.320	0.040	15.520	0.988	0.827	0.557
12	5.675	3.440	0.123	16.110	0.987	0.842	0.599
24	5.910	3.355	0.188	15.782	0.988	0.858	0.648
36	6.102	3.259	0.306	15.575	0.989	0.868	0.677
48	6.266	3.161	0.454	15.350	0.990	0.873	0.695
60	6.411	3.067	0.627	15.178	0.990	0.876	0.707
72	6.539	2.980	0.815	15.061	0.990	0.877	0.714
84	6.653	2.902	1.007	14.987	0.990	0.878	0.718
96	6.754	2.833	1.197	14.940	0.990	0.878	0.721
108	6.843	2.772	1.380	14.911	0.990	0.878	0.722
120	6.920	2.720	1.552	14.892	0.990	0.877	0.723
Level	7.437	2.379	2.631	14.347	0.989	0.866	0.700
Slope	-2.277	1.940	-5.824	4.522	0.954	0.492	-0.114
Curvature	-1.424	3.222	-8.948	5.282	0.903	0.634	0.369

Notes: This table presents descriptive statistics for monthly yields at 3- to 120-month maturity, and for the yield curve Level, Slope and Curvature factors extracted from the Nelson-Siegel model. The sample period is 1971:11–2013:11. We use following abbreviations. **Std. Dev.**: Standard Deviation; $\hat{\rho}(\mathbf{k})$: Sample Autocorrelation for Lag k.

3.1 Evidence on Parameter Instability

In this section we seek to validate the use of time-varying parameter methods. There is a vast selection of different tests of parameter instabilities and structural breaks in the literature from both a frequentist and a Bayesian perspective; see for example Chow (1960), Quandt (1960), Nyblom (1989), Andrews (1993), Andrews and Ploberger (1994), Hanson (2002) and Rossi (2005). McCulloch (2007) suggests a likelihood-based approach to test parameter instabilities in a TVP model, but the limiting distribution of the test statistics may not be standard and hence the critical values need to be bootstrapped. In the spirit of McCulloch (2007), we construct a likelihood-based test on the VAR system of the factor dynamics, using the 1983-2013 sample. We bootstrap 5000 samples to recover the test statistics following Feng and McCulloch (1996). Based on our test, the null hypothesis that the coefficients of the VAR are constant over time is rejected at 1% significance level, which means employing the TVP-VAR model is appropriate.

However, all the tests mentioned above are in-sample tests and fail to provide evidence concerning out-of-sample instabilities. Therefore, instead of explicitly specifying a test of parameter instability we follow a different strategy. First, note that in the case of our model specified in Section 2, the constant parameter Nelson-Siegel model can be obtained as a special case of our proposed time-varying specification, that is it is nested.[9] Since our ultimate purpose is to obtain optimal forecasts of the yield curve, "testing" for parameter instability can conveniently boil down to a comparison of predictability between the TVP-VAR and a constant parameter VAR. We employ the test proposed by Diebold and Mariano (1995) and evaluate the predictability of competing models across four forecast horizons ($h = 1, 3, 6, 12$ months) and at all twelve of our maturities. The p-values of the tests are reported in Table 2, which correspond to the test of the null hypothesis that the competing TVP-VAR model has equal expected square prediction error relative to the benchmark forecasting model constant parameter VAR (i.e. Diebold and Li (2006)), against the alternative hypothesis that the competing TVP-VAR forecasting model has a lower expected square prediction error than the benchmark forecasting model. Table 2 indicates the TVP-VAR consistently outperforms the constant parameter VAR. The test statistic rejects the null for most of the maturities, and especially at longer forecast horizons, so the time-varying parameter model should be preferred as it can provide more robust estimates.

[9] In particular, as Koop and Korobilis (2013) show, by setting the forgetting and decay factors $\Lambda = \delta = 1$, then β_t and Σ_t remain constant over the sample $t = 1, ..., T$.

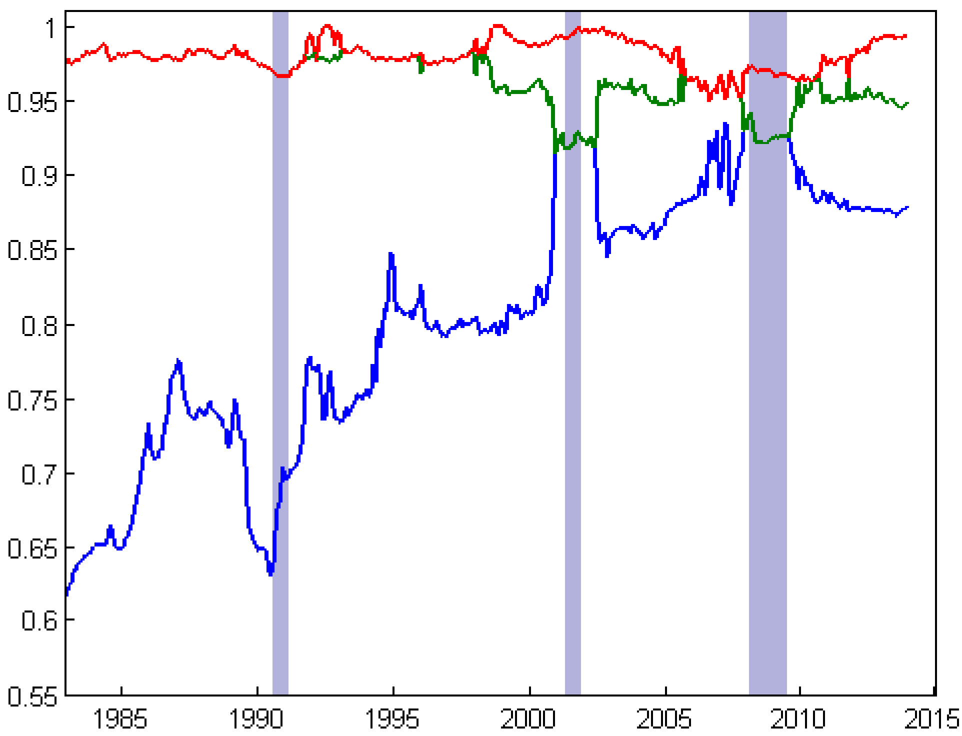
Table 2: **Parameter Instability Test**

P-Values: TVP-VAR vs. VAR												
Maturity	3	6	12	24	36	48	60	72	84	96	108	120
$h = 1$	0.02	0.00	0.54	0.14	0.02	0.00	0.00	0.00	0.01	0.08	0.33	0.68
$h = 3$	0.03	0.01	0.13	0.04	0.01	0.01	0.00	0.01	0.02	0.05	0.13	0.28
$h = 6$	0.00	0.00	0.04	0.02	0.01	0.01	0.01	0.01	0.02	0.04	0.08	0.16
$h = 12$	0.00	0.00	0.02	0.01	0.00	0.00	0.00	0.00	0.00	0.01	0.01	0.03

Notes: 1. This table reports the statistical significance for the relative forecasting performance, based on the Diebold and Mariano (1995) test. We conduct 1, 3, 9 and 12 months ahead forecasts for bond yields at maturities ranging from 3 months to 120 months. The predictive period is between 1983:10 and 2013:11.
2. Statistical significance for the relative MSFE statistics is based on the p-value for the Diebold and Mariano (1995) statistic; the statistic corresponds to the test of the null hypothesis that the competing TVP-VAR model has equal expected square prediction error relative to the benchmark forecasting model constant parameter VAR (i.e. Diebold and Li (2006)), against the alternative hypothesis that the competing forecasting model has a lower expected square prediction error than the benchmark forecasting model.

To highlight the importance of the TVP feature, we set out the persistence of the physical factor dynamics over time in Figure 1. This can be be examined by considering the behavior of the eigenvalues. We can detect significant changes in all eigenvalues, which reflects indispensable changes in the persistence of pricing factors over time. The first eigenvalue is relatively more stable than the other two, while there is a clear rising trend for the third eigenvalue. Moreover, we find that the second and third eigenvalues have important changes in near recession periods, which is connected to the shifting dynamics of Slope and Curvature factors.

Figure 1: Time-Varying Persistence of Physical Dynamics



Notes: The graph shows the largest three eigenvalues of the physical dynamics in the TVP model. The shaded areas are recession periods according to the NBER Recession Indicators.

3.2 Model Dynamics

In our Bayesian empirical analysis of the factor dynamics, we begin by selecting priors with Dynamic Prior Selection (DPS), then the best prior will be selected for each of the three VAR models. Next we update the model weights with Dynamic Model Averaging (DMA), and finally we update on the parameters using a Bayesian Kalman filter.

In the DPS step, we find that the best prior γ value in Eq. (2.10) is stable, i.e. fixed at 0.1, for all three VAR models, given the forgetting factor $\alpha = 0.99$. To ensure robustness, we decrease the values of α, as it controls the persistence of probabilities.[10] As α decreases the results do not change substantially: the best γ values is typically 0.1 for all three sizes

[10] In Appendix C, Figure 14 shows the prior selection results with different values of the forgetting factor.

of models. The evidence concludes that a relatively flexible and consistent prior can generate more accurate yield forecasts. For simplicity and tractability, we fix the value at $\gamma = 0.1$, and therefore the DPS procedure could be skipped in the following analysis. In fact, we find that holding γ constant at 0.1 slightly improves the forecasts, though the comparison of the forecasting results will not be reported in this paper due to limited space.

Graphical evidence of the usefulness of our approach is provided by Figure 2, which sets out the weights of the small, medium and large VAR models used in DMA. Interestingly our updating procedure implies we should use more macro-finance information in particular time periods. The following empirical observations are of economic importance.

Firstly, during recession periods, the approach tends to use more macro-finance information to generate forecasts. For instance, immediately before the financial crisis, the probability of the large-size (macro-finance) model rose steeply and then stayed at a high level throughout the whole crisis period, as indicated by the higher weights for the small NS model in Figure 2. In times of acute economic stress, macroeconomic and financial risk factors become more relevant for modeling yields, which is closely related to the 'financial accelerator' argument by Bernanke, Gertler and Gilchrist (1996).[11] The macro-finance model also displays considerable variability in importance, as displayed by the volatility of the probabilities in Figure 2.

Secondly, the small-size NS model generally has relatively high probability in the DMA except during recession periods. This is consistent with the viewpoint that only information from the bond market is used in pricing and predicting bond yields. It explains the effectiveness, at least during non-recession periods, of parsimonious yield curve models, such as Dai and Singleton (2003) and Diebold and Li (2006).

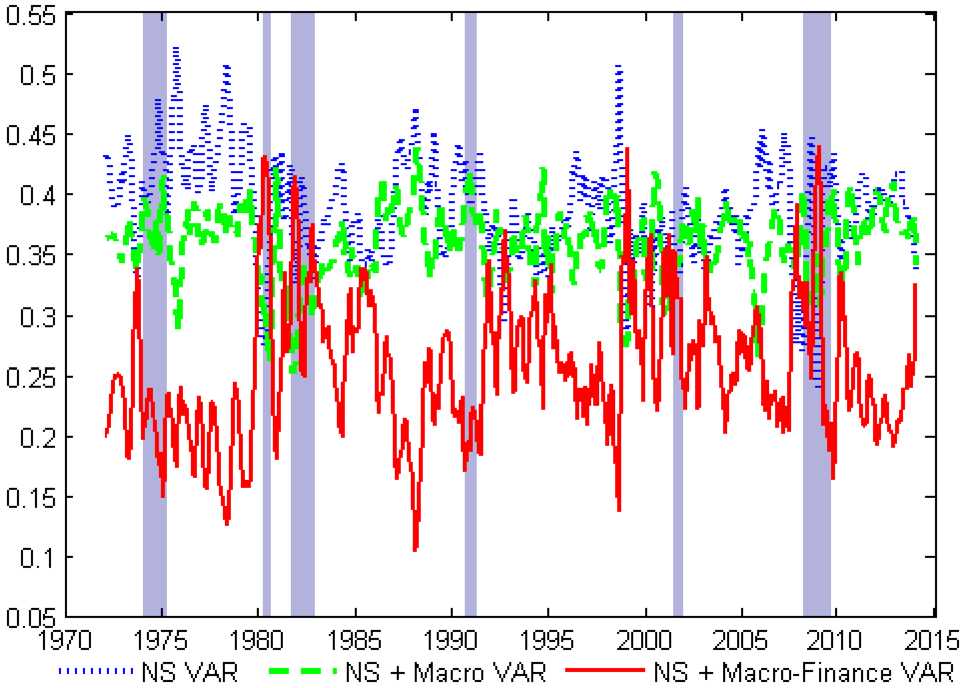
Thirdly, the probability of the medium-size (NS + macro) model is comparable to the small-size model since 1980s. This is consistent with the idea that macro variables are important

[11] This may also be explained by the construction in Fontaine and Garcia (2012): In the financial crisis, the arbitrageurs that use bond-market information only are capital-constrained and hence have funding stress, whereas the agents concerning more about macroeconomy and the whole financial condition, for example, the government, step in and drive the market. Hu, Pan and Wang (2013) have a related argument that the arbitrageurs help align the yields across maturities in normal periods but the pricing errors can be persistently high when arbitrage capital is low. These conjectures need to be confirmed with further evidence.

in determining yields since the start of 'Great Moderation', due to an active central bank, see Ang and Piazzesi (2003) and Bianchi, Mumtaz and Surico (2009).

Lastly, it seems that there is a slightly upward trend for the large macro-finance VAR since 1970, which implies that the agents tend to incorporate more and more macro and financial information when making investment decisions. This feature is consistent with the observation in Altavilla, Giacomini and Ragusa (2014) that the original version of the dynamic NS model without macro information has weaker predictive power in recent years.

Figure 2: Model Weights for NS, NS plus Macro and NS plus Macro-Finance VAR Models



Notes:
1. This figure sets out the time-varying probabilities of our three models in our Dynamic Model Averaging (DMA) approach. The probabilities for DMA are updated from a Kalman filter based on the predictive accuracy, see Eq. (A.6); the probabilities/weights of the VAR models sum up to 1. These updated estimates are smoothed using a 6-month moving average.
2. The three models are as follows. The small VAR contains the Nelson-Siegel (NS) factors. The medium VAR contains NS plus macro factors. The large VAR contains NS plus macro-finance factors.
3. The shaded areas the recession periods based on NBER Recession Indicators.

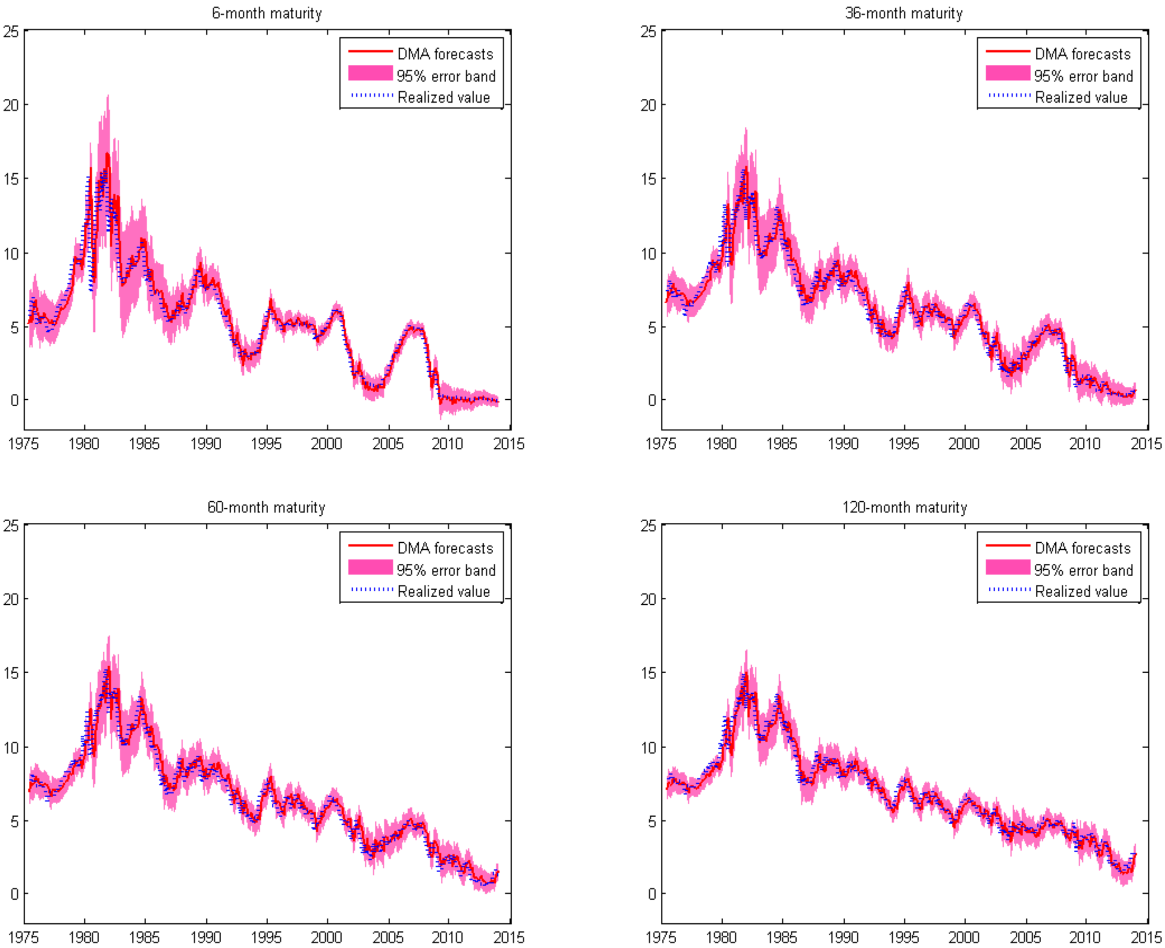
3.3 Forecasting Performance

As mentioned above, we use a NS-DMA model to predict the yields in a two-step estimation procedure. The first stage is using the Kalman filter to generate predictions of the three Nelson-Siegel yield factors with macro variables, using DMA. That is we use Eq. (2.3) with the predicted β_{t+1} to forecast our factors. The second stage is forecasting the yields with the predicted NS factors and the fixed NS loadings. The macro variables are not directly used to predict the yields in the second step, due to the consideration of Unspanned macro risks. The point forecasts of NS-DMA are compared to the realized yields across all maturities, and we also compute the predictive log-likelihood of forecasting models to evaluate the density forecasts. In terms of density forecasts, the comparison exercise using predictive likelihoods is similar to Geweke and Amisano (2010). The predictive duration is from 1983:10 to the 2013:11. Figure 3 displays the 3 months ahead forecasts of yields with 95% error bands against the realized values, generated by the NS-DMA model.

To better evaluate the predictive performance of NS factors and hence yields, we have the following seven benchmark models to compare with NS-DMA/DMS: Random Walk (RW) model, recursive estimations of factor dynamics using standard VAR following Diebold and Li (2006) (DL), 10-year rolling-window VAR estimations (DL-R10), recursive VAR estimations with three macro variables (DL-M), recursive estimations of standard VAR with macro principal components following Stock and Watson (2002) (DL-SW), time-varying parameter VAR estimations of factor dynamics without macro information (TVP) and time-varying parameter VAR estimations of factor dynamics with three macro variables (TVP-M).

The Random Walk (RW) model generates future yield predictions using the current information of the yields, as the current yield factors are the unbiased estimators of the future factor forecasts. The RW model is a challenging benchmark, as Duffee (2002) remarks it is hard for term structure models to beat it. This may be because yields are highly persistent and have a mean-reverting property. DL is the two-step forecasting model proposed by Diebold and Li (2006), which recursively estimates the factor dynamics using a standard VAR. In other words,

Figure 3: DMA Forecasts of Yields



Notes: These are 3 months ahead forecasts (95% error band) for yields against realized values with maturities 6, 36, 60 and 120 months, from early 1975 to late 2013. The forecasts are two-step forecasting using NS-DMA, which can be summarized by Eq. (2.1), (2.3) and (2.4).

DL estimates the VAR model of factors recursively with historical data, extending through all the following periods. We have four variations of the DL model: 10-year rolling-window estimations (DL-R10); recursive estimations with three macro variables of Fed Fund Rate, Inflation and Industrial Production (DL-M); and recursive estimations with three principal components of our whole macro-finance dataset (DL-SW). In the DL-SW model, three macro principal components are drawn using the method proposed by Stock and Watson (2002) to augment DL. Lastly, we include two extensions of DL using a time-varying parameter VAR without macro information and a time-varying parameter VAR with three macro variables to characterize the factor dynamics, denoted TVP and TVP-M, respectively; the latter has a similar model structure as in Bianchi, Mumtaz and Surico (2009), but here it is estimated in two steps with a fast algorithm proposed by Koop and Korobilis (2013). This obviates the need to employ the time-consuming Markov Chain Monte Carlo (MCMC) algorithm.

3.3.1 Point Forecasts

Table 3 and 4 display the 1-period and 3-period ahead Mean Squared Forecasting Error (MSFE) Performance of all forecasting models.[12] The core empirical results are very encouraging. As can be seen in Table 3 and 4, our preferred NS-DMA approach consistently outperforms the benchmark model. That is to say, the NS-DMA has a lower MSFE than the RW for nine of twelve maturities in the one-month ahead forecasts in Table 3.

Even at relatively long forecast horizons, the NS-DMA also performs better than the RW in average.[13] In the one-year long-term forecasts, without any further information, the NS-DMA performance is comparable to the RW. Therefore, our NS-DMA approach seems to better reflects the true dynamics of the yield factors by properly characterizing the nonlinear evolution of yield factors. In terms of density forecasts, the log-likelihood of NS-DMA is systematically the highest among all forecasting models, see Table 3. Among all models, NS-DMA is the only one comparable to, or better than, the RW. The DMS, TVP-M and the original DL have rea-

[12]More forecasting results are shown in Appendix C.

[13]See Appendix C for these details.

Table 3: **One-Month Ahead Relative MSFE of Term Structure Models**

MA	NS-DMA	DMS	TVP	TVP-M	DL	DL-R10	DL-M	DL-SW
3	**0.706**†	0.781	0.747	0.710	0.848	1.085	0.885	1.417
6	**0.818**†	0.927	0.894	0.908	1.068	1.313	1.130	1.668
12	0.971†	1.031	0.983	1.011	0.930	**0.897**	0.979	1.547
24	**1.000**†	1.075	1.044	1.060	1.064	1.105	1.103	1.461
36	**0.977**†	1.039	1.032	1.026	1.123	1.223	1.144	1.237
48	**0.965**†	1.008	1.016	1.002	1.130	1.266	1.143	1.099
60	**0.965**†	0.996	1.011	0.997	1.116	1.273	1.129	1.051
72	**0.971**†	0.998	1.015	1.006	1.096	1.259	1.114	1.055
84	**0.982**†	1.008	1.026	1.024	1.074	1.226	1.098	1.090
96	**0.996**†	1.023	1.040	1.046	1.052	1.173	1.083	1.139
108	**1.009**†	1.038	1.055	1.068	1.031	1.108	1.068	1.183
120	1.020†	1.050	1.065	1.084	**1.015**	1.043	1.053	1.214
Mean	**0.964**†	1.009	1.008	1.010	1.053	1.162	1.083	1.237

Notes: 1. This table shows 1-month ahead forecasts of bond yields with maturities ranging from 3 months to 120 months. The predictive duration is from 1983:10 to 2013:11.
2. We report the ratio of each models Mean Squared Forecast Errors (MSFE) relative to Random Walk MSFE, and the preferred values are in bold. The dagger (†) indicates, in terms of the sum of predictive log-likelihood, the model has the preferred value among all models at certain maturities (or in total), see Geweke and Amisano (2010) for details.
3. In this table, we use following abbreviations. **MA**: Maturity (Months); **MSFE**: Mean Squared Forecasting Error; **Mean**: Averaged MSFE across all sample maturities. In our proposed Nelson-Siegel (**NS**) framework, **DMA** (Dynamic Model Averaging) averages all the models with probabilities in each step, while **DMS** (Dynamic Model Selection) chooses the best model with the highest probability at any point in time. **TVP**: a time-varying parameter model without macro information; **TVP-M**: a time-varying parameter model with three macro variables: fund rate, inflation and industrial production, similar to Bianchi Mumtaz and Surico (2009) but estimated with a fast algorithm without the need of MCMC; **DL**: Diebold and Li (2006) model, i.e. constant coefficient Vector Autoregressive model with recursive (expanding) estimations; **DL-R10**: Diebold and Li (2006) estimates based 10-year rolling windows; **DL-M**: factor dynamics in Diebold and Li (2006) are augmented with three macro variables: fund rate, inflation and industrial production, using recursive estimations; **DL-SW**: factor dynamics in Diebold and Li (2006) are augmented with with three principal components (see Stock and Watson (2002)) of our macro/finance data, using recursive estimations; **RW**: Random Walk.

Table 4: **Three-Month Ahead Relative MSFE of Term Structure Models**

MA	NS-DMA	DMS	TVP	TVP-M	DL	DL-R10	DL-M	DL-SW
3	**0.765**†	0.873	0.864	0.845	1.105	1.514	1.070	1.795
6	**0.863**†	0.976	0.976	0.997	1.305	1.646	1.283	1.907
12	**0.931**†	1.003	0.997	1.019	1.131	1.231	1.119	1.727
24	**0.988**†	1.046	1.062	1.068	1.255	1.390	1.249	1.537
36	**1.002**†	1.044	1.073	1.060	1.295	1.482	1.292	1.358
48	**1.006**†	1.037	1.069	1.049	1.294	1.528	1.293	1.246
60	**1.006**†	1.032	1.063	1.043	1.269	1.539	1.272	1.196
72	**1.005**†	1.030	1.057	1.041	1.233	1.525	1.239	1.189
84	**1.002**†	1.029	1.053	1.044	1.190	1.488	1.201	1.207
96	**0.999**†	1.031	1.050	1.049	1.146	1.431	1.160	1.238
108	**0.996**†	1.033	1.049	1.055	1.102	1.360	1.120	1.272
120	**0.994**†	1.035	1.048	1.061	1.062	1.283	1.083	1.302
Mean	**0.969**†	1.018	1.035	1.032	1.205	1.449	1.205	1.405

Notes: 1. This table shows 3-month ahead forecasts of bond yields with maturities ranging from 3 months to 120 months. The predictive duration is from 1983:10 to 2013:11.
2. We report the ratio of each models Mean Squared Forecast Errors (MSFE) relative to Random Walk MSFE, and the preferred values are in bold. The dagger (†) indicates, in terms of the sum of predictive log-likelihood, the model has the preferred value among all models at certain maturities (or in total), see Geweke and Amisano (2010) for details.
3. In this table, we use following abbreviations. **MA**: Maturity (Months); **MSFE**: Mean Squared Forecasting Error; **Mean**: Averaged MSFE across all sample maturities. In our proposed Nelson-Siegel (**NS**) framework, **DMA** (Dynamic Model Averaging) averages all the models with probabilities in each step, while **DMS** (Dynamic Model Selection) chooses the best model with the highest probability at any point in time. **TVP**: a time-varying parameter model without macro information; **TVP-M**: a time-varying parameter model with three macro variables: fund rate, inflation and industrial production, similar to Bianchi Mumtaz and Surico (2009) but estimated with a fast algorithm without the need of MCMC; **DL**: Diebold and Li (2006) model, i.e. constant coefficient Vector Autoregressive model with recursive (expanding) estimations; **DL-R10**: Diebold and Li (2006) estimates based 10-year rolling windows; **DL-M**: factor dynamics in Diebold and Li (2006) are augmented with three macro variables: fund rate, inflation and industrial production, using recursive estimations; **DL-SW**: factor dynamics in Diebold and Li (2006) are augmented with with three principal components (see Stock and Watson (2002)) of our macro/finance data, using recursive estimations; **RW**: Random Walk.

sonable forecasting power among the remaining models, especially the DMS and TVP-M which have flexible time-varying parameter settings and incorporates some useful macroeconomic and financial information. It is worth noting that the rolling-window forecasts perform much less favorably, as rolling-window models discard some potentially helpful information that is not included in the windows. Hence, our TVP specification is always preferred in this sense.

In addition, the predictability of DL-SW is not satisfactory. As a variable shrinkage method, the macro principal components alone cannot provide useful information in terms of yield forecasting, since the method fails to exclude irrelevant information in a time-varying manner. Hence this result indicates the relative advantages of NS-DMA as a variable shrinkage method in forecasting.

Remarks on Predictive Gains Since the pricing dynamics are constrained by the NS restrictions, we conclude that the predictive gains are purely from the physical dynamics when taking parameter and model uncertainty into account. Here we would like to highlight different sources of predictive gains. As mentioned in the last section, the last four columns in Table 3 or 4 set out the predictive performance of constant-parameter models without stochastic volatility, which perform consistently worse than TVP models, no matter whether we include macro information or not. In contrast, our TVP models with stochastic volatility in the third and fourth columns provide significant gains in predictive performance, as they put more weights on the current observations and hence are robust to parameter uncertainty and structural changes. Moreover, introducing an extra layer of model uncertainty is also essential in improving forecast performance. It helps us properly incorporate macro-finance information in a time-varying manner, which is related to the 'scapegoat theory' by Bacchetta and Van Wincoop (2004). From the first two columns, we find further improvement over the TVP models if we allow for both parameter and model uncertainty. Hence, we believe that *the ensemble of these salient features – time-varying parameters, stochastic volatility and model averaging/selection, is the key to properly incorporate macro-finance information and hence can provide significant gains in predictability.*

To formalize the above arguments, we conduct a statistical test to evaluate the out-of-sample forecasting performance. In Table 5, we performs the test proposed by Diebold and Mariano (1995), in order to evaluate the forecasting performance of NS-DMA relative to DL, TVP-M and Random Walk, respectively. The Diebold and Mariano (1995) statistics are used by Diebold and Li (2006) and Altavilla, Giacomini and Ragusa (2014). The relative MSFE is shown at forecasting horizons of 1, 3, 6 and 12 months, i.e., $h = 1, 3, 6, 12$. It shows that the NS-DMA clearly outperforms the DL and TVP-M, though it only has statistical significance relative to the RW at shorter maturities. It implies the short rate forecasts of the NS-DMA are satisfactory. In general, the predictive performance of NS-DMA in some medium-term maturities is weaker than in other maturities, implying that some additional information may be needed to better capture the movements of the hump-shape Curvature factor.

Table 5: **MSFE from NS-DMA Relative to Other Models**

	NS-DMA vs. DL				NS-DMA vs. TVP-M				NS-DMA vs. RW			
Maturity	$h = 1$	$h = 3$	$h = 6$	$h = 12$	$h = 1$	$h = 3$	$h = 6$	$h = 12$	$h = 1$	$h = 3$	$h = 6$	$h = 12$
3	0.833***	0.693***	0.653***	0.843***	0.995	0.906*	0.860*	0.790**	0.706***	0.765***	0.871*	1.028
6	0.766***	0.661***	0.655***	0.846***	0.901**	0.865**	0.845**	0.800**	0.818**	0.863**	0.947	1.054
12	1.045	0.824**	0.743***	0.866***	0.961**	0.914**	0.897*	0.847**	0.971	0.931*	0.969	1.031
24	0.939**	0.788***	0.735***	0.849***	0.943***	0.925**	0.927*	0.890*	1.000	0.988	1.025	1.055
36	0.870***	0.774***	0.733***	0.845***	0.952***	0.945**	0.952	0.918	0.977	1.002	1.038	1.063
48	0.854***	0.777***	0.740***	0.842***	0.963**	0.959*	0.967	0.934	0.965	1.006	1.038	1.061
60	0.864***	0.793***	0.754***	0.844***	0.967**	0.965*	0.973	0.939	0.965	1.006	1.032	1.054
72	0.886***	0.815***	0.773***	0.846***	0.965**	0.965*	0.971	0.936	0.971	1.005	1.021	1.048
84	0.914***	0.842***	0.794***	0.849***	0.959**	0.960*	0.965	0.928	0.982	1.002	1.009	1.041
96	0.947**	0.872**	0.819**	0.851***	0.951**	0.953**	0.955	0.918	0.996	0.999	0.997	1.032
108	0.978*	0.904**	0.845**	0.854***	0.945***	0.944**	0.946	0.907	1.009	0.996	0.987	1.019
120	1.004	0.936	0.872*	0.860***	0.941***	0.937***	0.937	0.897	1.020	0.994	0.978	1.007

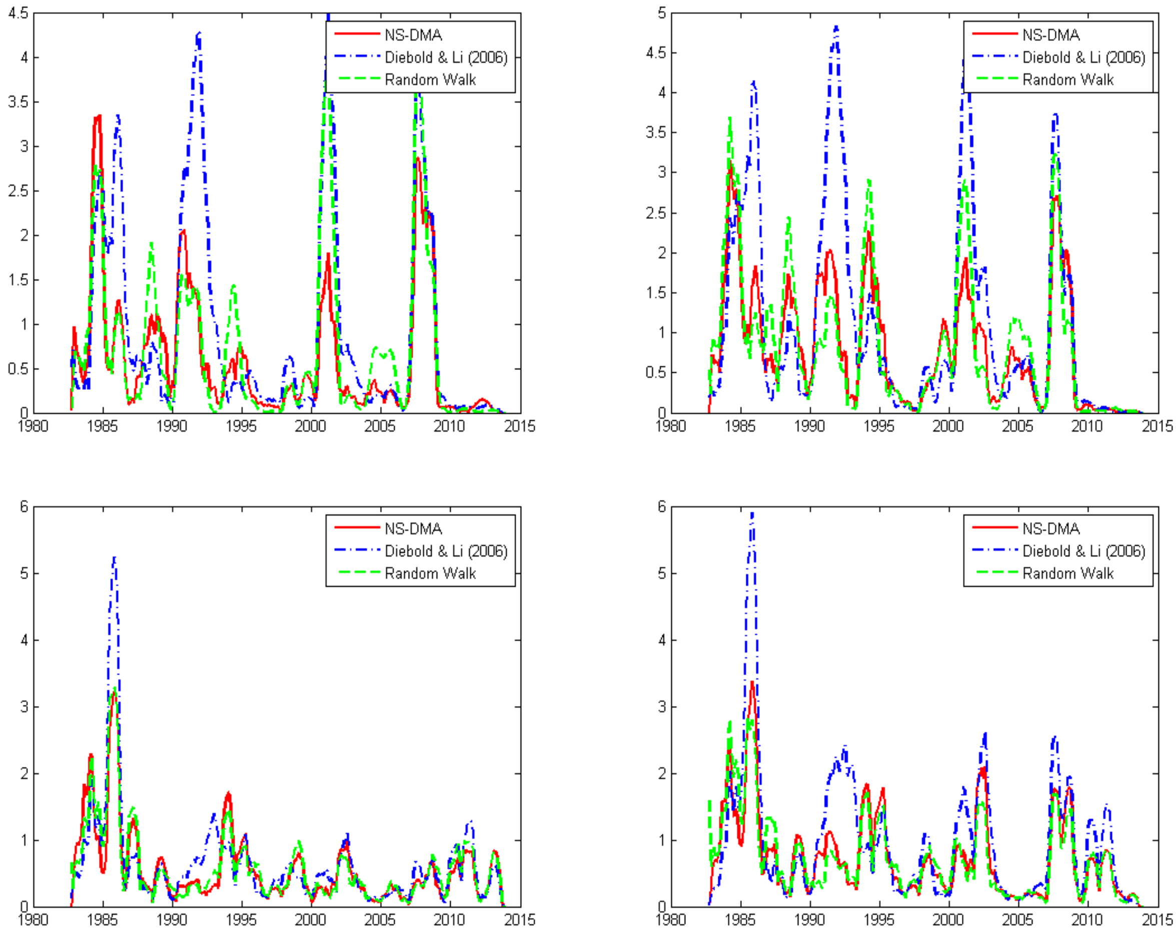
Notes: 1. This table reports MSFE-based statistics of NS-DMA forecasts of bond yields at maturities ranging from 3 months to 120 months, relative to the forecasts of Diebold and Li (2006) (DL), TVP-M (similar to Bianchi Mumtaz and Surico (2009)) or Random Walk (RW). The predictive period is between 1983:10 and 2013:11.
2. Statistical significance for the relative MSFE statistics is based on the p-value for the Diebold and Mariano (1995) statistic; the statistic corresponds to the test of the null hypothesis that the competing NS-DMA model has equal expected square prediction error relative to the benchmark forecasting model (DL, TVP-M or RW) against the alternative hypothesis that the competing forecasting model has a lower expected square prediction error than the benchmark forecasting model. *, ** and *** indicate significance at the 10%, 5%, and 1% levels, respectively.

One interesting observation about the NS-DMA is that at a long forecasting horizon (12-month ahead), the forecasts of long-end of the term structure are relatively better than the shorter-term bonds, and it is the opposite for a short forecasting horizon (1-month ahead). The following may explain the above observation. Generally, the long-term yields have lower volatility so the forecasts are stable. On the other hand, the short yields are anchored by the policy rates in a short period, so the forecasts of short yields in short horizon are vary accurate; however, without further information, the forecasts of short yields at a longer forecasting horizon are weaker, because the monetary policy target may change in the long run. In comparing our results to the existing literature, Diebold and Li (2006) beat a random walk using Diebold-Mariano test at 12-month forecasting horizons and for shorter maturities. However, Diebold and Rudebusch (2013) and Altavilla, Giacomini and Ragusa (2014) imply reduced ability of NS models to beat RW in recent years. We consistently improve upon DL across all horizons and maturities, which is confirmed by Relative MSFE, predictive log-likelihoods and Diebold-Mariano test.

Predictive Performance over Time To display the how the superior performance of our our method arises, Figure 4 shows the 6-month ahead Squared Forecasting Errors of DMA, DL and RW across the predictive period. It is clear that the DMA significantly and consistently outperforms the DL across all maturities and the RW at shorter maturities. It seems benchmark models perform much worse in near recession periods, while NS-DMA has stable performance due to its robustness to parameter and model uncertainty.

Note that all the models in this section are estimated via a two-step method, of which the first step is applying NS model, so the previous comparison is based on the NS framework. We do not include the type of Affine Term Structure Models (ATSM) such as in Ang and Piazzesi (2003) and Ang, Dong and Piazzesi (2007) for comparison for the following reasons. Theoretically, these models can be used for forecasting. However in practice, as indicated in Ang and Piazzesi (2003), the likelihood function is flat and hence the identification is very time-consuming, even though with additional restrictions in parameters. In addition, we perform

Figure 4: Squared Forecasting Error for Yields of 3-, 12-, 60- and 120-Month Maturities



Notes: 1. These are 6 months ahead Squared Forecasting Errors for predicted yields from early 1983 to late 2013. From top left clockwise we have maturities of 3, 12, 60 and 120 months. The models present here are DMA (solid), Diebold-Li (dashed and dotted) and RW (dashed).
2. The first two graphs show the errors for yields of maturities 3 and 12 months, in which the DMA significantly outperforms the DL and RW.

out-of-sample forecasts at a long horizon with a relatively small training sample, so we may fail to identify the parameters at some points when estimating the model recursively. The reason may be that the economic structure is changing over time and if we include the data before and after a structural change, the likelihood function might be even flatter. Besides, the restrictions are not time-varying, so the identification may be infeasible when facing a changing economic structure. Moreover, the forecast performance of ATSM are close to or even weaker than the ordinary NS model, see for example Christensen, Diebold and Rudebusch (2011), Duffee (2011*a*) and Joslin, Singleton and Zhu (2011). Indeed, the out-of-sample performance of ATSM-type models can be quite weak so they are not suitable as benchmark models.

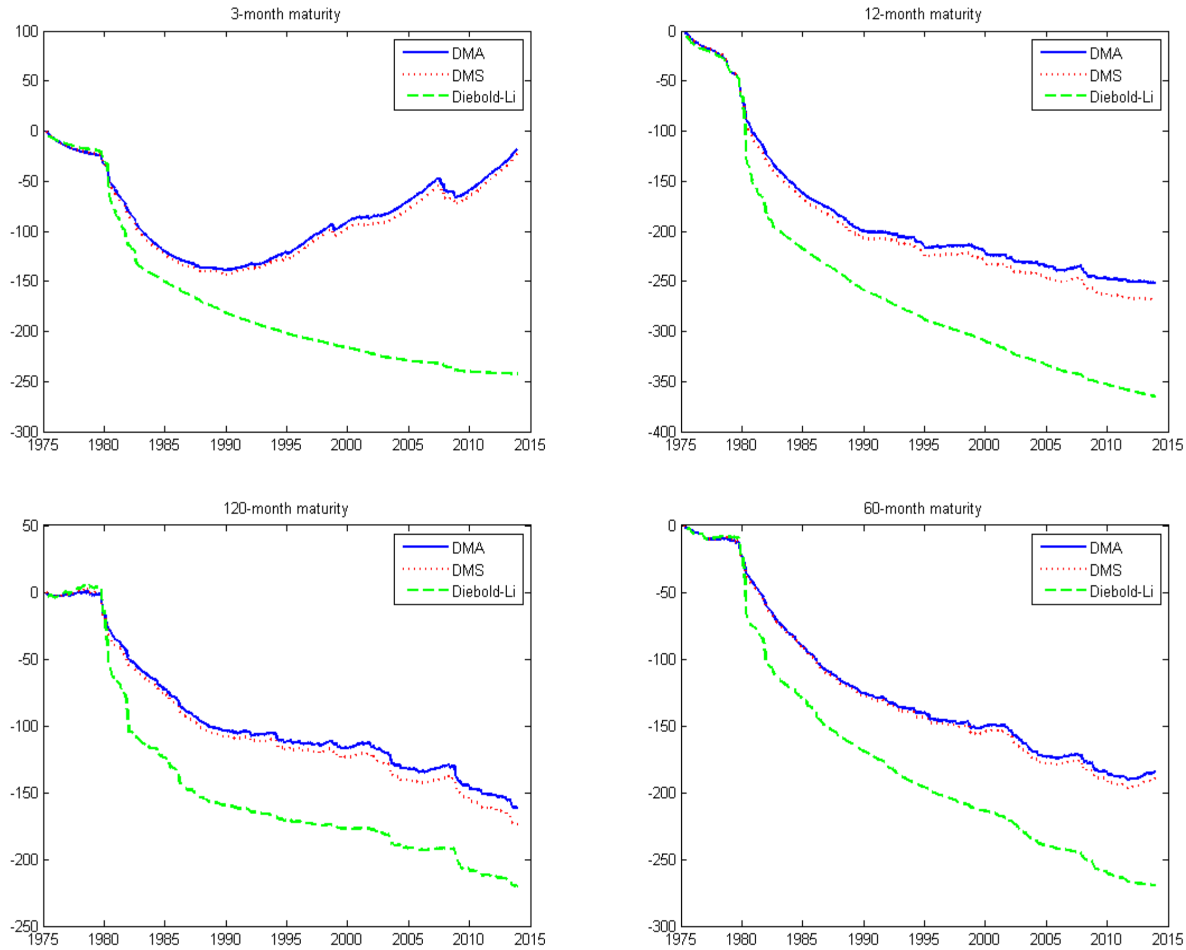
3.3.2 Density Forecasts and Time-Varying Volatility

It has been indicated by Bianchi, Mumtaz and Surico (2009) that homoskedasticity is a frequent and potentially inappropriate assumption in much of the macro-finance literature. Cieslak and Povala (2015) show that stochastic volatility can have a non-trivial influence on the conditional distribution of interest rates. Piazzesi (2010) indicates that fat tails in the distribution of the bond factors can be modeled by specifying an appropriate time-varying volatility. The dynamics of the bond yields therefore exhibit a heavy-tailed property in the unconditional distribution, as the conditional volatility is higher when the yields deviate more from the unconditional mean. The property of asymmetry/skewness is also implied by the yields dynamics due to the evolution of the innovation variances.

Our model relaxes the unrealistic homoskedasticity assumption and hence provides favorable density forecast performance, which is consistent with the evidence of Hautsch and Yang (2012). The cumulative sum of predictive log-likelihood displayed in Figure 5 shows the preferred predictive density of NS-DMA over DL across all maturities, especially for short rates. Hence, the NS-DMA should be preferred, as the important and realistic feature of stochastic volatility cannot be characterized by ordinary constant parameter models.

The NS-DMA not only provides more sensible results in terms of density forecasts, but also captures the desirable evolutionary dynamics of the economic structure. Figure 6 shows the

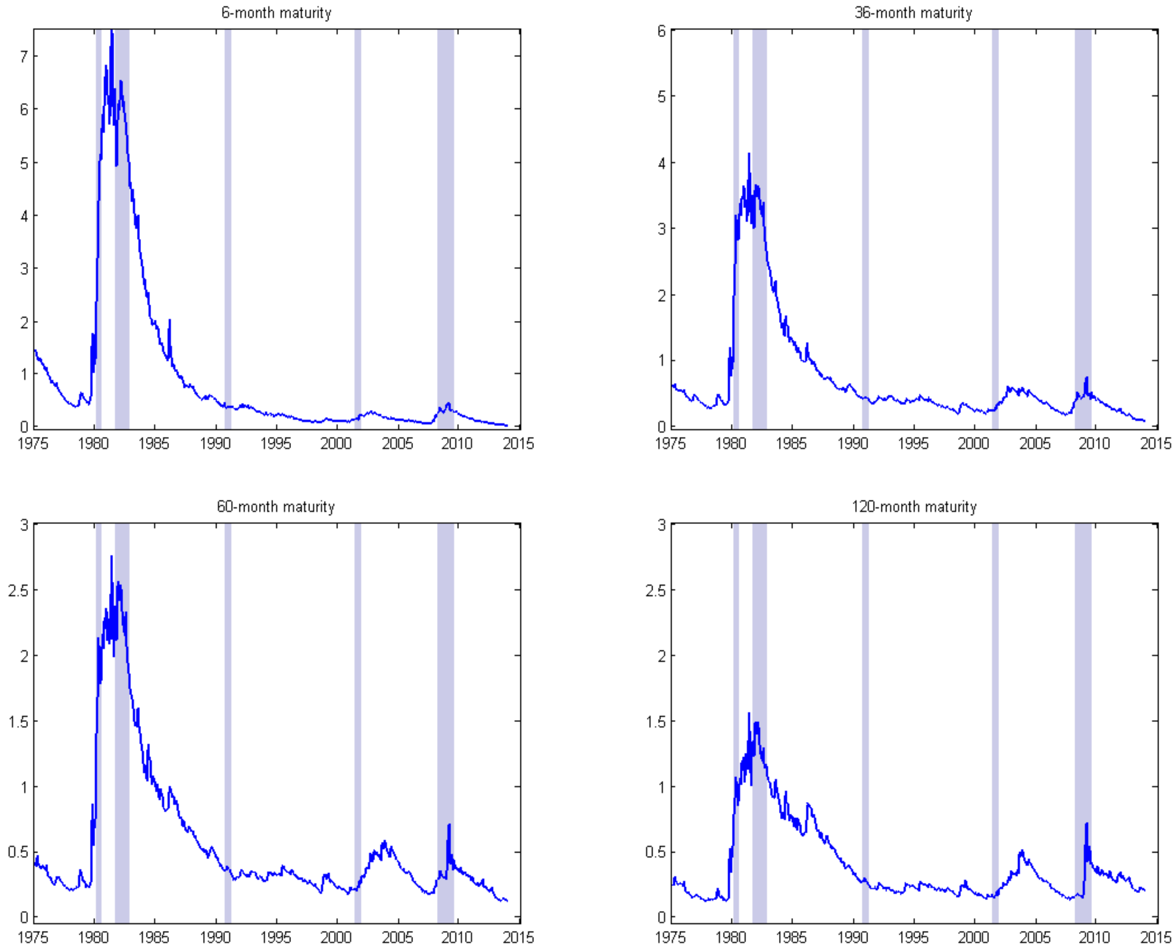
Figure 5: Cumulative Sum of Predictive Log-Likelihood of 3-, 12-, 60- or 120-Month Maturities



Notes: These are 1-month ahead cumulative sum of predictive log-likelihood for predicted yields from early 1975 to late 2013. From top left clockwise we have maturities of 3, 12, 60 and 120 months. The models present here are DMA (solid), DMS (dotted) and Diebold-Li (dashed). A higher log-likelihood implies improved density predictability.

time-varying second moments of 3 months ahead forecasts from the NS-DMA model. The figure displays a distinct time variation feature in the volatility evolution. The stable declining path of the volatility before the financial crisis matches the conclusions of Bianchi, Mumtaz and Surico (2009), in which they regarded the observation as the 'Great Moderation' of term structure. We can observe that the yields with longer maturities have lower volatilities. This feature is not intuitive. Theoretically, the long yield movements are mainly driven by three components: the expected future (real) short yields, future inflation expectations and the term premia. On the one hand, inflation expectations, depending on the current state of economic information, can change flexibly and frequently in a short time, so is the expected future short yields. On the other hand, term premia is also very volatile. Therefore, summing up the movements of these three components, the variance of long yields should be larger than the short yields; but the empirical result implies the opposite. As indicated in Duffee (2011*b*), the reason causing this result is that the factor driving up the expected future short yields or inflation expectations may drive down the term premia, so offsetting the variations of these components.

Figure 6: Time-Varying Second Moment



Notes: These are time-varying second moments of 3 months ahead forecasts for bonds at maturities 6, 36, 60 and 120 months, from early 1975 to late 2013. The variance of NS factors is estimated from Eq. (A.2), and then the variances of yield forecasts generated by each candidate model in the NS-DMA, can be easily calculated as linear combinations of factor variances.

From the perspective of time dimension, the volatilities of yields (especially shorter-term) are high in the 1980s, while the bond yield level is also relatively high. The high volatiities are due to large forecast variances of forecast models as well as a high degree of forecast dispersion in forecasts. It is clear that the volatilities are declining during the Great Moderation, and therefore the variances of bond forecasts are rather small between 1990 and 2007, except during the 2004-05 episode of 'Greenspan's Conundrum'. In around 2009, the volatilities surge to a high level since the 1990's, although the short yields stay at a relatively low level (restricted by zero lower bound) among all periods. Even after the financial crisis, ambiguity in yield forecasts still exists as the volatilities remain at a relatively high level.

3.3.3 Robustness: Do We Need Strict Arbitrage-Free Restrictions?

As we have discussed in Section 2, we impose NS restrictions on the pricing dynamics and leave the physical dynamics unconstrained. By allowing for parameter and model uncertainty in the physical dynamics, we are able to acquire significant predictive gains. The sources of these gains are also revealed in the last section.

Our NS-DMA approach does not explicitly impose 'hard' arbitrage-free restrictions.[14] The reason is that our focus here is not on the dynamic structure of market price of risks, as Duffee (2014) indicates that the no-arbitrage restrictions are unimportant if a model aims to pin down physical dynamics but not equivalent-martingale dynamics that specify the pricing of risk. In order to capture robust expectations of investors, we aim to improve forecasts of the interest rate term structure, and Joslin, Singleton and Zhu (2011) show that no-arbitrage cross-sectional restrictions are irrelevant to out-of-sample forecasts if the factor dynamics are unrestricted.[15]

To ensure the robustness of our NS-DMA approach, we extend the three-factor arbitrage-free Nelson-Siegel model proposed by Christensen, Diebold and Rudebusch (2011) and evaluate the forecast performance of the arbitrage-free version of NS-DMA.[16] The forecast performances

[14]From a theoretical perspective, Filipović (1999) and Björk and Christensen (1999) show that the Nelson-Siegel family does not impose the restrictions necessary to eliminate opportunities for riskless arbitrage. From a practical perspective, our implementation allows all bond yields to be priced with errors, which naturally breaks their original assumptions of the Nelson-Siegel family in their papers. Therefore, the potential loss of not imposing arbitrage-free restrictions may be mitigated.

[15]In practice, the arbitrage-free restrictions are not important in terms of forecasting in models assuming bond yields are priced with errors, see for example, Coroneo, Nyholm and Vidova-Koleva (2011) and Carriero and Giacomini (2011).

[16]The key difference between arbitrage-free NS-DMA and NS-DMA is an 'yield-adjustment term', which only

of two models are very close, which implies that the NS-DMA is almost arbitrage-free, which is consistent with theoretical evidence in Feunou, Fontaine and Le (2014) and Krippner (2015) that the NS models are near arbitrage-free. Hence, following Duffee (2014), we choose not to impose arbitrage-free restrictions to avoid potential misspecification.

4 Dynamics of Term Premia

4.1 Expectation Hypothesis and Term Premium

Within our empirical framework we shall set out the formal modeling of the term premia, which has been used to explain the failure of the Expectations Hypothesis and provides important information for the conduct of monetary policy, see Gürkaynak and Wright (2012).[17]

Based on the weak form of the Expectation Hypothesis, the long-term yield is average of expected future short term rates $y_t(\tau)^{EH}$ plus a constant Risk Premium, $constant^{EH}$:

$$y_t(\tau) = y_t(\tau)^{EH} + constant^{EH}, \tag{4.1}$$

where the Expectations Hypothesis (EH) consistent bond yield $y_t(\tau)^{EH}$ is given by:[18]

$$y_t(\tau)^{EH} = \frac{1}{\tau}\sum_{i=0}^{\tau-1} E_t y_{t+i}(1), \tag{4.2}$$

where $y_t(\tau)$ is the yield at time t for a bond of τ-period maturity. That is to say, the EH consistent long yield is equal to the average of expected short yields $E_t y_{t+i}(1)$.

The Expectation Hypothesis is closely related to the concept of *excess holding period return.*

depends on the maturity and factor volatility. See Christensen, Diebold and Rudebusch (2011) and Diebold and Rudebusch (2013) for more details.

[17]A simple approach to modeling the term structure is the Expectations Hypothesis (EH) that expected future short rates explain long rates. Campbell and Shiller (1991) indicates the empirical evidence fails to justify the strong form of Expectations Hypothesis and the idea that long-term interest rate are simply determined by the average of current and future expected short-term rates. However, EH could be resuscitated in weak form allowing for a constant Term Premia, consistent with an upward sloping yield curve. But risk, and hence the Term Premia, is unlikely to be constant while underlying variables are changing.

[18]The expectation here is under the physical measure. Our model can potentially be made arbitrage free, but this is beyond the scope of this paper. For further discussion see Joslin, Priebsch and Singleton (2014)

First, we define the *holding period return* as the return on buying an τ-year zero coupon bond at time t and then selling it, as an $(\tau - m)$-year zero coupon bond, at time $t + m$. This holding period return is given by:

$$HPR_{t+m}(\tau, m) = \frac{1}{m}[p_{t+m}(\tau - m) - p_t(\tau)] \tag{4.3}$$

where $p_t(\tau)$ is the log price of τ-year zero coupon bond at time t and $p_{t+m}(\tau - m)$ is the log price of $(\tau - m)$-year zero coupon bond at time $t + m$. The difference between holding period return and the m-year continuously compounded short yield is the *excess holding period return*:

$$EXR_{t+m}(\tau, m) = HPR_{t+m}(\tau, m) - y_t(m). \tag{4.4}$$

If the weak form of the Expectation Hypothesis holds, then with some simple algebra, we can prove that the expected excess holding period returns are the constant Risk Premia. In other words, we should not be able to predict the excess returns in the future. However, Cochrane and Piazzesi (2009) construct a test by regressing the excess bond returns on the forward rates, and show that the forward rates have significant predictive power. The Expectation Hypothesis is therefore rejected, implying that the term premium should be time-varying.[19] The time-varying term premium is therefore,

$$TP_t(\tau) = y_t(\tau) - y_t(\tau)^{EH}. \tag{4.5}$$

Alternatively, we can rewrite Eq. (4.5) by relating the term premium to the *excess holding period return*:

$$TP_t(\tau) = \frac{1}{\tau}E_t\Big(\sum_{i=0}^{\tau-2} EXR_{t+1+i}(\tau, 1)\Big). \tag{4.6}$$

By the linearity of expectation, we can write the 1-period ahead expected *excess holding*

[19] Similar evidence can be found in Duffee (2002), Cochrane and Piazzesi (2005), Sarno, Thornton and Valente (2007), Tang and Xia (2007) and Gürkaynak and Wright (2012).

period return as

$$E_t\Big(EXR_{t+1}(\tau,1)\Big) = -(\tau-1)E_t\ TP_{t+1}(\tau-1) + \tau\ TP_t(\tau). \tag{4.7}$$

Therefore, under the weak form of the Expectation Hypothesis, the expected *excess holding period return* is a constant as we mentioned:

$$E_t\Big(EXR^{EH}_{t+1}(\tau,1)\Big) = constant^{EH}. \tag{4.8}$$

In contrast, if the term premia is time-varying, then the predictability of *excess holding period return* stems from an element x^p_t that is orthognal to the EH term premia $constant^{EH}$,

$$E_t\Big(EXR_{t+1}(\tau,1)\Big) = constant^{EH} + x^p_t. \tag{4.9}$$

If we have a model that can generate time-varying term premia, then it is straightforward to obtain x^p_t. We can simply use the results from Eq. (4.7) and subtract the expected *excess holding period returns* from the mean. In the next section we use our estimates of the term premia to model excess returns in the bond market.

4.2 Predictability of Excess Returns

The term premia is closely related to the real economy. The behavior of investors is influenced by their future expectations, which can be reflected in the term premia. Harvey (1988) indicates that agents tend to buy long-term assets in 'good' times in order to smooth their consumption during the 'bad' time, and hence the long yields decline causing the negative term premia. It is also noted in Kim (2009), risk-aversion could vary with the business cycle. Close to recessions, agents tend to consider bonds an 'insurance' to maintain consumption levels in the downturn. During the recession periods, the agents may have different expectations of the future, so they may reallocate the assets frequently, and hence premia level is affected by the behavior of the agents.

Therefore, the decisions of the investors are governed by the expectations of future returns. Particularly, an investor can select different strategies according to different forecasts, to maximize expected excess returns. Based on a weak form of the Expectation Hypothesis, the term premia should be a constant; therefore an investor who favors the Expectation Hypothesis, chooses the historical average of excess holding period returns of bonds as the future forecasts.[20] In contrast, an investor makes different decisions in portfolio formation based upon alternative Term Premia estimates.

Formally, we assume an investor forms the expectations of h-period (month) ahead excess returns using the following model:

$$r_{t+h} = \alpha_m + \beta_m x_{t,m} + \epsilon_{t+h}, \tag{4.10}$$

where r_{t+h} is the excess bond return after h months, $x_{t,m}$ is the independent variable, ϵ_{t+h} is the error term and α_m and β_m are parameters. Following Cochrane and Piazzesi (2005) and Duffee (2011*a*), we calculate 3-, 6-, 9- and 12-month excess holding period returns of 2- to 10-year bonds at time t as r_t using Eq. (4.4).

For the Expectation Hypothesis investor, the model of h-period ahead forecasts is a restricted version ($\beta = 0$)

$$\bar{r}_{t+h} = \hat{\alpha}_m, \tag{4.11}$$

where $\hat{\alpha}_m$ is the historical mean of excess bond returns up to and including time t.

For the investor using another forecasting model j, the forecast of future excess returns is

$$\hat{r}_{j,t+h} = \hat{\alpha}_m + \hat{\beta}_m x_{t,m}, \tag{4.12}$$

where $\hat{\alpha}_m$ and $\hat{\beta}_m$ are the ordinary least squares (OLS) estimates of α_m and β_m in Eq. (4.10). We generate h-month-ahead out-of-sample forecasts of excess bond returns using a recursive

[20] The construction is similar to Thornton and Valente (2012) and Zhou and Zhu (2014).

expanding estimation method. As we mentioned in the last section, we can calculate the predictable element for each model by regressing the expected *excess holding period return* on a constant using Eq. (4.9).[21] The errors of the regressions are used as independent variables $x_{t,m}$, which can evaluate the predictability of different models and the robustness of Term Prmia estimates.

To compare the predictive performance of our forecasting models against the EH benchmark, we have two groups of candidates. The first group includes the predictable elements estimated from the smoothed NS-DMA (denoted NS-DMA*) conditional on the information of the whole sample, as well as the models proposed by Bauer, Rudebusch and Wu (2014) and Wright (2011).[22] Specifically, we estimate the NS-DMA* by conducting backward smoothing conditional on the information of the whole sample.[23] That is, we consider in-sample forecasts of excess returns $\hat{r}_{j,t+h}$ where $j = 1, 2, 3$, based upon NS-DMA*, BRW and Wright. We then consider the out-of-sample group, which includes the predictable elements estimated from the NS-DMA and DL (Diebold and Li (2006)), and the recursively estimated forward rate factor proposed by Cochrane and Piazzesi (2005) (denoted CP);[24] in this group we only use the information up to and including time t to obtain the variables, so it is a true out-of-sample forecasting exercise. After obtaining the predictable elements estimated from the candidate term structure models, each time we use one model-implied predictable element as the independent variable when forecasting the excess returns. Therefore, in addition to the EH forecasts $\bar{r}_{t+h}$, we have in total five kinds of excess return forecasts denoted as NS-DMA*, BRW, Wright, NS-DMA, DL and CP.

If the Expectation Hypothesis holds, the implied term premia and the future excess returns

[21]In our recursive estimation, the implementation is equivalent to subtracting the calculated expected *excess holding period return* from its historical mean.

[22]We use the whole sample to estimate the implied term premia of Bauer, Rudebusch and Wu (2014) (BRW) and Wright (2011) because the maximum likelihood estimation of these two models may fail to converge with subsamples. This is an in-sample forecasting as an extra layer of forward-looking information is introduced when estimating the parameters. This may bring about more significant performance but may not correctly reflect the true conditional expectations of investors as the information of realized expectations is contained.

[23]We use Rauch-Tung-Striebel (RTS) smoother, see Appendix A.3.

[24]Cochrane and Piazzesi (2005) show that the CP factor has significant predictive performance for the 1-year holding period excess returns. The CP factor in our implementation is recursively constructed using one-year yield and 2- to 10-year forward rates.

should be constant. Hence the forecasting models based on Eq. (4.12) should not have higher predictive power than the Expectation Hypothesis model Eq. (4.8). Otherwise, the expectation of excess bond returns is time-varying, and the model-implied term premia at time t may provide useful information in forecasting the excess bond returns at time $t+1$ and forward.

We evaluate the forecasting performance of the models using *out-of-sample* R^2, R^2_{OS}, proposed by Campbell and Thompson (2008),

$$R^2_{OS} = 1 - \frac{\sum_{t=1}^{T}(r_t - \hat{r}_{j,t})^2}{\sum_{t=1}^{T}(r_t - \bar{r}_t)^2} \tag{4.13}$$

where $\hat{r}_{j,t}$ is the fitted value from a predictive regression model j estimated through period $t-h$, and $\bar{r}_t$ is the historical average return estimated through period $t-h$. If the R^2_{OS} is positive, then the predictive regression has lower average mean-squared prediction error than the historical average return.[25] Formally, we test the null hypothesis that $R^2_{OS} \leq 0$ against the alternative hypothesis that $R^2_{OS} > 0$. We employ the statistic developed by Clark and West (2007) to evaluate the significance of the out-of-sample forecasts. Clark and West (2007) adjust the statistic of Diebold and Mariano (1995), as the previous version has a nonstandard distribution when comparing forecasts from nested models. When setting $\beta_{t,m}$ in Eq. (4.12) to zero we have the historical mean model, so using the *MSPE-adjusted* statistic of Clark and West (2007) is more appropriate here. To sum up, in this section we mainly discuss the statistical evaluation, and we will proceed with the discussion about economic evaluation in the next section.

4.3 Economic Value

The above evaluation of out-of-sample predictability does not consider the risk borne by an investor. It raises the issue of economic value of a forecasting model, as statistical significance does

[25] Clark and West (2006) indicate that the expected out-of-sample R^2 under the null of unpredictability is negative for series that are truly unpredictable, because in a finite sample the predictive regression will on average have a higher mean squared prediction error as it must estimate an additional coefficient. In contrast, the positive out-of-sample R^2 can be interpreted as evidence for predictability.

not measure its economic significance. This section evaluates whether the model predictability is sufficiently large to be of economic value to risk averse investors. Following Campbell and Thompson (2008), Welch and Goyal (2008), and Rapach, Strauss and Zhou (2009), we assume each investor, who is small and hence with no market impact, chooses portfolio weights based on the return forecasts. In this paper, we assume the investor only has two assets for selection: the short-term (1-year) and long-term (2 to 10 years) bonds. We then calculate realized utility gains for a mean-variance investor on a real-time basis.

To demonstrate the evaluation of the above strategies, we firstly discuss the case of an Expectation Hypothesis (EH) investor. We can compute the average utility for the mean-variance investor with relative risk aversion parameter γ_R who allocates his or her portfolio monthly between the short-term and long-term bonds using forecasts of the excess returns based on the historical average. This exercise requires the investor to forecast the variance of excess returns. Following Campbell and Thompson (2008), we assume that the investor estimates the variance $\hat{\sigma}^2_{t+1}$ using a 5-year rolling window using monthly data of excess annually returns. A mean-variance investor who forecasts the excess bond returns using the historical average $\bar{r}_{t+1}$ will decide at the end of period t to allocate the following share of his or her portfolio to bonds in period $t+1$:

$$w_{0,t} = (\frac{1}{\gamma_R})(\frac{\bar{r}_{t+1}}{\hat{\sigma}^2_{t+1}}) \tag{4.14}$$

where $\hat{\sigma}^2_{t+1}$ is the 5-year rolling-window estimate of the variance of excess returns.[26]

Over the out-of-sample period, the average of the realized utility of the investor is given by

$$\hat{v}_0 = \hat{\mu}_0 - (\frac{1}{2})\gamma_R \hat{\sigma}^2_0 \tag{4.15}$$

where $\hat{\mu}_0$ and $\hat{\sigma}^2_0$ are respectively the sample mean and variance of the excess holding period

[26]Following Campbell and Thompson (2008), Rapach, Strauss and Zhou (2009) and Thornton and Valente (2012), we constrain the portfolio weight on bonds to lie between -100% and 200% each month, so in Eq. (4.14) $w_{0,t} = -1$ ($w_{0,t} = 2$) if $w_{0,t} < -1$ ($w_{0,t} > 2$).

returns on the benchmark portfolio of the EH investor, which is constructed using forecasts of the excess returns based on the historical average.

Similarly, we can calculate the average utility for the same investor, when his or her decision is made by using a model to forecast the excess bond returns. The share chosen by the investor is

$$w_{j,t} = (\frac{1}{\gamma_R})(\frac{\hat{r}_{j,t+1}}{\hat{\sigma}^2_{t+1}}) \tag{4.16}$$

where $\hat{r}_{j,t+1}$ is the excess return forecast from model j. The resulting realized average utility level is

$$\hat{v}_j = \hat{\mu}_j - (\frac{1}{2})\gamma_R \hat{\sigma}^2_j \tag{4.17}$$

where $\hat{\mu}_j$ and $\hat{\sigma}^2_j$ are the sample mean and variance of the excess holding period returns on the portfolio indexed by j. The investor forms the portfolio j using forecasts of the excess returns of bonds according to the jth forecasting model.

We can compute the utility gain, or certainty equivalent return, as the difference between $\hat{v}_j$ in Eq. (4.17) and $\hat{v}_0$ Eq. (4.15)

$$\Delta = \hat{v}_j - \hat{v}_0. \tag{4.18}$$

The utility gain that is expressed in average annualized percentage return, can be interpreted as the portfolio management fee that an investor would be willing to pay to have access to the additional information available in a predictive model relative to the information in the historical term premia alone. We report results for risk aversion parameters $\gamma_R = 3$ and $\gamma_R = 6$; the results are qualitatively similar for other reasonably values (ranging from 1 to 10).

4.4 In-Sample and Out-of-Sample Performance

In Table 6, we report the in-sample and out-of-sample performance of our excess return forecasts and the utility gain Δ for the 5-year bonds. The statistical and economic evaluations for the in-sample group across maturities (2 to 10 years) are summarized in Figures 7 and 8, and where the evaluations for the out-of-sample group are displayed in Figure 9 and 10.

In Table 6, we find that across all forecast horizons NS-DMA* produce higher out-of-sample explanatory power and economic value than the other models. The R^2_{OS} of NS-DMA* ranges from 17% to 57% across four forecast horizons, and both the R^2_{OS} and economic value increase with forecast horizons. The economic value reaches 0.83% for the 5-year bond when $\gamma_R = 6$. In general, the NS-DMA*, BRW and Wright have relatively higher predictive performance, which makes sense as these three models are estimated with the whole sample; the information of realized excess returns is included in their term premia estimates.[27] In other words, the model-implied term premia at time t estimated from these three models can potentially be distorted by the realized information at time time $t + 1$ and forward, so the estimates may not fully reflect the current expectations of agents at each point in time.

Conversely, the NS-DMA does not have the distortions as it does not include any information in the future. Surprisingly, the NS-DMA with information up to time t, also has significant forecasting performance, and economic value quantitatively similar to, or even higher than, BRW and Wright. When the NS-DMA is compared with the CP factor or DL, the advantage is more distinct as the out-of-sample performance of the CP or DL is even worse than the EH benchmark. The results distinguish the robustness of NS-DMA* and NS-DMA in revealing the term structure dynamics. Moreover, the excess return forecasts of NS-DMA significantly (at 1 % for 9- and 12-month forecasting horizons) outperform the Expectation Hypothesis model, the economic value remains positive for all forecast horizons. The statistical significance implies that the Expectation Hypothesis does not hold, which has been well indicated in the previous

[27] In fact, it is a 'pseudo' forecasting exercise to generate forecasts using these three models, as the full-sample estimation of these models introduces an extra layer of forward-looking information. The reason we do this is to provide benchmarks to evaluate the in-sample performance of the NS-DMA*, the out-of-sample performance and the NS-DMA, and the economic significance of both models.

Table 6: **Predictive Results of 5-Year Bond Excess Returns**

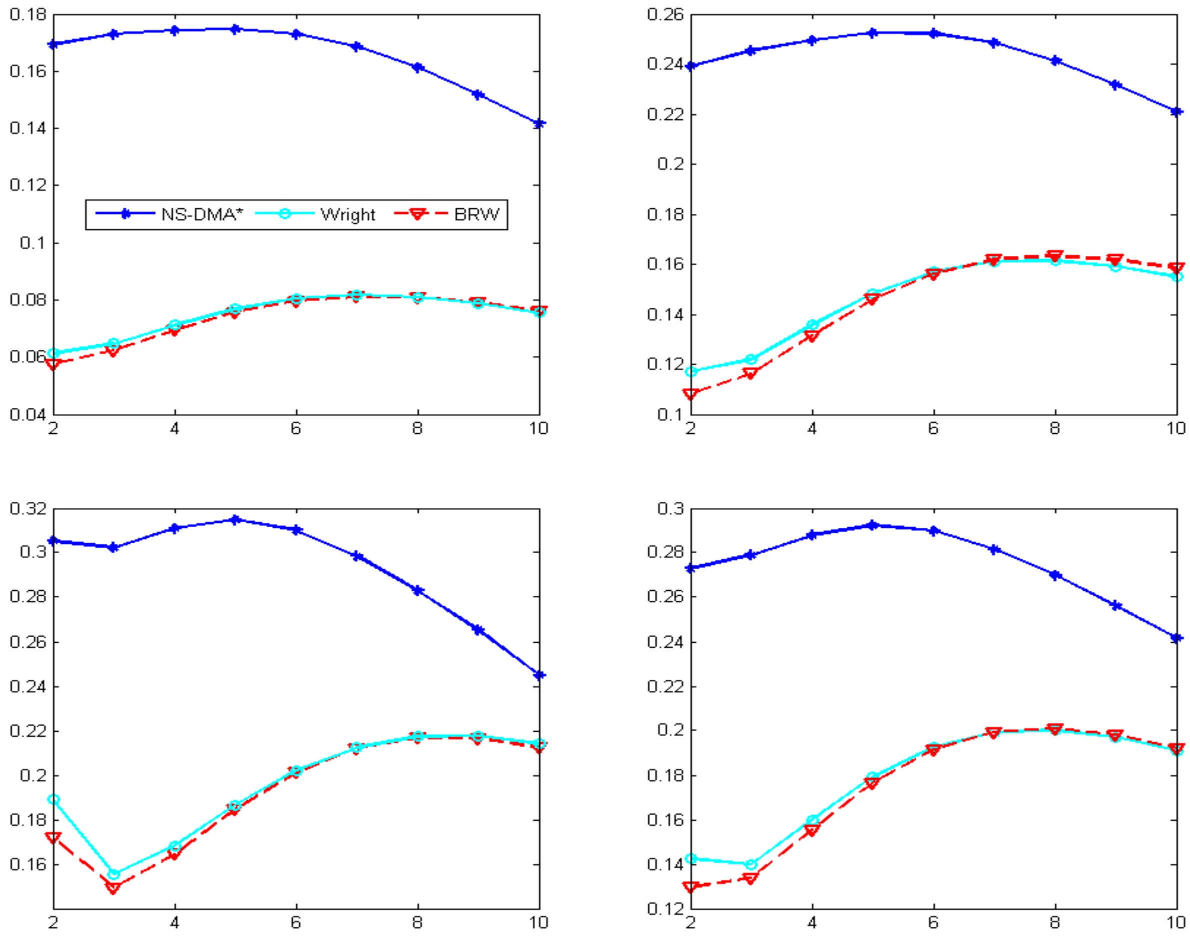
		$h = 12$			$h = 9$		
		R^2_{OS}	Δ ($\gamma_R = 3$)	Δ ($\gamma_R = 6$)	R^2_{OS}	Δ ($\gamma_R = 3$)	Δ ($\gamma_R = 6$)
In	NS-DMA*	31.49***	0.75	0.83	29.23***	0.54	0.56
Sample	Wright	18.66***	0.21	0.24	17.91***	0.12	0.14
	BRW	18.46***	0.14	0.16	17.66***	0.08	0.10
Out of	NS-DMA	4.70***	0.15	0.15	5.64***	0.09	0.08
Sample	DL	-3.44	–	–	-1.64	–	–
	CP	-28.35	–	–	-21.76	–	–
		$h = 6$			$h = 3$		
		R^2_{OS}	Δ ($\gamma_R = 3$)	Δ ($\gamma_R = 6$)	R^2_{OS}	Δ ($\gamma_R = 3$)	Δ ($\gamma_R = 6$)
In	NS-DMA*	25.24***	0.31	0.30	17.46***	0.20	0.21
Sample	Wright	14.83***	0.07	0.07	7.68***	0.04	0.03
	BRW	14.59***	0.05	0.05	7.57***	0.04	0.03
Out of	NS-DMA	3.58**	0.03	0.02	0.95*	0.02	0.01
Sample	DL	-1.33	–	–	0.39	–	–
	CP	-10.81	–	–	-13.12	–	–

Notes: 1. The table reports the out-of-sample R^2 statistics (unit %) for log excess bond returns (holding 3, 6, 9 and 12 months) on the 5-year long-term Treasury bonds over the forecast evaluation period from 1983:12 to 2013:11. The forecasting horizons (holding periods) are $h = 3, 6, 9, 12$ months, respectively.
2. Utility gain (Δ) is the portfolio management fee (in annualized percentage return) that an investor with mean-variance preferences would be willing to pay to have access to the forecasting model. The utility gain is computed at two risk aversion levels, i.e., $\gamma_R = 3, 6$. Higher utility gain is preferred.
3. Statistical significance for the R^2 statistic is based on the p-value for the Clark and West (2007) out-of-sample MSPE-adjusted statistic; the statistic corresponds to the test of the null hypothesis that the competing model has equal expected square prediction error relative to the benchmark forecasting model (historical mean) against the alternative hypothesis that the competing forecasting model has a lower expected square prediction error than the benchmark forecasting model. *, ** and *** indicate significance at the 10%, 5% and 1% levels, respectively.
4. The in-sample group includes NS-DMA*, Wright and BRW, and the full-sample information is used to estimate predictable elements in Eq. (4.9). The out-of-sample group includes NS-DMA, DL and CP, and the information up to and including time t are used for estimation. CP factor is recursively constructed using one-year yield and 2- to 10-year forward rates.

literature. Although the NS-DMA has both statistical power and economic value, the utility gain from the predictability is not sufficiently large, especially for shorter forecast horizons. The modest finding in economic value is consistent with Della Corte, Sarno and Thornton (2008). For bond investors and practitioners, the EH still plays an important role in out-of-sample forecasts of interest rate term structure, especially at short forecast horizons. However, the EH cannot fully reflect the real-time expectations of agents when facing economic uncertainty despite its conveniency.

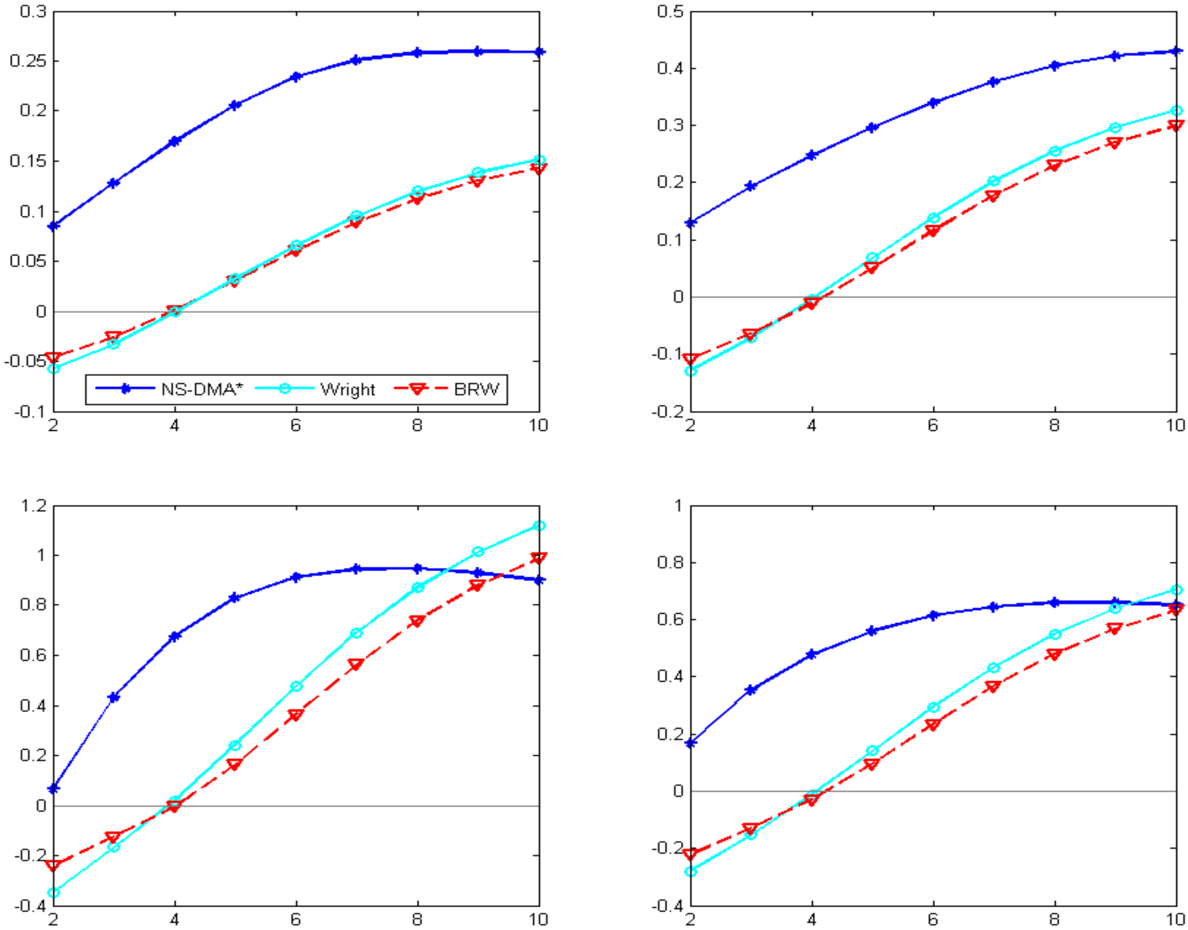
To further elaborate on our results, Figures 7 and 8 provide a general summary for the excess return forecasts of the in-sample group, across 9 bond maturities (from 2 to 10 years). The NS-DMA* consistently outperforms the Wright and BRW models across maturities and forecast horizons, both statistically and economically. The results suggest more robust full-sample estimates of the term premia using the NS-DMA* model.

Figure 7: Statistical Evaluation



Notes: These are the out-of-sample R^2 statistics for bonds at maturities from 2 to 10 years. The predictive duration is from 1983:10 to 2013:11. From top left clockwise we have forecast horizons/holding periods 3, 6, 9 and 12 months. The models present here are NS-DMA*, Wright and BRW. The statistics are consistently the highest for the NS-DMA*.

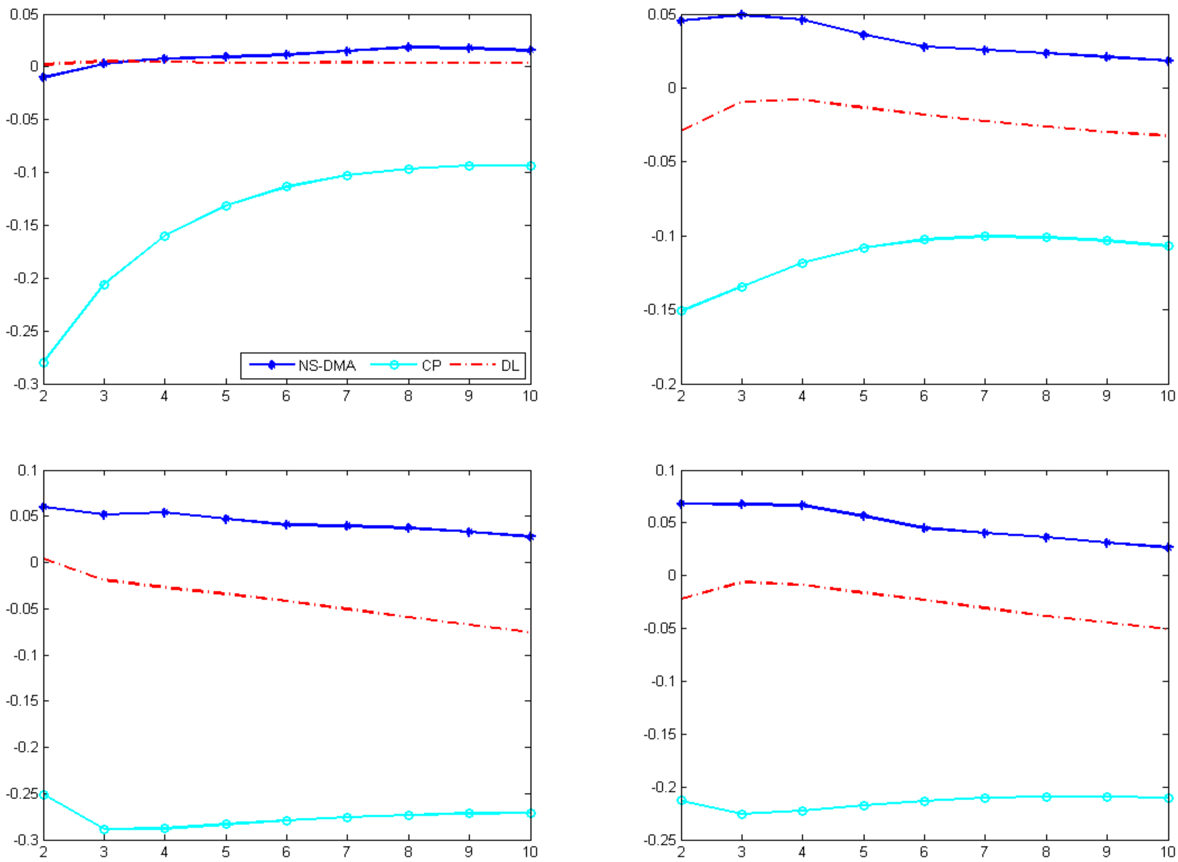
Figure 8: Economic Evaluation



Notes: These are the utility gain statistics Δ (unit %) at risk averse level $\gamma_R = 6$. The investors can choose one long-term bond as the risky asset from 9 maturities (2 to 10 years). The predictive duration is from 1983:10 to 2013:11. From top left clockwise we have forecast horizons/holding periods 3, 6, 9 and 12 months. The models present here are NS-DMA*, Wright and BRW. The NS-DMA typically has the highest Δ.

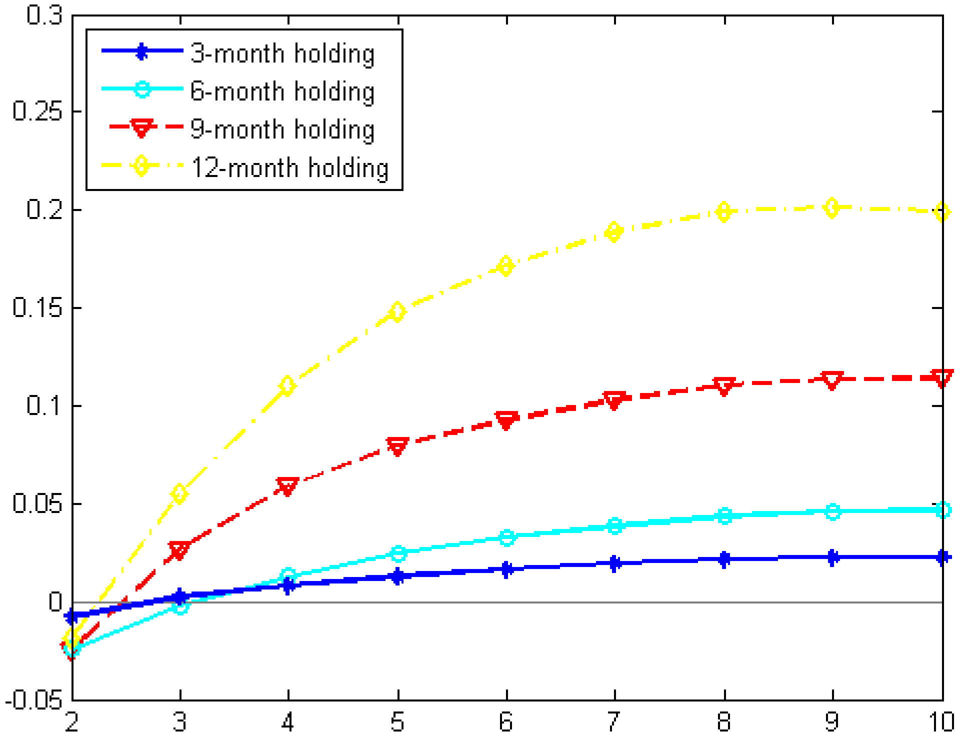
Figure 9 and 10 summarize the statistical analysis and economic evaluation for the excess return forecasts of the out-of-sample group, across 9 bond maturities (from 2 to 10 years). In seems the recursively constructed CP factor has no predictability gains against the EH benchmark. The NS-DMA has significantly positive out-of-sample R^2, especially for longer holding periods. The utility gain from the NS-DMA is not very significant when compared to the full-sample estimates from NS-DMA*, because the realization of agents' expectations may be contaminated by the market disturbances. Although the NS-DMA may not provide large economic gains for our constructed portfolio, *it is useful in revealing the agents' expectations in* *real time*. The NS-DMA is more reliable than the unrealistic EH or other full-sample estimates, as full-sample estimation unavoidably includes the information of realized expectations. In particular, NS-DMA allows for parameter and model uncertainty and hence is robust to learning and structural breaks, see Piazzesi and Schneider (2007) and Gürkaynak and Wright (2012). Therefore, our adaptive term structure model can provide plausible estimates in reflecting changes in investors' conditional expectations concerning the future path of monetary policy as well as the risk compensation the investors require.

Figure 9: Statistical Evaluation



Notes: 1. This figure shows the out-of-sample R^2 statistics for bonds at maturities from 2 to 10 years. The three models present here are NS-DMA, DL and CP.
2. The predictive duration is from 1983:10 to 2013:11. From top left clockwise we have forecast horizons/holding periods 3, 6, 9 and 12 months.

Figure 10: Economic Evaluation



Notes: 1. This figure shows the utility gain statistics Δ (unit %) of the NS-DMA at risk averse level $\gamma_R = 6$. The investors can choose one long-term bond as the risky asset from 9 maturities (2 to 10 years). The economic value of the CP or DL is not calculated due to insignificant or negative out-of-sample R^2.
2. The predictive duration is from 1983:10 to 2013:11. We have forecast horizons/holding periods 3, 6, 9 and 12 months.

4.5 Model-Implied Term Premia

In this section we set out a visual comparison of our term premium estimates. We plot the NS-DMA time-varying Risk Premia from 1985 for a medium-term bond (maturity 36 months) and a long-term bond (maturity 120 months) in Figure 11.[28] For comparison, we also plot the model-implied term premia estimated from other approaches proposed by Kim and Wright (2005), Wright (2011) and Bauer, Rudebusch and Wu (2014).[29]

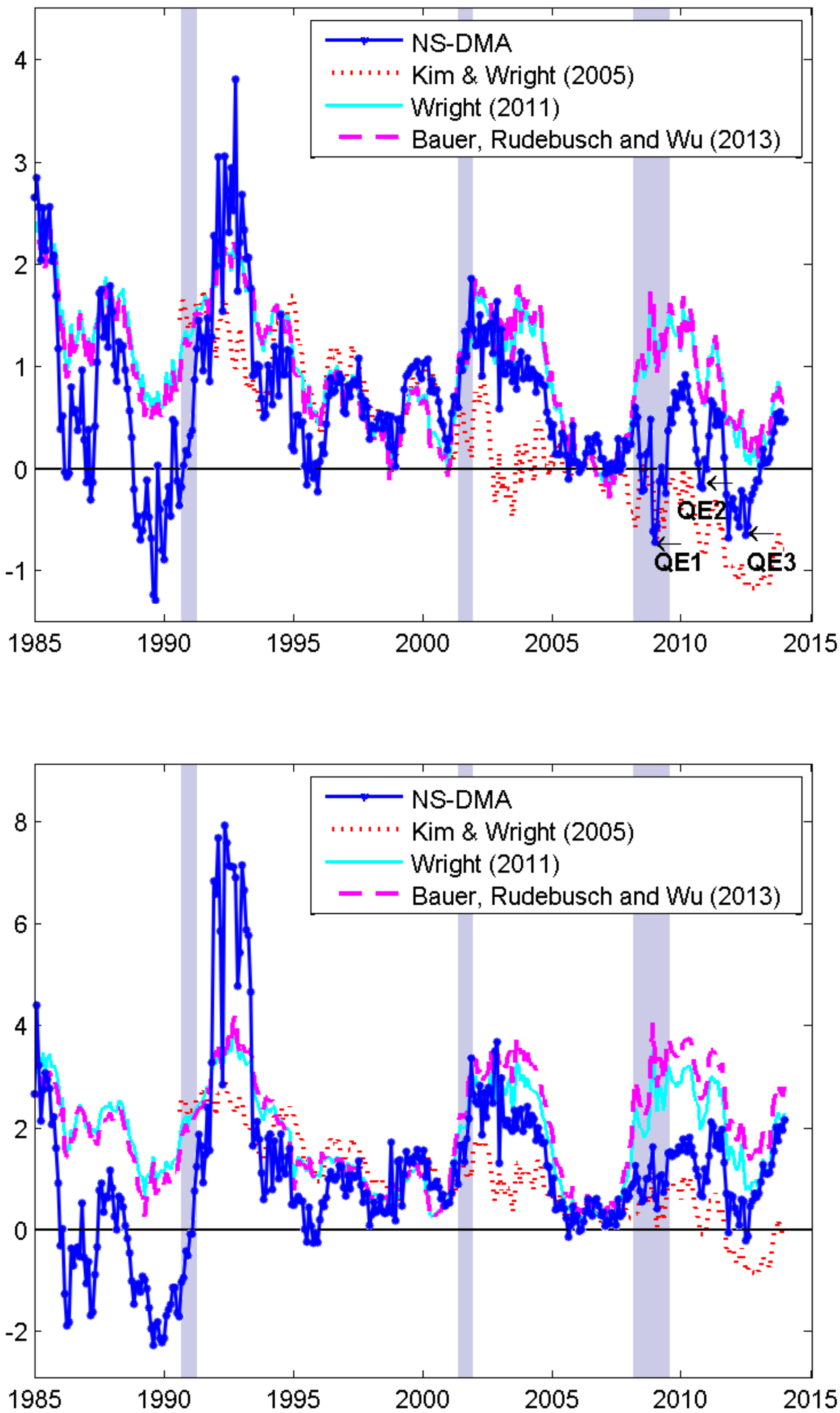
[28]The Risk Premia at other maturities show similar patterns because of their high correlations, but the results are not displayed here for sake of brevity.

[29]The comparison between the NS-DMA term premia and recursively estimated term premia from dynamic Nelson-Siegel is shown in Appendix D. The NS-DMA approach seems to be more robust than the constant-parameter dynamic Nelson-Siegel model, as the dynamic Nelson-Siegel model proposed by Diebold and Li (2006)

As it is shown in Figure 11, the NS-DMA seems to have captured the level and volatility of the Risk Premium. The estimates from NS-DMA have a consistent trend with the estimates of Kim and Wright (2005), Wright (2011) and Bauer, Rudebusch and Wu (2014), especially at the medium-term maturity, where the degree of term premia correlation between NS-DMA and Kim and Wright (2005) is 0.55 and the correlation between NS-DMA and Wright (2011) (or Bauer, Rudebusch and Wu (2014)) is more than 0.70. In general the term premia shows countercyclical pattern, as they rise in and around US recessions, except the estimates of Kim and Wright (2005). The difference between the estimates of Kim and Wright (2005) (KW) and other models is due to the estimated expectation of future short rate. As indicated in Christensen and Rudebusch (2012), there could be potential inaccuracy in the KW measure, because their factor dynamics tend to display much less persistence than the true process. According to the observations here, future short rates from KW would be expected to revert to their mean too quickly, and estimated risk-neutral rates would be too stable, so the KW term premia has a relatively lower variance and may display an acyclical pattern.

tends to overestimate the future short rates and hence underestimate the term premia.

Figure 11: Time-Varying Term Premia of 36-and 120-Month Bonds



Notes:
1. The top panel is the 36-month term premia and the bottom is the 120-month term premia. The EH consistent 36- and 120-month and bond yields are estimated using Eq. (4.2); we then calculate the term premia using Eq. (4.5).
2. In addition to NS-DMA, we use the whole sample to separately estimate two types of term premia employing the methods proposed by Wright (2011) and Bauer, Rudebusch and Wu (2014). The Kim and Wright (2005) term premia can be obtained from the Federal Reserve Board website.
3. Shaded areas are recession periods based on the NBER Recession Indicators. The unit is percentage.

Among all measures considered, the NS-DMA term premia seems to be more sensitive to changes in the economic environment, which can be seen more clearly from the lower panel in Figure 11 of the long-term term premia. The reason is that expectations of the future short rates and, hence, the term premia can change severely. Our empirical evidence shows that the NS-DMA has good performance in forecasting the future short rates, by utilizing a time-varying approach and appropriately including the information of macro-finance variables. For example, the short rate was continuously decreasing from 1990 to 1993 so the expectation of future short rates were also deceasing. Long rates were relatively stable in contrast, which leads to the increasing Risk Premia that peaked in 1993.

Specifically, our measures seem to capture the 'Greenspan's Conundrum' in 2005,[30] as the premia level fell substantially. The effects of three rounds of QE in recent years are also captured. The top panel in Figure 11 shows that the QE significantly increases the premia level, as the expected future short rates fall more sharply than the long rates. Between 2012 and early 2013, recession risk existed due to a fear of the rise in future short rates, which is consistent with the low level of premia; it explains why QE was launched in that period. Towards the end of 2013, the term premia was positive, consistent with the Fed tapering QE. Note that the effects of QE for the 3-year bond is more significant than the 10-year bond, because investors' expectations of short rates for the long run tend to be relatively stable and usually higher than 3%, according to the Blue Chip Financial Forecast survey data.[31] Accurately estimating term premia can provide valuable information for facilitating a prudent monetary policy, and NS-DMA estimates of the term premium are quite promising in serving this objective.

Lastly, we can observe that a divergence between the estimates of NS-DMA and other estimates from Wright (2011) and Bauer, Rudebusch and Wu (2014), lies in the financial crisis period. Christensen, Lopez and Rudebusch (2010) indicate that during the financial crisis,

[30]Federal Reserve Chairman Alan Greenspan observed that long-term yields had trended lower despite the fact that the Federal Open Market Committee's target for the federal funds rate had risen. A variety of possible explanations were considered implausible and, hence, he called it a 'conundrum'.

[31]We thank Jonathan Wright for pointing this out and sharing the survey data.

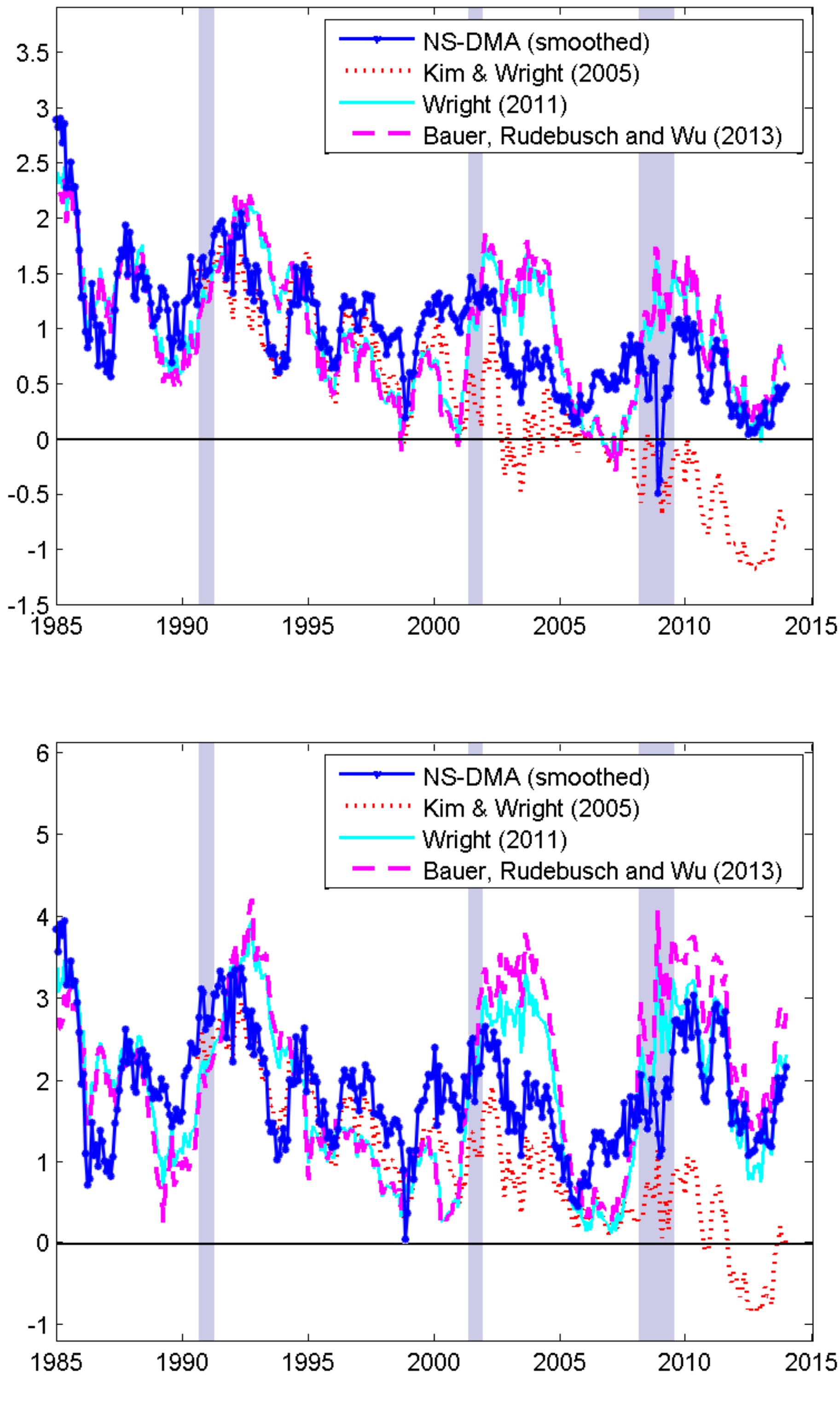
financial markets encountered intense selling pressure because of fears of credit and liquidity risks. The surge in risk aversion creates increased global demand for safe and highly liquid assets, for example, the nominal U.S. Treasury securities. This 'flight-to-quality' could lead to a sharp decline in their yields and therefore result in downward pressure on term premia. Bauer, Rudebusch and Wu (2014) argue, meanwhile, that the procyclical flight-to-quality pressure could not completely offset the usually countercyclical pattern of risk. Based on our estimates, we believe the flight-to-quality demand is evident and can suppress the countercyclical pattern. This makes a distinction between the financial crisis and the previous recessions, as global marketa are more unified than ever before and hence capital flows to a safe heaven.[32]

It is worth noting that the models of Wright (2011) and Bauer, Rudebusch and Wu (2014) are estimated with the whole sample of data, so the estimates of current term premia implicitly absorb the information from the future, which may be the potential reason for the divergence between NS-DMA and the two models. Therefore, to evaluate the robustness of the 'flight-to-quality' demand in financial crisis, we also use the full-sample estimates of the NS-DMA*.[33] The smoothed estimates are plotted in Figure 12; the smoothed Term Premium estimates of NS-DMA are less volatile and more consistent with the estimates of the other models, but the 'flight-to-quality' demand is still obvious as shown in the top panel.

[32]The countercyclical patterns of term premia before recessions have been researched in previous literature, such as Estrella and Mishkin (1998), Wright (2006), Kim (2009) and Wheelock and Wohar (2009), but the behavior during recession of term premia is not thoroughly discussed. D'Agostino, Giannone and Surico (2006) suggest that the term spread may become a weaker indicator of the real economy after the Great Moderation, which potentially supports our conclusion that the 'flight-to-quality' demand can suppress the countercyclical patterns of term premia.

[33]See Appendix A.3, the estimates of NS-DMA in Figure 11 reflect the expectations in real-time while the NS-DMA* estimates from the Rauch-Tung-Striebel (RTS) smoother contain the information of realized expectations.

Figure 12: Time-Varying Term Premia of 36-and 120-Month Bonds with Smoothed NS-DMA



Notes:
1. The top panel is the 36-month term premia and the bottom is the 120-month term premia. The EH consistent 36- and 120-month and bond yields are estimated using Eq. (4.2); we then calculate the term premia using Eq. (4.5).
2. We plot the NS-DMA term premia estimated from the RTS smoother conditional on the information of the whole sample. In addition to the smoothed estimates, we use the whole sample to separately estimate two types of term premia employing the methods proposed by Wright (2011) and Bauer, Rudebusch and Wu (2014). The Kim and Wright (2005) term premia can be obtained from the Federal Reserve Board website.
3. Shaded areas are recession periods based on the NBER Recession Indicators. The unit is percentage.

5 Conclusion

The Nelson-Siegel approach of yield curve modeling has been extended by Diebold and Li (2006), Diebold, Rudebusch and Aruoba (2006) and Bianchi, Mumtaz and Surico (2009). We further extend the literature using a Dynamic Model Averaging approach (NS-DMA), in order to characterize the nonlinear dynamics of yield factors, as Duffee (2002) suggests nonlinearity can potential improve yield forecasts. The framework we propose generalizes some frontier econometric techniques, and is augmented with many (unspanned) macro-finance factors as in Dewachter and Iania (2012). The NS-DMA method significantly improves the predictive accuracy and successfully identifies the dynamics of term premia, on grounds that it seems to have appropriately incorporated the information in the macro-economy. We then explore the predictive power of our term structure model regarding the future excess holding period returns. Our approach allows for potential structural breaks and model uncertainty, and hence, our real-time term premia forecasts are plausible and have both statistical power and economic value. According to the empirical results, we specifically discuss some informative responses of bond yields to monetary policy implementations in different periods, such as the Great Moderation and the QE after the financial crisis. Moreover, a distinct 'flight-to-quality' demand in the financial crisis is revealed.

To correctly specify the interactions between the yield factors and macro variables, some realistic assumptions are introduced to enhance our model, such as the settings of unspanned macro risks and time-varying parameters; but these assumptions cause econometric challenges in terms of model tractability. The challenges are addressed here by bringing in an efficient estimation technique. The NS-DMA model is believed to be robust, as it is highly consistent with the theoretical and empirical findings in the previous yield curve literature.

Future research could employ a one-step approach to provide forecasts with higher accuracy, in which case a trade-off should be made between predictive accuracy and estimation efficiency. Discussing the real part of the term structure is meaningful and desirable, but it is beyond the scope of this paper and will be our further work.

References

Altavilla, Carlo, Raffaella Giacomini, and Giuseppe Ragusa. 2014. "Anchoring the Yield Curve Using Survey Expectations." European Central Bank Working Paper Series 1632.

Anderson, B., Peter J. Hammond, and Cyrus A. Ramezani. 2010. "Affine models of the joint dynamics of exchange rates and interest rates." *Journal of Financial and Quantitative Analysis*, 45(5): 1341–1365.

Andrews, Donald WK. 1993. "Tests for parameter instability and structural change with unknown change point." *Econometrica: Journal of the Econometric Society*, 61(4): 821–856.

Andrews, Donald W.K., and Werner Ploberger. 1994. "Optimal tests when a nuisance parameter is present only under the alternative." *Econometrica*, 62(6): 1383–1414.

Ang, Andrew, and Monika Piazzesi. 2003. "A no-arbitrage vector autoregression of term structure dynamics with macroeconomic and latent variables." *Journal of Monetary Economics*, 50(4): 745–787.

Ang, Andrew, Sen Dong, and Monika Piazzesi. 2007. "No-arbitrage Taylor rules." National Bureau of Economic Research Working Paper 13448.

Anh, Le, and Scott Joslin. 2013. "Interest rate volatility and no-arbitrage affine term structure models." University of North Carolina at Chapel Hill Working Paper.

Avramov, Doron. 2002. "Stock return predictability and model uncertainty." *Journal of Financial Economics*, 64(3): 423–458.

Bacchetta, Philippe, and Eric Van Wincoop. 2004. "A scapegoat model of exchange-rate fluctuations." *American Economic Review*, 94(2): 114–118.

Bansal, Naresh, Robert A. Connolly, and Chris Stivers. 2014. "The stock-bond return relation, the term-structure's slope, and asset-class risk dynamics." *Journal of Financial and Quantitative Analysis*, FirstView: 1–44.

Bauer, Michael D., Glenn D. Rudebusch, and Jing Cynthia Wu. 2014. "Comment on 'term premia and inflation uncertainty: empirical evidence from an international panel dataset'." *American Economic Review*, 104(1): 323–337.

Bekaert, Geert, Seonghoon Cho, and Antonio Moreno. 2010. "New Keynesian macroeconomics and the term structure." *Journal of Money, Credit and Banking*, 42(1): 33–62.

Bernanke, Ben, Mark Gertler, and Simon Gilchrist. 1996. "The financial accelerator and the flight to quality." *The Review of Economics and Statistics*, 78(1).

Bernanke, Ben, Vincent Reinhart, and Brian Sack. 2004. "Monetary policy alternatives at the zero bound: An empirical assessment." *Brookings Papers on Economic Activity*, 2004(2): 1–100.

Bianchi, Francesco, Haroon Mumtaz, and Paolo Surico. 2009. "The great moderation of the term structure of UK interest rates." *Journal of Monetary Economics*, 56(6): 856–871.

Björk, Tomas, and Bent Jesper Christensen. 1999. "Interest rate dynamics and consistent forward rate curves." *Mathematical Finance*, 9(4): 323–348.

Campbell, John Y., and Robert J. Shiller. 1991. "Yield spreads and interest rate movements: A bird's eye view." *The Review of Economic Studies*, 58(3): 495–514.

Campbell, John Y., and Samuel B. Thompson. 2008. "Predicting excess stock returns out of sample: Can anything beat the historical average?" *Review of Financial Studies*, 21(4): 1509–1531.

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

Chow, Gregory C. 1960. "Tests of equality between sets of coefficients in two linear regressions." *Econometrica*, 28(3): 591–605.

Christensen, Jens H. E., and Glenn D. Rudebusch. 2012. "The response of interest rates to US and UK quantitative easing." *The Economic Journal*, 122(564): F385–F414.

Christensen, Jens H.E., Francis X. Diebold, and Glenn D. Rudebusch. 2011. "The affine arbitrage-free class of Nelson-Siegel term structure models." *Journal of Econometrics*, 164(1): 4–20.

Christensen, Jens H.E., Jose A. Lopez, and Glenn D. Rudebusch. 2010. "Inflation expectations and risk premiums in an arbitrage-free model of nominal and real bond yields." *Journal of Money, Credit and Banking*, 42(s1): 143–178.

Cieslak, Anna, and Pavol Povala. 2014. "Expecting the fed." Northwestern University Working Paper.

Cieslak, Anna, and Pavol Povala. 2015. "Information in the term structure of yield curve volatility." *Journal of Finance*, Forthcoming.

Clark, Todd E., and Kenneth D. West. 2006. "Using out-of-sample mean squared prediction errors to test the martingale difference hypothesis." *Journal of Econometrics*, 135(1): 155–186.

Clark, Todd E., and Kenneth D. West. 2007. "Approximately normal tests for equal predictive accuracy in nested models." *Journal of Econometrics*, 138(1): 291–311.

Cochrane, John H., and Monika Piazzesi. 2005. "Bond risk premia." *American Economic Review*, 95(1): 138–160.

Cochrane, John H., and Monika Piazzesi. 2009. "Decomposing the yield curve." *AFA 2010 Atlanta Meetings Paper.*

Coroneo, Laura, Domenico Giannone, and Michèle Modugno. 2014. "Unspanned macroeconomic factors in the yields curve." Board of Governors of the Federal Reserve System Working Paper 57.

Coroneo, Laura, Ken Nyholm, and Rositsa Vidova-Koleva. 2011. "How arbitrage-free is the Nelson-Siegel model?" *Journal of Empirical Finance*, 18(3): 393–407.

Creal, Drew D., and Jing Cynthia Wu. 2015. "Estimation of affine term structure models with spanned or unspanned stochastic volatility." *Journal of Econometrics*, 185(1): 60–81.

Cremers, K.J. Martijn. 2002. "Stock return predictability: A Bayesian model selection perspective." *Review of Financial Studies*, 15(4): 1223–1249.

D'Agostino, Antonello, Domenico Giannone, and Paolo Surico. 2006. "(Un)Predictability and macroeconomic stability." European Central Bank Working Paper Series 0605.

Dai, Qiang, and Kenneth J. Singleton. 2003. "Term Structure Dynamics in Theory and Reality." *Review of Financial Studies*, 16(3): 631–678.

Della Corte, Pasquale, Lucio Sarno, and Daniel L. Thornton. 2008. "The expectation hypothesis of the term structure of very short-term rates: Statistical tests and economic value." *Journal of Financial Economics*, 89(1): 158–174.

Dewachter, Hans, and Leonardo Iania. 2012. "An extended macro-finance model with financial factors." *Journal of Financial and Quantitative Analysis*, 46(06): 1893–1916.

Dewachter, Hans, and Marco Lyrio. 2008. "Learning, macroeconomic dynamics and the term structure of interest rates." In *Asset Prices and Monetary Policy.* 191–245. University of Chicago Press.

Dewachter, Hans, Leonardo Iania, and Marco Lyrio. 2014. "Information in the yield curve: A Macro-Finance approach." *Journal of Applied Econometrics*, 29(1): 42–64.

Diebold, Francis X., and Canlin Li. 2006. "Forecasting the term structure of government bond yields." *Journal of Econometrics*, 130(2): 337–364.

Diebold, Francis X., and G.D. Rudebusch. 2013. *Yield Curve Modeling and Forecasting: The Dynamic Nelson-Siegel Approach. The Econometric and Tinbergen Institutes Lectures*, Princeton University Press.

Diebold, Francis X., and Roberto S. Mariano. 1995. "Comparing Predictive Accuracy." *Journal of Business and Economic Statistics*, 13(3).

Diebold, Francis X., Glenn D. Rudebusch, and S. Borağan Aruoba. 2006. "The macroeconomy and the yield curve: A dynamic latent factor approach." *Journal of Econometrics*, 131(1): 309–338.

Duffee, Gregory R. 2002. "Term premia and interest rate forecasts in affine models." *Journal of Finance*, 57(1): 405–443.

Duffee, Gregory R. 2011*a*. "Forecasting with the term structure: The role of no-arbitrage restrictions." Johns Hopkins University, Department of Economics Working Paper.

Duffee, Gregory R. 2011*b*. "Information in (and not in) the term structure." *Review of Financial Studies*, 24(9): 2895–2934.

Duffee, Gregory R. 2013. "Bond pricing and the macroeconomy." In *Handbook of the Economics of Finance*. Vol. 2, Part B, , ed. George M. Constantinides, Milton Harris and Rene M. Stulz, 907–967. Elsevier.

Duffee, Gregory R. 2014. "Expected inflation and other determinants of Treasury yields." Johns Hopkins University, Department of Economics Working Paper.

Elliott, Graham, and Allan Timmermann. 2008. "Economic forecasting." *Journal of Economic Literature*, 46(1): 3–56.

Estrella, Arturo, and Frederic S. Mishkin. 1998. "Predicting US recessions: Financial variables as leading indicators." *Review of Economics and Statistics*, 80(1): 45–61.

Evans, Charles L., and David A. Marshall. 2007. "Economic determinants of the nominal treasury yield curve." *Journal of Monetary Economics*, 54(7): 1986–2003.

Feng, Ziding D., and Charles E. McCulloch. 1996. "Using bootstrap likelihood ratios in finite mixture models." *Journal of the Royal Statistical Society. Series B (Methodological)*, 609–617.

Feunou, Bruno, Jean-Sébastien Fontaine, and Anh Le. 2014. "Term structure modeling when monetary policy is unconventional: A new approach." Bank of Canada Working Paper.

Filipović, Damir. 1999. "A note on the Nelson–Siegel family." *Mathematical Finance*, 9(4): 349–359.

Fontaine, Jean-Sébastien, and René Garcia. 2012. "Bond liquidity premia." *Review of Financial Studies*, 25(4): 1207–1254.

Geweke, John, and Gianni Amisano. 2010. "Comparing and evaluating Bayesian predictive distributions of asset returns." *International Journal of Forecasting*, 26(2): 216–230.

Gürkaynak, Refet S., and Jonathan H. Wright. 2012. "Macroeconomics and the term structure." *Journal of Economic Literature*, 50(2): 331–367.

Gürkaynak, Refet S., Brian Sack, and Jonathan H. Wright. 2007. "The US Treasury yield curve: 1961 to the present." *Journal of Monetary Economics*, 54(8): 2291–2304.

Hanson, Bruce E. 2002. "Tests for parameter instability in regressions with I(1) processes." *Journal of Business and Economic Statistics*, 20(1): 45–59.

Harvey, Campbell R. 1988. "The real term structure and consumption growth." *Journal of Financial Economics*, 22(2): 305–333.

Hautsch, Nikolaus, and Fuyu Yang. 2012. "Bayesian inference in a stochastic volatility Nelson-Siegel model." *Computational Statistics and Data Analysis*, 56(11): 3774–3792.

Hördahl, Peter, Oreste Tristani, and David Vestin. 2006. "A joint econometric model of macroeconomic and term-structure dynamics." *Journal of Econometrics*, 131(1): 405–444.

Hu, Grace Xing, Jun Pan, and Jiang Wang. 2013. "Noise as information for illiquidity." *The Journal of Finance*, 68(6): 2341–2382.

Joslin, Scott, Anh Le, and Kenneth J. Singleton. 2013. "Gaussian macro-finance term structure models with lags." *Journal of Financial Econometrics*, 11(4): 581–609.

Joslin, Scott, Kenneth J. Singleton, and Haoxiang Zhu. 2011. "A new perspective on Gaussian dynamic term structure models." *Review of Financial Studies*, 24(3): 926–970.

Joslin, Scott, Marcel Priebsch, and Kenneth J. Singleton. 2014. "Risk premiums in dynamic term structure models with unspanned macro risks." *The Journal of Finance*, 69(3): 1197–1233.

Kim, Don H. 2009. "Challenges in Macro-Finance modeling." *Federal Reserve Bank of St. Louis Review*, 91(5, Part 2): 519–44.

Kim, Don H., and Athanasios Orphanides. 2012. "Term structure estimation with survey data on interest rate forecasts." *Journal of Financial and Quantitative Analysis*, 47(01): 241–272.

Kim, Don H., and Jonathan H. Wright. 2005. "An arbitrage-free three-factor term structure model and the recent behavior of long-term yields and distant-horizon forward rates." Board of Governors of the Federal Reserve System Working Paper.

Koop, Gary, and Dimitris Korobilis. 2012. "Forecasting inflation using Dynamic Model Averaging." *International Economic Review*, 53(3): 867–886.

Koop, Gary, and Dimitris Korobilis. 2013. "Large time-varying parameter VARs." *Journal of Econometrics*, 177(2): 185–198.

Koopman, Siem Jan, Max I.P. Mallee, and Michel Van der Wel. 2010. "Analyzing the term structure of interest rates using the dynamic Nelson-Siegel model with time-varying parameters." *Journal of Business and Economic Statistics*, 28(3): 329–343.

Krippner, Leo. 2015. "A theoretical foundation for the Nelson-Siegel class of yield curve models." *Journal of Applied Econometrics*, 30(1): 97–118.

Kurmann, André, and Christopher Otrok. 2013. "News shocks and the slope of the term structure of interest rates." *American Economic Review*, 103(6): 2612–32.

Litterman, Robert B., José Scheinkman, and Laurence Weiss. 1991. "Volatility and the yield curve." *The Journal of Fixed Income*, 1(1): 49–53.

Ludvigson, Sydney C., and Serena Ng. 2009. "Macro factors in bond risk premia." *Review of Financial Studies*, 22(12): 5027–5067.

McCulloch, J. Huston. 2007. "The Kalman foundations of adaptive least squares, with application to US inflation." Ohio State University Working Paper.

Neftci, Salih. 2004. *Principles of Financial Engineering.* Academic Press Inc.

Nelson, Charles R., and Andrew F. Siegel. 1987. "Parsimonious modeling of yield curves." *Journal of Business*, 60(4): 473–489.

Nyblom, Jukka. 1989. "Testing for the constancy of parameters over time." *Journal of the American Statistical Association*, 84(405): 223–230.

Orphanides, Athanasios, and Min Wei. 2012. "Evolving macroeconomic perceptions and the term structure of interest rates." *Journal of Economic Dynamics and Control*, 36(2): 239–254.

Piazzesi, Monika. 2010. "Affine term structure models." In *Handbook of Financial Econometrics*. Vol. 1, 691–766. Elsevier.

Piazzesi, Monika, and Martin Schneider. 2007. "Equilibrium yield curves." In *NBER Macroeconomics Annual 2006, Volume 21*. 389–472. MIT Press.

Prado, R., and M. West. 2010. *Time Series: Modeling, Computation, and Inference. Chapman & Hall/CRC Texts in Statistical Science*, Taylor & Francis.

Quandt, Richard E. 1960. "Tests of the hypothesis that a linear regression system obeys two separate regimes." *Journal of the American statistical Association*, 55(290): 324–330.

Raftery, Adrian E., Miroslav Kárnỳ, and Pavel Ettler. 2010. "Online prediction under model uncertainty via dynamic model averaging: Application to a cold rolling mill." *Technometrics*, 52(1): 52–66.

Rapach, David E., Jack K. Strauss, and Guofu Zhou. 2009. "Out-of-sample equity premium prediction: Combination forecasts and links to the real economy." *Review of Financial Studies*, 23(2): 821–862.

Rauch, Herbert E., C.T. Striebel, and F. Tung. 1965. "Maximum likelihood estimates of linear dynamic systems." *AIAA Journal*, 3(8): 1445–1450.

Rossi, Barbara. 2005. "Optimal tests for nested model selection with underlying parameter instability." *Econometric theory*, 21(05): 962–990.

Sarno, Lucio, Daniel L. Thornton, and Giorgio Valente. 2007. "The empirical failure of the expectations hypothesis of the term structure of bond yields." *Journal of Financial and Quantitative Analysis*, 42(01): 81–100.

Siegel, Andrew F., and Charles R. Nelson. 1988. "Long-term behavior of yield curves." *Journal of Financial and Quantitative Analysis*, 23(01): 105–110.

Smith, Josephine M., and John B. Taylor. 2009. "The term structure of policy rules." *Journal of Monetary Economics*, 56(7): 907–917.

Stock, James H., and Mark W. Watson. 2002. "Forecasting using principal components from a large number of predictors." *Journal of the American Statistical Association*, 97(460): 1167–1179.

Tang, Huarong, and Yihong Xia. 2007. "An international examination of affine term structure models and the expectations hypothesis." *Journal of Financial and Quantitative Analysis*, 42(01): 41–80.

Thornton, Daniel L, and Giorgio Valente. 2012. "Out-of-sample predictions of bond excess returns and forward rates: An asset allocation perspective." *Review of Financial Studies*, 25(10): 3141–3168.

Van Dijk, Dick, Siem Jan Koopman, Michel Van der Wel, and Jonathan H. Wright. 2014. "Forecasting interest rates with shifting endpoints." *Journal of Applied Econometrics*, 29(5): 693–712.

Welch, Ivo, and Amit Goyal. 2008. "A comprehensive look at the empirical performance of equity premium prediction." *Review of Financial Studies*, 21(4): 1455–1508.

West, Mike, and Jeff Harrison. 1997. *Bayesian Forecasting and Dynamic Models.* Springer-Verlag New York, Inc.

Wheelock, David C., and Mark E. Wohar. 2009. "Can the term spread predict output growth and recessions? A survey of the literature." *Federal Reserve Bank of St. Louis Review*, 91(Part 1): 419–440.

Wright, Jonathan H. 2006. "The yield curve and predicting recessions." Divisions of Research and Statistics and Monetary Affairs, Federal Reserve Board Working Paper 07.

Wright, Jonathan H. 2011. "Term premia and inflation uncertainty: Empirical evidence from an international panel dataset." *American Economic Review*, 101(4): 1514–1534.

Zhou, Guofu, and Xiaoneng Zhu. 2014. "What drives the international bond risk premia?" Washington University Working Paper.

Data Appendix

Table 7: List of Yields and Macro-Finance Variables

Series ID	Description
TB	3- and 6-month Treasury Bills (Secondary Market Rate) (NSA) [1]
ZCY	Smoothed Zero-coupon Yield from Gürkaynak, Sack and Wright (2007) (NSA) [1]
IND	Industrial Production Index [5]
CPI	Consumer Price Index for All Urban Consumers: All Items Less Food & Energy [5]
FED	Effective Federal Funds Rate, End of Month (NSA) [1]
SP	S&P 500 Stock Price Index, End of Month (NSA) [5]
TCU	Capacity Utilization: Total Industry [1]
M1	M1 Money Stock [5]
TCC	Total Consumer Credit Owned and Securitized, Outstanding (End of Month) [5]
LL	Loans and Leases in Bank Credit, All Commercial Banks [5]
DOE	DOE Imported Crude Oil Refinery Acquisition Cost (NSA) [5]
MSP	Median Sales Price for New Houses Sold in the United States (NSA) [5]
TWX	Trade Weighted U.S. Dollar Index: Major Currencies (NSA) [1]
ED	Eurodollar Spread: 3m Eurodollar Deposit Rate - 3m Treasury Bill Rate (NSA) [1]
WIL	Wilshire 5000 Total Market Index (NSA) [5]
DYS	Default Yield Spread: Moodys BAA-AAA (NSA) [1]
NFCI	National Financial Conditions Index (NSA) [1]

Notes:
1. In square brackets [·] we have a code for data transformations used in this data set: [1] means original series is used; [5] means log first-order difference is used to detrend and ensure stationarity. The series are seasonally adjusted except the ones with NSA.
2. Data are obtained from St. Louis Federal Reserve Economic Data [http://research.stlouisfed.org/], spanning from Nov. 1971 to Nov. 2013. The smoothed zero-coupon yield is available on the Federal Reserve Board website [http://www.federalreserve.gov/pubs/feds/2006/200628/200628abs.html/].
3. National Financial Conditions Index, provided by the Chicago Fed, is available on the website [http://www.chicagofed.org/webpages/publications/nfci/].
4. The small-size VAR model includes no macro variables. The medium-size VAR model includes only three macro variables: IND, CPI and FED. The large-size VAR model uses all the macro and financial variables in this data list.

Appendix A Econometric Methods

A.1 Bayesian Kalman Filter with Forgetting Factor

We conduct the Kalman filter estimation for the state space model with Eq. (2.3) and Eq. (2.4):

$$z_t = X_t \beta_t + v_t,$$

$$\beta_{t+1} = \beta_t + \mu_t,$$

where z_t is a $n \times 1$ vector of variables, $X_t = I_n \otimes \left[z'_{t-1}, ..., z'_{t-p}\right]'$, β_t are VAR coefficients, $v_t \sim N(0, \Sigma_t)$ with Σ_t an $n \times n$ covariance matrix, and $\mu_t \sim N(0, Q_t)$.

Given data at time t denoted as $D_t = (z_t, X_t)$ and all the data from time 1 to t denoted as $D_{1:t}$, the Bayesian solution to updating about the coefficients β_t takes the form

$$\begin{aligned} p(\beta_t|D_t) &\propto \mathbf{L}(z_t; \beta_t, X_t, D_{1:t-1})\, p(\beta_t|D_{t-1}), \\ p(\beta_t|D_{t-1}) &= \int_{\wp} p(\beta_t|D_{1:t-1}, \beta_{t-1})\, p(\beta_{t-1}|D_{t-1})\, d\beta_t, \end{aligned}$$

where $\wp$ is the support of β_t. The solution to this problem can be defined using a Bayesian generalization of the typical Kalman filter recursions. Given an initial condition $\beta_0 \sim N(m_0, \Phi_0)$ we can define (cf. West and Harrison (1997))[34]:

1. Posterior at time $t-1$

$$\beta_{t-1}|D_{t-1} \sim N(m_{t-1}, \Phi_{t-1}),$$

2. Prior at time t

$$\beta_t|D_{t-1} \sim N\left(m_{t|t-1}, \Phi_{t|t-1}\right),$$

[34] For a parameter θ we use the notation $\theta_{t|s}$ to denote the value of parameter θ_t given data up to time s (i.e. $D_{1:s}$) for $s > t$ or $s < t$. For the special case where $s = t$, I use the notation $\theta_{t|t} = \theta_t$

where $m_{t|t-1} = m_{t-1}$ and $\Phi_{t|t-1} = \Phi_{t-1} + Q_t$.

3. Posterior at time t

$$\beta_t|D_t \sim N\left(m_t, \Phi_t\right), \tag{A.1}$$

where $m_t = m_{t|t-1} + \Phi_{t|t-1}X_t'(V_t^{-1})'\widetilde{v}_t$ and $\Phi_t = \Phi_{t|t-1} - \Phi_{t|t-1}X_t'(V_t^{-1})'X_t\Phi_{t|t-1}'$, with $\widetilde{v}_t = z_t - X_t m_{t|t-1}$ the prediction error and $V_t = X_t\Phi_{t|t-1}X_t' + \Sigma_t$ its covariance matrix.

Following the discussion above, we need to find estimates for Σ_t and Q_t in the formulas above. We define the time t prior for Σ_t to be

$$\Sigma_t|D_{t-1} \sim iW\left(S_{t-1}, \delta n_{t-1}\right), \tag{A.2}$$

while the posterior takes the form

$$\Sigma_t|D_t \sim iW\left(S_t, n_t\right),$$

where $n_t = \delta n_{t-1}+1$ and $S_t = (1-a_t)\, S_{t-1} + a_t \left(S_{t-1}^{0.5}\Sigma_{t-1}^{-0.5}v_t v_t'\Sigma_{t-1}^{-0.5}S_{t-1}^{0.5}\right)$, with $a_t = n_t^{-1}$. In this formulation, v_t is replaced with the one-step ahead prediction error $\widetilde{v}_{t|t-1} = z_t - m_{t|t-1}X_t$. The estimate for S_t is approximately equivalent numerically to the Exponentially Weighted Moving Average (EWMA) filter $S_t = \delta S_{t-1} + (1-\delta)\, v_t v_t'$. The parameter δ is the decay factor, where for $0 < \delta < 1$. In fact, Koop and Korobilis (2013) apply such a scheme directly to the covariance matrix Σ_t, which results in a point estimate. In this case by applying variance discounting methods to the scale matrix S_t, we are able to approximate the full posterior distribution of Σ_t.

Regarding Q_t, we use the forgetting factor approach in Koop and Korobilis (2013); see also West and Harrison (1997) for a similar discounting approach. In this case Q_t is set to be proportionate to the filtered covariance $\Phi_{t-1} = cov\left(\beta_{t-1}|D_{t-1}\right)$ and takes the following form

$$Q_t = \left(\Lambda^{-1} - 1\right)\Phi_{t-1}, \tag{A.3}$$

for a given forgetting factor Λ.

The brief interpretation of forgetting factors is that they control how much 'recent past' information will be used. With the exponential decay for the forgetting factors, if it takes a value of 0.99, the information 24 periods ago (two years for monthly data) receives around 80% as much weight as the information of last period. If forgetting factor takes 0.95, then forecast performance 24 periods ago receives only about 30% as much weight. The similar implication holds for the decay factor.

A.2 Probabilities for Dynamic Selection and Averaging

To obtain the desire probabilities for dynamic selection or averaging, we need updating at each point in time. In papers such as Raftery, Kárnỳ and Ettler (2010) or Koop and Korobilis (2012) the models are TVP regressions with different sets of explanatory variables. The analogous result of the model prediction equation, when doing DMA or DPS, is

$$p(\beta_{t-1}|z^{t-1}) = \sum_{k=1}^{K} p(\beta_{t-1}^{(k)}|L_{t-1} = k, z^{t-1})\mathbf{Pr}(L_{t-1} = k|z^{t-1}), \tag{A.4}$$

where $L_{t-1} = k$ means the k_{th} model[35] is selected and $p(\beta_{t-1}^{(k)}|L_{t-1} = k, z^{t-1})$ is given by the Kalman filter (Eq. A.1). To simplify notation, let $\pi_{t|s}^{(l)} = \mathbf{Pr}(L_t = l|z^s)$.

Raftery, Kárnỳ and Ettler (2010) used an empirically sensible simplification that

$$\pi_{t|t-1}^{(i)} = \frac{\left(\pi_{t-1|t-1}^{(i)}\right)^{\alpha}}{\sum_{l=1}^{K}\left(\pi_{t|t}^{(l)}\right)^{\alpha}}, \tag{A.5}$$

where $0 < \alpha \leq 1$. A forgetting factor is also employed here, of which the meaning is discussed in the last section. The huge advantage of using the forgetting factor α is that it does not require an MCMC algorithm to draw transitions between models or a simulation algorithm over model space.

The model updating equation is

$$\pi_{t|t}^{(i)} = \frac{\pi_{t|t-1}^{(i)} p^{(i)}(z_t|z^{t-1})}{\sum_{l=1}^{K} \pi_{t|t-1}^{(l)} p^{(l)}(z_t|z^{t-1})}, \tag{A.6}$$

where $p^{(i)}(z_t|z^{t-1})$ is the predictive likelihood. When proceeding with Dynamic Model Selection, the model with the highest probability is the best model we would like to select. Alternatively, we can conduct Dynamic Model Averaging, which average the predictions of all models with respective probabilities.

[35]For example, the k_{th} model in Dynamic Model Selection/Averaging, or the k_{th} candidate γ value in Dynamic Prier Selection.

A.3 A Backward Smoother for the TVP-VAR

In Appendix A.1, the algorithm for estimating the VAR with time-varying parameters uses only forward recursions, i.e. parameters at time t are updated as new data become available. Such algorithm is very convenient when using the VAR for forecasting or real-time monetary policy analysis. One can obtain more accurate estimates of β_t by complementing the Kalman filter with a smoothing algorithm. Smoothing algorithms are based on backward recursions where information at time $t+1$ or forward is used to update coefficients at time t. It becomes obvious that such algorithms are not suitable for monetary policy in real-time, however they can be used for ex-post analysis of monetary phenomena.

Here we can use a fixed-interval smoother, such as the RTS (Rauch-Tung-Striebel) smoother developed by Rauch, Striebel and Tung (1965), which does not depend on the decay factors Λ and δ thus providing a mimimum mean square estimator of β_t without the need to optimize with respect to the decay factors:

$$\beta_{t|T} = E\left(\beta_t | D_{1:T}\right).$$

The solution to this optimization problem takes the following form (conditional on the information at time T)

$$\begin{aligned} m_{t|T} &= m_{t|t} + A_t \left(m_{t+1|T} - m_{t+1|t} \right), \\ \Phi_{t|T} &= \Phi_{t|t} + A_t \left(\Phi_{t+1|T} - \Phi_{t+1|t} \right) A_t', \end{aligned}$$

where $A_t = \Phi_{t|t} \Phi_{t+1|t}^{-1}$, and for t 'running backwards' from $T-1$ to 1.

As well as smoothing the large parameter vector β_t, one can obtain smoothed estimates of the covariance matrix Σ_t. West and Harrison (1997) provide such a backward (smoothing) algorithm. By iterating t from $T-1$ to 1, we can estimate

$$
\begin{aligned}
S_{t|T}^{-1} &= (1-\delta)\, S_t^{-1} + \delta S_{t+1|T}^{-1}, \\
n_{t|T} &= (1-\delta)\, n_t + \delta n_{t+1|T},
\end{aligned}
$$

in which case we can obtain $\Sigma_{t|T}$ using

$$\Sigma_t | D_{1:T} \sim iW\left(S_{t|T}, n_{t|T}\right).$$

Appendix B Interpretation of Factor Dynamics

We illustrate the factor dynamics in this section and try to shed light on the economic implications of the latent factors. The extracted NS factors are shown in Figure 13. The Level factor has a downward trend since the early 1980s. The Level factor also has greater persistence compared with the other more volatile factors. The downward trend in the Level factor is consistent with the descriptive statistics in Table 1 and the results of Koopman, Mallee and Van der Wel (2010). The latter suggest a strong link between the Level factor and (expected) inflation, as they share high persistence. Evans and Marshall (2007) also indicate that there is a link between the level of yields and inflation with structural VAR evidence. In particular, the Level factor fall significantly after the financial crisis, which may indicate that the market had low inflation expectations. The Level factor rises in 2013, potentially associated with rising inflation and the impact of the Fed's Quantitative Easing (QE) pattern.

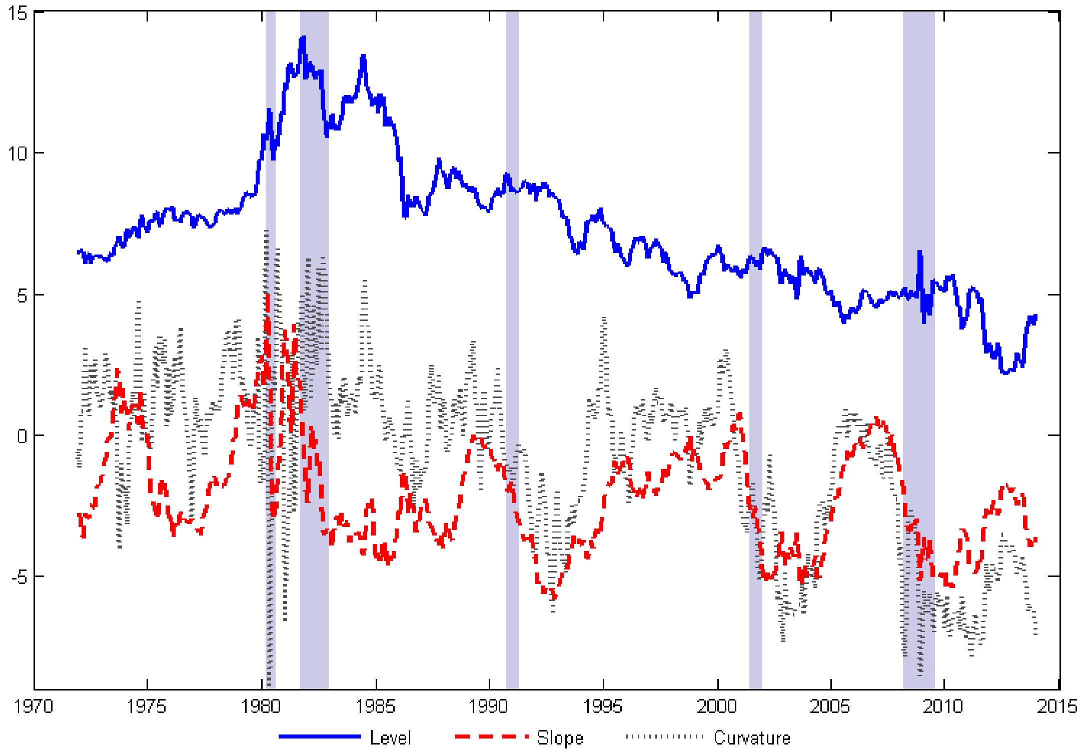
The Slope factor tends to fall sharply within recession periods, as indicated in Figure 13 by the shaded areas. Therefore, this factor could be closely related to real activity. The Slope factor is often considered as a proxy for the term spread, see Diebold, Rudebusch and Aruoba (2006). It can also be considered as a proxy for the stance of monetary policy, as the short end is influenced by policy rates.[36]

Lastly, the Curvature factor is harder to interpret and Diebold and Rudebusch (2013) indicate that this factor is less important than the other factors. On one hand, Litterman, Scheinkman and Weiss (1991) link the Curvature factor to the volatility of the level factor, via the argument of yield curve convexity, which can also be seen in Neftci (2004).[37] On the other hand, medium rates can be linked to expect short rates in the future, and therefore should be linked to current and expected future policies, which may potentially contain useful macro information missing in the first two factors.

[36] Recent research relates the Slope of term structure to news shocks on total factor productivity and asset-class risk, see Kurmann and Otrok (2013) and Bansal, Connolly and Stivers (2014).

[37] Generally, higher convexity means higher price-volatility or risk, and vice versa.

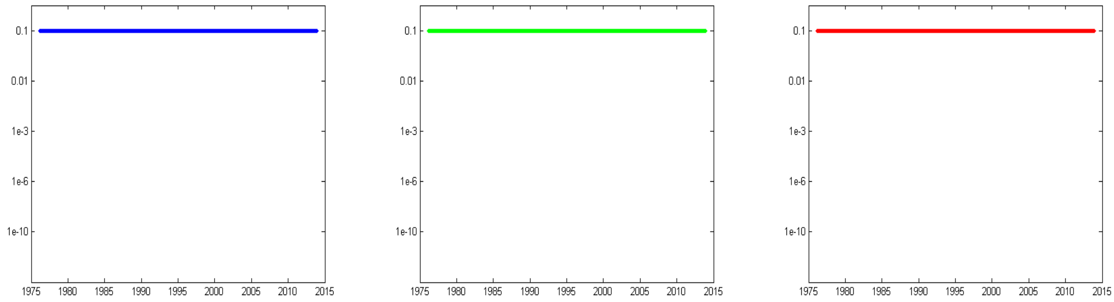
Figure 13: Nelson-Siegel Factor Dynamics



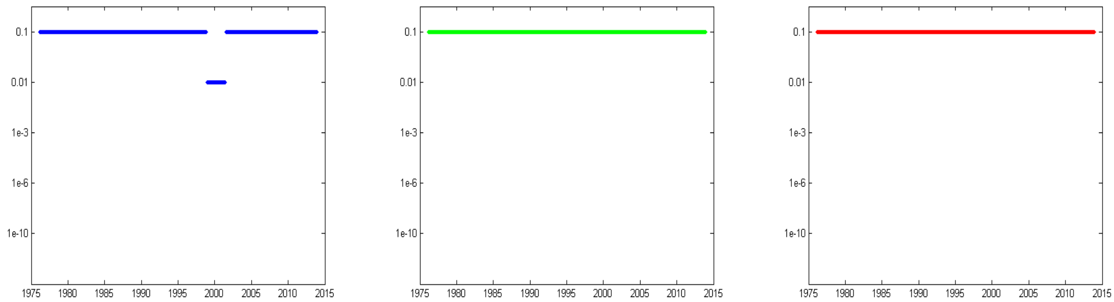
Notes: The graph shows the Nelson-Siegel Level, Slope and Curvature factors, which are drawn from Eq. (2.1). The shaded areas are recession periods according to the NBER Recession Indicators.

Appendix C Dynamic Prior Selection and Forecasting Results

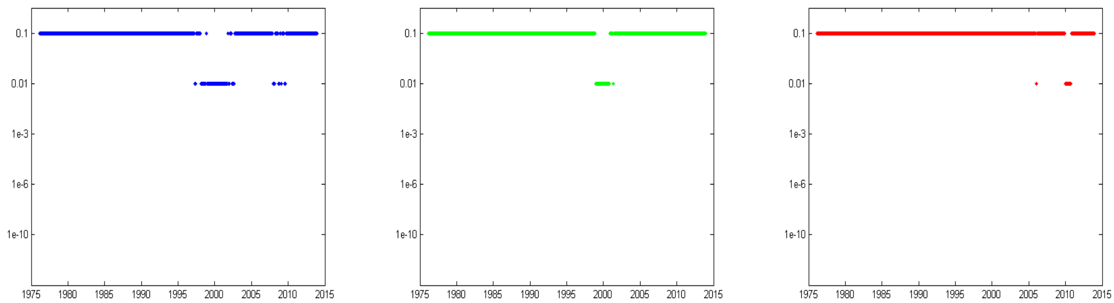
Figure 14: DPS results



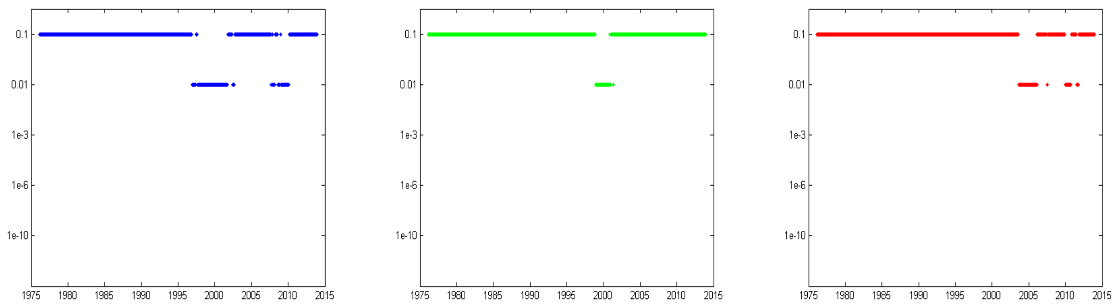
Notes: The above three graphs are the prior selection results given $\alpha = 0.99$. The blue, green and red lines (i.e. the lines in the left, middle and right) are the selections for the small, medium and large VAR models, respectively. The y-axis shows the candidate prior values $[10^{-10}, 10^{-6}, 0.001, 0.005, 0.01, 0.05, 0.1]$, and the x-axis indicates the time horizon from 1975.



Notes: The above three graphs are the prior selection results given $\alpha = 0.98$.



Notes: The above three graphs are the prior selection results given $\alpha = 0.97$.



Notes: The above three graphs are the prior selection results given $\alpha = 0.96$.

Table 8: Relative MSFE Performance of Term Structure Models

FH	6								12							
Maturity	**DMA**	**DMS**	**TVP**	**TVPM**	**DL**	**DLR10**	**DLM**	**DLSW**	**DMA**	**DMS**	**TVP**	**TVPM**	**DL**	**DLR10**	**DLM**	**DLSW**
3	0.871†	0.976	0.974	1.012	1.332	1.703	1.405	1.908	0.98†	1.073	1.021	1.240	1.349	1.605	1.517	1.677
6	0.947†	1.051	1.053	1.120	1.446	1.796	1.514	1.999	1.034†	1.128	1.079	1.292	1.419	1.703	1.579	1.784
12	0.969†	1.072	1.057	1.080	1.304	1.501	1.322	1.825	1.025†	1.139	1.082	1.210	1.353	1.592	1.458	1.661
24	1.025†	1.109	1.106	1.105	1.393	1.623	1.407	1.707	1.075†	1.191	1.139	1.208	1.474	1.757	1.573	1.664
36	1.038†	1.107	1.110	1.090	1.416	1.685	1.427	1.574	1.091†	1.202	1.152	1.188	1.528	1.848	1.623	1.625
48	1.038†	1.097	1.101	1.073	1.403	1.709	1.414	1.481	1.087†	1.193	1.145	1.163	1.532	1.885	1.625	1.591
60	1.032†	1.085	1.088	1.060	1.368	1.702	1.381	1.432	1.070†	1.175	1.127	1.140	1.505	1.884	1.596	1.577
72	1.021†	1.073	1.076	1.051	1.322	1.673	1.336	1.417	1.049†	1.154	1.106	1.121	1.459	1.858	1.549	1.581
84	1.009†	1.063	1.064	1.046	1.270	1.627	1.286	1.422	1.025†	1.133	1.086	1.105	1.405	1.816	1.494	1.599
96	0.997†	1.055	1.056	1.044	1.218	1.568	1.236	1.438	1.003†	1.115	1.068	1.092	1.349	1.766	1.437	1.623
108	0.987†	1.048	1.049	1.043	1.167	1.502	1.187	1.458	0.983†	1.100	1.054	1.083	1.294	1.711	1.381	1.649
120	0.978†	1.043	1.045	1.044	1.122	1.433	1.142	1.477	0.966†	1.089	1.043	1.077	1.243	1.655	1.329	1.673
Mean	0.994†	1.067	1.067	1.067	1.323	1.632	1.348	1.607	1.035†	1.143	1.093	1.174	1.415	1.748	1.524	1.648

Notes: 1. This table shows 6-month and 12-month ahead forecasts of bond yields with maturities ranging from 3 months to 120 months. The predictive duration is from early 1983 to the end of 2013.
2. The MSFE-based statistics relative to the RW are reported. The dagger (†) indicates, in terms of the sum of predictive log-likelihood, the model has the preferred value among all models at certain maturities (or in total), see Geweke and Amisano (2010).
3. In this table, we use following abbreviations. **MSFE**: Mean Squared Forecasting Error; **Mean**: Averaged MAFE across all sample maturities. **DMA** (Dynamic Model Averaging) averages all the models with probabilities in each step, while **DMS** (Dynamic Model Selection) chooses the best model with the highest probability at any point in time. **TVP-M**: a time-varying parameter model with three macro variables: fund rate, inflation and industrial production, similar to Bianchi Mumtaz and Surico (2009) but estimated with a fast algorithm without the need of MCMC; **DL**: Diebold and Li (2006) model, i.e. constant coefficient Vector Autoregressive model with recursive (expanding) estimations; **DL-R10**: Diebold and Li (2006) estimates based 10-year rolling windows; **TVP**: a time-varying parameter model without macro information; **DL-M**: factor dynamics in Diebold and Li (2006) are augmented with three macro variables: fund rate, inflation and industrial production, using recursive estimations; **DL-SW**: factor dynamics in Diebold and Li (2006) are augmented with with three principal components (see Stock and Watson (2002)) of our macro-finance data, using recursive estimations; **RW**: Random Walk.

Table 9: Relative MAFE Performance of Term Structure Models

FH	1								3							
Maturity	**DMA**	**DMS**	**TVP**	**TVPM**	**DL**	**DLR10**	**DLM**	**DLSW**	**DMA**	**DMS**	**TVP**	**TVPM**	**DL**	**DLR10**	**DLM**	**DLSW**
3	0.851†	0.890	0.866	0.858	0.968	1.050	0.979	1.275	0.884†	0.948	0.939	0.926	1.077	1.301	1.092	1.527
6	0.976†	1.036	1.025	1.007	1.101	1.218	1.120	1.462	0.957†	1.019	1.021	1.028	1.189	1.391	1.206	1.621
12	1.023†	1.047	1.021	1.039	0.984	0.963	1.013	1.324	0.987†	1.031	1.017	1.041	1.059	1.128	1.067	1.408
24	1.028†	1.053	1.041	1.047	1.047	1.059	1.054	1.272	1.006†	1.034	1.039	1.053	1.122	1.201	1.127	1.289
36	0.990†	1.010	1.011	1.015	1.062	1.142	1.073	1.131	1.013†	1.033	1.047	1.044	1.145	1.242	1.151	1.194
48	0.973†	0.987	1.000	0.992	1.065	1.164	1.072	1.049	1.012†	1.020	1.041	1.031	1.145	1.255	1.149	1.128
60	0.976†	0.986	1.000	0.988	1.061	1.167	1.066	1.023	1.011†	1.013	1.035	1.028	1.130	1.254	1.134	1.096
72	0.984†	0.995	1.006	1.000	1.051	1.160	1.059	1.030	1.008†	1.011	1.031	1.028	1.110	1.243	1.113	1.090
84	0.993†	1.006	1.014	1.013	1.041	1.140	1.047	1.050	1.003†	1.011	1.027	1.028	1.082	1.218	1.091	1.094
96	0.997†	1.014	1.020	1.025	1.030	1.110	1.038	1.071	0.998†	1.011	1.023	1.028	1.059	1.186	1.068	1.108
108	1.004†	1.023	1.027	1.037	1.015	1.069	1.029	1.091	0.994†	1.014	1.021	1.031	1.039	1.150	1.048	1.126
120	1.010†	1.028	1.034	1.044	1.004	1.028	1.019	1.110	0.988†	1.013	1.017	1.030	1.018	1.109	1.029	1.139
Mean	0.988†	1.008	1.008	1.009	1.036	1.104	1.047	1.143	0.991†	1.015	1.023	1.027	1.098	1.220	1.106	1.224

Notes: 1. This table shows 1-month and 3-month ahead forecasts of bond yields with maturities ranging from 3 months to 120 months. The predictive duration is from 1983:10 to 2013:11.
2. The MAFE-based statistics relative to the RW are reported. The dagger (†) indicates, in terms of the sum of predictive log-likelihood, the model has the preferred value among all models at certain maturities (or in total), see Geweke and Amisano (2010).
3. In this table, we use following abbreviations. **MAFE**: Mean Absolute Forecasting Error; **Mean**: Averaged MSFE across all sample maturities. **DMA** (Dynamic Model Averaging) averages all the models with probabilities in each step, while **DMS** (Dynamic Model Selection) chooses the best model with the highest probability at any point in time. **TVP-M**: a time-varying parameter model with three macro variables: fund rate, inflation and industrial production, similar to Bianchi Mumtaz and Surico (2009) but estimated with a fast algorithm without the need of MCMC; **DL**: Diebold and Li (2006) model, i.e. constant coefficient Vector Autoregressive model with recursive (expanding) estimations; **DL-R10**: Diebold and Li (2006) estimates based 10-year rolling windows; **TVP**: a time-varying parameter model without macro information; **DL-M**: factor dynamics in Diebold and Li (2006) are augmented with three macro variables: fund rate, inflation and industrial production, using recursive estimations; **DL-SW**: factor dynamics in Diebold and Li (2006) are augmented with with three principal components (see Stock and Watson (2002)) of our macro-finance data, using recursive estimations; **RW**: Random Walk.

Table 10: Relative MAFE Performance of Term Structure Models

FH	6								12							
Maturity	**DMA**	**DMS**	**TVP**	**TVPM**	**DL**	**DLR10**	**DLM**	**DLSW**	**DMA**	**DMS**	**TVP**	**TVPM**	**DL**	**DLR10**	**DLM**	**DLSW**
3	0.951†	1.009	1.020	1.003	1.206	1.360	1.249	1.522	1.028†	1.074	1.078	1.127	1.220	1.321	1.296	1.374
6	0.999†	1.055	1.068	1.061	1.250	1.411	1.290	1.581	1.054†	1.098	1.101	1.141	1.245	1.365	1.313	1.438
12	0.995†	1.052	1.044	1.047	1.131	1.236	1.148	1.445	1.031†	1.085	1.073	1.082	1.190	1.308	1.229	1.389
24	0.999†	1.040	1.036	1.033	1.165	1.263	1.176	1.348	1.055†	1.095	1.093	1.068	1.242	1.392	1.274	1.363
36	1.017†	1.040	1.046	1.027	1.187	1.297	1.197	1.290	1.063†	1.092	1.097	1.054	1.258	1.433	1.287	1.313
48	1.022†	1.036	1.046	1.024	1.187	1.305	1.194	1.243	1.061†	1.086	1.089	1.046	1.259	1.448	1.287	1.291
60	1.017†	1.029	1.040	1.021	1.174	1.296	1.177	1.212	1.054†	1.079	1.080	1.039	1.250	1.448	1.276	1.287
72	1.008†	1.022	1.033	1.019	1.153	1.277	1.155	1.199	1.048†	1.073	1.070	1.034	1.239	1.442	1.266	1.295
84	0.999†	1.015	1.025	1.016	1.127	1.250	1.128	1.196	1.041†	1.069	1.062	1.029	1.227	1.431	1.254	1.313
96	0.993†	1.011	1.020	1.014	1.102	1.227	1.104	1.205	1.032†	1.066	1.054	1.025	1.212	1.415	1.239	1.336
108	0.985†	1.008	1.015	1.012	1.079	1.205	1.080	1.215	1.019†	1.061	1.045	1.020	1.193	1.393	1.219	1.357
120	0.979†	1.007	1.012	1.012	1.061	1.183	1.061	1.228	1.007†	1.056	1.038	1.016	1.171	1.369	1.198	1.378
Mean	0.998†	1.028	1.034	1.025	1.153	1.276	1.164	1.306	1.042†	1.079	1.075	1.060	1.227	1.395	1.264	1.346

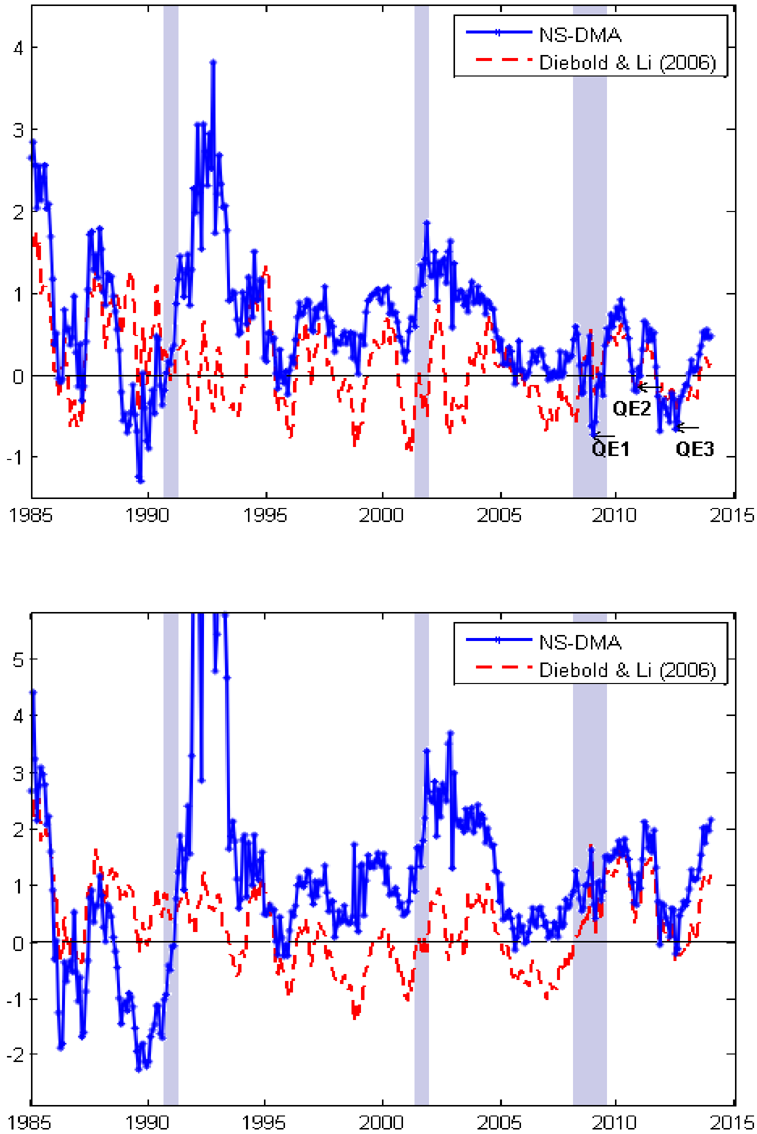
Notes: 1. This table shows 6-month and 12-month ahead forecasts of bond yields with maturities ranging from 3 months to 120 months. The predictive duration is from early 1983 to the end of 2013.

2. The MAFE-based statistics relative to the RW are reported. The dagger (†) indicates, in terms of the sum of predictive log-likelihood, the model has the preferred value among all models at certain maturities (or in total), see Geweke and Amisano (2010).

3. In this table, we use following abbreviations. **MAFE**: Mean Absolute Forecasting Error; **Mean**: Averaged MAFE across all sample maturities. **DMA** (Dynamic Model Averaging) averages all the models with probabilities in each step, while **DMS** (Dynamic Model Selection) chooses the best model with the highest probability at any point in time. **TVP-M**: a time-varying parameter model with three macro variables: fund rate, inflation and industrial production, similar to Bianchi Mumtaz and Surico (2009) but estimated with a fast algorithm without the need of MCMC; **DL**: Diebold and Li (2006) model, i.e. constant coefficient Vector Autoregressive model with recursive (expanding) estimations; **DL-R10**: Diebold and Li (2006) estimates based 10-year rolling windows; **TVP**: a time-varying parameter model without macro information; **DL-M**: factor dynamics in Diebold and Li (2006) are augmented with three macro variables: fund rate, inflation and industrial production, using recursive estimations; **DL-SW**: factor dynamics in Diebold and Li (2006) are augmented with with three principal components (see Stock and Watson (2002)) of our macro-finance data, using recursive estimations; **RW**: Random Walk.

Appendix D Term Premia of Diebold-Li and NS-DMA

Figure 15: Time-Varying Term Premia of 36-and 120-Month Bonds



Notes:
1. The top panel is the 36-month term premia and the bottom is the 120-month term premia. The EH consistent 36- and 120-month and bond yields are estimated using Eq. (4.2); we then calculate the term premia using Eq. (4.5).
2. In addition to NS-DMA, we plot the recursively estimated term premia employing the methods proposed by Diebold and Li (2006).
3. Shaded areas are recession periods based on the NBER Recession Indicators. The unit is percentage.