

Predicting Interest Rates Using Shrinkage Methods, Real-Time Diffusion Indexes, and Model Combinations *

Norman R. Swanson¹, Weiqi Xiong², and Xiye Yang³

^{1,3}Rutgers University and ²BlackRock

March 2020

Abstract

In the context of predicting the term structure of interest rates, we explore the marginal predictive content of real-time macroeconomic diffusion indexes extracted from a “data rich” real-time dataset, when used in dynamic Nelson-Siegel (NS) models of the variety discussed in [Svensson \(1994\)](#) (NSS) and [Diebold and Li \(2006\)](#) (DNS). Our diffusion indexes are constructed using principal component analysis with both targeted and un-targeted predictors, with targeting done using the lasso and elastic net. Our findings can be summarized as follows. First, the marginal predictive content of real-time diffusion indexes is significant for the preponderance of the *individual* models that we examine. The exception to this finding is the post “Great Recession” period. Second, forecast *combinations* that include only yield variables result in our most accurate predictions, for most sample periods and maturities. In this case, diffusion indexes do not have marginal predictive content for yields and do not seem to reflect unspanned risks. This points to the continuing usefulness of DNS and NSS models that are purely yield driven. Finally, we find that the use of fully revised macroeconomic data may have an important confounding effect upon results obtained when forecasting yields, as prior research has indicated that diffusion indexes are often useful for predicting yields when constructed using fully revised data, regardless of whether forecast combination is used, or not (see [Swanson and Xiong \(2018\)](#)). Nevertheless, our findings also underscore the potential importance of using machine learning, data reduction, and shrinkage methods in contexts such as term structure modeling.

JEL Classification: C12, C22, C53.

Keywords: Real-time Forecasting, Dynamic Nelson-Siegel Model, Term Structure of Interest Rates, Real-time Dataset, FRED-MD, Real-Time Diffusion Index.

* Norman R. Swanson, Department of Economics, Rutgers University, 75 Hamilton Street, New Brunswick, NJ 08901, USA, nswanson@econ.rutgers.edu. Weiqi Xiong, BlackRock, Three Garden Road, Central, Hong Kong, xiongweiqi@163.com. Xiye Yang, Department of Economics, Rutgers University, 75 Hamilton Street, New Brunswick, NJ 08901, USA, xiye yang@economics.rutgers.edu. The authors are grateful to Mingmian Cheng, Valentina Corradi, Eric Ghysels, Yuan Liao, Massimiliano Marcellino, Christian Schumacher, and Greg Tkacz for useful comments made on earlier drafts of this paper. The authors are also very grateful to the editor, Barbara Rossi, the associate editor, and three anonymous referees for many useful comments made during the revision process.

1 Introduction

The term structure of interest rates contains crucial information for forecasting macroeconomic variables and pricing interest rate contingent assets. As a consequence, forecasts of U.S. Treasury bill and bond yields remain important inputs in models used in industry and government. In this paper, we add to the literature on interest rate prediction by carrying out an extensive set of forecasting experiments in which we explore the marginal predictive content of so-called “data rich” or real-time latent macroeconomic factors (i.e., real-time macroeconomic diffusion indexes) in Nelson-Siegel (NS) type models. In our experiments, we specify diffusion indexes using various novel machine learning methods, including the lasso and elastic net. Summarizing, the questions that we attempt to answer include the following. Do macroeconomic, financial, and other non-yield variables contain marginal predictive content for interest rates, over and above that contained in the term structure? Does the use of targeted and/or un-targeted machine learning techniques lead to improved yield forecasts? Does the use of real-time data matter when forecasting interest rates using NS type models?¹

In addition to utilizing NS type models in our forecasting experiments, we utilize a number of alternative econometric models. Summarizing, we analyze dynamic Nelson-Siegel-Svensson (NSS) models (see (Svensson, 1994)), dynamic NS (DNS) models of the variety recently examined by Diebold and Li (2006), and various benchmark specifications, including vector autoregressive (VAR) and autoregressive (AR) models. Of note is that the real-time diffusion indexes that we construct are utilized both in the specification of the risk factors in our DNS and NSS models, as well as in the specification of our econometric models.² Real-time diffusion indexes, in turn, are extracted from a set of 130 macroeconomic variables, using principal component analysis with both targeted and un-targeted predictors, with targeting done using the lasso and elastic net.³ Our un-targeted PCA estimator is discussed in Stock and Watson (2002a,b) and Bai and Ng (2002, 2008, 2009), as is widely used in the empirical literature. The use of targeted predictors, on the other hand, is a relatively new method (see Bai and Ng (2008)), and involves extracting diffusion indexes from a *pre-selected* set of the variables, chosen from our real-time dataset. In our implementation of this method, we utilize machine learning methods, including the elastic net and the least absolute shrinkage operator (lasso) in order to pre-select variables.⁴

¹Our analysis is also meant to add to a number of important papers in the broader term structure forecasting literature, including: Diebold and Li (2006), Rudebusch and Wu (2008), Exterkate et al. (2013), Christensen and Rudebusch (2014), Krippner (2015), Christensen and Rudebusch (2016), Coroneo et al. (2016), Wu and Xia (2016), Ghysels and Marcellino (2018), and the references cited therein.

²The models utilized in this paper do not include a separate and important class called mixed-frequency models. For an interesting discussion of mixed frequency modeling and diffusion indexes, see Andreou et al. (2018).

³The real-time macroeconomic variable dataset that we utilize is available at the St. Louis Federal Reserve Bank website, and is discussed in McCracken and Ng (2016).

⁴See Stock and Watson (2012), Kim and Swanson (2014), and the references cited therein for a discussion of shrinkage methods used for forecasting with many predictors. For further discussion of targeted and un-targeted diffusion index

The particular data-rich environment that we examine is one in which real-time data that include the entire revision history for each variable are analyzed. Our focus on real-time data in our term structure forecasting experiments is closest to the work reported in [Exterkate et al. \(2013\)](#) and [Ghysels et al. \(2018\)](#). In order to better understand the importance of real-time data in forecasting, in general, consider the case of real-time GDP. Assume that observations on this variable that are collected for calendar date December 2000, say, include the first “reading” on 4th quarter 2000 GDP that was available in March 2001, and well as the 1st revised version of this datum that became available in June 2001, and so on, up until the present date.⁵ Thus, real-time datasets include entire sequences of revisions for each calendar dated observation. Data such as these allow researchers to simulate “truly” real-time forecasting environments, which differs from the common practice of using so-called fully revised data in forecasting experiments. This is important, as “fully revised” data consist of observations that were not actually available to market participants in real-time.

Broadly, speaking, our motivation for the research reported in this paper follows two strands of the literature. First, simple regression-based approaches to forecasting treasury yields have a historically good track record. Most notable among such models is probably the DNS model. Indeed, DNS type models have become the leading method for yield curve forecasting at many policy institutions (see e.g., [BIS \(2005\)](#)). However, findings in the recent literature suggest that DNS model performance has deteriorated in recent (post credit crisis) years (see e.g., [Moench \(2008\)](#), [Diebold and Rudebusch \(2013\)](#), and [Altavilla et al. \(2017\)](#)). This performance might be explained by regime change or structural breaks, for example. Other potential causes include generic model misspecification, model over-fitting, and measurement error. A further explanation involves the importance of the so-called zero lower bound in recent years (see e.g. [Christensen and Rudebusch \(2014\)](#), [Krippner \(2015\)](#), [Christensen and Rudebusch \(2016\)](#), and [Wu and Xia \(2016\)](#)). In this paper, we attempt to distinguish between these alternative explanations. This is done in part by examining the importance of forecast combinations in our experimental results.

We also address the related question of whether recent DNS model performance can be reconciled by looking beyond the cross section of yields, when pinning down the dynamic behavior of interest rates (see [Duffee \(2011\)](#)). Along these lines, recent research, including that discussed in the papers cited above, has shown that modeling the co-movements of the underlying economy by specifying diffusion indexes, or using key macroeconomic indicators, has proven useful for predicting yields. For example, using a dynamic factor model, [Coroneo et al. \(2016\)](#) find that real economic activity and real interest rates contain predictive content for government bond yields that is not spanned by the cross-section of yields. [Ang](#) construction methods, see [Schumacher \(2007, 2010\)](#), [Bai and Ng \(2008, 2009\)](#), [Kim and Swanson \(2014\)](#), [Carrasco and Rossi \(2016\)](#), and [Ghysels and Marcellino \(2018\)](#). For a discussion of forecasting using factor augmented Nelson-Siegel models, see [Exterkate et al. \(2013\)](#).

⁵For an informative discussion of the importance of data revision *and* the relevance of delayed release dates, see [Ghysels et al. \(2018\)](#).

and Piazzesi (2003) and Moench (2008) also report improved forecast accuracy, when diffusion indexes are introduced to their models. Additional recent studies consider enlarging the information set used in prediction with either observable macroeconomic factors (Diebold et al. (2006) and Rudebusch and Wu (2008)) or surveys (Altavilla et al. (2017)). Ludvigson and Ng (2009) find that adding macroeconomic factors helps when forecasting bond risk premia. Our objective in this paper is to provide further evidence on the importance of unspanned macroeconomic risks when predicting the terms structure by considering not only the use of real-time data in our analysis, but also by examining in the importance of machine learning methods, targeted PCA, and forecast combination when predicting yields.

A number of findings emerge upon examination of the results of our prediction experiments. First, we find that the marginal predictive content of real-time macroeconomic diffusion indexes is significant for many of the *individual* models that we examine, across various sample periods. Indeed, macroeconomic diffusion indexes are useful both for predicting the risk factors utilized in DNS and NSS models, as well as when added to purely econometric AR and VAR specifications. Moreover, DNS and NSS models that incorporate diffusion indexes are often the “mean square forecast error - best” (MSFE-best) performers. Summarizing, all model variants with diffusion indexes outperform less sophisticated models that do not include such indexes. The exception to this finding is the post “Great Recession” period, during which time diffusion indexes are not found to contain useful information for predicting yields. This suggests that there are un-spanned risks, at least pre Great Recession, and that these risks can be modeled using real-time data. This findings suggests that macroeconomic, financial, and other non-yield variables contain marginal predictive content for interest rates, over and above that contained in the term structure, when one utilizes individual models for forecasting.

Second, the picture changes when forecast combinations are added to the mix of models in our experiments. In particular, forecast *combinations* that include only yield variables result in our most accurate predictions, for *most* sample periods that we analyze. When combination forecasts are included in our analysis, thus, there are no unspanned risks (i.e., risks not spanned by bond prices or returns), and all relevant forecasting information is contained in the term structure. In the context of our experiments, this indicates that the use of fully revised macroeconomic data may have an important confounding effect upon results obtained when forecasting yields, as prior empirical evidence suggests that diffusion indexes are often useful for predicting yields when constructed using fully revised data, regardless of whether forecast combination is used, or not (see Swanson and Xiong (2018)).⁶ This suggests that one possible explanation for our finding concerning the usefulness of macroeconomic diffusion indexes pre Great Recession has to do with model misspecification, particularly since our MSFE-best combination utilize *all* of our purely yield driven models. Summarizing, when combinations are added to our mix

⁶Note that the methodology used in Swanson and Xiong (2018) differs from that used here in a number of dimensions. For example, those authors do not use targeted methods, such as the lasso and elastic net, when constructing diffusion indexes. Also, they consider only a small subset of the models examined here.

of models, we find no evidence of unspanned risks (i.e. no useful information in our real-time dataset), when forecasting the term structure, regardless of time period. This points to the continuing usefulness of DNS and NSS models that are purely yield driven. Of course, the variety of models examined in this paper is necessarily limited in scope, and further research is warranted in order to establish whether our findings are robust. For example, there are many recent and interesting machine learning methods that we have not utilized in our experiments.

Third, elastic net and lasso shrinkage operators are used for constructing diffusion indexes in many of our *individual* MSFE-best models. This is indicative of the potential importance of machine learning methods for prediction not only of the term structure, but also of macroeconomic variables in general, and adds to the burgeoning literature on this topic (see e.g., [Bai and Ng \(2009\)](#), [Schumacher \(2010\)](#), [Kim and Swanson \(2014\)](#), [Carrasco and Rossi \(2016\)](#), [Ghysels and Marcellino \(2018\)](#)). This finding suggests that targeted machine learning techniques lead to improved yield forecasts, although this finding no longer holds when combinations of different forecasts are compared with individual forecasts, as discussed above.

Finally, we find that the use of real-time data matters when forecasting interest rates using NS type models, in the sense that we overturn the findings of [Swanson and Xiong \(2018\)](#). In particular, using fully revised data, [Swanson and Xiong \(2018\)](#) find that big data such as that used in this paper “matter”, since variables other than yields enter into their MSFE-best NS type models, even when forecast combinations are considered.⁷ In contrast, when we use real-time data rather than fully revised data, we find that the use of forecast combinations renders our big data un-informative, as all MSFE-models contain only term structure information. Thus, the use of fully revised macroeconomic data may have an important confounding effect upon results obtained when forecasting yields.

The rest of the paper is organized as follows. In [Section 2](#), we describe the dynamic Nelson-Siegel models that we analyze, and in [Section 3](#) we discuss yield curve prediction with macroeconomic diffusion indexes. [Section 4](#) includes details describing our empirical setup. [Section 5](#) details the data used in our analysis, and [Section 6](#) contains a summary of our empirical findings. Concluding remarks are gathered in [Section 7](#).

2 Dynamic Nelson-Siegel Models

2.1 Three-factor Dynamic Nelson-Siegel Model

Motivated by rational expectation theory, [Nelson and Siegel \(1987\)](#) express spot interest rates in terms of instantaneous forward rates. Namely, the instantaneous forward interest rate of a bond with maturity m is denoted as $f(m)$, and the spot interest rate of a bond with maturity τ as $y(\tau)$. Then, the yield to

⁷See Tables 8A-8C in [Swanson and Xiong \(2018\)](#) for a summary of their forecast combination results using fully revised data.

maturity of a bond can be written as the average of forward rates:

$$y(\tau) = \frac{1}{\tau} \int_0^\tau f(m) dm.$$

In this context, [Nelson and Siegel \(1987\)](#) motivate the use of the following forward rate model (which can generate monotonically increasing, humped, S-shaped, and a variety of other yield curves):

$$f(m) = \beta_1 + \beta_2 \cdot \exp\left(\frac{m}{\theta_t}\right) + \beta_3 \cdot \left[\left(\frac{m}{\theta_t}\right) \exp\left(\frac{m}{\theta_t}\right)\right],$$

where $\lambda_t = \frac{1}{\theta_t}$ is the so-called decay parameter, which must be estimated, is assumed fixed in this model, and is time varying in the dynamic version of the model discussed below. It is then easy to derive the following model for bond yields:

$$y(\tau) = \beta_1 + \beta_2 \cdot \left[\frac{1 - \exp(-\frac{\tau}{\theta_t})}{\frac{\tau}{\theta_t}}\right] + \beta_3 \cdot \left[\frac{1 - \exp(-\frac{\tau}{\theta_t})}{\frac{\tau}{\theta_t}} - \exp(-\frac{\tau}{\theta_t})\right].$$

In the above model, the latent (risk) factors (i.e., the “betas”) are fixed. [Diebold and Li \(2006\)](#) generalize this model to allow for time-varying betas: $\beta_{1,t}$, $\beta_{2,t}$, and $\beta_{3,t}$. Their so-called Dynamic Nelson-Siegel (DNS) model is estimated using a two-step procedure. First, the rate of decay, λ_t , is set to a constant. Next, at each point in time, t , the yield cross section is linearly projected onto the set of factor loadings $(1, \frac{1 - \exp(-\lambda_t \tau)}{\lambda_t \tau}, \frac{1 - \exp(-\lambda_t \tau)}{\lambda_t \tau} - \exp(-\lambda_t \tau))$. In our experiments, various different dimensions are considered when specifying the yield cross section. Namely, we consider yield cross sections using 10, 12, and 30 different yield maturities (see below for further discussion). For example, with our cross section, we estimate the latent factors by fitting the following equation:

$$\begin{pmatrix} y_t(\tau_1) \\ y_t(\tau_2) \\ y_t(\tau_3) \\ \vdots \\ y_t(\tau_{12}) \end{pmatrix}_{12 \times 1} = \begin{pmatrix} 1 & \frac{1 - \exp(-\lambda_t \tau_1)}{\lambda_t \tau_1} & \frac{1 - \exp(-\lambda_t \tau_1)}{\lambda_t \tau_1} - \exp(-\lambda_t \tau_1) \\ 1 & \frac{1 - \exp(-\lambda_t \tau_2)}{\lambda_t \tau_2} & \frac{1 - \exp(-\lambda_t \tau_2)}{\lambda_t \tau_2} - \exp(-\lambda_t \tau_2) \\ 1 & \frac{1 - \exp(-\lambda_t \tau_3)}{\lambda_t \tau_3} & \frac{1 - \exp(-\lambda_t \tau_3)}{\lambda_t \tau_3} - \exp(-\lambda_t \tau_3) \\ \vdots & \vdots & \vdots \\ 1 & \frac{1 - \exp(-\lambda_t \tau_{12})}{\lambda_t \tau_{12}} & \frac{1 - \exp(-\lambda_t \tau_{12})}{\lambda_t \tau_{12}} - \exp(-\lambda_t \tau_{12}) \end{pmatrix}_{12 \times 3} \begin{pmatrix} \beta_{1,t} \\ \beta_{2,t} \\ \beta_{3,t} \end{pmatrix}_{3 \times 1}$$

The betas (i.e., $\hat{\beta}_{1,t}$, $\hat{\beta}_{2,t}$, and $\hat{\beta}_{3,t}$) are called the “level”, “slope”, and “curvature” factors. In particular, note that the loading on $\hat{\beta}_{1,t}$ is one, which is naturally interpreted as the “level” factor. The loading on $\hat{\beta}_{2,t}$ decreases as bond maturity increases, resulting in an increase of the “slope” of bond yield curve. Finally, the loading on the third latent factor, $\hat{\beta}_{3,t}$, starts from zero on the short end of yield curve, reaches its peak at some maturity in the middle, and gradually decays to zero as the maturity goes to infinity.⁸ In summary, the DNS model can be written as follows:

$$\hat{y}_t(\tau) = \hat{\beta}_{1,t} + \hat{\beta}_{2,t} \cdot \left[\frac{1 - \exp(-\lambda_t \tau)}{\lambda_t \tau}\right] + \hat{\beta}_{3,t} \cdot \left[\frac{1 - \exp(-\lambda_t \tau)}{\lambda_t \tau} - \exp(-\lambda_t \tau)\right]. \quad (2.1)$$

⁸An increase in the “level” component, $\beta_{1,t}$, affects all yields equally, and thus determines the level of the yield curve. Also, as maturity τ goes to infinity, $\beta_{1,t} = y_t(\infty)$, by definition. An increase in the “slope” component, $\beta_{2,t}$, affects short rates more than long rates, thereby changing the slope, or the so-called “term spread” of the yield curve. Finally, an

In order to construct 1-step ahead predictions using the DNS model, for example, we fit estimated factors to AR and VAR models, as follows.

$$\hat{\beta}_{i,t+1} = c_i + \gamma_i \hat{\beta}_{i,t} + \epsilon_{t+1} \quad i = 1, 2, 3 \quad \text{or}, \quad (2.2)$$

$$\hat{\beta}_{t+1} = \mathbf{c} + \mathbf{\Gamma} \beta_t + \epsilon_{t+1}, \quad (2.3)$$

where ϵ_t is a scalar stochastic disturbance term, ϵ_t is a 3×1 vector of stochastic disturbance terms, and c_i , \mathbf{c} , γ_i , and $\mathbf{\Gamma}$, $i = 1, \dots, 3$, are conformably defined constants, constant vectors and constant matrices, respectively. With these last two models, one can construct predictions of the $\hat{\beta}_{i,t}$, for $i = 1, \dots, 3$, which can in turn be inserted into the above model of $\hat{y}_t(\tau)$ in order to generate yield predictions. In the sequel, we consider two types of prediction models. In one, the decay parameter is fixed. In the other, the decay parameter is re-estimated prior to the construction of each new prediction. For further details, including a recent review of Treasury yield curve modeling using DNS models, see [Diebold and Rudebusch \(2013\)](#) and [De Pooter \(2007\)](#). For further discussion comparing arbitrage free dynamic latent factor and DNS models, see [Ang and Piazzesi \(2003\)](#), [Diebold et al. \(2006\)](#), [Christensen et al. \(2011\)](#), [Duffee \(2011\)](#), and the references cited therein.

2.2 Four-factor Nelson-Siegel-Svensson Model

[Svensson \(1994\)](#) extends the Nelson-Siegel-Svensson (NSS) model by adding a fourth term, which is designed to allow for a second “hump” in the yield curve. In particular, he proposes the following four-factor model for fitting the instantaneous forward interest rate:

$$f(m) = \beta_1 + \beta_2 \cdot \exp\left(\frac{m}{\theta_{1,t}}\right) + \beta_3 \cdot \left[\left(\frac{m}{\theta_{1,t}}\right) \cdot \exp\left(\frac{m}{\theta_{1,t}}\right)\right] + \beta_4 \cdot \left[\left(\frac{m}{\theta_{2,t}}\right) \cdot \exp\left(\frac{m}{\theta_{2,t}}\right)\right].$$

Notice that in the above equation there are now two different decay parameters, which control the double-hump shape of the forward curve. These are called θ_1 and θ_2 . Similar to the DNS model, we consider the following dynamic version of the NSS model.

$$\begin{aligned} \hat{y}_t(\tau) = & \hat{\beta}_{1,t} + \hat{\beta}_{2,t} \cdot \left[\frac{1 - \exp(-\lambda_{1,t}\tau)}{\lambda_{1,t}\tau}\right] + \hat{\beta}_{3,t} \cdot \left[\frac{1 - \exp(-\lambda_{1,t}\tau)}{\lambda_{1,t}\tau} - \exp(-\lambda_{1,t}\tau)\right] \\ & + \hat{\beta}_{4,t} \cdot \left[\frac{1 - \exp(-\lambda_{2,t}\tau)}{\lambda_{2,t}\tau} - \exp(-\lambda_{2,t}\tau)\right], \end{aligned}$$

where we now have two decay parameters, as discussed above. These are called $\lambda_{1,t}$ and $\lambda_{2,t}$. As detailed in [De Pooter \(2007\)](#), the second hump in the NSS model is difficult to identify without imposing additional increase in β_{3t} , the “curvature” component, increases medium-term yields and has little effect on the short and long end of the curve, inducing a hump shape in the yield curve. As demonstrated in [Diebold and Li \(2006\)](#), the “level” factor can be approximated with the 10-year bond yield, the “slope” factor can be approximated with 10-year - 3-month bond yield spreads, and the “curvature” factor moves closely with two times the 2-year yield minus the sum of the 3-month and 10-year yields.

restrictions. We adopt his approach to solving this issue, which includes assumptions that the two humps are at least one year apart, and that the second hump reaches its maximum for a maturity which is at least twelve months shorter than the first hump. Additionally, it is assumed that $\lambda_1 \neq \lambda_2$, in order to avoid multicollinearity. Finally, in our experiments, we estimate both static and dynamic rates of decay, with dynamic rates estimated using recursive nonlinear least squares. See Section 4 for further details on model estimation. Predictions of the betas and yields using this model follow the same approach that discussed above in the context of DNS models.

3 Unspanned Risks and Diffusion Indexes

Whether or not macroeconomic, financial and other non-yield information is useful for fitting and forecasting the yield curve remains an open question. As [Duffee \(2013\)](#) points out, yields are usually hypothesized to follow a Markov process, which implies that all information in fundamental economic variables should already be embedded in yield cross sections. This leaves no role for so-called “unspanned risks”, as proxied for by additional economic variables and/or diffusion indexes constructed in a data rich environment. Namely, he argues that, at least theoretically, it is redundant to add current non-yield information in forecast equations for interest rates. On the other hand, in the empirical literature there are many examples where adding economic variables and diffusion indexes has proven to be effective in improving yield curve fitting as well as forecasting. For example, [Ang and Piazzesi \(2003\)](#) find that macroeconomic variables are significant for explaining Treasury security yield dynamics, based on a VAR analysis. [Moench \(2008\)](#) shows that adding estimated diffusion indexes to an affine Gaussian term structure model improves out-of-sample forecast performance. [Diebold et al. \(2006\)](#) investigate bi-directional causality between yield betas and macroeconomic variables, and discovers strong evidence in favor of linkages between such variables and future yield curve dynamics.

Recently, focus has turned to so-called big data, and to the analysis of the usefulness of large scale datasets in the above context. As noted in [Bernanke and Boivin \(2003\)](#), monetary policy-makers and academics alike are very interested in examining the (predictive) usefulness of a wide range of variables in a data-rich environments. Indeed, the predictive content of diffusion indexes constructed using large scale datasets has been examined in countless academic papers in the past few years. The same is certainly true in industry, where the prevalence of big data and related machine learning methods is readily apparent.

The use of diffusion indexes in our forecasting models is motivated by a number of important papers, including [Stock and Watson \(2002a,b\)](#) and [Bai and Ng \(2002, 2008, 2009\)](#). Namely, let X_{tj} be the observed real-time datum for the j -th cross-sectional unit available at time t , for $t = 1, \dots, T$ and $j = 1, \dots, N$, and let:

$$X_{tj} = \Lambda_j' F_t^x + e_{tj}, \quad (3.1)$$

where F_t^x is an $r \times 1$ vector of common factors, Λ_j is an $r \times 1$ vector of factor loadings associated with F_t^x , and e_{tj} is the idiosyncratic component of X_{tj} .⁹ The product $\Lambda_j' F_t^x$ is the common component of X_{tj} , and is the dimension reducing factor representation of the data. In the sequel, our generic forecasting equation that utilizes latent factors from the above model is:

$$y_{t+h}(\tau) = c(\tau) + \delta_y' W_t + \delta_x' F_t^x + \varepsilon_{t+h}, \quad (3.2)$$

where h is the forecast horizon and W_t is a vector of explanatory variables. Additionally, the parameters δ_W and δ_F are defined conformably, and ε_{t+h} is a disturbance term. Following [Bai and Ng \(2002, 2006, 2008, 2009\)](#), the whole $T \times N$ panel of data, X , is first used to consistently estimate F_t^x , via either targeted or un-targeted PCA. Thereafter, (3.2) is estimated using least squares, and forecasts are constructed. Of note is that if $\sqrt{T}/N \rightarrow 0$, then the usual generated regressor problem does not arise, in the sense that $\hat{\delta}_x'$ and $\hat{\delta}_y'$ are \sqrt{T} consistent and asymptotically normal (see [Bai and Ng \(2008\)](#)).

In this setup, PCA is carried out using the standard eigenvalue-eigenvector decomposition or singular value decomposition method discussed in the above papers, and is based on the use of all of the predictors in our real-time dataset.

One important aspect of our setup is the use of real-time data. Recently, [McCracken and Ng \(2016\)](#) and St. Louis Federal Reserve Bank’s data desk created the FRED-MD, which is a large monthly real-time database that contains the entire revision history for over 130 macroeconomic variables. The dataset contains variables summarizing economic output and income, labor markets, consumption, money and credit, housing, and the stock market, for example. [McCracken and Ng \(2016\)](#) show that diffusion indexes extracted from their FRED-MD dataset contain the same predictive content as diffusion indexes constructed using the classic “fully revised” Stock and Watson dataset ([Stock and Watson, 2002a,b](#)). However, the FRED-MD is a real-time database, while the Stock and Watson dataset contains only fully revised data. Several studies have revealed the importance of collecting and updating such real-time datasets including [Diebold and Rudebusch \(1991\)](#), [Hamilton and Perez-Quiros \(1996\)](#), [Bernanke and Boivin \(2003\)](#), and the papers cited therein. We ask whether “information-rich” diffusion indexes are useful for predicting yields, when the data used to construct the indexes are purely “real-time”, rather than fully revised, as in [Swanson and Xiong \(2018\)](#). In this sense, we revisit the finding of [McCracken and Ng \(2016\)](#) that diffusion indexes extracted from real-time and “fully revised” datasets contain the same predictive content.

In addition, we consider the use of targeted predictors, when carrying out PCA, as discussed in [Bai and Ng \(2008\)](#). The use of targeted predictors involves the specification of a subset of X for use in PCA. In our experiments, this “subset” of variables is selected using either the lasso or the elastic net. Summarizing, “targeted PCA” is the same as PCA, except that a new subset of variables from the original

⁹All variables are standardized in this setup.

real-time dataset, X , is used at each point in time to construct the latent factors. This sort of variable “pre-selection” is computationally intensive, as the shrinkage operators must be applied at each point in time, prior to the estimation of each forecasting model and construction of each prediction; and this must be done for each targeted interest rate being predicted. Nevertheless, the computations involved are simple, as implementation of the lasso and elastic net simply involve constructing the following penalized regression estimators:

$$\hat{\theta}_{lasso} = \arg \min_{\theta} \left\{ \|y - \sum_{i=1}^N X_i \theta_i\|^2 + \lambda_1 \sum_{i=1}^N |\theta_i| \right\},$$

and

$$\hat{\theta}_{elastic\ net} = (1 + \lambda_2) \arg \min_{\theta} \left\{ \|y - \sum_{i=1}^N X_i \theta_i\|^2 + \lambda_1 \sum_{j=1}^N |\theta_j| + \lambda_2 \sum_{j=1}^N \theta_j^2 \right\},$$

where y denotes the $T \times 1$ vector of time series observations (i.e., $(y_{1+h}, \dots, y_{T+h})$) on the target variable being forecasted, X_i (i.e., $(X_{i,1}, \dots, X_{i,T})$), denotes the time series for a particular variable in X , and the θ_i are the coefficients to be estimated. In this context, the subset of variables used in targeted PCA is simply the set of variables that are associated with non-zero estimates of θ_i , after implementing one of the above estimators. These estimators facilitate variable selection because the use of an *L1-norm* penalty function ensures that parameter estimators may be identically zero. More specifically, for the elastic net, there are two tuning parameters (rather than 1, as is the case with the lasso). These parameters control the weights placed on the *L1*- and *L2-norm* penalty functions in the estimators. Ten-fold cross validation is used, in real-time, to estimate the tuning parameter in the lasso operator and the first tuning parameter in the elastic net operator, while the second parameter in the elastic net is fixed at different values, including 0.2, 0.4, 0.6, and 0.8. For a complete description of these operators and targeted prediction in general, refer to [Schumacher \(2007, 2010\)](#), [Bai and Ng \(2008, 2009\)](#), [Kim and Swanson \(2014\)](#), [Carrasco and Rossi \(2016\)](#), [Ghysels and Marcellino \(2018\)](#).

For further discussion of diffusion indexes in macroeconomic forecasting, see [Boivin and Ng \(2005\)](#), [Banerjee et al. \(2008\)](#), and [Kim and Swanson \(2014\)](#). For a discussion of alternative modeling approaches that are of interest in the current context, refer to [Coroneo et al. \(2016\)](#).

4 Empirical Setup

All models analyzed in this paper are estimated using rolling windows of 120 monthly observations. Thus, all models are re-estimated prior to the construction of each new h -month ahead forecast. In our experiments, 1-month, 3-month, and 12-month ahead yield predictions are constructed for 3-month, 1-year, 3-year, 5-year, and 10-year maturity bonds. We do not forecast maturities longer than 10-years, due to liquidity issues in bonds with maturity longer than 10 years. Empirical results are reported for four different prediction periods, including: 2001:1-2005:12 (Subsample 1), 2006:1-2010:12 (Subsample 2), 2011:1-2017:10 (Subsample 3), and 2001:1-2017:10 (Full Sample). In addition to analysis based on

mean square forecasts errors and Diebold and Mariano (1995) tests for these subsamples, we also carry out Giacomini and Rossi (2010) fluctuations tests, in order to address the issue of the selection of cut-off dates for our subsamples. All model parameters are estimated with maximum likelihood and PCA.

For a general analysis of the use of rolling versus recursive (and other) windowing techniques in the context of forecasting, see Clark and McCracken (2009), Hansen and Timmermann (2012), and Rossi and Inoue (2012). For an analysis of the bias that may arise in the context of small samples when estimating the autoregressive parameters of yield curve factors, see Bauer et al. (2012). In light of the potential for estimator bias to be prevalent in our experiments, we also estimated all of our models using recursive data windowing. Although not reported here (results are available upon request), our rolling window strategy yielded lower MSFEs in approximately 65% of cases considered, and in the majority of the remaining cases, MSFEs were not appreciably different, regardless of method used.

Finally, note that when estimating DNS and NSS models, we use three alternative yield cross sections, including yields from either 10, 12, or 30 different bond maturities (for further details, refer to Section 2.1). Since all of our empirical results are qualitatively similar regardless of the cross section used to estimate our DNS and NSS models, we only report results for the case where models are estimated using yields from 12 different bonds ranging from 3-month t-bills to the 10-year bonds.¹⁰

4.1 Models Used in Forecasting Experiments

A summary of the models used in our prediction experiments is given below.

4.1.1 Small Data Models

Autoregressive (AR) and Vector Autoregressive (VAR) Models:

Models in this section are summarized in Table 1, and include AR(1), VAR(1), AR(SIC), and VAR(SIC) models.

We utilize a number of benchmark time series models, specified as follows:

$$y_{t+h}(\tau) = c(\tau) + \delta'_y W_t + \varepsilon_{t+h}, \quad (4.1)$$

where τ denotes the maturity of the bond (bill) for which the scalar, $y_{t+h}(\tau)$, measures the annual yield. Additionally, W_t contains lags of $y_t(\tau)$ in autoregressive specifications, and contains lags of $y_t(\tau)$ and additional explanatory variables in vector autoregressive specifications, with δ_y a conformably defined coefficient vector, and $c(\tau)$ is a constant term.¹¹ In AR and VAR specifications, up to 5 lags of $y_t(\tau)$

¹⁰Bond maturities in the cross sections used to estimate the betas in our DNS and NSS models include 1 to 10 year maturity yields (10-dimensional dataset), 3 month, 6-month, and 1 to 10 year maturity yields (12-dimensional dataset), and 1 to 30 year maturity yields (30-dimensional dataset).

¹¹When specifying VAR models, the above equation constitutes only one (τ -maturity) equation in the VAR. As the same set of explanatory variables is utilized in each equation in the VAR, the well known SUR (seemingly unrelated regression)

are included, with the number of lags selected using the Schwarz information criterion (SIC). In addition to AR(SIC) and VAR(SIC) models, straw-man AR(1) and VAR(1) models are estimated. Additionally, in our unrestricted VAR models, W_t includes five bond yields, with maturities 3 months, 1 year, 3 years, 5 years, and 10 years.

Dynamic Nelson-Siegel (DNS) Models:

Models in this section are summarized in Table 1, and include DNS(1), DNS(2), DNS(3), DNS(4), DNS(5), and DNS(6) models.

As discussed above, the DNS model introduced by [Diebold and Li \(2006\)](#) is a dynamic version of the model introduced in [Nelson and Siegel \(1987\)](#), where cross-sectional movements in the term structure are summarized by the dynamics of three underlying latent factors (betas) interpreted as “level”, “slope”, and “curvature” factors. We refer to the three betas as “Nelson-Siegel factors” (NS-factors), and in our prediction experiments, both AR(1) and VAR(1) type models are specified in order to predict these factors for subsequent use in the prediction of $y_{t+h}(\tau)$.

We estimate the latent factors by fitting the following regression:

$$y_t(\tau) = \beta_{1,t} + \beta_{2,t} \cdot \left[\frac{1 - \exp(-\lambda_t \tau)}{\lambda_t \tau} \right] + \beta_{3,t} \cdot \left[\frac{1 - \exp(-\lambda_t \tau)}{\lambda_t \tau} - \exp(-\lambda_t \tau) \right] + \varepsilon_t, \quad (4.2)$$

As discussed above, we utilize yield cross sections that include 10, 12, and 30 different yield maturities. Predictions of y_{t+h} are constructed using the model:

$$\hat{y}_{t+h}(\tau) = \hat{\beta}_{1,t+h}^f + \hat{\beta}_{2,t+h}^f \cdot \left[\frac{1 - \exp(-\lambda_t \tau)}{\lambda_t \tau} \right] + \hat{\beta}_{3,t+h}^f \cdot \left[\frac{1 - \exp(-\lambda_t \tau)}{\lambda_t \tau} - \exp(-\lambda_t \tau) \right], \quad (4.3)$$

where $y_{t+h}(\tau)$ is a scalar, and $\hat{\beta}_{1,t+h}^f$, $\hat{\beta}_{2,t+h}^f$, and $\hat{\beta}_{3,t+h}^f$ are predictions constructed by specifying simple AR(1) models for $\hat{\beta}_{1,t}$, $\hat{\beta}_{2,t}$, and $\hat{\beta}_{3,t}$, including:

$$\hat{\beta}_{i,t+h}^f = \hat{c}_i + \hat{\gamma}_{y,i} \hat{\beta}_{i,t}, \quad \text{for } i = 1, 2, 3, \quad (4.4)$$

where $\hat{\beta}_{i,t+h}^f$, $\hat{\beta}_{i,t}$, \hat{c}_i and $\hat{\gamma}_i$ are scalars. We also construct predictions by using the following VAR(1) model:

$$\hat{\beta}_{t+h}^f = \hat{c}_y + \hat{\Gamma}_y \hat{\beta}_t, \quad (4.5)$$

where $\hat{\beta}_{t+h}^f = (\hat{\beta}_{1,t+h}^f, \hat{\beta}_{2,t+h}^f, \hat{\beta}_{3,t+h}^f)'$, \hat{c} is a 3×1 vector, and $\hat{\Gamma}_y = (\hat{\gamma}_1, \hat{\gamma}_2, \hat{\gamma}_3)$, with $\hat{\gamma}_j$ a 3×1 vector, for $j = 1, 2, 3$. In our experiments, the decay parameter is estimated both statically and dynamically (prior to the construction of each new prediction). For prediction models with a static rate of decay (i.e., models DNS(1) and DNS(4) in Table 1), λ_t is set equal to 0.0609, as in [Diebold and Li \(2006\)](#). DNS(2), DNS(3), DNS(5), and DNS(6)) utilize a dynamically estimated decay parameter, which is estimated as follows. First, a grid search for the decay parameter ($\frac{1}{\lambda_t}$) is carried out on the domain of (6.69, 33.46),

result ensures that consistent and efficient parameter estimates are obtained via equation by equation estimation using least squares.

which corresponds to the domain of a “curvature hump” of one to five years. The range for the grid search is selected on the basis of bond maturities.¹² Next, NS-factors are calculated with the selected rate of decay for the “curvature” factor that minimizes squared in-sample fitted errors. Finally, either an AR(1) or VAR(1) models are estimated in order to generate forecasts of the NS-factors, as discussed above.

Dynamic Nelson-Siegel-Svensson (NSS) Models:

Models in this section are summarized in Table 1, and include NSS(1), NSS(2), NSS(3), NSS(4), NSS(5), and NSS(6) models.

The dynamic Nelson-Siegel-Svensson (NSS) model is included in our prediction experiments because it is one of the most widely used in zero-coupon yield curve construction by major central banks (see BIS (2005)). As discussed above, in the model, Svensson (1994) adds an additional factor to the classic 3-factor Nelson-Siegel model that captures a second “curvature hump”. In our experiments, the four latent factors are referred to as “Nelson-Siegel-Svensson factors” (NSS-factors). Although Svensson did not analyze a dynamic version of his model in his original paper, we do so, following the approach of Diebold and Li (2006). The framework of our prediction experiments using the NSS model is, thus, analogous to that discussed above in the case of DNS model. In particular, estimates of the NSS-factors (i.e. $\beta_{1,t}$, $\beta_{2,t}$, $\beta_{3,t}$, and $\beta_{4,t}$) are constructed at each point in time by regressing $(1, \frac{1 - \exp(-\lambda_{1,t}\tau)}{\lambda_{1,t}\tau}, \frac{1 - \exp(-\lambda_{1,t}\tau)}{\lambda_{1,t}\tau} - \exp(-\lambda_{1,t}\tau), \frac{1 - \exp(-\lambda_{2,t}\tau)}{\lambda_{2,t}\tau} - \exp(-\lambda_{2,t}\tau))$ on $y_t(\tau)$. Additionally, the model is now:

$$y_t(\tau) = \beta_{1,t} + \beta_{2,t} \cdot \left[\frac{1 - \exp(-\lambda_{1,t}\tau)}{\lambda_{1,t}\tau} \right] + \beta_{3,t} \cdot \left[\frac{1 - \exp(-\lambda_{1,t}\tau)}{\lambda_{1,t}\tau} - \exp(-\lambda_{1,t}\tau) \right] + \beta_{4,t} \cdot \left[\frac{1 - \exp(-\lambda_{2,t}\tau)}{\lambda_{2,t}\tau} - \exp(-\lambda_{2,t}\tau) \right] + \varepsilon_t, \quad (4.6)$$

Resultant sequences of estimates, $\hat{\beta}_{1,t}$, $\hat{\beta}_{2,t}$, $\hat{\beta}_{3,t}$, and $\hat{\beta}_{4,t}$, for $t = 1, \dots, T$ are used to construct predictions of $y_{t+h}(\tau)$ using:

$$\hat{y}_{t+h}(\tau) = \hat{\beta}_{1,t+h}^f + \hat{\beta}_{2,t+h}^f \cdot \left[\frac{1 - \exp(-\lambda_{1,t}\tau)}{\lambda_{1,t}\tau} \right] + \hat{\beta}_{3,t+h}^f \cdot \left[\frac{1 - \exp(-\lambda_{1,t}\tau)}{\lambda_{1,t}\tau} - \exp(-\lambda_{1,t}\tau) \right] + \hat{\beta}_{4,t+h}^f \cdot \left[\frac{1 - \exp(-\lambda_{2,t}\tau)}{\lambda_{2,t}\tau} - \exp(-\lambda_{2,t}\tau) \right] \quad (4.7)$$

where $y_{t+h}(\tau)$ is a scalar, and $\hat{\beta}_{1,t+h}^f$, $\hat{\beta}_{2,t+h}^f$, $\hat{\beta}_{3,t+h}^f$, and $\hat{\beta}_{4,t+h}^f$ are predictions constructed by specifying simple AR models:

$$\hat{\beta}_{i,t+h}^f = \hat{c}_i + \hat{\gamma}_{y,i} \hat{\beta}_{i,t}, \quad \text{for } i = 1, 2, 3, 4 \quad (4.8)$$

where $\hat{\beta}_{i,t}^f$, $\hat{\beta}_{i,t}$, \hat{c}_i , and $\hat{\gamma}_{y,i}$ are scalars. We also construct NSS-factor predictions by using the following

¹²We find that setting domains too wide results in occasional ‘extreme’ estimates for NS-factors, which in turn leads to occasional poor yield forecasts. For further discussion, see below. For an extensive discussion of decay parameter estimation, refer to De Pooter (2007).

VAR(1) model:

$$\hat{\beta}_{t+h}^f = \hat{c}_y + \hat{\Gamma}_y \hat{\beta}_t, \quad (4.9)$$

where $\hat{\beta}_{t+h}^f = (\hat{\beta}_{1,t+h}^f, \hat{\beta}_{2,t+h}^f, \hat{\beta}_{3,t+h}^f, \hat{\beta}_{4,t+h}^f)'$, \hat{c}_y is a 4×1 vector, and $\hat{\Gamma}_y$ is a 4×4 matrix of constants. To estimate NSS model parameters, again, two estimation methods for the decay parameters are utilized. In the case of a fixed (static) decay parameter, $\lambda_{1,t}$ is equal to 0.0609, which is the same value as that used when estimating our three-factor DNS model; and the second rate of decay, $\lambda_{2,t}$, is set equal to 0.2985, corresponding to a second curvature hump at approximately 6 months.¹³ The subsequent forecasting procedure used to construct yield predictions is the same as that discussed above for our DNS models.

4.1.2 Big Data Models

AR and VAR Models with Macro Diffusion Indexes:

Models in this section are summarized in Table 1, and include AR(1)+FB1, AR(1)+FB2, AR(1)+FB3, VAR(1)+FB1, VAR(1)+FB2, and VAR(1)+FB3 models.

We utilize the prediction model given in equation (4.1), with added diffusion index (i.e. F_t^x) regressors. (Refer to Section 3 for a discussion of diffusion indexes.) In particular, we estimate variants of the following factor augmented forecasting model:

$$y_{t+h}(\tau) = c(\tau) + \delta_y' W_t + \delta_x' F_t^x + \varepsilon_{t+h}, \quad (4.10)$$

where F_t^x includes either 1, 2 or 3 diffusion indexes, and W_t is defined as above. Here, $c(\tau)$ is a constant term, and δ_y and δ_x are conformably defined vectors of coefficients. Just as in the case of our benchmark AR models, we also consider VAR variants of the above model. Note that although multiple yield lags were tried when specifying W_t in these models, our ‘MSFE-best’ models always included only the first lag of the yield(s). For this reason all empirical results discussed in the sequel use one lag of yield variables in AR and VAR models of this variety. Finally, as discussed in detail Section 3, the real-time diffusion indexes appearing in the above equation are constructed using both un-targeted and targeted PCA.

DNS Models with Macro Diffusion Indexes:

Models in this section are summarized in Table 1, and include DNS(1)+FB1, DNS(2)+FB1, DNS(3)+FB1, DNS(4)+FB1, DNS(5)+FB1, DNS(6)+FB1, DNS(1)+FB2, DNS(2)+FB2, DNS(3)+FB2, DNS(4)+FB2, DNS(5)+FB2, DNS(6)+FB2, DNS(1)+FB3, DNS(2)+FB3, DNS(3)+FB3, DNS(4)+FB3, DNS(5)+FB3, DNS(6)+FB3 models.¹⁴

¹³Restrictions on the decay parameters for the NSS model are imposed in order to ensure that the two curvature humps are at least one year apart, for identification purposes. In addition to the restriction that $\frac{1}{\lambda_{1,t}} \geq \frac{1}{\lambda_{2,t}} + 6.69$ (see De Pooter (2007)), restrictions are imposed on the domain of the two decay parameters. Namely, the grid search for the first decay parameter, $\frac{1}{\lambda_{1,t}}$, is over the domain of (6.69, 33.46); and for the second decay parameter, $\frac{1}{\lambda_{2,t}}$ is over the grid (0, 26.77). These restrictions ensure identification of two curvature factors individually, and avoid ‘extreme’ estimates for NSS-factors.

¹⁴FB1, FB2, and FB3 denote models that have been augmented to include either 1, 2, or 3 diffusion indexes, respectively.

For this class of models, real-time diffusion indexes are added to the models used to predict the betas, which are subsequently used in the construction of yield forecasts, as discussed at the beginning of this section. Namely, yield predictions are constructed using betas that are formed as follows:

$$\widehat{\beta}_{i,t+h}^f = \hat{c}_i + \hat{\gamma}'_{y,i} \widehat{\beta}_{i,t} + \hat{\gamma}'_{x,i} \hat{F}_t^x, \quad \text{for } i = 1, 2, 3,$$

where F_t^x again includes either 1, 2 or 3 diffusion indexes. All other terms are conformably defined. We also construct predictions by using the following VAR(1) variant of this model:

$$\widehat{\beta}_{t+h}^f = \hat{c}_y + \hat{\Gamma}_y \widehat{\beta}_t + \hat{\Gamma}_x \hat{F}_t^x,$$

where $\widehat{\beta}_{t+h}^f = (\widehat{\beta}_{1,t+h}^f, \widehat{\beta}_{2,t+h}^f, \widehat{\beta}_{3,t+h}^f)'$, \hat{c}_y is 3×1 vector, $\hat{\Gamma}_y = (\hat{\gamma}_1, \hat{\gamma}_2, \hat{\gamma}_3)$, with $\hat{\gamma}_j$ a 3×1 vector, for $j = 1, 2, 3$, and $\hat{\Gamma}_x$ is a conformably defined matrix of constants.

NSS Models with Macro Diffusion Indexes:

Models in this section are summarized in Table 1, and include: NSS(1)+FB1, NSS(2)+FB1, NSS(3)+FB1, NSS(4)+FB1, NSS(5)+FB1, NSS(6)+FB1, NSS(1)+FB2, NSS(2)+FB2, NSS(3)+FB2, NSS(4)+FB2, NSS(5)+FB2, NSS(6)+FB2, NSS(1)+FB3, NSS(2)+FB3, NSS(3)+FB3, NSS(4)+FB3, NSS(5)+FB3, NSS(6)+FB3.

Analogous to the DNS models discussed above, NSS-type yield predictions are constructed using betas formed as follows:

$$\widehat{\beta}_{i,t+h}^f = \hat{c}_i + \hat{\gamma}'_{y,i} \widehat{\beta}_{i,t} + \hat{\gamma}'_{x,i} F_t^x, \quad \text{for } i = 1, 2, 3, 4$$

where F_t^x includes either 1, 2 or 3 diffusion indexes, and all other terms are defined above. We also construct predictions using the following VAR(1) variant of this model:

$$\widehat{\beta}_{t+h}^f = \hat{c}_y + \hat{\Gamma}_y \widehat{\beta}_t + \hat{\Gamma}_x F_t^x,$$

where $\widehat{\beta}_{t+h}^f = (\widehat{\beta}_{1,t+h}^f, \widehat{\beta}_{2,t+h}^f, \widehat{\beta}_{3,t+h}^f, \widehat{\beta}_{4,t+h}^f)'$, \hat{c} is 4×1 vector, and $\hat{\Gamma}_y = (\hat{\gamma}_1, \hat{\gamma}_2, \hat{\gamma}_3, \hat{\gamma}_4)$, $\hat{\gamma}_j$ is a 4×1 vector, for $j = 1, 2, 3, 4$ and $\hat{\Gamma}_x$ is a conformably defined matrix of constants.

As a final note, it is worth mentioning that all macroeconomic variables are standardized to zero mean and unit variance before principal component analysis is carried out in order to construct diffusion indexes.

Forecast Combination

In our experiments, we also construct and analyze various forecast combinations. The particular combinations are detailed in Table 5. There are two reasons why we include combinations in our analysis. First, it is well known that forecast combination is useful in time series prediction. As shown in [Kim and Swanson \(2014\)](#), [Carrasco and Rossi \(2016\)](#), and [Hirano and Wright \(2017\)](#), much can be gained by combining forecasts from models estimated using a wide range of methods ranging from least squares

estimation to machine learning.¹⁵ More importantly, it turns out that while combination does not play an important role when comparing DNS and NSS type models with and without diffusion indexes if fully revised data are used in model and prediction construction, as discussed in [Swanson and Xiong \(2018\)](#), the same is not true when real-time data are used in our data rich environment. Indeed, we shall see that various forecast combinations dominate all of the models discussed above, when real-time data are utilized. This is important because it suggests that the use of fully revised data may be quite misleading in the types of experiments carried out in this paper.

Note that all of our forecast combinations are averages, as detailed in Table 5. However, we also considered some simple alternative weighting methods.¹⁶ First, we considered letting $w_{t+1} = \Sigma_t^{-1} \mathbf{1} / (\mathbf{1}^\top \Sigma_t^{-1} \mathbf{1})$, where Σ_t is the variance-covariance matrix of the out-of-sample forecasting errors, up to time t , and $\mathbf{1}$ is a column vector.¹⁷ In theory, this weight can minimize the variance of the forecast combination, provided that the matrix Σ_t is invertible and stable. However, in our forecasting experiments, the matrix Σ_t is nearly singular at many time periods. Hence, our empirical results upon implementation of this estimator were not very stable and sometimes generated large forecast errors. Another option which we considered is to replace Σ_t by its diagonal matrix. This method has been considered by [Timmermann \(2006\)](#). It is worth noting that in most cases where we tried this method, the equal weight method yielded smaller MSFEs, regardless of which models were combined. For this reason, these findings are not reported, although complete results are available upon request. In closing this section, it should be stressed that other methods can be utilized to construct combinations. Further investigation into this topic is left to future research, however.

4.2 Predictive Accuracy Testing

When comparing the predictive performance of the models detailed below, we report the MSFE, defined as:

$$\text{MSFE}(\tau) = \frac{1}{P} \sum_{t=1}^P (\hat{y}_{t+h}(\tau) - y_{t+h}(\tau))^2, \quad (4.11)$$

where $\hat{y}_{t+h}(\tau)$ is the h -step-ahead forecast of a Treasury bond yield, with maturity τ , and P is the number of ex ante predictions used in our analysis. Model MSFEs are compared using the [Diebold and Mariano \(1995\)](#) predictive accuracy test. The null hypothesis of the DM test is: $H_0 : E[L(\epsilon_{t+h}^{(1)})] - E[L(\epsilon_{t+h}^{(2)})] = 0$, where the $\epsilon_{t+h}^{(i)}$ is the prediction error of model i , for $i = 1, 2$. In our analysis, $L(\cdot)$ is the quadratic loss function. The DM test statistic is:

$$\text{DM}(\tau) = P^{-1} \sum_{t=1}^P \frac{d_{t+h}(\tau)}{\hat{\sigma}_d}, \quad (4.12)$$

¹⁵For a discussion of forecast combination using the types of factor augmented regressions discussed in this paper, see [Cheng and Hansen \(2015\)](#).

¹⁶We thank an anonymous referee for the suggestion that we broaden the scope of our analysis of combination methods.

¹⁷The length of the forecasting error sequence is the same as the window size in the recursive and rolling window scheme

where $d_{t+h}(\tau) = [\hat{\epsilon}_{t+h}^{(1)}(\tau)]^2 - [\hat{\epsilon}_{t+h}^{(2)}(\tau)]^2$, \bar{d} denotes the mean of $d_{t+h}(\tau)$, $\hat{\sigma}_{\bar{d}}$ is a heteroskedasticity and autocorrelation consistent estimate of the standard deviation of \bar{d} , and P denotes the number of ex ante predictions used to construct the test statistic. In the sequel, we assume that the DM_P test is asymptotically $N(0,1)$, although in cases where models being compared are nested, modified critical values tabulated by [McCracken \(2000\)](#) should be used (see [Corradi and Swanson \(2006\)](#) for complete details). For an interesting discussion of alternative approaches to assessing forecasting performance, see [Rossi and Sekhposyan \(2011\)](#). Additionally, a novel alternative approach for comparing predictive accuracy using fixed-smoothing asymptotics is discussed in [Coroneo and Iacone \(2018\)](#).

In a small experiment, noting that the performance of the DM test statistics may not be well approximated by its asymptotic distribution, we also constructed DM test critical values using the block resampling scheme of [Kunsch \(1989\)](#), as discussed in the predictive accuracy testing context in [Corradi and Swanson \(2007\)](#). The qualitative findings and conclusions based on our empirical analysis did not change when these alternative critical values were used. Complete results of these experiments are available upon request.¹⁸

Since our empirical findings are reported for various data subsamples, we also carry out a robustness check on our findings. Namely, we construct Giacomini-Rossi fluctuation tests due to [Giacomini and Rossi \(2010\)](#). Following their notation, we define $L^{(1)}(y_t, \hat{\theta}_{t-h,R})_{t=R+h}^T$ and $L^{(2)}(y_t, \hat{\theta}_{t-h,R})_{t=R+h}^T$ to be sequences of forecast losses associated with the benchmark model and alternative models, respectively. Let $\Delta L_t(\hat{\theta}_{t-h,R}, \hat{\gamma}_{t-h,R})$ be the difference of these two sequences. Then, using a local window of size m , we compute the following test statistics:

$$F_{t,m}^{\text{OOS}} = \hat{\sigma}^{-1} m^{-1/2} \sum_{s=t-m/2}^{t+m/2-1} \Delta L_s(\hat{\theta}_{s-h,R}, \hat{\gamma}_{s-h,R}),$$

where $\hat{\sigma}^2$ is a HAC estimator of the asymptotic variance of $m^{-1/2} \sum_{s=t-m/2}^{t+m/2-1} \Delta L_s(\hat{\theta}_{s-h,R}, \hat{\gamma}_{s-h,R})$. These statistics are Giacomini-White (GW) predictive accuracy test statistics, as discussed in [Giacomini and White \(2006\)](#). The null hypothesis of the fluctuation test is given by $H_0 : \mathbb{E}[\Delta L_t(\hat{\theta}_{t-h,R}, \hat{\gamma}_{t-h,R})] = 0$, for all $t = R+h, \dots, T$. The corresponding rejection rule for the one-sided test against the alternative $\mathbb{E}[\Delta L_t(\hat{\theta}_{t-h,R}, \hat{\gamma}_{t-h,R})] > 0$ is $\max_t F_{t,m}^{\text{OOS}} > \kappa_\alpha$, where κ_α solves:

$$\mathbb{P}\left(\sup_{\tau} \frac{1}{\sqrt{\mu}} (\mathcal{B}(\tau + \mu/2) - \mathcal{B}(\tau - \mu/2)) > \kappa_\alpha\right) = \alpha.$$

In our experiment, we set $\mu = 0.2$, which means that the local window size, m , is approximately 20% of the forecasting sample. According to Table 1 of [Giacomini and Rossi \(2010\)](#), the critical values are 2.938

¹⁸The valid use of the block bootstrap involves showing that

$$\sup_{v \in \mathbb{R}} |\mathbb{P}^*(\sqrt{P}(DM_P(\tau)^* - DM_P(\tau)) \leq v) - \mathbb{P}(\sqrt{P}(DM_P(\tau)) \leq v)| \xrightarrow{\mathbb{P}} 0,$$

where P^* denote the bootstrap probability measure, and $DM_P(\tau)^*$ is the bootstrap DM test statistic. Bootstrap critical values are obtained from inspection of the empirical distribution of M re-sampled values of $DM_P(\tau)^* - DM_P(\tau)$.

and 2.676, when $\alpha = 0.05$ and 0.10, respectively. It is worth noting that this is a supremum type test, in the spirit of Kolmogorov–Smirnov tests, for example. If we reject the null in favor of the above one-sided alternative, then we know that the alternative model significantly outperforms the benchmark model, at some point in the forecast sample. Since critical values are obtained from the distribution of the supreme statistics, they are not valid for assessing the relative performance of the two models at any given time point. Typically, the critical value for a supremum type test will be much larger than that associated with a corresponding point-wise test (of course, sequential testing bias becomes an issue when sequences of point-wise tests are used to make statistical inferences).

5 Data

Our term structure data are monthly U.S. zero-coupon (end of month) yield curve data reported by the Federal Reserve Board (see <https://www.quandl.com/data/FED/SVENY-US-Treasury-Zero-Coupon-Yield-Curve> and [Gürkaynak et al. \(2007\)](#) (GSW)). In particular, we utilize GSW monthly data for the period from August 1988 through October 2017, which contains data on 1 to 30 years maturity bond yields. In addition to GSW zero-yields, 3- and 6-months T-bill yields are utilized in order to “fill-out” the short end of the yield curve.¹⁹ Hence, we analyze a panel of data containing $N = 32$ yields and $T = 351$ monthly observations.

While Dickey-Fuller tests cannot reject the null hypothesis of a unit root in yields, preliminary forecast experiments using both yields and first-difference yields resulted in little difference when comparing MSFEs of yield predictions. Moreover, finance theory is not consistent with the presence of a unit root yield processes. For these reasons, we do not difference yields in our experiments.

As discussed above, real-time macroeconomic diffusion indexes are constructed using PCA. The dataset used is the FRED-MD dataset, which is a real-time monthly database of over 130 macroeconomic time series, as discussed in Section 2.²⁰ As discussed above, one advantage of the FRED-MD dataset is that all time series are updated monthly by the Federal Reserve Bank of St. Louis. Thus, researchers have truly real-time data available for conducting forecasting experiments, in which all vintages (revisions) of all variables are available. Use of such data ensures that future information cannot inadvertently be used to revise data from prior periods, which is a serious potential problem with non-real-time or fully revised data. (Refer to [Stark \(2010\)](#) for further discussion of real-time datasets.) Moreover, fully revised datasets “mix” vintages of observations, in the sense that the most recent observation in a fully revised

¹⁹3- and 6-months T-bill yields are constant-maturity, as reported in the FRED database of the Federal Reserve Bank of St. Louis. This “hybrid” zero-yield dataset that includes both GSW as well as FRED data is widely utilized in yield curve estimation (see e.g., [Gürkaynak and Wright \(2012\)](#) and [Hamilton and Wu \(2012\)](#)).

²⁰In our experiments, we removed all interest rates variables from this dataset, as we wanted to separate the information contained in our macroeconomic variables from that used in our DNS and NSS models. We thank one anonymous referee for making this suggestion.

dataset is a so-called “first release”, while earlier calendar dated observations have possibly been revised and re-released many times. In summary, note that only the use of real-time datasets make it possible to replicate truly real-time modeling and forecasting of U.S. Treasury yields, when using macroeconomic data that are subject to revision.

In the prediction experiments reported in this paper, it is important to note that data release delays may be relevant. As pointed out by [Ghysels et al. \(2018\)](#), for example, macroeconomic data are generally characterized not only by data revisions, but also by delayed releases. In their paper, these authors also extensively discuss the use of real-time data when estimating term structure models. Of note, in our context, is that any data release delays incorporated in our dataset would result in even less macroeconomic data being available in real time. This means that if we find evidence that real-time diffusion indexes are not useful for forecasting yields, then this evidence is not likely to be overturned if we account for data release delays. Still, the potential importance of such delays should not be overlooked when utilizing real-time data in forecasting experiments. For further discussion on the use of real-time data for predicting interest rates, see [Adrian et al. \(2013\)](#), in which paper an interesting regression-based approach to the pricing of interest rates using up to five principal components as yields is discussed.

Plots of the yield dataset, the FRED-MD real-time data, DNS (NSS) risk factors and various diffusion indexes used our analysis are contained in Figures 1-4. Figure 1 plots all (standardized) real-time FRED-MD variables, as well as the 1-year GSW yield. This figure illustrates how the range of values associated with our macroeconomic variables changes over time, relative to the range of values associated with our yield data. Notice, for example, that the range of the macroeconomic data is largest around 2008. In Figure 2, our yield data are plotted, together with three real-time diffusion indexes constructed using un-targeted PCA.²¹ Of note is that diffusion index volatility increases markedly around 2008, and is the lowest, historically, thereafter. Figures 3 and 4 plot the same yield data against DNS and NSS betas, constructed using our 12-dimensional yield cross section. While the first beta changes very little when constructed using either the DNS and NSS models, the other betas are evidently quite different. These differences might account in some part for the relative forecasting performance differences between the two models.

6 Empirical Findings

Our empirical investigation utilizes the models discussed in Section 4, and the methods discussed in Sections 2 and 3. Our objective is to predict U.S. Treasury yields at various maturities. Predictions are made using “small data” models (i.e. solely yield driven models), as well as “big data” models that include real-time diffusion indexes constructed from the real-time FRED-MD dataset discussed in Section 5. Our

²¹Plotted diffusion indexes are those associated with the largest three eigenvalues of the covariance matrix of the data.

small data models include autoregressive, vector autoregressive, DNS and NSS models, and our big data models are specified as pure diffusion index models or as hybrids that combine our small data models with diffusion indexes. In our experiments, diffusion indexes are constructed using targeted (machine learning based) PCA as well as un-targeted PCA, as discussed in Section 3.

A number of clear-cut conclusions emerge upon examination of the results collected in Tables 1—8. Turning first to Tables 2—4, note that these tables present the following results: (i) the three “MSFE-best” (i.e., lowest MSFE model) models for each yield maturity/forecast horizon permutation, in descending order from the first to the third (see Table 2); (ii) MSFEs for the three models listed in Table 2 (see Table 3); and relative MSFEs (where the RW benchmark is the numeraire) for the three models (see Table 4).²² All models used in our experiments are summarized in Table 1, and Tables 2-4 report results for DNS and NSS type models with betas constructed using our 12-dimensional historical yield dataset, unless otherwise noted. Results are presented for three forecast horizons ($h = 1, 3$, and 12 months), for six yield maturities (3- and 6-month, 1-year, 3-years, 5-years, and 10-years), and for four different forecasting periods, including: 2001:1-2005:12 (Subsample 1), 2006:1-2010:12 (Subsample 2), 2011:1-2017:10 (Subsample 3), and 2001:1-2017:10 (Full Sample). Illustrating the layout of the tables, we see in Table 2 that for 3-month T-bill yields, with a 1-month ahead forecasting horizon, the model NSS(4)+FB3(EN06) yields the lowest MSFE, for Subsample 1. This means that the real-time diffusion indexes constructed using the elastic net add marginal predictive content to the NSS model, and that this model is MSFE-best for 1-month ahead predictions of the 3-month Treasury bill yield. Additionally, note that this model has a three-star superscript, indicating rejection of the Diebold-Mariano null hypothesis of equal predictive accuracy, when compared with the RW benchmark, at the 1% level (p -values are calculated by using the block bootstrap method discussed above - see Section 4.2 for further details).

Notice that many of the models that are in the top three MSFE-best in Tables 2—4 include diffusion indexes (i.e., models with FB1, FB2, or FB3, in their names, corresponding to the addition of 1, 2, or 3 real-time diffusion indexes). In many of such models, the diffusion indexes are constructed by using the elastic net method (i.e., models with EN04, EN04 and EN06). In addition, many models are of the DNS and NSS variety. This pattern is quite prevalent across all forecast horizons and bond maturities, for the first two subsamples. This suggests that, for the 2001-2010 period, we have rather strong evidence that DNS and NSS models are useful (since they “beat” the benchmark model), as previously found by many authors, and also that the usefulness of all models, including DNS and NSS models, can often be

²²Results where AR(1) and AR(SIC) benchmarks are the numeraire were also tabulated, but due to the poor performance of these models, relative to the RW, in recent years, tabulated results are not reported (complete results are available upon request). In addition, note that all results reported in these tables have been compiled for each of the models analyzed, rather than only for the three MSFE-best models, as depicted in Tables 2-4. However, the number of tables is overwhelming when including all models, maturities and forecast horizons, and hence these results are omitted. for the sake of brevity. Complete results are available upon request.

improved by utilizing the information contained in real-time diffusion indexes. However, this pattern is less prevalent during the third subsample, from 2011-2017. In this time period, the RW benchmark or the simple AR(1) model often “wins”. This finding indicates a deterioration in predictive gains associated with using diffusion indexes, post Great Recession. Still, a broad conclusion based on Tables 2—4 is that macroeconomic, financial, and other non-yield variables contain marginal predictive content for interest rates. Moreover, this finding is most pronounced when one uses targeted machine learning techniques.

One possible explanation for the deteriorating performance of models with real-time diffusion indexes, post Great Recession, is that the so-called “zero lower bound” discussed extensively in the literature becomes important during this sample period (see e.g. [Christensen and Rudebusch \(2014\)](#), [Krippner \(2015\)](#), [Christensen and Rudebusch \(2016\)](#), and [Wu and Xia \(2016\)](#)). For this reason, it may be of interest to examine whether models might be improved via the use of shadow rates.²³ In particular, “sophisticated models” that include diffusion indexes might perform better post 2010, if shadow rates are utilized. This in turn would buttress our finding that sophisticated models are MSFE-best, prior to 2011. However, and as discussed below, if we use combination models, then combinations that *only* include purely yield driven models outperform individual sophisticated models as well as combinations of sophisticated models *and* combinations of both sophisticated and purely yield driven models, for most sample periods and maturities that we analyze. Moreover, combination models generally yield the lowest MSFEs, across all sub-samples, always beating individual models. Thus, our finding discussed below that unspanned risks are not found using real-time data, is not likely to be impacted by the introduction of shadow rates. This is because yield data are likely not impacted by shadow rate considerations in the earlier sample periods that we examine, and in these earlier sample periods, purely yield driven models are nevertheless MSFE-best, when combined using simple forecast averaging techniques. Still, this does not mean that shadow rates are not important. In consideration of this fact, we leave the investigation into the importance of shadow rates in our context to future research.²⁴

Turning back to our discussion of the results in Tables 2—4, note that inspection of these tables indicates that the DNS and NSS type models are often the MSFE-best models. In Table 2, DNS or NSS type models are MSFE-best in 13 of 15 maturity/horizon permutations for Subsample 1, and are in the top MSFE performers in 34 of 45 maturity/horizon permutations. In Subsample 2, results are similar, where analogous “wins” are 9 of 15 and 28 of 45, respectively. Finally, in Subsample 3, DNS or NSS type models “win” in only 4 of 15 cases and 22 of 45 cases, respectively. These findings are consistent with the findings in the recent literature suggesting that NS type model performance has deteriorated in recent post credit crisis years (see e.g., [Moench \(2008\)](#), [Diebold and Rudebusch \(2013\)](#), and [Altavilla et al. \(2017\)](#)), as discussed above.

²³We thank an anonymous referee for pointing this out to us.

²⁴For discussion of the zero lower bound issue affecting predictions from term structure models, see [Hamilton and Wu \(2012\)](#).

Our forecast combination analysis overturns our key findings from Tables 2-4, however. Note that Table 5 lists the combination models that were examined in our analysis.²⁵ Relative MSFEs for all combinations are given in Table 6 (1-month ahead forecasts), Table 7 (3-month ahead forecasts), and Table 8 (12-month ahead forecasts). Many model combinations significantly outperform the RW benchmark, for all five maturities and all three forecast horizons, in the first two subsamples, and in the full sample. However, *all* MSFE-best combinations involve purely yield driven models, regardless of maturity or forecast horizon. Thus, there appear to be no unspanned risks in our earlier sample periods. Namely real-time diffusion indexes no longer contain marginal predictive content for yields, in the early years of our sample period.

In our third, post Great Recession sample period, only a few combinations “beat” the RW benchmark. Interestingly, though, the particular combinations that do significantly outperform the benchmark in this period are always model combinations that involve purely yield driven specifications, including our DNS, NSS, and linear econometric models. This again indicates that there are no unspanned risks. Moreover, the MSFEs from MSFE-best forecast combination models are always lower than the MSFEs from the MSFE-best *individual* models, over *most* sample periods. Thus, our earlier finding that *individual* models with diffusion indexes are MSFE-best, when compared with *individual* models that are purely yield driven is no longer true, when combinations are considered. This in turn suggests that our previous finding concerning the importance of real-time diffusion indexes should be viewed with caution, as it might simply be an artifact of model misspecification, since these findings disappear when combinations of purely yield driven models are included in our experiments.

Another key result from our combination experiments is that the MSFE-best models are almost always “FS” type forecast combinations (compare the bolded entries in each column for each subsample in Tables 6–8). As noted in Table 5, FS forecast combination models utilize the average of all non-diffusion index type models (i.e., AR(1), AR(SIC), VAR(1), VAR(SIC), DNS(1) - DNS(6), NSS(1) - NSS(6)). Indeed, FS models “win” in 18 of 20 cases, when $h = 1$, across all five bond maturities and all four subsamples (see Table 6). The MSFE-best combinations in the other two cases are NSS-type models. When $h = 3$, the FS combination model “wins” in 19 of 20 cases (see Table 7). When $h = 12$, the FS combination model “wins” in 20 of 20 cases (see Table 8). Furthermore, all of the “best” point MSFEs in these tables are much lower than point MSFEs associated with the best individual models, as mentioned above. Thus, combination dominates under our real-time setup, and the best combinations do not utilize (macroeconomic) diffusion indexes. This result differs from that reported in Swanson and Xiong (2018), where big data matters for yield forecasting, even under forecast combination. Why this discrepancy? Using fully revised data, Swanson and Xiong (2018) find that big data such as that used in this paper “matter”, since variables other than yields enter into their MSFE-best NS type models,

²⁵Various other forecast combinations were also analyzed, but yielded poorer predictive accuracy than those based on the methods reported here, and are not reported, for the sake of brevity.

even when forecast combinations are considered.²⁶ In summary, two conclusions can be drawn from our combination experiments. First, fully revised data may have an important confounding effect upon results obtained when forecasting yields, as our experimental findings based on the use of real-time data differ markedly from those reported in Swanson and Xiong (2018). Second, while real-time diffusion indexes matter when comparing the predictive performance of *individual* models, they do not contain marginal predictive content when comparing combinations, indicating that evidence of unspanned risks based on *individual* models should be viewed with caution in our setup.

As a robustness check, we also constructed Giacomini-Rossi fluctuation tests (see Giacomini and Rossi (2010)). Results for selected model combinations are shown in Figures 7-9. The fluctuation test statistic (implemented using a Giacomini-White (GW) predictive accuracy test in each rolling window) is printed above each plot of GW test statistics in these figures (see Section 4.2 for further details). Consider Figure 7 as an example. In the first two rows of plots in this figure, we see that in the case of $h = 1$ -month ahead forecasts for 3-month T-bill yields, nearly all selected combination models significantly outperform the benchmark model, at least at one point in the forecast period, indicating test rejection, using one-sided critical values, which are 2.936 and 2.676, for 5% and 10% significance levels, respectively. Thus, the fluctuation test indicates a “break” in predictive performance, in the sense that the benchmark univariate model is rejected. Recall, however, that this test is not designed for carrying out inference on each GW statistic in the plots, rather it is a “max” type statistic. Returning to Figure 7, note that the lower panel shows results for 10-year bond yields. Seven out of nine model combinations depicted in these plots significantly outperform the benchmark model, at some point in the forecast sample. Inspection of Figures 8 and 9 indicates similar results. Thus, fluctuation test results are consistent with the findings reported in Tables 6-8.

7 Concluding Remarks

In this paper, we examine the usefulness of real-time macroeconomic diffusion indexes for forecasting the term structure of interest rates, when using dynamic Nelson-Siegel (DNS), dynamic Nelson-Siegel-Svensson (NSS), and various econometric models, constructed both with and without the aid of machine learning methods such as the lasso and elastic net. We find that the marginal predictive content of real-time diffusion indexes is significant for many of the *individual* models that we examine. However, we also find that *individual* model performance worsens, post Great Recession. While this points to the possible importance of the zero lower bound in recent years, for example, our analysis of forecast combinations

²⁶For a summary of the results in Swanson and Xiong (2018), refer to Tables 1 and 7 in their paper for a complete list of the individual and combination models used in their forecasting experiments, respectively; and refer to Tables 8A-8C in their paper for a summary of their predictive accuracy findings based on experiments comparing combination models using fully revised data. Compare these results with those presented in Table 6-8.

adds nuance to this finding. In particular, forecast combinations that include only yield variables result in the lowest MSFEs across various subsamples, including our post Great Recession sample period. Indeed, when combinations are utilized to construct forecasts, models with real-time diffusion indexes are not preferred for any of the sample periods that we analyze. When combination forecasts are included in our analysis, thus, there are no unspanned risks, and all relevant forecasting information is contained in the term structure. Moreover, the “MSFE-best” combinations almost always include DNS and NSS models. These results are consistent with two conclusions. First, fully revised data may have an important confounding effect upon results obtained when forecasting yields, as our experimental findings based on the use of real-time data differ markedly from those reported in [Swanson and Xiong \(2018\)](#). Second, while real-time diffusion indexes matter when comparing the predictive performance of *individual* models, they do not contain marginal predictive content when comparing combinations, indicating that evidence of unspanned risks based on *individual* models is likely spurious in our setup. The latter conclusion is quite strong, however, and further research may shed additional light on the issue. For example, we only utilize two shrinkage operators in our analysis. The use of additional machine learning, data reduction, and shrinkage methods may yield further evidence that we have not yet uncovered. Moreover, while our implementation of targeted and un-targeted PCA yields considerable evidence on the usefulness of such methods for predicting yields, there are many other methods, both linear and nonlinear, that might shed additional light on the usefulness of real-time macroeconomic data for predicting the term structure. Examples include the sparse principal components analysis and independent component analysis methods discussed in [Kim and Swanson \(2018\)](#). Finally, it would be of some interest to extend the important work in this area by authors including [Diebold and Li \(2006\)](#), [Rudebusch and Wu \(2008\)](#), [Exterkate et al. \(2013\)](#), [Christensen and Rudebusch \(2014\)](#), [Krippner \(2015\)](#), [Christensen and Rudebusch \(2016\)](#), [Coroneo et al. \(2016\)](#), [Wu and Xia \(2016\)](#), and [Ghysels and Marcellino \(2018\)](#), among others, to include the use of machine learning methods such as those discussed in this paper.

References

- ADRIAN, T., R. K. CRUMP, AND E. MOENCH (2013): “Pricing the term structure with linear regressions,” *Journal of Financial Economics*, 110–138.
- ALTAVILLA, C., R. GIACOMINI, AND G. RAGUSA (2017): “Anchoring the yield curve using survey expectations,” *Journal of Applied Econometrics*, 32, 1055–1068.
- ANDREOU, E., P. GAGLIARDINI, E. GHYSELS, AND M. RUBIN (2018): “Is industrial production still the dominant factor for the us economy? Evidence from a new class of mixed frequency (group) factor models,” Swiss Finance Institute Research Paper No. 16-11.
- ANG, A. AND M. PIAZZESI (2003): “A no-arbitrage vector autoregression of term structure dynamics with macroeconomic and latent variables,” *Journal of Monetary Economics*, 50, 745–787.
- BAI, J. AND S. NG (2002): “Determining the number of factors in approximate factor models,” *Econometrica*, 70, 191–221.
- (2006): “Evaluating latent and observed factors in macroeconomics and finance,” *Journal of Econometrics*, 131, 507–537.
- (2008): “Forecasting economic time series using targeted predictors,” *Journal of Econometrics*, 146, 304–317.
- (2009): “Boosting diffusion indices,” *Journal of Applied Econometrics*, 24, 607–629.
- BANERJEE, A., M. MARCELLINO, AND I. MASTEN (2008): “Forecasting Macroeconomic Variables Using Diffusion Indexes in Short Samples with Structural Change,” in *Forecasting in the presence of structural breaks and model uncertainty*, Emerald Group Publishing Limited, chap. 4, 149–194.
- BAUER, M. D., G. D. RUDEBUSCH, AND J. C. WU (2012): “Correcting estimation bias in dynamic term structure models,” *Journal of Business & Economic Statistics*, 30, 454–467.
- BERNANKE, B. S. AND J. BOIVIN (2003): “Monetary policy in a data-rich environment,” *Journal of Monetary Economics*, 50, 525–546.
- BIS (2005): “Zero-coupon yield curves,” Tech. rep., Bank for International Settlements, technical Report.
- BOIVIN, J. AND S. NG (2005): “Understanding and comparing factor-based forecasts,” Tech. rep., National Bureau of Economic Research.
- CARRASCO, M. AND B. ROSSI (2016): “In-sample inference and forecasting in misspecified factor models,” *Journal of Business & Economic Statistics*, 34, 313–338.

- CHENG, X. AND B. E. HANSEN (2015): “Forecasting with factor-augmented regression: A frequentist model averaging approach,” *Journal of Econometrics*, 186, 280–293.
- CHRISTENSEN, J. H., F. X. DIEBOLD, AND G. D. RUDEBUSCH (2011): “The affine arbitrage-free class of Nelson–Siegel term structure models,” *Journal of Econometrics*, 164, 4–20.
- CHRISTENSEN, J. H. AND G. D. RUDEBUSCH (2014): “Estimating shadow-rate term structure models with near-zero yields,” *Journal of Financial Econometrics*, 13, 226–259.
- (2016): “Modelling yields at the zero lower bound: Are shadow rates the solution?” in: *Dynamic Factor Models: Advances in Econometrics vol. 35*, ed. Eric Hillebrand and Siem Jan Koopman, 75–125.
- CLARK, T. E. AND M. W. MCCracken (2009): “Improving forecast accuracy by combining recursive and rolling forecasts,” *International Economic Review*, 50, 363–395.
- CORONEO, L., D. GIANNONE, AND M. MODUGNO (2016): “Unspanned macroeconomic factors in the yield curve,” *Journal of Business & Economic Statistics*, 34, 472–485.
- CORONEO, L. AND F. IACONE (2018): “Comparing predictive accuracy in small samples using fixed-smoothing asymptotics,” *Working Paper, University of York*.
- CORRADI, V. AND N. R. SWANSON (2006): “Predictive density evaluation,” *Handbook of Economic Forecasting*, 1, 197–284.
- (2007): “Nonparametric bootstrap procedures for predictive inference based on recursive estimation schemes,” *International Economic Review*, 48, 67–109.
- DE POOTER, M. (2007): “Examining the Nelson-Siegel class of term structure models: In-sample fit versus out-of-sample forecasting performance,” Tech. Rep. 07-043/4, Tinbergen Institute.
- DIEBOLD, F. X. AND C. LI (2006): “Forecasting the term structure of government bond yields,” *Journal of Econometrics*, 130, 337–364.
- DIEBOLD, F. X. AND R. S. MARIANO (1995): “Comparing predictive accuracy,” *Journal of Business & Economic Statistics*, 20, 134–144.
- DIEBOLD, F. X. AND G. D. RUDEBUSCH (1991): “Forecasting output with the composite leading index: A real-time analysis,” *Journal of the American Statistical Association*, 86, 603–610.
- (2013): *Yield curve modeling and forecasting: the dynamic Nelson-Siegel approach*, Princeton University Press.
- DIEBOLD, F. X., G. D. RUDEBUSCH, AND S. B. ARUOBA (2006): “The macroeconomy and the yield curve: a dynamic latent factor approach,” *Journal of Econometrics*, 131, 309–338.

- DUFFEE, G. R. (2011): “Information in (and not in) the term structure,” *The Review of Financial Studies*, 24, 2895–2934.
- (2013): “Forecasting interest rates,” in *Handbook of Economic Forecasting*, Elsevier, vol. 2, 385–426.
- EXTERKATE, P., D. V. DIJK, C. HEIJ, AND P. J. GROENEN (2013): “Forecasting the yield curve in a data-rich environment using the factor-augmented Nelson–Siegel model,” *Journal of Forecasting*, 32, 193–214.
- GHYSELS, E., C. HORAN, AND E. MOENCH (2018): “Forecasting through the rear view mirror: Data revision and bond return predictability,” *The Review of Financial Studies*, 31, 678–714.
- GHYSELS, E. AND M. MARCELLINO (2018): *Applied economic forecasting using time series methods*, Oxford University Press.
- GIACOMINI, R. AND B. ROSSI (2010): “Forecast comparisons in unstable environments,” *Journal of Applied Econometrics*, 25, 595–620.
- GIACOMINI, R. AND H. WHITE (2006): “Tests of conditional predictive ability,” *Econometrica*, 74, 1545–1578.
- GÜRKAYNAK, R. S., B. SACK, AND J. H. WRIGHT (2007): “The US Treasury yield curve: 1961 to the present,” *Journal of Monetary Economics*, 54, 2291–2304.
- GÜRKAYNAK, R. S. AND J. H. WRIGHT (2012): “Macroeconomics and the term structure,” *Journal of Economic Literature*, 50, 331–67.
- HAMILTON, J. D. AND G. PEREZ-QUIROS (1996): “What do the leading indicators lead?” *Journal of Business*, 27–49.
- HAMILTON, J. D. AND J. C. WU (2012): “The effectiveness of alternative monetary policy tools in a zero lower bound environment,” *Journal of Money, Credit and Banking*, 44, 3–46.
- HANSEN, P. R. AND A. TIMMERMAN (2012): “Choice of sample Split in out-of-sample forecast evaluation. European University Institute,” Working Paper.
- HIRANO, K. AND J. H. WRIGHT (2017): “Forecasting with model uncertainty: Representations and risk reduction,” *Econometrica*, 85, 617–643.
- KIM, H. H. AND N. R. SWANSON (2014): “Forecasting financial and macroeconomic variables using data reduction methods: New empirical evidence,” *Journal of Econometrics*, 178, 352–367.

- (2018): “Mining Big Data Using Parsimonious Factor, Machine Learning, Variable Selection and Shrinkage Methods,” *International Journal of Forecasting*, 34, 339–354.
- KRIPPNER, L. (2015): *Zero lower bound term structure modeling: A practitioner’s guide*, Springer.
- KUNSCH, H. R. (1989): “The jackknife and the bootstrap for general stationary observations,” *The Annals of Statistics*, 1217–1241.
- LUDVIGSON, S. C. AND S. NG (2009): “Macro factors in bond risk premia,” *The Review of Financial Studies*, 22, 5027–5067.
- MCCRACKEN, M. W. (2000): “Robust out-of-sample inference,” *Journal of Econometrics*, 99, 195–223.
- MCCRACKEN, M. W. AND S. NG (2016): “FRED-MD: A monthly database for macroeconomic research,” *Journal of Business & Economic Statistics*, 34, 574–589.
- MOENCH, E. (2008): “Forecasting the yield curve in a data-rich environment: A no-arbitrage factor-augmented VAR approach,” *Journal of Econometrics*, 146, 26–43.
- NELSON, C. R. AND A. F. SIEGEL (1987): “Parsimonious modeling of yield curves,” *Journal of Business*, 473–489.
- ROSSI, B. AND A. INOUE (2012): “Out-of-sample forecast tests robust to the choice of window size,” *Journal of Business & Economic Statistics*, 30, 432–453.
- ROSSI, B. AND T. SEKHPOSYAN (2011): “Understanding models’ forecasting performance,” *Journal of Econometrics*, 164, 158–172.
- RUDEBUSCH, G. D. AND T. WU (2008): “A macro-finance model of the term structure, monetary policy and the economy,” *The Economic Journal*, 118, 906–926.
- SCHUMACHER, C. (2007): “Forecasting German GDP using alternative factor models based on large datasets,” *Journal of Forecasting*, 26, 271–302.
- (2010): “Factor forecasting using international targeted predictors: The case of German GDP,” *Economics Letters*, 107, 95–98.
- STARK, T. (2010): “Realistic evaluation of real-time forecasts in the Survey of Professional Forecasters,” Tech. rep., Federal Reserve Bank of Philadelphia Research Division.
- STOCK, J. H. AND M. W. WATSON (2002a): “Forecasting using principal components from a large number of predictors,” *Journal of the American Statistical Association*, 97, 1167–1179.

- (2002b): “Macroeconomic forecasting using diffusion indexes,” *Journal of Business & Economic Statistics*, 20, 147–162.
- (2012): “Generalized shrinkage methods for forecasting using many predictors,” *Journal of Business & Economic Statistics*, 30, 481–493.
- SVENSSON, L. E. (1994): “Estimating and interpreting forward interest rates: Sweden 1992-1994,” Tech. rep., National Bureau of Economic Research.
- SWANSON, N. R. AND W. XIONG (2018): “Big data analytics in economics: What have we learned so far, and where should we go from here?” *Canadian Journal of Economics*, 51, 695–746.
- TIMMERMANN, A. (2006): “Forecast combinations,” *Handbook of Economic Forecasting*, 1, 135–196.
- WU, J. C. AND F. D. XIA (2016): “Measuring the macroeconomic impact of monetary policy at the zero lower bound,” *Journal of Money, Credit and Banking*, 48, 253–291.

Table 1: Models Used in Prediction Experiments

Model	Description
RW	Random walk model
AR(1)	Autoregressive model with one lag
AR(SIC)	Autoregressive model with lag(s) selected by the Schwarz information criterion
AR(1)+FB1	AR(1) model with diffusion index added, principal component analysis based on real-time macroeconomic dataset
AR(1)+FB2	AR(1) model with two diffusion indexes added, principal component analysis based on real-time macroeconomic dataset
AR(1)+FB3	AR(1) model with three diffusion indexes added, principal component analysis based on real-time macroeconomic dataset
VAR(1)	Five-dimensional vector autoregressive model with one lag
VAR(SIC)	Five-dimensional vector autoregressive model with lag(s) selected by the Schwarz information criterion
VAR(1)+FB1	VAR(1) model with one diffusion index added, principal component analysis based on real-time macroeconomic dataset
VAR(1)+FB2	VAR(1) model with two diffusion indexes added, principal component analysis based on real-time macroeconomic dataset
VAR(1)+FB3	VAR(1) model with three diffusion indexes added, principal component analysis based on real-time macroeconomic dataset
DNS(1)	Dynamic Nelson-Siegel (DNS) model with underlying AR(1) factor specifications fitted with twelve-dimensional yields: maturity $\tau = 12, 24, 36, 48, 60, 72, 84, 96, 108, 120$ months, with a static rate of decay parameter $\lambda = 0.0609$
DNS(2)	Dynamic Nelson-Siegel (DNS) model with underlying AR(1) factor specifications fitted with twelve-dimensional yields, with a dynamic rate of decay parameter λ_t (most recent λ_t are selected in generating predictions)
DNS(3)	Dynamic Nelson-Siegel (DNS) model with underlying AR(1) factor specifications fitted with twelve-dimensional yields, with a dynamic rate of decay parameter λ_t (median λ_t are selected in generating predictions)
DNS(4)	Dynamic Nelson-Siegel (DNS) model with underlying VAR(1) factor specifications fitted with twelve-dimensional yields, with a static rate of decay parameter $\lambda = 0.0609$
DNS(5)	Dynamic Nelson-Siegel (DNS) model with underlying VAR(1) factor specifications fitted with twelve-dimensional yields, with a dynamic rate of decay parameter λ_t (most recent λ_t are selected in generating predictions)
DNS(6)	Dynamic Nelson-Siegel (DNS) model with underlying VAR(1) factor specifications fitted with twelve-dimensional yields, with a dynamic rate of decay parameter λ_t (median λ_t are selected in generating predictions)
DNS(1-6)+FBk	DNS(1-6) model with k (1,2, or 3) number of diffusion index(es) added, principal component analysis based on real-time macroeconomic dataset
NSS(1)	Dynamic Nelson-Siegel-Svensson (NSS) model with underlying AR(1) factor specifications fitted with twelve-dimensional yields, with a static rate of decay parameter $\lambda = [0.0609, 0.2985]$
NSS(2)	Dynamic Nelson-Siegel-Svensson (NSS) model with underlying AR(1) factor specifications fitted with twelve-dimensional yields, with a dynamic rate of decay parameter λ_t (most recent λ_t are selected in generating predictions)
NSS(3)	Dynamic Nelson-Siegel-Svensson (NSS) model with underlying AR(1) factor specifications fitted with twelve-dimensional yields, with a dynamic rate of decay parameter λ_t (median λ_t are selected in generating predictions)
NSS(4)	Dynamic Nelson-Siegel-Svensson (NSS) model with underlying VAR(1) factor specifications fitted with twelve-dimensional yields, with a static rate of decay parameter $\lambda = 0.0609$
NSS(5)	Dynamic Nelson-Siegel-Svensson (NSS) model with underlying VAR(1) factor specifications fitted with twelve-dimensional yields, with a dynamic rate of decay parameter λ_t (most recent λ_t are selected in generating predictions)
NSS(6)	Dynamic Nelson-Siegel-Svensson (NSS) model with underlying VAR(1) factor specifications fitted with twelve-dimensional yields, with a dynamic rate of decay parameter λ_t (median λ_t are selected in generating predictions)
NSS(1-6)+FBk	NSS(1-6) model with k (1,2, or 3) number of diffusion index(es) added, principal component analysis based on real-time macroeconomic dataset

Notes: For models with real-time diffusion indexes, 5 additional variants of each model are analyzed. In these models, a subset of variables from the real-time dataset is used when constructing diffusion indexes, with the subset selected using either the elastic net (EN) or the LASSO. For complete details, refer to Section 4.

Table 2: Top 3 MSFE-Best Models

		3mo	1yr	3yr	5yr	10yr
2001:01-2005:12 1st subsample		NSS(4)+FB3(EN06)***	AR(1)+FB2(EN04)**	DNS(1)+FB1(EN06)	NSS(5)	NSS(5)+FB1(EN04)
	1-month	NSS(4)+FB1(EN06)***	AR(1)+FB3(EN04)**	DNS(1)+FB2(EN06)	RW	NSS(5)+FB1(EN02)
		VAR(1)+FB3(EN06)***	AR(1)+FB2*	DNS(1)+FB1(EN04)	NSS(5)+FB1(EN02)	NSS(5)+FB1(EN06)
		VAR(1)+FB1***	DNS(5)+FB2(EN06)	NSS(5)+FB2(EN04)	NSS(5)	NSS(2)+FB1(EN06)*
	3-month	VAR(1)+FB3(LASSO)***	VAR(1)+FB3(EN06)	NSS(5)+FB2(EN06)	NSS(5)+FB1	NSS(3)+FB1(EN06)*
		VAR(1)+FB2***	DNS(5)+FB3(EN08)	RW	RW	DNS(1)+FB1(EN06)*
		DNS(3)+FB1**	DNS(3)+FB1*	DNS(6)+FB1***	NSS(6)**	NSS(6)**
	12-month	DNS(2)+FB1**	DNS(2)+FB1	NSS(2)+FB1	NSS(6)+FB1**	NSS(6)+FB1**
		NSS(3)+FB1	DNS(6)+FB1***	DNS(6)+FB2***	NSS(6)+FB2*	NSS(6)+FB2**
		NSS(3)	DNS(1)	VAR(1)+FB3(EN02)	VAR(1)+FB3(EN08)*	VAR(1)+FB3(LASSO)
2006:01-2010:12 2nd subsample	1-month	DNS(5)+FB3(EN02)	DNS(5)	VAR(1)	VAR(1)+FB3(EN02)	NSS(1)
		DNS(6)+FB3(EN02)	NSS(1)	VAR(SIC)	VAR(1)	DNS(1)
		NSS(6)	DNS(5)+FB1	RW	NSS(3)	NSS(1)**
	3-month	NSS(3)	DNS(4)+FB1	AR(1)	VAR(1)	DNS(1)**
		NSS(6)+FB1	DNS(5)+FB2	NSS(1)	VAR(SIC)	DNS(4)
		VAR(1)+FB1*	NSS(2)+FB1	NSS(1)+FB1	AR(1)+FB1	DNS(4)
	12-month	VAR(1)+FB2	NSS(3)+FB1	DNS(1)+FB1	NSS(3)+FB1	DNS(4)+FB2
		VAR(1)+FB3	VAR(1)+FB1	NSS(3)+FB1	RW	DNS(4)+FB1
		RW	RW	RW	RW	RW
	1-month	AR(1)+FB2(EN08)	AR(1)	AR(1)+FB2(EN02)	AR(1)	NSS(1)+FB2(EN06)
2011:01-2017:10 3rd subsample		AR(SIC)	AR(1)+FB1	AR(1)	AR(SIC)	DNS(1)+FB2(EN06)
		RW	RW	RW	RW	DNS(1)+FB2(EN02)
	3-month	DNS(5)	DNS(5)	AR(1)	AR(1)	RW
		NSS(4)	DNS(5)+FB1	DNS(4)	AR(SIC)	DNS(1)+FB2(EN04)
		DNS(5)**	NSS(4)	RW	RW	NSS(6)+FB2(EN02)***
	12-month	DNS(6)**	NSS(4)+FB2	DNS(4)	DNS(4)	NSS(6)+FB1(EN06)***
		DNS(5)+FB1*	NSS(4)+FB1	DNS(4)+FB1	DNS(4)+FB2	NSS(6)+FB1(EN04)***
		VAR(1)+FB3(EN02)	AR(SIC)	RW	RW	RW
	1-month	VAR(1)+FB3(EN06)	DNS(5)+FB1	AR(1)	AR(1)	DNS(1)
		NSS(4)+FB3(EN04)	AR(1)+FB1	AR(1)+FB1	AR(1)+FB1	NSS(1)
2001:01-2017:10 full sample		DNS(5)+FB1**	DNS(5)+FB1*	RW	RW	RW
	3-month	DNS(6)+FB1**	DNS(5)+FB2*	DNS(4)+FB1	NSS(5)	DNS(1)
		DNS(5)+FB2**	DNS(5)+FB3	DNS(4)+FB2	NSS(5)+FB1	DNS(4)
		DNS(6)***	DNS(6)***	RW	RW	NSS(6)*
	12-month	DNS(6)+FB1***	DNS(6)+FB1**	NSS(5)	NSS(5)	NSS(6)+FB1
		DNS(6)+FB2***	DNS(6)+FB2***	DNS(6)	NSS(5)+FB3	NSS(6)+FB3

Notes: Entries in this table are the top three performing forecast models (based on MSFE), in descending order (from 1st to 3rd place), for various subsamples, horizons, and yield maturities. For a description of the model mnemonics, refer to Table 1. All DNS and NS models reported on are estimated using a historical yield dataset. Entries with LASSO, EN02, EN04, EN06, and EN08 denote cases where the *targeted* variant of our real-time dataset is used when constructing real-time macroeconomic diffusion indexes. Entries superscripted with *, **, or *** denote rejection of the DM predictive accuracy null hypothesis, in favor of the alternative that the listed model has significantly lower MSFE than the random walk benchmark, at a 10%, 5%, or 1% significance level, respectively. See Sections 4 and 6 for complete details.

Table 3: Point MSFEs of Top 3 MSFE-Best Models

		3mo	1yr	3yr	5yr	10yr
2001:01-2005:12	1-month	0.028***	0.057**	0.104	0.11	0.084
		0.028***	0.057**	0.108	0.111	0.085
		0.028***	0.057*	0.109	0.113	0.086
	3-month	0.158***	0.301	0.332	0.283	0.152*
		0.161***	0.306	0.338	0.289	0.158*
		0.164***	0.31	0.342	0.297	0.172*
	12-month	2.067**	2.338*	1.515***	0.918**	0.381**
		2.126**	2.559	1.52	0.931**	0.391**
		2.563	2.648***	1.521***	0.944*	0.393**
2006:01-2010:12	1-month	0.082	0.052	0.075	0.072*	0.089
		0.083	0.053	0.079	0.076	0.092
		0.083	0.055	0.079	0.078	0.093
	3-month	0.316	0.284	0.353	0.288	0.189**
		0.316	0.286	0.367	0.294	0.189**
		0.332	0.288	0.368	0.294	0.202
	12-month	1.899*	1.58	1.195	0.832	0.363
		2.169	1.699	1.206	0.858	0.369
		2.293	1.76	1.22	0.869	0.371
2011:01-2017:10	1-month	0.003	0.005	0.019	0.035	0.047
		0.005	0.006	0.02	0.036	0.05
		0.005	0.007	0.02	0.037	0.05
	3-month	0.01	0.011	0.045	0.101	0.161
		0.018	0.015	0.051	0.109	0.161
		0.018	0.015	0.052	0.111	0.165
	12-month	0.043**	0.064	0.134	0.316	0.325***
		0.044**	0.064	0.14	0.347	0.334***
		0.044*	0.064	0.141	0.35	0.335***
2001:01-2017:10	1-month	0.039	0.042	0.065	0.073	0.075
		0.04	0.042	0.069	0.076	0.075
		0.04	0.042	0.07	0.079	0.076
	3-month	0.182**	0.19*	0.225	0.217	0.189
		0.183**	0.193*	0.239	0.241	0.192
		0.183**	0.195	0.24	0.244	0.194
	12-month	1.723***	1.462***	1.012	0.727	0.428*
		1.731***	1.469**	1.025	0.739	0.439
		1.736***	1.473***	1.038	0.774	0.441

Notes: See notes to Table 2. Entries are point MSFEs.

Table 4: Relative MSFEs of Top 3 MSFE-Best Models

		3mo	1yr	3yr	5yr	10yr
2001:01-2005:12 1st subsample	1-month	0.484***	0.836**	0.956	0.987	0.943
		0.495***	0.836**	0.993	1	0.96
		0.496***	0.839*	0.995	1.01	0.966
	3-month	0.442***	0.85	0.97	0.953	0.754*
		0.449***	0.864	0.987	0.972	0.788*
		0.458***	0.874	1	1	0.853*
	12-month	0.619**	0.726*	0.824***	0.8**	0.732**
		0.637**	0.795	0.826	0.812**	0.751**
		0.767	0.822***	0.827***	0.823*	0.756**
2006:01-2010:12 2nd subsample	1-month	0.883	0.768	0.888	0.831*	0.889
		0.887	0.786	0.929	0.866	0.922
		0.894	0.808	0.929	0.89	0.935
	3-month	0.874	0.821	1	0.972	0.874**
		0.874	0.828	1.041	0.992	0.876**
		0.918	0.833	1.044	0.992	0.935
	12-month	0.625*	0.687	0.862	0.958	0.955
		0.714	0.739	0.87	0.987	0.972
		0.755	0.766	0.88	1	0.976
2011:01-2017:10 3rd subsample	1-month	1	1	1	1	1
		1.452	1.413	1.058	1.045	1.063
		1.486	1.481	1.061	1.076	1.067
	3-month	1	1	1	1	0.999
		1.748	1.309	1.124	1.089	1
		1.772	1.321	1.141	1.108	1.021
	12-month	0.676**	0.944	1	1	0.546***
		0.688**	0.947	1.044	1.098	0.56***
		0.691*	0.953	1.059	1.108	0.562***
2001:01-2017:10 full sample	1-month	0.845	0.984	1	1	1
		0.863	0.984	1.049	1.046	1.007
		0.88	0.985	1.069	1.083	1.009
	3-month	0.835**	0.895*	1	1	1
		0.841**	0.907*	1.063	1.111	1.015
		0.842**	0.919	1.068	1.126	1.023
	12-month	0.897***	0.877***	1	1	0.841*
		0.901***	0.882**	1.012	1.016	0.862
		0.904***	0.884***	1.026	1.064	0.867

Notes: See notes to Table 2. Entries are relative MSFEs, in which the numeraire is the random walk benchmark model.

Table 5: Forecast Combination Models

Model	Description
All	Average of all two hundred and sixty nine forecast models
Econometrics	Average of all forty one RW, AR and VAR type models
AR	Average of all twenty AR type models
VAR	Average of all twenty VAR type models
DNS	Average of all one hundred and fourteen DNS type models
NSS	Average of all one hundred and fourteen NSS type models
NS(ar)	Average of one hundred and fourteen DNS and NSS type models with underlying AR(1) factor specifications (those contain NNS(1-3) or DNS(1-3))
NS(var)	Average of one hundred and fourteen DNS and NSS type models with underlying VAR(1) factor specifications (those contain NNS(4-6) or DNS(4-6))
NS(ols)	Average of seventy six DNS and NSS type models with fixed decay parameter(s), estimated with OLS (those contain NNS(1,4) or DNS(1,4))
NS(nls)	Average of one hundred and fifty two DNS and NSS type models with dynamic decay parameter(s), estimated with NLS (those contain NNS(2,3,5,6) or DNS(2,3,5,6))
FB	Average of two hundred and fifty two models that contain macro diffusion index(es), principal component analysis based on all macroeconomic variables, this is, those models with “FB” in Table 1
FB(pure)	Average of thirty six models that contain diffusion indexes only, namely all “AR(1)+FB” and “VAR(1)+FB” models
FB(ols)	Average of one hundred and eight models that contain macro diffusion index(es), estimated with OLS (FB(pure) models plus NNS(1,4)+FBk and DNS(1,4)+FBk models)
FB(en)	Average of one hundred and sixty eight models that contain macro diffusion index(es), principal component analysis based on targeted macroeconomic variables pre-selected by the elastic net (all NNS(1-6)+FBk and DNS(1-6)+FBk models where the FB part is obtained by the elastic net)
FB(lasso)	Average of forty two models that contain macro diffusion index(es), principal component analysis based on targeted macroeconomic variables pre-selected by lasso (all NNS(1-6)+FBk and DNS(1-6)+FBk models where the FB part is obtained by lasso)
FS	Average of all non-FB type models, that is, those models without “FB” in Table 1

Notes: Entries in this table describe the forecast combination models utilized in our forecast combination experiments. Note that the historical yield dataset is used in the construction of the betas utilized in all DNS and NSS models. For further details, refer to Sections 4 and 6.

Table 6: 1-month-ahead Relative MSFEs of Forecast Combination Models

		3mo	1yr	3yr	5yr	10yr
2001:01-2005:12 1st subsample	All	0.33***	0.483***	0.441***	0.46***	0.406***
	Econometrics	0.586***	0.867**	1.055	1.081	1.077
	AR	0.701***	0.872**	1.046	1.07	1.017
	VAR	0.523***	0.901	1.105	1.141	1.188
	DNS	0.233***	0.376***	0.378***	0.429***	0.334***
	NSS	0.501***	0.553**	0.403***	0.391***	0.345***
	NS(ar)	0.452***	0.573**	0.398***	0.402***	0.329***
	NS(var)	0.234***	0.366***	0.365***	0.391***	0.339***
	NS(ols)	0.223***	0.399***	0.355***	0.391***	0.373***
	NS(nls)	0.368***	0.492***	0.405***	0.412***	0.339***
	FB	0.414***	0.597***	0.543***	0.567***	0.499***
	FB(pure)	0.578***	0.865**	1.058	1.087	1.081
	FB(ols)	0.369***	0.602***	0.632***	0.675***	0.641***
	FB(en)	0.84**	1.156	1.049	1.108	0.989
	FB(lasso)	0.821**	1.256	1.174	1.146	0.954
	FS	0.284***	0.35***	0.219***	0.235***	0.177***
2006:01-2010:12 2nd subsample	All	0.392*	0.523	0.489**	0.449***	0.411**
	Econometrics	0.946	0.944	0.983	0.961	1.003
	AR	1.038	1.047	1.057	1.066	1.049
	VAR	0.97	0.925	0.976	0.917	0.982
	DNS	0.346**	0.448*	0.422***	0.455***	0.411**
	NSS	0.369*	0.561	0.495**	0.418***	0.386**
	NS(ar)	0.357*	0.487	0.431***	0.402***	0.342**
	NS(var)	0.313**	0.508	0.455**	0.387***	0.347**
	NS(ols)	0.355*	0.522	0.411***	0.416***	0.382**
	NS(nls)	0.317*	0.482	0.444***	0.394***	0.34**
	FB	0.474*	0.625	0.593**	0.549***	0.502**
	FB(pure)	0.985	1.007	1.008	0.973	1.012
	FB(ols)	0.594*	0.725	0.665**	0.658***	0.626**
	FB(en)	0.893	1.167	1.123	1.069	0.98
	FB(lasso)	0.909	1.174	1.182	1.046	0.943
	FS	0.172**	0.164**	0.255***	0.252***	0.154**
2011:01-2017:10 3rd subsample	All	1.976	1.278	0.706**	0.672**	0.509***
	Econometrics	1.693	1.915	1.343	1.26	1.292
	AR	1.504	1.5	1.138	1.113	1.201
	VAR	3.014	3.548	1.92	1.602	1.48
	DNS	2.872	0.969	0.709*	0.74	0.492***
	NSS	1.899	1.832	0.738**	0.613**	0.456***
	NS(ar)	3.025	1.957	0.772*	0.664**	0.424***
	NS(var)	1.68	0.927	0.637**	0.64**	0.482***
	NS(ols)	2.104	0.99	0.515***	0.642**	0.394***
	NS(nls)	2.278	1.744	0.894	0.806	0.58***
	FB	2.565	1.62	0.886	0.847	0.632***
	FB(pure)	1.736	1.984	1.368	1.28	1.316
	FB(ols)	2.175	1.31	0.857	0.94	0.7***
	FB(en)	5.558	3.318	1.76	1.704	1.239
	FB(lasso)	5.542	3.436	1.885	1.753	1.252
	FS	2.377	1.104	0.492***	0.509***	0.3***
2001:01-2017:10 full sample	All	0.416**	0.536**	0.491***	0.497***	0.434***
	Econometrics	0.835*	0.949	1.062	1.073	1.102
	AR	0.928	0.983	1.062	1.077	1.076
	VAR	0.866*	1.028	1.153	1.151	1.18
	DNS	0.379**	0.436***	0.435***	0.498***	0.404***
	NSS	0.463**	0.612**	0.478***	0.444***	0.389***
	NS(ar)	0.471**	0.593**	0.455***	0.453***	0.358***
	NS(var)	0.324**	0.458***	0.432***	0.437***	0.378***
	NS(ols)	0.358**	0.484**	0.395***	0.448***	0.382***
	NS(nls)	0.394**	0.542**	0.478***	0.482***	0.4***
	FB	0.514**	0.655**	0.603***	0.615***	0.534***
	FB(pure)	0.857	0.981	1.076	1.084	1.113
	FB(ols)	0.558**	0.691**	0.671***	0.72***	0.65***
	FB(en)	1.012	1.256	1.162	1.209	1.048
	FB(lasso)	1.013	1.312	1.262	1.228	1.024
	FS	0.279***	0.294***	0.266***	0.294***	0.199***

Notes: Entries superscripted with *, **, or *** denote rejection of the DM predictive accuracy null hypothesis, in favor of the alternative that the listed model has significantly lower MSFE than the random walk benchmark, at a 10%, 5%, or 1% significance level, respectively. See Sections 4 and 6 for further details.

Table 7: 3-month-ahead Relative MSFEs of Forecast Combination Models

		3mo	1yr	3yr	5yr	10yr
2001:01-2005:12 1st subsample	All	0.295***	0.454***	0.5***	0.554***	0.544***
	Econometrics	0.542***	0.909	1.292	1.405	1.48
	AR	0.719***	1.023	1.293	1.339	1.305
	VAR	0.434***	0.889	1.391	1.571	1.772
	DNS	0.22***	0.346***	0.429***	0.504***	0.472***
	NSS	0.328***	0.458***	0.41***	0.428***	0.406***
	NS(ar)	0.4***	0.528***	0.46***	0.45***	0.37***
	NS(var)	0.2***	0.323***	0.407***	0.498***	0.526***
	NS(ols)	0.24***	0.376***	0.437***	0.522***	0.476***
	NS(nls)	0.281***	0.41***	0.406***	0.436***	0.422***
	FB	0.368***	0.561***	0.609***	0.673***	0.657***
	FB(pure)	0.536***	0.916	1.312	1.437	1.523
	FB(ols)	0.375***	0.607***	0.768**	0.885	0.857
	FB(en)	0.755***	1.056	1.145	1.264	1.19
	FB(lasso)	0.796**	1.42	1.293	1.309	1.632
	FS	0.238***	0.259***	0.18***	0.18***	0.156***
2006:01-2010:12 2nd subsample	All	0.461*	0.449	0.57**	0.534***	0.487***
	Econometrics	1.054	1	1.212	1.2	1.162
	AR	1.105	1.032	1.268	1.306	1.294
	VAR	1.171	1.064	1.233	1.172	1.091
	DNS	0.402*	0.387*	0.511**	0.491***	0.445***
	NSS	0.387*	0.399*	0.473**	0.418***	0.378***
	NS(ar)	0.383*	0.357*	0.472**	0.425***	0.383***
	NS(var)	0.431*	0.448	0.545*	0.504***	0.434**
	NS(ols)	0.396*	0.398*	0.51**	0.497***	0.412***
	NS(nls)	0.386*	0.381*	0.478**	0.428***	0.401***
	FB	0.555	0.539	0.691*	0.65**	0.59**
	FB(pure)	1.12	1.052	1.261	1.242	1.199
	FB(ols)	0.667	0.652	0.825	0.81*	0.71**
	FB(en)	1.015	1.046	1.275	1.205	1.11
	FB(lasso)	1.247	0.94	1.358	1.363	1.121
	FS	0.192**	0.163**	0.208***	0.212***	0.151***
2011:01-2017:10 3rd subsample	All	2.684	2.904	1.197	0.69*	0.493***
	Econometrics	2.671	4.189	2.146	1.495	1.351
	AR	2.301	2.513	1.565	1.303	1.325
	VAR	5.516	9.695	3.787	2.07	1.478
	DNS	3.266	2.284	1.141	0.682*	0.439***
	NSS	2.71	3.524	1.082	0.536**	0.361***
	NS(ar)	3.638	2.737	0.941	0.539***	0.363***
	NS(var)	2.612	3.548	1.494	0.74	0.447***
	NS(ols)	3.21	2.186	1.068	0.725*	0.407***
	NS(nls)	2.791	3.334	1.19	0.601**	0.406***
	FB	3.351	3.541	1.452	0.846	0.601***
	FB(pure)	2.784	4.518	2.292	1.549	1.379
	FB(ols)	3.303	3.05	1.571	1.09	0.755**
	FB(en)	6.8	6.486	2.617	1.603	1.15
	FB(lasso)	6.428	7.73	3.071	1.711	1.145
	FS	1.661	1.173	0.372***	0.266***	0.169***
2001:01-2017:10 full sample	All	0.422***	0.505**	0.59***	0.571***	0.507***
	Econometrics	0.835	1.025	1.325	1.339	1.328
	AR	0.939	1.06	1.304	1.319	1.308
	VAR	0.893	1.164	1.514	1.503	1.44
	DNS	0.367***	0.408***	0.526***	0.532***	0.452***
	NSS	0.402***	0.496**	0.495***	0.444***	0.381***
	NS(ar)	0.452***	0.493**	0.505***	0.457***	0.372***
	NS(var)	0.359***	0.454**	0.56***	0.546***	0.468***
	NS(ols)	0.373***	0.426***	0.523***	0.55***	0.43***
	NS(nls)	0.38***	0.459**	0.504***	0.464***	0.41***
	FB	0.517**	0.615**	0.717**	0.696***	0.615***
	FB(pure)	0.866	1.06	1.369	1.379	1.363
	FB(ols)	0.574**	0.681**	0.86*	0.893	0.772***
	FB(en)	0.997	1.169	1.327	1.304	1.149
	FB(lasso)	1.125	1.325	1.469	1.406	1.29
	FS	0.242***	0.233***	0.209***	0.209***	0.159***

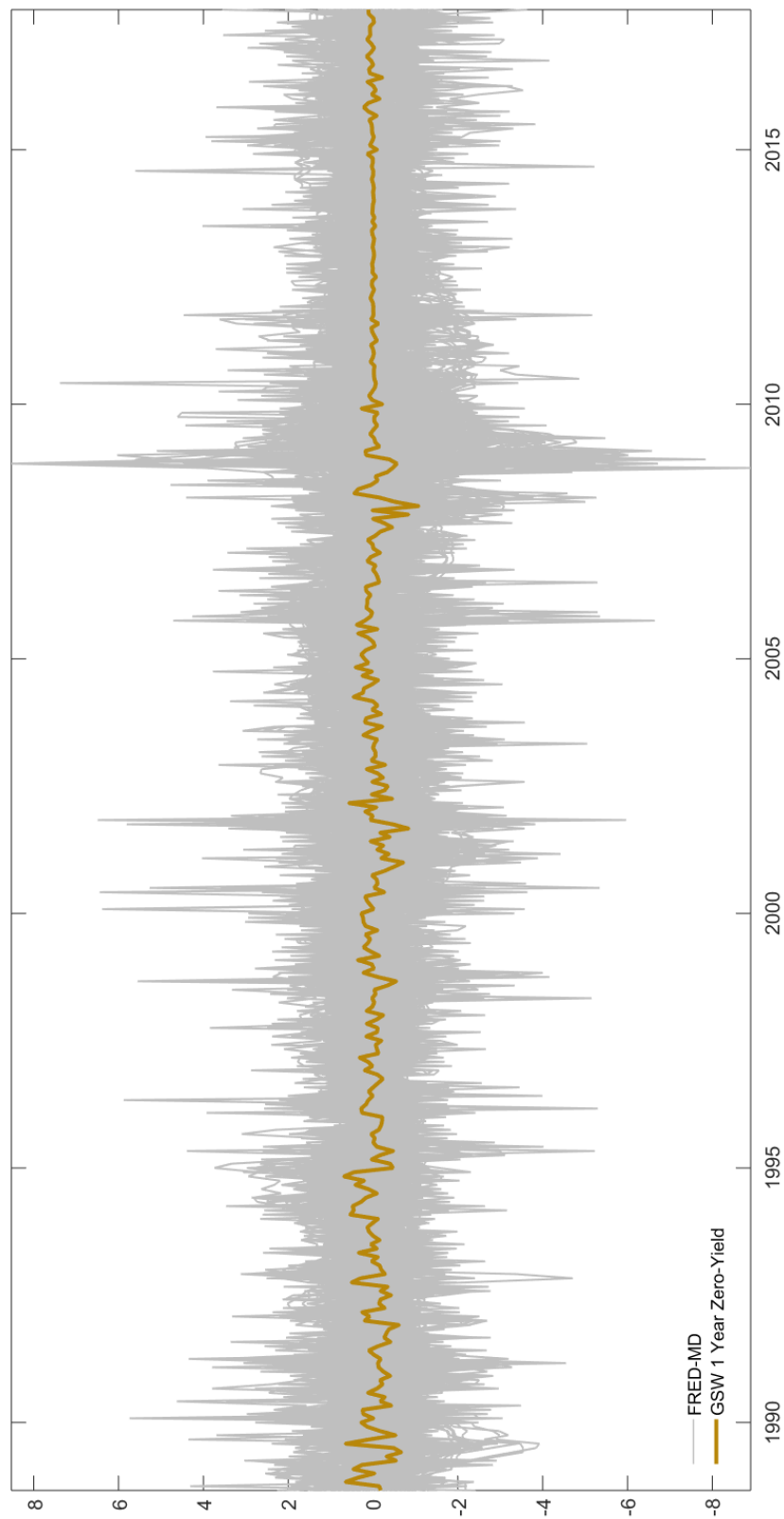
Notes: See notes to Table 6. Entries are MSFEs, relative to the benchmark random walk model, based on 3-month-ahead forecasts of monthly U.S. Treasury bond yields of various maturities.

Table 8: 12-month-ahead Relative MSFEs of Forecast Combination Models

		3mo	1yr	3yr	5yr	10yr
2001:01-2005:12 1st subsample	All	0.716	0.892	1.113	1.393	1.784
	Econometrics	1.696	2.436	3.576	4.266	4.939
	AR	1.488	1.944	2.766	3.608	4.275
	VAR	2.107	3.286	4.882	5.404	6.134
	DNS	0.585	0.702	0.93	1.166	1.462
	NSS	0.598	0.712	0.734*	0.944	1.353
	NS(ar)	0.549*	0.567*	0.441***	0.572***	0.982
	NS(var)	0.681	0.945	1.477	1.772	2.045
	NS(ols)	0.702	0.89	1.153	1.444	1.739
	NS(nls)	0.544*	0.628	0.69*	0.882	1.254
	FB	0.848	1.047	1.291	1.611	2.057
	FB(pure)	1.738	2.515	3.717	4.492	5.285
	FB(ols)	1.117	1.489	2.041	2.512	2.948
	FB(en)	1.501	1.823	2.191	2.71	3.303
	FB(lasso)	1.488	1.779	2.166	2.914	4.059
	FS	0.132***	0.111***	0.072***	0.067***	0.089***
2006:01-2010:12 2nd subsample	All	0.498**	0.619	0.605**	0.767*	1.106
	Econometrics	0.903	1.015	1.337	1.737	2.515
	AR	1.181	1.219	1.319	1.674	2.542
	VAR	0.836	1.131	1.568	1.981	2.815
	DNS	0.411**	0.553*	0.499**	0.687**	0.99
	NSS	0.484**	0.602*	0.559**	0.654**	0.914
	NS(ar)	0.471*	0.594	0.498**	0.595**	0.82
	NS(var)	0.434**	0.567*	0.555**	0.735*	1.106
	NS(ols)	0.42**	0.618	0.611**	0.805	1.145
	NS(nls)	0.458*	0.551*	0.478**	0.587**	0.841
	FB	0.593*	0.736	0.729*	0.929	1.31
	FB(pure)	1.032	1.135	1.433	1.835	2.766
	FB(ols)	0.655	0.851	0.947	1.243	1.76
	FB(en)	1.08	1.263	1.357	1.691	2.302
	FB(lasso)	1.072	1.491	1.288	1.733	2.056
	FS	0.2***	0.197***	0.215***	0.229***	0.18***
2011:01-2017:10 3rd subsample	All	5.749	6.018	3.396	1.428	0.703
	Econometrics	9.234	10.301	6.128	2.733	1.563
	AR	6.715	6.682	4.835	2.579	1.544
	VAR	18.457	20.014	9.288	3.35	1.703
	DNS	6.454	5.419	3.229	1.408	0.698*
	NSS	4.767	5.938	2.971	1.143	0.502**
	NS(ar)	2.915	2.975	1.453	0.844	0.547**
	NS(var)	10.567	11.113	5.846	2.027	0.682*
	NS(ols)	6.401	5.482	3.34	1.49	0.626*
	NS(nls)	5.039	5.644	2.968	1.178	0.593**
	FB	6.758	7.036	3.997	1.7	0.844
	FB(pure)	9.984	11.39	6.939	3.025	1.648
	FB(ols)	8.123	7.717	4.809	2.166	1.022
	FB(en)	11.833	12.255	6.975	2.89	1.528
	FB(lasso)	12.483	11.904	6.317	3.531	1.579
	FS	1.081	0.973	0.565***	0.303***	0.173***
2001:01-2017:10 full sample	All	0.681*	0.865	1.029	1.177	1.12
	Econometrics	1.425	1.983	2.802	3.098	2.798
	AR	1.414	1.725	2.288	2.74	2.594
	VAR	1.731	2.679	3.77	3.827	3.294
	DNS	0.583**	0.719	0.878	1.039	0.994
	NSS	0.601**	0.753	0.783*	0.876	0.851
	NS(ar)	0.544**	0.618**	0.518***	0.628***	0.74**
	NS(var)	0.698	0.958	1.336	1.449	1.189
	NS(ols)	0.647*	0.854	1.05	1.225	1.078
	NS(nls)	0.564**	0.679*	0.725**	0.829*	0.848
	FB	0.808	1.019	1.207	1.385	1.315
	FB(pure)	1.517	2.096	2.961	3.29	2.999
	FB(ols)	0.994	1.33	1.744	2	1.77
	FB(en)	1.443	1.765	2.108	2.38	2.238
	FB(lasso)	1.441	1.828	2.031	2.604	2.437
	FS	0.177***	0.161***	0.157***	0.166***	0.149***

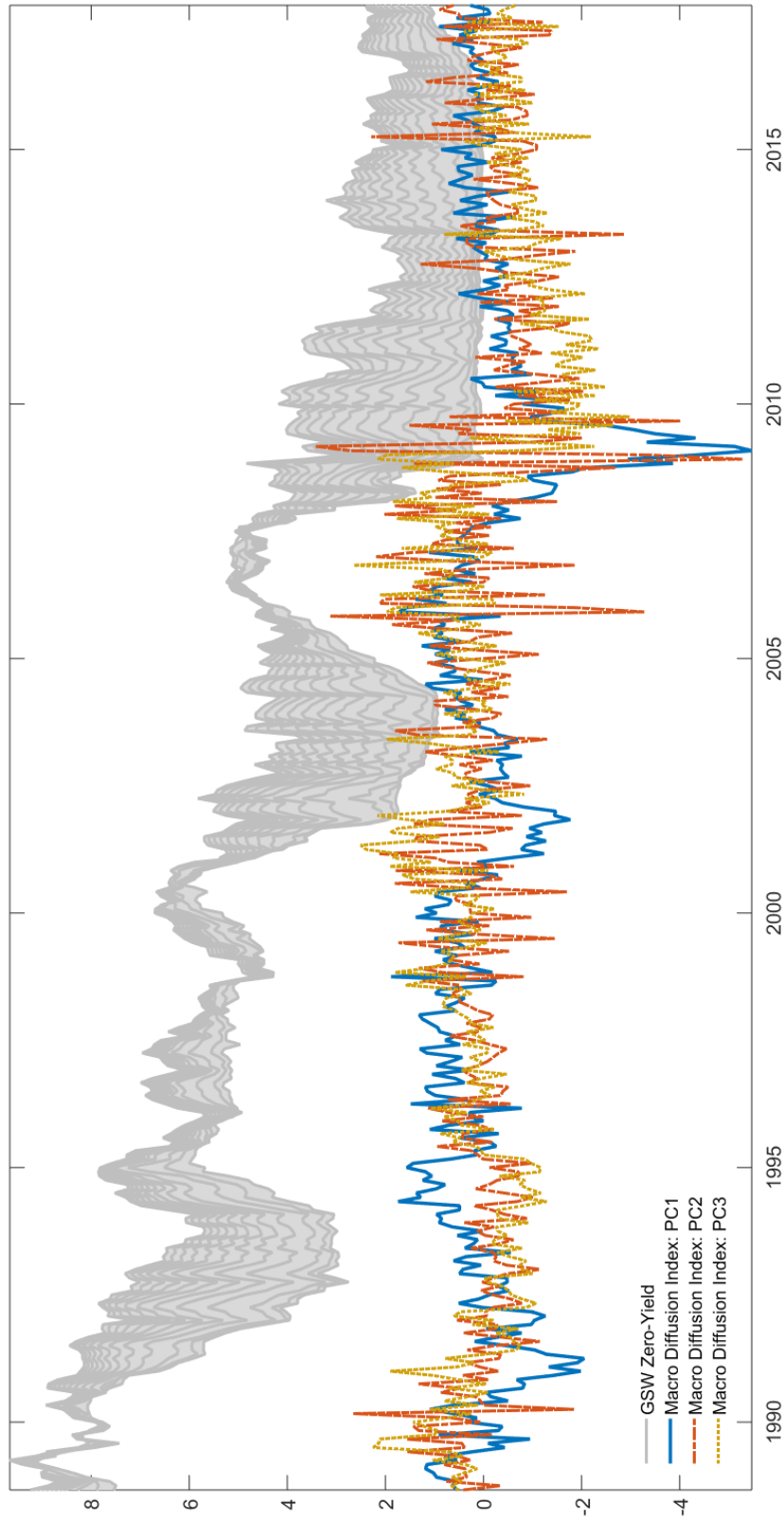
Notes: See notes to Table 6. Entries are MSFEs, relative to the benchmark random walk model, based on 12-month-ahead forecasts of monthly U.S. Treasury bond yields of various maturities.

Figure 1: FRED MD Dataset for Sample Period 1988:8 - 2017:10*



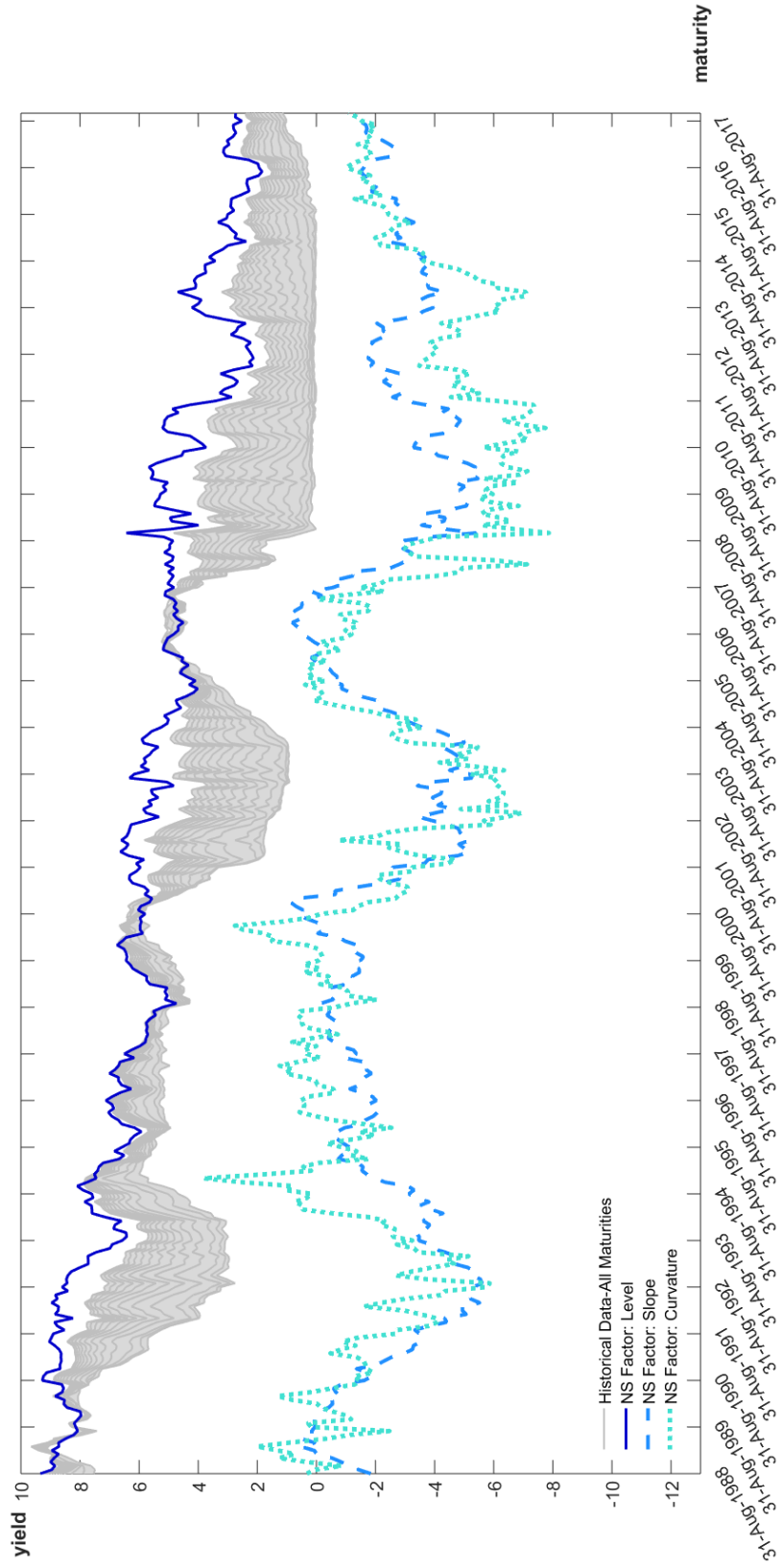
Notes: This figure contains plots of the entire set of FRED-MD variables, as well as a plot of the GSW 1-year yield (see Section 5 for details). All non-yield variables are transformed to ensure stationarity, and all variables are standardized to zero mean and unit variance.

Figure 2: Yields and Macroeconomic Diffusion Indexes for the Sample Period 1988:8 - 2017:10*



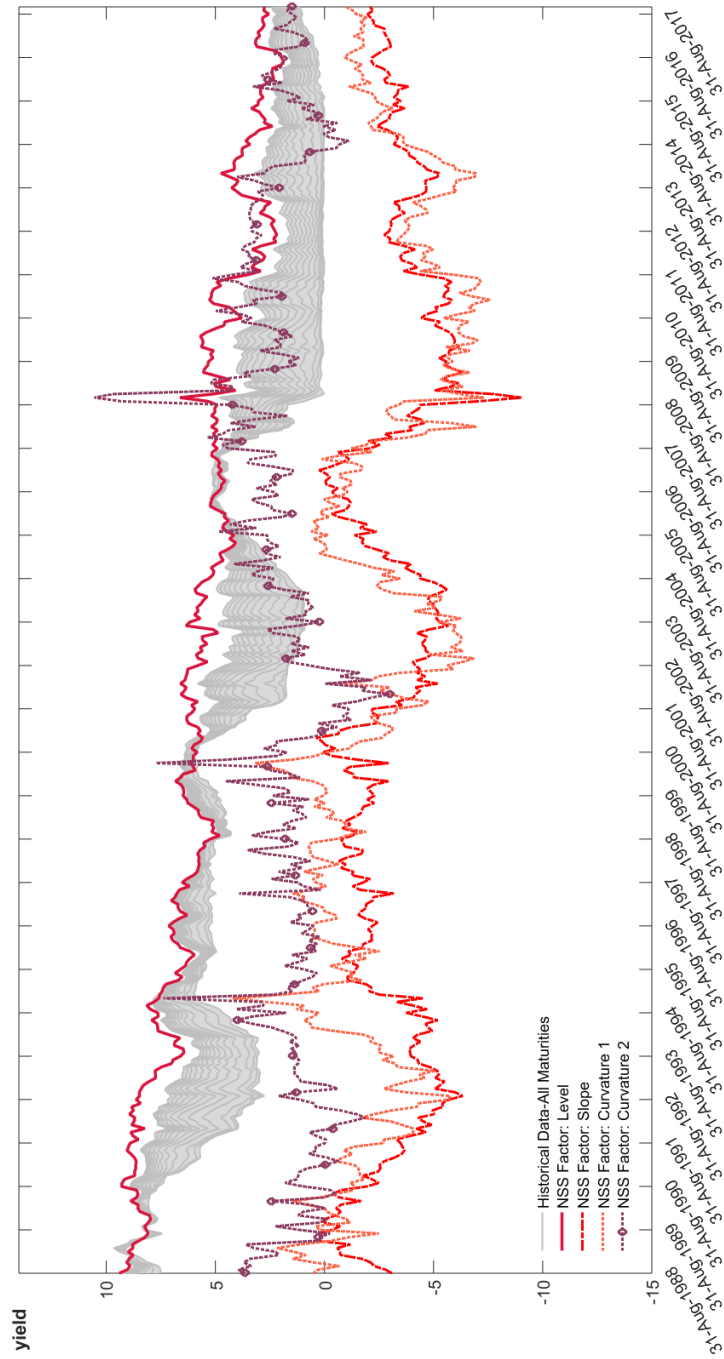
Notes: This figure displays the yield dataset used in the forecasting experiments summarized in Table 1-4. The yield dataset combines data from FRED H.15 and GSW datasets (see Section 5 for details). Plots of un-targeted real-time diffusion indexes estimated using data from 1988:8-2017:10 are also displayed.

Figure 3: Yields and Dynamic Nelson-Siegel Factors for the Sample Period 1988:8 - 2017:10*



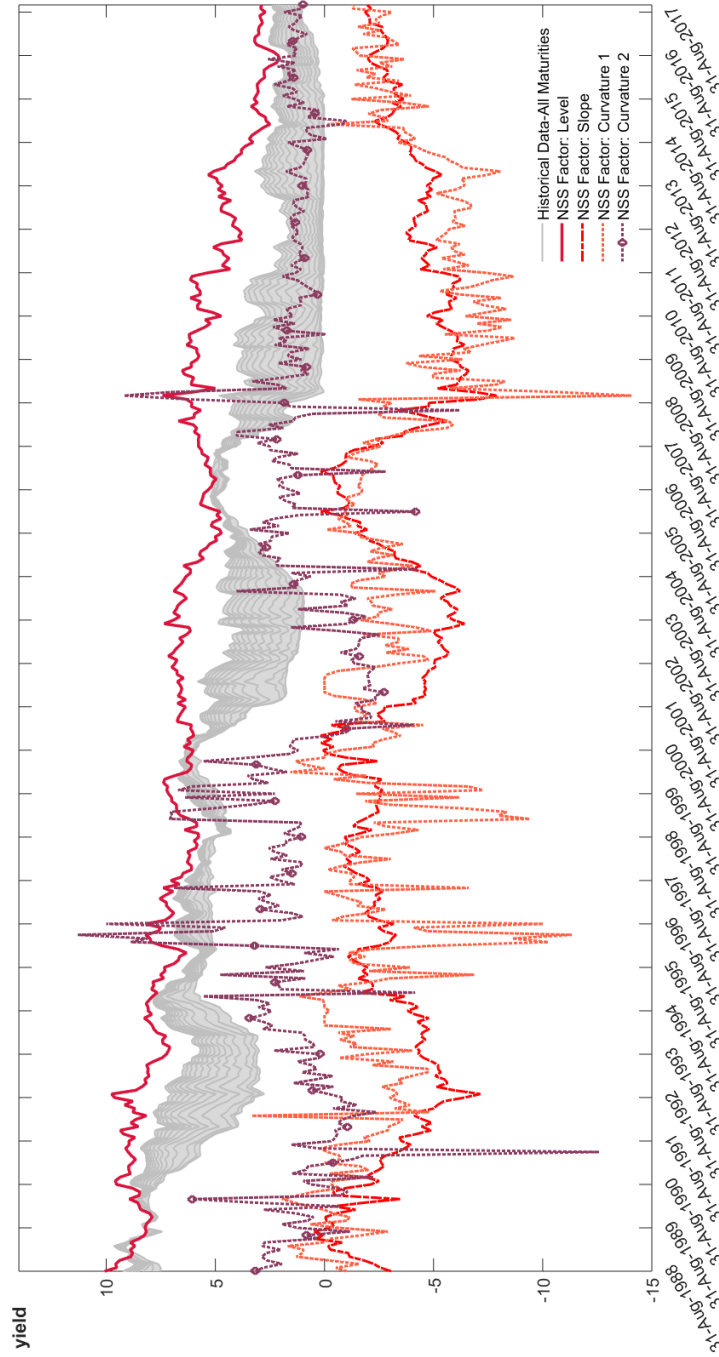
Notes: This figure displays the yield dataset used in the forecasting experiments summarized in Table 1-4. The yield dataset combines data from FRED H.15 and GSW datasets (see Section 5 for details). Plots of three DNS factors (level, slope, and curvature), constructed with a fixed rate of decay, and estimated using ordinary least squares, are also displayed. See Sections 2 and 4 for further details.

Figure 4: Yields and Dynamic Nelson-Siegel-Svensson Factors for the Sample Period 1988:8 - 2017:10*



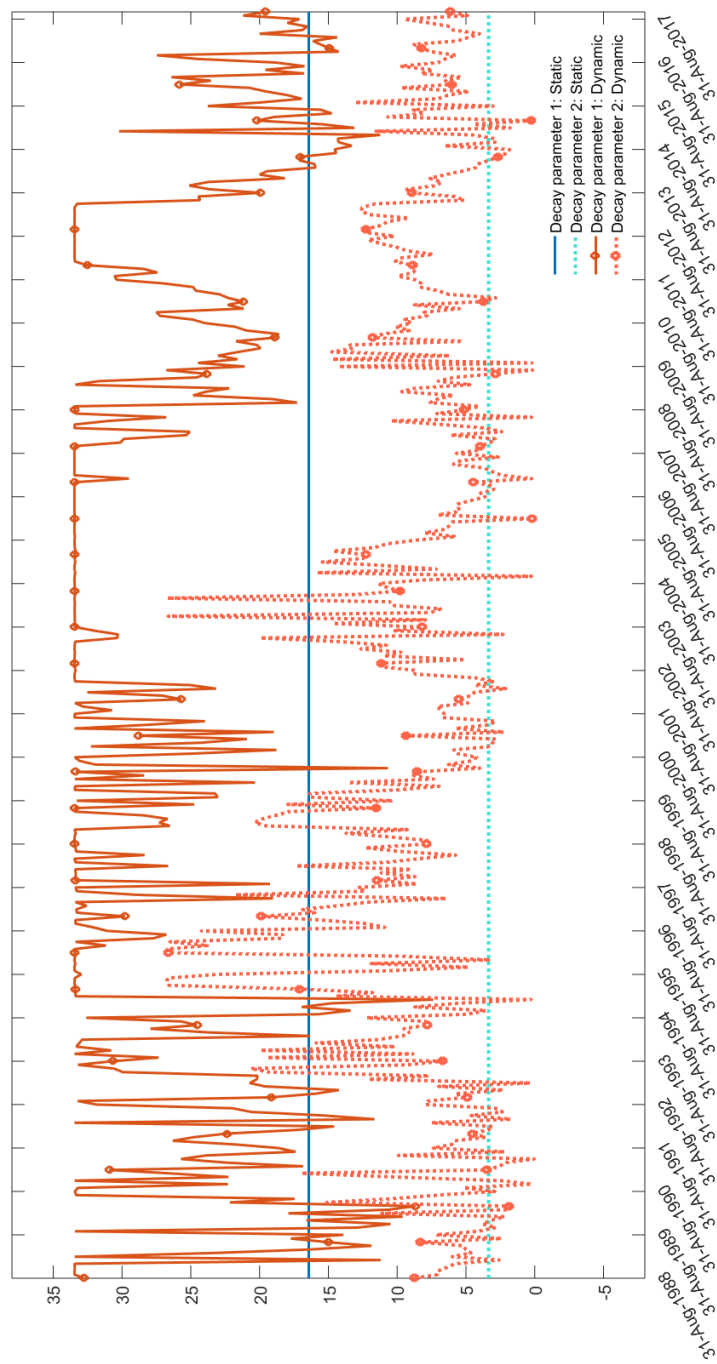
Notes: This figure displays the yield dataset used in the forecasting experiments summarized in Table 1-4. The yield dataset combines data from FRED H.15 and GSW datasets (see Section 5 for details). Plots of three NSS factors (level, slope, and curvature), constructed with a fixed rate of decay, and estimated using ordinary least squares, are also displayed. See Sections 2 and 4 for further details.

Figure 5: Yields and Dynamic Nelson-Siegel-Svensson Factors for the Sample Period 1988:8 - 2017:10*



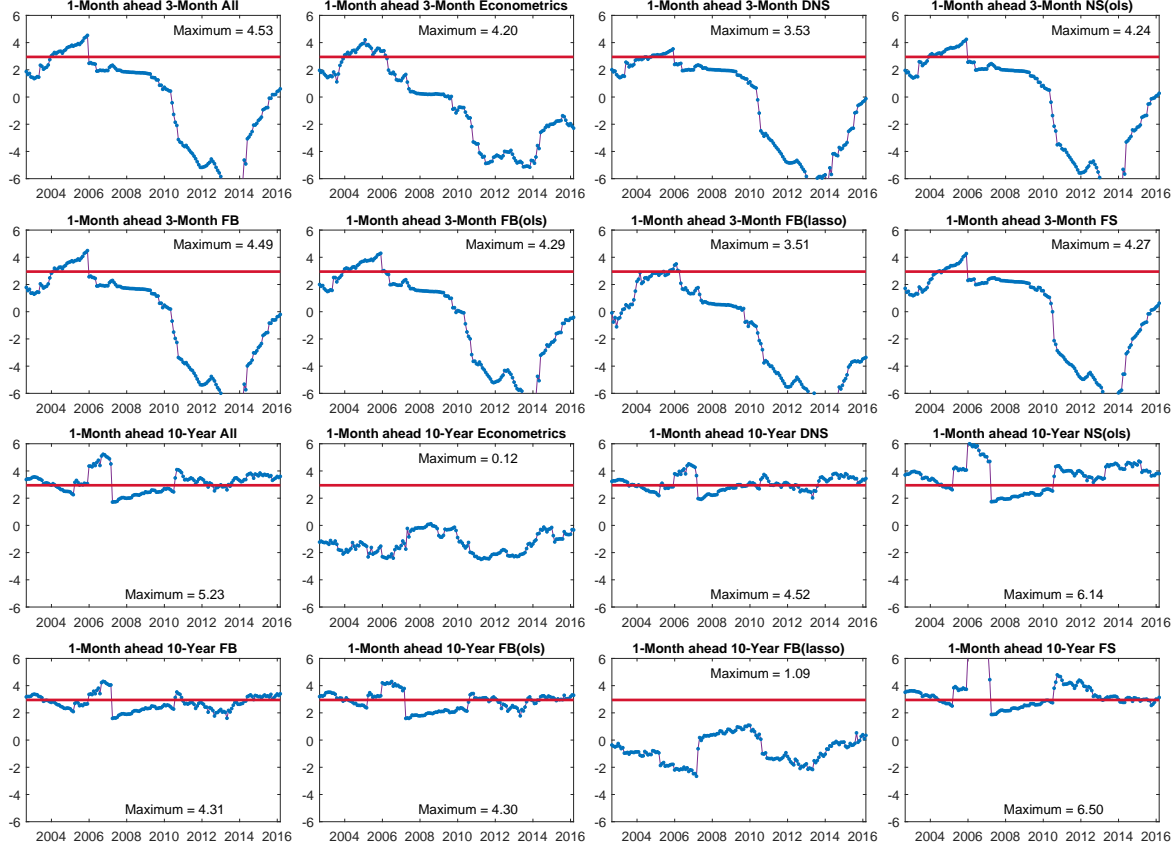
Notes: This figure displays the yield dataset used in the forecasting experiments summarized in Table 1-4. The yield dataset combines data from FRED H.15 and GSW datasets (see Section 5 for details). Plots of three NSS factors (level, slope, and curvature), constructed with a dynamic rate of decay, and estimated using ordinary least squares, are also displayed. See Sections 2 and 4 for further details.

Figure 6: Yields and Dynamic Nelson-Siegel-Svensson Factors for the Sample Period 1988:8 - 2017:10*



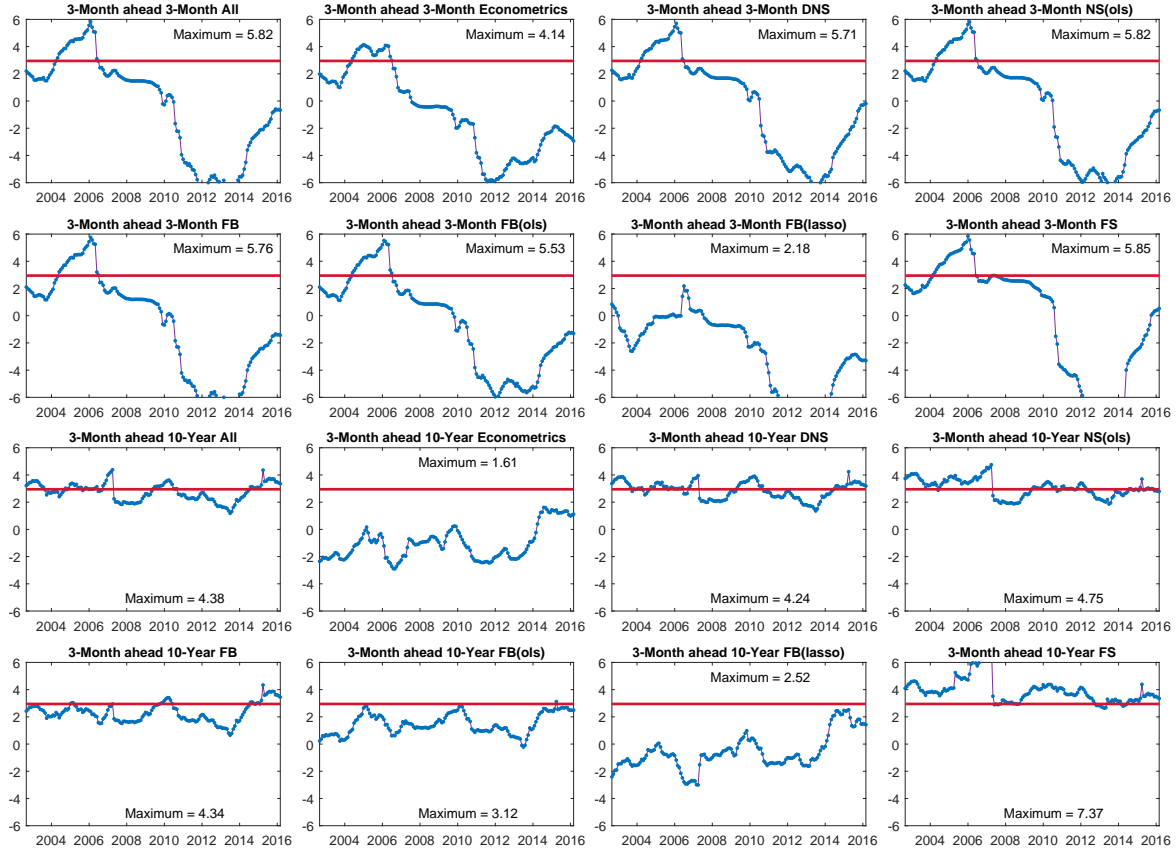
Notes: Static and dynamic decay rates.

Figure 7: Giacomini-Rossi Fluctuation Tests for Selected Model Combinations: $h = 1$ -Month Ahead Forecasts for 3-Month and 10-Year Maturity Yields*



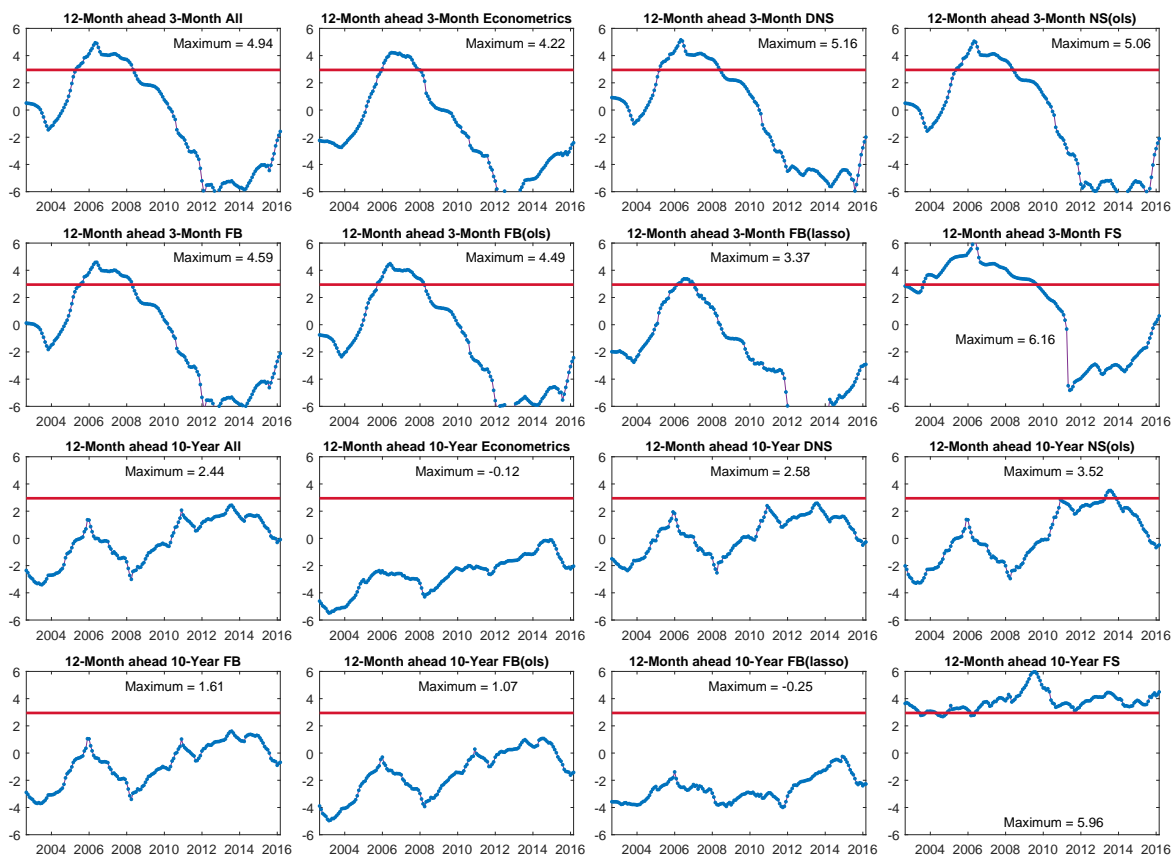
Notes: Refer to Section 4.2 and references therein for details about the fluctuation test. The fluctuation test statistic, which is the maximum value of the plotted statistics, is given at the top of each plot in this figure. See Section 6 for further discussion.

Figure 8: Giacomini-Rossi Fluctuation Tests for Selected Model Combinations: $h = 3$ -Month Ahead Forecasts for 3-Month and 10-Year Maturity Yields*



Notes: See notes to Figure 7.

Figure 9: Giacomini-Rossi Fluctuation Tests for Selected Model Combinations: $h = 12$ -Month Ahead Forecasts for 3-Month and 10-Year Maturity Yields*



Notes: See notes to Figure 7.