ESSAYS IN BIG DATA AND FORECASTING METHODS IN FINANCIAL ECONOMETRICS

by

WEIQI XIONG

A dissertation submitted to the

School of Graduate Studies

Rutgers, The State University of New Jersey

In partial fulfillment of the requirements

For the degree of

Doctor of Philosophy

Graduate Program in Economics

Written under the direction of

Norman Swanson

And approved by

New Brunswick, New Jersey

May, 2018

ABSTRACT OF THE DISSERTATION

Essays in Big Data and Forecasting Methods.

in Financial Econometrics

By WEIQI XIONG

Dissertation Director:

Professor Norman Swanson

This dissertation comprises two essays on big data and forecasting methods in financial econometrics. Methods for analyzing "big data" have received considerable attention by economists in recent years, given that applied practitioners now have an incredible amount of data available to them, and given that a whole host of new methods have been developed in various disciplines over the last 20 years or so. In the first essay, I discuss some of the latest (and most interesting) methods currently available for analyzing and utilizing big data when the objective is improved prediction. Additionally, I address predictive accuracy testing in the context of big data, and outline new loss function free methods that may be useful for forecast accuracy and model selection assessment. We also provide a brief empirical illustration of big-data in action, in which we show that big data are indeed useful when predicting the term structure of interest rates. This is done in a series of simple prediction experiments where the objective is to predict the term structure of interest rates, and models used include benchmark econometric models, dynamic Nelson Siegel (DNS) models (Diebold and Li, 2007) , diffusion index models (Stock and Watson, 2002), and hybrids of the three. The diffusion indexes in our experiments are estimates of the latent factors from principle component analysis of a macroeconomic dataset

including 103 U.S. variables. It is suggested that much remains to be learned regarding the ways in which extant econometric methods can be combined with dimension reduction methods in order to achieve improvements in prediction.

In the second essay, an extensive set of forecast experiments is conducted in order to explore the marginal predictive content of latent macroeconomic factors extracted from a so-called "data rich" or real-time dataset in dynamic Nelson-Siegel (DNS) type models. In particular, we assess the following classes of models: DNS type models of the variety, dynamic Nelson Siegel Svensson (NSS) type models (see Svensson (1994)), and various benchmark models, including vector autoregressive (VAR) and autoregressive (AR) models. The macroeconomic factors, or so-called "big data" diffusion indexes that we utilize are extracted using principle component analysis of 130 U.S macro-variables for which McCracken and Ng (2016) have constructed a real-time dataset. Experiments are carried out for various sub-samples between 2001 and 2018, and results are evaluated using a number of benchmark linear models. Additionally, various different dimensions are considered when specifying the yield cross section. Empirical results found are in contrast to the findings of Swanson and Xiong (2017), where including diffusion indexes always yields predictive improvement, although only fully revised macroeconomic data are utilized in that paper. Thus, the usefulness of diffusion indexes appears to hinge on whether or not a data-rich real-time environment is simulated in forecasting experiments or not.

Acknowledgements

This work would not have been possible without the financial support from Department of Economics and School of Graduate Studies at Rutgers University. I am especially indebted to my advisor and dissertation chair, Professor Norman R. Swanson, who have been very supportive of my career goals and who worked actively to provide me with suggestions, constructive criticism, guidance, encouragement throughout the period of study. I would like to express my gratitude to Professor Swanson for being a skilled supervisor and a great role model.

I would like to thank my committee members, Professor Yuan Liao and Professor Xiye Yang, and also outside member, Professor John Chao from University of Maryland. Thank you for investing time and providing inspiring and valuable feedback. I am proud and honored that you have accepted to be on my committee.

My sincerest thanks to Professor Klein, Professor Keister, and Professor Landon-Lane for generously sharing information and knowledge that enable me to develop skills in research and teaching. My gratitude also goes to the staffs of Economics Department, there are not enough words to describe your excellent work. Special thanks to Linda Zullinger, Janet Budge, Paula Seltzer, and Donna Ghilino.

My heartfelt thanks to my friends of the 2013 class: Ethan, Fatima, Geoff, Hyeon Ok, Jessica, Jieun, Maria, and Zheng. I am extremely grateful for their companionship and friendship over the years. Also, special thanks to Arpita, Hyunjoo, Mingmian, Ning, Ryuichiro, Sungkyung, Vladimir, Weijia, Won Hyung for your kind help and advice along the way. As it would be impossible to thank everyone who has helped, please accept this acknowledgment as a token of my appreciation and gratitude.

Finally, I am especially grateful to my parents, who supported me emotionally and financially. This journey would not have been possible without your constant love and encouragement.

Dedication

To my parents

Contents

List of Tables

Chapter 2

List of Tables (continue)

Chapter 3

List of Tables (continue)

Chapter 3

List of Figures

Chapter 2

Chapter 3

Chapter 1
Introduction

Methods for analyzing “big data”have received considerable attention by economists in recent years. This is not surprising, given that applied practitioners now have an incredible amount of data available to them, and given that a whole host of new methods have been developed in various disciplines over the last 20 years or so for processing these big data. Two key questions that economists continue to pose are, correspondingly, what are the forecasting gains associated with using big data, and which new methods should we use in our analyses? A third question, which is related, concerns which tools, such as predictive accuracy tests, to use for model selection with big data. In the context of forecasting, this third question is relevant because many critical advances have recently been made in the field of model selection and testing. In the first essay, we address all three questions. First, we discuss select state of the art methods for big data analysis. These include dimension reduction and shrinkage approaches that are currently being utilized not only in economics, but also in a whole host of other fields ranging from aerospace engineering to neuroscience. Second, we discuss recent advances in predictive accuracy testing and model selection, from the perspective of picking the “best” forecasting model. Finally, we tie our discussions together by considering the usefulness of big data when forecasting the term structure of interest rates.

In our empirical illustration, we show how important big data can be. Our objective is to predict U.S. Treasury yields of various maturities (i.e., the term structure of interest rates). Predictions will be made using “small data” models, including autoregressive, vector autoregressive, and dynamic Nelson-Siegel models, and “big data” models that utilize diffusion indexes estimated from a largescale macroeconomic dataset. The diffusion indexes in our experiments are estimates of the latent factors from principle component analysis of a macroeconomic dataset including 103 U.S. variables. Although the experimental setup that we utilize is limited in its scope, it is nevertheless interesting that the vast majority of mean square forecast error “best”models are hybrid DNS models that include diffusion indexes. Moreover, these hybrid models generally outperform standard econometric models, as well as various forecast combinations.

In the second essay, we add to the literature on interest rate prediction by carrying out an extensive set of forecast experiments in order to explore the marginal predictive content of so-called "data rich" or real-time latent macroeconomic factors in dynamic Nelson-Siegel (DNS) type models. In our context, data-rich environments contain real-time data, which are data that include the entire revision history for each variable. For example, real-time GDP observations for calendar date December 2000 include the first "reading" on 4th quarter 2000 GDP that was available in March 2001, and well as the 1st revised version of this datuum that became available in June 2001, and so on, up until the present date. Thus, real-time datasets include entire sequence 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. The macroeconomic factors, or so-called "big data" diffusion indexes that we utilize are extracted using principle component analysis of 130 U.S macro-variables for which McCracken and Ng (2016) have constructed a real-time dataset.

In particular, we examine the usefulness of real-time macroeconomic diffusion indexes when using dynamic Nelson-Siegel (DNS), dynamic Nelson-Siegel Svensson (NSS), and various econometric models for forecasting the term structure of interest rates. We find that the marginal predictive content of real-time diffusion indexes is significant for many of the models that we examine. We also find that model performance, across the board, is much worse post Great Recession. Indeed, not only does the predictive performance of DNS and NSS models worsen, in accord with the findings of various recent authors, but the performance of all of our models, including ones that utilize real-time diffusion indexes also worsens. Given the impressive predictive performance of these models prior to the great recession, we argue that new models need to be developed to address current economic conditions. Examples of models that might be useful include hybrid models in which the inclusion of diffusion indexes is triggered by variables such as predicted probabilities of recessions, economic variability, or the range of yields over some pre-defined prior period of time. We also present strong new evidence of the usefulness of forecast combination, and note that mean square error "best" (MSFE-best) forecast combinations often preclude the use of real-time diffusion indexes. This differs from earlier findings by Xiong and Swanson (2017), where it is found that if fully revised macroeconomic data are instead used in constructing diffusion indexes, then combinations that include diffusion index type models are MSFE-best. Thus, the usefulness of diffusion indexes appears to hinge on whether or not a data-rich real-time environment is simulated in forecasting experiments or not.

Chapter 2

Big Data Analytics in Economics: What Have We Learned so Far, and Where Should We Go From Here?

2.1 Introduction

Methods for analyzing "big data" have received considerable attention by economists in recent years. This is not surprising, given that applied practitioners now have an incredible amount of data available to them, and given that a whole host of new methods have been developed in various disciplines over the last 20 years or so for processing these big data. Two key questions that economists continue to pose are, correspondingly, what are the forecasting gains associated with using big data, and which new methods should we use in our analyses? A third question, which is related, concerns which tools, such as predictive accuracy tests, to use for model selection with big data. In the context of forecasting, this third question is relevant because many critical advances have recently been made in the field of model selection and testing. In this paper, we address all three questions. First, we discuss select state of the art methods for big data analysis. These include dimension reduction and shrinkage approaches that are currently being utilized not only in economics, but also in a whole host of other fields ranging from aerospace engineering to neuroscience. Second, we discuss recent advances in predictive accuracy testing and model selection, from the perspective of picking the "best" forecasting model. Finally, we tie our discussions together by considering the usefulness of big data when forecasting the term structure of interest rates.

In its inception, machine learning was a field of computer science concerned with designing computers (and computer programs) with the ability to learn, without the need for further programming. Many types of machine learning have been developed in recent years. For example, in computer science, key areas now include deep learning, shrinkage, and recall. Neural networks are perhaps the most ubiquitous variety of machine learning method that economists have, up until recently, been interested in. However, the landscape has changed dramatically in recent years, largely because of the explosion in big data. One strand of research in big data analysis uses dimension reduction methods, two main examples of which are principal components analysis (PCA) and partial least squares. A closely related strand considers shrinkage (penalized regression) methods, including the likes of ridge regression, the least absolute shrinkage selection operator (lasso), the elastic net, and the non-negative garrote. These and other shrinkage related methods are discussed in Bai and Ng (2008,2009), Schumacher (2009), Stock and Watson (2012), Kim and Swanson (2014,2016), and Hirano and Wright (2017), for example. Broadly speaking, the number of such methods available to empiricists is now immense.

In the first part of this paper, we discuss a very few of the latest such techniques, and suggest where we might go from here. For example, we discuss PCA and sparse PCA, in which the lasso is applied to PCA in order to induce sparseness in the number of observable variables utilized in the construction of latent factors or diffusion indexes resulting from application of PCA. We also discuss a related latent factor dimension reduction technique called independent component analysis, that takes the orthogonality condition imposed by PCA one step further by imposing statistically independence. Finally, we discuss ridge regression, the lasso, and the elastic net, in the context of penalized regression, where the number of regressors can be larger than the number of observations in a dataset.

In the second part of this paper, we discuss out-of-sample predictive accuracy testing, given the importance of accuracy assessment when comparing the many new "big data" methods available for constructing forecasts. There is now a rich literature on predictive accuracy testing. One of the most important contributions in the last 25 years is the seminal paper of Diebold and Mariano (1995, hereafter DM), in which tests of equal predictive accuracy between two competing models are proposed. Tests that generalize DM-type tests in order to account for parameter estimation error include West (1996) and West and McCracken (1998), McCracken (2000), and Corradi and Swanson (2007). Conditional predictive accuracy tests are developed in Giacomini and White (2006), in which the "estimated" model is conditioned on. Tests allowing for integrated and cointegrated variables are discussed in Clements and Hendry (1999,2001) and Corradi, Swanson and Olivetti (2001). The important issue of the joint comparison of more than two competing models is addressed in Sullivan, Timmermann and White (1999), White (2000), Hansen (2005), Romano and Wolf (2005), and Corradi and Distaso (2011). Papers that consider predictive accuracy testing via the use of encompassing and related tests include Phillips (1996), Harvey, Leybourne and Newbold (1997), Chao, Corradi and Swanson (2001), Clark and McCracken (2001), Corradi and Swanson (2002), and Giacomini and Komunjer (2005). Broadly speaking, predictive accuracy is assessed by comparing point measures such as mean square forecast error (MSFE) and mean absolute forecast error deviation (MAFD) in the above papers. The notion of considering predictive (error) densities rather than point error loss, model evaluation using predictive intervals, conditional quantiles, and predictive densities is addressed by Christoffersen (1998), Giacomini and Komunjer (2005), and Corradi and

Swanson (2005,2006a,b). For comprehensive surveys of this burgeoning literature, see West (2006), Clark and McCracken (2013), Corradi and Swanson (2013), and Diebold (2014).[1]

Recently, a new type of predictive accuracy tests have been devised that generalize the tests in all of the above papers, in one key dimension. In order to understand how this is done, note that most of the above papers consider forecast comparison based upon the examination of moments or conditional moments of the forecast errors, and researchers must specify the objective function (say, loss function or likelihood function) used in test formulation. As mentioned above, examples of relevant loss functions include MSFE and mean absolute forecast error MAFD. Unfortunately, the forecast superiority of one model, relative to other models, is dependent on the loss function that is specified. To circumvent this issue, Granger (1999a) proposes the use of generalized loss functions, $L(\cdot)$, with the following properties: (1) $L(e) = 0$, if the forecast error $e = 0$; (2) $L(e) \geq 0$ and $Min_e L(e) = 0$; and (3) $L(e)$ is monotonically non-decreasing as e moves away from zero (this means that $L(e_1) \geq L(e_2)$ if $e_1 > e_2 \geq 0$ or $e_1 < e_2 \leq 0$). Corradi, Jin and Swanson (2017, hereafter CJS) term the class of loss functions that satisfy the above three properties as general loss (GL or $\mathcal{L}_G$) functions. A second class of loss functions are defined as convex loss (CL or $\mathcal{L}_C$) functions, if in addition to satisfying the above three properties, they are convex. Examples of convex functions include MSFE and MAFD, as well as asymmetric functions including lin-lin and linex functions (see Elliott and Timmermann (2004) for further discussion). In CJS, it is supposed that there are l sets of forecasts, with corresponding sequences of one-step-ahead forecast errors, $\{e_{1t}\}$, $\{e_{2t}\}$..., $\{e_{lt}\}$, and the objective is to rank forecast sequences (or models), regardless of loss function. They establish links between tests for GL (CL) forecast superiority and tests for first (second) order stochastic dominance. This allows them to develop a forecast evaluation procedure that is based on an out-of-sample generalization of the stochastic dominance tests introduced by Linton, Maasoumi and Whang (2005, hereafter LMW), which is robust not only to the choice of loss function, but also to the possible presence of outliers. In addition to summarizing DM and related tests, the CJS test is discussed in detail below.[2]

In our empirical illustration, we show how important big data can be. This is done in a series of simple prediction experiments where the objective is to predict the term structure of interest rates,

[1] Alternatives to the use of traditional moment-based forecast evaluation methods include decision based approaches. For example, Granger and Pesaran (2000) argue in favor of a close link between the decision and the forecast evaluation problems. Pesaran and Skouras (2002) discuss a decision-based approach for evaluation and comparison of forecasts. Granger and Machina (2006) propose a class of realistic decision-based loss functions for forecast evaluation.

[2] The approach of using stochastic dominance to rank distributions of forecast errors was first introduced in Corradi and Swanson (2013), although they provide no theory, and their proposed tests are loss function specific. An alternative somewhat related measure called stochastic error loss is discussed in Diebold and Shin (2015).

and models used include benchmark econometric models, dynamic Nelson Siegel (DNS) models, diffusion index models, and hybrids of the three. The diffusion indexes in our experiments are estimates of the latent factors from principle component analysis of a macroeconomic dataset including 103 U.S. variables. Although the experimental setup that we utilize is limited in its scope, it is nevertheless interesting that the vast majority of mean square forecast error "best" models are hybrid DNS models that include diffusion indexes. Moreover, these hybrid models generally outperform standard econometric models, as well as various forecast combinations.

The rest of the paper is organized as follows. Section 2 summarizes recent advances in dimension reduction and penalized regression - both of which are key areas in machine learning. In Section 3, forecast evaluation is discussed, with emphasis on what the latest methods are, and where we need to go. An empirical illustration based on predicting the term structure of interest rates is given in Section 4. Finally, concluding remarks are gathered in Section 5.

2.2 Dimension Reduction and Penalized Regression

Dimension reduction and variable selection has never been more important in economics, given recent massive increases in the amount of data available to forecasters.[3] A key objective, given big data, is to remove redundant and irrelevant information from datasets. This problem has historically been be tackled via step-wise regression, for example. However, variables are typically highly correlated in time series applications. Hence, statistical significance tests used in many regression type algorithms suffer from severe size distortion issues. Ghysels, Hill, and Motegi (2017) address this issue by examining multiple parsimonious regressions, each with one key regressor, while jointly accounting for sequential testing problems.

A second solution to the dimension reduction problem with correlated regressors is the use of partial least squares (PLS), which was originally proposed by Herman Wold in the mid 1960s. Broadly speaking, PLS is a latent variable approach to modeling the covariance structure between two sets of variables. One set might be a target variable or variables to be predicted (say Y), while the other might be a very large set of correlated predictor variables, say X. More precisely, the model underlying PLS has

$$\begin{aligned} Y &= F_1 L_1 + E_1 \\ X &= F_2 L_2 + E_2, \end{aligned}$$

where F_1 and F_2 are projection matrices of X and Y; and $L_1$and L_2 are so-called factor loading

[3]See the 2015 issue of the *Journal of Econometrics* entitled **High Dimensional Problems in Econometrics**.

matrices that operate on the latent factors F_1 and F_2. Additionally, the error terms, E_1 and E_2 are assumed to be identically and independently distributed, and all matrices are conformably defined, given the dimensions of X and Y. In this setup, the decompositions of X and Y maximize the covariance between the latent factors F_1 and F_2.

A third solution uses principle components analysis (PCA), in which latent factors (often called diffusion indexes) are again estimated, but this time via use of an eigenvalue-eigenvector decomposition of the covariance or correlation matrix of the data, for example. Just as in PLS, the objective is to "explain" the data" using a reduced set of (latent) explanatory variables, with the idea being that the useful information in a large set of predictors is often contained in a (much smaller) set of latent factors, which are themselves simply linear combinations of the original variables. A key difference between PCA and PLS is that PLS directly attempts to account for correlation between the target variable and the predictors, while PCA is "unsupervised", in the sense that correlation with any given target variable is not emphasized in the construction of the latent factors. Rather, overall explanation of the entire dataset is the focus of PCA. Needless to say, this particular feature of PCA is of potential concern when targeting (predicting) a specific variable or variables. For this reason, many supervised versions of PCA have been developed. For example, Carrasco and Rossi (2016) use cross validation methods to supervise PCA, while Bai and Ng (2008) consider targeted forecasting using subsets of X (see also Armah and Swanson (2010a,b)) and Cheng, Swanson, And Yang (2017). Given its ease of application as well as recent empirical evidence on its usefulness, PCA (which is the oldest of the methods discussed in this paper; see Spearman (1904) and the discussion in Swanson (2016) for further details), has received the most attention in economics recently, and hence will be discussed in considerably more detail below.

Penalized regression or shrinkage methods, which reduce or shrink redundant or irrelevant variables are also important in big data analysis. Key examples include ridge regression, the lasso, and the elastic net. When viewed through the lens of multivariate regression analysis, all of these methods involve shrinking the magnitude of coefficients in regression models. When the "penalty functions" are carefully designed, and when the "regularization parameters" used to regulate the strength of the penalties in these functions are of sufficient magnitude, then substantial dimension reduction can be achieved. For example, when shrinkage is used in conjunction with PCA, factor loading matrices can be induced to be sparse, in the sense that certain coefficients in the linear combinations of the predictor variables are identically zero. This nice feature imposes parsimony on the number of variables used to form latent factors in PCA, whereas under standard PCA; all predictors receive non-zero weight in each latent factor. Just as in the case of PLS, the number of predictors may be greater than the number of observations in the dataset being analyzed using

PCA.

To fix ideas, let's consider the "original" shrinkage estimator. Namely, assume that we are interested in the model:

$$Y = X\theta + \varepsilon,$$

where Y contains data on a single variable, there are many (possibly highly correlated) variables represented in the data matrix, X, and ε is an error term. Later, we shall introduce the ridge estimator slightly differently, but for now, note that the ridge estimator can be expressed as:

$$\widehat{\theta}_{ridge} = (X'X + \lambda I)^{-1}X'Y.$$

The "ridge" down the diagonal in this estimator is equivalent to adding a penalty of $\lambda \sum_{i=1}^{N} \widehat{\theta}_i^2$ to the usual residual sum of squares term that is minimized in least squares estimation, where N is the number of predictors in X. Here, as $\lambda \to 0$, $\widehat{\theta}_{ridge} \to \widehat{\theta}_{ols}$, and as $\lambda \to \infty$, $\widehat{\theta}_{ridge} \to 0$. Evidently, applying the ridge penalty shrinks parameter estimates towards zero, which increase bias and reduces estimator variance. One very important feature of ridge regression is that invertibility problems associated with $X'X$ when the number of predictors is too large relative to the number of observations are no longer an issue, and there is always a unique solution (i.e., $\widehat{\theta}_{ridge}$). Other shrinkage estimators that shall be discussed in the sequel include one where the penalty function is $\lambda \sum_{i=1}^{N} \left|\widehat{\theta}_i\right|$ (the lasso) and another that combines both of the above penalty functions (the elastic net).

Another shrinkage estimator is based on bootstrap aggregation (bagging), and was introduced by Breiman (1996). Stock and Watson (2012) note that predictions of Y, at a point in time, $T+1$, conditional on information available up through period T, say $y^f_{T+1|T}$ can be constructed as follows:

$$y^f_{T+1|T} = \sum_{i=1}^{N} \psi(\lambda t_{\widehat{\theta}(i)})\widehat{\theta}(i)X_T(i),$$

where $X_T(i)$ is the datum on the i^{th} variable in X for period T, $\widehat{\theta}(i)$ is the least squares estimator from regressing $X_{T-1}(i)$ on Y_T, and $\psi(\lambda t_{\widehat{\theta}(i)})$ is a regularized (through λ) function of the t-statistic associated with the aforementioned regression.[4] For bagging $\lambda = 1$, while various Bayesian predictors, including Bayesian model averaging and empirical Bayes can also be formulated in this manner, by setting λ appropriately. Interestingly, Hirano and Wright (2017) show that forecasting models constructed using out-of-sample or split sample schemes perform well only when combined with

[4] In their setup, Stock and Watson (2012) assume that the predictors are zero mean random orthonormal variables. Also, Y_t is assumed to be zero mean, and the underlying model is assumed to be:

$$Y_t = \theta' X_{t-1} + \varepsilon_t,$$

where ε_t is an error term with fixed variance.

other methods, such as bagging. Broadly speaking, their results offer a glimpse into the benefits of using state of the art (asymptotic) statistical analysis in order to examine new methods that combine conventional out-of-sample approaches to model selection and estimation with algorithmic approaches such as bagging. In their paper, they show that out-of-sample schemes so regularly used for model selection (and estimation are inefficient when applied in the conventional manner. This finding is reversed when bagging or other risk reduction methods are combined with conventional out-of-sample schemes, however.

2.2.1 Static and Dynamic Factor Augmented Forecasting Models

Some of the most highly touted recent developments in forecasting center around estimation and asymptotic properties of diffusion indexes based on PCA; and the use of diffusion indexes in the construction of forecasting models. Following the discussion of Stock and Watson (2002a,b) and Armah and Swanson (2010a,b), we summarize key features of recent developments by considering static and dynamic factor models in order to motivate the use of diffusion indexes in forecasting.

Let y_{t+h} be the scalar target forecast variable and X_t be an N-dimensional vector of predictor variables, for $t = 1, \ldots, T$. Assume that (y_{t+1}, X_t) has a dynamic factor model representation with $\overline{r}$ common dynamic factors, f_t, which can be written as:

$$y_{t+h} = \beta' W_t + \alpha(L) f_t + \varepsilon_{t+h} \tag{2.1}$$

and

$$x_{it} = \lambda_i(L) f_t + e_{it}, \tag{2.2}$$

for $i = 1, 2, \ldots, N$, where W_t is an $l \times 1$ vector of observable variables with $l << N$, including lags of y_t; $\alpha(L) = \sum_{j=0}^{q} \alpha_j L^j$ and $\lambda_i(L) = \sum_{j=0}^{q} \lambda_{ij} L^j$ are finite order lag polynomials in nonnegative powers of L; and $h > 0$ is the forecast horizon. Thus, all variables in X_t can be expressed as a linear function of the dynamic factors (and an idiosyncratic shock, e_{it}). This dimension reducing feature of the model is the key feature worth noting. Now, we can write (2.1) and (2.2) in static form as:

$$y_{t+h} = \beta' W_t + \alpha' F_t + \varepsilon_{t+h} \tag{2.3}$$

and

$$x_{it} = \Lambda_i' F_t + e_{it}, \tag{2.4}$$

where $F_t = (f_t', \ldots, f_{t-q}')'$ is an $r \times 1$ vector of static factors, with $r = (q+1)\overline{r}$, α is an $r \times 1$ vector, and $\Lambda_i = (\lambda_{i0}', \ldots, \lambda_{iq}')'$ is a vector of factor loadings on the static factors, where λ_{ij} is an $\overline{r} \times 1$ vector for $j = 0, \ldots, q$ and $\beta = (\beta_1, \ldots, \beta_l)'$. The model in (2.3) is the “factor augmented forecasting model” presented in the diffusion index forecasting framework of Stock and Watson (2002a,b), and discussed further in Bai and Ng (2007). The static factor in (2.4) is thus named because the contemporaneous relationship between x_{it} and F_t. One major advantage of the static representation of the dynamic factor model is it enables us to use principal component analysis to estimate the factors. This involves estimating F_t using an eigenvalue-eigenvector decomposition of the sample covariance matrix of the data, after standardizing said data. Moreover, an important theoretical feature of the model in (2.3) is that consistent estimation of the factors in F_t, which can be achieved via simple application of PCA, allows for subsequent $\sqrt{T}$ consistent estimation of α and β in (2.3) using quasi-maximum likelihood, as long as $\sqrt{T}/N \to 0$, as $N, T \to \infty$. Thus, as shown in Bai and Ng (2006), F_t, when estimated using the PCA method outlined in Stock and Watson (2002a,b), can be treated as a vector of observed regressors, eschewing the need to address the generated regressor problem that often arises in applied econometrics. For a discussion of alternative methods for factor forecasting based on estimation of generalized dynamic factor (GDF) models, see Forni, Hallin, Lippi and Reichlin (2005) and Forni, Hallin, Lippi and Zaffaroni (2015). For further discussion of consistent estimation of factors in static as well as GDF models, see Ding and Hwang (1999), Forni, Hallin, Lippi and Reichlin (2000), Stock and Watson (2002b), Bai and Ng (2002) and Bai (2003), who show that the space spanned by both the static and dynamic factors can be consistently estimated when N and T are both large.

For forecasting purposes, little is gained from a clear distinction between static and dynamic factors (see Schumacher (2007) for a comparison of forecasts based on the use of factors estimated using static, dynamic, and other estimation methods). Moreover, Boivin and Ng (2005) compare alternative factor based forecast methodologies, and conclude that when the dynamic structure is unknown and the model is characterized by complex dynamics, the approach of Stock and Watson performs favorably.

Many important issues have been addressed in recent papers on diffusion index forecasting. For example, Bai and Ng (2006a) stress that the regressors (factors) in diffusion index models are estimated, which substantially increases forecast error variances, relative to a simpler setup where diffusion indexes are not estimated. In a related paper, Bai and Ng (2006b) examine whether observable economic variables can serve as proxies for the underlying unobserved factors. In particular, they use a variety of statistics to determine whether a group of observed variables yields the same information as that contained in the latent factors. Stock and Watson (2002a) have also attempted to link factors to observed variables. Armah and Swanson (2010) argue that if individual observable economic variables are indeed good proxies of the unobserved factors, then these proxies can be used in place of the factors in the diffusion index model for prediction. Once the set of factor proxies is fixed, one effectively eliminates the incremental increase in forecast error variance (i.e., uncertainty) associated with the use of estimated factors. Along these lines, they consider "smoothed"versions of the Bai and Ng (2006b) statistics that pre-select a set of factor proxies prior to the ex-ante construction of a sequence of predictions. Stock and Watson (1998,2009) demonstrate that when PCA is used in estimation, factors remain consistent even when there is some time variation in factor loadings and small amounts of data contamination, so long as the number of variables in the panel dataset or the number of predictors is very large (i.e., $N >> T$). The usefulness of factor augmented models that include cointegration restrictions is discussed in Banerjee, Marcellino and Marsten (2014). The importance of assessing and testing for structural breaks in these models is discussed in Banerjee, Marcellino and Marsten (2008), Stock and Watson (2009), and Chen, Dolado and Gonzalo (2014). Factor loading and parameter stability testing is addressed in Corradi and Swanson (2014), Breitung and Eickmeier (2011), Goncalves and Perron (2014), and Han and Inoue (2014). Finally, the empirical and theoretical properties of factor augmented VARMA models are investigated in Dufour and Stevanovic (2013).

For readers interested in estimation of factors used in (2.3), we close this section by outlining further details, drawing directly on Armah and Swanson (2010a,b). Let k ($k < \min\{N, T\}$) be an arbitrary number of factors, Λ^k be $N \times k$ factor loadings matrix, $(\Lambda_1^k, \ldots, \Lambda_N^k)'$, and F^k be the $T \times k$ matrix of factors $(F_1^k, \ldots, F_T^k)'$. From (2.4), estimates of Λ_i^k and F_t^k are obtained by solving the optimization problem:

$$V(k) = \min_{\Lambda^k, F^k} (NT)^{-1} \sum_{i=1}^{N} \sum_{t=1}^{T} (x_{it} - \Lambda_i^{k\prime} F_t^k)^2. \tag{2.5}$$

Let $\widetilde{F}^k$ and $\widetilde{\Lambda}^k$ be the minimizers of equation (2.5). Since Λ^k and F^k are not separately identifiable, if $N > T$, a computationally expedient approach would be to concentrate out $\widetilde{\Lambda}^k$ and minimize (2.5) subject to the normalization $F^{k\prime} F^k / T = I_k$. Minimizing (2.5) is equivalent to maximizing $tr[F^{k\prime}(XX')F^k]$. This optimization is solved by setting $\widetilde{F}^k$ to be the matrix of the k eigenvectors of XX' that correspond to the k largest eigenvalues of XX'. Note that $tr[\cdot]$ represents the matrix trace. Let $\widetilde{D}$ be a $k \times k$ diagonal matrix consisting of the k largest eigenvalues of XX'. The estimated factor matrix, denoted by $\widetilde{F}^k$, is $\sqrt{T}$ times the eigenvectors corresponding to the k largest eigenvalues of the $T \times T$ matrix XX'. Given $\widetilde{F}^k$ and the normalization $F^{k\prime}F^k/T = I_k$, $\widetilde{\Lambda}^{k\prime} = (\widetilde{F}^{k\prime}\widetilde{F}^k)^{-1}\widetilde{F}^{k\prime}X = \widetilde{F}^{k\prime}X/T$ is the corresponding factor loadings matrix.

The solution to the optimization problem in (2.5) is not unique. If $N < T$, it becomes computationally advantageous to concentrate out $\overline{F}^k$ and minimize (2.5) subject to $\overline{\Lambda}^{k\prime}\overline{\Lambda}^k/N = I_k$. This minimization is the same as maximizing $tr[\Lambda^{k\prime}X'X\Lambda^k]$, the solution of which is to set $\overline{\Lambda}^k$

equal to the eigenvectors of the $N \times N$ matrix $X'X$ that correspond to its k largest eigenvalues. One can thus estimate the factors as $\overline{F}^k = X'\overline{\Lambda}^k/N$. $\widetilde{F}^k$ and $\overline{F}^k$ span the same column spaces, hence for forecasting purposes, they can be used interchangeably. Given $\widetilde{F}^k$ and $\widetilde{\Lambda}^k$, let $\widehat{V}(k) = (NT)^{-1}\sum_{i=1}^{N}\sum_{t=1}^{T}(x_{it} - \widetilde{\Lambda}_i^{k\prime}\widetilde{F}_t^k)^2$ be the sum of squared residuals from regressions of X_i on the k factors, $\forall i$. A penalty function for over fitting, $g(N,T)$, is chosen such that the loss function

$$IC(k) = \log(\widehat{V}(k)) + kg(N,T) \tag{2.6}$$

can consistently estimate r. Let $kmax$ be a bounded integer such that $r \leq k\max$. Bai and Ng (2002) propose three versions of the penalty function $g(N,T)$, namely, $g_1(N,T) = \left(\frac{N+T}{NT}\right)\log\left(\frac{NT}{N+T}\right)$, $g_2(N,T) = \left(\frac{N+T}{NT}\right)\log C_{NT}^2$, and $g_3(N,T) = \left(\frac{\log(C_{NT}^2)}{C_{NT}^2}\right)$, all of which lead to consistent estimation of r. Additional details on the estimation of r are contained in Bai and Ng (2002). Alternative methods for selecting r are discussed in Chen, Huang, and Tu (2010), Onatski (2015), Carrasco and Rossi (2016), and the references cited therein.

For further reading in the area of factor models, including high dimensional covariance matrix estimation in approximate factor models and projected principal components analysis in factor models, see Fan, Liao and Wang (2016) and Fan, Laio and Mincheva (2011).

2.2.2 New Directions in Diffusion Index Estimation

As discussed earlier, ongoing research efforts in the study of factor augmented forecasting models include the analysis of problems associated with the "selection" of diffusion indexes that are most useful for predicting y_{t+1}. For example, see Bai and Ng (2008,2009) and Schumacher (2009), who discuss using targeted predictors based on quadratic principal components and thresholding rules for variable subset selection to estimate diffusion indexes. Armah and Swanson (2010a,b) also discuss this issue. Further, Carrasco and Rossi (2016) propose cross validation methods for selecting the "best" diffusion index for use in forecasting. A related area of research, which is the subject of this subsection, is the development of alternative diffusion index estimators, important examples of which use shrinkage methods in order to impose sparseness on the factor loadings used in the construction of diffusion indexes. Two of the many interesting new estimators in this context include sparse principal components analysis (SPCA) and independent component analysis (ICA).

Zou, Hastie, and Tibshirani (2006) note that diffusion indexes estimated using PCA are linear combinations of all underlying predictor variables, and factor loadings are hence all nonzero, which adversely affects the parsimony of forecasting models, a property known to be important in time series forecasting. Moreover, they stress that diffusion indexes are thus difficult to interpret. In light of this, they propose SPCA, in which the least absolute shrinkage selection operator (lasso) or the related shrinkage estimator called the elastic net is utilized in order to construct principal components with sparse loadings. This is done this by first reformulating PCA as a regression type optimization problem, and then by using a lasso (elastic net) on the coefficients in a suitably constrained regression model.

Before further discussing SPCA, it is worth noting that the lasso and elastic net are important techniques for big data analysis in and of themselves, and are related to the venerable ridge regression estimator. Using the above notation, say that

$$y_t = X_t'\theta + \varepsilon_t.$$

Here, penalized (shrinkage type) regression is carried out as follows: For the ridge estimator, construct:

$$\widehat{\theta}_{ridge} = \arg\min_{\theta} \left\{ \left\| y - \Sigma_{i=1}^{N} X_i \theta_i \right\|^2 + \lambda_2 \Sigma_{i=1}^{N} \theta_i^2 \right\},$$

where y is the Tx1 target variable, $X = [X_1, ..., X_N]$, $i = 1, ..., N$ is the TxN predictor matrix, with $X_i = (X_{1,i}, ..., X_{T,i})'$, and $\lambda > 0$ is the tuning parameter. Notice that this is an alternative formulation of $\widehat{\theta}_{ridge}$ to that given earlier. The more recently developed lasso and the elastic net estimators involve imposition of L_1 (lasso) and L_1+L_2 $-norm$ penalties on parameter magnitudes, and are formulated as:

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

and

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

The choice of regularization parameters can impact on the predictive performance of models specified using these sorts of methods. For a discussion further of the regularization parameters, including values to use thereof, please refer to Kim and Swanson (2017), as well as the papers cited in Kim and Swanson where the various estimation algorithms for these methods are developed.

Interestingly, SPCA follows directly by formulating PCA as a regression-type optimization problem, and then by subsequently imposing lasso (elastic net) constraints on the regression coefficients in the optimization problem. Put simply, factor loading can be recovered by regressing principal components on the N variables in X_t, as shown in Zou, Hastie, and Tibshirani (2006). Here, imposition of the L_2 $-norm$ penalty in ridge regression allows for $N > T$. Moreover, when the lasso or elastic net is utilized in this context, then large enough λ_1 yields sparse $\widehat{\theta}$. In this sense, SPCA is a natural data reduction method. Since the important paper by Zou et al., many authors have proposed modifications to SPCA, as discussed in Kim and Swanson (2017).

Broadly speaking, the lasso and elastic net constitute two of the most important penalized regression methods currently available, in which all predictor variables are retained in a model, but are constrained (regularized) by shrinking them towards zero. For important descriptions of these methods, see Tibshirani (1996), Zou and Hastie (2005), and Zou (2006).

All of the above penalized regression or shrinkage type methods are examples of machine learning. Other machine learning algorithms have also recently been explored in economics. Two examples are bagging and boosting. Bagging (also called bootstrap aggregation) involves first drawing bootstrap samples from an in-sample training dataset, and then constructing predictions, which are later combined. This algorithm is discussed above. Boosting is another so-called machine learning ensemble meta-algorithm algorithm that utilizes a supervised and user-determined set of functions or *learners* (e.g., least square estimators), and uses the set repeatedly on filtered data, which are typically outputs from previous iterations of the learning algorithm. Broadly speaking, boosting isolates which variables, from amongst a large group of variables, are useful for predicting a target variable. More specifically, boosting estimates an unknown function (e.g., the conditional mean) using sequential step-wise forward regression, with learners that may not only be least squares estimators, but may also be smoothing splines and kernel regressions, for example. For further discussion of boosting, see Freund and Schapire (1997), Bai and Ng (2009), Kim and Swanson (2014), and the references therein.

Two further examples include the non-negative garrote (see Breiman (1995) and Yuan and Lin (2007)) and least angle regression (see Efron, Hastie, Johnstone and Tibshirani (2004) and Bai and Ng (2008)), both of which are closely related to the elastic net.

Returning to the main subject of this section, we now discuss independent component analysis, which is predicated on the idea of "opening" the black box in which principal components often reside, and is an alternative to PCA and SPCA. ICA is used in many applications, from brain imaging to stock price return modeling. In all cases, there is a large set of observed individual signals, and it is assumed that each signal depends on several factors, which are unobserved. In this sense, the motivation is exactly the same as that used to justify PCA.

The starting point for ICA is the very simple assumption that the components, say F, are statistically independent in equation (2.3). This assumption is potentially much stronger than the orthogonality imposed under PCA. The key issue in ICA is the measurement of the "level" of independence between components. More specifically, ICA begins with statistically independent (and unobserved) source data, S, which are mixed according to an unknown "mixing matrix", Ω; and X, which is observed, is a mixture of S, weighted by Ω. For simplicity, we assume that the unknown mixing matrix, Ω, is square, although this assumption can be relaxed. Thus, it is assumed that $X = S\Omega$. Stated differently, assume that:

$$\begin{aligned} X_1 &= \omega_{11} S_1 + \cdots + \omega_{1N} S_N \\ X_2 &= \omega_{21} S_1 + \cdots + \omega_{2N} S_N \\ &\vdots \\ X_N &= \omega_{1N} S_1 + \cdots + \omega_{NN} S_N, \end{aligned} \tag{2.7}$$

where ω_{ij} is the (i, j) element of Ω. Since Ω and S are unobserved, one must estimate the "demixing matrix", Ψ, which transforms the observed X into the independent components, F. That is, $F = X\Psi$, or $F = S\Omega\Psi$. As detailed in Kim and Swanson (2017), if Ω is square, then so is Ψ, and $\Psi = \Omega^{-1}$, so that F is exactly the same as S, and perfect separation occurs. In general, it is only possible to find Ψ such that $\Omega\Psi = PD$, where P is a permutation matrix and D is a diagonal scaling matrix. The independent components, F are latent variables, and are analogous to the principal components discussed in the case of PCA. In summary, upon estimation of Ω and S, it is feasible to estimate the demixing matrix Ψ, and the independent components, F. However (2.7) is not identified unless several assumptions are made. The first assumption is that the sources, S, are statistically independent. Since various sources of information (for example, consumer's behavior, political decisions, etc.) may have an impact on the values of macroeconomic variables, this assumption is not strong. The second assumption is that the signals are stationary. For further details, see Tong, Liu, Soon, Huan (1991). ICA maps the N components of X into the rank N matrix, F. However, we can simply construct factors using up to r $(< N)$ components, without loss of generality, for comparability with PCA. Alternatively, one might carry out ICA using r principal components, hence further filtering diffusion indexes constructed using PCA in order to obtain statistically independent variants thereof (see Stone (2004) for further details). In general, the above model would be more realistic if there were noise terms added. See Hyvärinen and Oja (2000) for a detailed discussion of the noise-free model, and Hyvärinen (1998,1999) for a discussion of the model with noise added.

For a detailed comparison of ICA with PCA, see Kim and Swanson (2016), who note that the main difference between ICA and PCA is in the properties of the factors obtained. Principal components are uncorrelated and have descending variance so that they are naturally ordered in terms of their variances. While setting the diffusion index in equation (2.1) equal to the highest variance (correlation) principal components may well not equate with the specification of the indexes that are most useful for forecasting a given variable, say y_t, it is certainly the case that components explaining the largest share of the variance are often assumed to be the "relevant" ones. For simplicity, consider two observables, $X = (X_1, X_2)$. PCA finds a matrix which transforms X into uncorrelated components $F = (F_1, F_2)$, such that the uncorrelated components have a joint probability density function, $p_F(F)$ with:

$$E(F_1F_2) = E(F_1)E(F_2). \tag{2.8}$$

On the other hand, ICA finds a demixing matrix which transforms the observed $X = (X_1, X_2)$ into independent components $F^* = (F_1^*, F_2^*)$, such that the independent components have a joint pdf $p_{F^*}(F^*)$ with:

$$E\left[F_1^{*p}F_2^{*q}\right] = E\left[F_1^{*p}\right]E\left[F_2^{*q}\right], \tag{2.9}$$

for every positive integer value of p and q. Evidently, ICA is more restrictive, and it should thus not be surprising that implementation is much more difficult than PCA, in which estimation is much simpler, since it just involves finding a linear transformation of components which are uncorrelated. Moreover, there is no natural ordering of latent factors in ICA. This is perhaps a blessing in disguise. Namely, as stated above, there is no a priori reason why the ordinal (correlation) ranking of diffusion indexes corresponds to a ranking of their usefulness for predicting y_t (see Kim and Swanson (2014), Bai and Ng (2008) and Carrasco and Rossi (2016) for further discussion of this issue).

Even given all of the recent progress in the area, much remains to be done. There are innumerable possible estimators and algorithms than can potentially be utilized for machine learning (indeed we have touched in our discussion on only a very few of those already available). What will probably differentiate the “good methods” from the “not so good” is their ability to properly marry the latest tools in statistical inference with the latest algorithmic techniques. For example, step-wise methods now often rely on learning functions and thresholding variables (such as t-statistics) centered around conditional mean type prediction, while there is a clearly a need to fully incorporate conditional or predictive density type prediction in new methods. As another example, recall our earlier discussion on the use of asymptotic analysis to examine the combination of conventional out-of-sample schemes with bootstrap aggregation. Many of these sorts of analyses remain to be done in the context of combining conventional forecasting approaches with state of the art dimension reduction, machine learning, and penalized regression algorithms.

2.3 Forecast Evaluation

One of the reasons why machine learning has taken so long to “catch on” in economics is the problem of over-fitting. This issue is made very clear by considering the case of neural networks. We know, from Hornik, Stinchcombe, and White (1989) that neural networks are universal approximators, in the sense that properly designed neural networks with numbers of parameters that grow appropriately, as the sample grows, can approximate an arbitrary function arbitrarily well. However, we also know, from numerous empirical experiments, that more heavily parameterized models often tend to be outperformed, in a predictive sense, by more parsimonious models. The reasons for this are many, and include the effect of specifying models that are crude approximations of reality, and the fact that structural change is prevalent in time series models. Loosely speaking, then, it was the poor predictive accuracy of models that have been too heavily parameterized, or over-fitted, that led economists to eschew adopting machine learning and related big data methods. This is all changing, though, in part because a plethora of new tests for assessing predictive accuracy which account for over-fitting, have recently been developed. However, just as is the case in machine learning, much remains to be done in the area of predictive accuracy testing.

We begin this section by discussing standard predictive accuracy tests that are used every day by applied practitioners. Thereafter, we discuss novel new tests currently being developed that allow for model forecast comparison without specification of a loss function.

2.3.1 Loss Function Dependent Model Evaluation and Selection

As previously, assume that the objective is to predict y_t.The null hypothesis of equal predictive accuracy between two models of y_t, say model 0 and model 1, is specified as:

$$H_0 : E(L(u_{0,t+h}) - L(u_{1,t+h})) = 0$$

and

$$H_A : E(L(u_{0,t+h}) - L(u_{1,t+h})) \neq 0,$$

where $L(\cdot)$ is a loss function. In practice, we do not observe $u_{0,t+h}$ and $u_{1,t+h}$, which are assumed to be out-of-sample h-step ahead forecast errors, but only estimates thereof (i.e., say $\widehat{u}_{0,t+h}$ and $\widehat{u}_{1,t+h}$, respectively). When $P/R \to \pi = 0$, as $P, R \to \infty$ (asymptotically negligible parameter estimation error), where P is the number of forecast errors that we have constructed for each model being compared, and R is the initial "in-sample" estimation period (i.e., $P + R = T$), under recursive or rolling estimation, say, then we can construct the standard version of DM predictive accuracy test in order to test H_0. Namely:

$$DM_P = \frac{\overline{d}_t}{\widehat{\sigma}_{\overline{d}_t}} \xrightarrow{d} N(0,1),$$

where

$$\overline{d}_t = \frac{1}{P}\sum_{t=R+1}^{T} d_t,\ d_t = L(\widehat{u}_{0,t+h}) - L(\widehat{u}_{1,t+h}), \text{ and } \widehat{\sigma}_{\overline{d}_t} = \frac{\widehat{\sigma}_{d_t}}{\sqrt{P}}.$$

In the above test, for which a heteroscedasticity and autocorrelation consistent estimator of $\widehat{\sigma}_{d_t}$ is utilized whenever $h > 1$, the assumption that parameter estimation error is asymptotically negligible allows for the use of any loss function, $L(\cdot)$, including one that is non-differentiable. However, if accounting for parameter estimation error, one can consider only differentiable loss functions (see

Corradi and Swanson (2006b) for complete details). Moreover, regardless of loss function, the normal limiting distribution does not obtain if models 0 and 1 are nested; in which case non-standard critical values must be used, as outlined in McCracken (2000) and Clark and McCracken (2001,2013). An alternative test, which does not require correct dynamic specification and/or conditional homoskedasticity, and which is robust to nonnestedness is proposed by Chao, Corradi, and Swanson (2001). The test statistic is:

$$m_P = P^{-1/2} \sum_{t=R+1}^{T} \widehat{u}_{0,t+h} X_t, \tag{2.10}$$

where $\widehat{u}_{0,t+1}$ is the estimated prediction error, and X_t is some (possibly vector values) set of variables that might be useful for predicting our target variable, y_t. Here X_t may include lags. A simple example of where this sort of test is useful involves testing for linear (predictive) Granger causality, where the null and alternative models are (respectively):

$$y_t = \sum_{j=1}^{q} \beta_j y_{t-j} + u_{0,t}$$

and

$$y_t = \sum_{j=1}^{q} \beta_j y_{t-j} + \sum_{j=1}^{k} \alpha_j x_{t-j} + u_{1,t}$$

In this example, the practitioner estimates the null model, constructs (recursive or rolling, say) predictions, and utilizes the prediction errors (i.e., the $\widehat{u}_{0,t+1}$, for forecast horizon $h = 1$) in the construction of the test statistic, m_P, where P denotes the number of prediction errors. A key advantage of using this test is that models may be nested, thus avoiding issues associated with the testing of nested models that arise when implementing DM_P type tests.

More complex versions of this test that are consistent against generic nonlinear (Granger causal) alternatives are discussed in Corradi and Swanson (2002). In this test, the hypotheses of interest are:

$$\begin{aligned} \widetilde{H}_0 &: \quad E(u_{0,t+h} X_{t-j}) = 0, \;\; j = 0, 1, \ldots k. \\ \widetilde{H}_A &: \quad E(u_{0,t+h} X_{t-j}) \neq 0 \text{ for some } j, \quad j = 0, 1, \ldots k. \end{aligned}$$

As an example, note that if the model being tested does not include a variable, say Z_t, then inclusion of Z_t in X_t is equivalent to testing for out-of-sample Granger causality from Z_t to y_t. Notice also that this test is a variety of the well known Bierens specification test, rather than a test which directly compares two models, such as the DM test. When $P/R \to \pi = 0$, as $P, R \to \infty$, then $m_p' \widehat{S}_{11} m_P \xrightarrow{d} \chi^2_k$, where k is the number of new variables in X_t, and $\widehat{S}_{11}$ is an estimator of a $k \times k$ matrix S_{11}, with:

$$S_{11} = \sum_{j=-\infty}^{\infty} E\left((X_t u_{0,t+h} - \mu_1)(X_{t-j} u_{0,t+h-j} - \mu_1)' \right),$$

where $\mu_1 = E(X_t u_{t+h})$. In empirical applications, one estimates S_{11} as follows:

$$\begin{aligned}
\widehat{S}_{11} &= \frac{1}{P}\sum_{t=R}^{T-1}(\widehat{u}_{0,t+h}X_t - \widehat{\mu}_1)(\widehat{u}_{0,t+h}X_t - \widehat{\mu}_1)' \\
&\quad + \frac{1}{P}\sum_{t=\tau}^{l_T} w_\tau \sum_{t=R+\tau}^{T-1}(\widehat{u}_{0,t+h}X_t - \widehat{\mu}_1)(\widehat{u}_{0,t+h-\tau}X_{t-\tau} - \widehat{\mu}_1)' \\
&\quad + \frac{1}{P}\sum_{t=\tau}^{l_T} w_\tau \sum_{t=R+\tau}^{T-1}(\widehat{u}_{0,t+h-\tau}X_{t-\tau} - \widehat{\mu}_1)(\widehat{u}_{0,t+h}X_t - \widehat{\mu}_1)',
\end{aligned}$$

where $\widehat{\mu}_1 = \frac{1}{P}\sum_{t=R}^{T-1}\widehat{u}_{0,t+1}X_t$.

Alternatively, when comparing multiple different models, Sullivan, Timmermann and White (1999) and White (2000) proposes using the following test statistic:

$$S_P = \max_{k=1,\ldots,m} S_P(1,k),$$

where

$$S_P(1,k) = \frac{1}{\sqrt{P}}\sum_{t=R+1}^{T}\left(L(\widehat{u}_{0,t+h}) - L(\widehat{u}_{k,t+1})\right),\ k = 1,...,m.$$

The hypotheses are formulated as

$$H_0 : \max_{k=1,\ldots,m} E(L(u_{0,t+1}) - L(u_{k,t+1})) \leq 0.$$

$$H_A : \max_{k=1,\ldots,m} E(L(u_{0,t+1}) - L(u_{k,t+1})) > 0.$$

Thus, under the null hypothesis, no competitor model, amongst the set of the m alternatives, can provide a more (loss function specific) accurate prediction than the benchmark model (i.e., model 0). On the other hand, under the alternative, at least one competitor (and in particular, the best competitor) provides more accurate predictions than the benchmark. Critical values for this test can be constructed using the block bootstrap, as discussed in Corradi and Swanson (2007). An interesting extension of this test, in which rolling data windows are used in model estimation and all estimated parameters are conditioned on, is discussed in Giacomini and White (2006). For extensions of the above tests to predictive density evaluation, see Corradi and Swanson (2005,2006a,b).

2.3.2 Loss Function Free Model Evaluation and Selection

In this section we summarize new developments in forecast evaluation which is valid under generalized loss functions, and which is based directly on the evaluation of $F(u)$, the CDF of the forecast error. In particular, note that Corradi, Jin, and Swanson (2017) discuss testing for GL and CL forecast superiority. Their tests allow for parameter estimation error, data dependence, and comparison of multiple models, but require the underlying processes to be strictly stationary. To start, assume that the loss function (L) is defined such that $L : \mathbb{R} \to \mathbb{R}^+$ is continuously differentiable, except for finitely many points, with derivative L', such that $L'(z) \leq 0$, for all $z \leq 0$, and $L'(z) \geq 0$, for all $z \geq 0$.

Definition (Forecast Superiority): u_1 General-Loss (GL) outperforms u_2, denoted as $u_1 \succeq_G u_2$, if and only if $E(L(u_1)) \leq E(L(u_2))$, for all $L \in \mathcal{L}_G$; and u_1 Convex-Loss (CL) outperforms u_2, denoted as $u_1 \succeq_C u_2$, if and only if $E(L(u_1)) \leq E(L(u_2))$, for all $L \in \mathcal{L}_C$.

Here, u_1 and u_2 are sequences of forecast errors, as above. In order to connect the notion of forecast superiority to that of stochastic dominance, CJS establish a mapping between GL forecast superiority and first order stochastic dominance. They also establish linkages between CL forecast superiority and second order stochastic dominance. They then derive direct tests for GL/CL forecast superiority. Define:

$$G(x) = (F_2(x) - F_1(x))sgn(x), \tag{2.11}$$

where $sgn(x) = 1$ if $x \geq 0$, and $= -1$ if $x < 0$; and

$$C(x) = \int_{-\infty}^{x} (F_1(t) - F_2(t))dt1(x < 0) + \int_{x}^{\infty} (F_2(t) - F_1(t))dt1(x \geq 0), \tag{2.12}$$

where $1(\cdot)$ denotes the indicator function, which takes the value 1 if the condition is met, and 0 otherwise. CJS show that $E(L(u_1)) \leq E(L(u_2))$, for all $L \in \mathcal{L}_G$, if and only if $G(x) \leq 0$, for all $x \in \mathcal{X}$, where $\mathcal{X}$ is the union of the supports of all forecast errors; and $E(L(u_1)) \leq E(L(u_2))$, for all $L \in \mathcal{L}_C$, if and only if $C(x) \leq 0$ for all $x \in \mathcal{X}$.

Before implementing GL forecast superiority tests, one can construct a graph that contains a plot of $G(x)$ against x. When $u_1 \succeq_G u_2$, we expect all points to lie below or on the zero line. In other words, a crossing of the zero line in the graph indicates a violation of GL forecast superiority. Similarly, one can construct a graph that contains a plot of $C(x)$ against x. When $u_1 \succeq_C u_2$, we expect all points to lie below or on the zero line. In other words, a crossing of the zero line in the graph indicates a violation of CL forecast superiority.

Now, suppose that there are m sets of forecast errors $u_1, ..., u_m$, resulting from m forecasting models, and that we wish to test the null that $E(L(u_1)) \leq E(L(u_2))$, for all $L \in \mathcal{L}_G$, against the negation thereof (see CJS (2017) for complete details). When testing this null of no forecast superiority, it suffices to construct statistics as follows. For $k = 1, ..., m$, define:

$$\begin{aligned} F_k(x) &= P(u_{k,t} \leq x) \text{ and} \\ \overline{F}_{k,n}(x) &= P^{-1} \sum_{t=R}^{T} 1(u_{k,t} \leq x), \end{aligned}$$

The statistics discussed by CJS (2017) are constructed by calculating:

$$TG_n^+ = \max_{k=2,..,m} \sup_{x\in\mathcal{X}^+} \sqrt{n}G_{k,n}(x) \text{ and } TG_n^- = \max_{k=2,..,m} \sup_{x\in\mathcal{X}^-} \sqrt{n}G_{k,n}(x)$$

and

$$TC_n^+ = \max_{k=2,..,m} \sup_{x\in\mathcal{X}^+} \sqrt{n}C_{k,n}(x) \text{ and } TC_n^- = \max_{k=2,..,m} \sup_{x\in\mathcal{X}^-} \sqrt{n}C_{k,n}(x),$$

where $G_{k,n}(x) \ = \left(\overline{F}_{k,n}\left(x\right) - \overline{F}_{1,n}\left(x\right)\right) sgn(x)$

and

$C_{k,n}(x) = \Big\{\int_{-\infty}^{x} \left(\overline{F}_{1,n}\left(s\right)\right.$
$\left.-\overline{F}_{k,n}\left(s\right)\right) ds1(x<0) + \int_{x}^{\infty} \left(\overline{F}_{k,n}\left(s\right) - \overline{F}_{1,n}\left(s\right)\right) ds1(x \geq 0)\Big\}$.

Note that the positive and negative parts of $\mathcal{X}$ are treated separately in the above statistics. This is because stochastic equicontinuity of the empirical processes cannot be otherwise established, precluding inference based on statistics constructed without separately considering the positive and negative regions of the support.

For discussion of computation of the suprema in these statistics, as well as discussion of more general versions of the test statistics that explicitly account for parameter estimation error and different model estimation schemes (e.g., rolling versus recursive model estimation), see CJS (2017).

Critical values are constructed by using bootstrap methods, as discussed in CJS (2017). Although CJS make a substantial contribution in the nascent loss function robust forecast evaluation, their tests are not uniformly valid, as they have correct asymptotic size only under the least favorable case under the null hypothesis. It remains to develop tests that are uniformly asymptotically valid. Many theoretical questions of this sort remain unanswered in the predictive accuracy and model selection literature, and as new and increasingly complex machine learning methods are developed, theorists will have their hands full keeping up. For a key example of the type of analytically sophisticated analysis that is necessary in order to continue advancing our understanding of model selection, see Hirano and Wright (2017).

2.4 Empirical Illustration: Predicting Interest Rates Using Big Data versus Small Data Methods

In order to fix some of the ideas discussed in this paper, we carry out a small empirical investigation that utilizes a subset of the leading methods discussed above. Our objective is to predict U.S. Treasury yields of various maturities (i.e., the term structure of interest rates). Predictions will be made using "small data" models, including autoregressive, vector autoregressive, and dynamic

Nelson-Siegel models, and "big data" models that utilize diffusion indexes estimated from a largescale macroeconomic dataset.

2.4.1 Experimental Setup

All models in all experiments are re-estimated prior to the construction of each new prediction, using rolling 120 month windows of data; and estimation is carried out using least squares and principal components analysis. Monthly yield forecasts for horizons $h = 1-$, $3-$, and $12-$ steps ahead are constructed for a variety of bond maturities, and these are aggregated using mean square forecast error (MSFE) criteria, and evaluated using the DM_P predictive accuracy test discussed above. The development of a more exhaustive set of experiments is left to future research, and all conclusions made based on our experiments should thus be viewed with caution.

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

Small Data Models

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

(Models in this section are summarized in Table 2.1, and include: AR(1), VAR(1), AR(SIC), and VAR(SIC))

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

$$y_{t+h}(\tau) = c + \beta' W_t + \varepsilon_{t+h}, \tag{2.13}$$

where τ denotes the maturity of a bond (bill) for which the scalar, $y_{t+h}(\tau)$, measures the annual yield. Additionally, W_t contains lags of $y_{t+h}(\tau)$ in autoregressive specifications, and contains lags of $y_{t+h}(\tau)$ and additional explanatory variables in vector autoregressive specifications, with β a conformably defined coefficient vector.[5] In AR and VAR specifications, up to 5 lags of $y_{t+h}(\tau)$ are included in our models, 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 bonds of five different maturities (i.e. 1 year, 2 years, 3 years, 5 years, 10 years).

Dynamic Nelson Siegel (DNS) Models:

(Models in this section are summarized in Table 2.1, and include: DNS(1), DNS(2), DNS(3), DNS(4), DNS(5), and DNS(6))

[5] When specifying VAR models, equation (2.13) 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 SUR (seemingly unrelated regression) result ensures that consistent and efficient parameter estimates can be obtained via application of equation by equation least squares.

The DNS model introduced by Li and Diebold (2006) is a dynamic version of the term structure based upon Nelson and Siegel (1987), where the cross-sectional movement of the term structure model is summarized by the dynamics of three underlying latent factors interpreted as "level", "slope", and "curvature" factors. We refer to the three latent factors as "Nelson-Siegel factors", and in our prediction experiments, both AR(1) and VAR(1) DNS type models are specified in order to predict these factors for subsequent use in the prediction of $y_{t+h}(\tau)$. For a detailed discussion of yield curve modeling using the DNS models, see Diebold and Rudebusch (2013). For detailed discussions comparing arbitrage free dynamic latent factor models, arbitrage free DNS models, and DNS models, refer to Ang and Piazzesi (2003), Diebold, Rudebusch and Aruoba (2006), Christensen, Diebold, and Rudebusch (2011), Duffie (2011), and the references cited therein. For a discussion of the usefulness of survey information in related term structure modeling, see Altavilla, Giacomini, and Ragusa (2016).

In the DNS model, estimates of the Nelson-Siegel factors are constructed at each point in time by regressing $\{1, [\frac{1-\exp(-\lambda_t\tau)}{\lambda_t\tau}], [\frac{1-\exp(-\lambda_t\tau)}{\lambda_t\tau} - \exp(-\lambda_t\tau)]\}$ on $\mathbf{y}_t(\tau)$, where λ_t is a decay parameter (see below discussion). Namely, in a first step, the DNS model

$$\mathbf{y}_t(\tau) = \beta_{1,t} + \beta_{2,t}[\frac{1-\exp(-\lambda_t\tau)}{\lambda_t\tau}] + \beta_{3,t}[\frac{1-\exp(-\lambda_t\tau)}{\lambda_t\tau} - \exp(-\lambda_t\tau)] + \varepsilon_t, \tag{2.14}$$

is fitted at each point in time, t, yielding sequences of estimates, $\widehat{\beta}_{1,t}$, $\widehat{\beta}_{2,t}$, and $\widehat{\beta}_{3,t}$, for $t = 1, ..., T$. Note that in this step, 3 model variants are considered. One variant defines:

$$\mathbf{y}_t^{10}(\tau) = [y_t(12)\ \ y_t(24)\ \ y_t(36)\ \ y_t(48)\ \ y_t(60)\ \ y_t(72)\ \ y_t(84)\ \ y_t(96)\ \ y_t(108)\ \ y_t(120)]'.$$

In a second variant,

$$\mathbf{y}_t^{6}(\tau) = [y_t(12)\ \ y_t(24)\ \ y_t(36)\ \ y_t(60)\ \ y_t(84)\ \ y_t(120)]',$$

and in a third variant

$$\mathbf{y}_t^{4}(\tau) = [y_t(12)\ \ y_t(36)\ \ y_t(60) y_t(120)]'.$$

Predictions of y_{t+h} are constructed using the model:

$$y_{t+h}(\tau) = \widehat{\beta}^f_{1,t+h} + \widehat{\beta}^f_{2,t+h}[\frac{1-\exp(-\lambda_t\tau)}{\lambda_t\tau}] + \widehat{\beta}^f_{3,t+h}[\frac{1-\exp(-\lambda_t\tau)}{\lambda_t\tau} - \exp(-\lambda_t\tau)], \tag{2.15}$$

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

$$\hat{\beta}^f_{i,t+h} = \widehat{c}_i + \widehat{\gamma}_{ii}\widehat{\beta}_{i,t}, \quad \text{for } i = 1, 2, 3, \tag{2.16}$$

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

$$\hat{\beta}^f_{t+h} = \widehat{c} + \widehat{\gamma}\widehat{\beta}_t, \tag{2.17}$$

where $\hat{\beta}^f_{t+h} = \left(\widehat{\beta}^f_{1,t+h}, \widehat{\beta}^f_{2,t+h}, \widehat{\beta}^f_{3,t+h}\right)'$, $\widehat{c}$ is 3x1 vector, and $\widehat{\gamma} = (\widehat{\gamma}_1, \widehat{\gamma}_2, \widehat{\gamma}_3)$, with $\widehat{\gamma}_j$ a 3x1 vector, for $j = 1, 2, 3$. Note that the loading on $\hat{\beta}_{1,t}$ is one, so it is often interpreted as the "level" factor. Also, $\hat{\beta}_{2,t}$ decreases as maturity increases, resulting in an increase in the "slope" of bond yield curve. Finally, $\hat{\beta}_{3,t}$ has initial loading zero, on the short end of yield curve, and reaches its peak at around the 30 month maturity (when the rate of decay, λ_t, is fixed to 0.0609, as discussed by Diebold and Li (2006)), and gradually decays to zero as the maturity goes to infinity. We set the decay parameter equal to 0.0609. Since an increase in $\hat{\beta}_{3t}$ has a larger effect on medium-term yields than on short- and long-term yields, it is often called a "curvature" factor.

DNS Models with Macroeconomic Variables:

(Models in this section are summarized in Table 2.1, and include: DNS(1)+MAC, DNS(2)+MAC, DNS(3)+MAC, DNS(4)+MAC, DNS(5)+MAC, and DNS(6)+MAC)

DNS models of the variety discussed above are also estimated, where latent factor prediction models include macroeconomic variables. Namely, we consider predictions constructed using:

$$\hat{\beta}^f_{i,t+h} = \widehat{c}_i + \widehat{\gamma}_{ii}\widehat{\beta}_{i,t} + \widehat{\alpha}_i' M_t, \quad \text{for} \quad i = 1, 2, 3,$$

where M_t includes selected key macroeconomic variables discussed in Diebold and Li (2006), and $\widehat{\alpha}$ is a 3x1 vector. Here, M_t includes manufacturing capacity utilization, the federal funds rate, and the annual personal consumption expenditures price deflator. Analogous to the VAR(1) model given in (2.17), we additionally construct predictions according to:

$$\hat{\beta}^f_{t+h} = \widehat{c} + \widehat{\gamma}\widehat{\beta}_{i,t} + \widehat{\alpha} M_t, \quad \text{for} \quad i = 1, 2, 3,$$

where $\widehat{\alpha} = (\widehat{\alpha}_1, \widehat{\alpha}_2, \widehat{\alpha}_3)$, with $\widehat{\alpha}_j$ a 3x1 vector, for $i = 1, 2, 3$.

Diffusion Index Models:

(Models in this section are summarized in Table 2.1, and include: DIF(1), DIF(2), DIF(3))

We construct predictions using the diffusion index model discussed extensively above, where latent factors, F^s_t are estimated using PCA with a set of 10 yields given by $\mathbf{y}^{10}_t(\tau)$,

$$y_{t+h}(\tau) = c + \beta' W_t + \alpha' F^s_t + \varepsilon_{t+h}, \tag{2.18}$$

where F^s_t includes either 1, 2, or 3 latent factors corresponding to the largest eigenvalues of the eigenvalue/eigenvector decomposition of a small (standardized) yield dataset consisting of our 10-dimensional yield dataset, and W_t includes only one lag of the yield. This simple model is included in order to facilitate direct comparison with the DNS models given in equations (2.16) and (2.17).

Big Data Models

Diffusion Index Models:

(Models in this section are summarized in Table 2.1, include: DIF(4), DIF(5), DIF(6), VAR(1)+FB1, VAR(1)+FB2, VAR(SIC)+FB1, VAR(SIC)+FB2, DIF(1)+FB1, DIF(2)+FB1, DIF(3)+FB1, DIF(1)+FB2, DIF(2)+FB2, DIF(3)+FB2)

We utilize the prediction model given in equation (2.18), but with latent factors, say F_t^b, estimated using PCA with a set of 103 macroeconomic variables (see below data description for a discussion of the variables used). In particular, we estimate variants of the following factor augmented forecasting model:

$$y_{t+h}(\tau) = c + \beta' W_t + \alpha' F_t^b + \varepsilon_{t+h},$$

where setting $\beta = 0$ yields "pure" diffusion index models, and W_t is defined as above, yielding AR and VAR variants of these models. Inclusion of the lagged yield in W_t allows for direct comparison of our diffusion index models with our pure econometric AR and VAR models discussed at the beginning of this section. Here, F_t^b includes either 1 or 2 latent factors, and α and β are conformably defined vectors of coefficients. For a related discussion of so-called unspanned macroeconomic factors in the yield curve, see Bauer and de los Rios (2012) and Coroneo, Giannone and Modugno (2016).

Additionally, we construct predictions using diffusion index models of the following variety:

$$y_{t+h}(\tau) = c + \beta' W_t + \alpha_2' F_t^b + \alpha_2' F_t^s + \varepsilon_{t+h}.$$

Note that although multiple yield lags were tried when specifying W_t, "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.

DNS Models with Diffusion Indexes:

(Models in this section are summarized in Table 2.1, 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)

The DNS model discussed above is augmented to include diffusion indexes. Namely, we considered DNS type predictions constructed using:

$$\hat{\beta}^f_{i,t+h} = \widehat{c}_i + \widehat{\gamma}_i \widehat{\beta}_{i,t} + \widehat{\alpha}' F_t^b, \quad \text{for} \quad i = 1, 2, 3,$$

where F_t^b again includes either 1, 2 or 3 latent factors, and so is a scalar or a 3x1 vector. All other terms are conformably defined. Analogous to our above discussion of DNS models, we also construct predictions by using the following VAR(1) variant of this model:

$$\hat{\beta}^f_{t+h} = \widehat{c} + \widehat{\Gamma}\widehat{\beta}_t + \widehat{\Xi}F^b_t,$$

where $\hat{\beta}^f_{t+h} = \left(\widehat{\beta}^f_{1,t+h}, \widehat{\beta}^f_{2,t+h}, \widehat{\beta}^f_{3,t+h}\right)'$, $\widehat{c}$ is 3x1 vector, and $\widehat{\Gamma} = (\widehat{\gamma}_1, \widehat{\gamma}_2, \widehat{\gamma}_3)$, $\widehat{\gamma}_j$ is a 3x1 vector, for $j = 1, 2, 3$, and $\widehat{\Xi}$ is a 3x1 vector (if F^b_t is a scalar), or is a 3x2 matrix (if F^b_t is a 2x1 vector).

Forecast Combination

In our prediction experiments, we also construct and analyze a select set of forecast combinations. The particular combinations are detailed in Table 7. Although the focus of this paper is not forecast combination, there are two reasons why we include at least a small set of combinations. First, it is well known that forecast combination is useful in time series prediction. More importantly, inclusion of combinations in our empirical illustration serves to stress that an important area for future research involves combination of classical econometric and machine learning methods. Just as shown in Kim and Swanson (2014), Carrasco and Rossi (2016), and Hirano and Wright (2017), much can be gained via combination not only of forecasts, but also of methodologies.[6]

2.4.2 Data

Our term structure data are 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 Gurkaynak, Sack and Wright (2006)). In particular, we utilize monthly data for the period January 1982 through July 2016, for 1 through 10 year maturities. Hence, we analyze a panel of dataset containing $N = 10$ variables and $T = 415$ monthly observations. All yields are standardized to mean zero unit variance series before principle component analysis.

Macro factors are constructed using a balanced panel of 103 macroeconomic variables obtained from the FRED-MD dataset recently developed by the Federal Reserve Bank of St. Louis. A detailed explanation on how the data set is collected and adjusted is given in McCracken and Ng (2016). FRED-MD is maintained by FRED, is updated on a monthly basis, and can be accessed at

https://research.stlouisfed.org/econ/mccracken/fred-databases/. Our version of this dataset contains observations for the period January 1982 through July 2016.

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

2.4.3 Empirical Findings

Tables 2.2A - 2.2D contain relative MSFEs for yield forecasts constructed using the models listed in Table 2.1, for $h = 1$, for 1, 2, 3, 5, and 10 year maturities, and for 4 different forecasting periods, including: 1992:3-1999:12 (Subsample 1), 2000:1-2007:12 (Subsample 2), 2008:1-2016:7 (Subsample 3), and 1992:1-2016:7 (Subsample 4). The benchmark model used in the construction of relative MSFEs is the AR(1) forecasting model. Tabulated entries denoted in bold are the lowest (relative) point-MSFEs, for each maturity. Starred entries indicate rejection of the (DM_P test) null hypothesis of no difference between the benchmark and the alternative model listed in column 1 of the tables, in favor of the alternative model.[7] Tables 2.3A - D and 2.4A - D collect analogous results, but for $h = 3$ and $h = 12$, respectively. Additionally, the "MSFE-best" models for each bond maturity, each forecast horizon, and each subsample (i.e., the models denoted in bold in Tables 2.2A - 2.4D) are given in Table 2.5; and Table 2.6 is an analogous table, but with two alternative subsamples (i.e., expansionary and recessionary periods). Finally, the results of forecast combination experiments utilizing all of the models are summarized in Tables 2.7 and 2.8A - C.

Turning to the results based on Tables 2.2A through 2.4D, a number of clearcut conclusions emerge.

First, inspection of the results in Tables 2.2A - 2.2D indicates that for Subsamples 1 and 2, the MSFE-best model is usually a DNS model with added "big data" diffusion indexes. Namely, DNS+FB models usually "win". In particular, for forecast horizons of 1- and 3-steps ahead, this is true in 17 of 20 maturity/horizon permutations, across Subsamples 1 and 2. Interestingly, in the most recent subsample (i.e., Subsample 3), DNS+FB type models instead "win" in only 2 of 10 cases, for forecast horizons of 1- and 3-steps ahead. Thus, the post Great-recession period appears to have "confused" our models. Nevertheless, when results based on the entire prediction period (i.e., Subsample 4) are examined, it is noteworthy that DNS models with added "big data" diffusion indexes still "win" in 7 of 10 cases, for $h = 1$ and 3. For our longest forecast horizon (i.e., $h = 12$), the evidence in favor of using "big data" is not so clearcut, as baseline DNS models without diffusion indexes and straw-man AR and VAR models almost always "win".

Second, even cursory examination of Tables 2.2A - 2.4D indicates that models listed as MSFE-best in Table 2.5 are almost always significantly better than our benchmark AR(1) model, based upon application of the DM_P test.

Third, the DNS type models that "win" in our experiments are usually the vector variety (i.e., DNS(4), DNS(5) and DNS(6)). This suggests that the factors in the DNS model do not evolve

[7] *** entries indicate rejection at the 1% level, while ** and * denote rejection at the 5% and 10% levels, respectively.

independently of one another. Thus, not only can the factors (i.e., the "betas") be better predicted by utilizing "big data" diffusion indexes, as discussed above, but they can also be better predicted by modeling their cross-correlation dynamics.

We now turn to a discussion of the results in Tables 2.5 - 2.8.

In Table 2.5, where point "MSFE-best" models are listed by subsample and maturity, a number of further conclusions emerge. In this table, entries superscripted with ***, **, and * in Table 2.5 denote rejections of the null hypothesis of equal predictive accuracy at 0.01, 0.05, and 0.10 significance levels, respectively, based on application of the Diebold-Mariano test discussed in Section 3; and indicate that the listed model is predictively superior to a "benchmark" DNS(τ) model, based on MSFE loss. In particular, if the point "MSFE-best" model is DNS(τ)+*mod*, where *mod* denotes another component of the model (for example, *mod* may be FB1 or FB2, etc.) then the "benchmark" model is DNS(τ). If the point "MSFE-best" model is DNS(1), or if no DNS component appears in point "MSFE-best" model, then DNS(1) is the "benchmark" model. Finally, for entries denoted "DNS(1)", no predictive accuracy test was carried out. These test results are included to highlight the importance of incorporating "big data" in DNS type prediction models. Turning to the results of these tests, note that for forecast horizons of 1- and 3-steps ahead, DNS(τ)+FB models significantly outperform their DNS(τ) counterparts in almost all cases, across Subsamples 1 and 2. In Subsample 3 (2008:1-2016:7), the evidence is more mixed, with "less to choose" between the alternative models in our experiments. Additionally, and as discussed above, our "straw-man" models perform well at the 12-step ahead forecast horizon.

Needless to say, there are instances where AR type models outperform our more complex models. The reasons for this may be many. For example, structural breaks may play an important role that is not captured by any of our specifications, leading to cases where the "simplest" approximations (e.g., AR and VAR models) dominate, from the perspective of predictive accuracy. Of course, this does not preclude the possibility that more complex models than ours may outperform (V)AR models in such cases.

Additionally, note that (V)AR models perform better at longer horizons, which is not surprising, and is a well know stylized fact in empirical economics; again probably stemming from issues pertaining to the approximate nature of our models, and the ability of the most parsimonious models to dominate under increased uncertainty, due to model specification and parameter uncertainty issues.

In Table 2.6, we see that the evidence in favor of DNS+FB type models is both stronger and weaker when our prediction periods are broken into two alternative subsamples defined as "expansionary" and "recessionary", based upon application of NBER dating. In particular, in recessionary times, DNS+FB models win in 13 of 15 maturity/horizon permutations, including maturities of 1, 3, 5, and 10 years and horizons of $h = 1, 3$, and 12 months ahead. Thus, in recessionary times our DNS+FB models even "win" for $h = 12$, which was not the case based upon our earlier analysis of Subsamples 1-4. On the other hand, in expansionary times, DNS+FB models win in only 7 of 15 maturity/horizon permutations, and none of these wins occur when $h = 12$.

Finally, Table 2.7 lists a small number of different forecast combinations that were utilized in order to construct alternative prediction models to compare with those discussed above. The "MSFE-best" combination models are usually preferred to the AR(1) benchmark, based on application of the DM_P test, as might be expected, given our above discussion. However, it is noteworthy that point MSFEs associated with the best combination models are usually higher than point MSFEs associated with out best individual models. Indeed, combination models fail to "win" in 15 of 20 cases, for $h = 1$, Subsamples 1-4, and across all 5 bond maturities (see Table 2.8A). For $h = 3$, the case against forecast combination is even stronger, with combination models failing to "win" in 18 of 20 cases, for Subsamples 1-4 and across all 5 bond maturities (see Table 2.8B). Similarly, for $h = 12$, combination models fail to "win" in 17 of 20 cases (see Table 2.8C). Evidently, a richer set of combination models needs to be entertained if the usual result that combination works is to be found. Examination of this is left to future research.

2.5 Concluding Remarks

This paper discusses recent advances in the analysis of big data using latent factor type dimension reduction methods as well as various other machine learning and shrinkage approaches. It is suggested that much remains to be learned regarding the ways in which extant econometric methods can be combined with dimension reduction methods in order to achieve improvements in prediction. We show how readily standard econometric models can be augmented to include predictive error reducing information from big datasets, in an illustration in which the term structure of interest rates is predicted. Finally, we address predictive accuracy testing in the context of big data, and outline new loss function free methods that may be useful for forecast accuracy and model selection assessment.

Table 2.1: Models Used in Forecast Experiments*

Model	Description
AR(1)	Autoregressive model with one lag
VAR(1)	Five-dimensional vector autoregressive model with one lag
VAR(1)+FB1	VAR(1) model with one principle component added, principle component analysis based on all 103 macroeconomic variables
VAR(1)+FB2	VAR(1) model with two principle components added, principle component analysis based on all 103 macroeconomic variables
AR(SIC)	Autoregressive model with lag(s) selected by the Schwarz information criterion
VAR(SIC)	Five-dimensional vector autoregressive model with lag(s) selected by the Schwarz information criterion
VAR(SIC)+FB1	VAR(SIC) model with one principle component added, principle component analysis based on all 103 macroeconomic variables
VAR(SIC)+FB2	VAR(SIC) model with two principle components added, principle component analysis based on all 103 macroeconomic variables
DNS(1)	Dynamic Nelson-Siegel (DNS) model with underlying AR(1) factor specifications fitted with ten-dimensional yields:
	maturity $\tau = 12, 24, 36, 48, 60, 72, 84, 96, 108, 120$ months
DNS(2)	DNS model with underlying AR(1) factor specifications fitted with six-dimensional yields: maturity $\tau = 12, 24, 36, 60, 84, 120$ months
DNS(3)	DNS model with underlying AR(1) factor specifications fitted with four-dimensional yields: maturity $\tau = 12, 36, 60, 120$ months
DNS(4)	DNS model with underlying VAR(1) factor specifications fitted with ten-dimensional yields: maturity $\tau = 12, 24, 36, 48, 60, 72, 84, 96, 108, 120$ months
DNS(5)	DNS model with underlying VAR(1) factor specifications fitted with six-dimensional yields: maturity $\tau = 12, 24, 36, 60, 84, 120$ months
DNS(6)	DNS model with underlying VAR(1) factor specifications fitted with four-dimensional yields: maturity $\tau = 12, 36, 60, 120$ months
DNS(1)+FB1	DNS(1) model with one principle component added, principle component analysis based on all 103 macroeconomic variables
DNS(2)+FB1	DNS(2) model with one principle component added, principle component analysis based on all 103 macroeconomic variables
DNS(3)+FB1	DNS(3) model with one principle component added, principle component analysis based on all 103 macroeconomic variables
DNS(4)+FB1	DNS(4) model with one principle component added, principle component analysis based on all 103 macroeconomic variables
DNS(5)+FB1	DNS(5) model with one principle component added, principle component analysis based on all 103 macroeconomic variables
DNS(6)+FB1	DNS(6) model with one principle component added, principle component analysis based on all 103 macroeconomic variables
DNS(1)+FB2	DNS(1) model with two principle components added, principle component analysis based on all 103 macroeconomic variables
DNS(2)+FB2	DNS(2) model with two principle components added, principle component analysis based on all 103 macroeconomic variables
DNS(3)+FB2	DNS(3) model with two principle components added, principle component analysis based on all 103 macroeconomic variables
DNS(4)+FB2	DNS(4) model with two principle components added, principle component analysis based on all 103 macroeconomic variables
DNS(5)+FB2	DNS(5) model with two principle components added, principle component analysis based on all 103 macroeconomic variables
DNS(6)+FB2	DNS(6) model with two principle components added, principle component analysis based on all 103 macroeconomic variables
DNS(1)+MAC	DNS(1) model with three key macroeconomic variables added: manufacturing capacity utilization, the federal funds rate, and annual price inflation
DNS(2)+MAC	DNS(2) model with three key macroeconomic variables added: manufacturing capacity utilization, the federal funds rate, and annual price inflation
DNS(3)+MAC	DNS(3) model with three key macroeconomic variables added: manufacturing capacity utilization, the federal funds rate, and annual price inflation
DNS(4)+MAC	DNS(4) model with three key macroeconomic variables added: manufacturing capacity utilization, the federal funds rate, and annual price inflation
DNS(5)+MAC	DNS(5) model with three key macroeconomic variables added: manufacturing capacity utilization, the federal funds rate, and annual price inflation
DNS(6)+MAC	DNS(6) model with three key macroeconomic variables added: manufacturing capacity utilization, the federal funds rate, and annual price inflation
DIF(1)	Diffusion index model with one principle component estimator based on all ten-dimensional yields
DIF(2)	Diffusion index model with two principle component estimators based on all ten-dimensional yields
DIF(3)	Diffusion index model with three principle component estimators based on all ten-dimensional yields
DIF(4)	Diffusion index model with one principle component estimator based on all 103 macroeconomic variables
DIF(5)	Diffusion index model with two principle component estimators based on all 103 macroeconomic variables
DIF(6)	Diffusion index model with three principle component estimators based on all 103 macroeconomic variables
DIF(1)+FB1	DIF(1) model with one principle component added, principle component analysis based on all 103 macroeconomic variables
DIF(2)+FB1	DIF(2) model with one principle component added, principle component analysis based on all 103 macroeconomic variables
DIF(3)+FB1	DIF(3) model with one principle component added, principle component analysis based on all 103 macroeconomic variables
DIF(1)+FB2	DIF(1) model with two principle components added, principle component analysis based on all 103 macroeconomic variables
DIF(2)+FB2	DIF(2) model with two principle components added, principle component analysis based on all 103 macroeconomic variables
DIF(3)+FB2	DIF(3) model with two principle components added, principle component analysis based on all 103 macroeconomic variables

* Notes: This table summarizes the models utilized in all forecasting experiments.

Table 2.2A: 1-Step-Ahead Relative MSFEs of All Forecasting Models (Subsample 1: 1992:3-1999:12)*

Model	rMSFE				
Maturity	1 year	2 year	3 years	5 years	10 years
AR(1)	1.000	1.000	1.000	1.000	1.000
VAR(1)	1.099	1.108	1.103	1.098	1.141
VAR(1)+FB1	0.819**	0.868*	0.893*	0.927	1.045
VAR(1)+FB2	0.844	0.874	0.897	0.940	1.106
AR(SIC)	0.864**	0.942*	0.958	0.974	**0.972****
VAR(SIC)	1.099	1.108	1.103	1.098	1.141
VAR(SIC)+FB1	0.819**	0.868*	0.893*	0.927	1.045
VAR(SIC)+FB2	0.844	0.874	0.897	0.940	1.106
DNS(1)	1.032	1.097	1.061	1.039	1.067
DNS(2)	1.036	1.088	1.053	1.046	1.064
DNS(3)	1.040	1.123	1.066	1.045	1.037
DNS(4)	1.088	1.160	1.104	1.070	1.102
DNS(5)	1.095	1.147	1.095	1.081	1.098
DNS(6)	1.094	1.190	1.107	1.065	1.071
DNS(1)+FB1	0.900	0.862*	0.895	0.981	0.981
DNS(2)+FB1	0.891	0.865*	0.903	1.000	0.980
DNS(3)+FB1	0.876	0.868*	0.896	1.006	0.990
DNS(4)+FB1	0.784**	0.861**	0.870**	0.922	0.990
DNS(5)+FB1	0.785**	0.854**	0.867**	0.934	0.987
DNS(6)+FB1	0.775***	0.882**	0.872**	0.930	0.985
DNS(1)+FB2	0.960	0.908	0.948	1.053	1.053
DNS(2)+FB2	0.948	0.911	0.957	1.074	1.051
DNS(3)+FB2	0.933	0.911	0.948	1.081	1.073
DNS(4)+FB2	0.789**	0.844**	0.858**	0.920	0.988
DNS(5)+FB2	0.790**	**0.840****	**0.857****	0.934	0.985
DNS(6)+FB2	**0.775****	0.863**	0.860**	0.929	0.987
DNS(1)+MAC	1.028	1.099	1.073	1.056	1.095
DNS(2)+MAC	1.029	1.089	1.065	1.063	1.091
DNS(3)+MAC	1.032	1.123	1.079	1.062	1.063
DNS(4)+MAC	1.132	1.147	1.129	1.154	1.191
DNS(5)+MAC	1.130	1.140	1.125	1.164	1.184
DNS(6)+MAC	1.119	1.165	1.130	1.161	1.188
DIF(1)	3.048	2.655	1.926	**0.919****	2.245
DIF(2)	1.274	1.067	1.038	1.029	1.199
DIF(3)	0.973	1.046	1.044	1.049	1.128
DIF(4)	2.238	2.303	2.337	2.382	2.438
DIF(5)	2.253	2.338	2.386	2.455	2.588
DIF(6)	2.236	2.320	2.359	2.410	2.514
DIF(1)+FB1	2.208	2.182	1.717	0.950	2.239
DIF(2)+FB1	1.340	1.074	1.026	1.039	1.254
DIF(3)+FB1	0.958	1.006	1.021	1.060	1.164
DIF(1)+FB2	2.002	1.933	1.489	0.969	2.065
DIF(2)+FB2	1.269	1.052	1.016	1.029	1.247
DIF(3)+FB2	0.947	1.007	1.022	1.057	1.177

* Notes: Table 2.2A reports the mean squared forecast error (MSFE) relative to that from the benchmark AR(1) model based on 1-step-ahead forecasts of monthly U.S. Treasury bond yields of various maturities. The models, as listed in column 1, are summarized in Table 2.1. Entries in bold denote models with lowest mean square forecast error (MSFE) for a given bond maturity. Entries superscripted with ***, **, and * denote rejections of the null of equal predictive accuracy at 0.01, 0.05, and 0.10 significance levels, respectively, based on application of the Diebold-Mariano test discussed in Section 3; and indicate that the listed model is predictively superior to the AR(1) benchmark, based on MSFE loss. For complete details, refer to Section 4.

Table 2.2B: 1-Step-Ahead Relative MSFEs of All Forecasting Models (Subsample 2: 2000:1-2007:12) *

Model	rMSFE				
Maturity	1 year	2 year	3 years	5 years	10 years
AR(1)	1.000	1.000	1.000	1.000	1.000
VAR(1)	0.970	1.029	1.032	1.045	1.110
VAR(1)+FB1	0.733^{**}	0.858^{*}	0.906	0.971	1.084
VAR(1)+FB2	0.810	0.899	0.936	1.003	1.157
AR(SIC)	0.939	1.033	1.033	1.035	1.015
VAR(SIC)	0.970	1.029	1.032	1.045	1.110
VAR(SIC)+FB1	0.733^{**}	0.858^{*}	0.906	0.971	1.084
VAR(SIC)+FB2	0.810	0.899	0.936	1.003	1.157
DNS(1)	1.211	1.015	1.000	1.094	0.959^{*}
DNS(2)	1.182	1.016	1.012	1.121	0.958^{**}
DNS(3)	1.150	1.015	0.998	1.126	0.983
DNS(4)	1.017	1.067	1.031	1.082	1.026
DNS(5)	1.014	1.058	1.034	1.110	1.027
DNS(6)	1.021	1.099	1.037	1.099	1.032
DNS(1)+FB1	0.780^{*}	0.851	0.860	0.947	0.947
DNS(2)+FB1	0.773^{*}	0.842	0.859^{*}	0.966	0.944
DNS(3)+FB1	0.770^{*}	0.873	0.863	0.966	0.944
DNS(4)+FB1	0.708^{***}	0.853^{**}	0.866^{**}	0.962	0.959
DNS(5)+FB1	0.703^{***}	0.840^{**}	0.865^{**}	0.987	0.960
DNS(6)+FB1	0.713^{***}	0.884^{*}	0.872^{**}	0.979	0.965
DNS(1)+FB2	0.717^{**}	0.741^{**}	$\mathbf{0.763}^{**}$	**0.887**	0.855^{**}
DNS(2)+FB2	0.707^{**}	$\mathbf{0.734}^{**}$	0.766^{**}	0.912	$\mathbf{0.854}^{**}$
DNS(3)+FB2	$\mathbf{0.697}^{**}$	0.756^{**}	0.765^{**}	0.915	0.877^{**}
DNS(4)+FB2	0.727^{***}	0.793^{***}	0.824^{***}	0.961	0.933
DNS(5)+FB2	0.721^{***}	0.791^{***}	0.832^{**}	0.991	0.935
DNS(6)+FB2	0.703^{***}	0.810^{***}	0.824^{***}	0.983	0.960
DNS(1)+MAC	1.065	0.982	1.002	1.099	0.979
DNS(2)+MAC	1.037	0.983	1.011	1.125	0.977
DNS(3)+MAC	1.000	0.983	1.000	1.129	0.997
DNS(4)+MAC	0.972	1.040	1.056	1.165	1.064
DNS(5)+MAC	0.960	1.037	1.065	1.197	1.065
DNS(6)+MAC	0.949	1.057	1.056	1.190	1.097
DIF(1)	2.474	2.046	1.688	1.062	1.788
DIF(2)	1.288	1.112	1.104	1.061	1.214
DIF(3)	1.029	1.128	1.114	1.073	1.121
DIF(4)	1.566	1.733	1.830	1.930	1.961
DIF(5)	1.349	1.688	1.805	1.884	1.937
DIF(6)	1.389	1.697	1.804	1.868	1.919
DIF(1)+FB1	1.575	1.633	1.468	1.045	1.794
DIF(2)+FB1	1.093	1.001	1.027	1.038	1.227
DIF(3)+FB1	0.892	1.021	1.049	1.053	1.115
DIF(1)+FB2	1.435	1.673	1.521	1.039	1.667
DIF(2)+FB2	1.117	1.024	1.039	1.046	1.184
DIF(3)+FB2	0.875	1.023	1.058	1.059	1.122

* Notes: See notes to Table 2.2A.

Table 2.2C: 1-Step-Ahead Relative MSFEs of All Forecasting Models (Subsample 3: 2008:1-2016:7)

*

Model	rMSFE				
Maturity	1 year	2 year	3 years	5 years	10 years
AR(1)	1.000	**1.000**	**1.000**	1.000	1.000
VAR(1)	1.177	1.242	1.222	1.180	1.158
VAR(1)+FB1	1.249	1.311	1.293	1.281	1.297
VAR(1)+FB2	1.369	1.451	1.406	1.345	1.320
AR(SIC)	**0.920**	1.028	1.005	**0.999**	1.009
VAR(SIC)	1.177	1.242	1.222	1.180	1.158
VAR(SIC)+FB1	1.249	1.311	1.293	1.281	1.297
VAR(SIC)+FB2	1.369	1.451	1.406	1.345	1.320
DNS(1)	2.108	1.044	1.137	1.444	0.932*
DNS(2)	1.966	1.085	1.223	1.542	**0.928***
DNS(3)	1.710	1.003	1.098	1.521	0.980
DNS(4)	1.396	1.156	1.141	1.387	1.042
DNS(5)	1.317	1.127	1.167	1.459	1.030
DNS(6)	1.231	1.220	1.137	1.450	1.071
DNS(1)+FB1	2.132	1.536	1.390	1.511	1.078
DNS(2)+FB1	2.030	1.522	1.421	1.583	1.072
DNS(3)+FB1	1.914	1.577	1.379	1.582	1.110
DNS(4)+FB1	1.420	1.315	1.212	1.387	1.132
DNS(5)+FB1	1.373	1.275	1.223	1.452	1.123
DNS(6)+FB1	1.306	1.403	1.213	1.437	1.138
DNS(1)+FB2	2.259	1.661	1.469	1.523	1.075
DNS(2)+FB2	2.149	1.645	1.497	1.591	1.068
DNS(3)+FB2	2.044	1.707	1.462	1.595	1.106
DNS(4)+FB2	1.553	1.467	1.327	1.454	1.177
DNS(5)+FB2	1.503	1.423	1.332	1.513	1.166
DNS(6)+FB2	1.442	1.552	1.325	1.501	1.180
DNS(1)+MAC	1.720	1.051	1.094	1.331	0.943
DNS(2)+MAC	1.604	1.064	1.149	1.413	0.939
DNS(3)+MAC	1.429	1.060	1.078	1.406	0.966
DNS(4)+MAC	1.316	1.137	1.141	1.382	1.056
DNS(5)+MAC	1.228	1.108	1.162	1.447	1.041
DNS(6)+MAC	1.144	1.199	1.132	1.437	1.077
DIF(1)	2.521	2.326	2.076	1.245	1.621
DIF(2)	1.581	1.336	1.315	1.215	1.196
DIF(3)	1.058	1.392	1.388	1.259	1.257
DIF(4)	4.145	3.409	2.884	2.440	2.219
DIF(5)	4.718	3.707	2.960	2.305	1.980
DIF(6)	4.699	3.805	3.151	2.573	2.192
DIF(1)+FB1	4.053	3.207	2.411	1.280	1.575
DIF(2)+FB1	2.127	1.637	1.439	1.270	1.226
DIF(3)+FB1	1.367	1.666	1.532	1.326	1.320
DIF(1)+FB2	4.438	3.232	2.257	1.179	1.452
DIF(2)+FB2	2.100	1.615	1.403	1.182	1.106
DIF(3)+FB2	1.341	1.623	1.448	1.190	1.147

* Notes: See notes to Table 2.2A.

Table 2.2D: 1-Step-Ahead Relative MSFEs of All Forecasting Models (Subsample 4: 1992:3-2016:7)

*

Model	rMSFE				
Maturity	1 year	2 year	3 years	5 years	10 years
AR(1)	1.000	1.000	1.000	**1.000**	1.000
VAR(1)	1.063	1.103	1.101	1.102	1.139
VAR(1)+FB1	0.874*	0.955	0.990	1.046	1.165
VAR(1)+FB2	0.940	1.003	1.029	1.081	1.213
AR(SIC)	0.906*	0.998	0.999	1.004	1.001
VAR(SIC)	1.063	1.103	1.101	1.102	1.139
VAR(SIC)+FB1	0.874*	0.955	0.990	1.046	1.165
VAR(SIC)+FB2	0.940	1.003	1.029	1.081	1.213
DNS(1)	1.331	1.052	1.053	1.177	0.976
DNS(2)	1.291	1.058	1.075	1.218	**0.973**
DNS(3)	1.226	1.053	1.046	1.213	0.996
DNS(4)	1.123	1.120	1.083	1.166	1.053
DNS(5)	1.108	1.106	1.086	1.201	1.047
DNS(6)	1.093	1.158	1.085	1.189	1.059
DNS(1)+FB1	1.109	0.996	0.994	1.122	1.012
DNS(2)+FB1	1.081	0.991	1.003	1.155	1.008
DNS(3)+FB1	1.050	1.016	0.993	1.157	1.027
DNS(4)+FB1	0.886*	0.951	**0.946**	1.071	1.041
DNS(5)+FB1	0.874*	**0.935**	0.947	1.104	1.037
DNS(6)+FB1	**0.861****	0.990	0.950	1.095	1.045
DNS(1)+FB2	1.132	0.993	0.992	1.127	1.001
DNS(2)+FB2	1.100	0.988	1.003	1.162	0.998
DNS(3)+FB2	1.069	1.010	0.991	1.167	1.027
DNS(4)+FB2	0.924	0.951	0.951	1.090	1.052
DNS(5)+FB2	0.911	0.939	0.955	1.123	1.047
DNS(6)+FB2	0.885	0.982	0.951	1.115	1.061
DNS(1)+MAC	1.188	1.040	1.049	1.152	0.995
DNS(2)+MAC	1.153	1.040	1.063	1.187	0.991
DNS(3)+MAC	1.102	1.052	1.047	1.187	1.001
DNS(4)+MAC	1.105	1.101	1.102	1.224	1.094
DNS(5)+MAC	1.081	1.091	1.109	1.258	1.086
DNS(6)+MAC	1.055	1.127	1.100	1.252	1.112
DIF(1)	2.702	2.334	1.863	1.067	1.838
DIF(2)	1.344	1.141	1.128	1.095	1.203
DIF(3)	1.014	1.151	1.151	1.119	1.181
DIF(4)	2.361	2.293	2.256	2.229	2.198
DIF(5)	2.397	2.349	2.281	2.197	2.128
DIF(6)	2.404	2.366	2.314	2.253	2.193
DIF(1)+FB1	2.334	2.165	1.774	1.081	1.818
DIF(2)+FB1	1.403	1.159	1.120	1.105	1.234
DIF(3)+FB1	1.016	1.147	1.149	1.134	1.215
DIF(1)+FB2	2.279	2.092	1.677	1.056	1.681
DIF(2)+FB2	1.381	1.156	1.114	1.080	1.167
DIF(3)+FB2	1.000	1.140	1.134	1.096	1.147

* Notes: See notes to Table 2.2A.

Table 2.3A: 3-Step-Ahead Relative MSFEs of All Forecasting Models (Subsample 1: 1992:7-1999:12) *

Model	rMSFE				
Maturity	1 year	2 year	3 years	5 years	10 years
AR(1)	1.000	1.000	1.000	1.000	1.000
VAR(1)	1.014	1.062	1.074	1.063	1.035
VAR(1)+FB1	0.981	1.037	1.055	1.050	1.030
VAR(1)+FB2	1.070	1.119	1.134	1.130	1.114
AR(SIC)	0.885**	0.963	0.974	0.969	**0.933****
VAR(SIC)	1.014	1.062	1.074	1.063	1.035
VAR(SIC)+FB1	0.981	1.037	1.055	1.050	1.030
VAR(SIC)+FB2	1.070	1.119	1.134	1.130	1.114
DNS(1)	1.061	1.079	1.065	1.047	1.029
DNS(2)	1.067	1.082	1.069	1.054	1.031
DNS(3)	1.065	1.080	1.062	1.046	1.028
DNS(4)	0.990	1.075	1.063	1.024	1.003
DNS(5)	0.996	1.071	1.058	1.024	1.002
DNS(6)	1.000	1.086	1.063	1.017	0.989
DNS(1)+FB1	0.922	0.914	**0.954**	1.013	1.000
DNS(2)+FB1	0.923	0.924	0.964	1.024	1.004
DNS(3)+FB1	0.918	**0.914**	0.955	1.025	1.022
DNS(4)+FB1	**0.855****	0.961	0.973	0.966	0.959
DNS(5)+FB1	0.861**	0.957	0.969	0.966	0.956
DNS(6)+FB1	0.860**	0.968	0.972	**0.961**	0.951
DNS(1)+FB2	1.033	1.019	1.039	1.063	1.008
DNS(2)+FB2	1.032	1.027	1.048	1.074	1.013
DNS(3)+FB2	1.034	1.024	1.045	1.080	1.035
DNS(4)+FB2	0.920*	1.016	1.025	1.014	0.998
DNS(5)+FB2	0.930	1.017	1.025	1.019	0.999
DNS(6)+FB2	0.919*	1.018	1.021	1.008	0.990
DNS(1)+MAC	1.056	1.099	1.091	1.063	1.025
DNS(2)+MAC	1.063	1.102	1.093	1.068	1.028
DNS(3)+MAC	1.060	1.102	1.089	1.061	1.020
DNS(4)+MAC	0.942	1.043	1.052	1.040	1.031
DNS(5)+MAC	0.945	1.037	1.045	1.036	1.026
DNS(6)+MAC	0.950	1.053	1.054	1.038	1.025
DIF(1)	1.678	1.558	1.285	0.987	1.381
DIF(2)	1.272	1.252	1.227	1.207	1.248
DIF(3)	1.209	1.241	1.218	1.181	1.195
DIF(4)	1.217	1.278	1.312	1.355	1.430
DIF(5)	1.465	1.533	1.563	1.568	1.541
DIF(6)	1.489	1.558	1.587	1.587	1.538
DIF(1)+FB1	1.272	1.340	1.196	0.998	1.406
DIF(2)+FB1	1.216	1.199	1.178	1.176	1.240
DIF(3)+FB1	1.050	1.100	1.111	1.124	1.184
DIF(1)+FB2	1.395	1.513	1.388	1.200	1.408
DIF(2)+FB2	1.316	1.305	1.276	1.244	1.266
DIF(3)+FB2	1.140	1.203	1.209	1.199	1.229

* Notes: See notes to Table 2.2A.

Table 2.3B: 3-Step-Ahead Relative MSFEs of All Forecasting Models (Subsample 2: 2000:1-2007:12) *

Model			rMSFE		
Maturity	1 year	2 year	3 years	5 years	10 years
AR(1)	1.000	1.000	1.000	1.000	1.000
VAR(1)	0.832***	0.880***	0.885**	0.909*	1.013
VAR(1)+FB1	0.837***	0.880***	0.882**	0.905**	1.009
VAR(1)+FB2	0.839***	0.889**	0.895**	0.924*	1.043
AR(SIC)	0.819***	0.877**	0.873**	0.881**	0.937
VAR(SIC)	0.832***	0.880***	0.885**	0.909*	1.013
VAR(SIC)+FB1	0.837***	0.880***	0.882**	0.905**	1.009
VAR(SIC)+FB2	0.839***	0.889**	0.895**	0.924*	1.043
DNS(1)	1.250	1.068	1.009	1.051	0.913***
DNS(2)	1.233	1.069	1.019	1.066	0.913***
DNS(3)	1.232	1.063	1.009	1.078	0.977
DNS(4)	0.905***	0.919**	0.895**	0.915**	0.929
DNS(5)	0.900***	0.916**	0.897**	0.926**	0.929
DNS(6)	0.915***	0.930*	0.897**	0.920**	0.942
DNS(1)+FB1	0.676**	0.705**	0.724**	**0.827****	0.846**
DNS(2)+FB1	0.674**	0.706**	0.730**	0.840*	**0.846****
DNS(3)+FB1	**0.672****	**0.703****	**0.720****	0.842*	0.891*
DNS(4)+FB1	0.830***	0.857***	0.846***	0.884**	0.909*
DNS(5)+FB1	0.830***	0.857***	0.851***	0.898**	0.909*
DNS(6)+FB1	0.833***	0.863**	0.845***	0.889**	0.925*
DNS(1)+FB2	0.794*	0.755**	0.773**	0.898	0.921
DNS(2)+FB2	0.784*	0.754**	0.779**	0.912	0.920
DNS(3)+FB2	0.793	0.757**	0.777**	0.929	0.998
DNS(4)+FB2	0.833***	0.860***	0.849***	0.887**	0.915*
DNS(5)+FB2	0.833***	0.860***	0.854***	0.901**	0.916*
DNS(6)+FB2	0.836***	0.866**	0.848***	0.892**	0.931*
DNS(1)+MAC	1.073	1.026	1.026	1.097	0.963*
DNS(2)+MAC	1.055	1.028	1.036	1.111	0.962*
DNS(3)+MAC	1.049	1.018	1.025	1.122	1.020
DNS(4)+MAC	0.853***	0.897*	0.900*	0.952	0.972
DNS(5)+MAC	0.849***	0.896*	0.903*	0.964	0.970
DNS(6)+MAC	0.854***	0.901*	0.898*	0.959	0.996
DIF(1)	1.641	1.460	1.331	1.146	1.286
DIF(2)	1.234	1.217	1.191	1.188	1.354
DIF(3)	1.186	1.280	1.261	1.241	1.340
DIF(4)	0.946	1.084	1.188	1.330	1.358
DIF(5)	0.970	1.143	1.210	1.305	1.488
DIF(6)	1.007	1.182	1.247	1.331	1.511
DIF(1)+FB1	0.943	1.057	1.096	1.196	1.428
DIF(2)+FB1	0.915	1.018	1.082	1.184	1.462
DIF(3)+FB1	0.911	1.081	1.149	1.227	1.413
DIF(1)+FB2	1.090	1.273	1.266	1.189	1.728
DIF(2)+FB2	0.913	1.059	1.124	1.196	1.414
DIF(3)+FB2	0.905	1.098	1.171	1.242	1.432

* Notes: See notes to Table 2.2A.

Table 2.3C: 3-Step-Ahead Relative MSFEs of All Forecasting Models (Subsample 3: 2008:1-2016:7)

*

Model			rMSFE		
Maturity	1 year	2 year	3 years	5 years	10 years
AR(1)	1.000	1.000	1.000	1.000	1.000
VAR(1)	0.975	0.998	0.974	0.927*	0.926*
VAR(1)+FB1	0.937	0.976	0.961	0.924	0.935
VAR(1)+FB2	0.942	0.978	0.963	0.926	0.936
AR(SIC)	0.975	0.971	**0.946****	**0.921****	0.917**
VAR(SIC)	0.975	0.998	0.974	0.927*	0.926*
VAR(SIC)+FB1	0.937	0.976	0.961	0.924	0.935
VAR(SIC)+FB2	0.942	0.978	0.963	0.926	0.936
DNS(1)	1.825	1.278	1.304	1.407	0.971
DNS(2)	1.762	1.310	1.357	1.453	0.968
DNS(3)	1.652	1.203	1.260	1.431	1.022
DNS(4)	1.046	0.974	0.990	1.050	0.907*
DNS(5)	1.025	0.973	1.001	1.071	0.899*
DNS(6)	0.997	0.979	0.987	1.070	0.919*
DNS(1)+FB1	2.186	1.848	1.660	1.545	1.093
DNS(2)+FB1	2.154	1.846	1.675	1.571	1.089
DNS(3)+FB1	2.118	1.823	1.633	1.563	1.136
DNS(4)+FB1	0.957	0.929	0.948	1.014	0.914
DNS(5)+FB1	0.939	**0.927**	0.957	1.035	0.907
DNS(6)+FB1	**0.915**	0.939	0.947	1.032	0.920*
DNS(1)+FB2	2.215	1.871	1.643	1.478	1.041
DNS(2)+FB2	2.184	1.863	1.651	1.499	1.036
DNS(3)+FB2	2.158	1.853	1.619	1.493	1.074
DNS(4)+FB2	0.984	0.948	0.964	1.027	0.923
DNS(5)+FB2	0.966	0.945	0.972	1.047	0.916
DNS(6)+FB2	0.943	0.960	0.964	1.045	0.930
DNS(1)+MAC	1.864	1.393	1.372	1.423	1.006
DNS(2)+MAC	1.811	1.417	1.416	1.467	1.003
DNS(3)+MAC	1.721	1.337	1.339	1.451	1.056
DNS(4)+MAC	1.039	0.977	0.992	1.045	0.903*
DNS(5)+MAC	1.017	0.978	1.004	1.065	**0.895***
DNS(6)+MAC	0.988	0.981	0.987	1.061	0.911**
DIF(1)	1.367	1.366	1.315	1.124	1.230
DIF(2)	1.514	1.527	1.428	1.249	1.170
DIF(3)	1.483	1.602	1.535	1.369	1.219
DIF(4)	2.757	2.332	2.023	1.677	1.430
DIF(5)	2.918	2.444	2.107	1.728	1.429
DIF(6)	3.131	2.713	2.456	2.145	1.729
DIF(1)+FB1	2.756	2.260	1.817	1.282	1.270
DIF(2)+FB1	2.146	1.940	1.674	1.376	1.230
DIF(3)+FB1	1.867	1.948	1.741	1.460	1.272
DIF(1)+FB2	2.826	2.300	1.852	1.328	1.252
DIF(2)+FB2	2.111	1.912	1.662	1.377	1.244
DIF(3)+FB2	1.810	1.895	1.702	1.437	1.256

* Notes: See notes to Table 2.2A.

Table 2.3D: 3-Step-Ahead Relative MSFEs of All Forecasting Models (Subsample 4: 1992:7-2016:7)

*

Model	rMSFE				
Maturity	1 year	2 year	3 years	5 years	10 years
AR(1)	1.000	1.000	1.000	1.000	1.000
VAR(1)	0.927^{**}	0.974	0.977	0.972	0.987
VAR(1)+FB1	0.909^{***}	0.960	0.966	0.965	0.988
VAR(1)+FB2	0.941^{**}	0.994	1.001	1.002	1.027
AR(SIC)	0.878^{***}	0.930^{**}	0.928^{***}	**0.925*****	0.928^{***}
VAR(SIC)	0.927^{**}	0.974	0.977	0.972	0.987
VAR(SIC)+FB1	0.909^{***}	0.960	0.966	0.965	0.988
VAR(SIC)+FB2	0.941^{**}	0.994	1.001	1.002	1.027
DNS(1)	1.320	1.119	1.101	1.150	0.976
DNS(2)	1.300	1.129	1.118	1.171	0.976
DNS(3)	1.273	1.101	1.089	1.166	1.012
DNS(4)	0.967^{*}	0.989	0.981	0.994	0.946^{*}
DNS(5)	0.962^{*}	0.986	0.982	1.004	0.943^{*}
DNS(6)	0.963^{*}	0.999	0.981	0.999	0.949^{**}
DNS(1)+FB1	1.111	1.041	1.034	1.100	0.995
DNS(2)+FB1	1.103	1.045	1.044	1.116	0.995
DNS(3)+FB1	1.092	1.035	1.027	1.114	1.031
DNS(4)+FB1	0.868^{***}	0.912^{**}	0.918^{**}	0.952^{**}	0.928^{**}
DNS(5)+FB1	0.866^{***}	**0.910*****	0.920^{***}	0.962^{*}	**0.925****
DNS(6)+FB1	**0.861*****	0.919^{**}	**0.917****	0.956^{**}	0.932^{**}
DNS(1)+FB2	1.206	1.106	1.081	1.124	0.997
DNS(2)+FB2	1.194	1.106	1.089	1.139	0.997
DNS(3)+FB2	1.193	1.104	1.079	1.145	1.040
DNS(4)+FB2	0.898^{***}	0.938^{*}	0.943^{**}	0.974	0.947^{*}
DNS(5)+FB2	0.897^{***}	0.938^{**}	0.947^{**}	0.987	0.945^{*}
DNS(6)+FB2	0.889^{***}	0.944^{*}	0.940^{**}	0.979	0.951^{**}
DNS(1)+MAC	1.251	1.136	1.133	1.177	1.001
DNS(2)+MAC	1.234	1.144	1.148	1.196	1.001
DNS(3)+MAC	1.209	1.121	1.124	1.192	1.034
DNS(4)+MAC	0.927^{**}	0.969	0.979	1.011	0.966
DNS(5)+MAC	0.921^{**}	0.967	0.981	1.020	0.961
DNS(6)+MAC	0.918^{**}	0.976	0.978	1.017	0.974
DIF(1)	1.589	1.475	1.310	1.081	1.298
DIF(2)	1.312	1.300	1.261	1.212	1.246
DIF(3)	1.263	1.339	1.310	1.255	1.243
DIF(4)	1.460	1.438	1.435	1.438	1.411
DIF(5)	1.591	1.582	1.557	1.523	1.484
DIF(6)	1.665	1.668	1.664	1.658	1.604
DIF(1)+FB1	1.477	1.434	1.306	1.147	1.359
DIF(2)+FB1	1.304	1.294	1.260	1.235	1.295
DIF(3)+FB1	1.181	1.284	1.276	1.254	1.279
DIF(1)+FB2	1.598	1.594	1.452	1.233	1.433
DIF(2)+FB2	1.329	1.343	1.310	1.265	1.297
DIF(3)+FB2	1.195	1.318	1.312	1.281	1.293

* Notes: See notes to Table 2.2A.

Table 2.4A: 12-Step-Ahead Relative MSFEs of All Forecasting Models (Subsample 1: 1994:1-1999:12) *

Model	rMSFE				
Maturity	1 year	2 year	3 years	5 years	10 years
AR(1)	1.000	1.000	1.000	1.000	1.000
VAR(1)	1.312	1.302	1.280	1.207	1.049
VAR(1)+FB1	1.299	1.290	1.267	1.194	1.036
VAR(1)+FB2	1.292	1.282	1.260	1.189	1.031
AR(SIC)	1.208	1.217	1.206	1.132	0.967
VAR(SIC)	1.312	1.302	1.280	1.207	1.049
VAR(SIC)+FB1	1.299	1.290	1.267	1.194	1.036
VAR(SIC)+FB2	1.292	1.282	1.260	1.189	1.031
DNS(1)	0.635***	0.682***	0.730***	0.846**	**0.954**
DNS(2)	0.640***	0.688***	0.737***	0.853**	0.956
DNS(3)	**0.624******	**0.669******	**0.718******	**0.845****	0.973
DNS(4)	1.276	1.298	1.259	1.165	1.022
DNS(5)	1.283	1.298	1.258	1.168	1.025
DNS(6)	1.284	1.301	1.256	1.157	1.013
DNS(1)+FB1	0.905	0.889	0.952	1.097	1.159
DNS(2)+FB1	0.902	0.895	0.960	1.104	1.157
DNS(3)+FB1	0.897	0.882	0.950	1.113	1.200
DNS(4)+FB1	1.225	1.265	1.237	1.157	1.024
DNS(5)+FB1	1.232	1.265	1.236	1.158	1.027
DNS(6)+FB1	1.234	1.270	1.236	1.150	1.017
DNS(1)+FB2	1.055	0.980	1.049	1.226	1.295
DNS(2)+FB2	1.047	0.984	1.056	1.233	1.293
DNS(3)+FB2	1.059	0.984	1.060	1.256	1.351
DNS(4)+FB2	1.210	1.252	1.227	1.152	1.020
DNS(5)+FB2	1.216	1.250	1.225	1.153	1.022
DNS(6)+FB2	1.220	1.257	1.226	1.147	1.014
DNS(1)+MAC	0.685**	0.729**	0.776**	0.889	0.977
DNS(2)+MAC	0.689**	0.734**	0.782**	0.895	0.978
DNS(3)+MAC	0.672**	0.716**	0.765**	0.889	0.998
DNS(4)+MAC	1.228	1.275	1.253	1.181	1.050
DNS(5)+MAC	1.233	1.273	1.250	1.180	1.050
DNS(6)+MAC	1.237	1.280	1.253	1.176	1.045
DIF(1)	0.984	0.925*	0.838***	1.124	1.829
DIF(2)	1.328	1.346	1.493	1.748	1.905
DIF(3)	1.254	1.224	1.348	1.586	1.812
DIF(4)	1.122	1.156	1.184	1.225	1.292
DIF(5)	1.619	1.610	1.647	1.689	1.690
DIF(6)	1.718	1.695	1.723	1.747	1.712
DIF(1)+FB1	1.340	1.283	1.158	1.371	2.125
DIF(2)+FB1	1.622	1.749	1.871	2.054	2.170
DIF(3)+FB1	1.459	1.542	1.653	1.850	2.084
DIF(1)+FB2	1.487	1.503	1.548	1.960	2.324
DIF(2)+FB2	1.883	1.908	2.002	2.166	2.262
DIF(3)+FB2	1.593	1.637	1.752	1.961	2.176

* Notes: See notes to Table 2.2A.

Table 2.4B: 12-Step-Ahead Relative MSFEs of All Forecasting Models (Subsample 2: 2000:1-2007:12) *

Model	rMSFE				
Maturity	1 year	2 year	3 years	5 years	10 years
AR(1)	1.000	1.000	1.000	1.000	1.000
VAR(1)	0.567***	**0.475*****	**0.432*****	**0.431*****	0.557***
VAR(1)+FB1	0.583***	0.488***	0.444***	0.441***	0.561***
VAR(1)+FB2	0.587***	0.493***	0.451***	0.452***	0.584***
AR(SIC)	0.574***	0.485***	0.444***	0.445***	**0.547*****
VAR(SIC)	0.567***	0.475***	0.432***	0.431***	0.557***
VAR(SIC)+FB1	0.583***	0.488***	0.444***	0.441***	0.561***
VAR(SIC)+FB2	0.587***	0.493***	0.451***	0.452***	0.584***
DNS(1)	0.708***	0.602***	0.571***	0.631***	0.770***
DNS(2)	0.707***	0.605***	0.576***	0.638***	0.771***
DNS(3)	0.702***	0.599***	0.570***	0.639***	0.811***
DNS(4)	0.593***	0.507***	0.460***	0.454***	0.564***
DNS(5)	0.593***	0.506***	0.459***	0.456***	0.564***
DNS(6)	0.599***	0.510***	0.460***	0.452***	0.563***
DNS(1)+FB1	0.548***	0.507***	0.520***	0.658***	0.999
DNS(2)+FB1	0.547***	0.508***	0.523***	0.662***	0.996
DNS(3)+FB1	**0.543*****	0.502***	0.516***	0.661***	1.036
DNS(4)+FB1	0.595***	0.509***	0.461***	0.455***	0.564***
DNS(5)+FB1	0.595***	0.508***	0.461***	0.457***	0.565***
DNS(6)+FB1	0.598***	0.511***	0.461***	0.452***	0.562***
DNS(1)+FB2	1.023	1.049	1.143	1.530	2.724
DNS(2)+FB2	1.023	1.055	1.152	1.541	2.725
DNS(3)+FB2	1.010	1.036	1.130	1.521	2.739
DNS(4)+FB2	0.594***	0.510***	0.464***	0.460***	0.574***
DNS(5)+FB2	0.594***	0.510***	0.464***	0.462***	0.575***
DNS(6)+FB2	0.597***	0.511***	0.463***	0.456***	0.571***
DNS(1)+MAC	0.671***	0.609***	0.604***	0.700***	0.884***
DNS(2)+MAC	0.669***	0.611***	0.609***	0.706***	0.884***
DNS(3)+MAC	0.665***	0.606***	0.604***	0.708***	0.922**
DNS(4)+MAC	0.579***	0.495***	0.451***	0.449***	0.567***
DNS(5)+MAC	0.580***	0.495***	0.450***	0.451***	0.566***
DNS(6)+MAC	0.583***	0.497***	0.450***	0.447***	0.568***
DIF(1)	1.244	1.157	1.124	1.142	1.537
DIF(2)	1.895	1.456	1.321	1.372	1.801
DIF(3)	2.284	1.721	1.565	1.569	1.959
DIF(4)	0.842***	0.966	1.112	1.448	1.910
DIF(5)	1.019	1.109	1.243	1.765	3.530
DIF(6)	1.037	1.146	1.299	1.857	3.729
DIF(1)+FB1	1.001	1.056	1.129	1.539	2.403
DIF(2)+FB1	1.544	1.364	1.412	1.711	2.703
DIF(3)+FB1	1.798	1.664	1.720	2.028	2.985
DIF(1)+FB2	1.032	1.138	1.272	1.842	3.941
DIF(2)+FB2	1.850	1.804	1.980	2.589	4.305
DIF(3)+FB2	2.098	2.053	2.226	2.832	4.471

* Notes: See notes to Table 2.2A.

Table 2.4C: 12-Step-Ahead Relative MSFEs of All Forecasting Models (Subsample 3: 2008:1-2016:7) *

Model	rMSFE				
Maturity	1 year	2 year	3 years	5 years	10 years
AR(1)	1.000	1.000	1.000	1.000	1.000
VAR(1)	0.737***	0.716***	0.695***	**0.663*****	0.709***
VAR(1)+FB1	0.778***	0.756***	0.731***	0.686***	0.707***
VAR(1)+FB2	0.796***	0.773***	0.747***	0.697***	0.713***
AR(SIC)	**0.729*****	0.714***	**0.687*****	0.674***	0.737***
VAR(SIC)	0.737***	0.716***	0.695***	0.663***	0.709***
VAR(SIC)+FB1	0.778***	0.756***	0.731***	0.686***	0.707***
VAR(SIC)+FB2	0.796***	0.773***	0.747***	0.697***	0.713***
DNS(1)	1.468	1.389	1.483	1.559	1.118
DNS(2)	1.451	1.412	1.520	1.591	1.118
DNS(3)	1.423	1.356	1.467	1.583	1.179
DNS(4)	0.765***	0.689***	0.697***	0.737***	0.687***
DNS(5)	0.759***	0.694***	0.708***	0.748***	**0.683*****
DNS(6)	0.747***	0.679***	0.692***	0.747***	0.706***
DNS(1)+FB1	1.744	1.554	1.511	1.534	1.209
DNS(2)+FB1	1.734	1.557	1.524	1.552	1.206
DNS(3)+FB1	1.736	1.542	1.502	1.554	1.274
DNS(4)+FB1	0.813***	0.730***	0.737***	0.771***	0.703***
DNS(5)+FB1	0.804***	0.735***	0.746***	0.781***	0.699***
DNS(6)+FB1	0.791***	0.718***	0.730***	0.780***	0.723***
DNS(1)+FB2	1.895	1.779	1.792	1.842	1.427
DNS(2)+FB2	1.887	1.785	1.809	1.863	1.425
DNS(3)+FB2	1.877	1.757	1.771	1.851	1.486
DNS(4)+FB2	0.827***	0.742***	0.747***	0.777***	0.703***
DNS(5)+FB2	0.816***	0.745***	0.755***	0.786***	0.699***
DNS(6)+FB2	0.804***	0.730***	0.740***	0.786***	0.723***
DNS(1)+MAC	1.579	1.583	1.729	1.824	1.274
DNS(2)+MAC	1.555	1.600	1.761	1.850	1.269
DNS(3)+MAC	1.543	1.557	1.723	1.865	1.353
DNS(4)+MAC	0.758***	0.685***	0.695***	0.739***	0.694***
DNS(5)+MAC	0.751***	0.690***	0.705***	0.749***	0.691***
DNS(6)+MAC	0.739***	**0.675*****	0.689***	0.747***	0.711***
DIF(1)	1.190	1.280	1.277	1.357	1.155
DIF(2)	2.405	2.453	2.228	1.789	1.155
DIF(3)	2.787	2.711	2.641	2.260	1.472
DIF(4)	1.978	1.719	1.539	1.291	1.086
DIF(5)	1.859	1.677	1.611	1.564	1.457
DIF(6)	2.332	2.278	2.257	2.110	1.711
DIF(1)+FB1	2.036	1.789	1.563	1.316	1.167
DIF(2)+FB1	2.280	2.164	1.982	1.663	1.171
DIF(3)+FB1	2.490	2.487	2.410	2.139	1.569
DIF(1)+FB2	1.886	1.726	1.614	1.618	1.462
DIF(2)+FB2	2.289	2.228	2.123	1.907	1.430
DIF(3)+FB2	2.503	2.530	2.505	2.297	1.737

* Notes: See notes to Table 2.2A.

Table 2.4D: 12-Step-Ahead Relative MSFEs of All Forecasting Models (Subsample 4: 1994:1-2016:7) *

Model	rMSFE				
Maturity	1 year	2 year	3 years	5 years	10 years
AR(1)	1.000	1.000	1.000	1.000	1.000
VAR(1)	0.706***	0.660***	0.643***	0.665***	0.767***
VAR(1)+FB1	0.724***	0.674***	0.655***	0.673***	0.763***
VAR(1)+FB2	0.730***	0.680***	0.661***	0.680***	0.771***
AR(SIC)	**0.695***	**0.652***	**0.635***	**0.657***	**0.748***
VAR(SIC)	0.706***	0.660***	0.643***	0.665***	0.767***
VAR(SIC)+FB1	0.724***	0.674***	0.655***	0.673***	0.763***
VAR(SIC)+FB2	0.730***	0.680***	0.661***	0.680***	0.771***
DNS(1)	0.879**	0.784***	0.794***	0.910**	0.953
DNS(2)	0.875**	0.792***	0.806***	0.923*	0.954*
DNS(3)	0.863***	0.773***	0.788***	0.919*	0.994
DNS(4)	0.725***	0.673***	0.656***	0.686***	0.753***
DNS(5)	0.724***	0.674***	0.658***	0.690***	0.753***
DNS(6)	0.725***	0.674***	0.655***	0.685***	0.756***
DNS(1)+FB1	0.880	0.794**	0.810**	0.975	1.125
DNS(2)+FB1	0.876	0.796**	0.816**	0.983	1.122
DNS(3)+FB1	0.873	0.787**	0.805**	0.985	1.173
DNS(4)+FB1	0.730***	0.678***	0.661***	0.692***	0.759***
DNS(5)+FB1	0.729***	0.679***	0.663***	0.696***	0.759***
DNS(6)+FB1	0.729***	0.677***	0.659***	0.692***	0.764***
DNS(1)+FB2	1.234	1.195	1.264	1.539	1.807
DNS(2)+FB2	1.232	1.201	1.274	1.551	1.806
DNS(3)+FB2	1.222	1.183	1.254	1.543	1.850
DNS(4)+FB2	0.731***	0.679***	0.663***	0.695***	0.761***
DNS(5)+FB2	0.730***	0.680***	0.665***	0.699***	0.760***
DNS(6)+FB2	0.729***	0.678***	0.661***	0.695***	0.766***
DNS(1)+MAC	0.889*	0.838**	0.875**	1.022	1.054
DNS(2)+MAC	0.882**	0.844**	0.886*	1.032	1.052
DNS(3)+MAC	0.875**	0.828***	0.872**	1.036	1.101
DNS(4)+MAC	0.708***	0.662***	0.649***	0.687***	0.765***
DNS(5)+MAC	0.707***	0.662***	0.651***	0.690***	0.764***
DNS(6)+MAC	0.707***	0.661***	0.647***	0.687***	0.770***
DIF(1)	1.197	1.146	1.104	1.191	1.492
DIF(2)	1.941	1.653	1.546	1.561	1.602
DIF(3)	2.267	1.854	1.755	1.744	1.738
DIF(4)	1.149	1.159	1.216	1.358	1.419
DIF(5)	1.299	1.312	1.395	1.698	2.204
DIF(6)	1.435	1.478	1.581	1.895	2.367
DIF(1)+FB1	1.292	1.250	1.227	1.445	1.871
DIF(2)+FB1	1.729	1.598	1.617	1.777	1.984
DIF(3)+FB1	1.917	1.821	1.855	2.015	2.192
DIF(1)+FB2	1.295	1.323	1.395	1.813	2.540
DIF(2)+FB2	1.959	1.912	2.014	2.324	2.627
DIF(3)+FB2	2.127	2.088	2.199	2.502	2.764

* Notes: See notes to Table 2.2A.

Table 2.5: Top 3 Forecast Models with Lowest MSFE*

	Maturity	3 Months	1 Year	3 Years	5 Years	10 Years
Forecast Sample	Horizon					
		DNS(6)+FB2***	**DNS(5)+FB2*****	**DNS(5)+FB2*****	**DIF(1)****	**AR(SIC)****
	1 Step	DNS(6)+FB1***	DNS(4)+FB2***	DNS(4)+FB2***	DNS(4)+FB2**	DNS(2)+FB1*
		DNS(4)+FB1***	DNS(5)+FB1***	DNS(6)+FB2***	DNS(4)+FB1***	DNS(1)+FB1*
		DNS(4)+FB1***	**DNS(3)+FB1****	**DNS(1)+FB1***	**DNS(6)+FB1****	**AR(SIC)****
1992:3-1999:12	3 Step	DNS(6)+FB1***	DNS(1)+FB1**	DNS(3)+FB1*	DNS(5)+FB1**	DNS(6)+FB1*
‘1st Subsample		DNS(5)+FB1***	DNS(2)+FB1**	DNS(2)+FB1*	DNS(4)+FB1**	DNS(5)+FB1**
		DNS(3)***	**DNS(3)*****	**DNS(3)*****	**DNS(3)**	**DNS(1)**
	12 Step	DNS(1)	DNS(1)	DNS(1)	DNS(1)	DNS(2)
		DNS(2)	DNS(2)	DNS(2)	DNS(2)	AR(SIC)
		DNS(3)+FB2***	**DNS(2)+FB2****	**DNS(1)+FB2****	**DNS(1)+FB2*****	**DNS(2)+FB2****
	1 Step	DNS(5)+FB1***	DNS(1)+FB2**	DNS(3)+FB2**	DNS(2)+FB2***	DNS(1)+FB2**
		DNS(6)+FB2***	DNS(3)+FB2**	DNS(2)+FB2***	DNS(3)+FB2***	DNS(3)+FB2**
		DNS(3)+FB1***	**DNS(3)+FB1*****	**DNS(3)+FB1*****	**DNS(1)+FB1*****	**DNS(2)+FB1**
2000:1-2007:12	3 Step	DNS(2)+FB1***	DNS(1)+FB1***	DNS(1)+FB1***	DNS(2)+FB1***	DNS(1)+FB1
‘2nd Subsample		DNS(1)+FB1***	DNS(2)+FB1***	DNS(2)+FB1***	DNS(3)+FB1***	DNS(3)+FB1*
		DNS(3)+FB1**	**VAR(1)****	**VAR(1)*****	**VAR(1)*****	**AR(SIC)*****
	12 Step	DNS(2)+FB1**	VAR(SIC)**	VAR(SIC)***	VAR(SIC)***	VAR(SIC)***
		DNS(1)+FB1**	AR(SIC)**	VAR(SIC)+FB1***	VAR(SIC)+FB1***	VAR(1)***
		AR(SIC)***	**AR(1)**	**AR(1)****	**AR(SIC)*****	**DNS(2)**
	1 Step	AR(1)***	DNS(3)*	AR(SIC)**	AR(1)***	DNS(1)
		DIF(3)***	AR(SIC)	DNS(3)+MAC	DIF(1)+FB2*	DNS(2)+MAC
		DNS(6)+FB1*	**DNS(5)+FB1**	**AR(SIC)*****	**AR(SIC)*****	**DNS(5)+MAC**
2008:1-2016:7	3 Step	VAR(1)+FB1***	DNS(4)+FB1	DNS(6)+FB1	VAR(1)+FB1***	DNS(5)*
‘3rd Subsample		VAR(SIC)+FB1***	DNS(6)+FB1	DNS(4)+FB1	VAR(SIC)+FB1***	DNS(4)+MAC
		AR(SIC)***	**DNS(6)+MAC**	**AR(SIC)*****	**VAR(1)*****	**DNS(5)*****
	12 Step	VAR(1)***	DNS(6)***	DNS(6)+MAC	VAR(SIC)***	DNS(4)***
		VAR(SIC)***	DNS(4)+MAC	DNS(6)***	AR(SIC)***	DNS(5)+MAC
		DNS(6)+FB1	**DNS(5)+FB1**	**DNS(4)+FB1**	**AR(1)*****	**DNS(2)**
	1 Step	VAR(SIC)+FB1**	DNS(5)+FB2	DNS(5)+FB1	AR(SIC)***	DNS(1)
		VAR(1)+FB1**	DNS(4)+FB1	DNS(6)+FB1	VAR(SIC)+FB1	DNS(2)+MAC
		DNS(6)+FB1*	**DNS(5)+FB1**	**DNS(6)+FB1**	**AR(SIC)*****	**DNS(5)+FB1**
1992:3-2016:7	3 Step	DNS(5)+FB1*	DNS(4)+FB1	DNS(4)+FB1	DNS(4)+FB1	AR(SIC)
‘Whole Sample		DNS(4)+FB1*	DNS(6)+FB1	DNS(5)+FB1	DNS(6)+FB1	DNS(4)+FB1
		AR(SIC)***	**AR(SIC)*****	**AR(SIC)*****	**AR(SIC)*****	**AR(SIC)*****
	12 Step	VAR(1)***	VAR(1)***	VAR(1)***	VAR(1)***	DNS(5)***
		VAR(SIC)***	VAR(SIC)***	VAR(SIC)***	VAR(SIC)***	DNS(4)***

* Notes: See notes to Table 2.2A. This table reports the top three performing forecast models (based on MSFE) from lowest-MSFE to highest-MSFE, for all subsamples, horizons, and maturities, summarizing the results of Tables 2A-4D. Entries in bold denote models with lowest MSFE for a given maturity. Entries superscripted with ***, **, and * denote rejections of the null hypothesis of equal predictive accuracy at 0.01, 0.05, and 0.10 significance levels, respectively, based on application of the Diebold-Mariano test discussed in Section 3; and indicate that the listed model is predictively superior to a “benchmark” DNS(τ) model, based on MSFE loss. In particular, if the point “MSFE-best” model is DNS(τ)+mod, where mod denotes another component of the model (for example, mod may be FB1 or FB2, etc.) then the “benchmark” model is DNS(τ). If the point “MSFE-best” model is DNS(1), or if no DNS component appears in point “MSFE-best” model, then DNS(1) is the “benchmark” model. Finally, for entries denoted “DNS(1)”, no predictive accuracy test was carried out. These test results are included to highlight the importance of incorporating “big data”in DNS type prediction models. For complete details, refer to Section 4.

Table 2.6: Top 3 Forecast Models with Lowest MSFE in Expansionary and Recessionary Periods*

	Maturity	3 Months	1 Year	3 Years	5 Years	10 Years
Forecast Sample	Horizon					
Recession	1 Step	**DNS(4)+FB1**	**VAR(SIC)+FB1**	**VAR(1)**	**DNS(3)+FB2**	**DIF(2)+FB2**
		DNS(5)+FB1	VAR(1)+FB1	VAR(SIC)	DNS(2)+FB2	DNS(3)
		VAR(SIC)+FB1	DNS(2)+MAC	DNS(1)+MAC	VAR(SIC)	DNS(2)
	3 Step	**DNS(6)+FB1**	**DNS(6)+FB1**	**DNS(6)+FB1**	**DNS(2)+FB1**	**DNS(3)+FB1**
		DNS(6)+MAC	DNS(4)+FB1	DNS(4)+FB1	DNS(3)+FB1	DNS(2)
		DNS(1)+FB1	DNS(6)+MAC	DNS(5)+FB1	DNS(1)+FB1	DNS(1)
	12 Step	**DNS(3)+FB1**	**DNS(3)+FB1**	**DNS(3)+FB1**	**DNS(1)+FB1**	**VAR(1)**
		DNS(2)+FB1	DNS(1)+FB1	DNS(1)+FB1	DNS(2)+FB1	VAR(SIC)
		DNS(1)+FB1	DNS(2)+FB1	DNS(2)+FB1	DNS(3)+FB1	DNS(5)
Expansion	1 Step	**DNS(6)+FB2**	**DNS(5)+FB2**	**DNS(6)+FB2**	**AR(1)**	**DNS(2)+FB2**
		DNS(6)+FB1	DNS(4)+FB2	DNS(4)+FB2	AR(SIC)	DNS(1)+FB2
		VAR(1)+FB1	DNS(3)+FB2	DNS(5)+FB2	DIF(1)	DNS(2)+FB1
	3 Step	**DNS(5)+FB1**	**DNS(5)+FB1**	**AR(SIC)**	**AR(SIC)**	**DNS(5)+FB1**
		DNS(6)+FB1	DNS(4)+FB1	DNS(5)+FB1	DNS(4)+FB1	DNS(4)+FB1
		DNS(4)+FB1	AR(SIC)	DNS(4)+FB1	DNS(6)+FB1	AR(SIC)
	12 Step	**DNS(4)+MAC**	**AR(SIC)**	**AR(SIC)**	**AR(SIC)**	**AR(SIC)**
		DNS(5)+MAC	DNS(5)+MAC	DNS(5)+MAC	VAR(1)+FB1	DNS(6)
		DNS(6)+MAC	DNS(4)+MAC	DNS(6)+MAC	VAR(SIC)+FB1	DNS(4)

* Notes: See notes to Table 2.5. Recessions and expansion are defined according to NBER business cycle dates.

Table 2.7: Forecast Combination Models Used in Forecast Experiments*

Model	Description
All	Average of all forty four forecast models
FB	Average of twenty five models that contain principle component(s), principle component analysis based on all 103 macroeconomic variables
FS	Average of nineteen non-FB type models
Econometrics	Average of all eight AR and VAR type models
DNS	Average of all twenty two DNS type models
DI	Average of twelve diffusion index type models

* Notes: This table summarizes the combination models utilized in all forecast experiments.

Table 2.8A: 1-Step-Ahead Relative MSFEs of Forecast Combination Models*

	Model	rMSFE				
	Maturity	1 year	2 year	3 years	5 years	10 years
1992:3-1999:12 'Subsample 1'	All	0.922	0.976	0.971	0.999	1.066
	FB	0.906	0.949	0.958	1.008	1.101
	FS	1.003	1.062	1.030	1.020	1.063
	Econometrics	**0.842****	**0.893****	**0.912****	**0.930***	**0.993**
	DNS	0.861**	0.928**	0.932**	0.982	0.998
	DIF	1.293	1.292	1.196	1.142	1.468
2000:1-2007:12 'Subsample 2'	All	0.740***	0.866***	0.903**	0.981	0.967
	FB	**0.652*****	**0.812****	**0.864****	0.949	0.964
	FS	0.933	0.991	0.996	1.052	0.999
	Econometrics	0.769***	0.871**	0.899*	**0.933**	1.005
	DNS	0.751***	0.838***	0.866***	0.997	**0.931****
	DIF	0.926	1.076	1.091	1.047	1.201
2008:1-2016:7 'Subsample 3'	All	1.227	1.174	1.147	1.239	1.081
	FB	1.654	1.527	1.359	1.313	1.155
	FS	1.135	**1.010**	**1.067**	1.235	**1.013**
	Econometrics	**0.969**	1.079	1.092	**1.098**	1.125
	DNS	1.365	1.129	1.134	1.419	1.030
	DIF	1.841	1.702	1.466	1.216	1.335
1992:3-2016:7 'Subsample 4'	All	0.911	0.971	0.984	1.061	1.042
	FB	0.958	1.011	1.011	1.074	1.082
	FS	1.002	1.022	1.025	1.094	1.022
	Econometrics	**0.839*****	**0.922***	**0.947**	**0.980**	1.053
	DNS	0.922	0.932*	0.951	1.114	**0.991**
	DIF	1.257	1.286	1.215	1.127	1.329
Recession	All	0.814	0.991	0.968	0.920	0.996
	FB	1.052	1.287	1.190	0.997	1.063
	FS	0.887*	0.907	0.943*	0.999	**0.956**
	Econometrics	**0.692****	**0.805***	**0.841**	0.899	1.061
	DNS	0.707**	0.910	0.877	**0.893****	1.052
	DIF	1.506	1.517	1.404	1.109	0.984
Normal	All	0.948	0.966	0.987	1.089	1.054
	FB	0.923	0.938*	0.972	1.089	1.087
	FS	1.045	1.052	1.042	1.113	1.040
	Econometrics	**0.894***	0.953	0.971	**0.995**	1.051
	DNS	1.002	**0.938****	**0.967**	1.157	**0.975**
	DIF	1.163	1.225	1.174	1.131	1.419

* Notes: See notes to Table 2.2A. Forecast combination models are listed in Table 2.7.

Table 2.8B: 3-Step-Ahead Relative MSFEs of Forecast Combination Models*

	Model	rMSFE				
	Maturity	1 year	2 year	3 years	5 years	10 years
	All	0.943	1.011	1.021	1.024	1.027
	FB	0.941	1.013	1.032	1.046	1.057
1992:7-1999:12	FS	0.984	1.037	1.029	1.010	1.003
'Subsample 1'	Econometrics	0.989	1.041	1.055	1.050	1.028
	DNS	**0.881****	**0.953**	**0.970**	**0.982**	**0.968**
	DIF	1.168	1.228	1.194	1.174	1.278
	All	0.779***	0.842***	0.864***	0.935*	0.959**
	FB	**0.701*****	**0.789*****	0.828***	0.916*	0.971
2000:1-2007:12	FS	0.911**	0.930**	0.926**	0.970	0.960**
'Subsample 2'	Econometrics	0.844***	0.886***	0.888***	0.907**	0.996
	DNS	0.770***	0.790***	**0.802*****	**0.887****	**0.873*****
	DIF	0.881*	1.030	1.076	1.142	1.286
	All	1.049	1.067	1.088	1.125	0.995
	FB	1.284	1.280	1.230	1.187	1.040
2008:1-2016:7	FS	1.141	1.054	1.078	1.122	0.954
'Subsample 3'	Econometrics	**0.934**	**0.960**	**0.945**	**0.914****	**0.923****
	DNS	1.105	1.004	1.052	1.174	0.942
	DIF	1.614	1.597	1.472	1.314	1.262
	All	0.897**	0.955	0.976	1.022	0.997
	FB	0.918	0.983	1.001	1.041	1.027
1992:7-2016:7	FS	0.989	0.998	1.001	1.028	0.973**
'Subsample 4'	Econometrics	0.914***	0.960*	0.964*	**0.962***	0.979
	DNS	**0.885****	**0.899****	**0.925****	1.004	**0.933*****
	DIF	1.149	1.232	1.215	1.203	1.274
	All	0.794**	0.826*	0.831*	0.873**	0.973
	FB	0.856	0.897	0.875	0.880	1.009
Recession	FS	0.946*	0.910**	0.915**	0.962	0.972
	Econometrics	0.868***	0.874***	0.863***	0.863**	1.001
	DNS	**0.759*****	**0.763****	**0.770****	**0.828*****	**0.935**
	DIF	1.105	1.089	1.053	1.061	1.216
	All	0.962	1.016	1.032	1.060	1.000
	FB	0.957	1.024	1.048	1.082	1.030
Normal	FS	1.016	1.040	1.034	1.045	0.973*
	Econometrics	**0.943****	1.001	1.003	**0.988**	0.976
	DNS	0.965	**0.963**	**0.984**	1.049	**0.933*****
	DIF	1.177	1.299	1.277	1.239	1.283

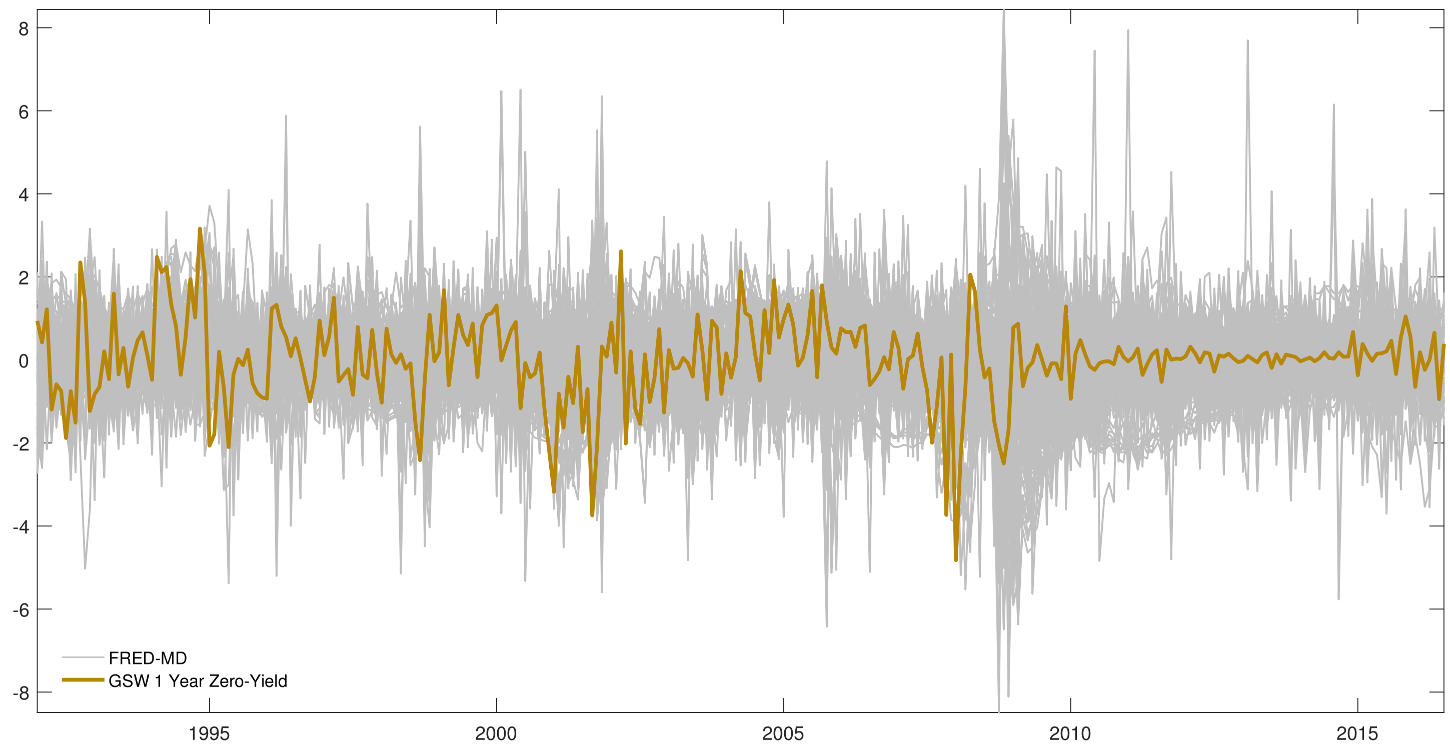
* Notes: See notes to Table 2.8A.

Table 2.8C: 12-Step-Ahead Relative MSFEs of Forecast Combination Models*

	Model	rMSFE				
	Maturity	1 year	2 year	3 years	5 years	10 years
	All	0.852	0.923	0.966	1.044	1.071
	FB	0.912	0.983	1.034	1.126	1.164
1994:1-1999:12	FS	0.889^{*}	0.933	0.951	0.988	0.990
'Subsample 1'	Econometrics	1.241	1.235	1.215	1.147	0.999
	DNS	**0.773****	**0.844****	**0.880***	**0.931**	**0.920****
	DIF	1.204	1.281	1.363	1.607	1.888
	All	0.673^{***}	0.619^{***}	0.616^{***}	0.717^{***}	1.006
	FB	0.706^{***}	0.676^{***}	0.699^{***}	0.868^{***}	1.341
2000:1-2007:12	FS	0.641^{***}	0.559^{***}	0.527^{***}	0.558^{***}	0.686^{***}
'Subsample 2'	Econometrics	0.591^{***}	0.509^{***}	**0.470*****	**0.467*****	**0.577*****
	DNS	**0.548*****	**0.486*****	0.470^{***}	0.535^{***}	0.712^{***}
	DIF	1.348	1.309	1.378	1.719	2.708
	All	0.809^{***}	0.868^{**}	0.965	1.081	0.950^{*}
	FB	0.986	1.022	1.084	1.158	1.008
2008:1-2016:7	FS	0.979	0.975	1.025	1.071	0.884^{***}
'Subsample 3'	Econometrics	0.768^{***}	**0.753*****	**0.733*****	**0.704*****	**0.738*****
	DNS	**0.746*****	0.764^{***}	0.881^{**}	1.056	0.894^{***}
	DIF	1.567	1.598	1.599	1.575	1.317
	All	0.729^{***}	0.722^{***}	0.754^{***}	0.882^{***}	1.006
	FB	0.800^{***}	0.800^{***}	0.842^{***}	0.998	1.166
1994:1-2016:7	FS	0.754^{***}	0.709^{***}	0.711^{***}	0.783^{***}	0.853^{***}
'Subsample 4'	Econometrics	0.720^{***}	0.678^{***}	**0.662*****	**0.680*****	**0.768*****
	DNS	**0.625*****	**0.603*****	0.633^{***}	0.754^{***}	0.843^{***}
	DIF	1.381	1.367	1.422	1.658	1.949
	All	1.033	1.016	1.023	1.072	1.063
	FB	**0.948**	0.963	0.999	1.103	1.254
Recession	FS	1.173	1.103	1.067	1.037	0.842^{***}
	Econometrics	1.191	1.087	0.995	**0.854*****	**0.631*****
	DNS	1.024	**0.950**	**0.945**	0.971	0.806^{***}
	DIF	0.972	1.117	1.222	1.476	2.163
	All	0.630^{***}	0.639^{***}	0.687^{***}	0.843^{***}	1.000
	FB	0.752^{***}	0.754^{***}	0.803^{***}	0.977	1.157
Normal	FS	0.618^{***}	0.598^{***}	0.622^{***}	0.731^{***}	0.854^{***}
	Econometrics	0.567^{***}	0.564^{***}	0.578^{***}	**0.645*****	**0.782*****
	DNS	**0.495*****	**0.506*****	**0.554*****	0.710^{***}	0.847^{***}
	DIF	1.514	1.437	1.472	1.695	1.927

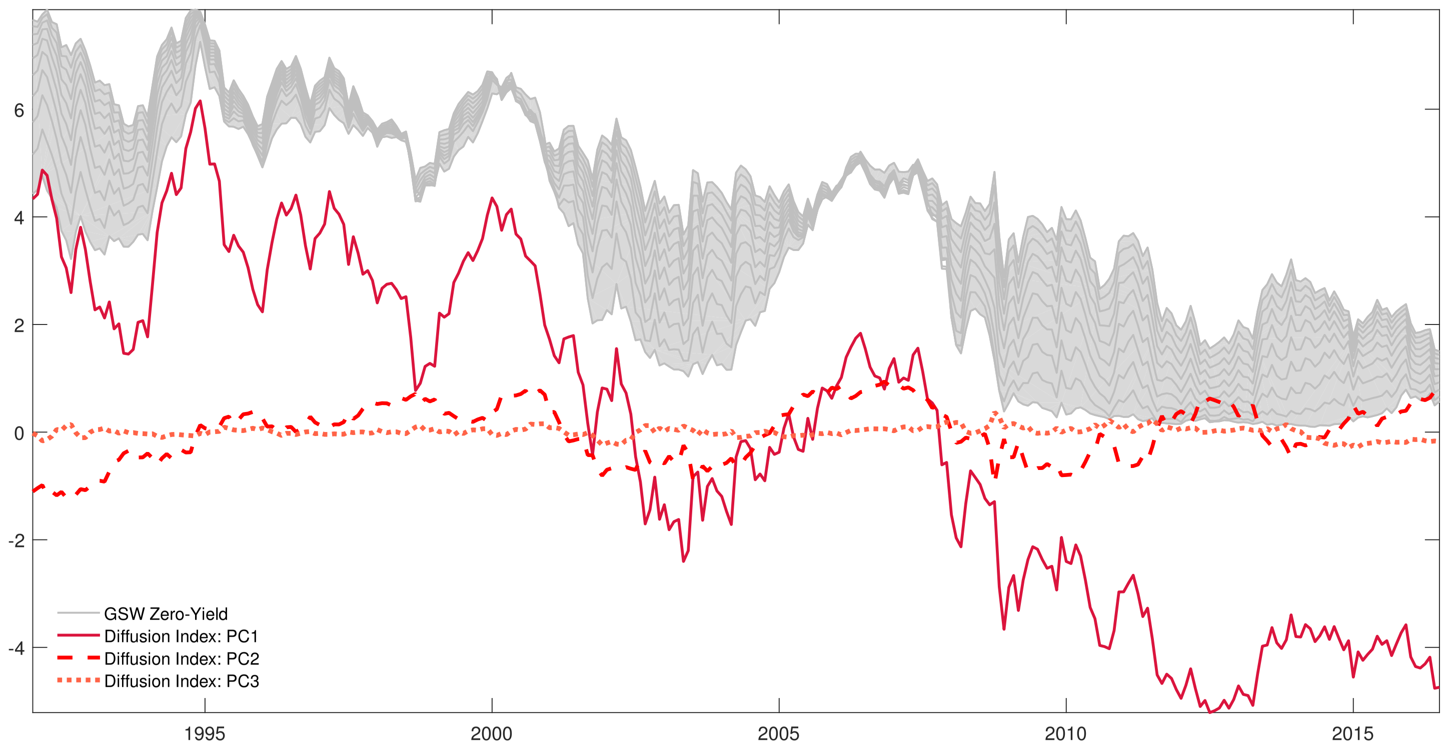
* Notes: See notes to Table 2.8A.

Figure 2.1: FRED MD Dataset for Sample Period 1992:1 - 2016:7*



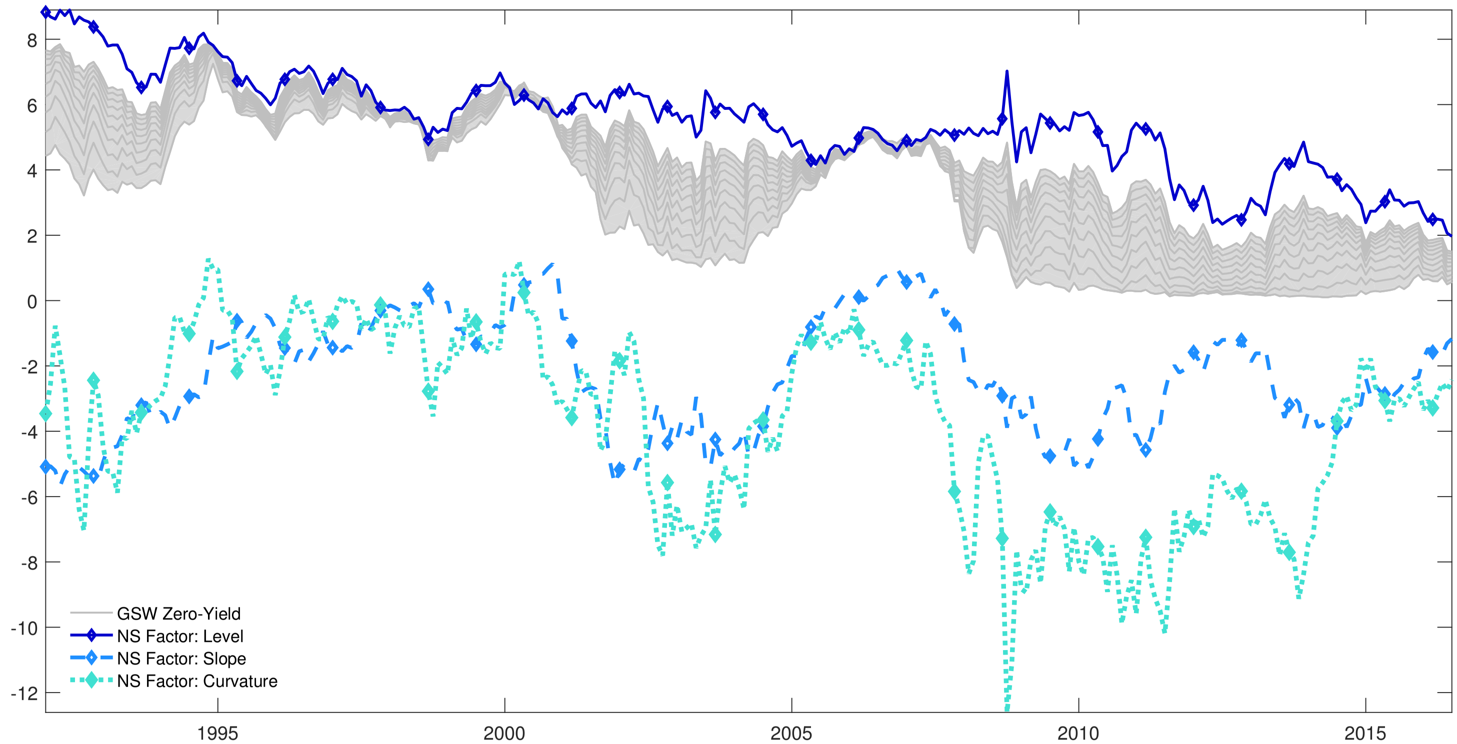
(*) Notes: The Figure displays all macroeconomic variables from the FRED-MD dataset and the 1 year zero-yield from the GSW dataset. All series transformed to ensure stationary and standardized to zero mean and unit variance.

Figure 2.2: Yields and Diffusion Indexes for Sample Period 1992:1 - 2016:7*



(*) Notes: This Figure displays all ten zero-yields from the GSW dataset and three principle components (diffusion indexes) in PCA.

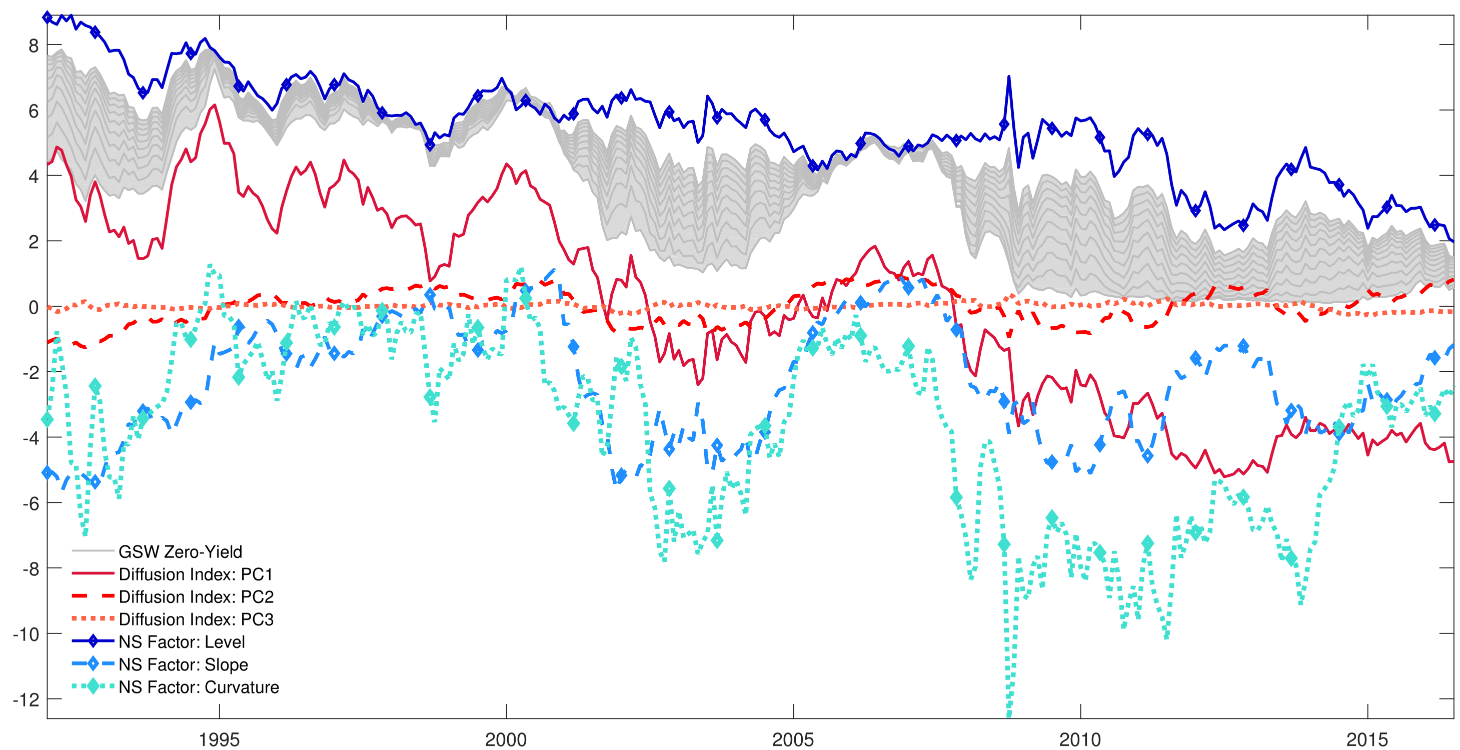
Figure 2.3: Yields and Dynamic Nelson Siegel Factors for Sample Period 1992:1 - 2016:7*



g

(*) Notes: This Figure displays all ten zero-yields from the GSW dataset and three Nelson Siegel latent factors (level, slope, and curvature).

Figure 2.4: Yields, Diffusion Indexes and Dynamic Nelson Siegel Factors for Sample Period 1992:1 - 2016:7*



(*) Notes: The Figure displays all ten zero-yields from the GSW dataset, three principle components (diffusion indexes) in PCA, and three Nelson Siegel latent factors (level, slope, and curvature).

Chapter 3

Data Rich Real-Time Dynamic Nelson Siegel Modeling

3.1 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 are, as always, as important input in models used in industry and government. Broadly speaking, then, researchers want to understand how interest rates are determined, with the hope of effectively forecasting future yield curves. In this paper, we add to the literature on interest rate prediction by carrying out an extensive set of forecast experiments in order to explore the marginal predictive content of so-called “data rich” or real-time latent macroeconomic factors in dynamic Nelson-Siegel (DNS) type models. In our context, data-rich environments contain real-time data, which are data that include the entire revision history for each variable. For example, real-time GDP observations for calendar date December 2000 include the first ”reading” on 4th quarter 2000 GDP that was available in March 2001, and well as the 1st revised version of this datuum that became available in June 2001, and so on, up until the present date. Thus, real-time datasets include entire sequence 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.

We carry out our prediction experiments for various sub-samples between 2001 and 2018, and results are evaluated using a number of benchmark linear models. In particular, we assess the following classes of models: (i) DNS type models of the variety recently examined by Diebold and Li (2007), (ii) dynamic Nelson Siegel Svensson (NSS) type models (see Svensson (1994)), and (iii) various benchmark models, including vector autoregressive (VAR) and autoregressive (AR) models. The macroeconomic factors, or so-called “big data” diffusion indexes that we utilize are extracted using principle component analysis of 130 U.S macro-variables for which McCracken and Ng (2016) have constructed a real-time dataset.

Although there many sophisticated models of the term structure have been examined in the literature, simpler regression-based approaches to forecasting treasury yields have the best track record for minimizing out-of-sample mean squared forecast error (MSFE). The most popular of these models is currently the Dynamic Nelson-Siegel model, as discussed in Diebold and Li (2006). Their DNS model is a dynamic version of the term structure model introduced by Nelson and Siegel (1987), where the cross-section movements of the term structure are summarized by the dynamics of level, slope, and curvature factors, assumed to follow AR(1) (or VAR(1)) processes. Although DNS type models have become the leading method for yield curve forecasting at many policy institutions (see BIS (2005)), and although this development is largely due to the successful empirical performance of these models, findings in the recent literature suggest that DNS model performance has deteriorated in recent (post credit crisis) years (see e.g. Altavilla, Giacomini and Ragusa (2014), Diebold, and Rudebusch (2012), and Mönch (2008)). This might be explained by the change of economic regime or structural breaks in time series models of interest rates, although research is ongoing in this area. Other potential causes include generic model misspecification, model over-fitting, and measurement error. One possible solution to this problem has centered around the introduction of new variants of DNS type models. One such model, which we examine in this paper is the so-called dynamic NSS model mentioned above.

Another possible solution to the problem mentioned above centers around the recent general consensus that has emerged in the literature stating that should look beyond the cross section of yields (as done in DNS models, for example) to pin down the dynamic behavior of interest rates (Duffee et al. 2012). Along these lines, modeling the co-movements of the underlying economy by specifying diffusion indexes (Stock and Watson, 2002a) or using key macroeconomic indicators has proven useful in predicting government bond yields. For example, using dynamic factor model framework, Coroneo et al. (2016) find that real economic activity and real interest rates contain predictive content for government bond yields that are not spanned by the cross-section of yields. Ang and Piazzesi (2003) and Mönch (2008) also report improved forecasts using affine models which include principal component-based macro factors. 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. (2014)). Ludvigson and Ng (2009) find that adding macro factors helps when forecasting bond risk premia. As discussed above, our empirical analysis adds to this literature by assessing the marginal predictive content of real-time diffusion indexes constructed in a data-rich environment. These indexes are used alone and as inputs into other models including DNS, NSS, and benchmark linear models; and predictions of yields at various maturities and for various forecast horizons are constructed. Additionally, a number of forecast combinations are examined, in which various permutations of our individual models are combined. Finally, results are tabulated for various sub-samples between 2001 and 2018 in order to assess whether model rankings are dependent upon sample period.

A number of clear-cut conclusions can be made based on our experiments. First, the use of real-time diffusion indexes increases the marginal predictive performance of all individual models considered in our analysis, for sample periods ranging from 2001 through 2010. Second, after 2010 no "data-rich" prediction models can beat an AR(1) benchmark. Third, DNS and NSS type models are the "best" MSFE "performers" over the 2001 through 2010 period. Fourth, forecast combination models that combine all models that do not include diffusion indexes (i.e., DNS, NSS, and benchmark models) yield the lowest overall MSFEs, dominating all other models across all sample periods, forecast horizons and bond maturities. This result is in contrast to the findings of Swanson and Xiong (2017), where including diffusion indexes always yields predictive improvement, although only fully revised macroeconomic data are utilized in that paper. Thus, the usefulness of diffusion indexes appears to hinge on whether or not a data-rich real-time environment is simulated in forecasting experiments or not.

The rest of the paper is organized as follows. Section 2 describes the Dynamic Nelson Siegel and Svensson models, and Section 3 discusses yield curve prediction with added macroeconomic diffusion indexes. Section 4 includes details describing our empirical setup, and discusses our empirical findings. Concluding remarks are gathered in Section 5.

3.2 The Dynamic Nelson Siegel Model

3.2.1 Three-factor Dynamic Nelson Siegel Model

Motivated by rational expectation theory, Nelson and Siegel (1985) 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 maturity of a bond can be written as the average of forward rates

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

Nelson and Siegel (1985) motivate the use of the following model of the forward rate that can generate monotonically increasing, humped, and occasionally S-shaped yield curves, a range of shapes for yield curves:

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

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:

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

In the above model, the latent 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. For example, with our 12-dimensional cross section, we estimate the latent factors by fitting the following regression:

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

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 maturity goes to infinity. Figures 3A - 3B exhibit the three NS factors estimated with ordinary least squares and non-linear least squares methods for sample period 1988:8 - 2017:10.[1] In summary, the DNS model can be written as follows:

$$\hat{y}_t(\tau) = \hat{\beta}_{1,t} + \hat{\beta}_{2,t} \cdot [\frac{1-\exp(-\lambda_t\tau)}{\lambda_t\tau}] + \hat{\beta}_{3,t} \cdot [\frac{1-\exp(-\lambda_t\tau)}{\lambda_t\tau} - \exp(-\lambda_t\tau)]. \tag{3.1}$$

In order to construct predictions using the DNS model, we fit estimated factors to AR and VAR models, as follows.

$$\hat{\beta}_{i,t+1} = c_i + \gamma_i\hat{\beta}_{i,t} + \epsilon_t \quad i = 1, 2, 3 \quad \text{or,} \tag{3.2}$$

[1] An increase in the "level" component, β_{1t}, affects all yields equally, thus it determines the level of the yield curve. Also, as maturity τ goes to infinity, $\beta_{1t} = y_t(\infty)$ by definition. An increase in "slope" component β_{2t} affects short rates more than long rates, thereby changing the slope, or the so-called "term spread" of the yield curve. Finally, an increase in β_{3t}, the "curvature" component, will increase medium-term yields and have little effect on the short and long end of the curve. Therefore, the yield curve will becomes more hump shaped. 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.

$$\hat{\boldsymbol{\beta}}_{t+1} = \mathbf{c} + \boldsymbol{\Gamma}\boldsymbol{\beta}_t + \boldsymbol{\epsilon}_t, \tag{3.3}$$

where ϵ_t is a scalar stochastic disturbance term, $\boldsymbol{\epsilon}_t$ is a 3×1 vector of stochastic disturbance terms, and c_i, $\mathbf{c}$, γ_i, and $\boldsymbol{\Gamma}$, $i = 1, ..., 3$, are conformably defined constants, constant vectors and constant matrices. With these last two models, one con construct predictions of the $\hat{\beta}_{i,t}$, for $i = 1, ..., 3$, which can in turn be inserted into the above model of $\hat{y}_t(\tau)$ in order to generate predictions thereof. In all experiments in the sequel, rolling estimation is carried out when estimating the above models (and all other models), using windows of length 120 months, so that "real-time" predictions are constructed in all cases. Additionally, we consider two types of prediction models. In one, the decay paraemter 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 discussios comparing arbitrage free dynamic latent factor and DNS models, see Ang and Piazzesi (2003), Diebold, Rudebusch and Aruoba (2006), Christensen, Diebold, and Rudebusch (2011), Duffie (2011), and the references cited therein.

3.2.2 Four-factor Nelson-Siegel-Svensson Model

Svensson (1994) extends the Nelson-Siegel Svensson (NSS) model by adding a fourth term, that allows for a second "hump" shape. In particular, he proposed the following four-factor model for fitting the instantaneous forward interest rate:

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

Notice that in the above equation there are now two different decay parameters controlling the double-hump shape of the forward curve, called θ_1 and θ_2. Similar to the DNS model, we consider a dynamic version of the NSS model. Thus, we utilize the following variant of the DNS model (factor estimation and prediction construction is carried out using the DNS modeling approach discussed above).

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

where we now have two decay parameters, as discussed above. These are called $\lambda_{1,t}$ and $\lambda_{2,t}$. As discussed in De Pooter (2007), the second hump in the NSS model is difficult to identify without imposing additional 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. Figures 4A - B plots the four NSS factors estimated with static and dynamic decay parameters, $\lambda_{1,t}$ and $\lambda_{2,t}$. Figure 3.4C plots estimated rates of decay used in the construction of the four Nelson-Siegel-Svensson factors, where the rates of decay $(\lambda_{1,t}, \lambda_{2,t})$ are either set to fixed numbers, or estimated recursively using nonlinear least squares. See Section 3.1.2 for details on model estimation.

3.3 Unspanned Risks and Diffusion Indexes

Whether or not macroeconomic, financial and other non-yield information is useful in fitting and forecasting the yield curve remains an open question. As Duffee (2012) points out, yields are usually hypothesized to follow Markov a 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 the above forecast equation 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. In particular, Ang and Piazzesi (2012) find that macroeconomic variables are significant for explaining Treasury security yield dynamics, based on a VAR analysis. Mönch (2008) shows that adding estimated diffusion indexes to an affine Gaussian term structure model improves out-of-sample forecast performance. Diebold, Rudebusch, and Aruoba (2006) investigate the bidirectional causality between yield betas and macro variables and discover strong evidence in favor of linkages between macroeconomic variables and future yield curve dynamics.

Recently, focus has turned to so-called big data, and to the analysis of the usefulness of largescale datasets in the above context. As noted in Bernanke (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. For example, the predictive usefulness of diffusion indexes constructed using largescale 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. One important aspect of big data in our context is the use of so-called real-time data, as discussed in the introduction. 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 over 130 macro-variables and all revisions of all of these variables. The dataset contains variables summarizing economic output and income, labor markets, consumption, money and credit, housing, and stock market, for example. Moreover, they show that diffusion indexes extracted from their FRED-MD dataset contain the same predictive content as diffusion indexes constructed using the classic Stock and Watson dataset (Stock and Watson (2002)). 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.

In this paper, we ask the following question: Are diffusion indexes 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, (2017), for example. We motivate the use of diffusion indexes by adopting a dynamic factor model framework resembling that used by Coroneo et al. (2016) and many others. Namely, we assume that yields curve factors, (which are the betas in the above discussion are are here called $F_{y,t}$), are driven by both past yield curve factors and macro factors, called $F_{x,t}$. Additionally, it is assumed that macroeconomic variables are driven only by $F_{x,t}$ only. Thus, we posit the following model:

$$\begin{pmatrix} F_{y,t+h} \\ x_t \end{pmatrix} = \begin{pmatrix} c_y \\ c_x \end{pmatrix} + \begin{bmatrix} \Gamma_y & \Gamma_x \\ 0 & \Gamma_{xx} \end{bmatrix} \begin{pmatrix} F_{y,t} \\ F_{x,t} \end{pmatrix} + \begin{pmatrix} e_{y,t+h} \\ e_{x,t} \end{pmatrix},$$

where c_y, c_x are vectors containing constant terms, h is the forecast horizon, Γ_y contains factor loadings on yield factors, Γ_{xx} contains factor loadings on the macro factors, and Γ_x summarizes the marginal effect of macro factors on yield factors. Additionally, $e_{y,t+h}$ and $e_{x,t}$ are idiosyncratic stochastic disturbance terms. In their paper, Coroneo et al. (2016) use a so-called expectation conditional restricted maximization algorithm for model estimation, and measure the effect of "unspanned" macroeconomic variables (risks) on the yield curve. We use principal component analysis (PCA) for estimating our macro diffusion indexes (i.e., macro factors), following Stock and Watson (2002a,b), and consider various alternative models that utilize macro diffusion indexes. For instance, we examine whether adding macro diffusion indexes to our DNS and NSS models improves the predictive accuracy of these models. Of course, we also consider baseline DNS (or NSS) models that contain only yield factors. More concretely, $h-$step ahead predictions for yield factors are constructed using the following model:

$$\widehat{F}^f_{y,t+h} = \hat{c}_y + \hat{\Gamma}'_y \widehat{F}_{y,t}, \tag{3.4}$$

where $\widehat{F}_{y,t}$ is our estimated DNS (or NSS) latent factor (i.e. $\widehat{F}_{y,t}$ are our betas in the above discussion), $\widehat{F}^f_{y,t+h}$ is our prediction constructed by specifying simple AR(1) or VAR(1) models, $\hat{c}_y$ is an estimate of c_y, and $\hat{\Gamma}_y$ is an estimate of Γ_y. We additionally add the first r_x principle components from a PCA analysis of our real-time dataset, denoted as $\widehat{F}_{x,t}$, to the above prediction model, yielding:

$$\widehat{F}^f_{y,t+h} = \hat{c}_y + \hat{\Gamma}'_y \widehat{F}_{y,t} + \hat{\Gamma}'_x \widehat{F}_{x,t} \tag{3.5}$$

where $\hat{\Gamma}_x$ is an estimate of Γ_x. When predicting yields, in addition to utilizing DNS and NSS models, we also examine whether adding macro diffusion indexes to benchmark AR and VAR models improves predictive accuracy. In particular, we consider the following model:

$$\begin{pmatrix} y_{t+h} \\ x_t \end{pmatrix} = \begin{pmatrix} c \\ c_x \end{pmatrix} + \begin{bmatrix} \Delta_y & \Delta_x \\ 0 & \Gamma_{xx} \end{bmatrix} \begin{pmatrix} y_t \\ F_{x,t} \end{pmatrix} + \begin{pmatrix} e_{t+h} \\ e_{x,t} \end{pmatrix},$$

where c is the vector containing constant terms, all coefficient matrices (i.e., Δ_y, Δ_x, and Γ_{xx}) are a conformably defined coefficient matrices, Δ_x summarizes the marginal effect of macro diffusion indexes on yields, and e_{t+h} is an idiosyncratic stochastic disturbance term. Summarizing, our focus of interest is on h-step ahead yield predictions constructed using the following model:

$$\hat{y}_{t+h}(\tau) = \hat{c}(\tau) + \hat{\delta}'_y y_t, \tag{3.6}$$

where $\hat{c}(\tau)$ is an estimate of $c(\tau)$, which is an element of c. Also, $\hat{\delta}_y$ is an estimate of δ_y, which is a row vector of Δ_y. y_t contains lags of $y_{t+1}(\tau)$ in autoregressive specifications, and contains lags of y_{t+1} in vector autoregressive specifications. We additionally add the macro diffusion indexes discussed above, F^x_t, to this model, yielding:

$$\hat{y}_{t+h}(\tau) = \hat{c}(\tau) + \hat{\delta}'_y y_t + \hat{\delta}'_x \widehat{F}_{x,t}, \tag{3.7}$$

where $\hat{\delta}_x$ is an estimate of δ_x, which is a row vector of Δ_x.

3.4 Empirical Results

We carry out an empirical investigation that utilizes the various models and methods discussed above. Our objective is to predict U.S. Treasury yields at various maturities (i.e., the term structure of interest rates). Predictions are made using "small data" models (i.e. models which only use historical yield cross sections for calibrating models), as well as "big data" models that include diffusion indexes constructed from the real-time FRED-MD dataset discussed above. Our small data models include autoregressive, vector autoregressive, dynamic Nelson-Siegel and dynamic Nelson-Siegel-Svensson models, and our "big data" models are specified as pure diffusion index models or as hybrids that cobmine our small data models with diffusion indexes. In the remaining subsections, we summarize our empirical setup, the data used in our analysis, and our experimental findings.

3.4.1 Empirical Setup

Predictive Accuracy Testing

When comparing the predictive performance of our models, we report the mean square forecast error (MSFE), defined as:

$$\text{MSFE}_h(\tau) = \sum_{t=1}^{P} (\hat{y}_{t+h}(\tau) - y_{t+h}(\tau))^2 \tag{3.8}$$

where $\hat{y}_{t+h}(\tau)$ is the h-step-ahead forecast of the Treasury bond yield, with maturity τ. P is the number of ex ante predictions used in our analysis. Additionally, all model parameters are estimated with maximum likelihood and PCA; and parameters are updated prior to the construction of each forecast using a rolling window of 120 months of historical data. Model MSFEs are compared using the predictive accuracy test introduced by Diebold and Mariano (1995). The null hypothesis of the DM test is: $H_0 : \text{E}[L(\epsilon_{t+h}^{(1)})] - \text{E}[L(\epsilon_{t+h}^{(2)})] = 0$, where the $\epsilon_{t+h}^{(i)}$ is the prediction error of model i, for $i = 1, 2$. In our analysis, $L(.)$ is a quadratic loss function. The DM test statistic is:

$$\text{DM}_P(\tau) = P^{-1} \sum_{t=1}^{P} \frac{d_{t+h}(\tau)}{\hat{\sigma}_{\bar{d}}} \tag{3.9}$$

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. If the DM_P statistic has a negative value, Model 1 is preferred to Model 2. If the DM_P statistic is significantly different from zero, the difference between Model 1 and Model 2 is statistically significant. 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).

Models Used in Forecasting Experiments

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

Small Data Models

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

Models in this section are summarized in Table 3.1, and include: AR(1), VAR(1), AR(SIC), and VAR(SIC).

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

$$y_{t+h}(\tau) = c(\tau) + \delta_y' W_t + \varepsilon_{t+h}, \tag{3.10}$$

where τ denotes the maturity of a 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. $c(\tau)$ contains the constant term. [2] In AR and VAR specifications, up to 5 lags of $y_t(\tau)$ 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 3.1, and include: DNS(1), DNS(2), DNS(3), DNS(4), DNS(5), and DNS(6).

As discussed above, the DNS model introduced by Li and Diebold (2006) is a dynamic version of the term structure based upon 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:

$$y_t(\tau) = \beta_{1,t} + \beta_{2,t} \cdot [\frac{1-\exp(-\lambda_t\tau)}{\lambda_t\tau}] + \beta_{3,t} \cdot [\frac{1-\exp(-\lambda_t\tau)}{\lambda_t\tau} - \exp(-\lambda_t\tau)] + \varepsilon_t, \tag{3.11}$$

which is discussed in Section 2.1. Again 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:

$$\widehat{y}_{t+h}(\tau) = \widehat{\beta}^f_{1,t+h} + \widehat{\beta}^f_{2,t+h} \cdot [\frac{1-\exp(-\lambda_t\tau)}{\lambda_t\tau}] + \widehat{\beta}^f_{3,t+h} \cdot [\frac{1-\exp(-\lambda_t\tau)}{\lambda_t\tau} - \exp(-\lambda_t\tau)], \tag{3.12}$$

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

$$\widehat{\beta}^f_{i,t+h} = \hat{c}_i + \hat{\gamma}_{y,i}\widehat{\beta}_{i,t}, \quad \text{for } i = 1, 2, 3, \tag{3.13}$$

where $\hat{\beta}^f_{i,t+h}$, $\widehat{\beta}_{i,t}$, $\hat{c}_i$ and $\hat{\gamma}_i$ are scalars. Note that $\hat{c}_i$ is an element of $\hat{c}_y$ in equation (3.1). Also, $\hat{\gamma}_{y,i}$ is an element of $\hat{\Gamma}_y$ as defined in equation (3.1). We also construct predictions by using the following

[2] When specifying VAR models, equation (4.3) 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 SUR (seemingly unrelated regression) result ensures that consistent and efficient parameter estimates can be obtained via application of equation by equation least squares.

VAR(1) model:

$$\widehat{\boldsymbol{\beta}}^f_{t+h} = \hat{c}_y + \hat{\Gamma}_y \widehat{\boldsymbol{\beta}}_t, \tag{3.14}$$

where $\widehat{\boldsymbol{\beta}}^f_{t+h} = (\hat{\beta}^f_{1,t+h}, \hat{\beta}^f_{2,t+h}, \hat{\beta}^f_{3,t+h})'$, $\hat{c}$ is a 3×1 vector, and $\hat{\Gamma}_y = (\widehat{\boldsymbol{\gamma}}_1, \widehat{\boldsymbol{\gamma}}_2, \widehat{\boldsymbol{\gamma}}_3)$, with $\widehat{\boldsymbol{\gamma}}_j$ 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 3.1), 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), 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.[3] 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 3.1, and include: NSS(1), NSS(2), NSS(3), NSS(4), NSS(5), and NSS(6).

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 consider 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:

$$\begin{aligned} y_t(\tau) = \beta_{1,t} + \beta_{2,t} \cdot [\frac{1-\exp(-\lambda_{1,t}\tau)}{\lambda_{1,t}\tau}] + \beta_{3,t} \cdot [\frac{1-\exp(-\lambda_{1,t}\tau)}{\lambda_{1,t}\tau} - \exp(-\lambda_{1,t}\tau)] \\ + \beta_{4,t} \cdot [\frac{1-\exp(-\lambda_{2,t}\tau)}{\lambda_{2,t}\tau} - \exp(-\lambda_{2,t}\tau)] + \varepsilon_t, \end{aligned} \tag{3.15}$$

[3] We 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).

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

$$\begin{aligned}\widehat{y}_{t+h}(\tau) = \widehat{\beta}^f_{1,t+h} + \widehat{\beta}^f_{2,t+h} \cdot [\frac{1-\exp(-\lambda_{1,t}\tau)}{\lambda_{1,t}\tau}] + \widehat{\beta}^f_{3,t+h} \cdot [\frac{1-\exp(-\lambda_{1,t}\tau)}{\lambda_{1,t}\tau} - \exp(-\lambda_{1,t}\tau)] \\ + \widehat{\beta}^f_{4,t+h} \cdot [\frac{1-\exp(-\lambda_{2,t}\tau)}{\lambda_{2,t}\tau} - \exp(-\lambda_{2,t}\tau)]\end{aligned} \tag{3.16}$$

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

$$\widehat{\beta}^f_{i,t+h} = \hat{c}_i + \hat{\gamma}_{y,i}\widehat{\beta}_{i,t}, \quad \text{for } i = 1, 2, 3, 4 \tag{3.17}$$

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

$$\widehat{\boldsymbol{\beta}}^f_{t+h} = \hat{c}_y + \hat{\Gamma}_y\widehat{\boldsymbol{\beta}}_t, \tag{3.18}$$

where $\widehat{\boldsymbol{\beta}}^f_{t+h} = (\hat{\beta}^f_{1,t+h}, \hat{\beta}^f_{2,t+h}, \hat{\beta}^f_{3,t+h}, \hat{\beta}^f_{4,t+h})'$, $hatc_y$ is 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.[4] The subsequent forecasting procedure used to construct yield predictions is the same as that discussed above for our DNS models.

Big Data Models

AR and VAR Models with Macro Diffusion Indexes:

Models in this section are summarized in Table 3.1, and include: AR(1)+FB1, AR(1)+FB2, AR(1)+FB3, VAR(1)+FB1, VAR(1)+FB2, and VAR(1)+FB3.

We utilize the prediction model given in equation (4.3), but with latent factors (i.e., diffusion indexes), F^x_t, estimated using PCA with a real-time macroeconomic dataset (see Section 3 for a discussion of diffusion indexes and Section 4.2 for a discussion of the data used in our analysis). In particular, we estimate variants of the following factor augmented forecasting model:

$$y_{t+h}(\tau) = c(\tau) + \delta'_y W_t + \delta'_x F^x_t + \varepsilon_{t+h}, \tag{3.19}$$

[4]Restrictions on the decay parameters for the NSS model are imposed to ensure that the two curvature humps are at least one year apart, for identification purposes. In addition to this 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 on (0, 26.77). These restrictions ensure identification of two curvature factors individually, and avoids 'extreme' estimates for NSS-factors.

where F_t^x includes either 1, 2 or 3 diffusion indexes, and W_t is defined as above, yielding AR and VAR variants of these models. Here, $c(\tau)$ is a constant term, and δ_y and δ_x are conformably defined vectors of coefficients, as discussed in Section 3. Note that although multiple yield lags were tried when specifying W_t, '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.

DNS Models with Macro Diffusion Indexes:

Models in this section are summarized in Table 3.1, 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.[5]

In this section, diffusion indexes (principle components) constructed using macro variables are augmented to DNS models discussed above. Namely, we considered DNS type predictions constructed using:

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

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

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

where $\widehat{\boldsymbol{\beta}}^f_{t+h} = (\widehat{\beta}^f_{1,t+h}, \widehat{\beta}^f_{2,t+h}, \widehat{\beta}^f_{3,t+h})')$, $\widehat{c_y}$ is 3×1 vector, $\hat{\Gamma}_y = (\widehat{\boldsymbol{\gamma}}_1, \widehat{\boldsymbol{\gamma}}_2, \widehat{\boldsymbol{\gamma}}_3)$, with $\widehat{\boldsymbol{\gamma}}_j$ 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 3.1, 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 our DNS models, NSS model predictions constructed using:

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

where F_t^x includes either 1, 2 or 3 latent factors. All other terms are conformably defined and analogous to our above discussion. We also construct predictions using the following VAR(1) variant

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

of this model:

$$\widehat{\boldsymbol{\beta}}^f_{t+h} = \hat{c}_y + \hat{\Gamma}_y \widehat{\boldsymbol{\beta}}_t + \hat{\Gamma}_x F^x_t,$$

where $\widehat{\boldsymbol{\beta}}^f_{t+h} = (\widehat{\beta}^f_{1,t+h}, \widehat{\beta}^f_{2,t+h}, \widehat{\beta}^f_{3,t+h}, \widehat{\beta}^f_{4,t+h})'$, $\hat{c}$ is 4×1 vector, and $\hat{\Gamma}_y = (\widehat{\boldsymbol{\gamma}}_1, \widehat{\boldsymbol{\gamma}}_2, \widehat{\boldsymbol{\gamma}}_3, \widehat{\boldsymbol{\gamma}}_4)$, $\widehat{\boldsymbol{\gamma}}_j$ 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 series before principle component analysis is utilized to construct diffusion indexes.

Forecast Combination:

In our experiments, we also construct and analyze various forecast combination models. The particular combinations are detailed in Table 3.8. Although the focus of this paper is not forecast combination, 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 via combination not only of forecasts, but also of methodologies.[6] 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 Xiong and Swanson (2017), 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.

Finally, note that in all experiments, models are estimated using rolling windows of 120 monthly observations, as discussed above. Thus, all models are re-estimated prior to the construction of each new h-step ahead forecast. Additionally, the first observation in our dataset is August 1988, and experiments are carried out for 4 different prediction periods, as discussed in the next section.

3.4.2 Data

Yield 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, Sack and Wright (GSW: 2006)). In particular, we utilize GSW monthly data for the August 1988 through October 2017, which contains data on 1

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

to 30 years maturity bond yields. In addition to GSW zero-yields, 3- and 6-months T-bill yields[7] are utilized in order to "fill-out" the short end of the yield curve. Hence, we analyze a panel of dataset containing $N = 32$ dimensional yields and $T = 351$ monthly observations. When constructing betas, we consider three variants of this data. In one variant, we utilize 12 yields (i.e., 3- and 6-months, 1 year, 2 year, ..., and 10 year yields); in a second variant, we utilize 10 yields, as done in Xiong and Swanson (2017) (i.e., 1 year, 2 year, ..., and 10 year yields); and in a third variant we utilize 30 yields (i.e., 1 year, 2 year, ..., and 30 year yields).

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 differences when comparing MSFEs of yield predictions. Moreover, finance theory is not consistent with the presence of a unit root in yield processes. For these reasons, we use only yield "levels" data in our experiments.

Macroeconomics Variables:

Macroeconomic factors (i.e., 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 that covers categories ranging from output and income, to labor market, prices, and interest rates. The FRED-MD dataset is developed and maintained by the Federal Reserve Bank of St. Louis. For details, see McCracken and Ng (2016), and for access to the dataset, visit *https://research.stlouisfed.org/econ/mccracken/fred-databases/*. In their paper, McCracken and Ng (2016) conduct an empirical research which shows that diffusion indexes extracted from the dataset share the same predictive content as those based on the classic (non-real-time) Stock and Watson (2002) dataset used so frequently in analyses such as ours. As discussed above, one advantage of FRED-MD 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. Moreover, fully revised datasets "mix" vintages of observations, in the sense that the most recent observation in a fully revised dataset is a so-called "first release", while earlier calendar dated observations have possibly been revised and re-released many times. Interestingly, we find that real-time data does matter, as findings from Xiong and Swanson (2017) supporting the use of macro diffusion indexes in DNS type models are reversed when real-time instead of fully revised data are used in index construction. Finally, note

[7] 3- 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 is widely utilized in yield curve estimation, see Gürkaynak and Wright (2012), Hamilton and Wu (2012).

that only the use of real-time datasets makes it possible to replicate truly real-time modeling and forecasting of U.S. Treasury yields, when using macroeconomic data that are subject to revision.

3.4.3 Empirical Findings

A number of clear-cut conclusions emerge upon examination of the results collected in Tables 3.1 - 3.9C.

First, let us turn to a discussion of the results in Tables 3.2A - 3.4C. Tables 3.2A - 3.2C include: (i) the three"MSFE-best" (i.e., lowest MSFE model) models for each yield maturity/forecast horizon permutation, in descending order from 1st to 3rd (see Table 3.2A); (ii) MSFE for the three models listed in Table 3.2A (see Table 3.2B); and relative MSFEs (relative to the AR(1) benchmark) for the 3 models (see Table 3.2C). All of these tables report results for DNS and NSS type models with betas constructed using our 12-dimensional historical yield dataset (see Section 4.2 for details), and results are presented for 3 forecast horizons ($h = 1, 3, 12$), for 6 yield maturities (3- and 6-month, 1 year, 3 years, 5 years, and 10 years), and for 4 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 (Subsample 4). Analogous results, but with different historical yield datasets are reported in Tables 3A-C (for our 12-dimensional historical yield dataset) and Tables 4A-C (for our 12-dimensional historical yield dataset). In this first set of tables, the MSFE-best across all 3 historical yield datasets, and for each maturity, horizon, and subsample is denoted in bold. Thus, for example, we see in Table 3.5A that there are many bolded entries, indicating that the 12-dimensional historical yield dataset is yielding many MSFE-best models. For example, at the $h = 1$-step ahead horizon, and for $\tau = 3$-months, 1 year and 3-years, the MSFE best models (which are all based on the 12-dimensional historical yield dataset) are VAR(1)+FB2, AR(1)+FB2, and DNS(1)+FB1, respectively. (See Table 3.1 for a list of model mnemonics.) All of these models include diffusion indexes, two include linear benchmark type AR and VAR models, and one utilizes a DNS type model.

Notice that many of the models that are MSFE-best in Tables 3.2A - 3.4C include diffusion indexes (i.e., models with FB1, FB2, or FB3 in their names). Additionally, many models are of the DNS and NSS variety. This pattern prevails across all forecast horizons and bond maturities. Thus, we have strong evidence that DNS and NSS models are useful, as previously found by many authors, and that the usefulness of all models including DNS type models can often be improved by including diffusion indexes. This result is not completely ubiquitous, however. For example, in the third subsample (post Great Recession), diffusion indexes are not in any MSFE-best models, when considering our 12-dimensional historical yield dataset (see Table 3.2A), indicating a deterioration in predictive gains associated with using diffusion indexes, post Great Recession. However, they are in the 2nd and 3rd "best" models, in many for many yield maturities. Thus, the usefulness of diffusion indexes appears to be somewhat sample dependent. This is not surprising, and suggests, for example, a possible use for hybrid models in which the inclusion of diffusion indexes is triggered by variables such as predicted probabilities of recessions, economic variability, or the range of yields over some pre-defined prior period of time.

As stated above, Tables 3.2A - 3.4D indicate that DNS or NSS type models are often the MSFE-best models. In Table 3.2A DNS or NSS type models are MSFE-best in 12 of 15 maturity/horizon permutations for Subsample 1, and are top three MSFE performers in 37 of 45 maturity/horizon permutations. However, in Subsample 2, the results deteriorate slightly, where analogous "wins" are 10 of 15 and 28 of 45. Finally, in Subsample 3, DNS or NSS type models "win" in only 7 of 15 cases and 19 of 45 cases, respectively. These are consistent with findings in the recent literature suggesting that DNS type model performance has deteriorated in recent post credit crisis years (see, e.g. Altavilla, Giacomini and Ragusa (2014), Diebold, and Rudebusch (2012), and Mönch (2008)).

On a different note, DNS and NSS model performance is optimized when our 12-dimensional historical yield dataset is used to construct historical betas. Adding further information from the long end of the yield curve appears, thus, to add more noisiness than information when estimating betas. This is not altogether surprising, given the well known difficulty in "pinning-down" the long end of the yield curve in empirical settings.

Second, let us turn to a discussion of the results in Tables 3.5A - 3.7D. Tables 3.5A - 3.5D contain relative MSFEs (relative to the AR(1) benchmark model) for all models (rather than just the top 3) examined in our 12-dimensional historical yield dataset for Subsample 1 (Table 3.5A), Subsample 2 (Table 3.5B), Subsample 3 (Table 3.5C), Subsample 4 (Table 3.5D), for the $h = 1$ case. Analogous results for $h = 4$ are contained in Tables 6A-6D, and for $h = 12$ results are contained in Tables 7A-7D. In Tables 5A-7D, the results of DM_P tests are also reported, where the benchmark is the AR(1) model, and the alternative is model listed in the first column of the tables. In particular, starred entries indicate rejection of the (DM_P test) null hypothesis of no difference between the benchmark and the alternative model.[8] Similar to Tables 2A-4C, in these tables, the "MSFE-best" (i.e., lowest MSFE model) for the 12-dimensional historical yield dataset for each maturity, horizon, and subsample is denoted in bold. Thus, for example, we see in Table 3.5A that the MSFE-best model for τ = 3-months is VAR(1)+FB2, which is our VAR(1) model with 2 diffusion indexes. The key take-away from these tables is that many of our models are significantly better than our AR(1) benchmark. This is particularly true for our first and second subsamples, when comparing

[8] *** entries denote rejection at the 1% level, while ** and * denote rejection at the 5% and 10% levels, respectively.

predictions for $h = 1$ and $h = 3$, and for all subsamples when comparing predictions for $h = 12$. On the other hand, there is less to choose between models when comparing predictive performance during subsample 3, for $h = 1$ and $h = 3$. This suggests that it is not only diffusion index based models that perform poorly, post Great Recession (see below discussion), but all models perform more poorly, relative to our simple AR(1) benchmark model, post recession. Thus, is is perhaps not surprising that DNS type models do not perform as well as previously, as no models seem to. This in turn indicates the need for the development of new models (such as the hybrid models discussed above) for addressing the unique set of economic conditions that characterize the post Great Recession period.

Third, note that Table 3.8 lists the forecast combination models that were utilized in our experiments. Relative MSFEs for all combinations are given in Table 3.9A (h = 1-step ahead forecasts), Table 3.9B (h = 3-step ahead forecasts), and Table 3.9C (h = 12-step ahead forecasts). Results in these tables are quite interesting. For instance, models with diffusion indexes (e.g., FB1, FB2, and FB3) often significantly outperform the AR(1) benchmark, for all five maturities and all three forecast horizons in Subsamples 1, 2, and 4. However, results are mixed for Subsample 3 (2011:1-2017:10), again indicating a deterioration in predictive gains associated with using diffusion indexes, post Great Recession. Another conclusion emerges when comparing NS(AR) and NS(VAR) forecast combination models. As discussed previously, NS-factors and NSS-factors can be better predicted by modeling their cross-correlation dynamics. This is borne out in the data. For example, in Subsample 1, 3, and 4, forecasts generated by combination model NS(VAR) have lower point MSFEs than NS(AR) models in 43/45 cases, across all three forecast horizons, and across all five bond maturities.

A key result to emerge from our combination experiments is that the MSFE-best models are almost always the "FS" type forecast combinations. As noted in Table 3.8, FS forecast combination models utilize the average of 16 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 19 of 20 cases, for $h = 1$, across all five bond maturities and all four subsamples (see Table 3.9A). For $h = 3$ and $h = 12$, FS combination model again "win" in 19 of 20 cases (see Tables 9B and 9C). Furthermore, all of the "best" MSFEs are much lower than point MSFEs associated with the best individual models. Thus, combination dominates under our real-time setup, and the best combinations do not utilize macro diffusion indexes. This result differs markedly from that reported in Xiong and Swanson (2017), where big data matters. However, Xiong and Swanson (2017) carries out experiments using a fully revised macroeconomic dataset rather than a real-time macroeconomic dataset. This indicates that fully revised data may have an important confounding effect upon results obtained be carrying out real-time prediction experiments. Examination of this issue using a richer set of individual and combination models is left to future research.

3.5 Concluding Remarks

In this paper, we examine the usefulness of real-time macroeconomic diffusion indexes when using dynamic Nelson-Siegel (DNS), dynamic Nelson-Siegel Svensson (NSS), and various econometric models for forecasting the term structure of interest rates. We find that the marginal predictive content of real-time diffusion indexes is significant for many of the models that we examine. We also find that model performance, across the board, is much worse post Great Recession. Indeed, not only does the predictive performance of DNS and NSS models worsen, in accord with the findings of various recent authors, but the performance of all of our models, including ones that utilize real-time diffusion indexes also worsens. Given the impressive predictive performance of these models prior to the great recession, we argue that new models need to be developed to address current economic conditions. Examples of models that might be useful include hybrid models in which the inclusion of diffusion indexes is triggered by variables such as predicted probabilities of recessions, economic variability, or the range of yields over some pre-defined prior period of time. We also present strong new evidence of the usefulness of forecast combination, and note that mean square error "best" (MSFE-best) forecast combinations often preclude the use of real-time diffusion indexes. This differs from earlier findings by Xiong and Swanson (2017), where it is found that if fully revised macroeconomic data are instead used in constructing diffusion indexes, then combinations that include diffusion index type models are MSFE-best.

Table 3.1: Models Used in Forecast Experiments*

Model	Description
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 one diffusion index added, principle component analysis based on real-time macroeconomic dataset
AR(1)+FB2	AR(1) model with two diffusion indexes added, principle component analysis based on real-time macroeconomic dataset
AR(1)+FB3	AR(1) model with three diffusion indexes added, principle 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, principle component analysis based on real-time macroeconomic dataset
VAR(1)+FB2	VAR(1) model with two diffusion indexes added, principle component analysis based on real-time macroeconomic dataset
VAR(1)+FB3	VAR(1) model with three diffusion indexes added, principle 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)+FB1	DNS(1) model with one diffusion index added, principle component analysis based on real-time macroeconomic dataset
DNS(2)+FB1	DNS(2) model with one diffusion index added, principle component analysis based on real-time macroeconomic dataset
DNS(3)+FB1	DNS(3) model with one diffusion index added, principle component analysis based on real-time macroeconomic dataset
DNS(4)+FB1	DNS(4) model with one diffusion index added, principle component analysis based on real-time macroeconomic dataset
DNS(5)+FB1	DNS(5) model with one diffusion index added, principle component analysis based on real-time macroeconomic dataset
DNS(6)+FB1	DNS(6) model with one diffusion index added, principle component analysis based on real-time macroeconomic dataset
DNS(1)+FB2	DNS(1) model with two diffusion indexes added, principle component analysis based on real-time macroeconomic dataset
DNS(2)+FB2	DNS(2) model with two diffusion indexes added, principle component analysis based on real-time macroeconomic dataset
DNS(3)+FB2	DNS(3) model with two diffusion indexes added, principle component analysis based on real-time macroeconomic dataset
DNS(4)+FB2	DNS(4) model with two diffusion indexes added, principle component analysis based on real-time macroeconomic dataset
DNS(5)+FB2	DNS(5) model with two diffusion indexes added, principle component analysis based on real-time macroeconomic dataset
DNS(6)+FB2	DNS(6) model with two diffusion indexes added, principle component analysis based on real-time macroeconomic dataset
DNS(1)+FB3	DNS(1) model with three diffusion indexes added, principle component analysis based on real-time macroeconomic dataset
DNS(2)+FB3	DNS(2) model with three diffusion indexes added, principle component analysis based on real-time macroeconomic dataset
DNS(3)+FB3	DNS(3) model with three diffusion indexes added, principle component analysis based on real-time macroeconomic dataset
DNS(4)+FB3	DNS(4) model with three diffusion indexes added, principle component analysis based on real-time macroeconomic dataset
DNS(5)+FB3	DNS(5) model with three diffusion indexes added, principle component analysis based on real-time macroeconomic dataset
DNS(6)+FB3	DNS(6) model with three diffusion indexes added, principle 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)+FB1	NSS(1) model with one principle component added, principle component analysis based on real-time macroeconomic dataset
NSS(2)+FB1	NSS(2) model with one principle component added, principle component analysis based on real-time macroeconomic dataset
NSS(3)+FB1	NSS(3) model with one principle component added, principle component analysis based on real-time macroeconomic dataset
NSS(4)+FB1	NSS(4) model with one principle component added, principle component analysis based on real-time macroeconomic dataset
NSS(5)+FB1	NSS(5) model with one principle component added, principle component analysis based on real-time macroeconomic dataset
NSS(6)+FB1	NSS(6) model with one principle component added, principle component analysis based on real-time macroeconomic dataset
NSS(1)+FB2	NSS(1) model with two diffusion indexes added, principle component analysis based on real-time macroeconomic dataset
NSS(2)+FB2	NSS(2) model with two diffusion indexes added, principle component analysis based on real-time macroeconomic dataset
NSS(3)+FB2	NSS(3) model with two diffusion indexes added, principle component analysis based on real-time macroeconomic dataset
NSS(4)+FB2	NSS(4) model with two diffusion indexes added, principle component analysis based on real-time macroeconomic dataset
NSS(5)+FB2	NSS(5) model with two diffusion indexes added, principle component analysis based on real-time macroeconomic dataset
NSS(6)+FB2	NSS(6) model with two diffusion indexes added, principle component analysis based on real-time macroeconomic dataset
NSS(1)+FB3	NSS(1) model with three diffusion indexes added, principle component analysis based on real-time macroeconomic dataset
NSS(2)+FB3	NSS(2) model with three diffusion indexes added, principle component analysis based on real-time macroeconomic dataset
NSS(3)+FB3	NSS(3) model with three diffusion indexes added, principle component analysis based on real-time macroeconomic dataset
NSS(4)+FB3	NSS(4) model with three diffusion indexes added, principle component analysis based on real-time macroeconomic dataset
NSS(5)+FB3	NSS(5) model with three diffusion indexes added, principle component analysis based on real-time macroeconomic dataset
NSS(6)+FB3	NSS(6) model with three diffusion indexes added, principle component analysis based on real-time macroeconomic dataset

* Notes: This table summarizes the models utilized in all forecasting experiments.

Table 3.2A: Top 3 MSFE-Best Models with Betas Calculated Using a 12-Dimensional Historical Yield Dataset*

Maturity	Horizon	3 Months	1 Year	3 Years	5 Years	10 Years
2001:Jan - 2005:Dec 1st Subsample	1 Step	**VAR(1)+FB2**	**AR(1)+FB2**	**DNS(1)+FB1**	**NSS(5)**	**NSS(5)+FB3**
		NSS(4)+FB1	AR(1)+FB3	DNS(1)	NSS(5)+FB1	NSS(5)
		NSS(4)+FB3	AR(1)+FB1	DNS(1)+FB2	NSS(5)+FB3	NSS(5)+FB1
	3 Step	**DNS(6)+FB2**	**AR(1)+FB1**	NSS(5)	NSS(5)	**NSS(5)**
		DNS(6)+FB1	VAR(SIC)	NSS(5)+FB1	NSS(5)+FB1	NSS(5)+FB1
		DNS(5)+FB2	VAR(1)+FB1	VAR(SIC)	NSS(5)+FB2	DNS(1)
	12 Step	**DNS(3)+FB1**	**DNS(3)+FB1**	DNS(6)+FB1	NSS(6)	NSS(6)
		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
2006:Jan - 2010:Dec 2nd Subsample	1 Step	**NSS(3)**	**DNS(1)**	**VAR(SIC)**	**VAR(SIC)**	NSS(1)
		NSS(2)	DNS(5)	VAR(1)	VAR(1)	DNS(1)
		VAR(1)+FB2	NSS(1)	VAR(1)+FB2	VAR(1)+FB3	AR(SIC)
	3 Step	**NSS(6)**	**DNS(5)+FB1**	**VAR(1)+FB1**	**VAR(1)+FB1**	NSS(1)
		NSS(3)	DNS(4)+FB1	VAR(SIC)	VAR(SIC)	DNS(1)
		NSS(6)+FB1	DNS(5)+FB2	VAR(1)	VAR(1)	DNS(4)
	12 Step	DNS(1)+FB1	**NSS(2)+FB1**	NSS(1)+FB1	AR(1)+FB1	DNS(4)
		NSS(2)+FB1	NSS(3)+FB1	DNS(1)+FB1	VAR(1)+FB1	DNS(4)+FB2
		NSS(3)+FB1	DNS(1)+FB1	NSS(3)+FB1	VAR(SIC)	DNS(4)+FB1
2011:Jan - 2017:Oct 3rd Subsample	1 Step	**AR(SIC)**	**AR(1)**	**AR(1)**	**AR(1)**	**DNS(1)**
		AR(1)	AR(1)+FB1	AR(1)+FB1	AR(SIC)	DNS(1)+FB1
		AR(1)+FB1	AR(SIC)	AR(SIC)	AR(1)+FB1	NSS(1)
	3 Step	**VAR(1)**	DNS(5)	**AR(1)**	**AR(1)**	DNS(4)
		VAR(SIC)	DNS(5)+FB1	DNS(4)	AR(SIC)	DNS(4)+FB3
		VAR(1)+FB1	DNS(5)+FB2	DNS(4)+FB2	NSS(5)+FB2	DNS(4)+FB2
	12 Step	**DNS(5)**	NSS(4)	DNS(4)	VAR(SIC)	NSS(6)
		DNS(6)	NSS(4)+FB2	DNS(4)+FB1	VAR(1)	NSS(6)+FB3
		DNS(5)+FB1	NSS(4)+FB1	DNS(4)+FB2	DNS(4)	NSS(6)+FB1
2001:Jan - 2017:Oct Whole Sample	1 Step	**VAR(1)+FB2**	**VAR(1)+FB2**	**AR(1)**	**AR(1)**	DNS(1)
		VAR(1)+FB3	AR(SIC)	AR(1)+FB1	AR(1)+FB1	NSS(1)
		VAR(1)+FB1	DNS(5)+FB1	AR(1)+FB2	AR(SIC)	AR(SIC)
	3 Step	**DNS(5)+FB1**	**DNS(5)+FB1**	**VAR(1)+FB1**	**VAR(SIC)**	DNS(1)
		DNS(6)+FB1	DNS(5)+FB2	VAR(SIC)	VAR(1)	DNS(4)
		DNS(5)+FB2	DNS(5)+FB3	VAR(1)	VAR(1)+FB1	DNS(4)+FB1
	12 Step	**DNS(6)**	**DNS(6)**	**VAR(1)**	VAR(1)	**NSS(6)**
		DNS(6)+FB1	DNS(6)+FB1	VAR(SIC)	VAR(SIC)	NSS(6)+FB1
		DNS(6)+FB2	DNS(6)+FB2	VAR(1)+FB1	VAR(1)+FB1	NSS(6)+FB3

* Notes: This table reports top three performing forecast models (based on MSFE), in descending order, for various subsamples, horizons, and yield maturities. For a description of the models listed in this table, refer to Section 4.1 and Table 3.1. Entries in bold denote lowest MSFE models for a given forecast horizon and yield maturity, across all three historical datasets (i.e., 10-, 12-, and 30-dimensional datasets) used in the construction of the betas utilized in DNS and NSS type models. For further discussion, refer to Section 4.3.

Table 3.2B: Point MSFEs of Top 3 MSFE-Best Models with Betas Calculated Using a 12-Dimensional Historical Yield Dataset*

	Horizon					
Maturity		3 Months	1 Year	3 Years	5 Years	10 Years
		0.029	**0.057**	**0.112**	**0.110**	**0.086**
	1 Step	0.029	0.059	0.112	0.113	0.087
		0.029	0.059	0.114	0.115	0.089
		0.217	**0.313**	0.345	0.283	**0.192**
2001:Jan - 2005:Dec	3 Step	0.218	0.316	0.349	0.289	0.197
1st Subsample		0.219	0.317	0.355	0.297	0.199
		2.067	**2.338**	1.515	0.918	0.381
	12 Step	2.126	2.559	1.520	0.931	0.391
		2.563	2.648	1.521	0.944	0.393
		0.082	**0.052**	**0.079**	**0.077**	0.092
	1 Step	0.086	0.053	0.079	0.078	0.093
		0.089	0.055	0.079	0.081	0.094
		0.316	**0.284**	**0.342**	**0.277**	0.189
2006:Jan - 2010:Dec	3 Step	0.316	0.286	0.349	0.279	0.189
2nd Subsample		0.332	0.288	0.349	0.280	0.202
		2.330	**1.580**	1.195	0.832	0.363
	12 Step	2.339	1.699	1.206	0.844	0.369
		2.367	1.799	1.220	0.846	0.371
		0.005	**0.006**	**0.020**	**0.036**	**0.052**
	1 Step	0.005	0.007	0.022	0.037	0.052
		0.006	0.007	0.022	0.038	0.053
		0.016	0.015	**0.051**	**0.109**	0.172
2011:Jan - 2017:Oct	3 Step	0.016	0.015	0.052	0.111	0.173
3rd Subsample		0.016	0.015	0.054	0.116	0.173
		0.043	0.064	0.140	0.339	0.484
	12 Step	0.044	0.064	0.141	0.340	0.487
		0.044	0.064	0.142	0.347	0.492
		0.039	**0.041**	**0.069**	**0.076**	0.075
	1 Step	0.040	0.042	0.070	0.079	0.076
		0.040	0.042	0.072	0.080	0.078
		0.182	**0.190**	**0.234**	**0.226**	0.192
2001:Jan - 2017:Oct	3 Step	0.183	0.193	0.234	0.227	0.194
Whole Sample		0.183	0.195	0.234	0.228	0.197
		1.723	**1.462**	**0.989**	0.721	**0.428**
	12 Step	1.731	1.469	0.990	0.721	0.439
		1.736	1.473	0.993	0.727	0.441

* Notes: See notes to Table 3.2A. Entires are point MSFEs.

Table 3.2C: Relative MSFEs of Top 3 MSFE-Best Models with Betas Calculated Using a 12-Dimensional Historical Yield Dataset*

	Horizon					
Maturity		3 Months	1 Year	3 Years	5 Years	10 Years
		0.488	**0.793**	**0.955**	**0.933**	**0.937**
	1 Step	0.491	0.810	0.959	0.957	0.948
		0.492	0.815	0.976	0.975	0.966
		0.583	**0.773**	0.789	0.759	**0.847**
2001:Jan - 2005:Dec	3 Step	0.585	0.780	0.797	0.775	0.869
1st Subsample		0.589	0.782	0.812	0.797	0.878
		0.424	**0.401**	0.314	0.288	0.306
	12 Step	0.436	0.439	0.315	0.293	0.314
		0.525	0.454	0.315	0.297	0.316
		0.896	**0.783**	**0.913**	**0.859**	0.906
	1 Step	0.935	0.802	0.915	0.864	0.918
		0.964	0.824	0.920	0.906	0.929
		0.883	**0.834**	**0.931**	**0.869**	0.828
2006:Jan - 2010:Dec	3 Step	0.883	0.841	0.949	0.876	0.830
2nd Subsample		0.928	0.846	0.949	0.877	0.886
		0.690	**0.592**	0.699	0.755	0.768
	12 Step	0.693	0.637	0.706	0.765	0.781
		0.701	0.674	0.714	0.767	0.785
		0.978	**1.000**	**1.000**	**1.000**	**0.972**
	1 Step	1.000	1.048	1.054	1.030	0.978
		1.119	1.091	1.074	1.055	0.986
		0.784	0.691	**1.000**	**1.000**	0.877
2011:Jan - 2017:Oct	3 Step	0.785	0.697	1.016	1.017	0.882
3rd Subsample		0.801	0.706	1.053	1.062	0.884
		0.154	0.243	0.438	0.644	0.576
	12 Step	0.157	0.244	0.444	0.647	0.580
		0.158	0.245	0.445	0.660	0.586
		0.837	**0.947**	**1.000**	**1.000**	0.953
	1 Step	0.853	0.950	1.019	1.036	0.955
		0.862	0.950	1.045	1.045	0.981
		0.809	**0.826**	**0.900**	**0.906**	0.896
2001:Jan - 2017:Oct	3 Step	0.814	0.836	0.901	0.907	0.903
Whole Sample		0.816	0.848	0.902	0.911	0.920
		0.672	**0.556**	**0.478**	0.485	**0.503**
	12 Step	0.675	0.559	0.478	0.485	0.516
		0.677	0.560	0.480	0.489	0.519

* Notes: See notes to Table 3.2A. Entries are relative MSFEs, when compared with the AR(1) benchmark model.

Table 3.3A: Top 3 MSFE-Best Models with Betas Calculated Using a 10-Dimensional Historical Yield Dataset*

	Horizon					
Maturity		3 Months	1 Year	3 Years	5 Years	10 Years
		VAR(1)+FB2	**AR(1)+FB2**	DNS(1)+FB1	AR(1)	DNS(1)
	1 Step	VAR(1)+FB1	AR(1)+FB3	DNS(1)+FB2	AR(1)+FB1	DNS(1)+FB1
		VAR(1)+FB3	AR(1)+FB1	DNS(1)	AR(1)+FB2	DNS(2)
		AR(1)+FB3	**AR(1)+FB1**	**NSS(3)+FB1**	**NSS(3)+FB1**	DNS(5)
2001:Jan - 2005:Dec	3 Step	AR(1)+FB2	VAR(SIC)	NSS(2)+FB1	NSS(3)	DNS(5)+FB1
1st Subsample		VAR(1)+FB1	VAR(1)+FB1	NSS(3)	NSS(2)+FB1	DNS(5)+FB2
		DNS(3)+FB1	DNS(3)+FB1	DNS(6)+FB1	DNS(6)+FB1	DNS(6)+FB2
	12 Step	DNS(2)+FB1	NSS(2)+FB1	DNS(3)+FB1	DNS(6)	DNS(6)+FB3
		VAR(1)	DNS(2)+FB1	DNS(6)	DNS(6)+FB2	DNS(6)
		VAR(1)+FB2	NSS(4)	**VAR(SIC)**	**VAR(SIC)**	DNS(1)
	1 Step	VAR(1)+FB3	DNS(4)+FB3	VAR(1)	VAR(1)	DNS(3)
		VAR(1)+FB1	VAR(1)+FB3	VAR(1)+FB2	VAR(1)+FB3	AR(SIC)
		AR(1)	VAR(1)+FB1	**VAR(1)+FB1**	**VAR(1)+FB1**	DNS(3)
2006:Jan - 2010:Dec	3 Step	VAR(1)+FB1	NSS(4)	VAR(SIC)	DNS(6)	DNS(6)
2nd Subsample		AR(SIC)	DNS(4)+FB1	VAR(1)	VAR(SIC)	DNS(6)+FB1
		DNS(1)+FB1	DNS(1)+FB1	**NSS(1)+FB1**	**NSS(1)+FB1**	**DNS(6)**
	12 Step	DNS(2)+FB1	DNS(2)+FB1	DNS(1)+FB1	DNS(6)	DNS(6)+FB1
		DNS(3)+FB1	DNS(3)+FB1	DNS(3)+FB1	AR(1)+FB1	DNS(6)+FB2
		AR(SIC)	**AR(1)**	**AR(1)**	**AR(1)**	AR(1)
	1 Step	AR(1)	AR(1)+FB1	AR(1)+FB1	AR(SIC)	AR(SIC)
		AR(1)+FB1	AR(SIC)	AR(SIC)	AR(1)+FB1	DNS(1)
		VAR(1)	**DNS(5)**	**AR(1)**	**AR(1)**	DNS(4)
2011:Jan - 2017:Oct	3 Step	VAR(SIC)	DNS(5)+FB2	DNS(5)	AR(SIC)	DNS(4)+FB3
3rd Subsample		VAR(1)+FB1	DNS(5)+FB1	DNS(5)+FB1	NSS(4)	DNS(4)+FB2
		VAR(1)+FB3	**DNS(5)**	**DNS(4)**	**NSS(6)**	**NSS(6)**
	12 Step	VAR(1)	DNS(5)+FB1	DNS(4)+FB1	NSS(4)+FB3	NSS(6)+FB1
		VAR(SIC)	DNS(5)+FB2	DNS(4)+FB2	NSS(4)	NSS(6)+FB3
		VAR(1)+FB2	**VAR(1)+FB2**	**AR(1)**	**AR(1)**	**DNS(1)**
	1 Step	VAR(1)+FB3	AR(SIC)	AR(1)+FB1	AR(1)+FB1	AR(SIC)
		VAR(1)+FB1	AR(1)+FB1	AR(1)+FB2	AR(SIC)	AR(1)
		VAR(1)+FB1	VAR(1)+FB1	**VAR(1)+FB1**	**VAR(SIC)**	DNS(4)
2001:Jan - 2017:Oct	3 Step	VAR(1)+FB2	NSS(4)	VAR(SIC)	VAR(1)	DNS(4)+FB1
Whole Sample		AR(SIC)	VAR(1)+FB2	VAR(1)	VAR(1)+FB1	DNS(4)+FB2
		VAR(1)+FB1	VAR(1)+FB1	**VAR(1)**	**DNS(6)**	DNS(6)
	12 Step	VAR(1)	VAR(1)	VAR(SIC)	VAR(1)	DNS(6)+FB2
		VAR(SIC)	VAR(SIC)	VAR(1)+FB1	VAR(SIC)	DNS(6)+FB1

* Notes: See notes to Table 3.2A.

Table 3.3B: Point MSFEs of Top 3 MSFE-Best Models with Betas Calculated Using a 10-Dimensional Historical Yield Dataset*

	Horizon					
Maturity		3 Months	1 Year	3 Years	5 Years	10 Years
		0.029	**0.057**	0.112	0.118	0.088
	1 Step	0.029	0.059	0.114	0.120	0.090
		0.030	0.059	0.115	0.122	0.092
		0.233	**0.313**	**0.258**	**0.256**	0.202
2001:Jan - 2005:Dec	3 Step	0.240	0.316	0.337	0.291	0.207
1st Subsample		0.252	0.317	0.340	0.302	0.207
		2.383	2.462	1.687	1.062	0.511
	12 Step	2.405	2.542	1.693	1.066	0.514
		3.052	2.587	1.698	1.066	0.515
		0.089	0.055	**0.079**	**0.077**	0.092
	1 Step	0.089	0.056	0.079	0.078	0.094
		0.090	0.057	0.079	0.081	0.094
		0.358	0.317	**0.342**	**0.277**	0.161
2006:Jan - 2010:Dec	3 Step	0.358	0.322	0.349	0.279	0.177
2nd Subsample		0.359	0.325	0.349	0.279	0.189
		1.994	1.613	**1.087**	**0.752**	**0.316**
	12 Step	2.344	1.767	1.130	0.830	0.332
		2.370	1.820	1.271	0.832	0.334
		0.005	**0.006**	**0.020**	**0.036**	0.054
	1 Step	0.005	0.007	0.022	0.037	0.054
		0.006	0.007	0.022	0.038	0.054
		0.016	**0.014**	**0.051**	**0.109**	0.180
2011:Jan - 2017:Oct	3 Step	0.016	0.014	0.053	0.111	0.180
3rd Subsample		0.016	0.014	0.054	0.114	0.181
		0.062	**0.056**	**0.140**	**0.317**	**0.412**
	12 Step	0.062	0.059	0.140	0.324	0.433
		0.062	0.060	0.142	0.326	0.442
		0.039	**0.041**	**0.069**	**0.076**	**0.075**
	1 Step	0.040	0.042	0.070	0.079	0.078
		0.040	0.042	0.072	0.080	0.079
		0.188	0.197	**0.234**	**0.226**	0.199
2001:Jan - 2017:Oct	3 Step	0.192	0.199	0.234	0.227	0.202
Whole Sample		0.193	0.201	0.234	0.228	0.203
		1.797	1.566	**0.989**	**0.713**	0.439
	12 Step	1.797	1.568	0.990	0.721	0.448
		1.798	1.569	0.993	0.721	0.448

* Notes: See notes to Table 3.2B.

Table 3.3C: Relative MSFEs of Top 3 MSFE-Best Models with Betas Calculated Using a 10-Dimensional Historical Yield Dataset*

	Horizon					
Maturity		3 Months	1 Year	3 Years	5 Years	10 Years
		0.488	**0.793**	0.957	1.000	0.960
	1 Step	0.503	0.810	0.977	1.019	0.977
		0.508	0.815	0.980	1.039	1.000
		0.626	**0.773**	**0.590**	**0.686**	0.890
2001:Jan - 2005:Dec	3 Step	0.646	0.780	0.770	0.782	0.913
1st Subsample		0.676	0.782	0.777	0.809	0.914
		0.488	0.423	0.350	0.334	0.411
	12 Step	0.493	0.436	0.351	0.335	0.413
		0.625	0.444	0.352	0.335	0.414
		0.964	0.838	**0.913**	**0.859**	0.904
	1 Step	0.970	0.850	0.915	0.864	0.920
		0.978	0.856	0.920	0.906	0.929
		1.000	0.931	**0.931**	**0.869**	0.705
2006:Jan - 2010:Dec	3 Step	1.000	0.946	0.949	0.875	0.778
2nd Subsample		1.003	0.956	0.949	0.876	0.827
		0.591	0.604	**0.636**	**0.683**	**0.669**
	12 Step	0.695	0.662	0.661	0.753	0.702
		0.702	0.682	0.743	0.755	0.706
		0.978	**1.000**	**1.000**	**1.000**	1.000
	1 Step	1.000	1.048	1.054	1.030	1.006
		1.119	1.091	1.074	1.055	1.006
		0.784	**0.642**	**1.000**	**1.000**	0.916
2011:Jan - 2017:Oct	3 Step	0.785	0.655	1.032	1.017	0.917
3rd Subsample		0.801	0.655	1.049	1.041	0.920
		0.221	**0.215**	**0.438**	**0.603**	**0.491**
	12 Step	0.221	0.225	0.439	0.617	0.516
		0.221	0.229	0.445	0.620	0.527
		0.837	**0.947**	**1.000**	**1.000**	**0.951**
	1 Step	0.853	0.950	1.019	1.036	0.981
		0.862	0.950	1.045	1.045	1.000
		0.834	0.857	**0.900**	**0.906**	0.926
2001:Jan - 2017:Oct	3 Step	0.855	0.863	0.901	0.907	0.941
Whole Sample		0.859	0.874	0.902	0.911	0.944
		0.700	0.595	**0.478**	**0.480**	0.516
	12 Step	0.701	0.596	0.478	0.485	0.526
		0.701	0.597	0.480	0.485	0.526

* Notes: See notes to Table 3.2C.

Table 3.4A: Top 3 MSFE-Best Models with Betas Calculated Using a 30-Dimensional Historical Yield Dataset*

Maturity	Horizon	3 Months	1 Year	3 Years	5 Years	10 Years
2001:Jan - 2005:Dec 1st Subsample	1 Step	**VAR(1)+FB2**	**AR(1)+FB2**	AR(1)+FB1	AR(1)	AR(1)
		VAR(1)+FB1	AR(1)+FB3	DNS(1)+FB1	AR(1)+FB1	AR(SIC)
		VAR(1)+FB3	AR(1)+FB1	AR(1)	AR(1)+FB2	NSS(4)
	3 Step	AR(1)+FB3	**AR(1)+FB1**	VAR(SIC)	DNS(3)	DNS(4)
		AR(1)+FB2	VAR(SIC)	VAR(1)	DNS(4)	NSS(4)
		VAR(1)+FB1	VAR(1)+FB1	VAR(1)+FB2	VAR(SIC)	DNS(4)+FB2
	12 Step	DNS(3)+FB1	DNS(3)+FB1	**DNS(3)+FB1**	**DNS(3)+FB1**	**DNS(3)**
		DNS(2)+FB1	DNS(1)+FB1	VAR(1)	DNS(3)	DNS(3)+FB1
		DNS(1)+FB1	DNS(2)+FB1	VAR(SIC)	VAR(1)	DNS(5)
2006:Jan - 2010:Dec 2nd Subsample	1 Step	VAR(1)+FB2	VAR(1)+FB3	**VAR(SIC)**	**VAR(SIC)**	**NSS(1)**
		VAR(1)+FB3	VAR(1)+FB2	VAR(1)	VAR(1)	DNS(1)
		VAR(1)+FB1	VAR(1)	VAR(1)+FB2	VAR(1)+FB3	DNS(2)
	3 Step	AR(1)	NSS(4)	**VAR(1)+FB1**	**VAR(1)+FB1**	**NSS(4)+FB1**
		VAR(1)+FB1	VAR(1)+FB1	VAR(SIC)	VAR(SIC)	NSS(4)+FB2
		AR(SIC)	VAR(1)+FB2	VAR(1)	VAR(1)	DNS(1)
	12 Step	DNS(3)+FB1	DNS(3)+FB1	DNS(1)+FB1	AR(1)+FB1	DNS(3)
		DNS(2)+FB1	DNS(2)+FB1	AR(1)+FB1	VAR(1)+FB1	DNS(4)
		AR(SIC)	DNS(1)+FB1	DNS(3)+FB1	VAR(SIC)	NSS(4)+FB1
2011:Jan - 2017:Oct 3rd Subsample	1 Step	**AR(SIC)**	**AR(1)**	**AR(1)**	**AR(1)**	AR(1)
		AR(1)	AR(1)+FB1	AR(1)+FB1	AR(SIC)	AR(SIC)
		AR(1)+FB1	AR(SIC)	AR(SIC)	AR(1)+FB1	NSS(1)+FB1
	3 Step	**VAR(1)**	NSS(4)	**AR(1)**	**AR(1)**	**DNS(5)**
		VAR(SIC)	AR(1)	AR(SIC)	AR(SIC)	DNS(5)+FB1
		VAR(1)+FB1	VAR(1)	NSS(4)	DNS(4)	DNS(5)+FB2
	12 Step	VAR(1)+FB3	DNS(6)+FB2	DNS(6)	VAR(SIC)	DNS(5)+FB3
		VAR(1)	DNS(6)+FB1	DNS(6)+FB3	VAR(1)	DNS(4)+FB3
		VAR(SIC)	DNS(6)	NSS(4)+FB1	DNS(6)	DNS(4)+FB2
2001:Jan - 2017:Oct Whole Sample	1 Step	**VAR(1)+FB2**	**VAR(1)+FB2**	**AR(1)**	**AR(1)**	NSS(1)
		VAR(1)+FB3	AR(SIC)	AR(1)+FB1	AR(1)+FB1	AR(SIC)
		VAR(1)+FB1	AR(1)+FB1	AR(1)+FB2	AR(SIC)	AR(1)
	3 Step	VAR(1)+FB1	VAR(1)+FB1	**VAR(1)+FB1**	**VAR(SIC)**	**DNS(5)**
		VAR(1)+FB2	VAR(1)+FB2	VAR(SIC)	VAR(1)	DNS(5)+FB1
		AR(SIC)	VAR(SIC)	VAR(1)	VAR(1)+FB1	DNS(5)+FB2
	12 Step	VAR(1)+FB1	VAR(1)+FB1	**VAR(1)**	VAR(1)	DNS(4)
		VAR(1)	VAR(1)	VAR(SIC)	VAR(SIC)	DNS(4)+FB1
		VAR(SIC)	VAR(SIC)	VAR(1)+FB1	VAR(1)+FB1	DNS(5)

* Notes: See notes to Table 3.2A.

Table 3.4B: Point MSFEs of Top 3 MSFE-Best Models with Betas Calculated Using a 30-Dimensional Historical Yield Dataset*

	Horizon					
Maturity		3 Months	1 Year	3 Years	5 Years	10 Years
		0.029	**0.057**	0.115	0.118	0.092
	1 Step	0.029	0.059	0.116	0.120	0.094
		0.030	0.059	0.117	0.122	0.094
		0.233	**0.313**	0.355	0.306	0.205
2001:Jan - 2005:Dec	3 Step	0.240	0.316	0.357	0.320	0.209
1st Subsample		0.252	0.317	0.360	0.323	0.210
		2.628	2.489	**1.375**	**0.869**	**0.374**
	12 Step	2.690	2.742	1.750	1.044	0.405
		2.949	2.832	1.755	1.114	0.477
		0.089	0.057	**0.079**	**0.077**	**0.075**
	1 Step	0.089	0.058	0.079	0.078	0.082
		0.090	0.059	0.079	0.081	0.087
		0.358	0.316	**0.342**	**0.277**	**0.179**
2006:Jan - 2010:Dec	3 Step	0.358	0.317	0.349	0.279	0.187
2nd Subsample		0.359	0.327	0.349	0.280	0.187
		2.149	1.791	1.188	0.832	0.365
	12 Step	2.169	1.807	1.278	0.844	0.365
		2.434	2.066	1.336	0.846	0.368
		0.005	**0.006**	**0.020**	**0.036**	0.054
	1 Step	0.005	0.007	0.022	0.037	0.054
		0.006	0.007	0.022	0.038	0.054
		0.016	0.018	**0.051**	**0.109**	**0.163**
2011:Jan - 2017:Oct	3 Step	0.016	0.022	0.054	0.111	0.165
3rd Subsample		0.016	0.022	0.055	0.114	0.168
		0.062	0.062	0.143	0.339	0.588
	12 Step	0.062	0.062	0.148	0.340	0.591
		0.062	0.063	0.154	0.342	0.592
		0.039	**0.041**	**0.069**	**0.076**	0.078
	1 Step	0.040	0.042	0.070	0.079	0.078
		0.040	0.042	0.072	0.080	0.079
		0.188	0.197	**0.234**	**0.226**	**0.187**
2001:Jan - 2017:Oct	3 Step	0.192	0.201	0.234	0.227	0.190
Whole Sample		0.193	0.202	0.234	0.228	0.193
		1.797	1.566	**0.989**	0.721	0.501
	12 Step	1.797	1.568	0.990	0.721	0.503
		1.798	1.569	0.993	0.727	0.505

* Notes: See notes to Table 3.2B.

Table 3.4C: Relative MSFEs of Top 3 MSFE-Best Models with Betas Calculated Using a 30-Dimensional Historical Yield Dataset*

Maturity	Horizon	3 Months	1 Year	3 Years	5 Years	10 Years
		0.488	**0.793**	0.983	1.000	1.000
	1 Step	0.503	0.810	0.992	1.019	1.020
		0.508	0.815	1.000	1.039	1.021
		0.626	**0.773**	0.812	0.821	0.906
2001:Jan - 2005:Dec	3 Step	0.646	0.780	0.816	0.858	0.920
1st Subsample		0.676	0.782	0.822	0.866	0.927
		0.538	0.427	**0.285**	**0.273**	**0.300**
	12 Step	0.551	0.471	0.363	0.328	0.325
		0.604	0.486	0.364	0.350	0.383
		0.964	0.856	**0.913**	**0.859**	**0.739**
	1 Step	0.970	0.877	0.915	0.864	0.807
		0.978	0.890	0.920	0.906	0.860
		1.000	0.929	**0.931**	**0.869**	**0.786**
2006:Jan - 2010:Dec	3 Step	1.000	0.931	0.949	0.876	0.822
2nd Subsample		1.003	0.961	0.949	0.877	0.822
		0.637	0.671	0.695	0.755	0.772
	12 Step	0.643	0.677	0.748	0.765	0.773
		0.721	0.774	0.781	0.767	0.779
		0.978	**1.000**	**1.000**	**1.000**	1.000
	1 Step	1.000	1.048	1.054	1.030	1.006
		1.119	1.091	1.074	1.055	1.007
		0.784	0.813	**1.000**	**1.000**	**0.832**
2011:Jan - 2017:Oct	3 Step	0.785	1.000	1.055	1.017	0.843
3rd Subsample		0.801	1.018	1.083	1.041	0.856
		0.221	0.236	0.450	0.644	0.700
	12 Step	0.221	0.236	0.463	0.647	0.703
		0.221	0.239	0.483	0.651	0.705
		0.837	**0.947**	**1.000**	**1.000**	0.980
	1 Step	0.853	0.950	1.019	1.036	0.981
		0.862	0.950	1.045	1.045	1.000
		0.834	0.857	**0.900**	**0.906**	**0.873**
2001:Jan - 2017:Oct	3 Step	0.855	0.874	0.901	0.907	0.886
Whole Sample		0.859	0.876	0.902	0.911	0.898
		0.700	0.595	**0.478**	0.485	0.589
	12 Step	0.701	0.596	0.478	0.485	0.591
		0.701	0.597	0.480	0.489	0.594

* Notes: See notes to Table 3.2C.

Table 3.5A: h=1-Step Ahead Relative MSFEs of All Forecasting Models (Subsample 1: 2001:1-2005:12) *

Model			rMSFE		
Maturity	3 month	1 year	3 years	5 years	10 years
AR(1)	1.000	1.000	1.000	1.000	1.000
AR(SIC)	0.758^{***}	0.941	1.089	1.086	1.020
AR(1)+FB1	0.721^{**}	0.815^{*}	0.983	1.019	1.027
AR(1)+FB2	0.632^{***}	$\mathbf{0.793}^{*}$	1.019	1.039	1.049
AR(1)+FB3	0.631^{***}	0.810^{*}	1.037	1.048	1.052
VAR(1)	0.536^{***}	0.952	1.052	1.088	1.173
VAR(SIC)	0.523^{***}	0.941	1.042	1.082	1.176
VAR(1)+FB1	0.503^{***}	0.860	1.039	1.086	1.169
VAR(1)+FB2	$\mathbf{0.488}^{***}$	0.855	1.059	1.115	1.217
VAR(1)+FB3	0.508^{***}	0.860	1.070	1.125	1.221
DNS(1)	1.156	1.151	0.959	1.091	0.975
DNS(2)	1.062	1.409	1.073	1.141	0.978
DNS(3)	1.113	1.859	2.357	2.774	2.225
DNS(4)	0.639^{***}	1.209	1.018	1.114	1.102
DNS(5)	0.653^{**}	1.085	1.022	1.084	1.015
DNS(6)	0.792	1.567	1.984	2.291	1.917
DNS(1)+FB1	0.885	0.867	**0.955**	1.083	0.988
DNS(2)+FB1	0.778^{*}	1.032	1.007	1.089	0.976
DNS(3)+FB1	0.871	1.628	2.444	2.844	2.294
DNS(4)+FB1	0.575^{***}	1.036	0.985	1.109	1.076
DNS(5)+FB1	0.554^{**}	0.941	0.980	1.068	1.005
DNS(6)+FB1	0.683^{*}	1.402	1.939	2.287	1.928
DNS(1)+FB2	0.870	0.846	0.976	1.107	1.026
DNS(2)+FB2	1.026	1.294	1.082	1.150	1.017
DNS(3)+FB2	1.044	1.593	2.100	2.509	2.061
DNS(4)+FB2	0.588^{***}	1.038	1.001	1.128	1.089
DNS(5)+FB2	0.585^{**}	0.975	0.988	1.069	1.008
DNS(6)+FB2	0.714^{*}	1.432	1.938	2.277	1.920
DNS(1)+FB3	0.886	0.904	1.030	1.147	1.048
DNS(2)+FB3	1.011	1.346	1.161	1.220	1.078
DNS(3)+FB3	1.023	1.647	2.169	2.542	2.058
DNS(4)+FB3	0.566^{***}	1.034	1.001	1.123	1.092
DNS(5)+FB3	0.617^{**}	0.986	0.994	1.066	1.009
DNS(6)+FB3	0.752	1.447	1.928	2.250	1.898
NSS(1)	2.301	1.954	1.118	1.173	0.986
NSS(2)	3.136	3.138	1.500	1.217	1.054
NSS(3)	3.573	3.348	1.627	1.331	1.180
NSS(4)	0.517^{***}	0.980	1.025	1.121	1.091
NSS(5)	1.590	1.546	0.998	**0.933**	0.948
NSS(6)	1.751	2.046	1.416	1.244	1.178
NSS(1)+FB1	1.577	1.517	1.089	1.154	1.007
NSS(2)+FB1	2.238	2.094	1.273	1.139	1.057
NSS(3)+FB1	2.554	2.254	1.442	1.301	1.211
NSS(4)+FB1	0.491^{***}	0.919	1.020	1.121	1.091
NSS(5)+FB1	1.605	1.576	1.026	0.957	0.966
NSS(6)+FB1	1.754	2.072	1.450	1.273	1.200
NSS(1)+FB2	2.530	2.134	1.244	1.213	1.055
NSS(2)+FB2	3.312	2.662	1.226	1.038	0.966
NSS(3)+FB2	3.516	2.862	1.371	1.189	1.150
NSS(4)+FB2	0.494^{***}	0.939	1.055	1.151	1.127
NSS(5)+FB2	1.665	1.688	1.098	0.995	0.969
NSS(6)+FB2	1.814	2.197	1.560	1.358	1.253
NSS(1)+FB3	2.397	2.154	1.320	1.272	1.088
NSS(2)+FB3	3.078	2.586	1.298	1.098	0.980
NSS(3)+FB3	3.257	2.799	1.462	1.273	1.187
NSS(4)+FB3	0.492^{***}	0.943	1.050	1.135	1.118
NSS(5)+FB3	1.680	1.699	1.087	0.975	**0.937**
NSS(6)+FB3	1.830	2.221	1.554	1.341	1.222

* Notes: See notes to Table 3.2A. Entries are MSFEs, relative to the benchmark AR(1) MSFE. A 10-dimensional historical yield dataset is used in the construction of betas in all DNS and NSS type models reported on in this table. Entries in bold denote models with lowest MSFE, for a given maturity. Starred entries denote rejection of the null of equal predictive accuracy, based on application of the Diebold-Mariano test discussed in Section 4.1.1, where the benchmark model is an AR(1) process. Significance levels for the test are reported as $^{***}p < 0.01$, $^{**}p < 0.05$, and $^{*}p < 0.1$.

Table 3.5B: h=1-Step Ahead Relative MSFEs of All Forecasting Models (Subsample 2: 2006:1-2010:12) [*]

Model			rMSFE		
Maturity	3 month	1 year	3 years	5 years	10 years
AR(1)	1.000	1.000	1.000	1.000	1.000
AR(SIC)	1.025	0.941	1.003	0.999	0.929
AR(1)+FB1	1.101	1.085	1.057	1.047	1.069
AR(1)+FB2	1.197	1.115	1.031	1.012	1.030
AR(1)+FB3	1.154	1.058	0.993	0.975	1.011
VAR(1)	0.996	0.890	0.915	0.864*	0.963
VAR(SIC)	0.999	0.891	**0.913**	**0.859***	0.956
VAR(1)+FB1	0.978	0.917	0.968	0.955	1.097
VAR(1)+FB2	0.964	0.877	0.920	0.907	1.039
VAR(1)+FB3	0.970	0.856	0.920	0.906	1.056
DNS(1)	1.831	**0.783****	1.229	1.385	0.918
DNS(2)	1.331	0.896	1.209	1.108	1.022
DNS(3)	1.329	0.954	2.257	2.905	2.621
DNS(4)	1.177	0.899	1.137	1.309	0.958
DNS(5)	1.010	0.802*	1.128	1.115	1.074
DNS(6)	1.039	0.847	2.050	2.823	2.726
DNS(1)+FB1	1.226	1.461	1.166	1.305	1.052
DNS(2)+FB1	1.077	1.286	1.230	1.170	1.105
DNS(3)+FB1	1.049	1.056	1.526	2.230	2.588
DNS(4)+FB1	1.081	1.211	1.125	1.250	1.084
DNS(5)+FB1	1.001	0.902	1.156	1.164	1.143
DNS(6)+FB1	1.018	0.867	1.896	2.645	2.617
DNS(1)+FB2	1.222	1.522	1.135	1.245	1.027
DNS(2)+FB2	1.111	1.350	1.259	1.211	1.161
DNS(3)+FB2	1.073	1.075	1.607	2.410	2.807
DNS(4)+FB2	1.060	1.216	1.058	1.181	1.036
DNS(5)+FB2	1.002	0.882	1.119	1.118	1.096
DNS(6)+FB2	1.020	0.866	1.919	2.678	2.634
DNS(1)+FB3	1.273	1.391	1.098	1.206	1.043
DNS(2)+FB3	1.072	1.305	1.218	1.182	1.147
DNS(3)+FB3	1.025	1.014	1.547	2.348	2.758
DNS(4)+FB3	1.011	1.205	1.022	1.137	1.037
DNS(5)+FB3	0.991	0.862	1.095	1.097	1.088
DNS(6)+FB3	1.007	0.836	1.877	2.638	2.611
NSS(1)	1.003	0.824*	0.986	1.232	**0.906**
NSS(2)	0.935	1.311	1.975	2.017	1.521
NSS(3)	**0.896**	1.825	1.618	1.986	1.990
NSS(4)	1.012	0.923	1.054	1.243	0.988
NSS(5)	1.088	1.328	1.766	1.761	1.426
NSS(6)	1.011	2.511	2.508	3.050	2.750
NSS(1)+FB1	1.232	1.365	1.083	1.185	1.039
NSS(2)+FB1	1.091	1.520	1.941	2.055	1.709
NSS(3)+FB1	1.281	1.776	1.834	2.561	2.374
NSS(4)+FB1	1.074	1.027	1.126	1.293	1.055
NSS(5)+FB1	1.271	1.482	1.816	1.803	1.476
NSS(6)+FB1	1.133	2.766	2.864	3.433	3.045
NSS(1)+FB2	1.208	1.377	1.075	1.151	1.018
NSS(2)+FB2	1.082	1.638	1.994	2.043	1.667
NSS(3)+FB2	1.152	1.743	1.993	2.756	2.521
NSS(4)+FB2	1.083	1.030	1.073	1.218	1.001
NSS(5)+FB2	1.310	1.529	1.835	1.803	1.446
NSS(6)+FB2	1.168	2.815	2.912	3.474	3.037
NSS(1)+FB3	1.335	1.602	1.040	1.071	1.040
NSS(2)+FB3	1.260	1.919	1.737	1.777	1.513
NSS(3)+FB3	1.253	2.149	2.170	2.892	2.635
NSS(4)+FB3	1.089	1.029	1.043	1.140	1.013
NSS(5)+FB3	1.332	1.490	1.799	1.773	1.444
NSS(6)+FB3	1.216	2.822	3.040	3.644	3.188

[*] Notes: See notes to Table 3.5A.

Table 3.5C: h=1-Step Ahead Relative MSFEs of All Forecasting Models (Subsample 3: 2011:1-2017:10) *

Model			rMSFE		
Maturity	3 month	1 year	3 years	5 years	10 years
AR(1)	1.000	**1.000**	**1.000**	**1.000**	1.000
AR(SIC)	**0.978**	1.091	1.074	1.030	1.006
AR(1)+FB1	1.119	1.048	1.054	1.055	1.024
AR(1)+FB2	1.534	1.459	1.201	1.149	1.077
AR(1)+FB3	1.613	1.515	1.333	1.291	1.171
VAR(1)	2.277	2.490	1.827	1.538	1.281
VAR(SIC)	2.314	2.533	1.833	1.538	1.280
VAR(1)+FB1	2.350	2.566	1.876	1.573	1.304
VAR(1)+FB2	2.087	2.236	1.738	1.469	1.216
VAR(1)+FB3	2.210	2.472	1.866	1.555	1.276
DNS(1)	7.708	2.545	1.660	1.998	**0.972**
DNS(2)	8.505	4.500	2.584	2.015	1.396
DNS(3)	8.694	4.316	5.046	5.754	3.669
DNS(4)	5.397	3.026	1.391	1.797	1.019
DNS(5)	3.364	1.532	1.948	1.862	1.547
DNS(6)	3.307	1.691	5.550	6.508	4.326
DNS(1)+FB1	10.007	2.557	2.005	2.215	0.978
DNS(2)+FB1	17.098	9.410	4.575	3.115	1.949
DNS(3)+FB1	17.318	8.148	5.395	5.656	3.622
DNS(4)+FB1	6.258	2.554	1.517	1.936	1.025
DNS(5)+FB1	3.169	1.388	1.935	1.851	1.550
DNS(6)+FB1	3.108	1.570	5.561	6.505	4.326
DNS(1)+FB2	9.315	3.068	2.042	2.168	1.035
DNS(2)+FB2	14.695	8.377	4.211	2.880	1.789
DNS(3)+FB2	14.876	7.251	5.350	5.681	3.621
DNS(4)+FB2	6.136	2.801	1.497	1.883	1.008
DNS(5)+FB2	3.179	1.541	1.876	1.775	1.491
DNS(6)+FB2	3.101	1.730	5.552	6.466	4.282
DNS(1)+FB3	9.208	2.614	1.825	2.008	1.021
DNS(2)+FB3	12.533	6.652	3.460	2.449	1.575
DNS(3)+FB3	12.713	5.893	5.127	5.641	3.643
DNS(4)+FB3	6.073	2.707	1.561	1.936	1.022
DNS(5)+FB3	3.454	1.598	2.042	1.895	1.546
DNS(6)+FB3	3.403	1.940	5.879	6.700	4.401
NSS(1)	4.293	2.824	1.493	1.909	0.986
NSS(2)	7.286	8.454	3.760	2.406	1.572
NSS(3)	7.686	8.427	7.105	7.646	4.721
NSS(4)	2.971	2.379	1.501	1.962	1.077
NSS(5)	2.771	3.675	2.051	1.496	1.140
NSS(6)	4.015	4.253	8.703	9.549	5.592
NSS(1)+FB1	6.521	4.326	1.856	2.149	1.004
NSS(2)+FB1	10.959	12.740	5.440	3.294	1.976
NSS(3)+FB1	11.011	12.108	6.637	6.660	4.113
NSS(4)+FB1	3.622	2.937	1.577	1.959	1.101
NSS(5)+FB1	2.961	3.260	1.928	1.447	1.117
NSS(6)+FB1	4.221	3.793	8.627	9.578	5.643
NSS(1)+FB2	6.312	4.577	1.939	2.099	1.054
NSS(2)+FB2	10.456	12.243	5.361	3.257	1.903
NSS(3)+FB2	10.634	11.740	6.949	6.971	4.267
NSS(4)+FB2	3.410	2.766	1.487	1.841	1.050
NSS(5)+FB2	3.173	3.730	2.132	1.524	1.117
NSS(6)+FB2	4.434	4.085	9.040	9.820	5.732
NSS(1)+FB3	5.209	3.915	1.683	1.899	1.033
NSS(2)+FB3	7.969	9.702	4.775	3.016	1.797
NSS(3)+FB3	8.205	9.518	7.595	7.718	4.702
NSS(4)+FB3	3.993	3.199	1.749	2.068	1.099
NSS(5)+FB3	3.389	4.266	2.409	1.674	1.191
NSS(6)+FB3	4.909	4.726	9.308	9.966	5.794

* Notes: See notes to Table 3.5A.

Table 3.5D: h=1-Step Ahead Relative MSFEs of All Forecasting Models (Whole Sample: 2001:1-2017:10) *

Model	rMSFE				
Maturity	3 month	1 year	3 years	5 years	10 years
AR(1)	1.000	1.000	**1.000**	**1.000**	1.000
AR(SIC)	0.923	0.950	1.055	1.045	0.981
AR(1)+FB1	0.961	0.950	1.019	1.036	1.042
AR(1)+FB2	1.002	0.977	1.045	1.051	1.049
AR(1)+FB3	0.980	0.964	1.056	1.070	1.069
VAR(1)	0.882*	1.015	1.094	1.096	1.123
VAR(SIC)	0.881*	1.013	1.089	1.092	1.121
VAR(1)+FB1	0.862*	0.987	1.114	1.134	1.179
VAR(1)+FB2	**0.837****	**0.947**	1.089	1.111	1.149
VAR(1)+FB3	0.853	0.954	1.110	1.131	1.174
DNS(1)	1.841	1.068	1.144	1.368	**0.953**
DNS(2)	1.550	1.362	1.306	1.298	1.110
DNS(3)	1.576	1.597	2.644	3.394	2.773
DNS(4)	1.164	1.178	1.107	1.314	1.024
DNS(5)	0.982	0.984	1.173	1.245	1.184
DNS(6)	1.048	1.250	2.439	3.290	2.888
DNS(1)+FB1	1.489	1.234	1.161	1.379	1.009
DNS(2)+FB1	1.677	1.644	1.521	1.508	1.293
DNS(3)+FB1	1.704	1.758	2.458	3.172	2.771
DNS(4)+FB1	1.123	1.205	1.101	1.318	1.065
DNS(5)+FB1	0.931	0.950	1.161	1.253	1.207
DNS(6)+FB1	0.986	1.171	2.360	3.225	2.850
DNS(1)+FB2	1.450	1.282	1.164	1.360	1.029
DNS(2)+FB2	1.682	1.740	1.526	1.505	1.284
DNS(3)+FB2	1.675	1.696	2.308	3.086	2.774
DNS(4)+FB2	1.109	1.223	1.082	1.292	1.046
DNS(5)+FB2	0.943	0.967	1.144	1.223	1.174
DNS(6)+FB2	0.999	1.195	2.367	3.224	2.842
DNS(1)+FB3	1.481	1.225	1.152	1.334	1.039
DNS(2)+FB3	1.558	1.643	1.460	1.443	1.241
DNS(3)+FB3	1.543	1.614	2.294	3.072	2.761
DNS(4)+FB3	1.070	1.210	1.076	1.284	1.052
DNS(5)+FB3	0.961	0.967	1.158	1.237	1.187
DNS(6)+FB3	1.019	1.201	2.386	3.243	2.858
NSS(1)	1.632	1.497	1.114	1.335	0.955
NSS(2)	2.035	2.631	1.950	1.726	1.375
NSS(3)	2.192	2.964	2.285	2.777	2.463
NSS(4)	0.915	1.037	1.094	1.326	1.048
NSS(5)	1.350	1.574	1.412	1.330	1.183
NSS(6)	1.419	2.386	2.703	3.476	2.991
NSS(1)+FB1	1.595	1.615	1.179	1.357	1.018
NSS(2)+FB1	1.955	2.468	2.025	1.874	1.558
NSS(3)+FB1	2.186	2.624	2.215	2.774	2.453
NSS(4)+FB1	0.970	1.087	1.127	1.342	1.080
NSS(5)+FB1	1.470	1.634	1.429	1.347	1.202
NSS(6)+FB1	1.501	2.486	2.844	3.628	3.125
NSS(1)+FB2	1.926	1.939	1.265	1.362	1.040
NSS(2)+FB2	2.327	2.770	2.012	1.817	1.491
NSS(3)+FB2	2.452	2.886	2.276	2.851	2.530
NSS(4)+FB2	0.967	1.088	1.114	1.307	1.058
NSS(5)+FB2	1.525	1.738	1.498	1.379	1.192
NSS(6)+FB2	1.553	2.587	2.967	3.728	3.165
NSS(1)+FB3	1.901	2.010	1.259	1.323	1.055
NSS(2)+FB3	2.234	2.708	1.881	1.705	1.408
NSS(3)+FB3	2.307	2.905	2.466	3.080	2.706
NSS(4)+FB3	0.996	1.115	1.132	1.317	1.073
NSS(5)+FB3	1.553	1.757	1.512	1.388	1.200
NSS(6)+FB3	1.608	2.640	3.044	3.808	3.229

* Notes: See notes to Table 3.5A.

Table 3.6A: h=3-Step Ahead Relative MSFEs of All Forecasting Models (Subsample 1: 2001:1-2005:12) *

Model			rMSFE		
Maturity	3 month	1 year	3 years	5 years	10 years
AR(1)	1.000	1.000	1.000	1.000	1.000
AR(SIC)	0.703***	0.863*	1.068	1.111	1.155
AR(1)+FB1	0.688***	**0.773***	1.002	1.143	1.170
AR(1)+FB2	0.646***	0.897	1.112	1.150	1.200
AR(1)+FB3	0.626***	0.897	1.119	1.155	1.207
VAR(1)	0.682***	0.783***	0.816***	0.869**	1.035
VAR(SIC)	0.680***	0.780***	0.812***	0.866**	1.032
VAR(1)+FB1	0.676***	0.782***	0.825***	0.884**	1.058
VAR(1)+FB2	0.692***	0.788***	0.822***	0.876**	1.038
VAR(1)+FB3	0.717***	0.806***	0.832***	0.883**	1.043
DNS(1)	1.183	1.145	0.938	1.013	0.878***
DNS(2)	1.084	1.254	1.017	1.029	0.957
DNS(3)	1.060	1.258	1.350	1.628	1.687
DNS(4)	0.679***	0.890***	0.817***	0.872***	0.949
DNS(5)	0.624***	0.864***	0.826***	0.858***	0.885**
DNS(6)	0.617***	0.873*	1.036	1.253	1.403
DNS(1)+FB1	0.784*	0.811*	0.829*	0.942	0.880**
DNS(2)+FB1	0.717**	0.874	0.857	0.918	0.939
DNS(3)+FB1	0.720**	0.972	1.307	1.609	1.731
DNS(4)+FB1	0.654***	0.859***	0.814***	0.883**	0.956
DNS(5)+FB1	0.592***	0.829***	0.813***	0.854***	0.884**
DNS(6)+FB1	0.585***	0.840**	1.031	1.263	1.423
DNS(1)+FB2	0.723**	0.932	1.024	1.095	1.006
DNS(2)+FB2	0.755***	1.103	1.136	1.203	1.249
DNS(3)+FB2	0.734***	1.100	1.405	1.700	1.872
DNS(4)+FB2	0.668***	0.875***	0.821***	0.887**	0.959
DNS(5)+FB2	0.589***	0.838***	0.828***	0.869***	0.903*
DNS(6)+FB2	**0.583***	0.851*	1.046	1.277	1.439
DNS(1)+FB3	0.687**	0.920	1.042	1.116	1.040
DNS(2)+FB3	0.721***	1.099	1.156	1.228	1.304
DNS(3)+FB3	0.697***	1.090	1.428	1.723	1.912
DNS(4)+FB3	0.688***	0.893**	0.830***	0.893**	0.964
DNS(5)+FB3	0.607***	0.851***	0.832***	0.871***	0.904*
DNS(6)+FB3	0.603***	0.868*	1.053	1.282	1.444
NSS(1)	1.666	1.573	1.062	1.087	0.913***
NSS(2)	1.899	1.855	1.053	0.900	0.913
NSS(3)	1.857	1.930	1.112	0.906	0.911
NSS(4)	0.688***	0.818***	0.823***	0.875**	0.945
NSS(5)	1.099	1.121	**0.789****	**0.759****	**0.847****
NSS(6)	1.056	1.251	0.896	0.808	0.895
NSS(1)+FB1	0.981	1.049	0.897	0.979	0.903**
NSS(2)+FB1	1.168	1.169	0.909	0.909	1.030
NSS(3)+FB1	1.131	1.231	0.945	0.899	1.006
NSS(4)+FB1	0.694***	0.832***	0.836***	0.891**	0.961
NSS(5)+FB1	1.080	1.108	0.797**	0.775**	0.869**
NSS(6)+FB1	1.038	1.233	0.898	0.818	0.912
NSS(1)+FB2	1.311	1.502	1.221	1.189	1.043
NSS(2)+FB2	1.731	1.640	1.027	0.991	1.195
NSS(3)+FB2	1.558	1.751	1.087	0.973	1.162
NSS(4)+FB2	0.713***	0.841***	0.837***	0.882**	0.952
NSS(5)+FB2	1.093	1.144	0.828**	0.797**	0.886*
NSS(6)+FB2	1.054	1.268	0.931	0.845	0.936
NSS(1)+FB3	1.235	1.463	1.233	1.208	1.074
NSS(2)+FB3	1.585	1.524	1.021	1.001	1.219
NSS(3)+FB3	1.450	1.617	1.056	0.970	1.184
NSS(4)+FB3	0.747***	0.863***	0.844***	0.886**	0.953
NSS(5)+FB3	1.111	1.162	0.836*	0.802**	0.889*
NSS(6)+FB3	1.074	1.290	0.941	0.851	0.939

* Notes: See notes to Table 3.5A.

Table 3.6B: h=3-Step Ahead Relative MSFEs of All Forecasting Models (Subsample 2: 2006:1-2010:12) *

Model	rMSFE				
Maturity	3 month	1 year	3 years	5 years	10 years
AR(1)	1.000	1.000	1.000	1.000	1.000
AR(SIC)	1.003	0.995	1.013	1.012	1.007
AR(1)+FB1	1.358	1.220	1.150	1.106	1.089
AR(1)+FB2	1.456	1.331	1.187	1.133	1.119
AR(1)+FB3	1.476	1.355	1.179	1.113	1.099
VAR(1)	1.060	0.977	0.949	0.877^{**}	0.912^{*}
VAR(SIC)	1.061	0.978	0.949	0.876^{**}	0.910^{*}
VAR(1)+FB1	1.000	0.931	**0.931**	**0.869***	0.913
VAR(1)+FB2	1.027	0.961	0.957	0.899^{*}	0.957
VAR(1)+FB3	1.029	0.964	0.958	0.898^{*}	0.950
DNS(1)	1.477	0.914^{**}	1.162	1.216	0.830^{**}
DNS(2)	1.233	0.966	1.132	1.088	0.991
DNS(3)	1.249	1.045	1.494	1.691	1.660
DNS(4)	1.152	0.847^{**}	1.047	1.088	0.886
DNS(5)	1.077	0.851^{**}	1.026	1.015	0.993
DNS(6)	1.098	0.912^{*}	1.292	1.490	1.596
DNS(1)+FB1	1.186	1.280	1.183	1.209	1.004
DNS(2)+FB1	1.264	1.343	1.328	1.262	1.079
DNS(3)+FB1	1.241	1.196	1.127	1.196	1.403
DNS(4)+FB1	1.058	0.841	1.021	1.057	0.924
DNS(5)+FB1	1.027	**0.834****	1.011	1.009	1.001
DNS(6)+FB1	1.045	0.878^{*}	1.230	1.413	1.516
DNS(1)+FB2	1.256	1.315	1.177	1.201	1.025
DNS(2)+FB2	1.339	1.409	1.347	1.275	1.093
DNS(3)+FB2	1.306	1.228	1.104	1.170	1.410
DNS(4)+FB2	1.067	0.851	1.024	1.057	0.929
DNS(5)+FB2	1.043	0.846^{**}	1.018	1.015	1.009
DNS(6)+FB2	1.061	0.890^{*}	1.237	1.419	1.521
DNS(1)+FB3	1.287	1.320	1.177	1.188	1.035
DNS(2)+FB3	1.376	1.466	1.373	1.283	1.076
DNS(3)+FB3	1.342	1.283	1.112	1.147	1.350
DNS(4)+FB3	1.074	0.852	1.027	1.062	0.926
DNS(5)+FB3	1.047	0.848^{**}	1.020	1.016	1.004
DNS(6)+FB3	1.064	0.891^{*}	1.241	1.423	1.522
NSS(1)	1.036	0.903^{*}	1.003	1.116	**0.828*****
NSS(2)	1.005	1.104	1.384	1.438	1.273
NSS(3)	0.883	1.108	1.017	0.904	0.913
NSS(4)	1.077	1.007	1.032	1.078	0.917
NSS(5)	1.016	1.061	1.140	1.151	1.102
NSS(6)	**0.883***	1.301	1.352	1.494	1.658
NSS(1)+FB1	1.440	1.301	1.115	1.117	0.978
NSS(2)+FB1	1.341	1.256	1.260	1.359	1.387
NSS(3)+FB1	1.373	1.363	1.242	1.246	1.230
NSS(4)+FB1	1.112	1.056	1.073	1.117	0.952
NSS(5)+FB1	1.076	1.120	1.172	1.175	1.123
NSS(6)+FB1	0.928	1.367	1.428	1.579	1.749
NSS(1)+FB2	1.661	1.423	1.110	1.100	0.997
NSS(2)+FB2	1.606	1.469	1.230	1.329	1.378
NSS(3)+FB2	1.637	1.628	1.260	1.243	1.231
NSS(4)+FB2	1.178	1.117	1.117	1.166	1.013
NSS(5)+FB2	1.154	1.223	1.259	1.258	1.205
NSS(6)+FB2	1.019	1.506	1.512	1.641	1.803
NSS(1)+FB3	1.877	1.611	1.132	1.082	1.009
NSS(2)+FB3	1.850	1.892	1.303	1.313	1.324
NSS(3)+FB3	1.887	2.138	1.407	1.290	1.248
NSS(4)+FB3	1.207	1.143	1.131	1.177	1.019
NSS(5)+FB3	1.154	1.221	1.255	1.255	1.204
NSS(6)+FB3	1.018	1.492	1.511	1.647	1.815

* Notes: See notes to Table 3.5A.

Table 3.6C: h=3-Step Ahead Relative MSFEs of All Forecasting Models (Subsample 3: 2011:1-2017:10) *

Model			rMSFE		
Maturity	3 month	1 year	3 years	5 years	10 years
AR(1)	1.000	1.000	**1.000**	**1.000**	1.000
AR(SIC)	1.088	1.085	1.055	1.017	1.024
AR(1)+FB1	1.576	1.528	1.236	1.135	1.052
AR(1)+FB2	1.738	1.671	1.284	1.162	1.058
AR(1)+FB3	2.174	2.031	1.554	1.338	1.123
VAR(1)	**0.784****	1.018	1.193	1.063	0.938*
VAR(SIC)	0.785*	1.029	1.202	1.067	0.939*
VAR(1)+FB1	0.801	1.030	1.203	1.069	0.942*
VAR(1)+FB2	0.819	1.045	1.213	1.074	0.943*
VAR(1)+FB3	0.935	1.148	1.255	1.097	0.960
DNS(1)	5.613	2.500	2.072	1.775	0.967
DNS(2)	5.712	4.778	2.978	1.922	1.279
DNS(3)	5.857	4.762	3.530	2.695	1.533
DNS(4)	1.420	1.175	1.016	1.148	**0.877**
DNS(5)	0.885	**0.691******	1.292	1.227	1.053
DNS(6)	0.906	0.919	2.797	2.662	1.667
DNS(1)+FB1	8.158	3.500	2.698	2.049	1.008
DNS(2)+FB1	10.102	8.219	4.713	2.680	1.581
DNS(3)+FB1	10.290	7.865	4.507	3.011	1.669
DNS(4)+FB1	1.598	1.070	1.056	1.189	0.885
DNS(5)+FB1	0.905	0.697***	1.300	1.231	1.058
DNS(6)+FB1	0.928	0.930	2.806	2.665	1.670
DNS(1)+FB2	8.037	3.475	2.637	2.005	1.003
DNS(2)+FB2	9.954	8.148	4.691	2.671	1.578
DNS(3)+FB2	10.137	7.812	4.551	3.041	1.677
DNS(4)+FB2	1.577	1.078	1.053	1.186	0.884
DNS(5)+FB2	0.921	0.706**	1.309	1.236	1.059
DNS(6)+FB2	0.946	0.945	2.820	2.672	1.672
DNS(1)+FB3	8.636	3.932	2.460	1.828	0.943
DNS(2)+FB3	9.185	8.066	4.401	2.476	1.462
DNS(3)+FB3	9.341	7.854	4.559	3.009	1.631
DNS(4)+FB3	1.649	1.090	1.072	1.197	0.882
DNS(5)+FB3	1.145	0.817*	1.387	1.275	1.071
DNS(6)+FB3	1.172	1.053	2.882	2.702	1.681
NSS(1)	5.816	4.389	2.066	1.791	1.017
NSS(2)	7.439	8.268	3.923	2.153	1.371
NSS(3)	7.200	8.484	4.277	2.903	1.439
NSS(4)	0.897	0.869	1.075	1.214	0.903
NSS(5)	0.983	1.641	1.322	1.082	0.924*
NSS(6)	1.321	1.876	4.034	3.509	1.744
NSS(1)+FB1	8.693	6.816	2.775	2.100	1.069
NSS(2)+FB1	11.022	12.049	5.413	2.752	1.594
NSS(3)+FB1	10.611	12.160	4.657	2.739	1.343
NSS(4)+FB1	0.986	0.928	1.072	1.200	0.898
NSS(5)+FB1	0.975	1.472	1.267	1.065	0.917*
NSS(6)+FB1	1.310	1.691	3.974	3.498	1.746
NSS(1)+FB2	8.538	6.707	2.714	2.053	1.061
NSS(2)+FB2	10.630	11.796	5.500	2.825	1.610
NSS(3)+FB2	10.167	11.801	4.796	2.854	1.377
NSS(4)+FB2	1.014	0.967	1.086	1.201	0.901
NSS(5)+FB2	0.978	1.466	1.264	1.062	0.914*
NSS(6)+FB2	1.306	1.692	4.002	3.515	1.752
NSS(1)+FB3	8.409	7.008	2.506	1.838	0.991
NSS(2)+FB3	10.381	12.676	5.981	2.932	1.583
NSS(3)+FB3	9.887	12.696	6.214	3.583	1.599
NSS(4)+FB3	1.308	1.175	1.159	1.256	0.902
NSS(5)+FB3	1.103	1.603	1.297	1.070	0.918*
NSS(6)+FB3	1.490	1.855	4.062	3.549	1.771

* Notes: See notes to Table 3.5A.

Table 3.6D: h=3-Step Ahead Relative MSFEs of All Forecasting Models (Whole Sample: 2001:1-2017:10) *

Model			rMSFE		
Maturity	3 month	1 year	3 years	5 years	10 years
AR(1)	1.000	1.000	1.000	1.000	1.000
AR(SIC)	0.859^{*}	0.930	1.044	1.057	1.060
AR(1)+FB1	1.037	0.997	1.083	1.128	1.101
AR(1)+FB2	1.068	1.117	1.157	1.146	1.122
AR(1)+FB3	1.084	1.141	1.179	1.172	1.142
VAR(1)	0.864^{***}	0.877^{***}	0.902^{***}	0.907^{**}	0.960^{*}
VAR(SIC)	0.864^{***}	0.876^{***}	0.901^{***}	$\mathbf{0.906}^{**}$	0.959^{*}
VAR(1)+FB1	0.834^{***}	0.857^{***}	$\mathbf{0.900}^{**}$	0.911^{**}	0.969
VAR(1)+FB2	0.855^{***}	0.874^{***}	0.910^{**}	0.920^{**}	0.977
VAR(1)+FB3	0.872^{***}	0.888^{***}	0.919^{**}	0.927^{**}	0.983
DNS(1)	1.481	1.095	1.123	1.226	$\mathbf{0.896}^{***}$
DNS(2)	1.321	1.262	1.222	1.210	1.087
DNS(3)	1.322	1.298	1.585	1.842	1.622
DNS(4)	0.930^{*}	0.882^{***}	0.929^{**}	1.003	0.903^{*}
DNS(5)	0.847^{***}	0.852^{***}	0.947^{*}	0.983	0.981
DNS(6)	0.854^{***}	0.892^{**}	1.284	1.593	1.562
DNS(1)+FB1	1.240	1.119	1.127	1.240	0.967
DNS(2)+FB1	1.314	1.359	1.363	1.362	1.222
DNS(3)+FB1	1.311	1.332	1.487	1.702	1.604
DNS(4)+FB1	0.879^{***}	0.859^{**}	0.920^{**}	1.003	0.920^{*}
DNS(5)+FB1	$\mathbf{0.809}^{***}$	$\mathbf{0.826}^{***}$	0.935^{*}	0.980	0.985
DNS(6)+FB1	0.814^{***}	0.860^{***}	1.256	1.570	1.544
DNS(1)+FB2	1.239	1.196	1.217	1.297	1.011
DNS(2)+FB2	1.363	1.505	1.509	1.491	1.322
DNS(3)+FB2	1.344	1.411	1.530	1.738	1.654
DNS(4)+FB2	0.889^{***}	0.872^{**}	0.925^{**}	1.005	0.922
DNS(5)+FB2	0.816^{***}	0.836^{***}	0.946^{*}	0.990	0.994
DNS(6)+FB2	0.822^{***}	0.872^{**}	1.268	1.579	1.551
DNS(1)+FB3	1.257	1.210	1.212	1.270	1.002
DNS(2)+FB3	1.336	1.525	1.506	1.471	1.291
DNS(3)+FB3	1.313	1.431	1.545	1.733	1.630
DNS(4)+FB3	0.905^{**}	0.882^{**}	0.932^{**}	1.011	0.922
DNS(5)+FB3	0.834^{***}	0.848^{***}	0.955	0.998	0.998
DNS(6)+FB3	0.841^{***}	0.885^{**}	1.278	1.588	1.556
NSS(1)	1.518	1.386	1.117	1.223	0.925^{**}
NSS(2)	1.676	1.770	1.421	1.327	1.197
NSS(3)	1.589	1.818	1.325	1.260	1.107
NSS(4)	0.879^{***}	0.903^{***}	0.931^{**}	1.012	0.920^{*}
NSS(5)	1.056	1.115	0.979	0.965	0.956
NSS(6)	0.984	1.297	1.338	1.548	1.451
NSS(1)+FB1	1.476	1.379	1.139	1.231	0.988
NSS(2)+FB1	1.605	1.621	1.416	1.407	1.352
NSS(3)+FB1	1.587	1.704	1.367	1.358	1.202
NSS(4)+FB1	0.902^{**}	0.933^{**}	0.954	1.032	0.934
NSS(5)+FB1	1.074	1.127	0.992	0.978	0.967
NSS(6)+FB1	0.996	1.309	1.367	1.583	1.485
NSS(1)+FB2	1.737	1.665	1.294	1.309	1.035
NSS(2)+FB2	1.993	1.951	1.470	1.445	1.407
NSS(3)+FB2	1.906	2.079	1.456	1.410	1.264
NSS(4)+FB2	0.944	0.967	0.974	1.046	0.952
NSS(5)+FB2	1.118	1.191	1.044	1.019	0.997
NSS(6)+FB2	1.047	1.388	1.420	1.622	1.512
NSS(1)+FB3	1.797	1.739	1.292	1.272	1.023
NSS(2)+FB3	2.027	2.109	1.536	1.462	1.387
NSS(3)+FB3	1.961	2.266	1.616	1.556	1.358
NSS(4)+FB3	0.985	0.998	0.990	1.062	0.955
NSS(5)+FB3	1.131	1.205	1.049	1.021	0.999
NSS(6)+FB3	1.062	1.400	1.430	1.632	1.524

* Notes: See notes to Table 3.5A.

Table 3.7A: h=12-Step Ahead Relative MSFEs of All Forecasting Models (Subsample 1: 2001:1-2005:12) *

Model			rMSFE		
Maturity	3 month	1 year	3 years	5 years	10 years
AR(1)	1.000	1.000	1.000	1.000	1.000
AR(SIC)	0.945	0.995	1.077	1.065	1.015
AR(1)+FB1	0.658***	0.805***	1.070	1.410	1.814
AR(1)+FB2	0.871**	1.046	1.171	1.557	2.745
AR(1)+FB3	0.888*	1.083	1.223	1.621	2.838
VAR(1)	0.625***	0.513***	0.363***	0.350***	0.451***
VAR(SIC)	0.626***	0.515***	0.364***	0.351***	0.454***
VAR(1)+FB1	0.631***	0.518***	0.367***	0.355***	0.458***
VAR(1)+FB2	0.637***	0.522***	0.370***	0.358***	0.460***
VAR(1)+FB3	0.634***	0.520***	0.369***	0.357***	0.464***
DNS(1)	0.767***	0.626***	0.533***	0.596***	0.679***
DNS(2)	0.638***	0.565***	0.479***	0.505***	0.554***
DNS(3)	0.624***	0.523***	0.436***	0.504***	0.670***
DNS(4)	0.645***	0.542***	0.397***	0.388***	0.448***
DNS(5)	0.609***	0.524***	0.387***	0.382***	0.416***
DNS(6)	0.577***	0.457***	0.315***	0.330***	0.455***
DNS(1)+FB1	0.535***	0.493***	0.515***	0.672***	0.995
DNS(2)+FB1	0.436***	0.439***	0.443***	0.528***	0.718***
DNS(3)+FB1	**0.424*****	**0.401*****	0.396***	0.514***	0.810***
DNS(4)+FB1	0.638***	0.536***	0.394***	0.387***	0.451***
DNS(5)+FB1	0.605***	0.522***	0.386***	0.382***	0.417***
DNS(6)+FB1	0.573***	0.454***	**0.314*****	0.330***	0.456***
DNS(1)+FB2	0.978	0.940	1.000	1.260	1.940
DNS(2)+FB2	0.739***	0.733***	0.765***	0.935	1.402
DNS(3)+FB2	0.727***	0.693***	0.703***	0.892**	1.443
DNS(4)+FB2	0.643***	0.540***	0.396***	0.387***	0.449***
DNS(5)+FB2	0.610***	0.525***	0.389***	0.384***	0.419***
DNS(6)+FB2	0.578***	0.457***	0.315***	0.330***	0.454***
DNS(1)+FB3	1.012	0.974	1.036	1.300	1.993
DNS(2)+FB3	0.765***	0.759***	0.784***	0.949	1.411
DNS(3)+FB3	0.753***	0.718***	0.720***	0.906*	1.451
DNS(4)+FB3	0.643***	0.540***	0.396***	0.387***	0.450***
DNS(5)+FB3	0.610***	0.525***	0.388***	0.383***	0.418***
DNS(6)+FB3	0.578***	0.458***	0.316***	0.331***	0.455***
NSS(1)	0.795***	0.695***	0.554***	0.615***	0.704***
NSS(2)	0.832***	0.660***	0.400***	0.374***	0.497***
NSS(3)	0.783***	0.667***	0.417***	0.363***	0.449***
NSS(4)	0.632***	0.538***	0.392***	0.384***	0.447***
NSS(5)	0.721**	0.588***	0.349***	0.303***	0.344***
NSS(6)	0.686***	0.602***	0.360***	**0.288*****	**0.306*****
NSS(1)+FB1	0.526***	0.526***	0.513***	0.667***	1.001
NSS(2)+FB1	0.570***	0.486***	0.315***	0.328***	0.566***
NSS(3)+FB1	0.525***	0.489***	0.326***	0.312***	0.508***
NSS(4)+FB1	0.658***	0.558***	0.405***	0.398***	0.461***
NSS(5)+FB1	0.728**	0.593***	0.351***	0.307***	0.352***
NSS(6)+FB1	0.693***	0.607***	0.363***	0.293***	0.314***
NSS(1)+FB2	0.943	0.961	0.994	1.261	1.976
NSS(2)+FB2	1.016	0.820***	0.525***	0.547***	0.948
NSS(3)+FB2	0.916	0.848**	0.571***	0.550***	0.894
NSS(4)+FB2	0.665***	0.564***	0.409***	0.400***	0.463***
NSS(5)+FB2	0.736**	0.599***	0.356***	0.311***	0.354***
NSS(6)+FB2	0.700***	0.614***	0.368***	0.297***	0.316***
NSS(1)+FB3	0.964	0.990	1.023	1.293	2.024
NSS(2)+FB3	1.042	0.841**	0.541***	0.565***	0.969
NSS(3)+FB3	0.934	0.873*	0.593***	0.571***	0.918
NSS(4)+FB3	0.657***	0.558***	0.406***	0.399***	0.468***
NSS(5)+FB3	0.733**	0.599***	0.356***	0.311***	0.356***
NSS(6)+FB3	0.697***	0.613***	0.369***	0.298***	0.319***

* Notes: See notes to Table 3.5A.

Table 3.7B: h=12-Step Ahead Relative MSFEs of All Forecasting Models (Subsample 2: 2006:1-2010:12) *

Model			rMSFE		
Maturity	3 month	1 year	3 years	5 years	10 years
AR(1)	1.000	1.000	1.000	1.000	1.000
AR(SIC)	0.721***	0.816**	0.984	1.020	0.970**
AR(1)+FB1	0.914	0.777*	0.748**	**0.755****	0.909**
AR(1)+FB2	1.095	0.889	0.866	0.973	1.442
AR(1)+FB3	1.119	0.907	0.888	0.997	1.474
VAR(1)	0.864***	0.815***	0.794***	0.769***	0.794**
VAR(SIC)	0.863***	0.814***	0.793***	0.767***	0.789**
VAR(1)+FB1	0.854***	0.801***	0.786***	0.765***	0.799*
VAR(1)+FB2	0.860***	0.807***	0.791***	0.771***	0.807*
VAR(1)+FB3	0.860***	0.808***	0.793***	0.774***	0.815*
DNS(1)	1.127	1.021	1.235	1.366	1.015
DNS(2)	0.998	0.963	1.094	1.115	1.019
DNS(3)	1.004	1.000	1.261	1.431	1.472
DNS(4)	0.893**	0.768***	0.839***	0.881**	**0.768****
DNS(5)	0.867**	0.796***	0.839***	0.836***	0.887
DNS(6)	0.867***	0.806***	0.946	1.075	1.305
DNS(1)+FB1	**0.690****	0.674**	0.706***	0.888**	1.012
DNS(2)+FB1	0.793*	0.766*	0.809	0.884	0.913
DNS(3)+FB1	0.786*	0.734**	0.743***	0.871**	1.257
DNS(4)+FB1	0.876**	0.756***	0.825***	0.866**	0.785*
DNS(5)+FB1	0.880**	0.809***	0.854***	0.851**	0.900
DNS(6)+FB1	0.880**	0.820***	0.965	1.098	1.335
DNS(1)+FB2	0.876	0.923	1.039	1.243	1.434
DNS(2)+FB2	0.993	0.992	1.046	1.114	1.128
DNS(3)+FB2	0.973	0.931	0.949	1.063	1.437
DNS(4)+FB2	0.878**	0.759***	0.826***	0.867**	0.781*
DNS(5)+FB2	0.879**	0.808***	0.852***	0.848**	0.889
DNS(6)+FB2	0.879**	0.818***	0.961	1.091	1.318
DNS(1)+FB3	0.901	0.955	1.067	1.264	1.452
DNS(2)+FB3	1.016	1.017	1.070	1.134	1.135
DNS(3)+FB3	0.995	0.955	0.967	1.071	1.420
DNS(4)+FB3	0.880**	0.761***	0.830***	0.871**	0.788*
DNS(5)+FB3	0.880**	0.810***	0.854***	0.851**	0.895
DNS(6)+FB3	0.880**	0.820***	0.964	1.094	1.323
NSS(1)	1.020	1.053	1.174	1.329	1.043
NSS(2)	1.070	1.320	1.753	1.947	1.981
NSS(3)	1.002	1.266	1.510	1.492	1.312
NSS(4)	0.872**	0.826***	0.824***	0.872**	0.789*
NSS(5)	0.869**	0.853**	0.852***	0.864***	0.896
NSS(6)	0.781***	0.885*	0.941	0.907	0.846
NSS(1)+FB1	0.796	0.728**	**0.699*****	0.852**	0.987
NSS(2)+FB1	0.693**	**0.592*****	0.811**	1.140	1.913
NSS(3)+FB1	0.701**	0.637***	0.714***	0.778**	0.934
NSS(4)+FB1	0.897**	0.854***	0.847***	0.896**	0.792**
NSS(5)+FB1	0.932	0.929	0.937	0.957	0.999
NSS(6)+FB1	0.838***	0.964	1.019	0.972	0.879
NSS(1)+FB2	1.163	1.115	1.078	1.208	1.365
NSS(2)+FB2	0.907	0.813*	0.958	1.249	1.930
NSS(3)+FB2	0.918	0.900	0.938	0.980	1.070
NSS(4)+FB2	0.919*	0.872**	0.864**	0.909*	0.793*
NSS(5)+FB2	0.937	0.932	0.940	0.959	0.993
NSS(6)+FB2	0.842***	0.967	1.026	0.981	0.886
NSS(1)+FB3	1.212	1.162	1.109	1.224	1.372
NSS(2)+FB3	0.938	0.841	0.976	1.265	1.949
NSS(3)+FB3	0.948	0.936	0.973	1.015	1.105
NSS(4)+FB3	0.923*	0.877**	0.869**	0.915*	0.804*
NSS(5)+FB3	0.941	0.937	0.945	0.964	1.001
NSS(6)+FB3	0.846***	0.972	1.031	0.988	0.896

* Notes: See notes to Table 3.5A.

Table 3.7C: h=12-Step Ahead Relative MSFEs of All Forecasting Models (Subsample 3: 2011:1-2017:10) *

Model			rMSFE		
Maturity	3 month	1 year	3 years	5 years	10 years
AR(1)	1.000	1.000	1.000	1.000	1.000
AR(SIC)	2.008	1.704	1.201	1.052	1.107
AR(1)+FB1	3.138	3.030	2.200	1.474	1.100
AR(1)+FB2	3.210	3.090	2.281	1.524	1.134
AR(1)+FB3	4.989	4.567	3.213	1.832	1.141
VAR(1)	0.221***	0.315***	0.511***	0.647***	0.756***
VAR(SIC)	0.221***	0.316***	0.508***	**0.644*****	0.754***
VAR(1)+FB1	0.225***	0.320***	0.526***	0.662***	0.766***
VAR(1)+FB2	0.224***	0.319***	0.527***	0.662***	0.767***
VAR(1)+FB3	0.221***	0.315***	0.529***	0.666***	0.771***
DNS(1)	2.391	1.614	1.993	1.680	1.039
DNS(2)	3.369	3.185	3.151	2.207	1.580
DNS(3)	3.380	3.035	2.527	1.652	1.135
DNS(4)	0.202***	0.284***	**0.438*****	0.660***	0.702***
DNS(5)	**0.154*****	0.247***	0.520***	0.704***	0.823***
DNS(6)	0.157***	0.306***	0.823*	0.955	0.861***
DNS(1)+FB1	4.706	3.596	3.530	2.442	1.263
DNS(2)+FB1	7.672	7.268	6.286	3.861	2.313
DNS(3)+FB1	7.712	7.097	5.412	3.115	1.795
DNS(4)+FB1	0.206***	0.280***	0.444***	0.667***	0.707***
DNS(5)+FB1	0.158***	0.252***	0.543***	0.723***	0.836***
DNS(6)+FB1	0.161***	0.311***	0.847*	0.976	0.876***
DNS(1)+FB2	4.689	3.569	3.515	2.443	1.283
DNS(2)+FB2	7.490	7.106	6.164	3.795	2.278
DNS(3)+FB2	7.529	6.936	5.311	3.068	1.772
DNS(4)+FB2	0.207***	0.287***	0.445***	0.666***	0.706***
DNS(5)+FB2	0.158***	0.253***	0.545***	0.725***	0.838***
DNS(6)+FB2	0.161***	0.314***	0.853	0.981	0.879**
DNS(1)+FB3	5.602	4.207	3.691	2.407	1.227
DNS(2)+FB3	7.016	6.645	5.712	3.486	2.044
DNS(3)+FB3	7.071	6.549	4.987	2.846	1.583
DNS(4)+FB3	0.194***	0.319***	0.470***	0.679***	0.718***
DNS(5)+FB3	0.162***	0.261***	0.542***	0.720***	0.833***
DNS(6)+FB3	0.165***	0.327***	0.869	0.990	0.881**
NSS(1)	2.188	2.018	1.962	1.697	1.085
NSS(2)	2.880	3.157	2.934	2.041	1.437
NSS(3)	2.872	3.067	1.898	1.046	0.591***
NSS(4)	0.203***	**0.243*****	0.483***	0.704***	0.733***
NSS(5)	0.245***	0.483***	0.716***	0.793***	0.811***
NSS(6)	0.269***	0.493***	0.995	0.965	**0.576*****
NSS(1)+FB1	4.678	4.614	3.654	2.533	1.331
NSS(2)+FB1	5.235	5.914	5.220	3.230	1.904
NSS(3)+FB1	5.197	5.727	3.585	1.805	0.916
NSS(4)+FB1	0.200***	0.245***	0.496***	0.716***	0.740***
NSS(5)+FB1	0.235***	0.463***	0.715***	0.798***	0.815***
NSS(6)+FB1	0.255***	0.473***	1.008	0.981	0.586***
NSS(1)+FB2	4.582	4.526	3.617	2.524	1.348
NSS(2)+FB2	5.074	5.743	5.107	3.182	1.891
NSS(3)+FB2	5.038	5.552	3.495	1.774	0.911
NSS(4)+FB2	0.201***	0.244***	0.498***	0.721***	0.742***
NSS(5)+FB2	0.239***	0.469***	0.719***	0.801***	0.817***
NSS(6)+FB2	0.259***	0.480***	1.016	0.987	0.590***
NSS(1)+FB3	4.993	4.918	3.671	2.439	1.279
NSS(2)+FB3	5.321	5.968	5.119	3.079	1.755
NSS(3)+FB3	5.210	5.846	3.704	1.812	0.860
NSS(4)+FB3	0.229***	0.257***	0.488***	0.716***	0.731***
NSS(5)+FB3	0.273***	0.505***	0.692***	0.768***	0.799***
NSS(6)+FB3	0.298***	0.525***	1.007	0.970	0.580***

* Notes: See notes to Table 3.5A.

Table 3.7D: h=12-Step Ahead Relative MSFEs of All Forecasting Models (Whole Sample: 2001:1-2017:10) *

Model			rMSFE		
Maturity	3 month	1 year	3 years	5 years	10 years
AR(1)	1.000	1.000	1.000	1.000	1.000
AR(SIC)	0.904*	0.970	1.062	1.053	1.045
AR(1)+FB1	0.868*	0.887**	1.061	1.275	1.379
AR(1)+FB2	1.062	1.081	1.166	1.423	1.885
AR(1)+FB3	1.160	1.171	1.265	1.514	1.933
VAR(1)	0.701***	0.596***	**0.478*****	**0.485*****	0.630***
VAR(SIC)	0.701***	0.597***	0.478***	0.485***	0.629***
VAR(1)+FB1	0.700***	0.595***	0.480***	0.489***	0.638***
VAR(1)+FB2	0.706***	0.600***	0.483***	0.493***	0.640***
VAR(1)+FB3	0.704***	0.599***	0.483***	0.494***	0.645***
DNS(1)	0.979	0.785***	0.797***	0.921	0.879***
DNS(2)	0.900**	0.791***	0.797***	0.884*	1.042
DNS(3)	0.895**	0.769***	0.769***	0.873**	0.989
DNS(4)	0.722***	0.600***	0.508***	0.535***	0.603***
DNS(5)	0.689***	0.595***	0.506***	0.528***	0.657***
DNS(6)	**0.672*****	**0.556*****	0.502***	0.584***	0.758***
DNS(1)+FB1	0.781***	0.674***	0.750***	0.974	1.105
DNS(2)+FB1	0.896	0.815**	0.898	1.085	1.389
DNS(3)+FB1	0.888	0.773***	0.794**	0.966	1.278
DNS(4)+FB1	0.712***	0.592***	0.503***	0.533***	0.609***
DNS(5)+FB1	0.693***	0.597***	0.511***	0.534***	0.665***
DNS(6)+FB1	0.675***	0.559***	0.507***	0.592***	0.769***
DNS(1)+FB2	1.103	1.041	1.167	1.426	1.593
DNS(2)+FB2	1.137	1.069	1.171	1.385	1.707
DNS(3)+FB2	1.124	1.018	1.051	1.242	1.574
DNS(4)+FB2	0.716***	0.596***	0.504***	0.533***	0.607***
DNS(5)+FB2	0.695***	0.599***	0.512***	0.535***	0.664***
DNS(6)+FB2	0.677***	0.560***	0.507***	0.591***	0.767***
DNS(1)+FB3	1.172	1.100	1.210	1.451	1.597
DNS(2)+FB3	1.140	1.075	1.162	1.354	1.619
DNS(3)+FB3	1.127	1.026	1.048	1.221	1.499
DNS(4)+FB3	0.716***	0.598***	0.507***	0.536***	0.613***
DNS(5)+FB3	0.696***	0.600***	0.512***	0.535***	0.663***
DNS(6)+FB3	0.678***	0.562***	0.509***	0.594***	0.769***
NSS(1)	0.945*	0.857***	0.794***	0.927	0.913***
NSS(2)	1.016	0.960	0.891	0.960	1.118
NSS(3)	0.961	0.945	0.778***	0.710***	0.648***
NSS(4)	0.706***	0.613***	0.504***	0.537***	0.618***
NSS(5)	0.758***	0.664***	0.495***	0.497***	0.622***
NSS(6)	0.705***	0.683***	0.542***	0.522***	**0.503*****
NSS(1)+FB1	0.815**	0.752***	0.755***	0.975	1.131
NSS(2)+FB1	0.824**	0.738***	0.743***	0.924	1.325
NSS(3)+FB1	0.801***	0.746***	0.625***	0.629***	0.742***
NSS(4)+FB1	0.731***	0.634***	0.519***	0.553***	0.627***
NSS(5)+FB1	0.786***	0.689***	0.518***	0.521***	0.644***
NSS(6)+FB1	0.730***	0.709***	0.564***	0.541***	0.516***
NSS(1)+FB2	1.190	1.152	1.178	1.430	1.624
NSS(2)+FB2	1.153	1.018	0.918	1.080	1.488
NSS(3)+FB2	1.099	1.054	0.844**	0.820***	0.930
NSS(4)+FB2	0.744***	0.644***	0.526***	0.558***	0.629***
NSS(5)+FB2	0.792***	0.694***	0.522***	0.524***	0.645***
NSS(6)+FB2	0.736***	0.715***	0.570***	0.547***	0.520***
NSS(1)+FB3	1.240	1.201	1.209	1.443	1.618
NSS(2)+FB3	1.191	1.049	0.934	1.080	1.446
NSS(3)+FB3	1.129	1.093	0.881*	0.847**	0.925
NSS(4)+FB3	0.742***	0.642***	0.525***	0.559***	0.629***
NSS(5)+FB3	0.794***	0.697***	0.521***	0.521***	0.640***
NSS(6)+FB3	0.737***	0.718***	0.571***	0.546***	0.519***

* Notes: See notes to Table 3.5A.

Table 3.8: Forecast Combination Models*

Model	Description
All	Average of all fifty eight forecast models
FB	Average of forty two models that contain macro diffusion index(es), principle component analysis based on all macroeconomic variables
FB(1)	Average of fourteen models that contain one diffusion index, namely all "+FB1" models.
FB(2)	Average of fourteen models that contain two diffusion indexes, namely all "+FB2" models.
FB(3)	Average of fourteen models that contain three diffusion indexes, namely all "+FB3" models.
FS	Average of sixteen non-FB type models
Econometrics	Average of all ten AR and VAR type models
AR	Average of all five AR type models
VAR	Average of all five VAR type models
DNS	Average of all twenty four DNS type models
NSS	Average of all twenty four NSS type models
NS(AR)	Average of all twenty four DNS and NSS type models with underlying AR(1) factor specifications
NS(VAR)	Average of all twenty four DNS and NSS type models with underlying VAR(1) factor specifications
NS(OLS)	Average of all sixteen DNS and NSS type models with fixed decay parameter(s), estimated with OLS.
NS(NLS)	Average of all thirty two DNS and NSS type models with dynamic decay parameter(s), estimated with NLS.

* Notes: This table summarizes the combination models utilized in all forecast experiments.

Table 3.9A: h=1-Step Ahead Relative MSFEs of Forecast Combination Models*

	Model	rMSFE				
	Maturity	3 months	1 year	3 years	5 years	10 years
	All	0.457***	0.574***	0.442***	0.466***	0.412***
	FB	0.512***	0.651***	0.521***	0.545***	0.489***
	FB(1)	0.477***	0.601***	0.525***	0.556***	0.495***
	FB(2)	0.572**	0.697**	0.522***	0.542***	0.491***
	FB(3)	0.527***	0.687**	0.535***	0.550***	0.496***
	FS	**0.370*****	**0.441*****	**0.284*****	**0.301*****	**0.246*****
2001:Jan - 2005:Dec	Econometrics	0.559***	0.833**	1.007	1.037	1.078
1st Subsample	AR	0.675***	0.820**	0.999	1.016	1.013
	VAR	0.498***	0.883	1.046	1.093	1.183
	DNS	0.624***	0.993	1.059	1.215	1.012
	NSS	1.531	1.493	1.015	0.973	0.956
	NS(AR)	1.322	1.481	1.105	1.131	0.972
	NS(VAR)	0.717***	1.120	0.972	0.998	0.959
	NS(OLS)	0.660***	1.035	0.999	1.112	1.044
	NS(NLS)	1.057	1.324	1.057	1.065	1.018
	All	0.349***	0.401***	0.478***	0.470***	0.383***
	FB	0.416**	0.510**	0.560***	0.546***	0.462***
	FB(1)	0.442**	0.524**	0.575***	0.557***	0.463***
	FB(2)	0.434***	0.527***	0.582***	0.564***	0.467***
	FB(3)	0.408**	0.527**	0.548***	0.536***	0.470***
	FS	**0.253*****	**0.239*****	**0.345*****	**0.331*****	**0.219*****
2006:Jan - 2010:Dec	Econometrics	0.900	0.825	0.897	0.891	0.968
2nd Subsample	AR	0.963	0.897	0.961	0.968	0.975
	VAR	0.953	0.847	0.897	0.869	0.996
	DNS	0.998	0.878	1.174	1.358	1.138
	NSS	0.773*	1.022	1.051	0.919	0.889
	NS(AR)	0.847	0.997	1.020	0.996	0.886
	NS(VAR)	0.938	0.913	1.141	1.081	0.894
	NS(OLS)	0.908*	0.946	1.009	1.181	0.988
	NS(NLS)	0.787**	0.931	1.100	1.007	0.922
	All	2.816	1.507	0.904	0.893	0.509***
	FB	3.291	1.788	1.060	1.029	0.590***
	FB(1)	3.718	1.980	1.075	1.041	0.591***
	FB(2)	3.330	1.909	1.086	1.034	0.592***
	FB(3)	3.095	1.719	1.103	1.056	0.614***
	FS	1.799	**0.931**	**0.576*****	**0.599*****	**0.337*****
2011:Jan - 2017:Oct	Econometrics	1.248	1.342	1.284	1.208	1.107
3rd Subsample	AR	**1.085**	1.061	1.075	1.060	1.010
	VAR	2.128	2.364	1.798	1.518	1.260
	DNS	6.462	2.202	2.093	2.165	1.259
	NSS	4.303	3.891	2.046	1.727	1.089
	NS(AR)	8.315	5.336	2.424	2.043	1.114
	NS(VAR)	2.984	1.644	1.689	1.707	1.124
	NS(OLS)	4.479	2.259	1.541	1.936	0.998
	NS(NLS)	5.224	3.806	2.444	2.251	1.447
	All	0.499***	0.552***	0.512***	0.549***	0.427***
	FB	0.579***	0.655***	0.601***	0.639***	0.507***
	FB(1)	0.601**	0.648***	0.610***	0.650***	0.510***
	FB(2)	0.614***	0.692***	0.612***	0.645***	0.509***
	FB(3)	0.571***	0.676***	0.609***	0.643***	0.518***
	FS	**0.365*****	**0.379*****	**0.342*****	**0.369*****	**0.261*****
2001:Jan - 2017:Oct	Econometrics	0.789***	0.860**	1.000	1.019	1.044
Whole Sample	AR	0.861*	0.869*	0.994	1.008	0.998
	VAR	0.836**	0.955	1.081	1.097	1.133
	DNS	1.101	1.013	1.227	1.448	1.128
	NSS	1.212	1.423	1.153	1.099	0.967
	NS(AR)	1.355	1.492	1.232	1.259	0.979
	NS(VAR)	0.947	1.058	1.122	1.164	0.980
	NS(OLS)	0.975	1.068	1.068	1.295	1.010
	NS(NLS)	1.084	1.294	1.241	1.273	1.099

* Notes: See notes to Table 3.5A. Forecast combination models are listed in Table 3.8.

Table 3.9B: h=3-Step Ahead Relative MSFEs of Forecast Combination Models*

	Model	rMSFE				
	Maturity	3 months	1 year	3 years	5 years	10 years
	All	0.370***	0.442***	0.363***	0.378***	0.394***
	FB	0.415***	0.510***	0.436***	0.454***	0.483***
	FB(1)	0.387***	0.456***	0.401***	0.424***	0.452***
	FB(2)	0.444***	0.548***	0.458***	0.474***	0.504***
	FB(3)	0.431***	0.542***	0.464***	0.480***	0.514***
	FS	**0.294*****	**0.317*****	**0.217*****	**0.222*****	**0.212*****
2001:Jan - 2005:Dec	Econometrics	0.661***	0.801***	0.906**	0.959	1.065
1st Subsample	AR	0.658***	0.837***	1.028	1.087	1.126
	VAR	0.688***	0.787***	0.821***	0.875**	1.040
	DNS	0.628***	0.850***	0.876*	0.986	1.012
	NSS	1.049	1.121	0.830**	0.806**	0.880**
	NS(AR)	0.941	1.140	0.976	1.006	1.018
	NS(VAR)	0.750***	0.923**	0.800**	0.822**	0.898*
	NS(OLS)	0.728***	0.896**	0.860**	0.933*	0.927*
	NS(NLS)	0.852**	1.018	0.849*	0.872*	0.979
	All	0.378***	0.360***	0.422***	0.422***	0.367***
	FB	0.467***	0.462***	0.509***	0.503***	0.448***
	FB(1)	0.462***	0.460***	0.518***	0.510***	0.447***
	FB(2)	0.470***	0.459***	0.507***	0.504***	0.452***
	FB(3)	0.483***	0.484**	0.510***	0.500***	0.449***
	FS	**0.276*****	**0.238*****	**0.286*****	**0.279*****	**0.204*****
2006:Jan - 2010:Dec	Econometrics	0.898	0.896	0.963	0.937	0.965
2nd Subsample	AR	0.994	0.976	1.027	1.026	1.042
	VAR	1.033	0.959	0.946	0.881**	0.924
	DNS	0.966	0.847	1.025	1.065	0.961
	NSS	0.835	0.846	0.945	0.914	0.855
	NS(AR)	1.012	0.979	0.984	0.967	0.875*
	NS(VAR)	1.035	0.961	1.064	1.021	0.886
	NS(OLS)	0.909	0.850	1.008	1.085	0.933
	NS(NLS)	0.866	0.841	0.963	0.925	0.866*
	All	1.829	1.435	0.838*	0.613***	0.370***
	FB	2.172	1.715	0.995	0.722**	0.443***
	FB(1)	2.295	1.767	1.013	0.735**	0.449***
	FB(2)	2.200	1.702	1.009	0.735**	0.449***
	FB(3)	2.116	1.798	1.023	0.719**	0.438***
	FS	1.096	**0.849**	**0.504*****	**0.376*****	**0.215*****
2011:Jan - 2017:Oct	Econometrics	0.892	0.982	1.109	1.057	0.982
3rd Subsample	AR	1.334	1.255	1.131	1.086	1.034
	VAR	**0.819**	1.049	1.211	1.073	0.944*
	DNS	3.423	2.053	1.855	1.512	0.980
	NSS	3.369	3.446	1.738	1.204	0.807**
	NS(AR)	7.732	6.656	3.089	1.843	1.016
	NS(VAR)	0.967	0.867	1.177	1.083	0.821**
	NS(OLS)	3.118	1.776	1.517	1.485	0.924
	NS(NLS)	3.413	3.385	2.024	1.405	0.927
	All	0.426***	0.444***	0.426***	0.437***	0.377***
	FB	0.503***	0.535***	0.511***	0.520***	0.457***
	FB(1)	0.491***	0.508***	0.499***	0.512***	0.449***
	FB(2)	0.520***	0.553***	0.522***	0.532***	0.467***
	FB(3)	0.516***	0.564***	0.528***	0.530***	0.465***
	FS	**0.315*****	**0.303*****	**0.269*****	**0.271*****	**0.211*****
2001:Jan - 2017:Oct	Econometrics	0.781***	0.849***	0.946*	0.968	1.002
Whole Sample	AR	0.841*	0.914	1.036	1.064	1.065
	VAR	0.856***	0.872***	0.905**	0.912**	0.967
	DNS	0.889**	0.895*	1.017	1.110	0.984
	NSS	1.032	1.088	0.951	0.918*	0.845***
	NS(AR)	1.220	1.279	1.148	1.140	0.972
	NS(VAR)	0.892***	0.937**	0.941	0.944	0.866***
	NS(OLS)	0.900*	0.909	0.975	1.089	0.928*
	NS(NLS)	0.951	1.030	0.991	0.987	0.924*

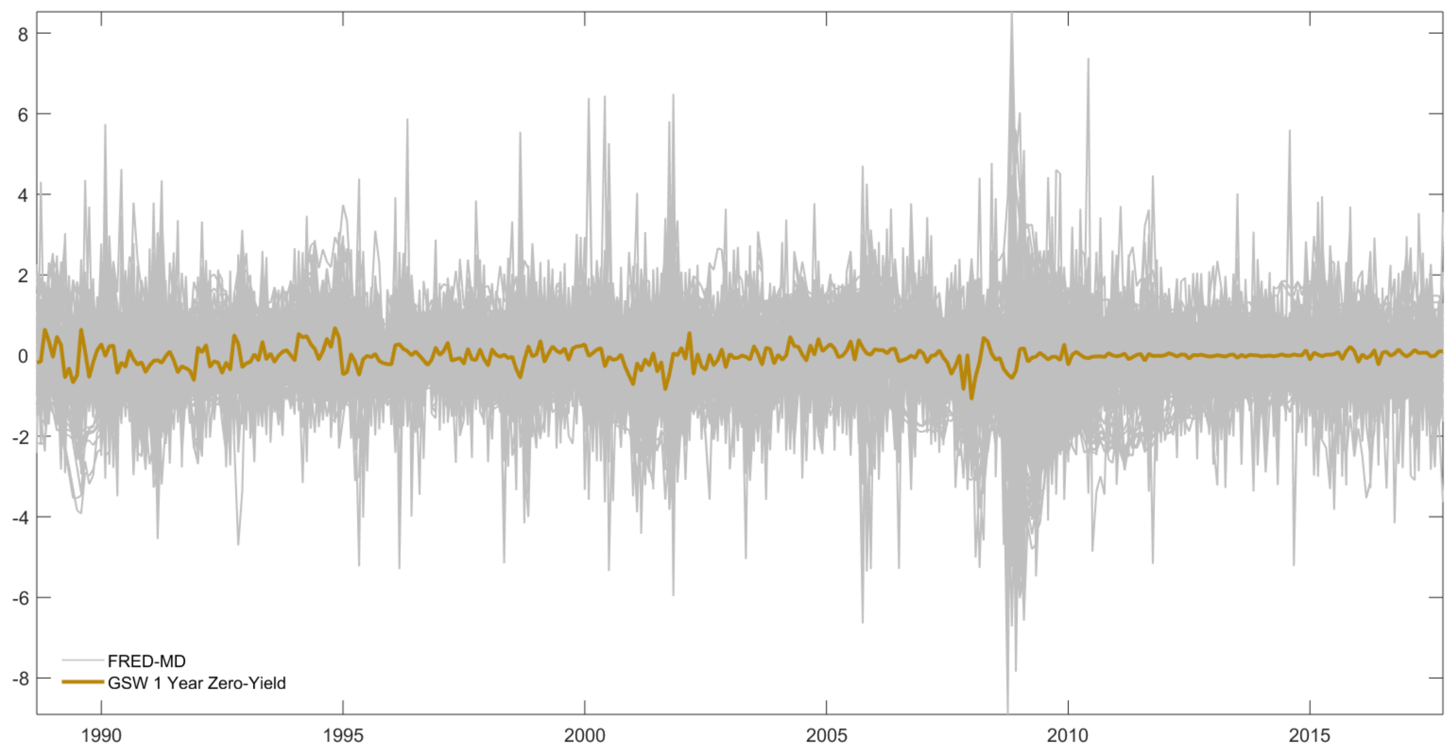
* Notes: See notes to Table 3.9A.

Table 3.9C: h=12-Step Ahead Relative MSFEs of Forecast Combination Models*

	Model	rMSFE				
	Maturity	3 months	1 year	3 years	5 years	10 years
	All	0.261***	0.222***	0.157***	0.163***	0.215***
	FB	0.318***	0.276***	0.203***	0.214***	0.287***
	FB(1)	0.264***	0.228***	0.162***	0.169***	0.227***
	FB(2)	0.348***	0.302***	0.227***	0.242***	0.327***
	FB(3)	0.354***	0.309***	0.232***	0.245***	0.329***
	FS	**0.143*****	**0.111*****	**0.065*****	**0.065*****	**0.084*****
2001:Jan - 2005:Dec	Econometrics	0.638***	0.627***	0.590***	0.649***	0.874***
1st Subsample	AR	0.831***	0.958	1.084	1.294	1.746
	VAR	0.631***	0.518***	0.367***	0.354***	0.457***
	DNS	0.569***	0.505***	0.425***	0.480***	0.632***
	NSS	0.658***	0.568***	0.383***	0.378***	0.506***
	NS(AR)	0.709***	0.658***	0.566***	0.647***	0.943
	NS(VAR)	0.647***	0.542***	0.362***	0.337***	0.387***
	NS(OLS)	0.618***	0.552***	0.469***	0.538***	0.695***
	NS(NLS)	0.610***	0.529***	0.374***	0.378***	0.510***
	All	0.299***	0.292***	0.354***	0.398***	0.390***
	FB	0.361***	0.353***	0.416***	0.464***	0.473***
	FB(1)	0.322***	0.309***	0.371***	0.425***	0.444***
	FB(2)	0.381***	0.377***	0.442***	0.489***	0.498***
	FB(3)	0.387***	0.383***	0.447***	0.493***	0.502***
	FS	**0.239*****	**0.235*****	**0.271*****	**0.288*****	**0.229*****
2006:Jan - 2010:Dec	Econometrics	0.688***	0.661***	0.741***	0.776***	0.842**
2nd Subsample	AR	0.649***	0.616***	0.728***	0.820***	1.047
	VAR	0.860***	0.809***	0.791***	0.769***	0.799*
	DNS	0.708***	0.669***	0.798***	0.894**	0.910*
	NSS	0.681***	0.685***	0.818***	0.895**	0.882**
	NS(AR)	0.665***	0.645***	0.791***	0.940	1.099
	NS(VAR)	0.877**	0.839***	0.883**	0.892**	0.829*
	NS(OLS)	0.700***	0.668***	0.804***	0.933	0.827**
	NS(NLS)	0.691***	0.680***	0.808***	0.872***	0.934*
	All	0.701***	0.661***	0.658***	0.508***	0.355***
	FB	0.905	0.858**	0.835***	0.632***	0.438***
	FB(1)	0.955	0.902*	0.872**	0.658***	0.453***
	FB(2)	0.927	0.876**	0.857***	0.651***	0.452***
	FB(3)	0.892	0.860*	0.822**	0.613***	0.422***
	FS	0.312***	**0.292*****	**0.307*****	**0.251*****	**0.180*****
2011:Jan - 2017:Oct	Econometrics	0.831***	0.867***	0.948*	0.907***	0.886***
3rd Subsample	AR	2.543	2.364	1.760	1.271	1.044
	VAR	0.222***	0.317***	0.520***	0.656***	0.763***
	DNS	1.595	1.268	1.509	1.248	0.932
	NSS	1.342	1.488	1.312	0.997	0.747***
	NS(AR)	4.543	4.391	3.570	2.222	1.281
	NS(VAR)	**0.194*****	0.311***	0.536***	0.628***	0.650***
	NS(OLS)	1.232	0.940	1.286	1.253	0.884***
	NS(NLS)	1.572	1.641	1.468	1.055	0.822***
	All	0.295***	0.261***	0.236***	0.264***	0.300***
	FB	0.361***	0.323***	0.295***	0.329***	0.378***
	FB(1)	0.317***	0.280***	0.258***	0.295***	0.353***
	FB(2)	0.387***	0.348***	0.319***	0.355***	0.405***
	FB(3)	0.391***	0.353***	0.322***	0.353***	0.395***
	FS	**0.188*****	**0.156*****	**0.131*****	**0.141*****	**0.147*****
2001:Jan - 2017:Oct	Econometrics	0.666***	0.647***	0.649***	0.714***	0.873***
Whole Sample	AR	0.835***	0.912**	1.039	1.186	1.350
	VAR	0.702***	0.597***	0.480***	0.489***	0.636***
	DNS	0.669***	0.585***	0.584***	0.681***	0.798***
	NSS	0.697***	0.641***	0.548***	0.581***	0.665***
	NS(AR)	0.862**	0.806***	0.809***	0.938	1.104
	NS(VAR)	0.717***	0.622***	0.501***	0.501***	0.565***
	NS(OLS)	0.677***	0.602***	0.602***	0.728***	0.792***
	NS(NLS)	0.684***	0.620***	0.549***	0.584***	0.705***

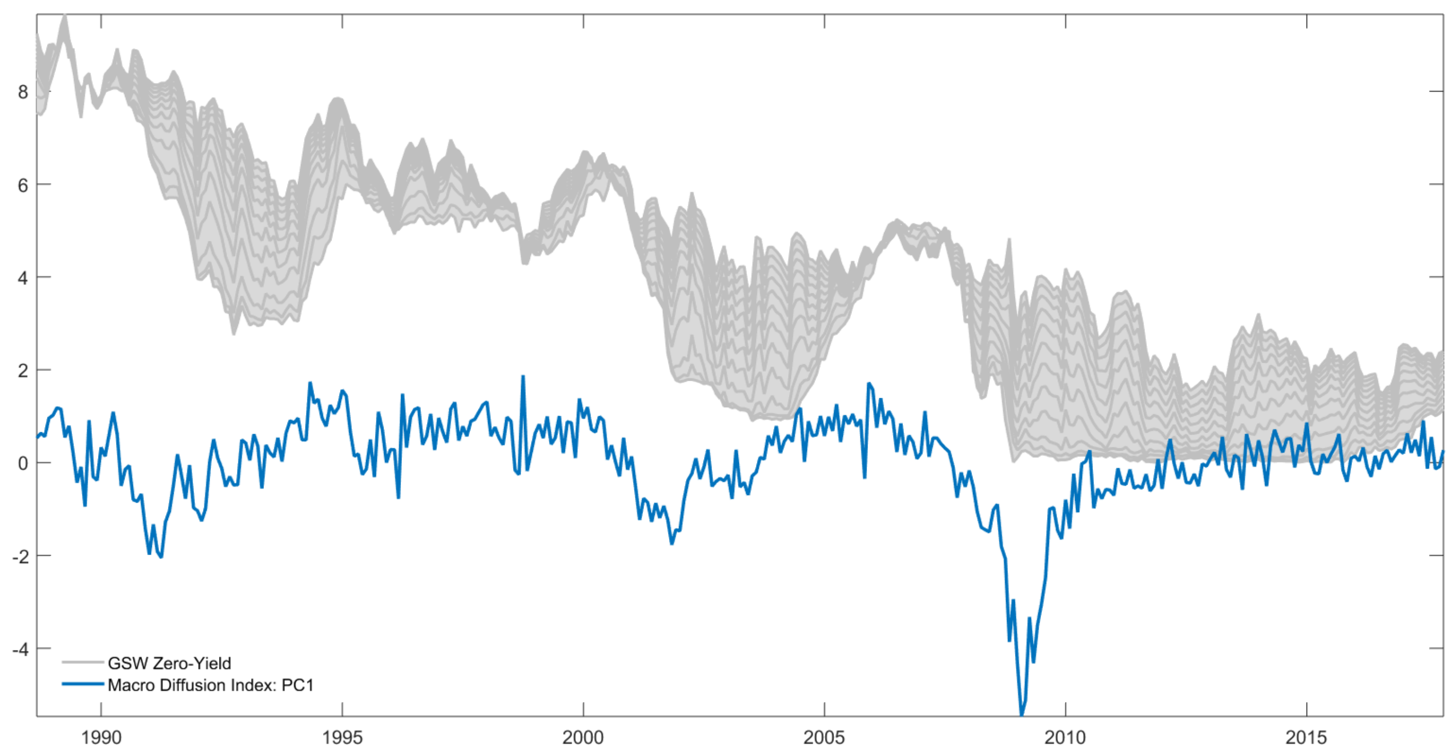
* Notes: See notes to Table 3.9A.

Figure 3.1: FRED MD Dataset for Sample Period 1988:8 - 2017:10*



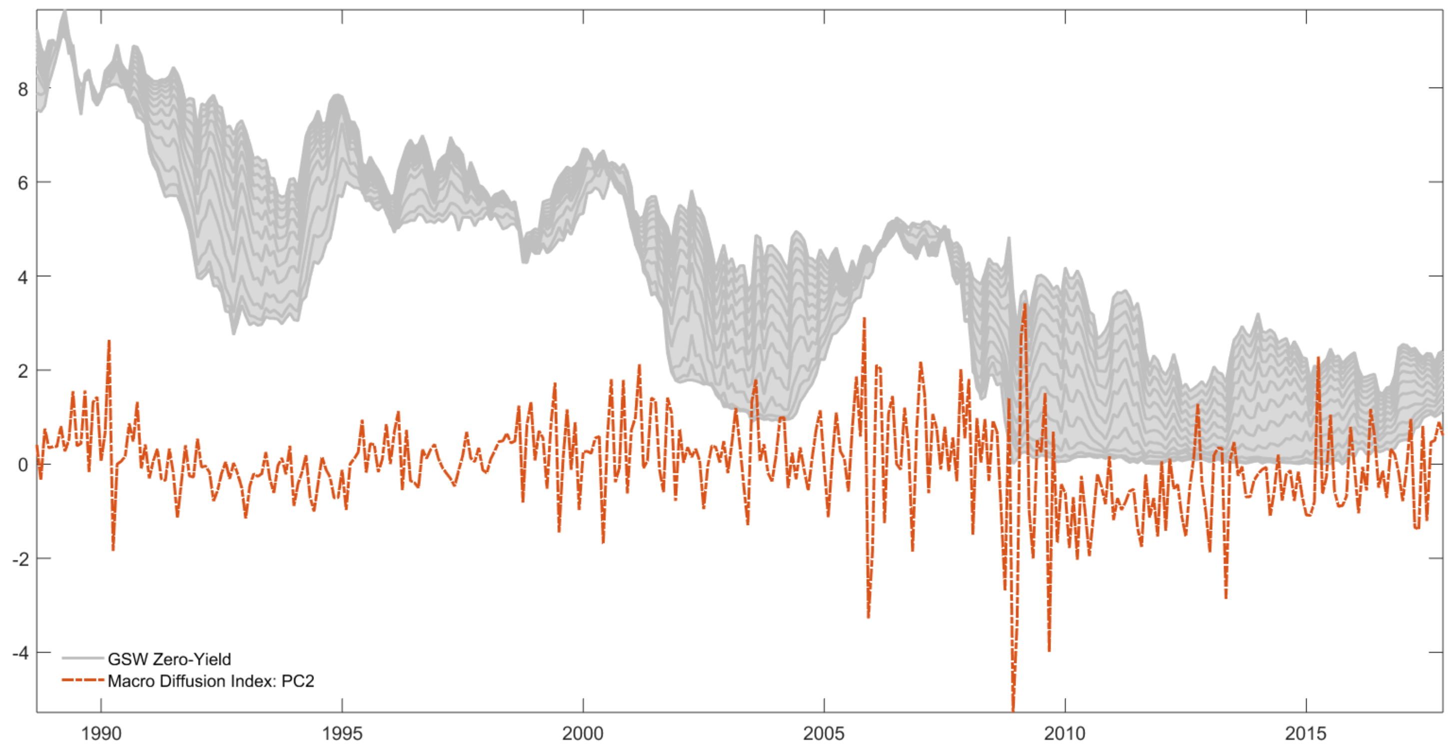
(*) Notes: This figure displays all macroeconomic variables from the FRED-MD dataset and the 12-dimensional zero-yield data from FRED H.15 and the GSW dataset. All non-yield variables are transformed to ensure stationary and all variables are standardized to zero mean and unit variance.

Figure 3.2A: Yields and Macroeconomic Diffusion Indexes (First Principal Component: PC1) for the Sample Period 1988:8 - 2017:10*



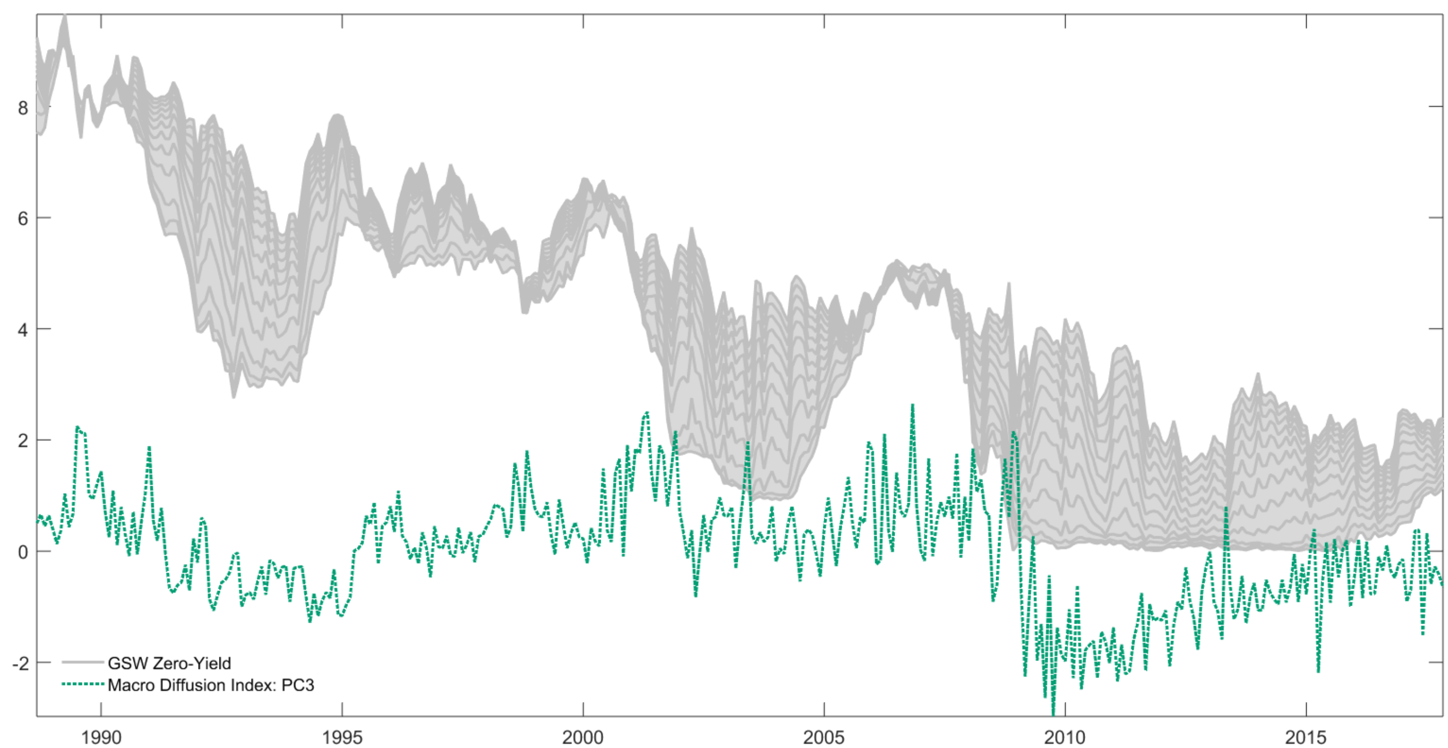
(*) Notes: This figure displays all twelve-dimensional zero-yield data collected from FRED H.15 and from the GSW dataset and one macroeconomic diffusion index (PC1).

Figure 3.2B: Yields and Macro Diffusion Indexes (Second Principal Component: PC2) for the Sample Period 1988:8 - 2017:10*



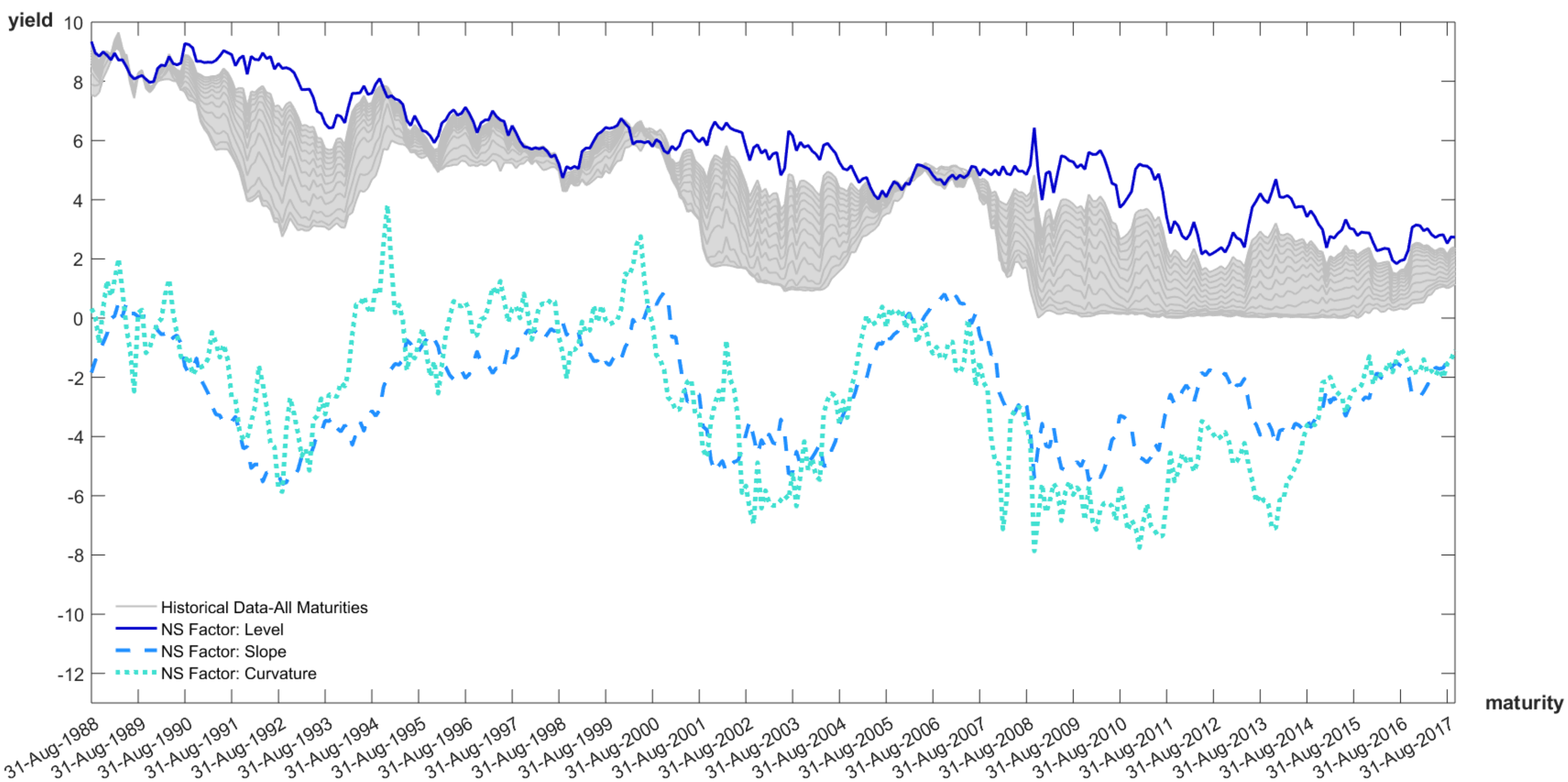
(*) Notes: This figure displays all twelve-dimensional zero-yield data collected from FRED H.15 and from the GSW dataset and one macroeconomic diffusion index (PC2).

Figure 3.2C: Yields and Macro Diffusion Indexes (Third Principal Component: PC3) for Sample Period 1988:8 - 2017:10*



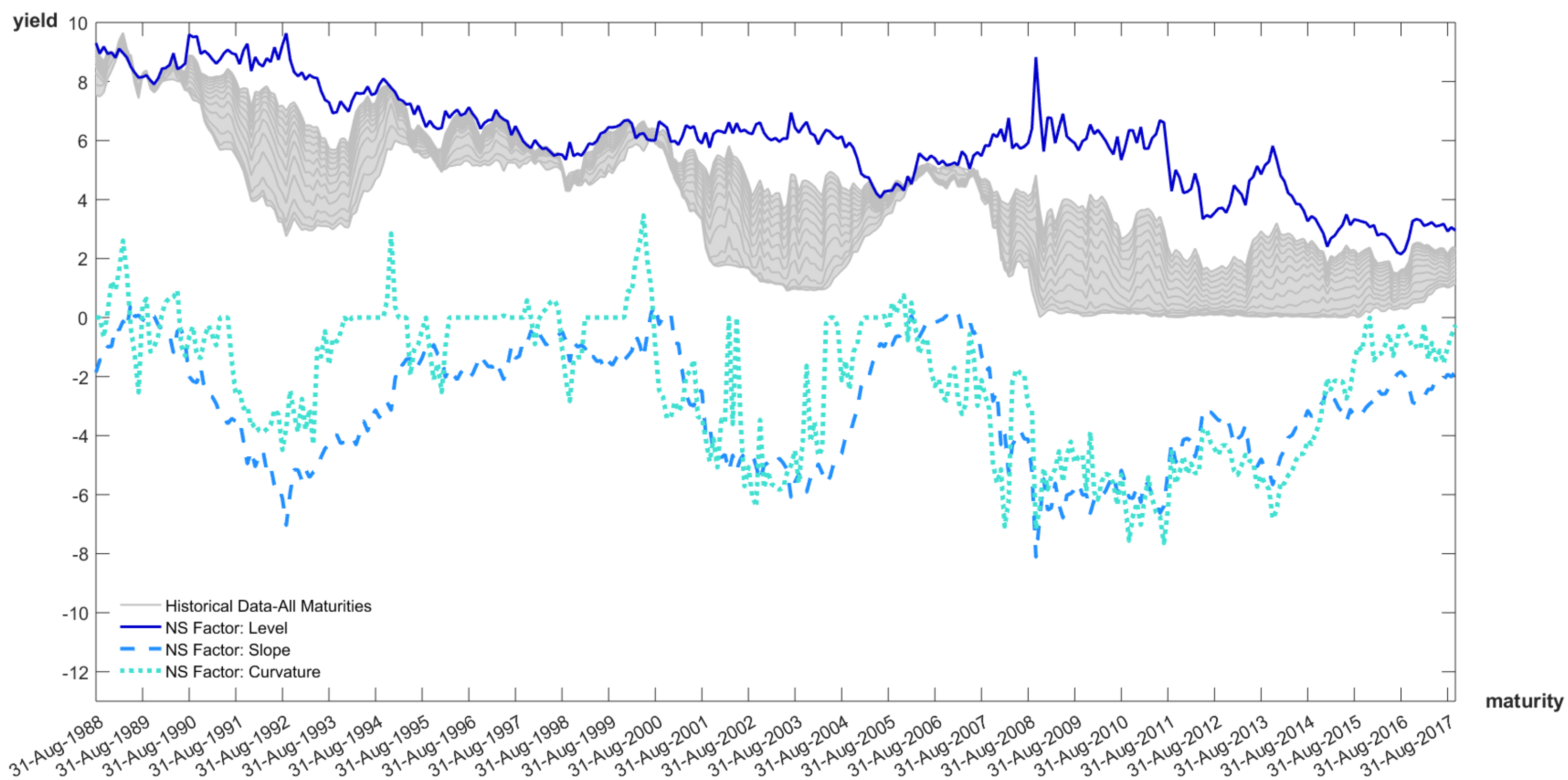
(*) Notes: This figure displays all twelve-dimensional zero-yield data collected from FRED H.15 and from the GSW dataset and one macroeconomic diffusion index (PC3).

Figure 3.3A: Yields and Dynamic Nelson Siegel Factors for the Sample Period 1988:8 - 2017:10*



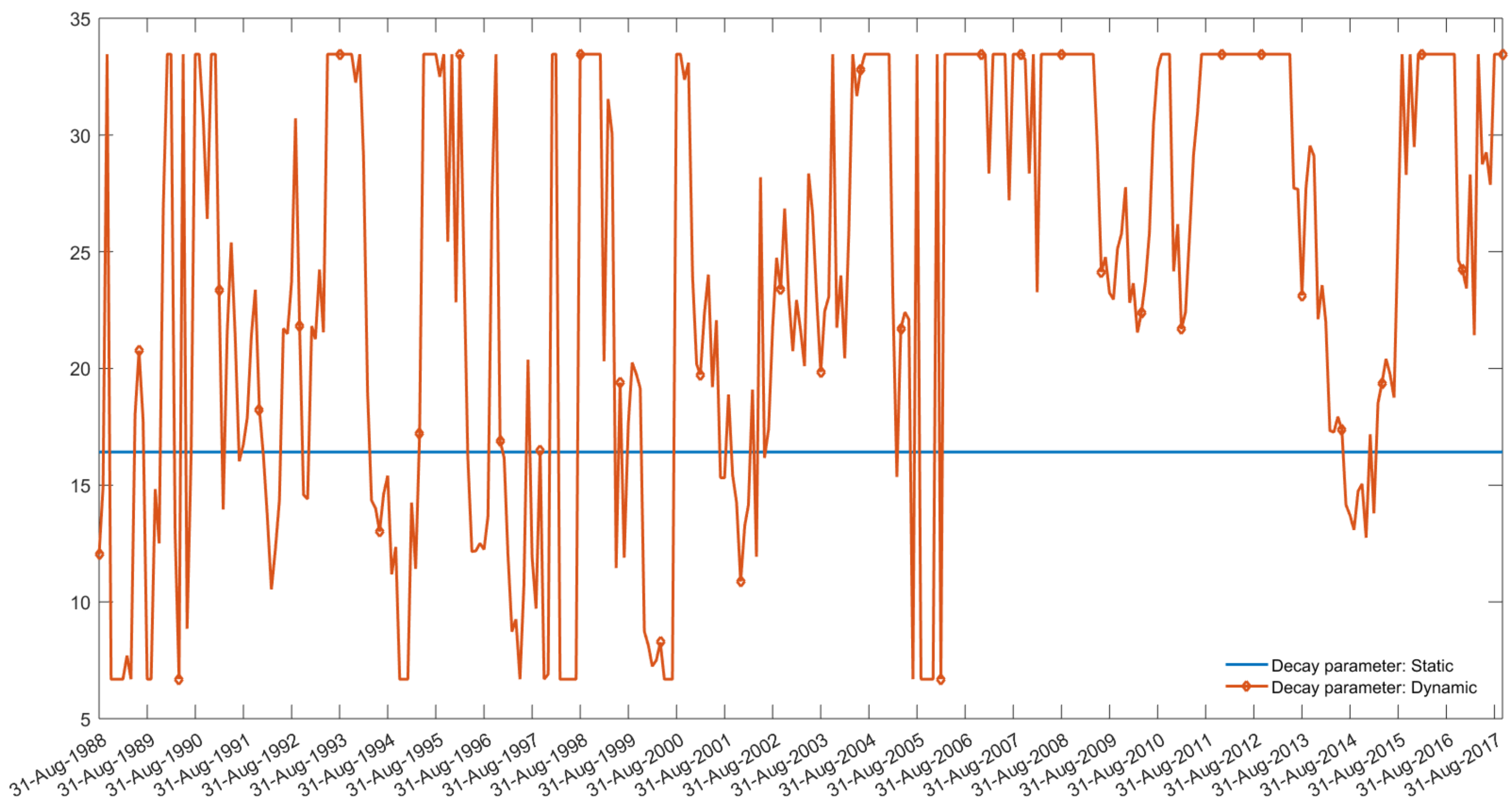
(*) Notes: This figure displays all twelve zero-yields from FRED H.15 and the GSW dataset and three Nelson-Siegel latent factors (level, slope, and curvature) with a fixed rate of decay estimated using ordinary least squares method.

Figure 3.3B: Yields and Dynamic Nelson Siegel Factors for the Sample Period 1988:8 - 2017:10*



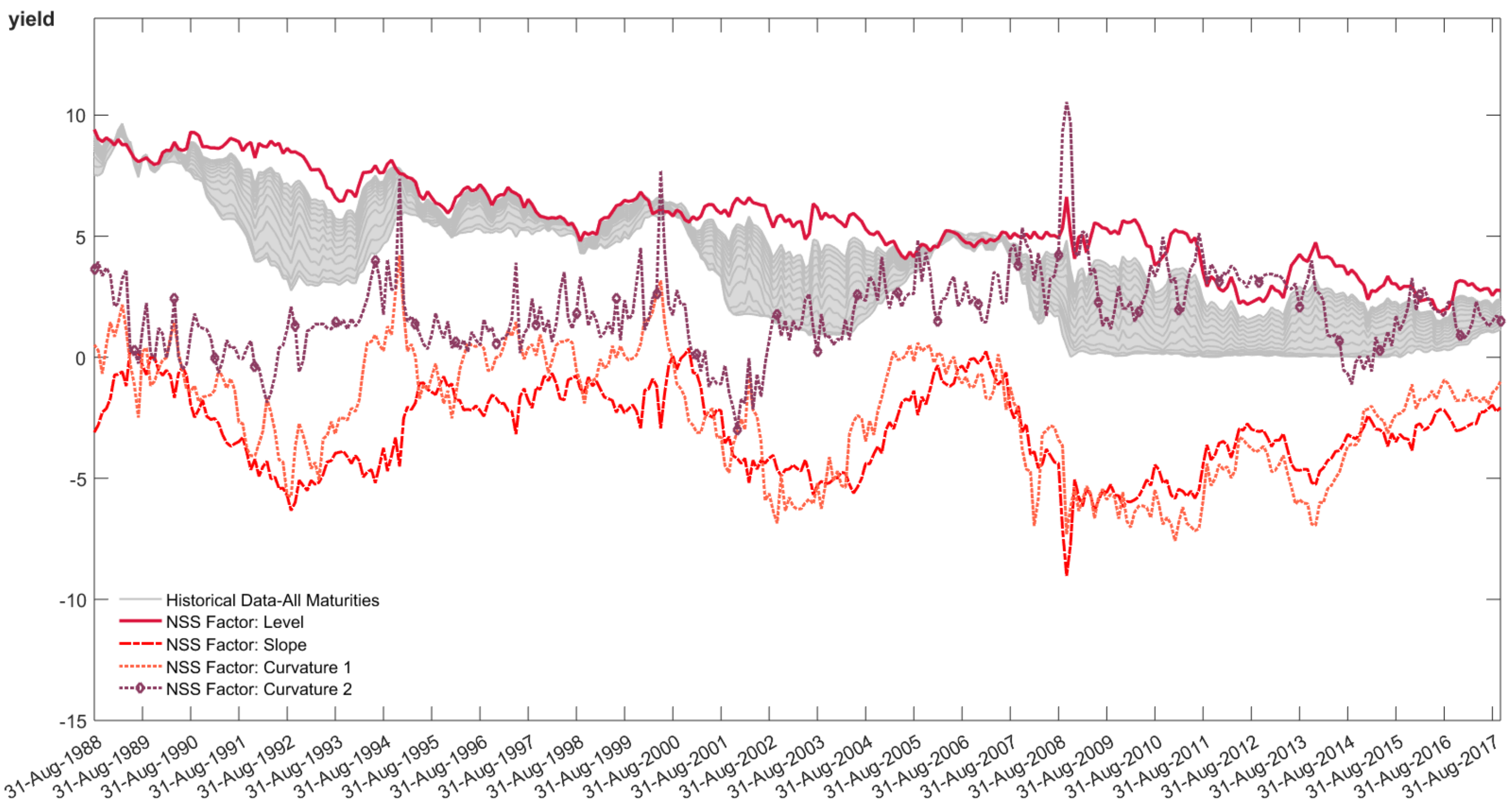
(*) Notes: This figure displays all twelve zero-yields from FRED H.15 and the GSW dataset and three Nelson-Siegel latent factors (level, slope, and curvature) estimated with dynamic rate of decay estimated using non-linear least squares method.

Figure 3.3C: Inverse Decay Parameter in Dynamic Nelson Siegel Model for the Sample Period 1988:8 - 2017:10*



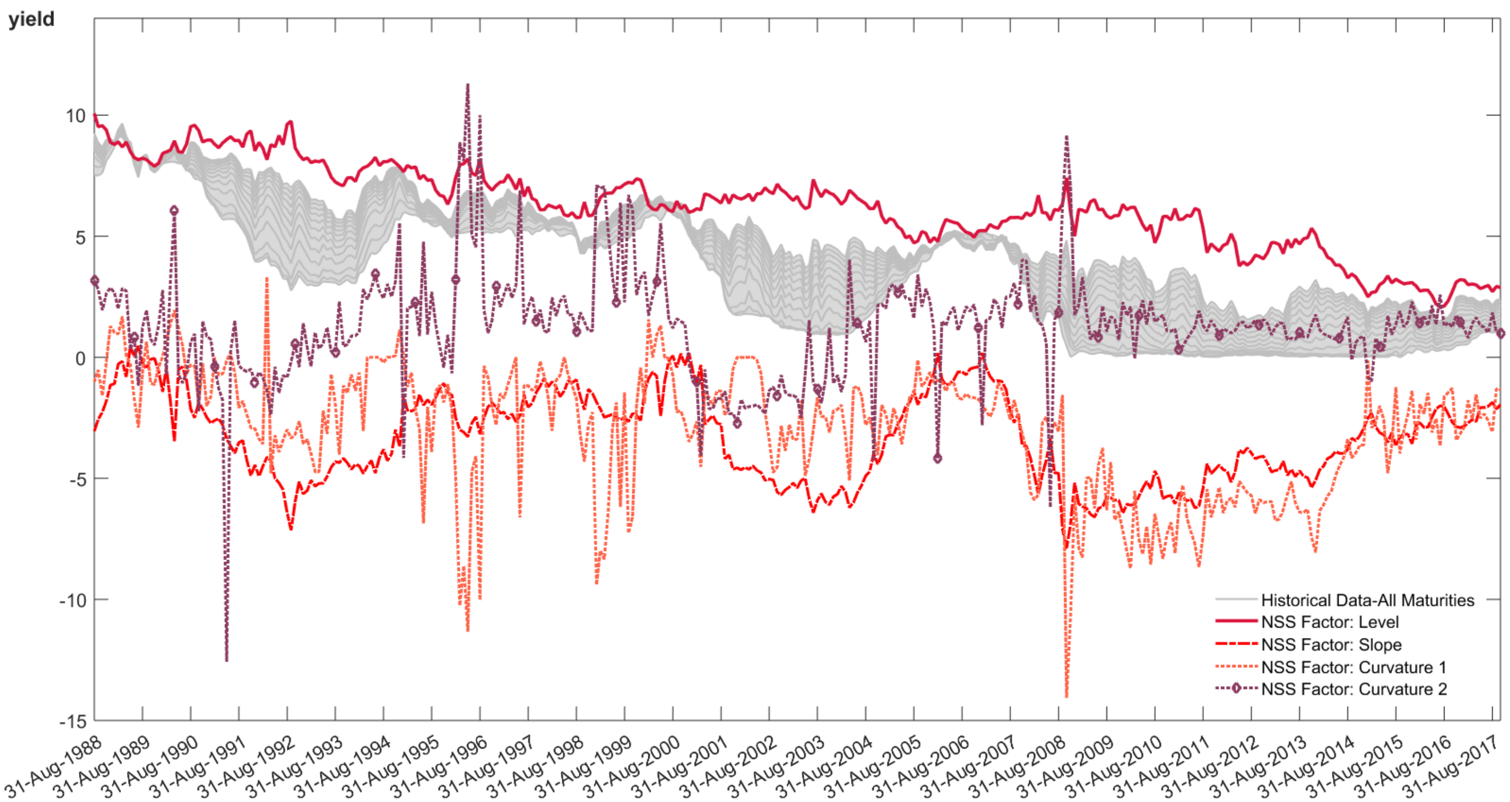
(*) Notes: This figure displays both static and (estimated) dynamic inverse rates of decay used in dynamic Nelson-Siegel modeling. See discussion in Section 4.1.2.

Figure 3.4A: Yields and Dynamic Nelson Siegel Svensson Factors for the Sample Period 1988:8 - 2017:10*



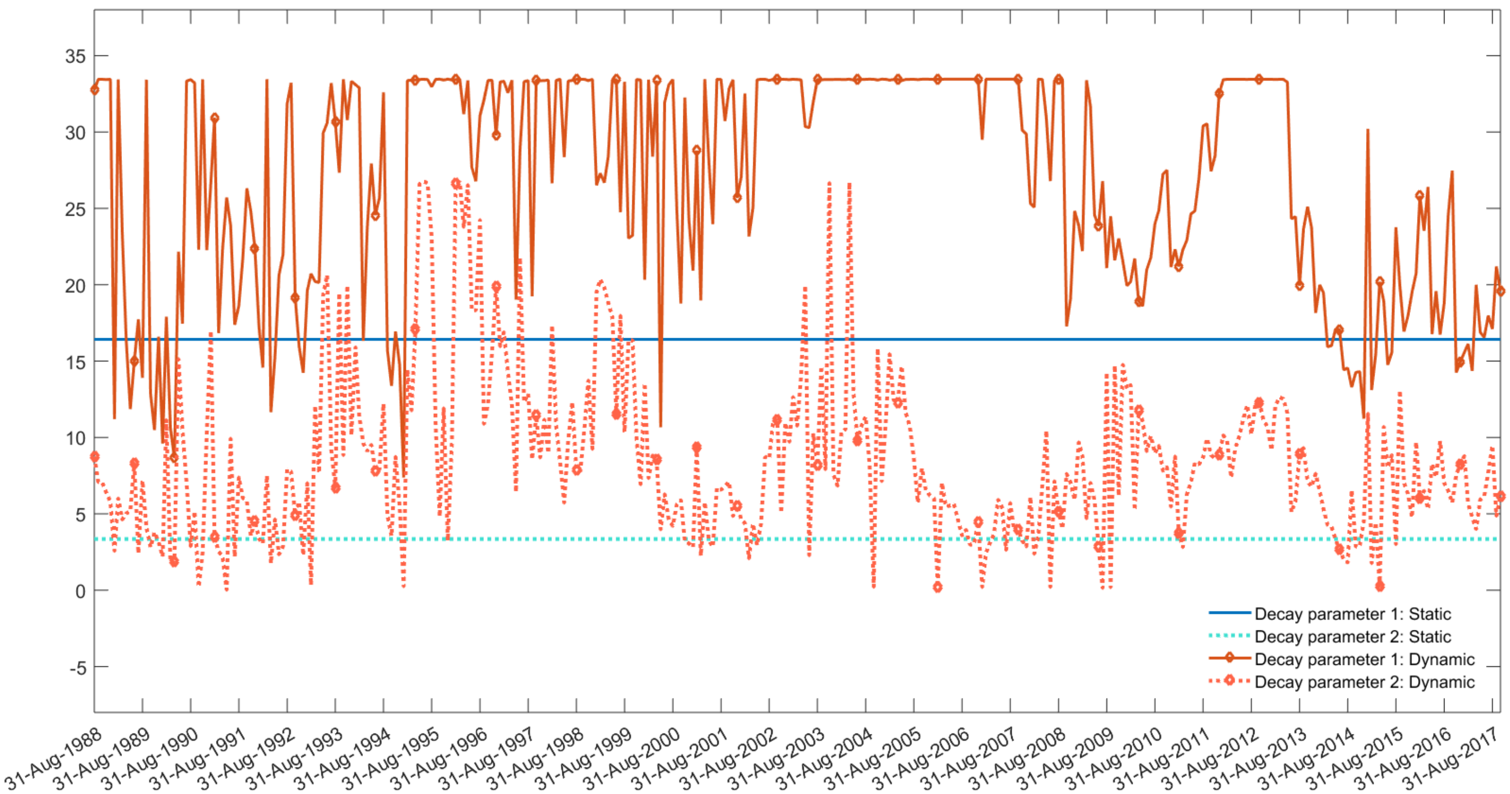
(*) Notes: This figure displays all twelve zero-yields from FRED H.15 and the GSW dataset and four Nelson-Siegel-Svensson latent factors (level, slope, and two curvatures) with fixed rates of decay estimated using ordinary least squares method.

Figure 3.4B: Yields and Dynamic Nelson Siegel Svensson Factors for the Sample Period 1988:8 - 2017:10*



(*) Notes: This figure displays all twelve zero-yields from FRED H.15 and the GSW dataset and four Nelson-Siegel-Svensson latent factors (level, slope, and two curvatures) estimated with dynamic rates of decay estimated using non-linear least squares method.

Figure 3.4C: Inverse Decay Parameter in Dynamic Nelson Siegel Svensson Model for the Sample Period 1988:8 - 2017:10*



(*) Notes: This figure displays both static and (estimated) dynamic inverse rates of decay used in dynamic Nelson-Siegel-Svensson modeling. See discussion in section 4.1.2.

3.6 References

Altavilla, C., R. Giacomini and G. Ragusa (2016), Anchoring the Yield Curve Using Survey Expectations, Working Paper, University College London.

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.

Armah, N.A. and N.R. Swanson (2010a), Seeing Inside the Black Box: Using Diffusion Index Methodology to Construct Factor Proxies in Largescale Macroeconomic Time Series Environments, *Econometric Reviews* 29, 476-510.

Armah, N.A. and N.R. Swanson (2010b), Diffusion Index Models and Index Proxies: Recent Results and New Directions, *European Journal of Pure and Applied Mathematics* 3, 2010, 478-501.

Aruoba, S. B., Diebold, F. X., and Scotti, C. (2012), Real-time measurement of business conditions, *Journal of Business & Economic Statistics*, 27(4), 417-427.

Backus, D., S. Foresi, and C. Telmer (1998), Discrete-Time Models of Bond Pricing, *NBER Working paper 6736*.

Banerjee, A., M. Marcellino and I. Marsten (2008) Chapter 4 Forecasting Macroeconomic Variables Using Diffusion Indexes in Short Samples with Structural Change. In Forecasting in the Presence of Structural Breaks and Model Uncertainty (pp. 149-194). Emerald Group Publishing Limited.

Bauer, G. and A.D. de los Rios (2012), An International Dynamic Term Structure Model with Economic Restrictions and Unspanned Risks, Working Paper #2012-5, Bank of Canada.

Bai, J. and Ng, S. (2002), Determining the Number of Factors in Approximate Factor Models, *Econometrica*, 70, 191-221.

Bai, J. and Ng, S. (2006a), Confidence Intervals for Diffusion Index Forecasts and Inference for Factor-Augmented Regressions, *Econometrica*, 74(4), 1133-1150.

Bai, J. and Ng, S. (2006b), Evaluating Latent and Observed Factors in Macroeconomics and Finance, *Journal of Econometrics*, 113, 507-537.

Bai, J. and Ng, S. (2007), Determining the Number of Primitive Shocks in Factor Models, *Journal of Business and Economic Statistics* 25, 52-60.

Bai, J. and Ng, S. (2008), Forecasting Economic Time Series Using Targeted Predictors, *Journal of Econometrics* 146, 304-317.

Bai, J. and Ng, S. (2009), Boosting Diffusion Indices, *Journal of Applied Econometrics* 24, 607-629.

Bekaert, G., Hodrick, R. J. and Marshall, D. A. (1997), On biases in tests of the expectations hypothesis of the term structure of interest rates, *Journal of Financial Economics*, 44(3), 309-348.

BIS (2005), Zero-Coupon Yield Curves: Technical Documentation, *BIS papers*, 25.

Boivin, J. and S. Ng (2005), Understanding and Comparing Factor Based Macroeconomic Forecasts, *International Journal of Central Banking* 1, 117-152.

Carrasco, M. and B. Rossi (2016), In-Sample Inference and Forecasting in Misspecified Factor Models, *Journal of Business and Economic Statistics* 34, 313-338.

Castle, J.A., M.P. Clements and D.F. Hendry (2013), Forecasting by Factors, by Variables or Both?, *Journal of Econometrics* 177, 305-319.

Chao, J.C., V. Corradi and N.R. Swanson (2001), An out of Sample Test for Granger Causality, *Macroeconomic Dynamics* 5, 598-620.

Chamberlain, G. and Rothschild M. (1983), Arbitrage, Factor Structure, and Mean-Variance Analysis on Large Asset Markets, *Econometrica*, Vol. 51, No. 5, pp. 1281-1304.

Christensen, J.H.E., 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.

Clark, T. E. and McCracken, M. W. (2009), Improving Forecast Accuracy by Combining Recursive and Rolling Forecast, *International Economic Review*, 50(2), 363-395.

Corradi, V., S. Jin, and N.R. Swanson (2017), Robust Forecast Comparison, *Econometric Theory*, forthcoming.

Corradi, V. and N.R. Swanson (2002), A Consistent Test for Nonlinear out of Sample Predictive Accuracy, *Journal of Econometrics* 110, 353-381.

Corradi, V. and N.R. Swanson (2005), A Test for Comparing Multiple Misspecified Conditional Interval Models, *Econometric Theory* 21, 991-1016.

Corradi, V. and N.R. Swanson (2006a), Predictive Density and Conditional Confidence Interval Accuracy Tests, *Journal of Econometrics* 135, 187-228.

Corradi, V. and N.R. Swanson (2006b), Predictive Density Evaluation, in G. Elliot, C. W. J. Granger, and A. Timmermann, (eds.), **Handbook of Economic Forecasting**, Volume 1, Elsevier, Amsterdam.

Corradi, V. and N.R. Swanson (2007), Nonparametric Bootstrap Procedures For Predictive Inference Based On Recursive Estimation Schemes, *International Economic Review* 48, 67–109.

Corradi, V. and W. Distaso (2011), Multiple Forecast Evaluation, in D. F. Hendry and M. P. Clements (eds.), **Oxford Handbook of Economic Forecasting**, Oxford University Press, Oxford.

Corradi, V. and N.R. Swanson (2013), A Survey of Recent Advances in Forecast Accuracy Comparison Testing, with an Extension to Stochastic Dominance, in X. Chen and N.R. Swanson (eds.), **Causality, Prediction, and Specification Analysis: Recent Advances and Future Directions, Essays in honor of Halbert L. White, Jr.**, Springer.

Corradi, V. and N.R. Swanson (2014), Testing for Structural Stability of Factor Augmented Forecasting Models, *Journal of Econometrics* 182, 2014, 100-118.

Corradi, V., N.R. Swanson and C. Olivetti (2001), Predictive Ability with Cointegrated Variables, *Journal of Econometrics* 104, 315-358.

Coroneo, L., D. Giannone and M. Modugno (2016), Unspanned Macroeconomic Factors in the Yield Curve, *Journal of Business and Economic Statistics* 34, 472-485.

Cox, J., J. E. Ingersoll, and S. A. Ross (1981), A Reexamination of Traditional Hypotheses About the Term Structure of Interest Rates, *Journal of Finance* 36, 769-799.

Cox, J., J. E. Ingersoll, and S. A. Ross (1985a), An intertemporal general equilibrium model of asset prices, *Econometrica* 53, 363-384.

Cox, J., J. E. Ingersoll, and S. A. Ross (1985b), A Theory of the Term Structure of Interest Rates, *Econometrica*, 53, 385-407.

Culberison, J. M. (1957), The Term Structure of Interest Rates, *Quarterly Journal of Economics*, 71(1957), 485-517.

De Pooter, M., Examining the Nelson-Siegel Class of Term Structure Models: In-Sample Fit versus Out-of-Sample Forecasting Performance (2007), Available at *SSRN*: https://ssrn.com/abstract=992748 or http://dx.doi.org/10.2139/ssrn.992748.

De Pooter, M., Ravazzolo, F., and Van Dijk, D. J. (2010), Term structure forecasting using macro factors and forecast combination, *FRB International Finance Discussion Paper*, (993).

Diebold, F.X. and Li, C. (2006), Forecasting the Term Structure of Government Bond Yields, *Journal of Econometrics*, 130(2), 337-364.

Diebold, F.X. and R.S. Mariano (1995), Comparing Predictive Accuracy, *Journal of Business and Economic Statistics* 13, 253-263.

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.

Diebold, F.X. and G.D. Rudebusch (2013), Yield Curve Modeling and Forecasting: The Dynamic Nelson-Siegel Approach, *Princeton University Press: Princeton.*

Ding, A. A., and Hwang, J. T. (1999), Prediction Intervals, Factor Analysis Models, and High-Dimensional Empirical Linear Prediction, *Journal of the American Statistical Association*, 94, 446455.

Duffee, G. R. (2002), Term premia and interest rate forecasts in affine models, *The Journal of Finance*, 57(1), 405-443.

Duffee, G.R. (2011), Information In (and Not In) the Term Structure, *Review of Financial Studies* 24, 2895–2934.

Duffee, G. R. (2012), Forecasting interest rates, *Handbook of Economic Forecasting*, Vol. 2.

Duffie, D. and R. Kan (1996), A Yield-Factor Model of Interest Rates, *Mathematical Finance*, 6, 379-406.

Exterkate, P., Dijk, D.V., Heij, C., and Groenen, P.J. (2013), Forecasting the Yield Curve in a DataRich Environment Using the FactorAugmented NelsonSiegel Model, *Journal of Forecasting*, 32(3), 193-214.

Faust, J. and Leeper, E. M. (2015), The myth of normal: The bumpy story of inflation and monetary policy. In *Prepared for the Federal Reserve Bank of Kansas Citys Jackson Hole Symposium.*

Forni, M. and Reichlin, L. (1996), Dynamic Common Factors in Large Cross-Sections, *Empirical Economics*, 21, 27-42.

Forni, M. and Reichlin, L. (1998), Lets Get Real: A Dynamic Factor Analytical Approach to Disaggregated Business Cycle, *Review of Economic Studies*, 65, 453-474.

Fernandez A. and N.R. Swanson (2017), Real-Time Datasets Really Do Make a Difference: Definitional Change, Data Release and Forecasting, *Quantitative Finance and Economics* 1(1): 2-25.

Geweke, J. (1977), The Dynamic Factor Analysis of Economic Time Series," in *Latent Variables in Socio-Economic Models*, eds. D. J. Aigner and A. S. Goldberger, Amsterdam: North-Holland, Ch. 19.

Giacomini, R. and H. White (2006), Tests of Conditional Predictive Ability, *Econometrica* 74, 1545-1578.

Giacomini R. and B. Rossi (2009), Detecting and Predicting Forecast Breakdowns, *Review of Economic Studies* 76, 669-705.

Goncalves S. and B. Perron (2014), Bootstrapping Factor-Augmented Regression Models, *Journal of Econometrics* 182, 156-173

Goncalves S. and H. White (2004), Maximum Likelihood and the Bootstrap for Nonlinear Dynamic Models, *Journal of Econometrics* 119, 199-219

Granger, C.W.J. (1999a), Outline of Forecast Theory using Generalized Cost Functions, *Spanish Economic Review* 1, 161-173

Granger, C.W.J. (1999b), **Empirical Modeling in Economics: Specification and Evaluation**, Cambridge University Press, New York.

Granger, C.W.J. and M. H. Pesaran (2000), Economic and Statistical Measures of Forecast Accuracy, *Journal of Forecasting* 19, 537-560.

Gürkaynak, R.S., B. Sack, and J.H. Wright (2006), The U.S. Treasury Yield Curve: 1961 to the Present, Finance and Economics Discussion Series 2006-28.

Gürkaynak, R.S. and Wright, J. H. (2012), Macroeconomics and the term structure, *Journal of Economic Literature*, 50(2), 331-67.

Hamilton, J. D. and Wu, J. C. (2012), The effectiveness of alternative monetary policy tools in a zero lower bound environment, *Journal of Money, Credit and Banking*, 44(s1), 3-46.

Han, X. and A. Inoue (2014), Tests for Parameter Instability in Dynamic Factor Models, *Econometric Theory* 31, 1117-1152.

Hansen, P.R. and A. Timmermann (2012), Choice of Sample Split in Out-of-Sample Forecast Evaluation, Working Paper, UCSD, Rady School of Management.

Hicks, J. R. (1946), Value and Capital, 2nd edition. London: Oxford University Press.

Hirano, K. and J.H. Wright (2017), Forecasting With Model Uncertainty: Representations and Risk Reduction, *Econometrica* 85, 617-643.

Hyvärinen, A. (1998), Independent Component Analysis in the Presence of Gaussian Noise by Maximizing Joint Likelihood, *Neurocomputing* 22, 49-67.

Hyvärinen, A. (1999), Survey on Independent Component Analysis, *Neural Computing Surveys* 2, 94-128.

Hyvärinen, A. and E. Oja (2000), Independent Component Analysis: Algorithms and Applications, *Neural Networks* 13, 411-430.

Kim, H. H. and Swanson, N. R. (2014), Forecasting financial and macroeconomic variables using data reduction methods: New empirical evidence, *Journal of Econometrics*, 178, 352-367.

Kim H.H. and N.R. Swanson (2017), Mining Big Data Using Parsimonious Factor, Machine Learning, Variable Selection, and Shrinkage Methods, *International Journal of Forecasting,* forthcoming.

Lawley, D. N. and Maxwell, A. E. (1971), Factor Analysis as a Statistical Method," New York: American Elsevier Publishing.

Longstaff, F. A. and Schwartz, E. S. (1992), Interest rate volatility and the term structure: A two-factor general equilibrium model, *The Journal of Finance*, 47(4), 1259-1282.

Ludvigson, S. C. and Ng, S. (2009), Macro Factors in Bond Risk Premia, *Review of Financial Studies*, 22(12), 5027-5067. bf

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.

McCulloch, J. H. (1975), The TaxAdjusted Yield Curve, *The Journal of Finance*, 30(3), 811-830.

Modigliani, F., and Sutch, R. (1966), Innovations in Interest Rate Policy, *American Economic Review*, 56(1966), 178-197.

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

Nelson, C. and Siegel, A. (1985), Parsimonious modeling of yield curves for US treasury bills, *NBER Working Paper Series*, 1594.

Nelson, C. and Siegel, A. (1987), Parsimonious Modeling of Yield Curves, *Journal of Business* 60, 473-489.

Rossi, B. and A. Inoue (2012), Out of Sample Forecast Tests Robust to Window Size Choice, *Journal of Business and Economic Statistics* 30, 432-453.

Rudebusch, G. D. and Wu, T. (2008), A Macro-Finance Model of the Term Structure, Monetary Policy and the Economy, *The Economic Journal*, 118(530), 906-926.

Sargent, T. J. and Sims, C. A. (1977), Business Cycle Modeling Without Pretending to Have Too Much A Priori Economic Theory, in *New Methods in Business Cycle Research*, eds. C. Sims et al., Minneapolis: Federal Reserve Bank of Minneapolis.

Stark, T. (2010), Realistic evaluation of real-time forecasts in the Survey of Professional Forecasters, *Federal Reserve Bank of Philadelphia Research Rap, Special Report.*

Schumacher, C. (2007), Forecasting German GDP Using Alternative Factor Models Based on Large Datasets, *Journal of Forecasting* 26, 271-302.

Schumacher, C. (2009), Factor Forecasting Using International Targeted Predictors: The Case of German GDP, *Economics Letters* 107, 95-98.

Stock, J. and Watson, M. (1998), Diffusion Indexes, Working Paper 6702, National Bureau of Economic Research.

Stock, J. and Watson, M. (1999), Forecasting Inflation, *Journal of Monetary Economics* 44, 293-335.

Stock, J.H. and M.W. Watson (2002a), Macroeconomic Forecasting Using Diffusion Indexes, *Journal of Business and Economic Statistics* 20, 147-162.

Stock, J.H. and M.W. Watson (2002b), Forecasting Using Principal Components from a Large Number of Predictors, *Journal of the American Statistical Association* 97, 1167-1179.

Stock, J.H. and M.W. Watson (2004), Combination Forecasts of Output Growth in a Seven-Countries Data-Set, *Journal of Forecasting* 23, 405-430.

Stock, J.H. and M.W. Watson (2009), Forecasting in Dynamic Factor Models Subject to Structural Instability, in J. Castle and N. Shephard (eds.), **The Methodology and Practice of Econometrics: Festschrift in Honour of D.F. Hendry**, Oxford University Press.

Stock, J. H. and M.W. Watson (2012), Generalized Shrinkage Methods for Forecasting Using Many Predictors, *Journal of Business and Economic Statistics* 30, 481-493.

Svensson, L. E. (1994), Estimating and interpreting forward interest rates: Sweden 1992-1994 (No. w4871), *National Bureau of Economic Research.*

Swanson, N.R. (2016), Comment on: In Sample Inference and Forecasting in Misspecified Factor Models, *Journal of Business and Economic Statistics* 34, 348-353.

Swanson, N.R. and Xiong, W. (2017), Big Data Analytics In Economics: What Have We Learned So Far, And Where Should We Go From Here?, *Canadian Journal of Economics*, forthcoming.

Timmermann, A. (2006), Forecast Combinations, in G. Elliot, C. W. J. Granger, and A. Timmermann, (eds.), Handbook of Economic Forecasting, Elsevier, Amsterdam.

Vasicek, O. A. (1977), An Equilibrium Characterization of the Term Structure, *Journal of Financial Economics*, 5, 177-188

West, K. D. (1996), Asymptotic Inference about Predictive Ability, *Econometrica* 64, 1067-1084.

West, K. D. and M. W. McCracken (1998), Regression Based Tests of Predictive Ability, *International Economic Review* 39, 817-840.

West, K. D. (2006), Forecast Evaluation, in G. Elliot, C. W. J. Granger, and A. Timmermann, (eds.), Handbook of Economic Forecasting, Elsevier, Amsterdam.

White, H. (2000), A Reality Check for Data Snooping, *Econometrica* 68, 1097-1126.

Yuan, M. and Y. Lin (2007), On the Non-Negative Garrotte Estimator, *Journal of the Royal Statistical Society* 69, 143-161.

Zou, H. (2006), The Adaptive Lasso and Its Oracle Properties, *Journal of the American Statistical Association* 101, 1418-1429.

Zou, H. and T. Hastie (2005), Regularization and Variable Selection via the Elastic Net, *Journal Of The Royal Statistical Society Series B* 67, 301-320.

Zou, H., T. Hastie, and R. Tibshirani (2006), Sparse Principal Component Analysis, *Journal of Computational and Graphical Statistics* 15, 262-286.