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

Norman R. Swanson and Weiqi Xiong
Rutgers University

September 2017

Abstract

Research into predictive accuracy testing remains at the forefront of the forecasting field. One reason for this is that rankings of predictive accuracy across alternative models, which under misspecification are loss function dependent, are universally utilized to assess the usefulness of econometric models. A second reason, which corresponds to the objective of this paper, is that researchers are currently focusing considerable attention on so-called big data, and on new (and old) tools that are available for the analysis of this data. One of the objectives in this field is the assessment of whether big-data leads to improvement in forecast accuracy. In this survey paper, we discuss some of the latest (and most interesting) methods currently available for analyzing and utilizing big data when the objective is improved prediction. Our discussion includes a summary of various so-called dimension reduction, shrinkage, and machine learning methods, as well as a summary of recent tools that are useful for ranking prediction models associated with the implementation of these methods. 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.

JEL Classification: C12, C22, C53.

Keywords: Convex loss function, Empirical processes, Forecast superiority, General loss function.

* Corresponding Author: Norman R. Swanson, Department of Economics, Rutgers University, 75 Hamilton Street, New Brunswick, NJ 08901, USA, nswanson@econ.rutgers.edu.

Weiqi Xiong, Department of Economics, Rutgers University, 75 Hamilton Street, New Brunswick, NJ 08901, USA, wxiong@economics.rutgers.edu.

This paper has been prepared for a State of the Art Lecture at the 2017 Canadian Economic Association Conference at St. Francis Xavier University. We are grateful to the editor, Francisco Ruge-Murcia and an anonymous referee for useful comments on earlier drafts of this paper. We also thank Mingmian Cheng, Valentina Corradi, Frank Diebold, Xu Jiang, Hyun Hak Kim, Dan Levenstein, Yuan Liao, Matt Lightwood, Hal Pedersen, Byung-Dong Seo, Rachel Swanson, Greg Tkacz, Xiye Yang, and participants of the 2017 Deutsche Bundesbank Workshop on Forecasting for useful comments and suggestions.

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 $e_1 < e_2 \leq 0$).

[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.

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, 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 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

[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).

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

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 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 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.

2.1 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

[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.

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 $r = (q+1)\bar{r}$, 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 $\bar{r} \times 1$ 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 $tr[\Lambda^{k\prime}X^{\prime}X\Lambda^{k}]$, the solution of which is to set $\overline{\Lambda}^{k}$ equal to the eigenvectors of the $N \times N$ matrix $X^{\prime}X$ that correspond to its k largest eigenvalues. One can thus estimate the factors as $\overline{F}^{k} = X^{\prime}\overline{\Lambda}^{k}/N$. $\widetilde{F}^{k}$ 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 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:

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

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.

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.

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:

$$m_P = P^{-1/2}\sum_{t=R+1}^{T} \widehat{u}_{0,t+h} X_t, \tag{3.1}$$

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:

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

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:

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

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

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

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

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).

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:

$$G(x) = (F_2(x) - F_1(x))sgn(x), \tag{3.2}$$

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

$$C(x) = \int_{-\infty}^{x} (F_1(t) - F_2(t))dt1(x < 0) + \int_{x}^{\infty} (F_2(t) - F_1(t))dt1(x \geq 0), \tag{3.3}$$

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

and

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

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).

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.

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 1, and include: AR(1), VAR(1), AR(SIC), and VAR(SIC))

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

$$y_{t+h}(\tau) = c + \beta' W_t + \varepsilon_{t+h}, \tag{4.1}$$

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 1, and include: DNS(1), DNS(2), DNS(3), DNS(4), DNS(5), and DNS(6))

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

$$\mathbf{y}_t(\tau) = \beta_{1,t} + \beta_{2,t}[\frac{1-\exp(-\lambda_t\tau)}{\lambda_t\tau}] + \beta_{3,t}[\frac{1-\exp(-\lambda_t\tau)}{\lambda_t\tau} - \exp(-\lambda_t\tau)] + \varepsilon_t, \tag{4.2}$$

[5] When specifying VAR models, equation (4.1) 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.

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:

$$y_{t+h}(\tau) = \widehat{\beta}^f_{1,t+h} + \widehat{\beta}^f_{2,t+h}[\frac{1-\exp(-\lambda_t\tau)}{\lambda_t\tau}] + \widehat{\beta}^f_{3,t+h}[\frac{1-\exp(-\lambda_t\tau)}{\lambda_t\tau} - \exp(-\lambda_t\tau)], \tag{4.3}$$

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:

$$\hat{\beta}^f_{i,t+h} = \widehat{c}_i + \widehat{\gamma}_{ii}\widehat{\beta}_{i,t}, \quad \text{for } i = 1, 2, 3, \tag{4.4}$$

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:

$$\hat{\beta}^f_{t+h} = \widehat{c} + \widehat{\gamma}\widehat{\beta}_t, \tag{4.5}$$

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 $\widehat{\beta}_{1,t}$ 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 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:

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

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 (4.5), 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 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 $\mathbf{y}^{10}_t(\tau)$,

$$y_{t+h}(\tau) = c + \beta' W_t + \alpha' F^s_t + \varepsilon_{t+h}, \tag{4.6}$$

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 (4.4) and (4.5).

Big Data Models

Diffusion Index Models:

(Models in this section are summarized in Table 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 (4.6), but with latent factors, say F^b_t, 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:

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

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^b_t 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 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:

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

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:

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

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_t^b is a scalar), or is a 3x2 matrix (if F_t^b 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]

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

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.

4.3 Empirical Findings

Tables 2A - 2D contain relative MSFEs for yield forecasts constructed using the models listed in Table 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 3A-D and 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 2A-4D) are given in Table 5; and Table 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 7 and 8A-C.

Turning to the results based on Tables 2A through 4D, a number of clearcut conclusions emerge.

First, inspection of the results in Tables 2A-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

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

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 2A-4D indicates that models listed as MSFE-best in Table 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 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 5-8.

In Table 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 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 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 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 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 8B). Similarly, for $h = 12$, combination models fail to "win" in 17 of 20 cases (see Table 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.

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.

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.

Bai, J. (2003), Inferential Theory for Factor Models of Large Dimensions, *Econometrica* 71, 135-171.

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

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

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

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

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

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

Banerjee, A., M. Marcellino and I. Marsten (2008), Forecasting Macroeconomic Variables Using Diffusion Indexes in Short Samples with Structural Changes, in D. Rapach and M. Wohar (eds.), **Forecasting in the Presence of Structural Breaks and Model Uncertainty**, Emerald Group Publishing, London.

Banerjee, A., M. Marcellino and I. Marsten (2014), Forecasting with Factor-Augmented Error Correction Models, *International Journal of Forecasting* 30, 589-612.

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.

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

Breiman, L. (1995), Better Subset Regression Using the Nonnegative Garrote, *Technometrics* 37, 373–384.

Breiman, L. (1996), Bagging Predictors, *Machine Learning* 24, 123–140.

Breitung, J. and S. Eickmeier (2011), Testing for Structural Breaks in Dynamic Factor Models, *Journal of Econometrics* 163, 71-84.

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.

Chen, L., J.J. Dolado and J. Gonzalo, (2014), Detecting Big Structural Breaks in Large Factor Models, *Journal of Econometrics* 180, 30-48.

Chen, Y.-P., H.-C. Huang and I.-P. Tu (2010), A New Approach for Selecting the Number of Factors, *Computational Statistics and Data Analysis* 54, 2990-2998.

Cheng, X. and B.E. Hansen (2015), Forecasting with Factor Augmented Regression: A Frequentist Model Averaging Approach, *Journal of Econometrics* 186, 280-293.

Cheng, M., N.R. Swanson and X. Yang (2017), Latent Common Return Volatility Factors: Capturing Elusive Predictive Accuracy Gains When Forecasting Volatility, Working Paper, Rutgers University.

Carrasco, M., V. Chernozhukov, S. Goncalves and E. Renault (2015), **High Dimensional Problems in Econometrics**, *Special Issue*, *Journal of Econometrics* 186, 277-476.

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.

Christoffersen, P. (1998), Evaluating Interval Forecasts, *International Economic Review* 39, 841-862.

Clark, T.E. and M.W. McCracken (2001), Tests of Equal Forecast Accuracy and Encompassing for Nested Models, *Journal of Econometrics* 105, 85-110.

Clark, T. E. and M.W. McCracken (2013), Advances in Forecast Evaluation, in G. Elliott and A. Timmermann (eds.), **Handbook of Economic Forecasting**, Volume 2, Elsevier, Amsterdam.

Clements, M.P. and D.F. Hendry (1993), On the Limitations of Comparing Mean Square Forecast Errors, *Journal of Forecasting* 12, 617-637.

Clements, M.P. and D.F. Hendry (1999), On Winning Forecasting Competitions in Economics, *Spanish Economic Review* 1, 123-160.

Clements, M.P. and D.F. Hendry (2001), Forecasting with Difference and Trend Stationary Models, *Econometrics Journal* 4, S1–S19.

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

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.

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

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

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

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 M. Shin (2015), Assessing Point Forecast Accuracy by Stochastic Loss Distance, *Economics Letters* 130, 37-38.

Ding, A. and J. Hwang (1999), Prediction Intervals, Factor Analysis Models, and High-Dimensional Empirical Linear Prediction, *Journal of the American Statistical Association* 94, 446-455.

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

Dufour, J.-M. and D. Stevanovic (2013), Factor Augmented VARMA Models: Identification, Estimation, Forecasting and Impulse Responses, *Journal of Business and Economic Statistics* 31, 491-506.

Efron, B., T. Hastie, L. Johnstone and R. Tibshirani (2004), Least Angle Regression, *Annals of Statistics* 32, 407-499.

Elliott, G. and A. Timmermann (2004), Optimal Forecast Combinations Under General Loss Functions and Forecast Error Distributions, *Journal of Econometrics* 122, 47–79.

Fan, J., Y. Liao, and M. Mincheva (2011), High Dimensional Covariance Matrix Estimation in Approximate Factor Models, *Annals of Statistics* 39, 3320-3356.

Fan, J., Y. Liao, and W. Wang (2016), Projected Principal Components Analysis in Factor Models, *Annals of Statistics* 44, 219-254.

Forni, M., M. Hallin, M. Lippi, and L. Reichlin (2000), The Generalized Dynamic Factor Model: Identification and Estimation, *The Review of Economics and Statistics* 82, 540-552.

Forni, M., M. Hallin, M. Lippi, and L. Reichlin (2005), The Generalized Dynamic Factor Model, One Sided Estimation and Forecasting, *Journal of the American Statistical Association* 100, 830–840.

Forni, M., M. Hallin, M. Lippi, L. Reichlin and P. Zaffaroni (2015), Dynamic Factor Models with Infinite Dimensional Factor Spaces: One Sided Representations, *Journal of the Econometrics* 185, 359-371.

Freund, Y. and R.E. Schapire (1997), A Decision-Theoretic Generalization of On-Line Learning and an Application to Boosting, *Journal of Computer and System Sciences* 55, 119–139.

Ghysels, E., J.B. Hill and K. Motegi (2017), Testing a Large Set of Zero Restrictions in Regression Models, with an Application to Mixed Frequency Granger Causality, Working Paper, University of North Carolina, Chapel Hill.

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.

Gurkaynak, 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.

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.

Hendry D.F. and M.P. Clements (2002), Pooling of Forecasts, *Econometrics Journal* 5, 1-26.

Hendry D.F. and G. Mizon (2005), Forecasting in the Presence of Structural Breaks and Policy Regime Shifts, in D.W.K. Andrews and J.H. Stock (eds.) **Identification and Inference for Econometric Models: Essays in Honour of Thomas Rothemberg**, Cambridge University Press.

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

Hornik, K., M. Stinchcombe, and H. White (1989), Multilayer Feedforward Networks are Universal Approximators, *Neural Networks* 2, 359-366.

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 N.R. Swanson (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.

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.

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

Onatski, A. (2015), Asymptotic Analysis of the Squared Estimation Error in Misspecified Factor Models, *Journal of Econometrics* 186, 388-406.

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.

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.

Spearman, C. (1904), General Intelligence Objectively Determined and Measured, *American Journal of Psychology* 15, 201-293.

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

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.

Stone, J.V. (2004), **Independent Component Analysis**, MIT Press, Boston.

Sullivan, R., A. Timmermann and H. White (1999), Data-snooping, Technical Trading Rule Performance, and the Bootstrap, *Journal of Finance* 54, 1647-1691.

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

Tibshirani, R. (1996), Regression Shrinkage and Selection via the Lasso, *Journal of the Royal Statistical Society, Series B* 58, 267–288.

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

Tong, L., R.-W. Liu, V.C. Soon, and Y.-F. Huan (1991), Indeterminacy and Identifiability of Blind Identification, *IEEE Transactions on Circuits and Systems* 38, 499-509.

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.

Table 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 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 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 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 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 2A.

Table 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 2A.

Table 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 2A.

Table 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 2A.

Table 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 2A.

Table 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 2A.

Table 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 2A.

Table 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 2A.

Table 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 2A.

Table 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^{***}	$\mathbf{0.663}^{***}$	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)	$\mathbf{0.729}^{***}$	0.714^{***}	$\mathbf{0.687}^{***}$	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^{***}	$\mathbf{0.683}^{***}$
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^{***}	$\mathbf{0.675}^{***}$	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 2A.

Table 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 2A.

Table 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 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 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 5. Recessions and expansion are defined according to NBER business cycle dates.

Table 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 8A: 1-Step-Ahead Relative MSFEs of Forecast Combination Models*

	Model	rMSFE				
	Maturity	1 year	2 year	3 years	5 years	10 years
	All	0.922	0.976	0.971	0.999	1.066
	FB	0.906	0.949	0.958	1.008	1.101
1992:3-1999:12	FS	1.003	1.062	1.030	1.020	1.063
'Subsample 1'	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
	All	0.740***	0.866***	0.903**	0.981	0.967
	FB	**0.652*****	**0.812****	**0.864****	0.949	0.964
2000:1-2007:12	FS	0.933	0.991	0.996	1.052	0.999
'Subsample 2'	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
	All	1.227	1.174	1.147	1.239	1.081
	FB	1.654	1.527	1.359	1.313	1.155
2008:1-2016:7	FS	1.135	**1.010**	**1.067**	1.235	**1.013**
'Subsample 3'	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
	All	0.911	0.971	0.984	1.061	1.042
	FB	0.958	1.011	1.011	1.074	1.082
1992:3-2016:7	FS	1.002	1.022	1.025	1.094	1.022
'Subsample 4'	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
	All	0.814	0.991	0.968	0.920	0.996
	FB	1.052	1.287	1.190	0.997	1.063
Recession	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
	All	0.948	0.966	0.987	1.089	1.054
	FB	0.923	0.938*	0.972	1.089	1.087
Normal	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 2A. Forecast combination models are listed in Table 7.

Table 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

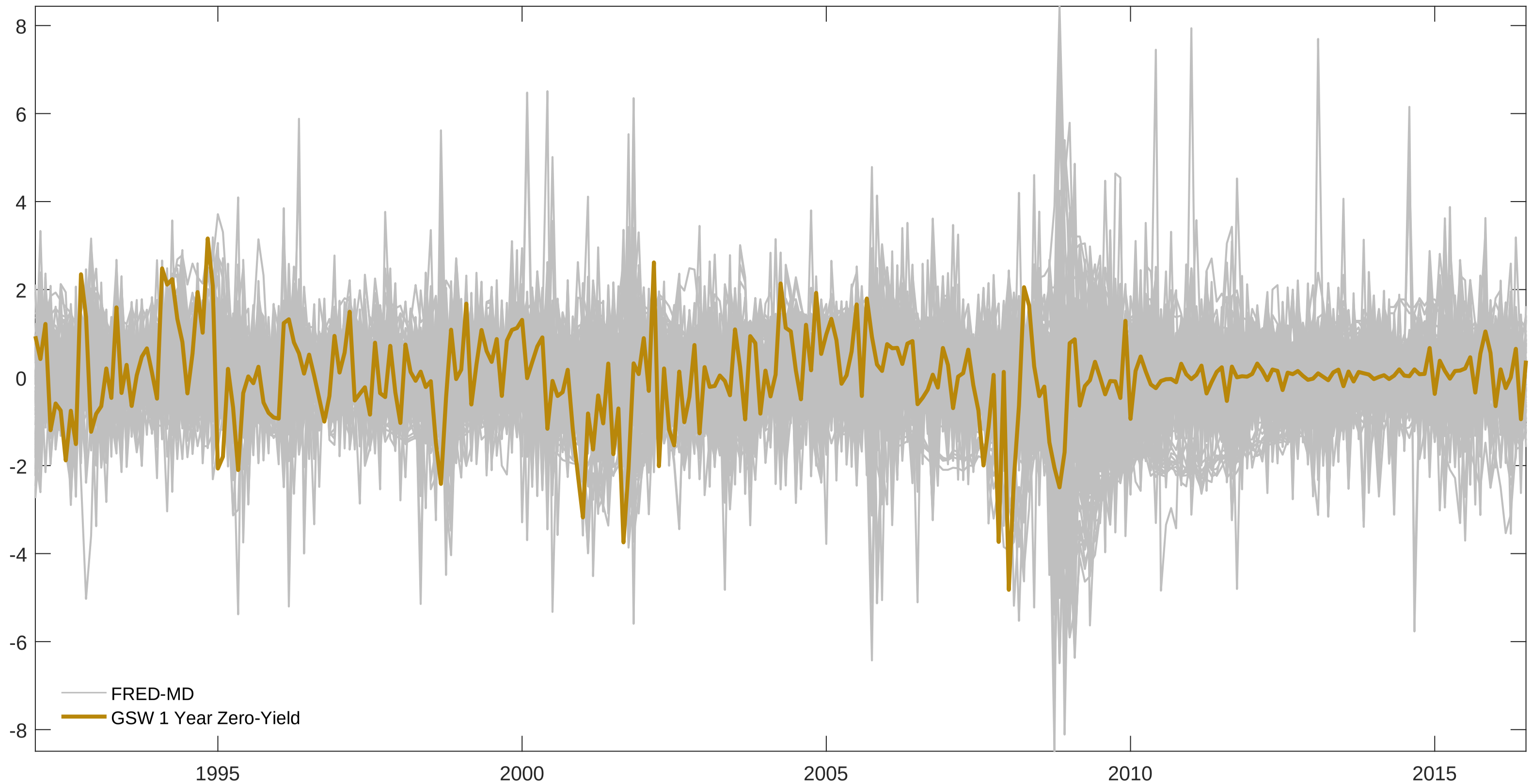
* Notes: See notes to Table 8A.

Table 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

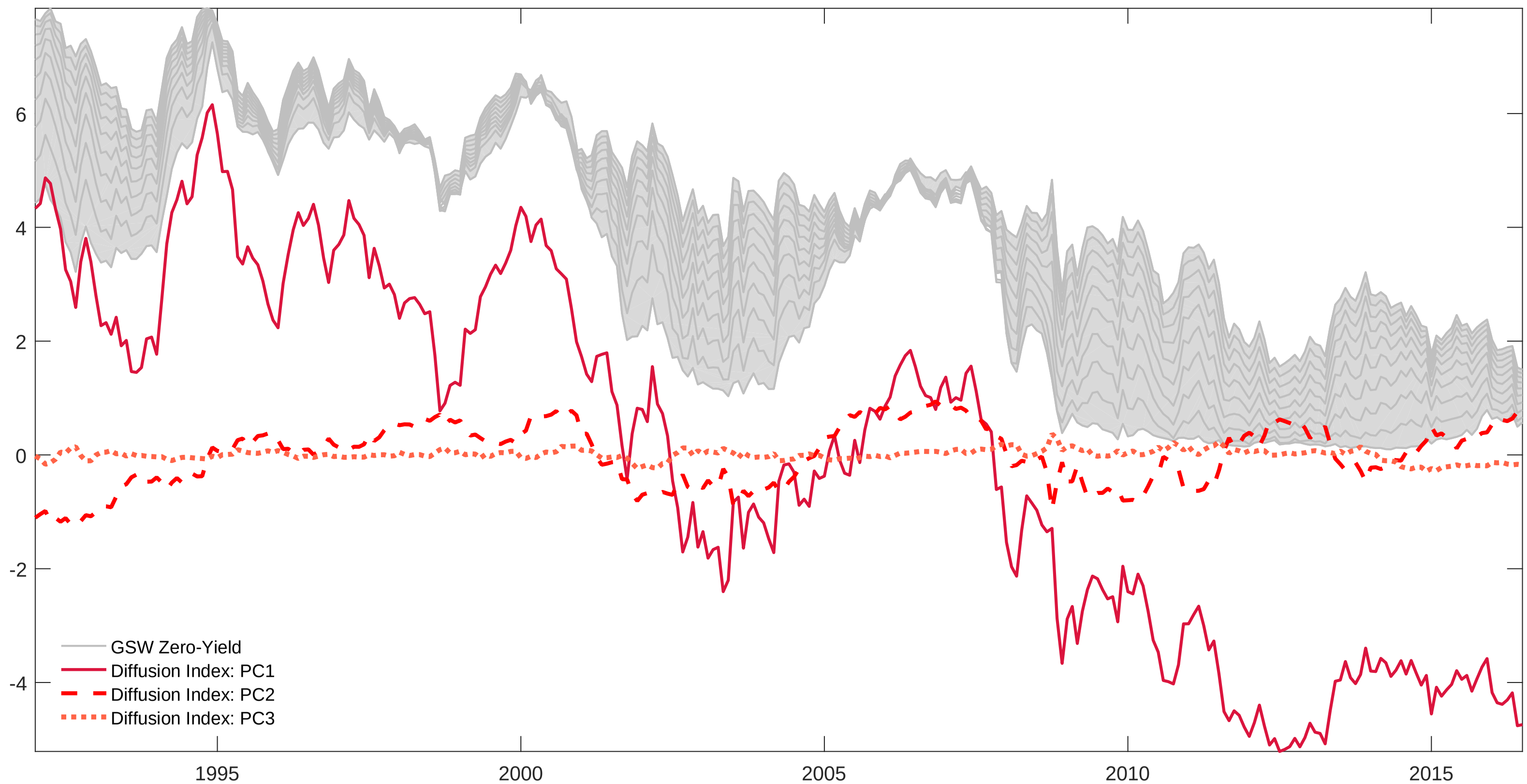
* Notes: See notes to Table 8A.

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



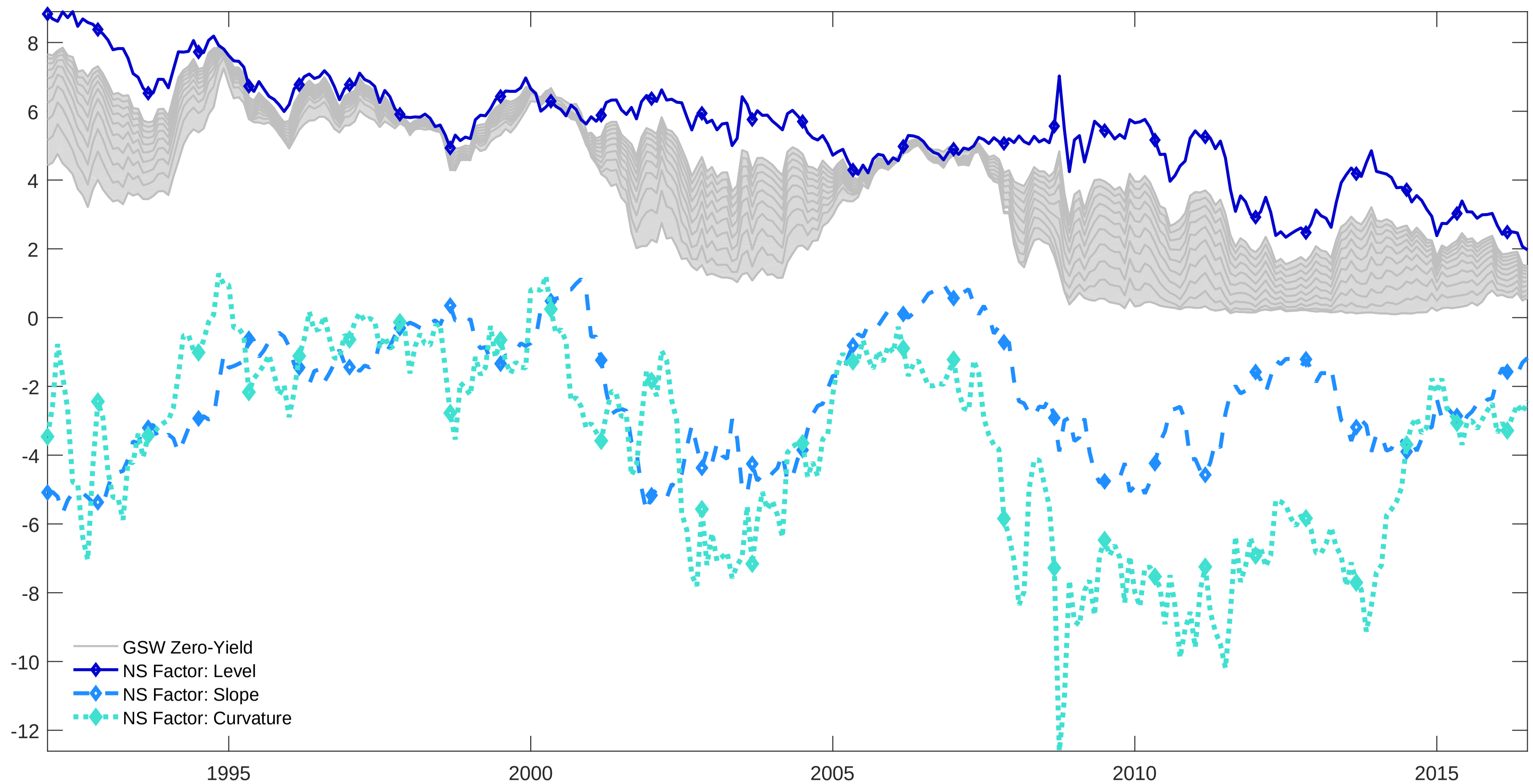
(*) Notes: The Figure displays all macroeconomic variables from the FRED-MD dataset and the 1 year zero-yield from the GSW dataset. All series transformed to ensure stationary and standardized to zero mean and unit variance.

Figure 2: 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.

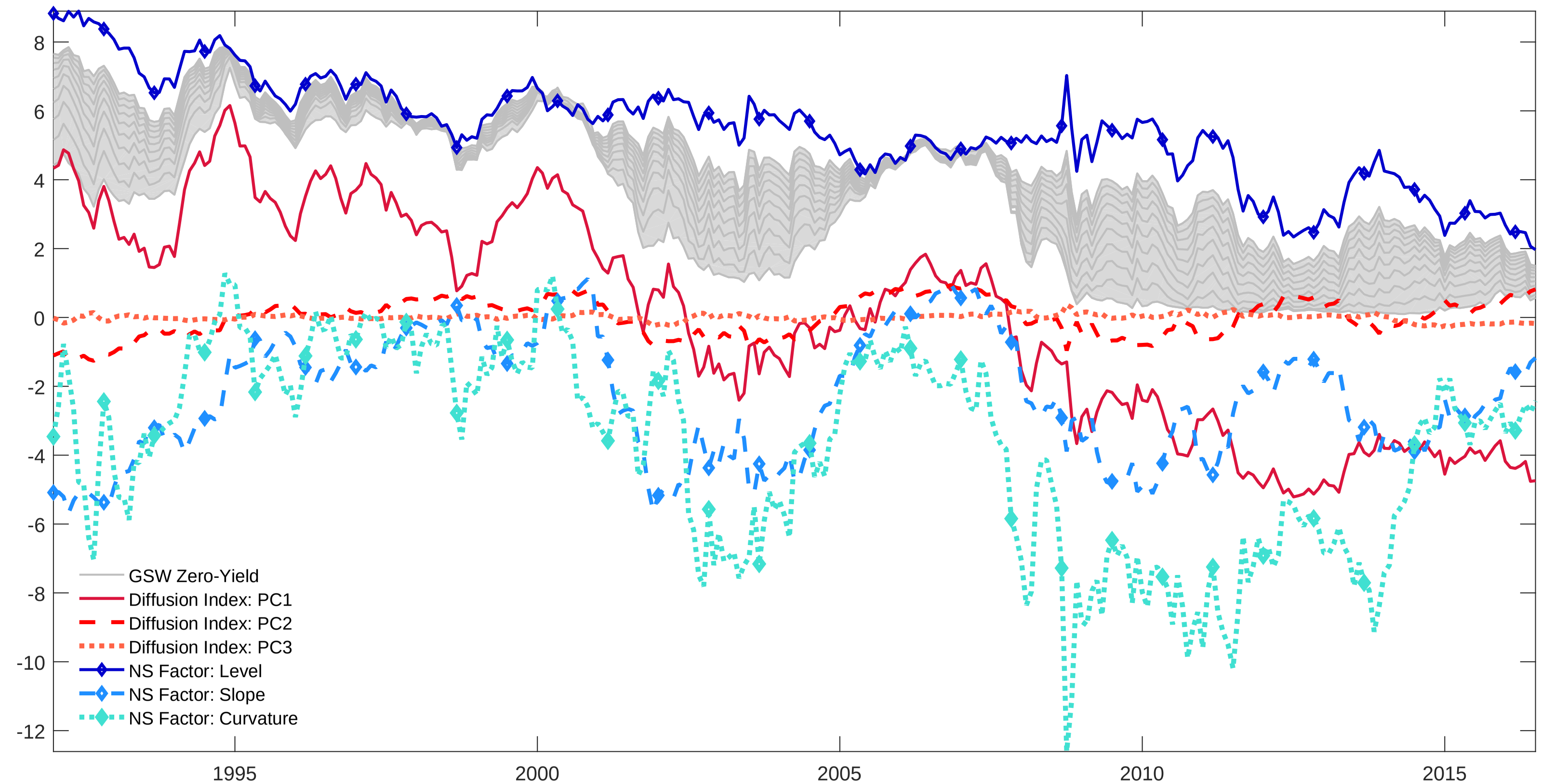
Figure 3: Yields and Dynamic Nelson Siegel Factors for Sample Period 1992:1 - 2016:7*



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

g

Figure 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).