Essex Finance Centre
Working Paper Series

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

"Joseph P. Byrne, Shuo Cao and Dimitris Korobilis"

Term Structure Dynamics, Macro-Finance Factors and Model Uncertainty

Joseph P. Byrne[*], Shuo Cao[†] and Dimitris Korobilis[‡]

Abstract

This paper models and predicts the term structure of US interest rates in a data rich environment. We allow the model dimension and parameters to change over time, accounting for model uncertainty and sudden structural changes. The proposed time-varying parameter Nelson-Siegel Dynamic Model Averaging (DMA) predicts yields better than standard benchmarks. DMA performs better since it incorporates more macro-finance information during recessions. The proposed method allows us to estimate plausible real-time term premia, whose countercyclicality weakened during the financial crisis.

---

[*]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: shuo.cao@outlook.com.
[‡]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. Three pricing factors can capture most of the variation in bond yield data, as indicated in Nelson and Siegel (1987) and Litterman and Scheinkman (1991). Diebold and Li (2006) propose a dynamic Nelson-Siegel (NS) model and successfully predict the yield curve. Our paper builds upon previous work and proposes a term structure model with the ensemble of several salient features. Firstly, to fully capture the factor dynamics, both parameter instability and stochastic volatility in a large system are taken into account. We utilize the dynamic Nelson-Siegel setup with time-varying parameters following Bianchi, Mumtaz and Surico (2009).[1] Our time-varying macro-finance model builds upon a large vector autoregressive (VAR) system with macroeconomic and financial factors in the spirit of Carriero, Kapetanios and Marcellino (2012) and Coroneo, Giannone and Modugno (2015). By extending Koop and Korobilis (2013) a Bayesian method is developed that allows a fast estimation of large systems with many variables.

Secondly, in a reduced-form representation we incorporate financial information in addition to traditional macro variables. Ang and Piazzesi (2003) introduce inflation and the output gap to augment the term structure model and show that macro factors can explain large variation in bond yields. This evidence is echoed by other researchers such as Diebold, Rudebusch and Aruoba (2006), who also stress the importance of key macro variables for the yield curve. Moreover, Moench (2008) shows that a term structure model augmented with a broad macro-finance information set can provide superior forecasts, and the global financial crisis, as an abrupt nonlinear shock, highlighted the importance of the financial market for macroeconomic activity and bond yields more generally. In this paper, we incorporate a substantial range of macro-finance risk factors with modeling techniques that distill large datasets.

Lastly, the proposed model accommodates different degrees of structural changes. Following Koop and Korobilis (2012) we employ Dynamic Model Averaging (DMA) methods in order to determine in a data-based way which macroeconomic or financial risks are relevant for the yield curve.[2] We can choose, at different points in time, between three models: i) one with three pricing factors only; ii) pricing factors plus three key macroeconomic indicators; and iii) pricing factors augmented using up to 15 macro and financial

---

[1]The term structure model is also similar to Van Dijk et al. (2014), where they show drifting parameters are helpful in improving forecast performance.

[2]Bayesian model averaging accounts for model uncertainty, see Bauer (2016) for the implementation in a static setup.

factors. The 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) in 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). Using DMA probabilities are assigned to each of the models at each point in time and thus averaging is dynamically implemented. When compared with alternative time-varying parameter models, this method is more robust as it encompasses moderate to sudden changes in economic conditions. DMA allows agents to flexibly shift to a more plausible model specification over time, and Elliott and Timmermann (2008) indicate this method can reduce the total forecast risk associated with using only a single 'best' model.

We empirically examine U.S. term structure dynamics using monthly observations from 1971 to 2013. The proposed approach has useful empirical properties in yield forecasting, as it considers parameter and model uncertainty and is robust to potential structural breaks. We compare the forecast performance of DMA to a basic dynamic Nelson-Siegel model and several variants, and show that gains in predictability are due to the ensemble of salient features – time-varying coefficients, stochastic volatility and dynamic model averaging. We find that the predictability of term structure models is time-varying and tends to be procyclical, and macro-finance information is important during recessions. The superior out-of-sample forecasting performance of DMA, especially for short rates, reveals plausible expectations of market participants in real time, and the indicators of real activity and the stock market are particularly helpful in explaining the movements.[3] Using only conditional information, DMA provides successful term premium alternatives to full-sample estimates produced by the no-arbitrage term structure models of Kim and Wright (2005), Wright (2011) and Bauer, Rudebusch and Wu (2014). The estimated term premia has a significant countercyclical pattern, but it appears this pattern is weakened in the global financial crisis possibly because of 'flight-to-quality' demand for US bonds.

This paper is structured as follows. Section 2 describes the framework and the estimation method for modeling bond yield dynamics. Section 3 describes the data and discusses the results. Specifically, the first two subsections test the parameter instability and elaborate on the usefulness of employing DMA. Section 3.3 displays the point and density forecasting performance of our term structure model. Section 3.4 presents that the evidence of time-varying predictability and reveals important macro-finance sources that drive the bond yields. Section 3.5 shows the model-implied term premia has informative

---

[3]This is consistent with Kurmann and Otrok (2013) and Bansal, Connolly and Stivers (2014), who relate the changes in the term structure to news shocks on total factor productivity and asset-class risk, respectively.

economic implications. Section 4 concludes.

2 Methods

2.1 The Cross-Sectional 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. 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. Let $y_t(\tau)$ denote yields at maturity τ, then the factor model we use is of the form:[4]

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

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 above Nelson-Siegel restrictions on loadings are cross-sectional restrictions. Feunou et al. (2014) show that the NS model is the continuous time limit of their near arbitrage-free class with a unit root in the pricing dynamics. In light of their findings, we specify the cross-sectional loadings with NS restrictions and focus on time-series variation of yield factors, in order to improve the forecast performance.[5]

The time series or physical dynamics of factors are augmented with macro-finance information in an unrestricted VAR. In this setup, the macro variables only affect the unobserved NS factors and do not interact contemporaneously with the observed yields, so that they are unspanned by the yields. In other words, a 'knife-edge' restriction is

---

[4]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.

[5]Joslin, Singleton and Zhu (2011) show that no-arbitrage cross-sectional restrictions cannot improve out-of-sample forecasts in the context of canonical Gaussian affine term structure models. We test the robustness of core results to the no-arbitrage restrictions in Appendix C.3.

imposed on the coefficients of macro variables in the cross section, while the time-series dynamics are left unconstrained, see Joslin, Priebsch and Singleton (2014) for details.

2.2 Yield Factor Dynamics

In the 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 is of considerable empirical importance. The Level factor L_t^{NS} loads on all maturities evenly. The Slope factor S_t^{NS} approximates the long-short spread, 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).[6]

An important and 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 (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, $B_{1t}, ..., B_{pt}$ are time-varying autoregressive coefficients, M_t is a vector of macro-finance risk factors, and v_t is the error term. Following Coroneo, Giannone and Modugno (2015) 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 $z_t = \begin{bmatrix} L_t^{NS}, S_t^{NS}, C_t^{NS}, M_t' \end{bmatrix}'$, M_t is a $q \times 1$ vector of macro-finance factors, $X_t = I_n \otimes \begin{bmatrix} z_{t-1}', ..., z_{t-p}' \end{bmatrix}$ for $n = q+3$, $\beta_t = \begin{bmatrix} c_t, vec(B_{1t})', \cdots, vec(B_{pt})' \end{bmatrix}'$ is a vector summarizing all VAR coefficients, $v_t \sim N(0, \Sigma_t)$ 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

---

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

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 the VAR coefficients,

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

based upon a prior β_0 discussed below, and $\mu_t \sim N(0, Q_t)$. Following Koop and Korobilis (2013) we set $Q_t = (\Lambda^{-1} - 1) \, cov(\beta_{t-1} | \mathcal{D}_{t-1})$ 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:

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

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 for model structural instability and regime changes in the joint dynamics of the NS factors and the macroeconomic and financial factors. In Bayesian inference if Markov Chain Monte Carlo is employed, it will be computationally demanding especially in a recursive forecasting context. Here we extend the methodology of Koop and Korobilis (2013) and conduct a fast 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

This paper argues that the possible set of risk factors relevant for characterizing the yield curve can change over time. We are faced, therefore, with multiple potential yield curve models. Hence, we focus on Eq. (2.3) and work with three different model 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, therefore $q = 0$ 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 choose among different models at each point in time. That is, for
each t we chose the optimal $\mathcal{M}^{(i)}$ which maximizes the probability/weight

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

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
the updating step

$$\pi_{t|t}^{(i)} \propto \pi_{t|t-1}^{(i)} p^{(i)}\left(z_t | D_{t-1}\right). \tag{2.8}$$

where the quantity $p^{(i)}\left(z_t | D_{t-1}\right)$ 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. In this paper we focus on the predictive likelihoods of the three
Nelson-Siegel factors when implementing DMA. The time t prior $\pi_{t|t-1}^{(i)}$ is given by

$$\pi_{t|t-1}^{(i)} = \frac{\left(\pi_{t-1|t-1}^{(i)}\right)^{\alpha}}{\sum_{i=1}^{K}\left[\left(\pi_{t-1|t-1}^{(i)}\right)^{\alpha}\right]} \tag{2.9}$$

where $0 < \alpha \leq 1$ is a decay factor which allows discounting exponentially past forecasting
performance, see Koop and Korobilis (2013) for more information. When $\alpha \to 0$ we have
the case that at each point in time we update our beliefs with a prior of equal weights
for each model. 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. In this
paper a sufficiently small value is used for α such that the time t prior is flat, and we will
show later this can capture the changing economic conditions and increase the predictive
performance.

2.3.2 Prior Selection

We define a Minnesota prior for our VAR, which provides 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

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

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 Dynamic 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), the forgetting factor Λ in Eq. (A.3), and the decay factor δ in Eq. (A.2).[7] Regarding the forgetting factor and the decay factor, we follow recommendations in Koop and Korobilis (2013). Intuitively, these parameters control the discounting of past information, which occurs at an exponential rate. When these parameters are equal to one, the model becomes a constant parameter model. Values smaller than one discount past data at a faster rate, allowing faster switches of model parameters. However, too small values may induce sudden changes to outliers, so the state space system is not stable and the results will not be reliable. Hence, following Koop and Korobilis (2013), we choose relatively high values (but less than one) to ensure stability while still allowing for flexibility: The Λ and δ are set to 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

---

[7]Following Diebold and Li (2006), Bianchi, Mumtaz and Surico (2009) and Van Dijk et al. (2014) we set $\lambda^{NS} = 0.0609$.

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 (i.e. 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 (i.e. NS + macro-finance) includes all 15 macro and financial variables, which should comprehensively include the information the market players are able to acquire.

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 instability and structural breaks in the literature from both a frequentist and a Bayesian perspective; see for example, Andrews and Ploberger (1994), Hanson (2002) and Rossi (2005). McCulloch (2007) suggests a likelihood-based approach to test parameter instability in a TVP model. The limiting distribution of the test statistics may not be standard and, consequently, its critical values need to be bootstrapped. In the spirit of McCulloch (2007), we construct a likelihood-based test on the small 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

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.

concerning out-of-sample instability. 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 it is nested.[8] 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

---

[8]In particular, as Koop and Korobilis (2013) show, by setting the forgetting and decay factors $\Lambda = \delta = 1$, our model is equivalent to the recursive estimation of a model with constant parameters and volatility.

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.

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 time-varying physical factor dynamics of the small-size VAR in Figure 1. This can 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 seem relatively stable, but the mild variation in the eigenvalue would translate into sufficiently large changes in long-term expectations. Another observation is the clear rising trend for the third eigenvalue, which implies the third pricing factor is becoming more persistent. 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. This is evidence of sudden structural changes. As macro-finance information is considered important during recessions as suggested by Bernanke, Gertler and Gilchrist (1996), it is uncertain whether the small-size VAR can still produce plausible forecasts when faced with structural instability.

Figure 1: Time-Varying Persistence of Physical Dynamics
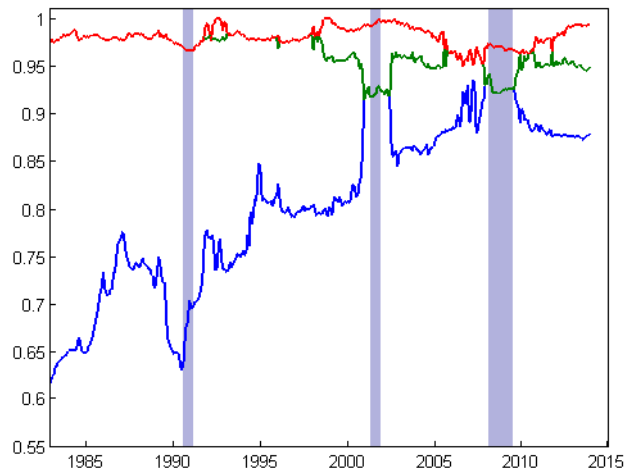
Notes: The graph shows the largest three eigenvalues of the physical dynamics in the small-size TVP model. The shaded areas are recession periods according to the NBER Recession Indicators.

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 Dynamic Prior Selection 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 associated forgetting factor fixed. The associated forgetting factor controls the persistence of probabilities, and the results do not change substantially as long as it is sufficiently large: the best γ values is relatively stable for all three sizes of models when the forgetting factor is larger than 0.90. 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, possibly because of the fact that fixing γ reduces posterior parameter uncertainty which in turn can affect uncertainty of posterior predictive densities.

Graphical evidence of the usefulness of our model averaging approach is provided by the Figure 2. The upper two panels set out the relative importance of the small, medium and large VAR models used in DMA. In general, there is substantial time variation in the weights, and the empirical observations are of economic importance.

Figure 2: Model Weights for NS, NS plus Macro and NS plus Macro-Finance VAR Models
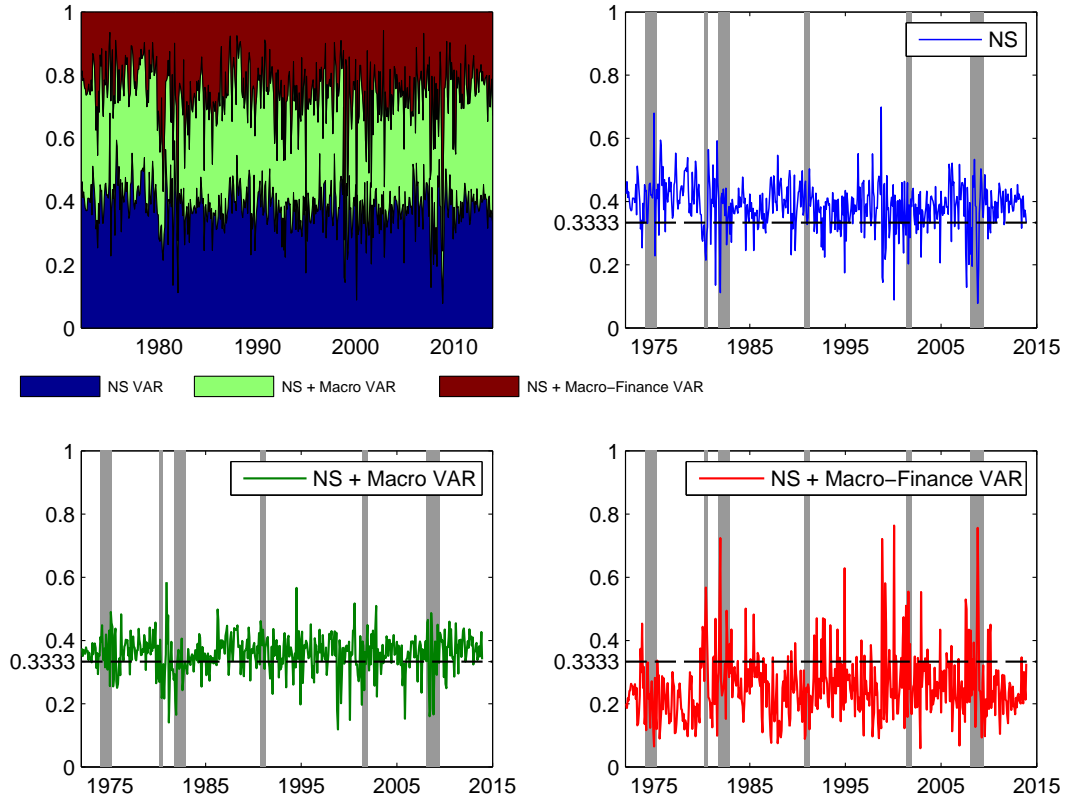
Notes:
1. This figure sets out the time-varying probabilities of our three models in our Dynamic Model Averaging (DMA) approach. The probabilities for DMA are updated from a Kalman filter based on the predictive accuracy, see Eq. (A.5); the probabilities/weights of the VAR models sum up to 1.
2. The upper left panel shows the probability weights of all models. The upper right and the lower panels display the weights of the NS VAR, NS + Macro VAR and the NS + Macro-Finance VAR, respectively. The shaded areas are the recession periods based on NBER Recession Indicators.

Firstly, during recession periods, the approach tends to use more macro-finance information to generate forecasts. The probability of the large-size (macro-finance) model rose steeply and then stayed at a high level during macroeconomic recessions. This is indicated by the higher weights for the macro-finance model during recession periods in the lower right panel of Figure 2. In times of acute economic stress, macroeconomic and financial risk factors become more relevant for modeling yields, which is supported by the 'financial accelerator' argument of Bernanke, Gertler and Gilchrist (1996). Among the three, the macro-finance model displays the largest variability in terms of the assigned weights. Hence the additional macro-finance information used to predict yields is appropriately modeled using the DMA approach.

Additionally, the allocated weights of small-size NS model are similar to the medium-size (NS + macro) model. These two models generally have higher weights in the DMA

during non-recession periods, but the medium-size model tends to be more stable. This means parsimonious yield curve models with macroeconomic variables, such as Ang and Piazzesi (2003) and Diebold, Rudebusch and Aruoba (2006), are generally effective except during recession periods.

It is worth reiterating the importance of the large macro-finance VAR, as Altavilla, Giacomini and Ragusa (2014) indicate that the original version of the dynamic NS model without macro information has weaker predictive power in recent years. We show the large-size VAR significantly boosts the forecast performance because of its superior performance during the recession periods. Moreover, model averaging expands the model set when compared with a single-model setup or model selection, and potentially mitigates the misspecification problem. Intuitively, the consideration of models with richer information allows us to effectively 'hedge' the risk of using a single model as Elliott and Timmermann (2008) suggest.

Since the changes in model weights are very sensitive to new information, DMA allows us to react to sudden, rather than smooth, changes in coefficients. Without model averaging or selection, a time-varying parameter model with a specific information set may have volatile performance in forecasting, as the true dynamics may not be well captured during certain periods. Our approach encompasses moderate to sudden changes in the economic environment and accordingly is promising in producing more stable forecasting performance.

3.3   Forecasting Performance

We now consider the forecasting performance of our approach. We use the Dynamic Model Averaging (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, with the addition of 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, because of the consideration of unspanned macro risks. The predictive duration is from 1983:10 to the 2013:11.

To better evaluate the predictive performance of DMA, we have the following seven variants of dynamic Nelson-Siegel models: recursive estimation of factor dynamics using standard VAR following Diebold and Li (2006) (DL), 10-year rolling-window VAR estimations (DL-R10), recursive VAR estimation with three macro variables (DL-M), recursive estimations of standard VAR with macro-finance principal components following

14

Stock and Watson (2002) (DL-SW), time-varying parameter VAR estimations of factor dynamics without macro information (TVP), time-varying parameter VAR estimations of factor dynamics with three macro variables (TVP-M), and Dynamic Model Selection (DMS).

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, 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 is essentially the model estimated in Bianchi, Mumtaz and Surico (2009) using MCMC methods. We report the performance of all models relative to the Random Walk (RW) model so that we can evaluate whether the term structure models successfully capture the high persistence in bond yields.

We assess all models' predictive properties in Table 3 which displays the one-period and three-period ahead Mean Squared Forecasting Error (MSFE) Performance for all forecasting models. The core empirical results are very encouraging for the proposed method. As can be seen in Table 3, our preferred DMA model consistently outperforms all the benchmark models. Table 4 shows the DMA is also preferred at relatively long forecast horizons.[9] The cumulative sum of predictive log-likelihood is displayed in Figure 3. It shows that the predictive density of the DMA is more accurate compared to the predictive density of the Diebold-Li (DL) across all maturities, especially for short rates.

Among all models, the results indicate DMA is the only one comparable in forecasting performance to, or better than, the RW. In fact, DMA not only successfully captures the persistence in bond yields, but also reveals robust short rate expectations and risk premium estimates because of its superior performance in short rate forecasts. It is worth noting that the rolling-window forecasts perform much less favorably. In addition, the predictability of DL-SW is not satisfactory. 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. That is, the common information in

---

[9]The density forecast performance is also reported in Tables 3 and 4, the log-likelihood of DMA is systematically the highest among all forecasting models.

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

	DMA	DMS	TVP	TVP-M	DL	DL-R10	DL-M	DL-SW
MA	One-Month Ahead Relative MSFE							
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
MA	Three-Month Ahead Relative MSFE							
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 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. We report the ratio of each model Mean Squared Forecast Errors (MSFE) relative to Random Walk MSFE, and the preferred values are in bold. The dagger ($\dagger$) 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 three principal components (see Stock and Watson (2002)) of our macro/finance data, using recursive estimations; **RW**: Random Walk.

Table 4: Relative MSFE Performance of Term Structure Models

FH	h=6								h=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 six-month and twelve-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 three principal components (see Stock and Watson (2002)) of our macro-finance data, using recursive estimations; **RW**: Random Walk.

17

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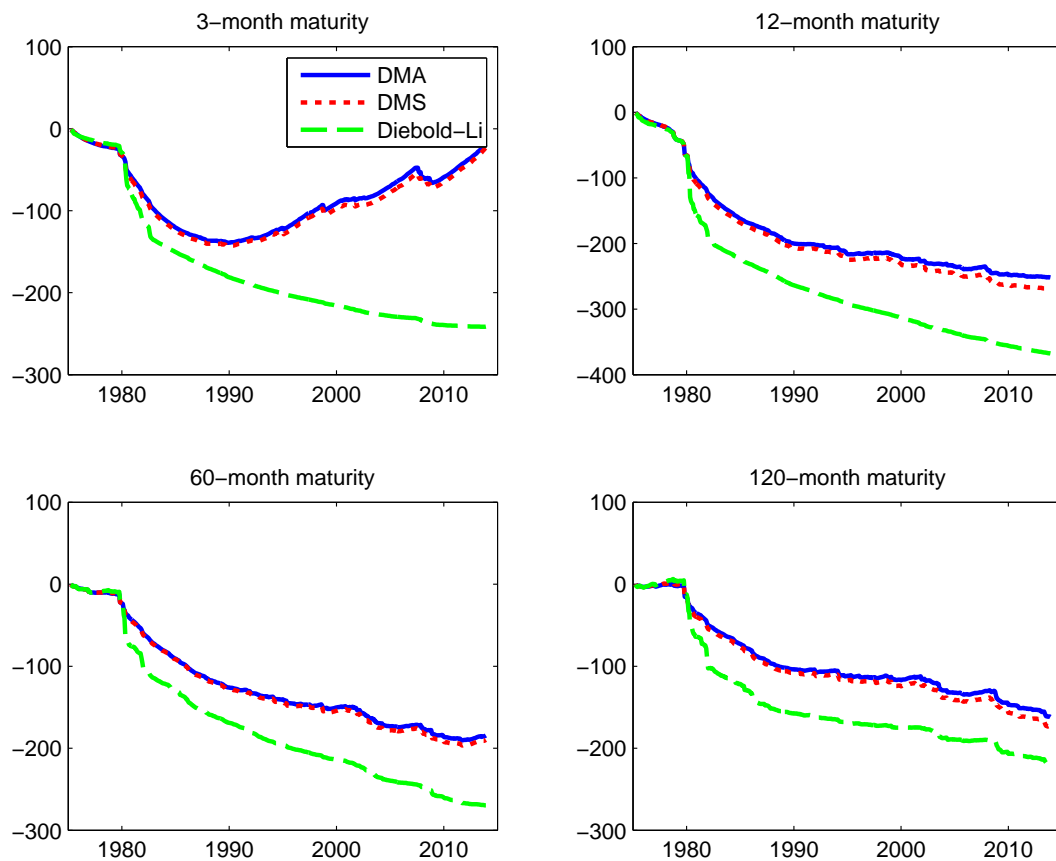
Notes: These are 1-month ahead cumulative sums of predictive log-likelihood for predicted yields from early 1975 to late 2013. From top left clockwise we have maturities of 3, 12, 120 and 60 months. The models are DMA (solid), DMS (dotted) and Diebold-Li (dashed). A higher log-likelihood implies improved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

macro-finance variables may not be useful in forecasting. Hence this result indicates the relative advantages of DMA as a plausible shrinkage method.

In the Nelson-Siegel setup, the long-term yields are almost exclusively driven by the Level factor which is very persistent and has relatively lower volatility, so long-rate forecasts at longer horizons should be quite stable for capable term structure models. For long yields, the forecast performance of a term structure model should be very close to the random walk if the model successfully captures the high persistence as suggested by Duffee (2011a). In contrast, if short yields are anchored by policy rates, this implies short-horizon forecasts of short yields are accurate as long as monetary policy is predictable in the short run. However, without further information, forecasts of short yields at longer forecast horizons deteriorate substantially, given that the monetary policy target or market expectations may shift in the long run. In comparing our results to the existing literature, Diebold and Li (2006) shows the DL beats the RW for forecast horizons up to 12 months before 2000. But Diebold and Rudebusch (2013) and Altavilla, Giacomini and Ragusa (2014) imply NS can no longer beat a RW, which is in line with the increased persistence as we showed previously. Our extended NS model consistently improves upon DL across all horizons and maturities, which is confirmed by Relative MSFEs, predictive log-likelihoods, and the Diebold-Mariano test. Moreover, and at least for shorter horizons, our proposed method improves upon the RW.

Remarks on Predictive Gains Since the pricing dynamics are constrained by the NS restrictions, we can conclude that the predictive gains are purely from the physical dynamics especially by 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 set out the predictive performance of constant-parameter models without stochastic volatility, which are 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 weight on the current observations and hence are robust to parameter uncertainty and structural changes.[10] Moreover, introducing an extra layer of model uncertainty is also essential in improving forecast performance. It helps us properly assimilate macro-finance information in a time-varying manner and more importantly, react to abrupt changes, which parallels the 'scapegoat theory' in Bacchetta and Van Wincoop (2004). From the first two columns in Table 3, we find further improvement over the TVP models if we allow for both parameter

---

[10]Additional results about stochastic volatility can be found in Appendix C.2.

and model uncertainty. Hence, we believe that *the ensemble of these salient features – time-varying coefficients, 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 show results of the Diebold and Mariano (1995) test, in order to evaluate the forecasting performance of DMA relative to DL and TVP-M. The Diebold and Mariano (1995) statistic is also used by Diebold and Li (2006) and Altavilla, Giacomini and Ragusa (2014). The relative MSFE is shown in Table 5 for forecasting horizons 1, 3, 6 and 12 months. These results indicate that the DMA clearly outperforms the DL and TVP-M, not only since MSFE are consistently lower but the differences are statistically significant.

Table 5: MSFE from DMA Relative to Other Models

Maturity	DMA vs. DL				DMA vs. TVP-M			
	$h = 1$	$h = 3$	$h = 6$	$h = 12$	$h = 1$	$h = 3$	$h = 6$	$h = 12$
3	0.833***	0.693***	0.653***	0.843***	0.995	0.906*	0.860*	0.790**
6	0.766***	0.661***	0.655***	0.846***	0.901**	0.865**	0.845**	0.800**
12	1.045	0.824**	0.743***	0.866***	0.961**	0.914**	0.897*	0.847**
24	0.939**	0.788***	0.735***	0.849***	0.943***	0.925**	0.927*	0.890*
36	0.870***	0.774***	0.733***	0.845***	0.952***	0.945**	0.952	0.918
48	0.854***	0.777***	0.740***	0.842***	0.963**	0.959*	0.967	0.934
60	0.864***	0.793***	0.754***	0.844***	0.967**	0.965*	0.973	0.939
72	0.886***	0.815***	0.773***	0.846***	0.965**	0.965*	0.971	0.936
84	0.914***	0.842***	0.794***	0.849***	0.959**	0.960*	0.965	0.928
96	0.947**	0.872**	0.819**	0.851***	0.951**	0.953**	0.955	0.918
108	0.978*	0.904**	0.845**	0.854***	0.945***	0.944**	0.946	0.907
120	1.004	0.936	0.872*	0.860***	0.941***	0.937***	0.937	0.897

Notes: 1. This table reports MSFE-based statistics of DMA forecasts of bond yields at maturities ranging from 3 months to 120 months, relative to the forecasts of Diebold and Li (2006) (DL) or TVP-M (similar to Bianchi Mumtaz and Surico (2009)). 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 DMA model has equal expected square prediction error relative to the benchmark forecasting model (DL or TVP-M) 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.

3.4 Time-Varying Predictability and Macro-Finance Sources

Figure 4 shows six-month ahead Squared Forecasting Errors of DL and DMA across the whole out-of-sample forecast period. It is evident that the DMA significantly and

consistently outperforms the DL across all maturities. We detect a pattern that the predictability of term structure models, DL in particular, tends to be procyclical. The forecast errors are in general higher during periods when economic conditions deteriorate, especially for short-term rates. Economic theories suggest that central banks can influence short rates to achieve policy goals, so the deteriorated predictability implies unexpected or abrupt changes in the behavior of policy makers. For long-term yields, the predictability seems more acyclical, as the movements in long yields are affected not only by short rate expectations but also by the expected risk compensation.

Figure 4: Squared Forecasting Errors for Yields of 3-, 12-, 60- and 120-Month Maturities
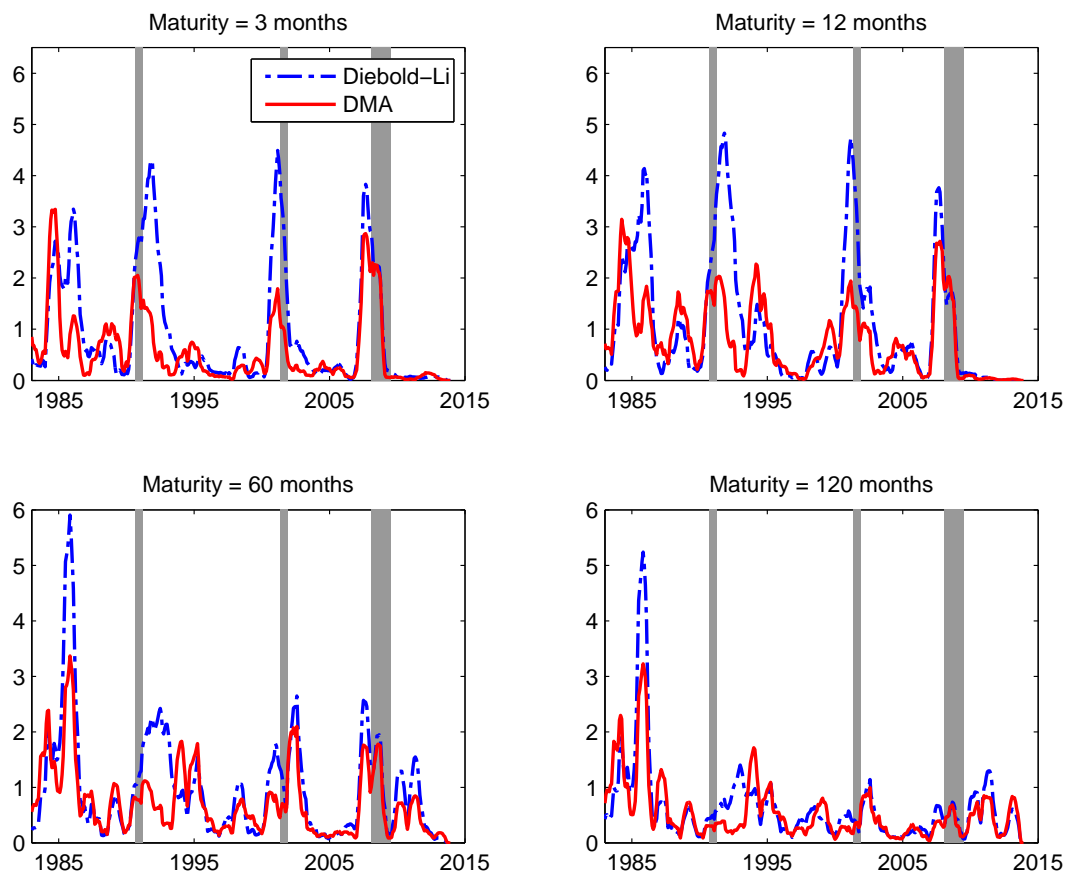
Notes: These are 6 months ahead Squared Forecasting Errors for predicted yields from early 1983 to late 2013. We calculate 9-month moving averages for clarity and plot the statistics for maturities of 3, 12, 60 and 120 months. The models are DMA (solid) and Diebold-Li (dashed and dotted).

As we have discussed earlier, the DL fails to account for a larger information set and parameter instability, which reduces its forecasting performance. Additionally, our approach allows for model uncertainty, and the large macro-finance VAR significantly contributes to the superior performance of DMA during recession periods. It is of importance to include the large-size VAR, as the increase in the weight assigned to this model

significantly reduces forecast errors of DMA when compared with the DL benchmark.[11] Moreover, the DMA has better performance than TVP or TVP-M models especially for short rates as shown in Table 3. As we have discussed, DMA allows the model to capture the sudden changes, which in this case are potentially related to the Fed's policy targets.

We are very interested in why the large-size model has distinctive performance during contraction periods. The question is: What are the underlying economic sources that contribute to the pricing factor movements? Following Koop, Pesaran and Potter (1996) and Diebold and Yilmaz (2014), we conduct the *generalized forecast error variance decomposition* to evaluate the contributions of shocks to respective macro-finance variables.[12] Among 15 variables, our results in Figure 5 suggest that the most important variables driving large-size VAR predictability are indicators of real activity and the stock market. In particular, real activity and stock markets contribute to more than 80% of the 60-month forecast error variance of bond factors during the recent three recessions. There is substantial time variation in the role of these variables, and the contributions of two groups tend to be negatively correlated. Specifically, the economic content of Slope and Curvature factors can be largely explained by real activity since the Great Moderation, but the stock market condition is still indispensable. This observation is in line with Kurmann and Otrok (2013) and Bansal, Connolly and Stivers (2014), but contrasts with the evidence from the UK economy provided by Bianchi, Mumtaz and Surico (2009). In the Nelson-Siegel framework, pricing factors are closely related to short rate expectations and term premia, which we will discuss in details in the following.

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). The Expectations Hypothesis (EH) consistent bond yield $y_t(\tau)^{EH}$ is given by:[13]

$$y_t(\tau)^{EH} = \frac{1}{\tau} \sum_{i=0}^{\tau-1} E_t y_{t+i}(1), \tag{3.1}$$

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

---

[11]The regression results are not shown for the sake of brevity but are available upon request.

[12]We encourage readers to consult the original papers for motivation and background. The generalized variance decomposition is invariant to the ordering of the variables in the VAR, but sums of forecast error variance contributions are not necessarily unity. Here we calculate the normalized weights which add up to unity following Diebold and Yilmaz (2014).

[13]The expectation here is under the physical measure.

Figure 5: Variance Decomposition of Bond Pricing Factors
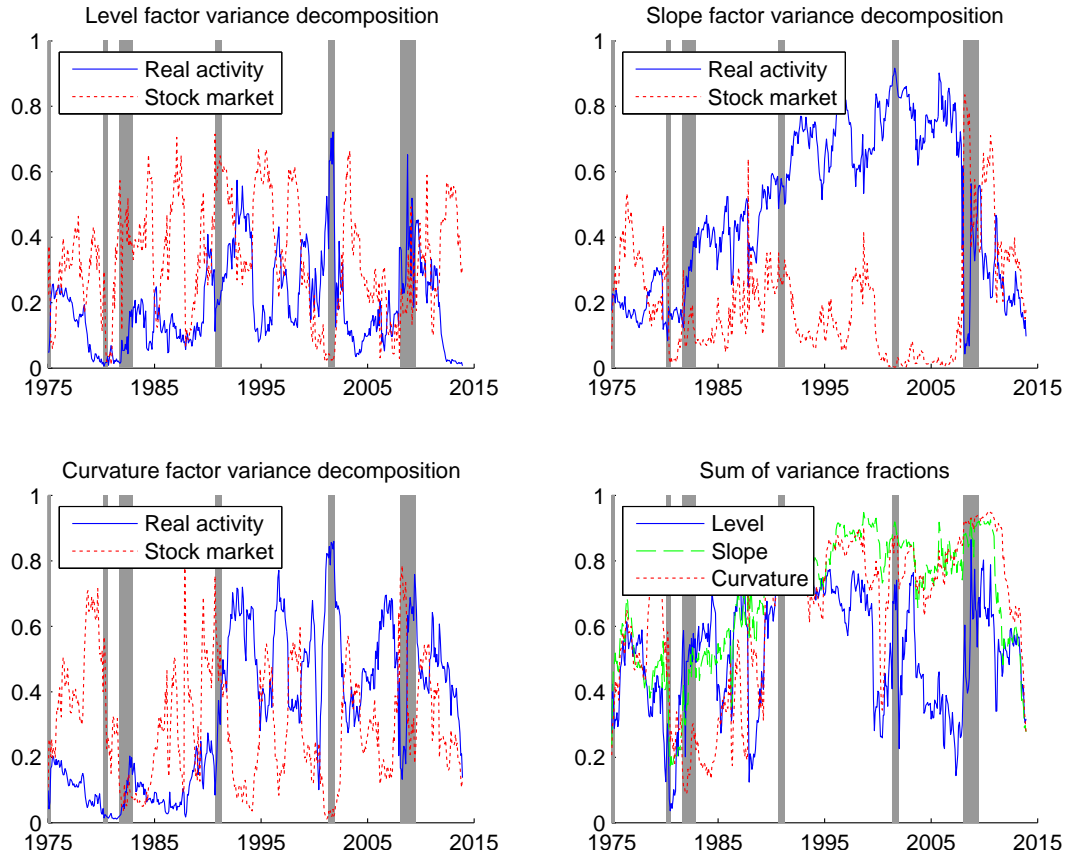
Notes:

1. This figure sets out the generalized forecast error variance decomposition of pricing factors using the large-size VAR model. The upper panels and the bottom left panel show the average contributions of our target variables to the forecast error variance of the respective bond factors over time. At each point in time, the fractions are calculated based on the 60-month forecast error variance. *Real activity* corresponds to the information of Industrial Production Index and Total Industry Capacity Utilization, and *Stock market* corresponds to the information of S&P 500 Stock Price Index and Wilshire 5000 Total Market Index.

2. The lower right panel displays for each pricing factor the sum of the variance fractions of the two groups of target variables shown in the previous panels. The shaded areas are the recession periods based on NBER Recession Indicators.

time-varying term premium is therefore,

$$TP_t(\tau) = y_t(\tau) - y_t(\tau)^{EH}. \tag{3.2}$$

In the large VAR system, both the short rate expectations and the term premia are linear functions of pricing factors and macro and finance variables, see Diebold, Rudebusch and Aruoba (2006). By the linearity of expectation, we can directly employ the generalized variance decomposition for these quantities.

The patterns in variance decompositions displayed in Figure 6 have intuitive appeal, revealing the relative importance of macro-finance variables in driving short rate expectations and risk permia. Standard theory such as the Taylor rule suggests that policy rates should react at least partially to real activity, and our evidence shows short rate expectations are indeed mainly driven by real activity indicators. In contrast, we find that there is strong time variation regarding the main source of risk compensation required by investors, and the underlying sources differ sharply for different horizons. In particular, short-term risk permia is largely explained by real activity shocks during recessions, while long-term risk premia is much less sensitive to real activity during the same periods and more related to the stock market condition in normal times. This observation is interesting but not surprising: As suggested by finance theories, investors' risk attitude influences the demand for bonds and stocks, and Bansal, Connolly and Stivers (2014) show there is a strong link between these two types of assets.

3.5 Model-Implied Term Premia

In this section we set out a visual comparison of our term premium estimates. We plot the 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 7. For comparison, we also plot the model-implied term premia estimated from no-arbitrage term structure models proposed by Kim and Wright (2005), Wright (2011) and Bauer, Rudebusch and Wu (2014), all of which use full-sample data.[14] Note that we use monthly data when applying the methods of Wright (2011) and Bauer, Rudebusch and Wu (2014), and the physical VAR dynamics are all augmented with three macro variables as in our medium-size model in this paper. As a result, the term premium measures from these two methods are similar, which helps resolve a discrepancy indicated in Bauer, Rudebusch and Wu (2014).

---

[14]The comparison between the DMA term premia and recursively estimated term premia from dynamic Nelson-Siegel is shown in Appendix C.4. The 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) tends to overestimate the future short rates and hence underestimate the term premia.

Figure 6: Variance Decomposition of Short Rate Expectations and Term Premia
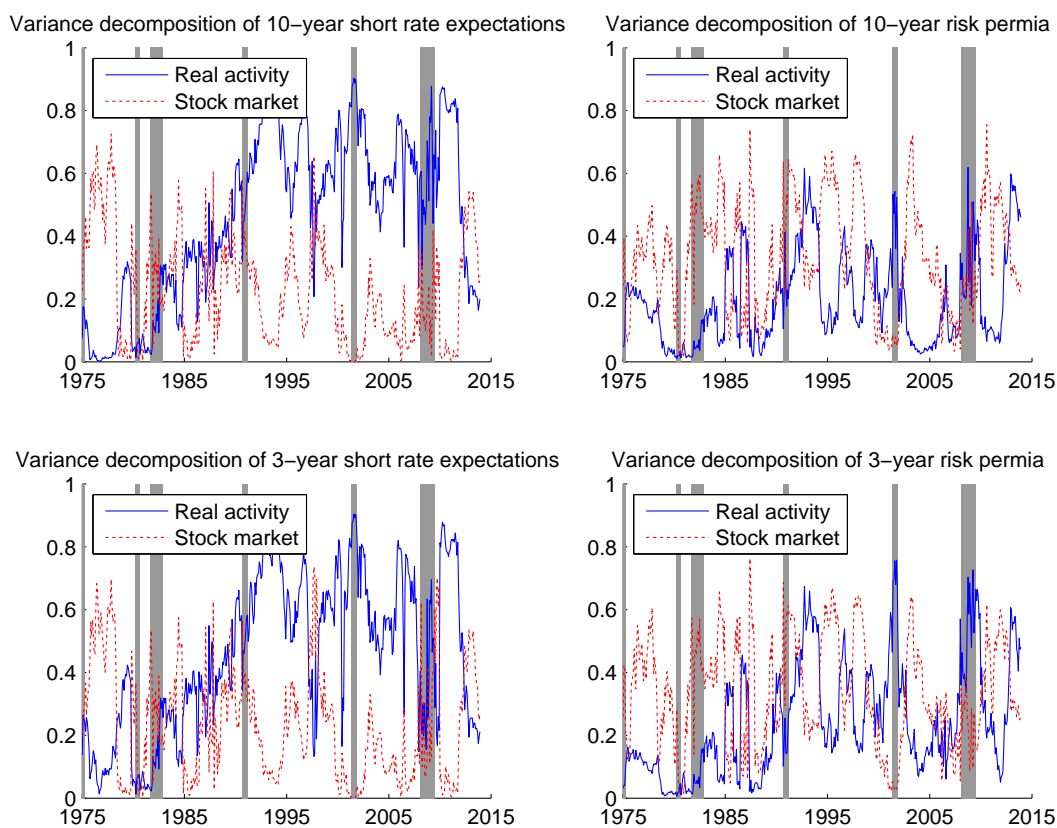
Notes:

This figure sets out the generalized forecast error variance decomposition of short rate expectations and risk premia using the large-size VAR model. The left panels show at each point in time, the average contributions of our target variables to the forecast error variance of 10-year and 3-year short rate expectations, respectively. The right panels show at each point in time, the average contributions of our target variables to the forecast error variance of 10-year and 3-year risk premia, respectively. The time-varying fractions are calculated based on the 60-month forecast error variance. *Real activity* corresponds to the information of Industrial Production Index and Total Industry Capacity Utilization, and *Stock market* corresponds to the information of S&P 500 Stock Price Index and Wilshire 5000 Total Market Index. The shaded areas the recession periods based on NBER Recession Indicators.

It is worth emphasizing that DMA captures plausible term premia using conditional information only. As it is shown in the upper panel of Figure 7, the 36-month term premium estimates of DMA are highly consistent with the full-sample estimates of Wright (2011) and Bauer, Rudebusch and Wu (2014). In general the term premia displays countercyclical behavior, as they rise in and around US recessions, apart from 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), in the KW measure the factor dynamics tend to display distinctively different persistence from other measures because of the augmentation of survey data. According to the observations here, the expected future short rates from the survey tend to be very stable, so the KW term premia has a relatively lower variance and may display an acyclical pattern.

Figure 7: Time-Varying Term Premia of 36-and 120-Month Bonds
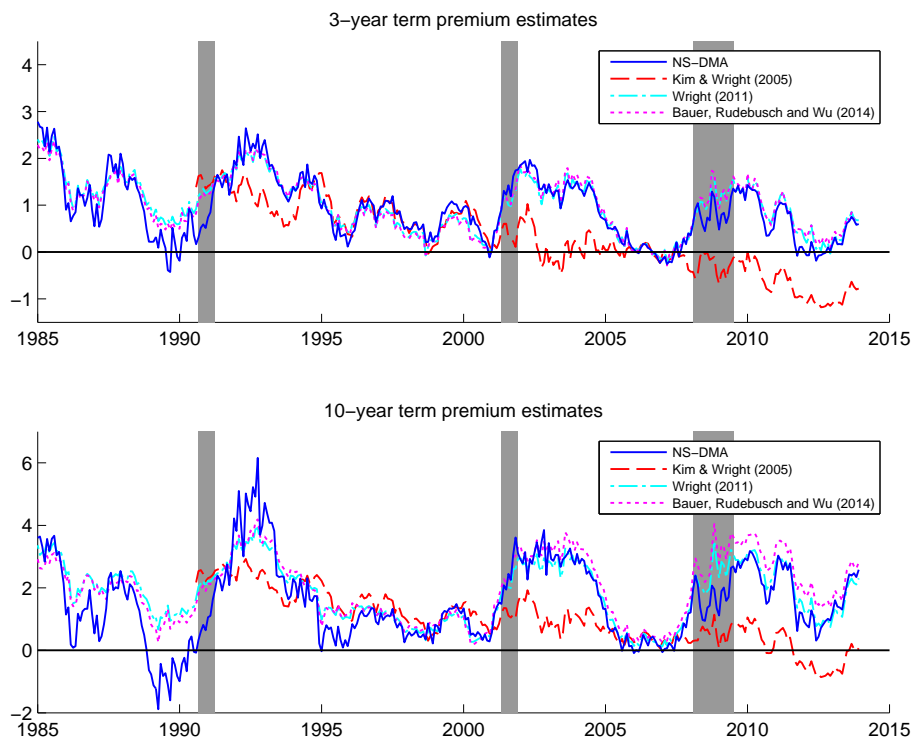
Notes:
1. The top panel is the 36-month term premia and the bottom is the 120-month term premia. The EH consistent 36- and 120-month bond yields are estimated using Eq. (3.1); we then calculate the term premia using Eq. (3.2).
2. In addition to DMA, we use the whole sample to separately estimate two 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 DMA term premia seems to be more sensitive

to changes in the economic environment, which can be seen more clearly from the lower panel of Figure 7 of the long-term term premia. The reason is that expectations of the future short rates move flexibly in DMA and, hence, the 10-year term premia presents a more significant countercyclical pattern. For example, the short rate was continuously decreasing from 1990 to 1993 so the expectation of future short rates was also decreasing. Long rates were relatively stable in contrast, which leads to the increasing risk premia that peaked in 1993.

We can also observe that a divergence between the estimated term premia of DMA and that of 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, 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, the flight-to-quality demand is evident as shown in the graphs. This makes a distinction between the financial crisis and the previous recessions, as global markets are more unified than ever before and hence capital flows to a safe heaven.

The countercyclical pattern of term premia has been identified in previous literature, such as Estrella and Mishkin (1998), Wright (2006), Kim (2009) and Wheelock and Wohar (2009). 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 parallels the evidence shown in Figure 6. In this paper, we present positive evidence that the 'flight-to-quality' demand potentially suppresses the countercyclical pattern of term premia.

4 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 (DMA) approach with the consideration of a large set of macro-finance factors, in order to better characterize the nonlinear dynamics of yield factors and further improve yield forecasts. We explore time-varying predictability of term structure models and unfold the time variation of

economic sources that drive short rate expectations and risk premia. The DMA method significantly improves the predictive accuracy for bond yields, short rates in particular, and successfully identifies plausible dynamics of term premia in real time. We specifically discuss the countercyclical behavior of term premia and reveal a distinct 'flight-to-quality' demand in the recent financial crisis.

To correctly specify the interactions between the yield factors and macro-finance information, realistic specifications are introduced to enhance this model, such as the settings of unspanned macro risks and time-varying parameters, but these assumptions cause econometric challenges in terms of model tractability. These challenges are addressed here by bringing in a fast and simple estimation technique. The proposed 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. Disentangling the real part of the term structure from inflation expectations is meaningful and desirable, but it is beyond the scope of this paper and will be considered for 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.

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.

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. 2016. "Restrictions on risk prices in dynamic term structure models." *Journal of Business & Economic Statistics*, Forthcoming.

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.

Bernanke, Ben, Mark Gertler, and Simon Gilchrist. 1996. "The financial accelerator and the flight to quality." *The Review of Economics and Statistics*, 78(1): 1–15.

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.

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.

Carriero, Andrea, George Kapetanios, and Massimiliano Marcellino. 2012. "Forecasting government bond yields with large Bayesian vector autoregressions." *Journal of Banking & Finance*, 36(7): 2026–2047.

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. 2016. "Information in the term structure of yield curve volatility." *Journal of Finance*, 71(3): 1393–1436.

Coroneo, Laura, Domenico Giannone, and Michele Modugno. 2015. "Unspanned macroeconomic factors in the yield curve." *Journal of Business and Economic Statistics*.

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.

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

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, 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 Glenn 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 Kamil Yilmaz. 2014. "On the network topology of variance decompositions: Measuring the connectedness of financial firms." *Journal of Econometrics*, 182(1): 119–134.

Diebold, Francis X., and Roberto S. Mariano. 1995. "Comparing Predictive Accuracy." *Journal of Business and Economic Statistics*, 13(3): 253–263.

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. 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)*, 58(3): 609–617.

Feunou, Bruno, Jean-Sébastien Fontaine, Anh Le, and Christian Lundblad. 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.

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.

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

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 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.

Koop, Gary, M. Hashem Pesaran, and Simon M. Potter. 1996. "Impulse response analysis in nonlinear multivariate models." *Journal of Econometrics*, 74(1): 119–147.

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., and Jose Scheinkman. 1991. "Common factors affecting bond returns." *The Journal of Fixed Income*, 1(1): 54–61.

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.

Moench, Emanuel. 2008. "Forecasting the yield curve in a data-rich environment: A no-arbitrage factor-augmented VAR approach." *Journal of Econometrics*, 146(1): 26–43.

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.

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

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

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.

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

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

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.

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.

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.

Data Appendix

<div align="center">

Table 6: List of Yields and Macro-Finance Variables

</div>

Series ID	Description
TB	3- and 6-month Treasury Bills (Secondary Market Rate) [1]
ZCY	Smoothed Zero-coupon Yield from Gürkaynak, Sack and Wright (2007) [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 [1]
SP	S&P 500 Stock Price Index, End of Month [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 [5]
MSP	Median Sales Price for New Houses Sold in the United States [5]
TWX	Trade Weighted U.S. Dollar Index: Major Currencies [1]
ED	Eurodollar Spread: 3m Eurodollar Deposit Rate - 3m Treasury Bill Rate [1]
WIL	Wilshire 5000 Total Market Index [5]
DYS	Default Yield Spread: Moodys BAA-AAA [1]
NFCI	National Financial Conditions Index [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 when appropriate.
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 an $n \times 1$ 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 that all the data from time 1 to t denoted as D_t, the Bayesian solution to updating about the coefficients β_t takes the form

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

where $\wp$ is the support of β_{t-1}. 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))[15]:

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

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

---

[15]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 $S_t = \delta S_{t-1} + n_t^{-1}\left(S_{t-1}^{0.5} V_{t-1}^{-0.5} \widetilde{v}_{t|t-1} \widetilde{v}_{t|t-1}' V_{t-1}^{-0.5} S_{t-1}^{0.5}\right)$. 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 Σ_t is approximately equivalent numerically to the Exponentially Weighted Moving Average (EWMA) filter $\hat{\Sigma}_t = \delta \hat{\Sigma}_{t-1} + (1 - \delta) v_t v_t'$. 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

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

where $L_{t-1} = k$ means the k_{th} model[16] is selected and $p(\beta_{t-1}^{(k)}|L_{t-1} = k, D_{t-1})$ is given by the Kalman filter (Eq. A.1). To simplify notation, let $\pi_{t|s}^{(l)} = \mathbf{Pr}(L_t = l|D_s)$.

The model updating equation is

$$\pi_{t|t}^{(i)} = \frac{\pi_{t|t-1}^{(i)}p^{(i)}(z_t|D_{t-1})}{\sum_{l=1}^{K} \pi_{t|t-1}^{(l)}p^{(l)}(z_t|D_{t-1})}, \tag{A.5}$$

where $p^{(i)}(z_t|D_{t-1})$ is the predictive likelihood. Raftery, Kárnỳ and Ettler (2010) used an empirically sensible simplification that

$$\pi_{t|t-1}^{(i)} = \frac{\left(\pi_{t-1|t-1}^{(i)}\right)^{\alpha}}{\sum_{l=1}^{K} \left(\pi_{t-1|t-1}^{(l)}\right)^{\alpha}}, \tag{A.6}$$

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.

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

---

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

Model Averaging, which average the predictions of all models with respective probabilities.

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 8. 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 fell 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.

Figure 8: Nelson-Siegel Factor Dynamics
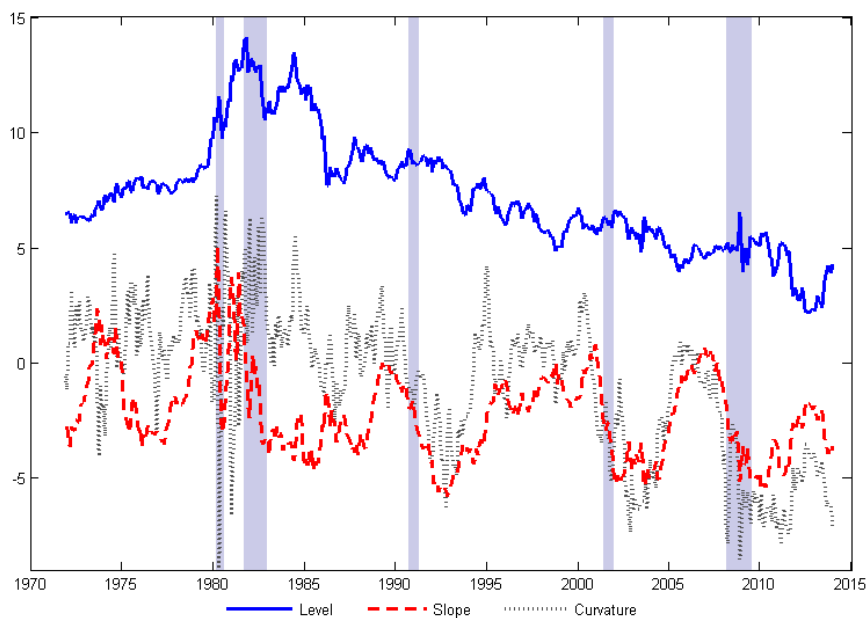
Notes: The graph shows the Nelson-Siegel Level, Slope and Curvature factors, which are drawn from Eq. (2.1). The shaded areas are recession periods according to the NBER Recession Indicators.

The Slope factor tends to fall sharply within recession periods, as indicated in Figure 8 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.

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).[17] 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.

---

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

Appendix C   Additional Results

C.1   Forecasting Results

Figure 9: DMA Forecasts of Yields
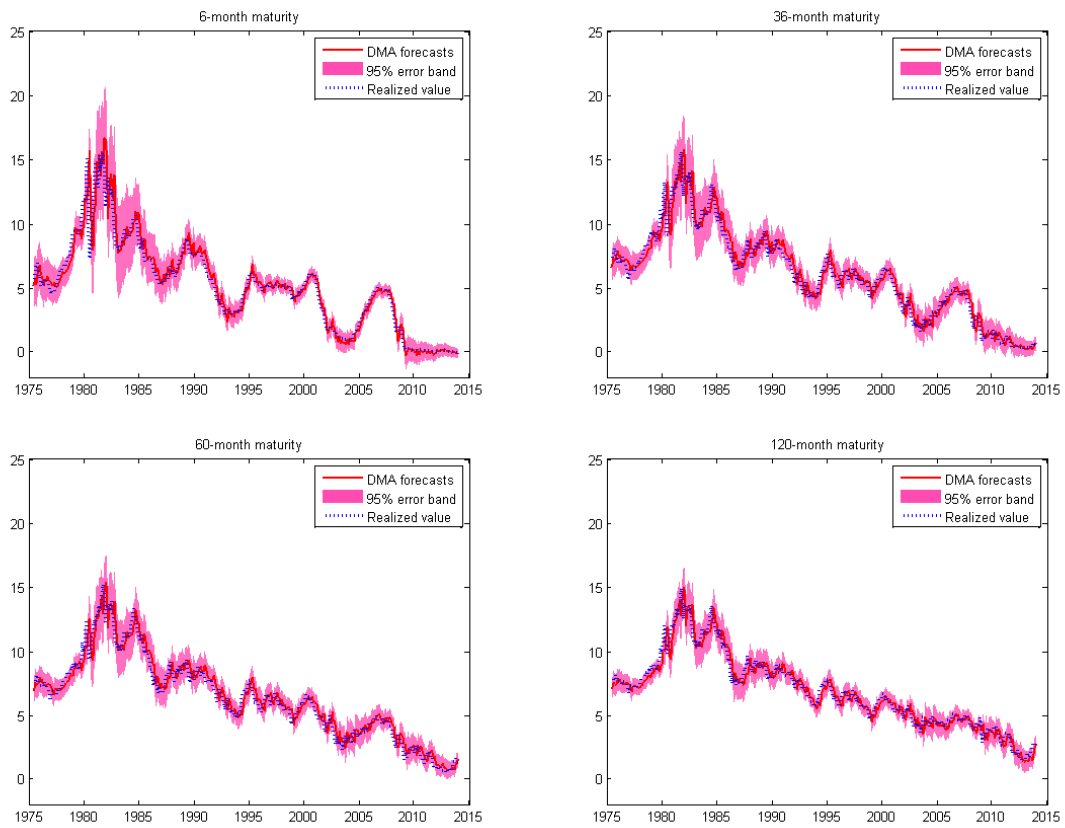
Notes: These are 3 months ahead forecasts (95% error band) for yields against realized values with maturities 6, 36, 60 and 120 months, from early 1975 to late 2013. The forecasts are two-step forecasting using DMA, which can be summarized by Eq. (2.1), (2.3) and (2.4).

Table 7: Relative MAFE Performance of Term Structure Models

FH	h=1								h=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 three principal components (see Stock and Watson (2002)) of our macro-finance data, using recursive estimations; **RW**: Random Walk.

Table 8: Relative MAFE Performance of Term Structure Models (Continued)

FH	h=6								h=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 six-month and twelve-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 at any point in time. **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 three principal components (see Stock and Watson (2002)) of our macro-finance data, using recursive estimations; **RW**: Random Walk.

44

C.2 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 (2016) 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 bond factors can be modeled by specifying an appropriate time-varying volatility. The DMA model allows for heteroskedastic variances and this assumption is crucial for its good density forecast performance; this evidence is consistent with Hautsch and Yang (2012).

The DMA not only provides more sensible results in terms of density forecasts, but also captures the desirable evolutionary dynamics of the economic structure. Figure 10 shows the time-varying second moments of 3 month ahead forecasts from the DMA model. The figure displays distinct time variation in the evolution of volatility. The stable decline of volatility before the financial crisis matches the conclusions of Bianchi, Mumtaz and Surico (2009), who refer to this empirical result as the 'Great Moderation' of the term structure. We observe that yields with longer maturities have lower volatilities. This feature is counter-intuitive. Theoretically, long yields are mainly driven by three components: the expected future (real) short yields; inflation expectations; and the term premia. Inflation expectations may change abruptly and frequently during a short period of time, so do the expected future short yields. At the same time, term premia can also be quite volatile. Therefore, summing up the movements of these three components, the variance of long yields should be larger than the short yields; nevertheless, the empirical result implies the opposite. As indicated in Duffee (2011b), 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, thus, offsetting the variation in these components.

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

Figure 10: Time-Varying Second Moment
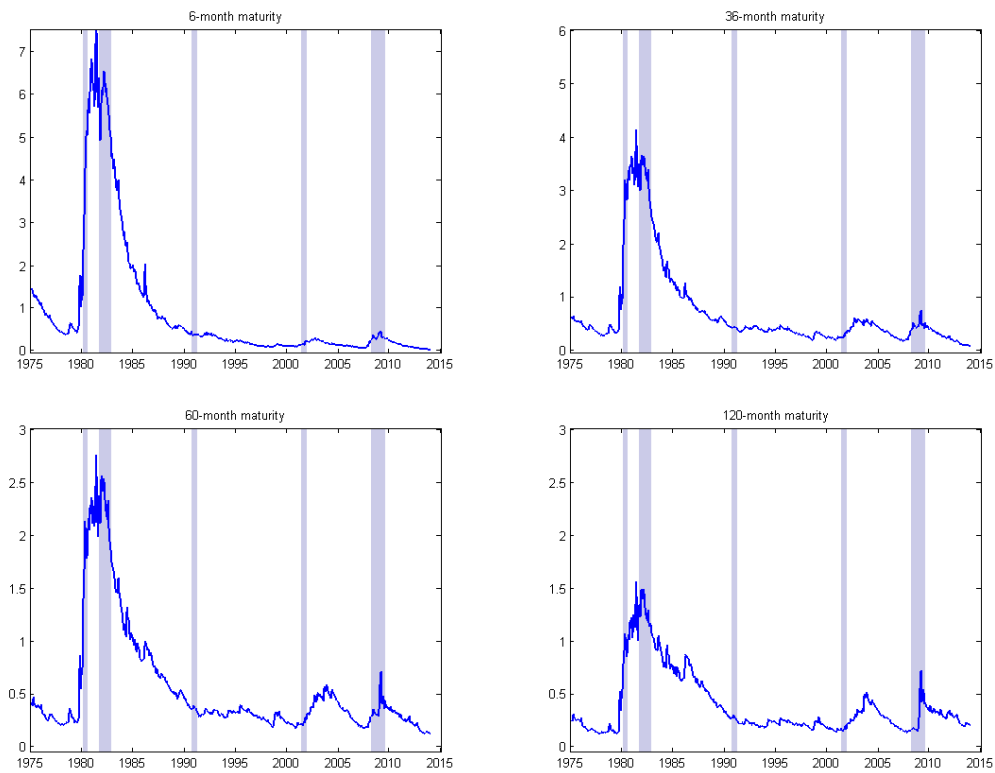
Notes: These are time-varying second moments of 3 months ahead forecasts for bonds at maturities 6, 36, 60 and 120 months, from early 1975 to late 2013. The variance of NS factors is estimated from Eq. (A.2), and then the variances of yield forecasts generated by each candidate model in the DMA, can be easily calculated as linear combinations of factor variances.

volatilities 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 the 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.

C.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 DMA approach does not explicitly impose 'hard' arbitrage-free restrictions. 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. The reason is that our focus here is not on the dynamic structure of market price of risks. 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 expectations of investors, we aim to improve forecasts of the interest rate term structure. Joslin, Singleton and Zhu (2011) show that no-arbitrage cross-sectional restrictions are irrelevant for out-of-sample forecasts if the factor dynamics are unrestricted. 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).

To ensure the robustness of our 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 DMA. The key difference between arbitrage-free DMA and DMA is a 'yield-adjustment term', which only depends on the maturity and factor volatility. See Christensen, Diebold and Rudebusch (2011) and

Diebold and Rudebusch (2013) for more details. The forecast performances of two models are very close, implying that the DMA is almost arbitrage-free, which is consistent with theoretical evidence in Feunou et al. (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.

C.4 Term Premia of Diebold-Li and DMA

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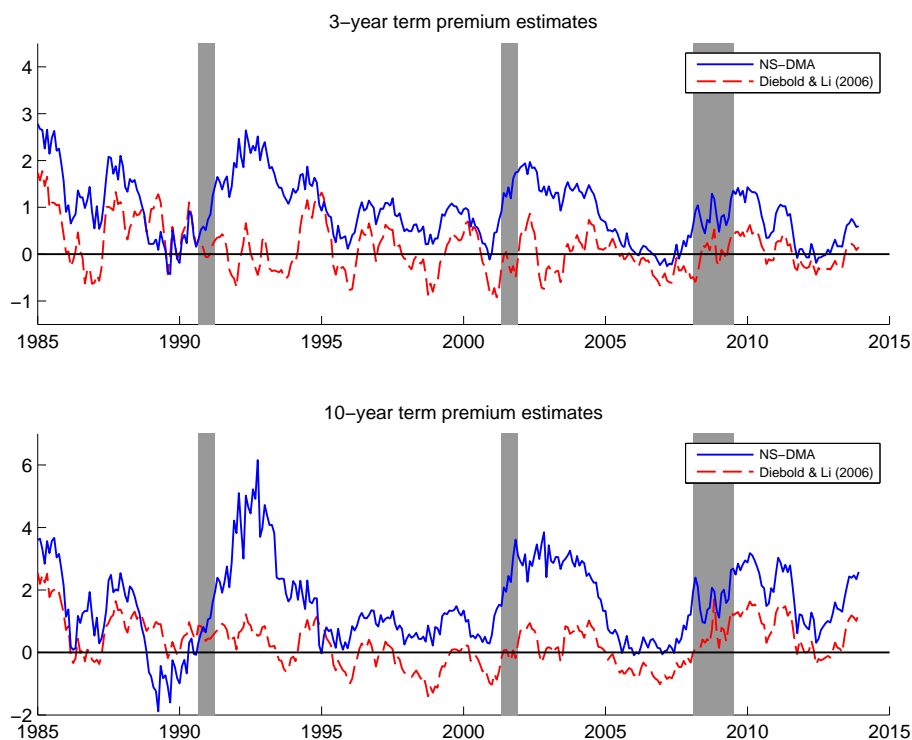
Notes:

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 bond yields are estimated using Eq. (3.1); we then calculate the term premia using Eq. (3.2).

2. In addition to 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.