Accepted Manuscript

Forecasting the term structure of government bond yields in unstable environments

Joseph P. Byrne, Shuo Cao, Dimitris Korobilis

PII: S0927-5398(17)30080-4
DOI: https://doi.org/10.1016/j.jempfin.2017.09.004
Reference: EMPFIN 997

To appear in: *Journal of Empirical Finance*

Received date : 20 June 2016
Revised date : 12 June 2017
Accepted date : 26 September 2017

Please cite this article as: Byrne J.P., Cao S., Korobilis D., Forecasting the term structure of government bond yields in unstable environments. *Journal of Empirical Finance* (2017), https://doi.org/10.1016/j.jempfin.2017.09.004

Forecasting the Term Structure of Government Bond Yields in Unstable Environments*

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

Sunday 1st October, 2017

Abstract

In this paper we model and predict the term structure of US interest rates in a data-rich and unstable environment. The dynamic Nelson-Siegel factor model is extended to allow the model dimension and the parameters to change over time, in order to account for both model uncertainty and sudden structural changes, in one setting. The proposed specification performs better than several alternatives, since it incorporates additional macro-finance information during hard times, while it allows for more parsimonious models to be relevant during normal periods. A dynamic variance decomposition measure constructed from our model shows that parameter uncertainty and model uncertainty regarding different choices of predictors explain a large proportion of the predictive variance of bond yields.

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

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

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

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

‡Research Institute, Shenzhen Stock Exchange, Shenzhen, China. Email: shuo.cao@outlook.com.

§Essex Business School, University of Essex, Colchester, UK. Email: d.korobilis@essex.ac.uk.

1 Introduction

Modeling the term structure of interest rates using risk factors is a vast and expanding research frontier in financial economics.[1] Diebold and Li (2006) propose a dynamic Nelson-Siegel (NS) model in order to predict the yield curve.[2] However, Altavilla, Giacomini and Ragusa (2014) indicate that the original version of the dynamic NS model without macro information has weaker predictive power in the past twenty years. Various attempts have been made to tackle the weakening predictive power of term structure models. While Ang and Piazzesi (2003) and Diebold, Rudebusch and Aruoba (2006) stress the importance of key macro variables for the yield curve modeling, Dewachter and Iania (2012) and Dewachter, Iania and Lyrio (2014) show financial factors to be more prominent for explaining yields. This evidence is supported by other researchers, such as Moench (2008), who show that a term structure model augmented with a broad macro-finance information set can provide superior forecasts.

That said, the majority of term structure studies rely on fixed information sets, that is preselection of all possible predictors, and they rarely question whether the introduction of different predictors *per se* can become a source of forecast uncertainty as suggested in Dangl and Halling (2012). Moreover, it is well documented by many econometricians that economic and financial predictors have short-lived predictive content and econometric relationships are unstable; see the general review by Rossi (2013), as well as Dangl and Halling (2012), Koop and Korobilis (2012) and Byrne, Korobilis and Ribeiro (2018) for examples pertaining to stock return, inflation and exchange rate predictability, respectively. In this paper, therefore, we raise and attempt to answer the following questions: To what extent does the consideration of various predictors proliferate forecast uncertainty of term structure models? How do we improve bond yield forecasts when predictors are possibly unstable and short-lived? To answer these questions, our paper builds upon previous work and proposes a modeling framework for term structure forecasting that has several salient features.

Firstly, we incorporate financial information in addition to traditional macro variables, as the global financial crisis highlighted the importance of the financial market for macroeconomic activity and bond yields more generally. We incorporate a substantial range of macro-finance risk factors with model combination techniques that distil large datasets.[3] Estimating a large vector autoregressive model (VAR) with macroeconomic

[1]See Piazzesi (2010), Gürkaynak and Wright (2012), Duffee (2013) and Diebold and Rudebusch (2013) for extensive reviews.

[2]They use three pricing factors to capture most of the variation in bond yield data, which has been well established in Nelson and Siegel (1987) and Litterman and Scheinkman (1991).

[3]Allowing for a large macro-finance information set fully accounts for, and extends, the point of

and financial factors is a non-trivial task due to the proliferation of parameters in large dimensions, see Carriero, Kapetanios and Marcellino (2012) and Coroneo, Giannone and Modugno (2015). Therefore, following Koop and Korobilis (2013) a Bayesian shrinkage estimation methodology is adopted, in order to estimate our large system with many variables. Secondly, we employ time-varying parameter (TVP) VARs to fully capture different degrees of structural change and both coefficient instability and stochastic volatility are taken into account. Van Dijk et al. (2014) suggest that a term structure model considering time-varying parameters can significantly improve its predictive power. In the same spirit, our forecasting exercises are conducted in a TVP dynamic Nelson-Siegel setup that extends Bianchi, Mumtaz and Surico (2009).

Our key methodological contribution is to adopt a model averaging methodology that can potentially mitigate forecast uncertainty originating from different choices of predictors. The proposed Dynamic Model Averaging (DMA) method can determine in a data-based manner which macroeconomic or financial risks are relevant for the yield curve at each point in time. That is, unlike traditional model averaging approaches that select important predictors that are relevant during the whole data sample[4], DMA is a probabilistic framework that allows at each point in time different predictors to be relevant for forecasting. Our application of DMA extends Koop and Korobilis (2013) from the VAR setting to a more general factor-augmented system for term structure modeling, which is new to the literature. We choose at different points in time between three candidate models: i) one with three pricing factors only; ii) pricing factors plus three key macroeconomic indicators; and iii) pricing factors augmented using up to fifteen macro and financial factors. The third macro-finance model is like a 'kitchen sink' model allowing for much more macro-finance information to be incorporated in the spirit of Moench (2008). Model probabilities are assigned to each of the models at each point in time and, thus, averaging is dynamically implemented. When compared to alternative time-varying parameter estimation methodologies, 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 conditional on the most recent information, and Elliott and Timmermann (2008) indicate that model averaging methods in general can reduce the total forecast risk associated with using only a single 'best' model.

Moreover, our setup allows a quantitative evaluation of various sources of forecast uncertainty, by employing an informative variance decomposition following Dangl and Halling (2012). We quantify the relative importance of parameter uncertainty and model uncertainty in terms of different predictors, and show that both parameter uncertainty

Ludvigson and Ng (2009) that large datasets can improve forecasting power.

[4]See Bauer (2016) for an example of Bayesian model averaging in a static setup.

and model uncertainty are important and in total account for one third of predictive variance. Therefore, when choice of predictors is uncertain, DMA that builds upon TVP models is promising in assimilating macro-finance information dynamically.

We examine empirically U.S. term structure dynamics using over fourty years of monthly observations. 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 substantial gains in yield predictability are due to the ensemble of salient features – time-varying parameters and dynamic model averaging. The contribution of each feature we incorporate in our model is significant and time-varying, and we find 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.[5] 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).

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 3.4 displays the point and density forecasting performance of our term structure model. Section 3.6 shows the model-implied term premia has informative economic implications. Section 4 concludes.

2 Methods

2.1 Cross-Sectional Restrictions and Yield Factor Dynamics

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

[5] The indicators of real activity and the stock market are helpful in explaining the movements, see Appendix C.5. 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.

yields at maturity τ, then the factor model we use is of the form:[6]

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

where L_t^{NS} is the "Level" factor, S_t^{NS} is the "Slope" factor, C_t^{NS} is the "Curvature" factor and $\varepsilon_t(\tau)$ is the error term. In the formulation above, λ^{NS} is a parameter that controls the shapes of loadings for the NS factors. For estimation purposes, we can rewrite equation (2.1) in the equivalent 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 restrictions that apply on the cross-section of yields. We use simple ordinary least squares (OLS) to extract three NS factors, and following Diebold and Li (2006), Bianchi, Mumtaz and Surico (2009) and Van Dijk et al. (2014), we set $\lambda^{NS} = 0.0609$. 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).[7]

Time-series dynamics An important and novel aspect of our methodology is in modeling the factor dynamics. 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

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

[6]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.

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

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 additional restrictions on the VAR system above. In our framework instead macro-finance 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 imposed on the coefficients of macro-finance variables in the cross section, while the time-series dynamics (VAR model) are left unconstrained.[8]

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

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

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

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

based upon a prior β_0 discussed below, and $\mu_t \sim N\left(0, Q_t\right)$. Following Koop and Korobilis (2013) we set $Q_t = \left(\Lambda^{-1} - 1\right) cov\left(\beta_{t-1} | \mathcal{D}_{t-1}\right)$ where $\mathcal{D}_{t-1}$ denotes all the available data at time $t-1$ and scalar $\Lambda \in (0, 1]$ is a 'forgetting factor' discounting older observations. The covariance matrix Σ_t evolves according to a Wishart matrix discount process (Prado and West (2010)) of the form:

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

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

[8] Bauer and Rudebusch (2017) test the knife-edge restrictions and point out these restrictions, though statistically rejected, have only small effects on cross-sectional fit and estimated term premia. Therefore, we follow Joslin, Priebsch and Singleton (2014) and Coroneo, Giannone and Modugno (2015) for the tractability of our proposed model.

a specific function of the squared residuals of our model and explained in the Appendix A.1. The pair of forgetting/decay factors (Λ, δ) can be interpreted as a Bayesian prior on the amount of time-variation expected in the drifting coefficients β_t and the volatilities Σ_t, respectively. Following recommendations in Koop and Korobilis (2013) and Dangl and Halling (2012), we set the forgetting factor $\Lambda = 0.99$ and the decay factor $\delta = 0.95$.

To sum up, we have specified a VAR with drifting coefficients and stochastic volatility which allows for modeling structural instability and regime changes in the joint dynamics of the NS factors and the macroeconomic and financial factors. The Nelson-Siegel restrictions allow for straightforward estimation and interpretation, and shift our focus on the time-series dynamics of the estimated factors. Given ample empirical evidence about the importance of structural change and volatility in forecasting macroeconomic and financial variables,[9] our approach is more pragmatic: by specifying a flexible VAR structure with time-varying parameters and stochastic volatility, as well as relevant macro-finance information that can switch over time, we expect to obtain better forecasts of the underlying yield curve.

2.2 Model Uncertainty and the Role of Macro-Finance Factors

This paper argues that the possible set of risk factors relevant for characterizing the yield curve can change over time. We focus on Eq. (2.3) and work with three possible information sets: 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 inflation and Industrial Production, so $q = 3$. The large (NS + macro-finance) model includes $q = 15$ macroeconomic and financial variables.

Having three candidate models $i = 1, 2, 3$, 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 obtain the probability/weight for each model i

$$\pi_t^{(i)} = f\left(L_{t-1} = i | D_{t-1}\right)$$

under the regularity conditions $\sum_{i=1}^{K} \pi_t^i = 1$ and $\pi_t^i \in [0, 1]$, and where L_{t-1} is the model selected 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

[9]See for example, Dangl and Halling (2012), Koop and Korobilis (2012), Van Dijk et al. (2014) and Byrne, Korobilis and Ribeiro (2018).

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 subsequently show this can capture changing economic conditions and increase the predictive performance.

3 Empirics

3.1 Data

The smoothed yields provided from the US Federal Reserve by Gürkaynak, Sack and Wright (2007) are used in the term-structure model specified in the previous section. We also include 3- and 6-month Treasury Bills (Secondary Market Rate). The empirical analysis focuses on yields with maturities of 3, 6, 12, 24, 36, 48, 60, 72, 84, 96, 108 and 120 months. The key macroeconomic and financial variables that enter our Dynamic Model Averaging model are obtained from St. Louis Federal Reserve Economic Data (FRED). These include inflation, real activity indicators, monetary variables, as well as the stock market, exchange rate, house prices and other financial market indicators; the details can be found in the Data Appendix. The full sample is from November 1971 to November 2013 and we use end of the month yield data. 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 VAR sizes entering the dynamic model averaging framework. 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.

Table 1: Descriptive Statistics of Bond Yields

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

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

3.2 Evidence on Parameter Instability

In this section we seek to motivate 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 proposed test statistic 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.

Nevertheless, such tests of parameter instability are in-sample tests and fail to provide evidence concerning structural instability and predictability out-of-sample. Instead, we follow a different strategy and we note that the constant parameter version of the Nelson-Siegel model can be obtained as a special case of our proposed time-varying specification.[10] 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 pseudo out-of-sample predictive power between the TVP-VAR and a constant parameter VAR. We employ the test proposed by Diebold and Mariano (1995) and evaluate the predictive power of competing models across four forecast horizons ($h = 1, 3, 6, 12$ months) and for all of our maturities. The p-values of the tests are reported in Table 2, which correspond to the test of the null hypothesis that the competing TVP-VAR model has equal expected square prediction error relative to the benchmark forecasting model constant parameter VAR (i.e. Diebold and Li (2006)), against the alternative hypothesis that the competing TVP-VAR forecasting model has a lower expected square prediction error than the benchmark forecasting model. Table 2 indicates the TVP-VAR consistently outperforms the constant parameter VAR. The test statistic rejects the null for most of the maturities, and especially at longer forecast horizons, so the time-varying parameter model should be preferred as it can provide more robust estimates.

To highlight the importance of the TVP feature, Figure 1 sets out the time-varying persistence of factor dynamics in the small-size VAR. 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

[10] 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 coefficients and constant variance.

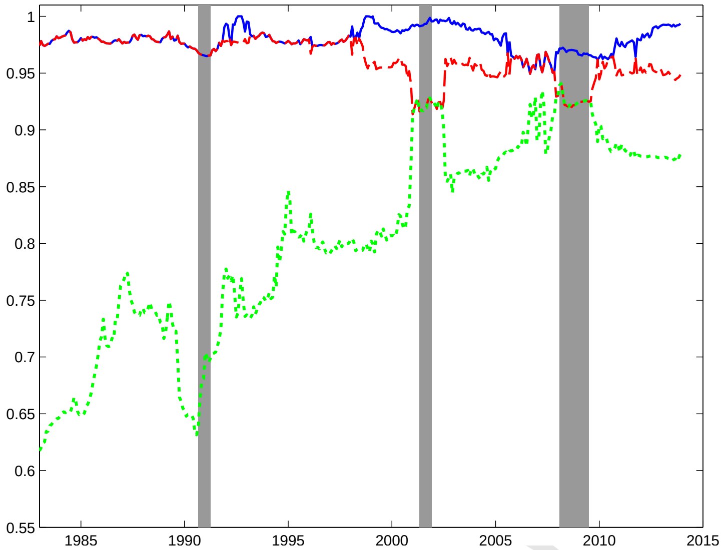
Table 2: Parameter Instability Test

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 1983:11 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 (that is, a model identical to the Diebold and Li (2006) model), against the alternative hypothesis that the competing TVP-VAR model has a lower expected square prediction error than the benchmark forecasting model.

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 Factor Dynamics



Notes: The graph shows the largest three eigenvalues of the factor time-series dynamics in the small-size TVP model. The shaded areas are recession periods according to the NBER Recession Indicators.

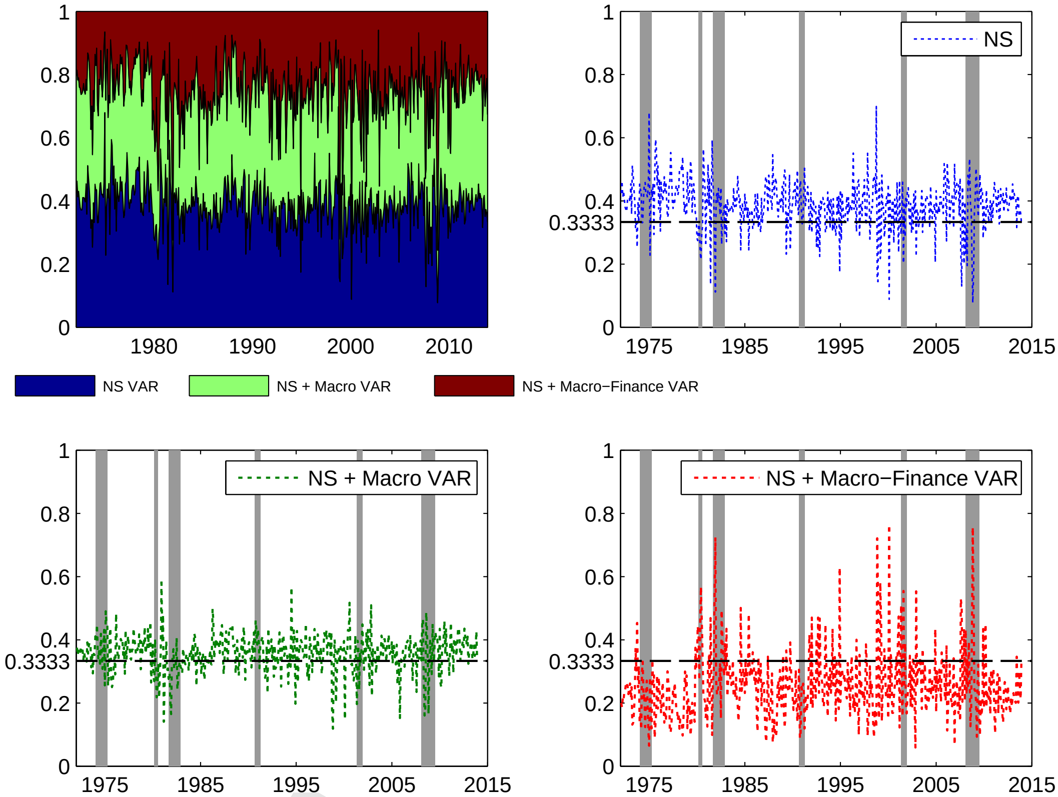
3.3 Model Dynamics

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.

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 sets allows 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 reacts 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 forecasting performance, as the true dynamics may not be captured well during certain periods. The DMA approach encompasses moderate to sudden changes in the economic environment and accordingly is promising in producing consistently superior forecasting

Figure 2: Model Weights for TVP, TVP-M and TVP-L 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 TVP(NS VAR), the TVP-M (NS + Macro VAR) and the TVP-L (NS + Macro-Finance VAR), respectively. The shaded areas are the recession periods based on NBER Recession Indicators.

performance, a point we discuss in detail in the following sections.

3.4 Forecasting Performance

At each time, we generate forecasts using the proposed DMA model with the data only up to that period. As described in the methodology section, we obtain the full posterior distribution of state variables and parameters, and the predictive density of yields, using only data up to time t. Specifically, the out-of-sample forecasts of yields are generated in the following two-step procedure.

The first stage is using the Kalman filter and DMA to generate predictions of the three Nelson-Siegel yield factors with macro variables. That is, at each time, β_t is estimated using the Kalman filter and Σ_t is estimated according to the forgetting factor method. For each of the three candidate models (the small, meddle or large), we use Eq. (2.3) with the predicted $\hat{\beta}^{(i)}_{t+1} = \beta^{(i)}_t$ to forecast pricing factors. Also at each time a model averaging method is recursively implemented to generate a weighted average of the forecasts by three candidate models, as decried in Section 2.2. The second stage is forecasting the yields with the predicted NS factors and the fixed NS loadings. This step is straightforward as bond yields are just linear combinations of three NS factors. The macro variables are not directly used to predict the yields in the second step, because of the knife-edge cross-sectional restrictions.

We use a training sample from 1971:11 to 1983:10, which gives a long forecast evaluation period from 1983:11 to the 2013:11. To thoroughly investigate the predictive power of our proposed model, we produce out-of-sample monthly forecasts for 30 years, and the predictive horizons range from one month to twelve months. 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 a 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 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) forecast, which allows us to evaluate whether the term structure models successfully capture the high persistence in bond yields.

The point forecast precision of all term structure models is assessed in Table 3 which displays the one and three period ahead Mean Squared Forecast Error (MSFE) of the competing specifications relative to the MSFE of the random walk. Lower values of the relative MSFEs indicate better performance in general, while values lower than one indicate better performance relative to the random walk in particular. The core empirical results are very encouraging for the proposed method. As can be seen in this table, the DMA model consistently outperforms all the benchmark models. In many cases the relative MSFEs are also below one, meaning that the model fairs well compared to the random walk. Nevertheless, in some cases the relative MSFEs are higher than one, or they are not substantially lower than one in order to be considered significant.[11] Table 4 shows relative MSFEs for horizons of six and twelve months ahead and it reveals that the DMA is also preferred at relatively long forecast horizons. Moving from point to density forecasts, that is forecasts that take into account the uncertainty in the predictive distribution, we find that also DMA performs well. In Tables 3 and 4 we find that DMA has the highest value of the log predictive score; see Geweke and Amisano (2010) for a definition of this metric. Given that DMA is outperforming all other specifications in all instances on average using this measure, we simply denote this in the two tables by using the symbol †. Figure 3 goes one step further and instead of quoting the average predictive log-scores, it plots the cumulative sum of predictive log-scores during the whole evaluation period, for selected maturities and for the DMA/DMS and Diebold-Li approaches. The predictive density of the DMA specification 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

[11] Significance of relative MSFEs can also be measured using the Diebold and Mariano (1995) statistic we used in Table 2 to measure the predictive differences between constant and time-varying parameter specifications.

Table 3: One and three month ahead relative MSFEs of competing term structure models

	DMA	**DMS**	**TVP**	**TVP-M**	**DL**	**DL-R10**	**DL-M**	**DL-SW**
Maturity	ONE MONTH AHEAD RELATIVE MSFE ($h = 1$)							
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
Maturity	THREE MONTH AHEAD RELATIVE MSFE ($h = 3$)							
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

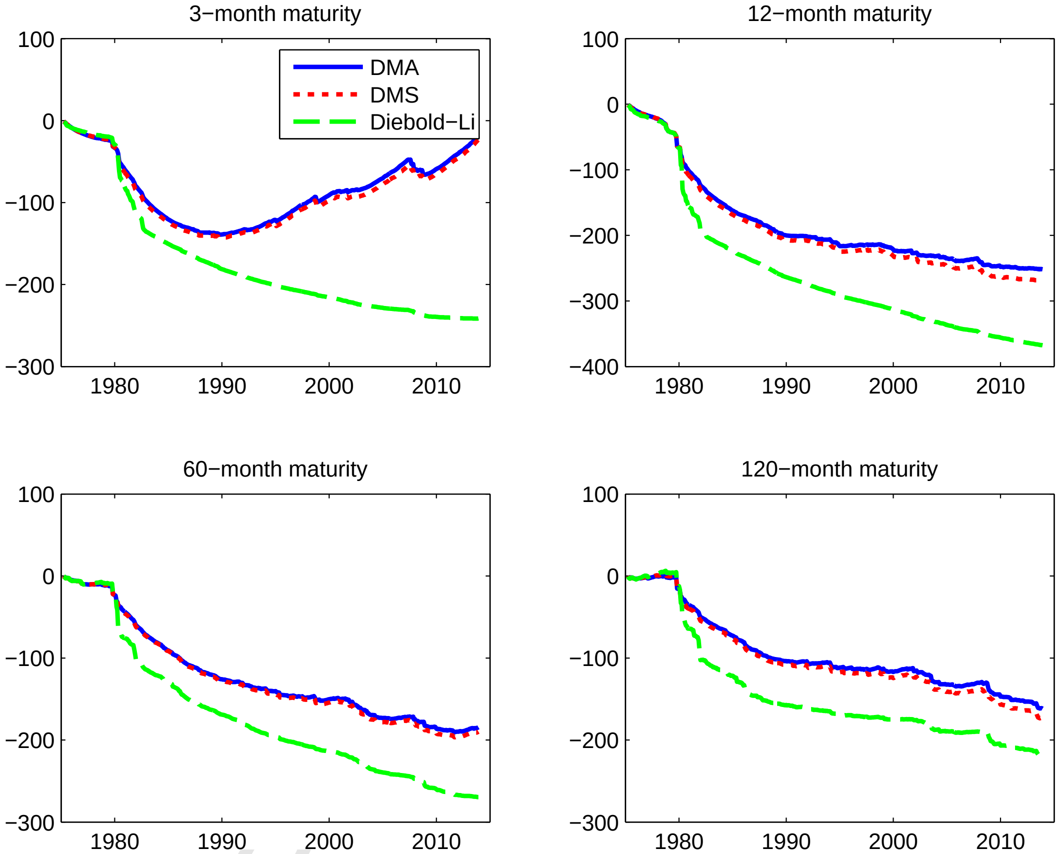
Notes: 1. This table presents one and three month ahead forecast statistics of bond yields with maturities ranging from three months to 120 months. The evaluation period is 1983:11 to 2013:11.
2. We report the ratio of each model's Mean Squared Forecast Errors (MSFE) relative to the MSFE of the random walk, and the preferred values are in bold. The symbol † indicates the model with the highest value of average predictive log-scores, a measure that takes into account the whole predictive density; see Geweke and Amisano (2010) for details.
3. In this table, we use following abbreviations: **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.

Table 4: Six and twelve month ahead relative MSFEs of competing term structure models

	DMA	**DMS**	**TVP**	**TVP-M**	**DL**	**DL-R10**	**DL-M**	**DL-SW**
Maturity	Six month ahead relative MSFE ($h = 6$)							
3	**0.871**†	0.976	0.974	1.012	1.332	1.703	1.405	1.908
6	**0.947**†	1.051	1.053	1.120	1.446	1.796	1.514	1.999
12	**0.969**†	1.072	1.057	1.080	1.304	1.501	1.322	1.825
24	1.025†	1.109	1.106	1.105	1.393	1.623	1.407	1.707
36	1.038†	1.107	1.110	1.090	1.416	1.685	1.427	1.574
48	1.038†	1.097	1.101	1.073	1.403	1.709	1.414	1.481
60	1.032†	1.085	1.088	1.060	1.368	1.702	1.381	1.432
72	1.021†	1.073	1.076	1.051	1.322	1.673	1.336	1.417
84	1.009†	1.063	1.064	1.046	1.270	1.627	1.286	1.422
96	**0.997**†	1.055	1.056	1.044	1.218	1.568	1.236	1.438
108	**0.987**†	1.048	1.049	1.043	1.167	1.502	1.187	1.458
120	**0.978**†	1.043	1.045	1.044	1.122	1.433	1.142	1.477
Mean	**0.994**†	1.067	1.067	1.066	1.323	1.632	1.348	1.607
Maturity	Twelve month ahead relative MSFE ($h = 12$)							
3	**0.980**†	1.073	1.021	1.240	1.349	1.605	1.517	1.677
6	1.034†	1.128	1.079	1.292	1.419	1.703	1.579	1.784
12	1.025†	1.139	1.082	1.210	1.353	1.592	1.458	1.661
24	1.075†	1.191	1.139	1.208	1.474	1.757	1.573	1.664
36	1.091†	1.202	1.152	1.188	1.528	1.848	1.623	1.625
48	1.087†	1.193	1.145	1.163	1.532	1.885	1.625	1.591
60	1.070†	1.175	1.127	1.140	1.505	1.884	1.596	1.577
72	1.049†	1.154	1.106	1.121	1.459	1.858	1.549	1.581
84	1.025†	1.133	1.086	1.105	1.405	1.816	1.494	1.599
96	1.003†	1.115	1.068	1.092	1.349	1.766	1.437	1.623
108	**0.983**†	1.100	1.054	1.083	1.294	1.711	1.381	1.649
120	**0.966**†	1.089	1.043	1.077	1.243	1.655	1.329	1.673
Mean	1.035†	1.143	1.093	1.174	1.415	1.748	1.524	1.648

Notes: 1. This table presents six and twelve month ahead forecast statistics of bond yields with maturities ranging from three months to 120 months. The evaluation period is 1983:11 to 2013:11.
2. We report the ratio of each model's Mean Squared Forecast Errors (MSFE) relative to the MSFE of the random walk, and the preferred values are in bold. The symbol † indicates the model with the highest value of average predictive log-scores, a measure that takes into account the whole predictive density; see Geweke and Amisano (2010) for details.
3. In this table, we use following abbreviations: **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.

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



Notes: These are one 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.

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 predictive power 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. This is an example of an unstable forecasting relationship, where the same information in macro-finance variables may not be useful in forecasting at all periods and forecast horizons. Hence, this result indicates the relative advantages of DMA as a plausible adaptive method for forecasting using only relevant information at each time period.

In the Nelson-Siegel setup 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 (2011*a*). 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 have shown previously. Our extended NS model consistently improves upon DL across all horizons and maturities, which is confirmed by relative MSFEs, predictive log-scores, and the Diebold-Mariano test. Moreover, at shorter horizons, and to some extent at longer horizons, our proposed method improves upon the RW.

Fluctuation Test In addition to the relative MSFE and the cumulative log predictive likelihoods, we formally test for forecasting ability over time in the presence of instabilities by implementing the Giacomini and Rossi (2010) one-sided Fluctuation test (Ft-test). Under the null hypothesis, the Ft-test gauges whether the local relative forecasting performance (based on Diebold and Mariano (1995) test) of the competing model and the benchmark model (say DL) is equal at each point in time. The alternative is that the competing model forecasts better than the DL. Hence, when the Ft-test statistic is above its critical value at the 10% level of significance, the competing model forecasts significantly better than the DL at that point in time. Otherwise, if the Ft-test is below its critical value, the evidence suggests the absence of forecasting ability of the model. To compute the test, we follow the recommendations in Giacomini and Rossi (2010) and set the size of moving local window to a third of the in-sample observations.[12]

The Ft-tests for the $h = 3, 6$ and 12 month forecast horizons are reported in Figures 4, 5 and 6. At the 3-month and 6-month horizons, both DMA and TVP(TVP-M) display significant forecasting ability except the very long maturity, as the Ft-test is usually above its critical value. At the twelve month forecast horizon, DMA and TVP(TVP-M) perform even better, as the Ft-test is consistently above the critical value for all maturities. Overall, DMA performs better than TVP or TVP-M as the test statistics are consistently higher, especially for shorter maturities. This is a further piece of evidence regarding the importance of systematically taking into account time-variation in parameters of term structure models, as well as model uncertainty.

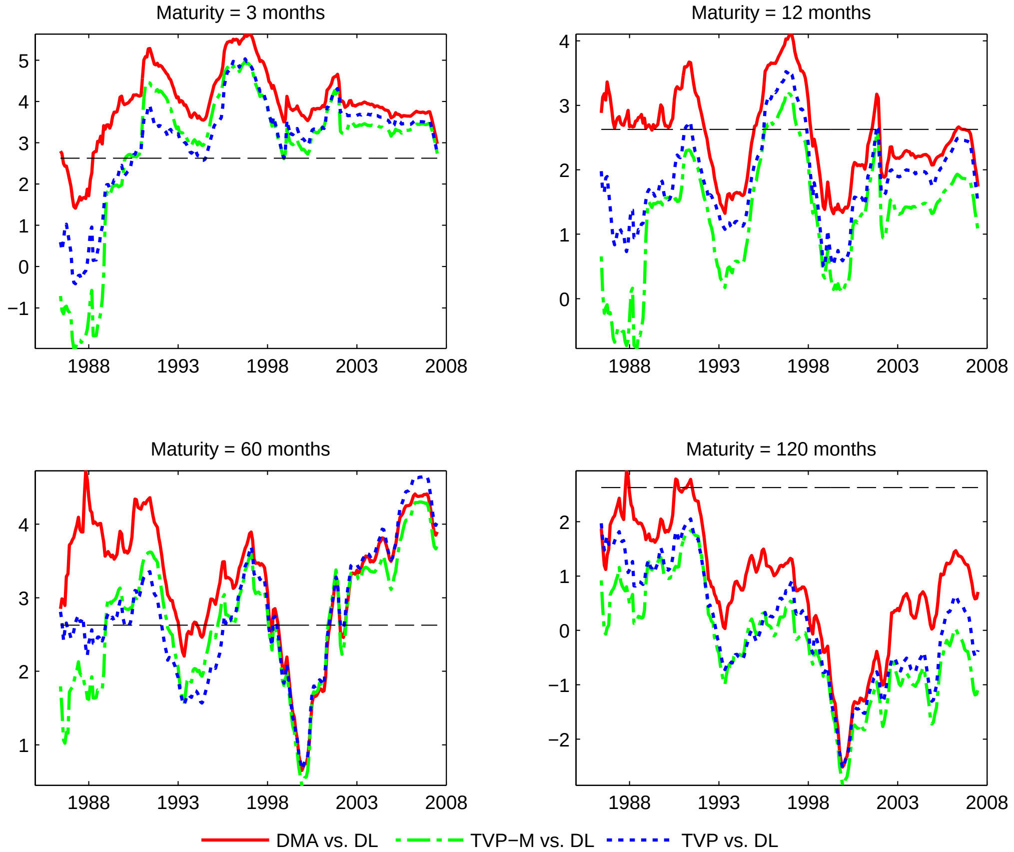
3.5 Predictive Gains and Sources of Instability

Since the pricing dynamics of most competing methods are fully characterized by the NS restrictions, we can conclude that the predictive gains in our preferred specification are purely from its time-series dynamics. Specifically these gains seem to be stemming from the fact that parameter and model uncertainty are fully taken into account. In this subsection our desire is to further highlight and decompose the various sources of these predictive gains. In the previous results we have found that time-varying parameter models improve over constant parameter versions, whether these include macro-finance information or not. Further improvements can be achieved by allowing the dimension of the model to switch over time. In order to pin down the exact sources of predictive gains we first 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 also the differences are statistically significant.

However, to what extent the constituent components of our DMA model contribute

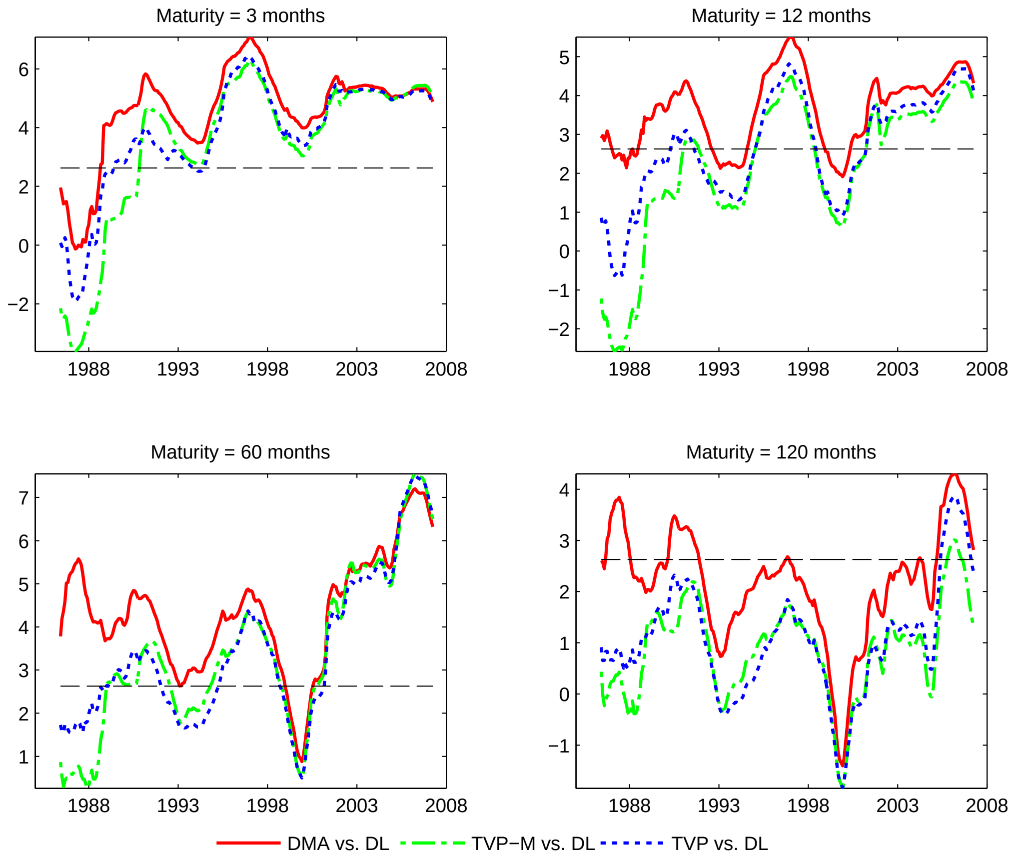
[12]According to Giacomini and Rossi (2010) Monte Carlo evidence, the Fluctuation test has good properties when implemented using a local moving window size that is a small, such as a third of the in-sample estimation window.

Figure 4: Fluctuation Test at 3-Month Forecasting Horizon



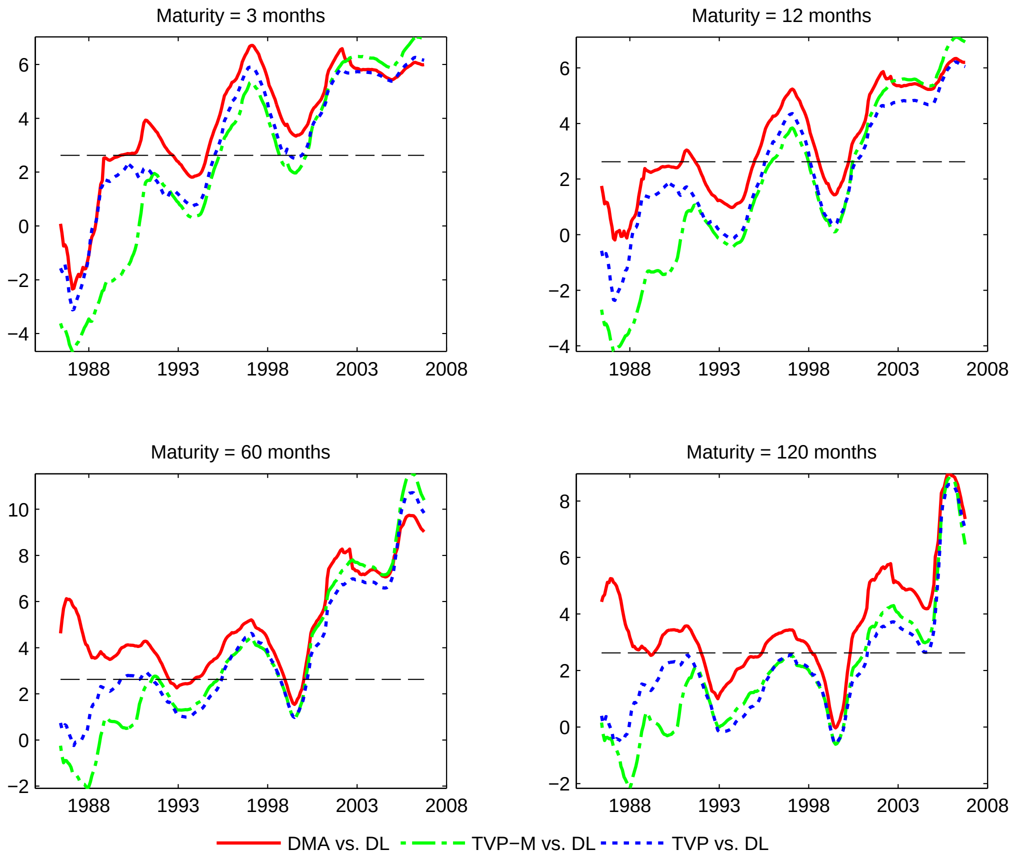
Notes: The Figure shows the Giacomini and Rossi's (2010) one-sided Fluctuation test (Ft-test) based on DMW-test for three competing models: DMA, TVP-M and TVP. The benchmark model is the Diebold and Li (2006) (DL). It also displays the one-sided Ft-test critical value at 10% level of significance. When the Ft-test statistic is above its critical value, we reject the null of equal local relative forecasting performance between the competing model and the DL, concluding that the method under consideration forecasts significantly better than the DL at that point in time. When the Ft-test is below its critical value, the evidence is consistent with the absence of forecasting ability of the method under consideration. The forecasting horizon is h = 3 months.

Figure 5: Fluctuation Test at 6-Month Forecasting Horizon



Notes: The Figure shows the Giacomini and Rossi's (2010) one-sided Fluctuation test (Ft-test) based on DMW-test for three competing models: DMA, TVP-M and TVP. The benchmark model is the Diebold and Li (2006) (DL). It also displays the one-sided Ft-test critical value at 10% level of significance. When the Ft-test statistic is above its critical value, we reject the null of equal local relative forecasting performance between the competing model and the DL, concluding that the method under consideration forecasts significantly better than the DL at that point in time. When the Ft-test is below its critical value, the evidence is consistent with the absence of forecasting ability of the method under consideration. The forecasting horizon is h = 6 months.

Figure 6: Fluctuation Test at 12-Month Forecasting Horizon



Notes: The Figure shows the Giacomini and Rossi's (2010) one-sided Fluctuation test (Ft-test) based on DMW-test for three competing models: DMA, TVP-M and TVP. The benchmark model is the Diebold and Li (2006) (DL). It also displays the one-sided Ft-test critical value at 10% level of significance. When the Ft-test statistic is above its critical value, we reject the null of equal local relative forecasting performance between the competing model and the DL, concluding that the method under consideration forecasts significantly better than the DL at that point in time. When the Ft-test is below its critical value, the evidence is consistent with the absence of forecasting ability of the method under consideration. The forecasting horizon is h = 12 months.

Table 5: MSFE from DMA Relative to Other Models

	DMA vs. DL				DMA vs. TVP-M			
Maturity	$h = 1$	$h = 3$	$h = 6$	$h = 12$	$h = 1$	$h = 3$	$h = 6$	$h = 12$
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 1983:11 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.

to the predictive gains is still unclear. To provide a straightforward answer to this question, we perform a simple but informative decomposition to pin down various sources of instability that might affect the out-of-sample forecasting performance of DMA. In particular, this decomposition quantifies exactly the relative contribution of time-varying parameters or model uncertainty to the prediction variance of bond yields, highlighting sources that potentially hinder the bond yield forecasts.

We use the law of total variance to decompose the variance of a random variable into its constituent parts. Following Dangl and Halling (2012), we begin with the decomposition of pricing factor forecasts with respect to different choices of forecasting model L:

$$Var(F^{NS}) = E_L\Big(Var(F^{NS}|L)\Big) + Var_L\Big(E(F^{NS}|L)\Big). \tag{3.1}$$

After some algebra and using the expressions detailed in previous sections, we have

$$\begin{aligned} Var(F^{NS}_{t+1}) \quad = \quad & \underbrace{\sum_i (\Sigma_t|L_i, D_t) P(L_i|D_t)}_{\text{Observational variance}} \\ & + \underbrace{\sum_i (X_t \Phi_{t|t-1} X_t'|L_i, D_t) P(L_i|\ D_t)}_{\text{Parameter uncertainty}} \\ & + \underbrace{\sum_i (\hat{F}^{NS}_{t+1,i} - \hat{F}^{NS}_{t+1})^2 P(L_i|D_t)}_{\text{Model uncertainty}}, \end{aligned}$$

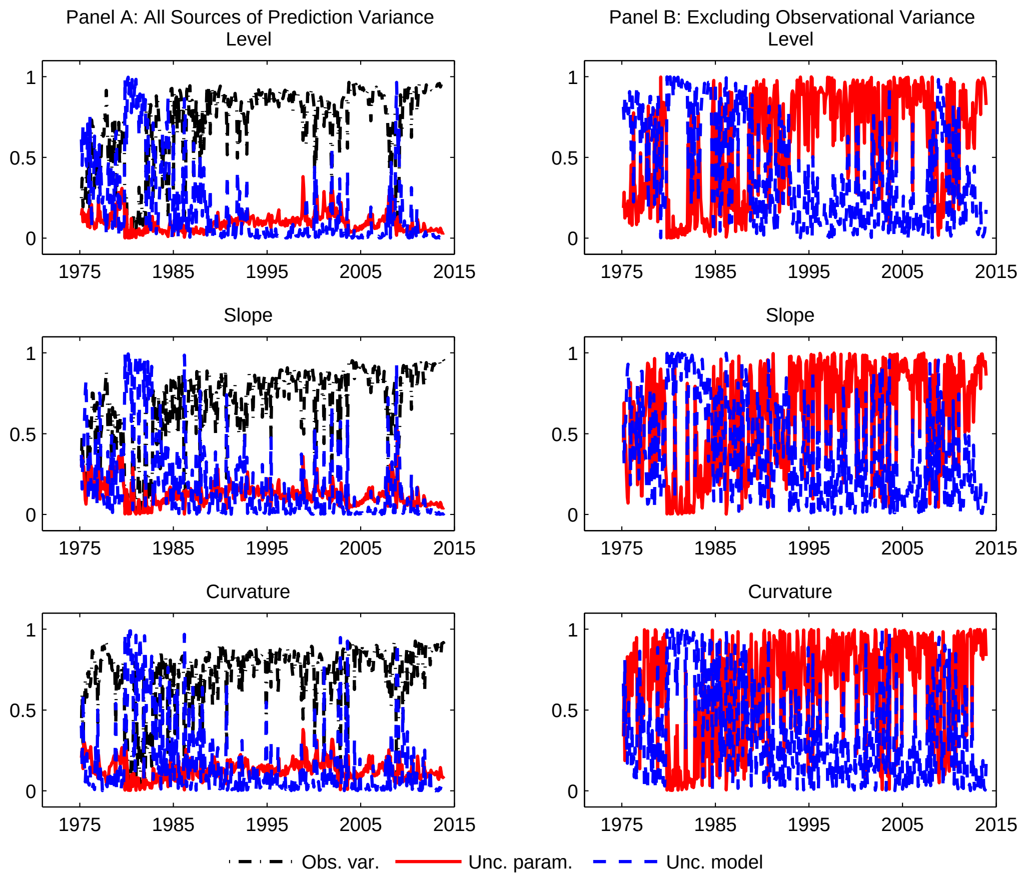
where Σ_t denotes the variance of the disturbance term in the observation equation, $\Phi_{t|t-1}$ denotes the unconditional variance of the time-t prior of the coefficient vector β_t, $\hat{F}^{NS}_{t+1,i}$ is the forecast conditional on L_i and $\hat{F}^{NS}_{t+1}$ is the weighted average over all candidate models.

The individual terms of Equation (3.2) state the sources of prediction uncertainty and have intuitive interpretations. The first term measures the expected observational variance, calculated over different choices of forecast model L. This term in fact captures the random fluctuations or risks in the pricing factors, relative to the predictable drift component. The second term is the expected variance from errors in the estimation of the coefficient vector, which can be interpreted as the source of estimation or parameter uncertainty. The third term captures model uncertainty, which can also be considered as the time variability of predictors used to generate forecasts.

Figure 7 displays the variance decomposition for three pricing factors, where Panel A shows that the predominant source of uncertainty is observational variance except during the 1980s. On average observational variance accounts for more than 70% of predictive variance for all three pricing factors. This is consistent with the findings of Dangl and

Halling (2012), as the asset prices frequently fluctuate randomly over their expected values. These fluctuations serve as the source of risk premia, and dominate the drift components in the term structure model. Therefore the fluctuations in fact contaminate the predictive power of term structure models, especially during the periods when pricing factors are highly persistent.[13]

Figure 7: Sources of Prediction Variance



Notes: This figure displays the decomposition of the prediction variance with respect to different sources. In Panel A, the prediction variance is split into observational variance (Obs. var.), variance caused by errors in the estimation of parameters (Unv. param.), variance caused by the model uncertainty (Unc. model). The illustration shows the relative weights of these components. Panel B masks out observational variance and shows relative weights of the remaining variance.

In Panel B of Figure 7, by excluding the observational variance we can focus upon the relative weights of the remaining sources of prediction uncertainty. The parameter uncertainty turns out to be a main source of prediction uncertainty after 1990, on average above 60%, which implies parameter instability is another crucial reason causing interest

[13]This does not at all mean term structure models are not useful. For instance, term structure models can reveal informative dynamics of market prices of risks and have reliable term premia of long-term bonds, which can not be offered by the random walk model.

rate unpredictability during that time. Therefore, a successful forecasting model should at least consider the feature of time-varying parameters. The model uncertainty is also important especially during certain periods. For example, model uncertainty rises steeply during the 1980s, accounting for more than 90% of total variance. This observation suggests without considering model uncertainty the predictive power of term structure models may be significantly compromised. The contribution of each source is time-varying but comparable for the three pricing factors. It highlights that *the consideration of both parameter uncertainty and model uncertainty regarding different choices of indicators* is a better way to produce more reliable interest rate forecasts.

3.6 Model-Implied Term Premia

In this section we set out a visual comparison of our term premium estimates.[14] 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 8. 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.[15] 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).

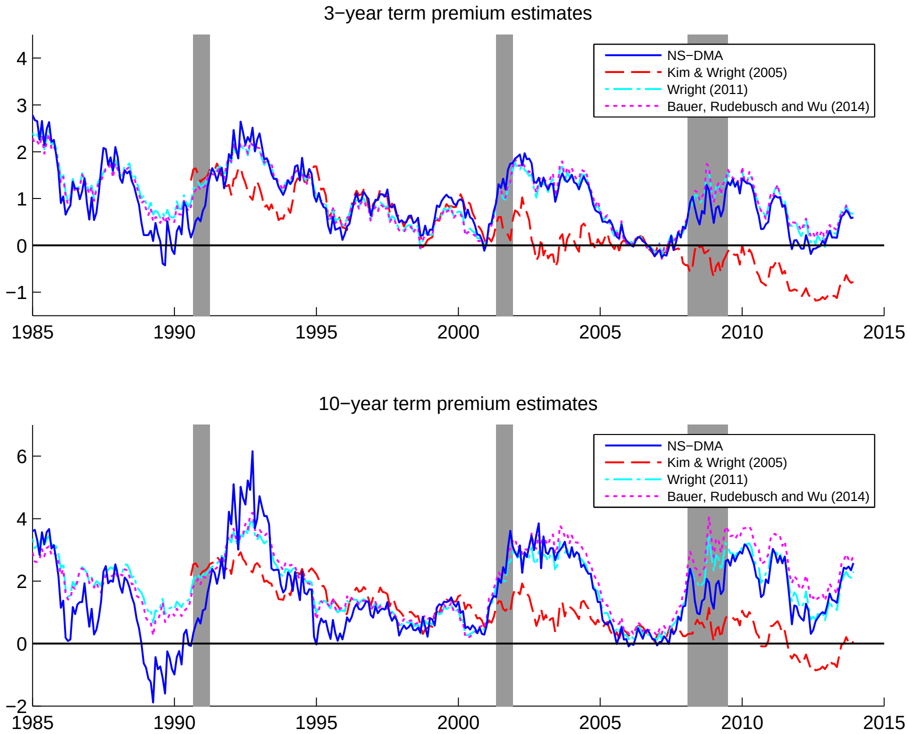
It is worth emphasizing that DMA captures plausible term premia using conditional information only. As it is shown in the upper panel of Figure 8, the 36-month term premium estimates of DMA are highly consistent with the full-sample estimates of Wright (2011) and Bauer, Rudebusch and Wu (2014). In general all term premia estimates display countercyclical behavior, as they rise in and around US recessions, except 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

[14]A more thorough discussion about term premia and risk-neutral rates, as well as the underlying drivers, can be found in Appendix C.5.

[15]The comparison between the DMA term premia and recursively estimated term premia from dynamic Nelson-Siegel is shown in Appendix C.5.1. 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.

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 8: 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. (C.1); we then calculate the term premia using Eq. (C.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 seem to be more sensitive to changes in the economic environment, which can be seen more clearly from the lower panel of Figure 8 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.

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 this literature by proposing 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 the time-varying predictive power of term structure models and unfold the time variation of sources that significantly drive the predictive variance of the yield curve. 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.

Specifying the interactions between the yield factors and macro-finance information using time varying parameters, stochastic volatility, and switching information set, causes additional econometric challenges in terms of tractability of estimation. Such challenges are addressed here by bringing in a fast and simple estimation technique. The proposed yield curve specification is robust to various sources of structural instabilities, and it is highly consistent with the theoretical and empirical findings in the previous yield curve literature. Future research could employ a one-step estimation approach to provide forecasts with higher accuracy, in which case a trade-off should be made between predictive accuracy and estimation efficiency. Finally, 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 can also be considered in 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.

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., and Glenn D. Rudebusch. 2017. "Resolving the spanning puzzle in macro-finance term structure models." *Review of Finance*, 21(2): 511.

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.

Byrne, Joseph, Dimitris Korobilis, and Pinho Ribeiro. 2018. "On the sources of uncertainty in exchange rate predictability." *International Economic Review*, 59(1): forthcoming.

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.

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.

Dangl, Thomas, and Michael Halling. 2012. "Predictive regressions with time-varying coefficients." *Journal of Financial Economics*, 106(1): 157–181.

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.

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.

Giacomini, Raffaella, and Barbara Rossi. 2010. “Forecast comparisons in unstable environments.” *Journal of Applied Econometrics*, 25(4): 595–620.

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., 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.

Rossi, Barbara. 2013. "Advances in Forecasting under Instability." Vol. 2 of *Handbook of Economic Forecasting*, Chapter 0, 1203–1324. Elsevier.

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.

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

Table 6: List of Yields and Macro-Finance Variables

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.

Predict term structure of US interest rates in a data rich environment
Allow the model dimension and parameters to change over time
More macro-finance information is chosen to be relevant during recessions
As a result forecasts of yield curve are superior relative to benchmarks
We also evaluate estimates of term premia using our approach