Forecasting Financial and Macroeconomic Variables Using Data Reduction Methods: New Empirical Evidence*

Hyun Hak Kim and Norman R. Swanson
Rutgers University

October 2010 (this version, November 2011)

Abstract

In this paper, we empirically assess the predictive accuracy of a large group of models based on the use of principle components and other shrinkage methods, including Bayesian model averaging and various bagging, boosting, least angle regression and related methods Our results suggest that model averaging does not dominate other well designed prediction model specification methods, and that using a combination of factor and other shrinkage methods often yields superior predictions. For example, when using recursive estimation windows, which dominate other "windowing" approaches in our experiments, prediction models constructed using pure principal component type models combined with shrinkage methods yield mean square forecast error "best" models around 70% of the time, when used to predict 11 key macroeconomic indicators at various forecast horizons. Baseline linear models (which "win" around 5% of the time) and model averaging methods (which win around 25% of the time) fare substantially worse than our sophisticated nonlinear models. Ancillary findings based on our forecasting experiments underscore the advantages of using recursive estimation strategies, and provide new evidence of the usefulness of yield and yield-spread variables in nonlinear prediction specification.

Keywords: prediction, bagging, boosting, Bayesian model averaging, ridge regression, least angle regression, elastic net and non-negative garotte.
JEL Classification: G1.

---

* Hyun Hak Kim (hykim@econ.rutgers.edu) and Norman R. Swanson (nswanson@econ.rutgers.edu), Department of Economics, Rutgers University, 75 Hamilton Street, New Brunswick, NJ 08901, USA. This paper was prepared for the Sir Clive W.J. Granger Memorial Conference held at Nottingham University in May 2010. We are very grateful to the organizers, Rob Taylor and David Hendry, for hosting this special event. The paper has been modified substantially due to numerous useful comments made on earlier versions by the editor, Graham Elliott and by an anonymous referee. In additional to thanking the editor and referee, the authors wish to thank to seminar participants at the conference, as well as at the Bank of Canada and Rutgers University, for useful comments. Additional thanks are owed to Nii Armah, Valentina Corradi, David Hendry, Gary Koop, John Landon-Lane, Fuchun Li, Greg Tkacz, Hiroki Tsurumi, and Hal White for numerous useful suggestions on an earlier version of this paper. The authors gratefully acknowledge financial support from a Rutgers University Research Council grant.

1 Introduction

Technological advances over the last five decades have led to impressive gains in not only computational power, but also in the quantity of available financial and macroeconomic data. Indeed, there has been something of a race going on in recent years, as technology, both computational and theoretical, has been hard pressed to keep up with the ever increasing mountain of data available for empirical use. From a computational perspective, this has helped spur the development of data shrinkage techniques, for example. In economics, one of the most widely applied of these is diffusion index methodology. Diffusion index techniques offer a simple and sensible approach for extracting common factors that underlie the dynamic evolution of large numbers of variables. To be more specific, let Y be a time series vector of dimension $(T \times 1)$ and let X be a time-series predictor matrix of dimension $(T \times N)$, and define the following factor model, where F_t denotes a $1 \times r$ vector of unobserved common factors that can be extracted from X_t. Namely, let $X_t = F_t\Lambda' + e_t$, where e_t is an $1 \times N$ vector of disturbances and Λ is an $N \times r$ coefficient matrix. Using common factors extracted from the above model, Stock and Watson (2002a,b) as well as Bai and Ng (2006a) examine linear autoregressive type forecasting models augmented by the inclusion of common factors.

In this paper, use the forecasting models of Stock and Watson (2002a,b) as a starting point in an analysis of diffusion index and other shrinkage methods. In particular, we first estimate the unobserved factors, F_t, and then forecast Y_{t+h} using observed variables and $\hat{F}_t$, where $\hat{F}_t$ is an estimator of F_t. However, even though factor models are now widely used, many issues remain outstanding, such as the determination of the number of factors to be used in subsequent prediction model specification (see e.g. Bai and Ng (2002, 2006b, 2008)). In light of this, and in order to add functional flexibility, we additionally implement prediction models where the numbers and functions of factors to be used is subsequently selected using a variety of additional shrinkage methods. Various other related methods, including targeted regressor selection based on shrinkage, are also implemented. In this sense, we add to the recent work of Stock and Watson (2005a) as well as Bai and Ng (2008, 2009), who survey several methods for shrinkage that are based on factor augmented autoregression models. Shrinkage methods considered in this paper include bagging, boosting, Bayesian model averaging, simple model averaging, ridge regression, least angle regression, elastic net and the non-negative garotte. We also evaluate various linear models, and hence add to the recent work of Pesaran et al. (2011), who carry out a broad examination of factor-augmented vector autoregression models.

In summary, the purpose of this paper is to empirically assess the predictive accuracy of various linear models; pure principal component type models; principal components models constructed using subsets of variables selected based on the elastic net and other shrinkage techniques; principle components models where the factors to be used in prediction are directly selected using shrinkage methods such as ridge regression and bagging; models constructed by directly applying shrinkage methods (other than principle components) to the data; and a number of model averaging methods. The horse-race that we carry out using all of the above approaches allows us to provide new evidence on the usefulness of factors in general as well as on various related issues such as whether model averaging still "wins" rather ubiquitously.

The variables that we predict include a variety of macroeconomic variables that are useful for evaluating the state of the economy. More specifically, forecasts are constructed for eleven series, including: the unemployment rate, personal income less transfer payments, the 10 year Treasury-bond yield, the consumer price index, the producer price index, nonfarm payroll employment, housing starts, industrial production, M2, the S&P 500 index, and gross domestic product. These variables constitute 11 of the 14 variables (for which long data samples are available) that the Federal Reserve takes into account, when formulating the nation's monetary policy. In particular, as noted in Armah and Swanson (2011) and on the Federal Reserve Bank of New York's website: *"In formulating the nation's monetary policy, the Federal Reserve considers a number of factors, including the economic and financial indicators which follow, as well as the anecdotal reports compiled in the Beige Book. Real Gross Domestic Product (GDP); Consumer Price Index (CPI); Nonfarm Payroll Employment Housing Starts; Industrial Production/Capacity Utilization; Retail Sales; Business Sales and Inventories; Advance Durable Goods Shipments, New Orders and Unfilled Orders; Lightweight Vehicle Sales; Yield on 10-year Treasury Bond; S&P 500 Stock Index; M2."*

Our finding can be summarized as follows. First, as might be expected, for a number of our target variables, we find that various sophisticated models, such as component-wise boosting, have lower mean square forecast errors (MSFEs) than benchmark linear autoregressive forecasting models constructed using only observable variables, hence suggesting that models that incorporate common factors constructed using diffusion index methodology offer a convenient way to filter the information contained in large-scale economic datasets. More specifically, models constructed using pure principal component type models combined with shrinkage methods yield MSFE-"best" models around 70% of the time, across multiple forecast horizons, and for various prediction periods. Moreover, a small subset of combined factor/shrinkage type models "win" approximately 50% of the time, including c-boosting, ridge regression, least angle regression, elastic net and the non-negative garotte, with c-boosting the clear overall "winner". Baseline linear models (which "win" around 5% of the time) and model averaging methods (which "win" around 25% of the time) fare substantially worse than our sophisticated nonlinear models. Ancillary findings based on our forecasting experiments underscore the advantages of using recursive estimation windowing strategies[1], and provide new evidence of the usefulness of yield and yield-spread variables in nonlinear prediction specification.

Although we leave many important issues to future research, such as the prevalence of structural breaks other than level shifts, and the use of even more general nonlinear methods for describing the data series that we examine, we believe that results presented in this paper add not only to the diffusion index literature, but also to the extraordinary collection of papers on forecasting that Clive W.J. Granger wrote during his decades long research career. Indeed, as we and others have said many times, we believe that Clive W.J. Granger is in many respects the father of time series forecasting, and we salute his innumerable contributions in areas from predictive accuracy testing, model selection analysis, and forecast combination, to forecast loss function analysis, forecasting using nonstationary data, and nonlinear forecasting model specification.

The rest of the paper is organized as follows. In the next section we provide a brief survey of factor models. In Section 3, we survey the robust shrinkage estimation methods used in our prediction experiments. Data, forecasting methods, and benchmark forecasting models are discussed in Section 4, and empirical results are presented in Section 5. Concluding remarks are given in Section 6.

2 Diffusion Index Models

Recent forecasting studies using large-scale datasets and pseudo out-of-sample forecasting include: Artis et al. (2002), Boivin and Ng (2005, 2006), Forni et al. (2005), and Stock and Watson (1999, 2002, 2005a,b, 2006). Stock and Watson (2006) discuss in some detail the literature on the use of diffusion indices for forecasting. In the following brief discussion of diffusion index methodology, we follow Stock and Watson (2002).

Let X_{tj} be the observed datum for the $j-$th cross-sectional unit at time t, for $t = 1, ..., T$

[1] For further discussion of estimation windows and the related issue of structural breaks, see Pesaran et al. (2011).

and $j = 1, ..., N$. We begin with the following model:

$$X_{tj} = F_t \Lambda_j' + e_{tj}, \tag{1}$$

where F_t is a $1 \times r$ vector of common factors, Λ_j is an $1 \times r$ vector of factor loadings associated with F_t, and e_{tj} is the idiosyncratic component of X_{tj}. The product $F_t \Lambda_j'$ is called the common component of X_{tj}. This is a useful dimension reducing factor representation of the data, particularly when $r << N$, as is usually assumed to be the case in the empirical literature. Following Bai and Ng (2002), the whole panel of data $X = (X_1, ..., X_N)$ can be represented as (1). Connor and Korajczyk (1986, 1988, 1993) note that the factors can be consistently estimated by principal components as $N \to \infty$, even if e_{tj} is weakly cross-sectionally correlated. Similarly, Forni et al. (2005) and Stock and Watson (2002) discuss consistent estimation of the factors when $N, T \to \infty$. We work with high-dimensional factor models that allow both N and T to tend to infinity, and in which e_{tj} may be serially and cross-sectionally correlated, so that the covariance matrix of $e_t = (e_{t1}, ..., e_{tN})$ does not have to be a diagonal matrix. We will also assume $\{F_t\}$ and $\{e_{tj}\}$ are two groups of mutually independent stochastic variables. Furthermore, it is well known that if $\Lambda = (\Lambda_1, ..., \Lambda_N)'$ for $F_t \Lambda' = F_t Q Q^{-1} \Lambda'$, a normalization is needed in order to uniquely define the factors, where Q is a nonsingular matrix. Assuming that $(\Lambda' \Lambda / N) \to I_r$, we restrict Q to be orthonormal. This assumption, together with others noted in Stock and Watson (2002) and Bai and Ng (2002), enables us to identify the factors up to a change of sign and consistently estimate them up to an orthonormal transformation.

With regard to choice of r, note that Bai and Ng (2002) provide one solution to the problem of choosing the number of factors. They establish convergence rates for factor estimates under consistent estimation of the number of factors, r, and propose panel criterion to consistently estimate the number of factors. Bai and Ng (2002) define selection criteria of the form $PC(r) = V\left(r, \hat{F}\right) + rh(N, T)$, where $h(\cdot)$ is a penalty function. In this paper, the following version is used (for discussion, see Bai and Ng (2002) and Armah and Swanson (2010)):

$$SIC(r) = V\left(r, \hat{F}\right) + r\hat{\sigma}^2 \left(\frac{(N + T - r)\ln(NT)}{NT} \right). \tag{2}$$

A consistent estimate of the true number of factors is $\hat{r} = \arg \min_{0 \leq r \leq r_{\max}} SIC(r)$. In a number of our models, we use this criteria for choosing the number of factors. However, as discussed above, we also use a variety of shrinkage methods to specify numbers and functions of factors to be used alternative prediction models. These shrinkage models, including bagging and other methods outlined in the introduction are also directly applied to our panel of data, without constructing factors.

The basic structure of the forecasting models examined in this paper is the same as that examined in Artis et al. (2002), Bai and Ng (2002, 2006a,b, 2008, 2009), Boivin and Ng (2005) and Stock and Watson (2002, 2005a,b, 2006). In particular, we consider models of the following generic form:

$$Y_{t+h} = W_t \beta_W + F_t \beta_F + \varepsilon_{t+h}, \tag{3}$$

where h is the forecast horizon, Y_t is the scalar valued "target" variable to be forecasted, W_t is a $1 \times s$ vector of observable variables, including lags of Y_t, ε_t is a disturbance term, and the β's are parameters estimated using least squares. In a predictive context, Ding and Hwang (1999) analyze the properties of forecasts constructed from principal components when N and T are large. They perform their analysis under the assumption that the error processes $\{e_{tj}, \varepsilon_{t+h}\}$ are cross-sectionally and serially *iid*. Forecasts of Y_{t+h} based on (3) involve a two step procedure because both the regressors and coefficients in the forecasting equations are unknown. The data X_t are first used to estimate the factors, $\hat{F}_t$, by means of principal components. With the estimated factors in hand, we obtain the estimators $\hat{\beta}_F$ and $\hat{\beta}_W$ by regressing Y_{t+h} on $\hat{F}_t$ and W_t. Of note is that if $\sqrt{T}/N \to 0$, then the generated regressor problem does not arise, in the sense that least squares estimates of $\hat{\beta}_F$ and $\hat{\beta}_W$ are $\sqrt{T}$ consistent and asymptotically normal (see Bai and Ng (2008)). In this paper, we try different methods for estimating $\hat{\beta}_F$ and then compare the predictive accuracy of the resultant forecasting models.[2].

3 Robust Estimation Techniques

We consider a variety of "robust" estimation techniques including statistical learning algorithms (bagging and boosting), as well as various penalized regression methods including ridge regression, least angle regression, elastic net, and the non-negative garotte. We also consider forecast combination in the form of Bayesian model averaging.

The following sub-sections provide summary details on implementation of the above methods in contexts where in a first step we estimate factors using the principal components

[2] We refer the reader to Stock and Watson (1999, 2002, 2005a,b) and Bai and Ng (2002, 2008, 2009) for a detailed explanation of this procedure, and to Connor and Korajczyk (1986, 1988, 1993), Forni et al. (2005) and Armah and Swanson (2010) for further detailed discussion of generic diffusion models.

analysis, while in a second step we select factor weights using shrinkage. Approaches in which we first directly implement shrinkage to select an "informative" set of variables for: (i) direct use in prediction model construction; or (ii) use in a second step where factors are constructed for subsequent use in prediction model construction, follow immediately. Note that all variables are assumed to be standardized in the sequel. Algorithms for the methods outlined below are given in the originating papers cited as well as discussed in some detail in Kim and Swanson (2011).

3.1 Statistical Learning (Bagging and Boosting)

3.1.1 Bagging

Bagging, which is a short for "bootstrap aggregation", was introduced by Breiman (1996). Bagging involves first drawing bootstrap samples from in-sample "training" data, and then constructing predictions, which are later combined. Thus, if a bootstrap sample based predictor is defined as $\hat{Y}_b^* = \hat{\beta}_b^* X_b^*$, where $b = 1, ..., B$ denotes the b-th bootstrap sample drawn from the original dataset, then the bagging predictor is $\hat{Y}^{Bagging} = \frac{1}{B}\sum_{b=1}^{B} \hat{Y}_b^*$. In this paper, we follow Bühlmann and Yu (2002) and Stock and Watson (2005a) who note that that, asymptotically, the bagging estimator can be represented in shrinkage form. Namely:

$$\hat{Y}_{t+h}^{Bagging} = W_t \hat{\beta}_W + \sum_{j=1}^{r} \psi\left(t_j\right) \hat{\beta}_{Fj} \hat{F}_{t,j} \tag{4}$$

where $\hat{Y}_{t+h}^{Bagging}$ is the forecast of Y_{t+h} made using data through time t, and $\hat{\beta}_W$ is the least squares (LS) estimator from a regression of Y_{t+h} on W_t, where W_t is a vector of lags of Y_t as in (3) including a vector of ones, $\hat{\beta}_{Fj}$ is a LS estimator from a regression of residuals, $Z_t = Y_{t+h} - W_t \hat{\beta}_W$ on $\hat{F}_{T-h,j}$, and t_j is the t-statistic associated with $\hat{\beta}_{Fj}$, defined as $\sqrt{T}\hat{\beta}_{Fj}/s_e$, where s_e, is a Newey-West standard error, and ψ is a function specific to the forecasting method. In the current context we set:

$$\psi\left(t\right) = 1 - \Phi\left(t + c\right) + \Phi\left(t - c\right) + t^{-1}[\phi\left(t - c\right) - \phi\left(t + c\right)], \tag{5}$$

where c is the pretest critical value, ϕ is the standard normal density and Φ is the standard normal CDF. In this paper, we follow Stock and Watson (2005a), and set the pretest critical value for bagging, c to be 1.96.

3.1.2 Boosting

Boosting (see Freund and Schapire (1997)) is a procedure that builds on a user-determined set of functions (e.g. least square estimators), often called "learners" and uses the set repeatedly on filtered data which are typically outputs from previous iterations of the learning algorithm. The output of a boosting algorithm generally takes the form:

$$\hat{Y}^M = \sum_{m=1}^{M} \kappa_m f\left(X; \beta_m\right),$$

where the κ_m can be interpreted as weights, and $f\left(X; \beta_m\right)$ are function of the panel dataset, X. Friedman (2001) introduce "L_2Boosting", which takes the simple approach of refitting "base learners" to residuals from previous iterations.[3] Bühlmann and Yu (2003) a boosting algorithm fitting "learners" using one predictor at a, in contexts where a large numbers of predictors are available, in the context of *iid* data. Bai and Ng (2009) modify this algorithm to handle time-series. We use their "Component-Wise L_2Boosting" algorithm in the sequel, with least squares "learners".

As an example, consider the case where boosting is done on the original W_t data as well as factors, $\hat{F}_t$, constructed using principal components analysis; and denote the output of the boosting algorithm as $\hat{\mu}^M\left(\hat{F}_t\right)$. Then, predictions are constructed using the following model:

$$\hat{Y}_{t+h}^{Boosting} = W_t\hat{\beta}_W + \hat{\mu}^M\left(\hat{F}_t\right). \tag{6}$$

Evidently, when shrinkage is done directly on X_t, then $\hat{F}_t$ in the above expression is suitably replaced with X_t.

3.2 Penalized Regression (Least Angle Regression, Elastic Net, and Non-Negative Garotte)

Ridge regression, which was introduced by Hoerl and Kennard (1970), is likely the most well known penalized regression method (see Kim and Swanson (2011)) for further discussion. Recent advances in penalized regression have centered to some extent on the penalty function. Ridge regression is characterized by an L_2 penalty function. More recently, there has been much research examining the properties of L_1 penalty functions, using the so called Lasso (least absolute shrinkage and selection operator) regression method, as introduced by Tibshirani (1996), and various hybrids and generalizations thereof. Examples of these in-

[3] Other extensions of the original boosting problem discussed by Friedman (2001) are given in Ridgeway et al. (1999) and Shrestha and Solomatine (2006).

clude least angle regression , the elastic net, and the non-negative garotte, all of which are implemented in our prediction experiments.

3.2.1 Least Angle Regression (LAR)

Least Angle Regression (LAR), as introduced by Efron et al. (2004), is based on a model-selection approach known as forward stage-wise regression, which has been extensively used to examine cross-sectional data (for further details, see Efron et al. (2004) and Bai and Ng (2008)). Gelper and Croux (2008) extend Bai and Ng (2008) to time series forecasting with many predictors. We implement the algorithm of Gelper and Croux (2008) when constructing the LAR estimator.

Like many other stagewise regression approaches, start with $\hat{\mu}^0 = \bar{Y}$, the mean of the target variable, use the residuals after fitting W_t to the target variable, and construct a first estimate, $\hat{\mu} = X_t\hat{\beta}$, in stepwise fashion, using standardized data, and in M iterations, say. Possible explanatory variables are incrementally examined, and their added to the estimator function, $\widehat{\mu}$, according to their explanatory power. Following the same notation as used above, in the case where shrinkage is done solely on common factors, the objective is to construct predictions,

$$\hat{Y}_{t+h}^{LAR} = W_t\hat{\beta}_W + \hat{\mu}^M(\hat{F}_t).$$

3.2.2 Elastic Net (EN)

Zou and Hastie (2005) point out that the lasso has undesirable properties when T is greater than N or when there is a group of variables amongst which all pairwise correlations are very high. They develop a new regularization method that they claim remedies the above problems. The so-called elastic net (EN) simultaneously carries out automatic variable selection and continuous shrinkage. Its name comes from the notion that it is similar in structure to a stretchable fishing net that retains "all the big fish". Zou and Hastie (2005) In this paper, we use the algorithm of Bai and Ng (2008), who modify the naive EN to use time series rather than cross sectional data. To fix ideas, assume again that we are interested in X and Y, and that variables are standardized. For any fixed non-negative η_1 and η_2, the elastic net criterion is defined as:

$$L(\eta_1, \eta_2, \beta) = |Y - X\beta|^2 + \eta_2 |\beta|^2 + \eta_1 |\beta|_1 , \tag{7}$$

where $|\beta|^2 = \sum_j^N (\beta_j)^2$ and $|\beta|_1 = \sum_j^N |\beta_j|$. The solution to this problem is the so-called naive elastic net, given as:

$$\hat{\beta}^{NEN} = \frac{\left(\left|\hat{\beta}^{LS}\right| - \eta_1/2\right)_{pos}}{1+\eta_2} sign\left\{\hat{\beta}^{LS}\right\}. \tag{8}$$

where $\hat{\beta}^{LS}$ is the least square estimator of β and $sign\left(\cdot\right)$ equals ± 1. Here, "*pos*" denotes the positive part of the term in parentheses. Zou and Hastie (2005), in the context of above naive elastic net, point out that there is double shrinkage in this criterion, which does not help to reduce the variance and may lead to additional bias so that they propose a version of the elastic net in which this double shrinkage is corrected. In this context, the elastic net estimator, $\hat{\beta}^{EN}$, is defined as:

$$\hat{\beta}^{EN} = \left(1+\eta_2\right)\hat{\beta}^{NEN}, \tag{9}$$

where η_2 is a constant, usually "optimized" via cross validation methods. Zou and Hastie (2005) propose an algorithm called "LAR-EN" to estimate $\hat{\beta}^{EN}$ using the LAR algorithm implemented in this paper.[4] In the current context, $\hat{\beta}^{EN}$ is either the coefficient vector associated with the $\hat{F}_t$ in a forecasting model of the variety given in (3), assuming that $\psi\left(\cdot\right) = 1$, or is a coefficient vector associated directly with the panel dataset, X_t.

3.2.3 NON-NEGATIVE GAROTTE (NNG)

The non-negative garotte (NNG), was introduced by Breiman (1995). This method is a scaled version of the least square estimator with shrinkage factors, and is closely related to the EN and LAR. Yuan and Lin (2007) develop an efficient garrotte algorithm and prove consistency in variable selection. As far as we know, this method has previously not been used in the econometrics literature. We follow Yuan and Lin (2007) and apply it to time series forecasting. As usual, we begin by considering standardized X and Y. Assume that the following shrinkage factor is given: $q\left(\zeta\right) = \left(q_1\left(\zeta\right), q_2\left(\zeta\right), ..., q_N\left(\zeta\right)\right)'$, where $\zeta > 0$ is a tuning parameter. The objective is to choose the shrinkage factor in order to minimize:

$$\frac{1}{2}\left\|Y - Gq\right\|^2 + T\zeta\sum_{j=1}^{N} q_j, \qquad \text{subject to } q_j > 0,\ j = 1, .., N, \tag{10}$$

where $G = (G_1, .., G_N)'$, $G_j = X_j\widehat{\beta}_j^{LS}$, and $\widehat{\beta}^{LS}$ is the least squares estimator. The NNG estimator of the regression coefficient vector is defined as $\hat{\beta}_j^{NNG} = q_j\left(\zeta\right)\hat{\beta}_j^{LS}$, and the estimate of Y is defined as $\widehat{\mu} = X\hat{\beta}^{NNG}\left(\zeta\right)$, so that predictions can be formed in a manner that is analogous to that discussed in the previous subsections. Assuming, for example, that $X'X =$

[4] We use their algorithm, which is discussed in more detail in Kim and Swanson (2011).

I, the minimizer of expression (10) has the following explicit form: $q_j(\zeta) = \left(1 - \frac{\zeta}{(\hat{\beta}_j^{LS})^2}\right)_+$, $j = 1, ..., N$. This ensures that the shrinking factor may be identically zero for redundant predictors. The disadvantage of the NNG is its dependence on the ordinary least squares estimator, which can be especially problematic in small samples. However, Zou (2006) shows that the NNG with ordinary least squares is also consistent, if N is fixed, as $T \to \infty$. Our approach is to start the algorithm with the least squares estimator, as in Yuan (2007), who outline a simple algorithm for the non-negative garotte that we use in the sequel.

3.3 Bayesian Model Averaging

In recent years, Bayesian Model Averaging (BMA) has been applied to many forecasting problems, and has been frequently shown to yield improved predictive accuracy, relative to approaches based on the use of individual models. For this reason, we include BMA in our prediction experiments; and we view it as one of our benchmark modeling approaches. For further discussion of BMA in a forecasting context, see Koop and Potter (2004), Wright (2008, 2009), and Kim and Swanson (2011)

In addition, for a concise discussion of general BMA methodology, see Hoeting et al. (1999) and Chipman et al. (2001). The basic idea of BMA starts with supposing interest focuses on Q possible models, denoted by $M_1, ..., M_Q$, say. In forecasting contexts, BMA involves averaging target predictions, Y_{t+h} from the candidate models, with weights appropriately chosen. In a very real sense, thus, it resembles bagging. The key difference is that BMA puts little weight on implausible models, as opposed to other varieties of shrinkage discussed above that operate directly on regressors. The algorithm that we use for implementation of BMA follows closely Chipman et al. (2001), Fernandez et al. (2001), and Koop and Potter (2004). For complete details, the reader is referred to Kim and Swanson (2011).

4 Data

Following a long tradition in the diffusion index literature, we examine monthly data observations on 144 U.S. macroeconomic time series for the period 1960:01 - 2009:5 ($N = 144, T = 593$)[5]. Forecasts are constructed for eleven variables, including: the unemployment rate, personal income less transfer payments, the 10 year Treasury-bond yield, the consumer price index, the producer price index, non-farm payroll employment, housing starts, industrial pro-

[5]This is an updated and expanded version of the Stock and Watson (2005a,b) dataset.

duction, M2, the S&P 500 index, and gross domestic product.[6]. These variables constitute 11 of the 14 variables (for which long data samples are available) that the Federal Reserve takes into account, when formulating the nation's monetary policy, as noted in Armah and Swanson (2011), Kim and Swanson (2011), and on the Federal Reserve Bank of New York's website. Table 1 lists the eleven variables. The third row of the table gives the transformation of the variable used in order to induce stationarity. In general, logarithms were taken for all nonnegative series that were not already in rates (see Stock and Watson (2002, 2005a) for complete details). Note that a full list of predictor variables is provided in the appendix to an earlier working paper version which is available upon request from the authors.

5 Forecasting Methodology

Using the transformed dataset, denoted above by X, factors are estimated by the method of principal component analysis discussed in Section 2. In Kim and Swanson (2011), factors are additionally estimated using independent component analysis and sparse principal component analysis. After estimating factors, the alternative methods outlined in the previous sections are used to form forecasting models and predictions. In particular, we consider three specification types when constructing shrinkage based prediction models: *Specification Type 1*: Principal components are first constructed, and then prediction models are formed using the shrinkage methods of Section 3 to select functions of and weights for the factors to be used in our prediction models of the type given in (3). *Specification Type 2:* Principal component models of the type given in (3) are constructed using subsets of variables from the largescale dataset that are first selected via application of the shrinkage methods of Section 3. This is different from the above approach of estimating factors using all of the variables. *Specification Type 3:* Prediction models are constructed using only the shrinkage methods discussed in Section 3, without use of factor analysis at any stage.

In our prediction experiments, pseudo out-of-sample forecasts are calculated for each variable and method, for prediction horizons $h = 1, 3$, and 12. All estimation, including lag selection, shrinkage, and factor construction is done anew, at each point in time, prior to the construction of each new prediction, using both recursive and rolling estimation windows. Note that at each estimation period, the number of factors included will be different, following the testing approach discussed in Section 2. Note also that lags of the target predictor

[6]Note that gross domestic product is reported quaterly. We interpolate these data to a monthly frequency following Chow and Lin (1971),

variables are also included in the set of explanatory variables, in all cases. Selection of the number of lags to include is done using the SIC. Out-of-sample forecasts begin after 13 years (e.g. the initial in-sample estimation period is R =156 observations, and the out-of-sample period consists of $P = T - R = 593 - 156 = 437$ observations, for $h = 1$). Moreover, the initial in-sample estimation period is adjusted so that the ex ante prediction sample length, P, remains fixed, regardless of the forecast horizon. For example, when forecasting the unemployment rate, when $h = 1$, the first forecast will be $\hat{Y}_{157}^{h=1} = \hat{\beta}_W W_{156} + \hat{\beta}_F \tilde{F}_{156}$, while in the case where $h = 12$, the first forecast will be $\hat{Y}_{157}^{h=12} = \hat{\beta}_W W_{145} + \hat{\beta}_F \tilde{F}_{145}$ In our rolling estimation scheme, the in-sample estimation period used to calibrate our prediction models is fixed at length 12 years. The recursive estimation scheme begins with the same in-sample period of 12 years (when $h = 12$), but a new observation is added to this sample prior to the re-estimation and construction of each new forecast, as we iterate through the ex-ante prediction period. Note, thus, that the actual observations being predicted as well as the number of predictions in our ex-ante prediction period remains fixed, regardless of forecast horizon, in order to facilitate comparison across forecast horizons as well as models.

Forecast performance is evaluated using mean square forecast error (MSFE), defined as:

$$MSFE_{i,h} = \sum_{t=R-h+2}^{T-h+1} \left(Y_{t+h} - \hat{Y}_{i,t+h} \right)^2, \tag{11}$$

where $\widehat{Y}_{i,t+h}$ is the forecast for horizon h for the $i-$th model. Forecast accuracy is evaluated using point MSFEs as well as the predictive accuracy test of Diebold and Mariano (DM: 1995), which is implemented using quadratic loss, and which has a null hypothesis that the two models being compared have equal predictive accuracy. DM test statistics have asymptotic $N(0, 1)$ limiting distributions, under the assumption that parameter estimation error vanishes as $T, P, R \to \infty$, and assuming that each pair of models being compared is nonnested. Namely, the null hypothesis of the test is $H_0 : E\left[l\left(\varepsilon_{t+h|t}^1\right)\right] - E\left[l\left(\varepsilon_{t+h|t}^2\right)\right] = 0$,where $\varepsilon_{t+h|t}^i$ is $i-$th model's prediction error and $l(\cdot)$ is the quadratic loss function. The actual statistic in this case is constructed as: $DM = P^{-1} \sum_{i=1}^{P} d_t / \hat{\sigma}_{\overline{d}}$, where $d_t = \left(\widehat{\varepsilon_{t+h|t}^1}\right)^2 - \left(\widehat{\varepsilon_{t+h|t}^2}\right)^2$, $\overline{d}$ is the mean of d_t, $\hat{\sigma}_{\overline{d}}$ is a heteroskedasticity and autocorrelation robust estimator of the standard deviation of $\overline{d}$, and $\widehat{\varepsilon_{t+h|t}^1}$and $\widehat{\varepsilon_{t+h|t}^2}$ are estimates of the true prediction errors $\varepsilon_{t+h|t}^1$and $\varepsilon_{t+h|t}^2$. Thus, if the statistic is negative and significantly different from zero, then Model 2 is preferred over Model 1.

In concert with the various forecast model specification approaches discussed above, we form predictions using the following benchmark models, all of which are estimated using least squares.

Univariate Autoregression: Forecasts from a univariate AR(p) model are computed as $\hat{Y}^{AR}_{t+h} = \hat{\alpha} + \hat{\phi}(L) Y_t$, with lags , p, selected using the SIC.

Multivariate Autoregression: Forecasts from an ARX(p) model are computed as $Y^{ARX}_{t+h} = \hat{\alpha} + \hat{\beta} Z_t + \hat{\phi}(L) Y_t$, where Z_t is a set of lagged predictor variables selected using the SIC. Dependent variable lags are also selected using the SIC. Selection of the exogenous predictors includes choosing up to six variables prior to the construction of each new prediction model, as the recursive or rolling samples iterate forward over time.

Principal Component Regression: Forecasts from principal component regression are computed as $\hat{Y}^{PCR}_{t+h} = \hat{\alpha} + \hat{\gamma}\hat{F}_t$, where $\hat{F}_t$ is estimated via principal components using $\{X_t\}^T_{t=1}$, as in equation (3).

Factor Augmented Autoregression: Based on equations (3), forecasts are computed as $Y^h_{t+h} = \hat{\alpha} + \hat{\beta}_F \hat{F}_t + \hat{\beta}_W(L) Y_t$. This model combines an AR(p) model, with lags selected using the SIC, with the above principal component regression model.

Combined Bivariate ADL Model: As in Stock and Watson (2005a), we implement a combined bivariate autoregressive distributed lag (ADL) model. Forecasts are constructed by combining individual forecasts computed from bivariate ADL models. The i-th ADL model includes $p_{i,x}$ lags of $X_{i,t}$, and $p_{i,y}$ lags of Y_t, and has the form $\hat{Y}^{ADL}_{t+h} = \hat{\alpha} + \hat{\beta}_i(L) X_{i,t} + \hat{\phi}_i(L) Y_t$. The combined forecast is $\hat{Y}^{Comb,h}_{T+h|T} = \sum\limits_{t=1}^{n} w_i \hat{Y}^{ADL,h}_{T+h|T}$. Here, we set $(w_i = 1/n)$, where $n = 146$. There are a number of studies that compare the performance of combining methods in controlled experiments, including: Clemen (1989), Diebold and Lopez (1996), Newbold and Harvey (2002), and Timmermann (2005); and in the literature on factor models, Stock and Watson (2004, 2005a, 2006), and the references cited therein. In this literature, combination methods typically outperform individual forecasts. This stylized fact is sometimes called the "forecast combining puzzle."

Mean Forecast Combination: To further examine the issue of forecast combination, we form forecasts as the simple average of the thirteen forecasting models summarized in Table 2.

6 Empirical Results

In this section, we discuss the results of our prediction experiments. For the case where models are estimated using recursive data windows, our results are gathered in Tables 3 to 6. Detailed results based on rolling estimation are omitted for the sake of brevity, although they are available upon request from the authors. Summary statistics based upon both estimation window types are contained in Tables 7 and 8.

Tables 3-6 report MSFEs and the results of DM predictive accuracy tests for all alternative forecasting models, using Specification Type 1 without lags (Table 3), Specification Type 1 with lags (Table 4), Specification Type 2 (Table 5), and Specification Type 3 (Table 6). Panels A-C contain results for h =1, 3 and 12 month ahead prediction horizons, respectively. In each panel, the first row of entries reports the MSFE of our AR(SIC) model, and all other rows report MSFEs relative to the AR(SIC) value. Thus, entries greater than unity imply point MSFEs greater than those of our AR(SIC) model. Entries in bold denote MSFE-"best" models for a given variable,forecast horizon, and specification type. For example, in Panel C of Table 3, the MSFE-best model for unemployment (UR), when h=12, is ridge regression, with a MSFE of 0.939. Recalling that all reported entries in Tables 3-6 are for recursively estimated models, note that in each table, dot-circled entries denote cases for which the MSFE-best model yields a lower MSFE than that based on using rolling estimation, under the same specification type. For example, the ridge regression MSFE of 0.939 discussed above (i.e. see Table 3, UR, h = 12) is not dot-circled because one of the models, under rolling window estimation, yields a lower MSFE, under Specification Type 1 without lags. However, the MSFE value for UR of 0.780 in Table 3, under h = 1 is dot-circled, denoting that no model yields a lower MSFE under rolling window estimation, for Specification Type 1 with no lags. This method of reporting allows us to compare rolling window estimation results without having to actually report rolling type MSFEs in our tables. Boxed entries denote cases where models are MSFE "winners" across all specification types (i.e. across Tables 3-6), when only viewing recursively estimated models. For example, in Panel A of Table 3, the MSFE-best value for HS is ARX(SIC), and the value is boxed, denoting the fact that this MSFE value is the lowest across all 4 tables (i.e. across all specification types), under recursive estimation. Note that we do not draw a box around ARX(SIC) in other specification since it is redundant, as the ARX(SIC) in each specification type is identical (only factor methods change across specifications types; benchmark linear models remain the same). However, the fact that the entry is not also dot-circled indicates that a lower MSFE arises for one of the models when estimated using rolling windows of data, for the specification reported on in this particular table. Finally, circled entries denote models that are MSFE-best across all specification and estimation window types. For example, in Panel A of Table 3 it is apparent that FAAR in the "universal" MSFE-best model for UR, at horizon h = 1.That is, if one method in recursive estimation "wins", we circle it and do not put a box around it, as this would be redundant information.

Results from DM predictive accuracy tests, for which the null hypothesis is that of equal predictive accuracy between the benchmark model (defined to be the AR(SIC) model), and the model listed in the first column of the tables, are reported with single starred entries denoting rejection at the 10% level, and double starred entries denoting rejection at the 5% level.

Various results are apparent, upon inspection of tables. For example, for Specification Type 1, notice that in Panel A of Table 3, every forecast model yields a lower MSFE than the AR(SIC) model except bagging, when predicting the unemployment rate (UR), regardless of forecast horizon, with one exception (i.e. for $h = 12$, the ARX(SIC) model also has higher MSFE than the AR(SIC) model). Indeed, for most variables, there are various models that have lower point MSFEs than the AR(SIC) model, regardless of forecast horizon. However, there are exceptions. For example, for TB10Y, there are few models that yield lower MSFEs that the AR(SIC) model, other than when $h = 1$, regardless of specification type (compare Table 3-6). Still, even in this case, there are some models that outperform the AR(SIC) model, even for horizons other than $h = 1$, including the Combined-ADL model under Specifications 1 and 2 (see Tables 3-5, Panels B and C), and LAR or EN under Specification 3 (see Table 6, Panels B and C). Additionally, comparison of the results in Tables 3 and 4 suggests that there is little advantage to using lags of factors when constructing predictions in our context. Instead, it appears that the more important determinant of model performance is the type of combination factor/shrinkage type model employed when constructing forecasts. Evidence of this will is discussed in some detail below.

There are no models that uniformly yield lowest MSFEs, across both forecast horizon and variable. However, various models perform quite well, including in particular FAAR and PCR models. This supports the oft reported result that models that incorporate common factors offer a convenient way to filter the information contained in large-scale economic datasets.

Turning to Table 7, notice that the results reported in Panel A summarize findings of Tables 1-3. In particular, "wins" are reported across all specification types, so that each row of entries in the panel sum to 11, the number of target variables in our experiments. When comparing results for $h = 1, 3$, and 12, we see that forecasts constructed using our model averaging specifications (Combined-ADL, BMA, and Mean) yield MSFE-best predictions for 1/11 ($h = 1$), 5/11 ($h = 3$), and 3/11 ($h = 12$) variables when using only recursive estimation, and for 0/11 ($h = 1$), 5/11 ($h = 3$), and 3/11 ($h = 12$) variables, when using both recursive and rolling estimation windows. This result is quite interesting, given the plethora of recent evidence indicating the superiority of model averaging methods in a variety of forecasting contexts; and is accounted for in part by our use of various relatively complicated combined factor/shrinkage models. In particular, when combining "wins" across all three forecast horizons in the right hand section of Panel A in Table 7, note that C-Boosting, Ridge, LAR, EN, and NNG "win" in 15/33 cases. Moreover, the majority of these "wins" are accounted for by Specifications 1 and 2, suggesting that our shrinkage type methods perform best when coupled with factor analysis. In contrast, pure factor models (FAAR and PCR) yield "wins" in 8/33 cases, model averaging methods yield "wins" in 8/33 cases, and our non-factor and non-shrinkage based models "win" in 2/33 cases. Thus, the dominant model type is the combination factor/shrinkage type model. Finally, models that involve factors, in aggregate, "win" in 23/33 cases; model averaging fares quite poorly; and pure linear models are almost never MSFE-best.

As evidenced in Panel B of Table 7, MSFE-best recursively estimated models dominate MSFE-best models estimated using rolling windows around 70% of the time, regardless of forecast horizon. This is perhaps not surprising, given the number of times that our more complicated combination factor/shrinkage type models are MSFE-best across all specification and estimation types; and suggests that structural breaks play a secondary role to parameter estimation error in determining the MSFE-"best" models.[7]

It should also be noted that DM test statistics yield ample evidence that a variety of models are statistically superior to our simple linear benchmark model, including many of our more sophisticated shrinkage based models. Such models are denoted as starred entries in the tables (see Section 4.2 for further details).

Finally, turning to the results in Table 8, notice that for a single forecast horizon, $h = 1$, results have been re-calculated for sub-samples corresponding to all of the NBER-dated expansionary periods in our sample, and to the combination of all recessionary and all expansionary periods. Although results drawn from inspection of this table are largely in accord with those reported above, one additional noteworthy finding is worth stressing. Namely, in Panel A of the table, note that, when MSFE-best models are tabulated by specification type, our model averaging specifications perform quite well, particularly for Specification Types 2 and 3. This conforms to the results that can be observed by individually looking at each

[7] In lieu of this finding, the experiments carried out in this paper were replicated using the approach proposed by Clements and Hendry for addressing level shifts in the underlying data generating processes of our target variables (for details, refer to Clements and Hendry (1994, 1995, 2008)). Adjusting for level shifts by using differences of differences did not lead to notably improved prediction performance, however. (Complete results are available upon request from the authors.)

of Tables 3-6 (i.e. compare the bolded MSFE-best models in each individual table). However, notice that when results are summarized across all specification types (see Panel B of the table), then the model averaging type specifications yield MSFE-best predictions in far fewer cases. This is because Specification Type 1, where model averaging clearly "wins" the least, is the predominant winner when comparing the three specification types, as mentioned previously. Namely, the model building approach whereby we first construct factors and thereafter use shrinkage methods to estimate functions of and weights for factors to be used in our prediction models is the dominant specification type. This result serves to further stress that when more complicated specification methods are used, model averaging methods fare worse, and combination factor/shrinkage based approaches fare better. Put differently, we have evidence that when simpler linear models are specified, model averaging does worse than when more sophisticated nonlinear models are specified. Additionally, pure factor type models also perform well, particularly for the long expansion period from 1982-1990.

Given the importance of factors in our forecasting experiments, it would seem worthwhile to examine which variables contribute to the estimated factors used in our MSFE-best models, across all specification and estimation window types. This is done in Figure 1, where we report the ten most frequently selected variables for a variety of MSFE-best models and forecast horizons. Keeping in mind that factors are re-estimated at each point in time, prior to each new prediction being constructed, a 45 degree line denotes cases for which a particular variables is selected every time. For example, in Panels A and B, the BAA Bond Yield - Federal Funds Rate spread is the most frequently selected predictor when constructing factors to forecast the Producer Price Index and Housing Starts, respectively. For Specification Type 1, variables are selected based on the $A(j)$ and $M(j)$ statistics following Bai and Ng (2006a) and Armah and Swanson (2010), and for Specification Type 2, we directly observe variables which are selected by shrinkage methods and then used to construct factors, prior to the construction of each new forecast. The list of selected variables does not vary much, for Specification Type 1. On the other hand, in Panels D and F, we see that the most frequently selected variables are not selected all the time. For example, in Panel D, CPI:Apparel is selected over all periods and the 3 month Treasury bill yield is selected continuously, after 1979. Of further note is that interest-rate related variables (i.e. Treasury bills rates, Treasury bond rates, and spreads with Federal Funds Rate) are frequently selected, across all specification type, estimation window types, and forecast horizons. This confirms that in addition to their well established usefulness in linear models, yields and spreads remain important in nonlinear modelling contexts.

7 Concluding Remarks

This paper empirically examines approaches to combining factor models and robust estimation, and presents the results of a "horse-race" in which mean-square-forecast-error (MSFE) "best" models are selected, in the context of a variety of forecast horizons, estimation window schemes and sample periods. In addition to pure common factor prediction models, the forecast model specification methods that we analyze include bagging, boosting, Bayesian model averaging, ridge regression, least angle regression, the elastic net and the non-negative garotte; as well as univariate autoregressive and autoregressive plus exogenous variables models. For the majority of the target variables that we forecast, we find that various of these shrinkage methods, when combined with simple factors formed using principal component analysis (e.g. component-wise boosting), perform better than all other models. This suggests that diffusion index methodology is particularly useful when combined with other shrinkage methods, thus adding to the extant evidence of this finding (see Bai and Ng (2008, 2009), and Stock and Watson (2005a)).

We also find that model averaging methods perform surprisingly poorly, given our prior that they would "win" in most cases. Given the rather extensive empirical evidence suggesting the usefulness of model averaging when specifying linear prediction models, this is taken as further evidence of the usefulness of more sophisticated nonlinear modelling approaches.

References

Armah, N. A. and Swanson, N. R. (2010). Seeing inside the black box: Using diffusion index methodology to construct factor proxies in large scale macroeconomic time series environments. *Econometric Reviews*, 29:476–510.

Armah, N. A. and Swanson, N. R. (2011). Some variables are more worthy than others: New diffusion index evidence on the monitoring of key economic indicator. *Applied Financial Economics*, 21:43–60.

Artis, M. J., Banerjee, A., and Marcellino, M. (2002). Factor forecasts for the UK. Discussion Paper 3119, Center for Economic Policy Research.

Bai, J. and Ng, S. (2002). Determining the number of factors in approximate factor models. *Econometrica*, 70:191–221.

Bai, J. and Ng, S. (2006a). Confidence intervals for diffusion index forecasts and inference for factor-augmented regressions. *Econometrica*, 74:1133–1150.

Bai, J. and Ng, S. (2006b). Evaluating latent and observed factors in macroeconomics and finance. *Journal of Econometrics*, 131:507–537.

Bai, J. and Ng, S. (2008). Forecasting economic time series using targeted predictors. *Journal of Econometrics*, 146:304–317.

Bai, J. and Ng, S. (2009). Boosting diffusion indices. *Journal of Applied Econometrics*, 4:607–629.

Boivin, J. and Ng, S. (2005). Understanding and comparing factor-based forecasts. *International Journal of Central Banking*, 1:117–152.

Boivin, J. and Ng, S. (2006). Are more data always better for factor analysis? *Journal of Econometrics*, 132:169–194.

Breiman, L. (1995). Better subset regression using the nonnegative garrote. *Technometrics*, 37(4):373–384.

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

Bühlmann, P. and Yu, B. (2002). Analyzing bagging. *Annals of Statistics*, 30:927–961.

Bühlmann, P. and Yu, B. (2003). Boosting with the L_2 loss: Regression and classification. *Journal of the American Statistical Association*, 98:324–339.

Chipman, H., George, E. I., and Mcculloch, R. E. (2001). The practical implementation of bayesian model selection. *Lecture Notes -Monograph Series*, 38:65–134.

Chow, G. C. and Lin, A. (1971). Best linear unbiased interpolation, distribution, and extrapolation of time series by related series. *The Review of Economics and Statistics*, 53:372–75.

Clemen, R. T. (1989). Combining forecasts: A review and annotated bibliography. *International Journal of Forecasting*, 5:559–583.

Clements, M. P. and Hendry, D. F. (1994). Towards a theory of economic forecasting. In Hargreaves, C., editor, *Non-stationary time series analyses and cointegration*, pages 9–52. Oxford University Press.

Clements, M. P. and Hendry, D. F. (1995). Macro-economic forecasting and modelling. *Economic Journal*, 105:1001–1003.

Clements, M. P. and Hendry, D. F. (2008). Intercept corrections and structural change. Working paper, Oxford University.

Connor, G. and Korajczyk, R. A. (1986). Performance measurement with the arbitrage pricing theory : A new framework for analysis. *Journal of Financial Economics*, 15:373–394.

Connor, G. and Korajczyk, R. A. (1988). Risk and return in an equilibrium APT : Application of a new test methodology. *Journal of Financial Economics*, 21:255–289.

Connor, G. and Korajczyk, R. A. (1993). A test for the number of factors in an approximate factor model. *Journal of Finance*, 48:1263–91.

Diebold, F. X. and Lopez, J. A. (1996). Forecast evaluation and combination. Working Paper 0192, National Bureau of Economic Research.

Ding, A. A. and Hwang, J. T. G. (1999). Prediction intervals, factor analysis models, and high-dimensional empirical linear prediction. *Journal of the American Statistical Association*, 94:446–455.

Efron, B., Hastie, T., Johnstone, L., and Tibshirani, R. (2004). Least angle regression. *Annals of Statistics*, 32:407–499.

Fernandez, C., Ley, E., and Steel, M. F. J. (2001). Benchmark priors for bayesian model averaging. *Journal of Econometrics*, 100:381–427.

Forni, M., Hallin, M., Lippi, M., and Reichlin, L. (2005). The generalized dynamic factor model: One-sided estimation and forecasting. *Journal of the American Statistical Association*, 100:830–840.

Freund, Y. and Schapire, R. E. (1997). A decision-theoretic generalization of on-line learning and an application to boosting. *Journal of Computer and System Sciences*, 55:119–139.

Friedman, J. H. (2001). Greedy function approximation: A gradient boosting machine. *The Annals of Statistics*, 29:1189–1232.

Gelper, S. and Croux, C. (2008). Least angle regression for time series forecasting with many predictors. Working paper, Katholieke Universiteit Leuven.

Hoerl, A. and Kennard, R. (1970). Ridge regression: Biased estimation for nonorthogonal problems. *Technometrics*, 12:55–67.

Hoeting, J. A., Madigan, D., Raftery, A. E., and Volinsky, C. T. (1999). Bayesian model averaging: A tutorial. *Statistical Science*, 14:382–417.

Kim, H. H. and Swanson, N. R. (2011). Diffusion indices sing nonlinear factor methods. Working paper, Rutgers University.

Koop, G. and Potter, S. (2004). Forecasting in dynamic factor models using bayesian model averaging. *Econometrics Journal*, 7:550–565.

Newbold, P. and Harvey, D. I. (2002). Forecast combination and encompassing. In Clements, M. P. and Hendry, D. F., editors, *A Companion to Economic Forecasting*, pages 268–283. Blackwell Press, Oxford.

Pesaran, M. H., Pick, A., and Timmermann, A. (2011). Variable selection, estimation and inference for multi-period forecasting problems. *Journal of Econometrics*, 164:173–187.

Ridgeway, G., Madigan, D., and Richardson, T. (1999). Boosting methodology for regression problems. In *The Seventh International Workshop on Artificial Intelligence and Statistics*, pages 152–161. Morgan Kaufmann: New York.

Shrestha, D. L. and Solomatine, D. P. (2006). Experiments with AdaBoost.RT, an improved boosting scheme for regression. *Neural Computation*, 18:1678–1710.

Stock, J. H. and Watson, M. W. (1999). Forecasting inflation. *Journal of Monetary Economics*, 44:293–335.

Stock, J. H. and Watson, M. W. (2002). Forecasting using principal components from a large number of predictors. *Journal of the American Statistical Association*, 97:1167–1179.

Stock, J. H. and Watson, M. W. (2004). Combination forecasts of output growth in a seven-country data set. *Journal of Forecasting*, 23:405–430.

Stock, J. H. and Watson, M. W. (2005a). An empirical comparison of methods for forecasting using many predictors. Working Paper, Princeton University.

Stock, J. H. and Watson, M. W. (2005b). Implications of dynamic factor models for var analysis. Working Paper 11467, National Bureau of Economic Research.

Stock, J. H. and Watson, M. W. (2006). Forecasting with many predictors. In Elliott, G., Granger, C., and Timmermann, A., editors, *Handbook of Economic Forecasting*, volume 1, chapter 10, pages 515–554. Elsevier: Armsterdam:.

Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society, Series B*, 58:267–288.

Timmermann, A. G. (2005). Forecast combinations. Discussion Paper 5361, Center for Economic Policy Research.

Wright, J. H. (2008). Bayesian model averaging and exchange rate forecasting. *Journal of Econometrics*, 146:329–341.

Wright, J. H. (2009). Forecasting u.s. inflation by bayesian model averaging. *Journal of Forecasting*, 28:131–144.

Yuan, M. (2007). Nonnegative garrote component selection in functional anova models. In *Proceedings of The Eleventh International Conference on Artificial Intelligence and Statistics*, volume 2, pp 660-666.

Yuan, M. and Lin, Y. (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 Hastie, T. (2005). Regularization and variable selection via the elastic net. *Journal of the Royal Statistical Society Series B*, 67:301–320.

Table 1: Target Variables For Which Forecasts Are Constructed*

Series	Abbreviation	Y_{t+h}
Unemployment Rate	UR	Z_{t+1}-Z_t
Personal Income Less Transfer Payments	PI	ln(Z_{t+1}-Z_t)
10 Year Treasury Bond Yield	TB10Y	Z_{t+1}-Z_t
Consumer Price Index	CPI	ln(Z_{t+1}-Z_t)
Producer Price Index	PPI	ln(Z_{t+1}-Z_t)
Nonfarm Payroll Employment	NNE	ln(Z_{t+1}-Z_t)
Housing Starts	HS	ln(Z_t)
Industrial Production	IPX	ln(Z_{t+1}-Z_t)
M2	M2	ln(Z_{t+1}-Z_t)
S&P 500 Index	SNP	ln(Z_{t+1}-Z_t)
Gross Domestic Product	GNP	ln(Z_{t+1}-Z_t)

* Notes : Data used in model estimation and prediction construction are monthly U.S. figures for the period 1960:1-2009:5. The transformation used in forecast model specification and forecast construction is given in the last column of the table. See Section 4.1 for complete details.

Table 2: Models and Methods Used In Real-Time Forecasting Experiments*

Method	Description
AR(SIC)	Autoregressive model with lags selected by the SIC
ARX	Autoregressive model with exogenous regressors
Combined-ADL	Combined autoregressive distributed lag model
FAAR	Factor augmented autoregressive model
PCR	Principal components regression
Bagging	Bagging with shrinkage, c = 1.96
Boosting	Component boosting, M = 50
BMA(1/T)	Bayesian model averaging with g-prior = 1/T
BMA(1/N^2)	Bayesian model averaging with g-prior = 1/N^2
Ridge	Ridge regression
LARS	Least angle regression
EN	Elastic net
NNG	Non-negative garotte
Mean	Arithmetic mean

* Notes: This table summarizes the model specification methods used in the construction of prediction models. In addition to the above pure linear, factor and shrinkage based methods, three different combined factor and shrinkage type prediction model specification methods are used in our forecasting experiments, including: Specification Type1 - Principal components are first constructed, and then prediction models are formed using the above shrinkage methods (ranging from bagging to NNG) to select functions of and weights for the factors to be used in our prediction moels. Specification Type 2 - Principal component models are constructed using subsets of variables from the large-scale dataset that are first selected via application of the above shrinkage methods (ranging from bagging to NNG). This is different from the above approach of estimating factors using all of the variables. Specification Type 3 - Prediction models are constructed using only the above shrinkage methods (ranging from bagging to NNG), without use of factor analysis at any stage. See Sections 3 and 4.3 for complete details.

Table 3: Relative Mean Square Forecast Errors: Recursive Estimation, Specification Type 1 (no lags)*

Panel A: Recursive, h = 1

Method	UR	PI	TB10Y	CPI	PPI	NPE	HS	IPX	M2	SNP	GDP
AR(SIC)	12.713	0.009	40.975	0.003	0.012	0.001	2.477	0.021	0.004	0.573	0.008
ARX(SIC)	0.897	0.974	1.038	0.939	1.031	0.989	**0.900**	0.874	1.120	1.104	**0.916**
Combined-ADL	0.957**	1.052	0.987	1.030	1.019	0.938	0.977**	0.944	1.101*	1.002	1.093**
FAAR	**0.780****	0.902	0.950	0.916	0.969	**0.811***	0.961	0.804**	0.953	1.023	0.965
PCR	0.830**	**0.870**	1.019	**0.875**	**0.943**	0.922	1.764**	**0.800****	1.43**	1.018	0.973
Bagging	1.025	1.062	0.977	1.341*	1.167**	0.913	1.084	1.080	0.985	1.019	0.958
C-Boosting	0.902*	0.969	0.953	0.963	0.989	0.875**	0.949	0.848**	0.958	0.978	1.006
BMA(1/T)	0.899	0.965	0.954	0.954	0.991	0.873**	0.960	0.851**	0.972	0.989	1.018
BMA($1/N^2$)	0.892*	0.969	0.947	0.954	0.991	0.866**	0.949	0.839**	0.969	0.987	1.012
Ridge	0.887**	0.964	**0.940**	0.963	1.000	0.885*	0.938	0.816**	0.969	1.006	0.996
LARS	0.913*	0.968	0.972**	0.977	0.984	0.954**	0.981	0.949**	0.977	0.982	0.995
EN	0.913*	0.969	0.972**	0.977	0.984	0.954**	0.981	0.95**	0.977	0.982	0.995
NNG	0.966**	0.98**	0.994	0.979*	0.984	0.95**	0.989	0.984*	0.989**	0.985	0.991
Mean	0.859**	0.933**	0.942**	**0.910	0.953	0.841**	0.910**	0.845**	**0.939****	**0.976**	0.940**

Panel B: Recursive, h = 3

Method	UR	PI	TB10Y	CPI	PPI	NPE	HS	IPX	M2	SNP	GDP
AR(SIC)	12.857	0.009	47.642	0.004	0.014	0.001	5.173	0.023	0.005	0.620	0.009
ARX(SIC)	0.988	0.902	1.016*	0.981	0.945	0.940	1.000	0.895	1.000	1.028	1.032
Combined-ADL	0.977**	1.058	**0.998**	1.059*	1.045	0.948	0.955**	0.948	1.233**	1.010	1.109
FAAR	0.915	0.867**	1.026	**0.929**	0.936	**0.818****	0.895	0.866	1.006	1.052	1.058
PCR	**0.912**	**0.865****	1.004	0.930	**0.909**	0.835*	1.447**	0.859	1.164*	1.043	1.020
Bagging	1.062	1.071	1.013	1.168**	1.096	1.016	0.899	0.938	1.017	1.004	1.025
C-Boosting	0.935	0.924*	1.004	0.977	0.984	0.883*	0.852*	0.880	0.988	1.005	0.983
BMA(1/T)	0.946	0.935	1.006	0.992	0.983	0.868*	**0.852***	0.888	0.996	1.006	0.994
BMA($1/N^2$)	0.932	0.920	1.008	0.988	0.984	0.861*	0.854*	0.881	0.994	1.011	0.996
Ridge	0.919	0.893**	1.012	0.982	0.991	0.866*	0.891	0.865	0.993	1.017	0.994
LARS	0.977	0.977**	1.003	0.992	0.993	0.984	0.926*	0.963	0.997	0.994	0.974
EN	0.977	0.977**	1.003	0.992	0.993	0.984	0.926*	0.963	0.996	**0.993**	0.974
NNG	0.980*	0.992*	1.005	0.990	0.990	0.989	0.984**	0.987*	0.996	1.003	0.985*
Mean	0.920*	0.898**	1.000	0.947	0.938**	0.858**	0.862**	**0.849****	**0.977**	0.998	**0.955**

Panel C: Recursive, h = 12

Method	UR	PI	TB10Y	CPI	PPI	NPE	HS	IPX	M2	SNP	GDP
AR(SIC)	14.951	0.009	46.773	0.004	0.014	0.002	20.916	0.026	0.006	**0.620**	0.009
ARX(SIC)	1.014	0.993	1.001	1.004	1.006	0.991	1.000	0.995	1.000	1.046	1.000
Combined-ADL	0.980**	1.064	**0.996**	1.043	1.037	0.966	0.952**	0.952	1.212**	1.010	1.172**
FAAR	0.956	1.009	1.032	**0.886****	0.939	**0.874**	**0.818****	0.972	0.989	1.022	1.045
PCR	0.958	1.003	1.021	0.929	0.948	0.887	0.956	0.962	1.061	1.023	1.034
Bagging	1.072**	0.968	1.035	0.895**	0.993	1.178**	0.932	1.052*	0.982	1.003	1.008
C-Boosting	0.950	0.986	1.005	0.901**	0.955*	0.909	0.85**	0.954	0.989	1.007	1.010
BMA(1/T)	0.960	1.000	1.002	0.901*	0.955	0.922	0.852**	0.956	0.994	1.003	1.015
BMA($1/N^2$)	0.959	0.997	1.004	0.903*	0.955	0.908	0.854**	0.955	0.995	1.005	1.020
Ridge	**0.939**	0.988	1.007	0.896**	0.954	0.892	0.875**	0.949	0.991	1.007	1.021
LARS	0.959	0.981	1.005	0.983**	0.985**	0.932	0.909**	0.936	0.993	1.008	1.001
EN	0.960	0.980	1.004	0.983**	0.985**	0.932	0.909**	0.936	0.992	1.008	1.001
NNG	0.975**	0.988*	1.010	0.992**	0.991**	0.975**	0.981**	0.967**	0.992	1.011	1.000
Mean	0.942	**0.955**	1.005	0.894**	**0.939****	0.875**	0.853**	**0.918**	**0.957****	1.001	**0.999**

*Notes: See notes to Tables 1 and 2. Numerical entries in this table are mean square forecast errors (MSFEs) based on the use of various recursively estimated prediction models. Forecasts are monthly, for the period 1974:3-2009:5. Models and target variables are predicted in Tables 1 and 2. Forecast horizons reported on include h=1,3 and 12. Entries in the first row, corresponding to our benchmark AR(SIC) model, are actual MSFEs, while all other entries are relative MSFEs, such that numerical values less than unity constitute cases for which the alternative model has lower point MSFE than the AR(SIC) model. Entries in bold denote point-MSFE "best" models for a given variable and forecast horizon. Dot-circled entries denote cases for which the Specification Type 1 (no lags) MSFE-best model using recursive estimation yields a lower MSFE than that based on using rolling estimation. Circled entries denote models that are MSFE-best across all specification types and estimation types (i.e. rolling and recursive). Boxed entries denote cases where models are "winners" across all specification types, when only viewing recursively estimated models. The results from Diebold and Mariano (1995) predictive accuracy tests, for which the null hypothesis is that of equal predictive accuracy between the benchmark model (defined to be the AR(SIC) model), and the model listed in the first column of the table, are reported with single starred entries denoting rejection at the 10% level, and double starred entries denoting rejection at the 5% level. See Sections 4 and 5 for complete details.

Table 4: Relative Mean Square Forecast Errors: Recursive Estimation, Specification Type 1 (with lags)*

Panel A: Recursive, h = 1

Method	UR	PI	TB10Y	CPI	PPI	NPE	HS	IPX	M2	SNP	GDP
AR(SIC)	12.713	0.009	40.975	0.003	0.012	0.001	2.477	0.021	0.004	0.573	0.008
ARX(SIC)	0.897	0.974	1.038	0.939	1.031	0.989	**0.900**	0.874	1.120	1.104	**0.916**
Combined-ADL	0.957**	1.052	0.987	1.030	1.019	0.938	0.977**	0.944	1.101*	1.002	1.093**
FAAR	**0.850***	0.926	1.044	0.888	1.008	1.005	1.079	0.851	0.968	1.095	1.050
PCR	0.908	**0.888**	1.058	**0.864**	1.002	0.999	1.646**	0.855	1.292**	1.091	1.076
Bagging	1.287**	1.017	1.069*	2.566**	1.545**	2.160**	1.851**	1.304**	1.028	1.131**	0.962
C-Boosting	0.903	0.968	0.961	0.951	1.002	0.910	0.945	0.827**	0.963	**0.975**	1.005
BMA(1/T)	0.910	0.972	0.988	0.942	1.018	0.904	0.956	**0.804****	0.959	1.012	1.019
BMA($1/N^2$)	0.907	0.962	0.996	0.955	1.023	0.904	0.954	0.816**	0.947	1.002	1.022
Ridge	0.911	0.959	0.988	0.919	1.014	0.944	0.992	0.821**	0.977	1.048	1.040
LARS	0.975**	0.977*	0.981	0.988	0.988	0.967*	0.974	0.948**	0.972*	0.989	0.995
EN	0.977**	0.978**	0.982	0.988	0.988	0.969*	0.975	0.949**	*0.970	0.989	0.992
NNG	0.972**	0.990	0.994	0.984	0.996	0.975	0.989	0.964**	0.993	0.993	0.994
Mean	0.867**	0.922**	**0.955**	0.889**	**0.944**	**0.879****	0.922*	0.821**	****0.930**	0.977	0.948*

Panel B: Recursive, h = 3

Method	UR	PI	TB10Y	CPI	PPI	NPE	HS	IPX	M2	SNP	GDP
AR(SIC)	12.857	0.009	47.642	0.004	0.014	0.001	5.173	0.023	0.005	0.620	0.009
ARX(SIC)	0.988	0.902	1.016*	0.981	0.945	0.940	**1.000**	0.895	1.000	1.028	**1.032**
Combined-ADL	0.977**	1.058	0.998	1.059*	1.045	0.948	0.955**	0.948	1.233**	1.010	1.109
FAAR	**1.014**	0.931	1.106	0.907	0.992	0.886	0.898	0.925	1.069	1.117*	1.144
PCR	0.999	**0.928**	1.092	**0.906**	0.975	0.898	1.404**	0.921	1.249**	1.107	1.115
Bagging	1.174**	1.017	1.141**	1.339**	1.204*	1.295**	1.050	1.010	0.995	1.007	1.087**
C-Boosting	0.951	0.914*	1.010	0.946	0.969	0.832**	0.879	0.868	1.006	**1.007**	0.967
BMA(1/T)	0.944	0.932	1.020	0.943	0.982	0.818**	0.903	**0.851**	1.027	1.030	0.990
BMA($1/N^2$)	0.954	0.942	1.011	0.953	0.981	0.836*	0.889	0.862	1.020	1.011	0.979
Ridge	0.944	0.917	1.047	0.933	0.992	0.844	0.891	0.869	1.046	1.064	1.033
LARS	0.979	0.973**	0.992	0.984	0.982	0.968	0.951**	0.962	0.996	1.000	0.969
EN	0.973*	0.975**	0.991	0.983	0.986	0.963	0.965**	0.962	0.996	1.000	0.969**
NNG	0.980	0.986*	1.001	0.991	0.995	0.963**	0.977**	0.967**	0.993	0.993	*0.970
Mean	0.924	0.891**	**0.988**	0.901**	**0.928****	**0.84****	0.851**	0.838**	**0.977**	0.997	0.962

Panel C: Recursive, h = 12

Method	UR	PI	TB10Y	CPI	PPI	NPE	HS	IPX	M2	SNP	GDP
AR(SIC)	14.951	0.009	46.773	0.004	0.014	0.002	20.916	0.026	0.006	0.621	0.009
ARX(SIC)	1.014	0.993	1.001	1.004	1.006	0.991	**1.000**	0.995	1.000	1.046	**1.000**
Combined-ADL	0.980**	1.064	0.997	1.043	1.037	0.966	0.952**	0.952	1.212**	1.010	1.172**
FAAR	**0.985**	1.070	1.087	0.938	0.951	0.932	0.841*	1.082	1.049	1.081*	1.145**
PCR	0.983	**1.069**	1.081	**0.932**	0.942	0.924	1.020	1.071	1.116*	1.081*	1.132**
Bagging	1.003	**1.050	1.053	1.137*	1.078	1.174**	0.900	1.104**	0.971	1.034	1.001
C-Boosting	0.913	0.985	0.988	0.89**	0.947	0.896	0.846**	0.947	0.941	**0.999**	1.031
BMA(1/T)	0.930	1.007	1.002	0.908	0.935*	0.888	0.853**	**0.975**	0.981	1.006	1.031
BMA($1/N^2$)	0.936	0.997	0.999	0.909*	0.952	0.907	0.833**	0.964	0.982	1.002	1.019
Ridge	0.926	1.005	1.029	0.897	0.931	0.867	0.887*	1.006	1.001	1.029	1.067
LARS	0.968**	0.973	0.992	0.974**	0.988	0.974*	0.923**	0.963*	0.973**	0.995	1.004
EN	0.969**	0.971	0.992	0.972**	0.989	0.963**	0.929**	0.965*	0.975**	0.994	1.003
NNG	0.979**	0.985	1.002	0.993	1.007	0.975**	0.967**	0.978*	0.994	0.998	0.999
Mean	0.902**	0.956	**0.995**	0.888**	**0.927****	****0.860**	0.829**	0.925	**0.943****	0.999	1.010

*Notes: See notes to Table 3. Dot-circled entries denote cases for which the Specification Type 1 (lags) MSFE-best model using recursive estimation yields lower MSFE than using rolling estimation. Circled entries denote models that are MSFE-best across all specification types and estimation types (i.e. rolling and recursive). Boxed entries denote cases where models are "winners" across all specification types, when only viewing recursively estimated models.

Table 5: Relative Mean Square Forecast Errors: Recursive Estimation, Specification Type 2*

Panel A: Recursive, h = 1

Method	UR	PI	TB10Y	CPI	PPI	NPE	HS	IPX	M2	SNP	GDP
AR(SIC)	12.713	0.009	40.975	0.003	0.012	0.001	2.477	0.021	0.004	0.573	0.008
C-Boosting	**0.891***	0.962	0.971	0.961	1.024	0.887	0.961	0.906	1.047	1.011	**0.865****
BMA(1/T)	0.896*	**0.956**	1.005	0.968	0.989	**0.870****	0.990	**0.864****	0.960	0.995	1.013
BMA(1/N^2)	0.900*	0.962	0.986	**0.945**	0.983	0.899*	**0.942**	0.893**	**0.926**	1.019	1.012
LARS	0.914**	0.994	0.972**	0.998	1.008	0.916**	0.978	0.996	0.982**	0.983	0.876**
EN	1.149*	1.217	1.118	3.646**	1.464**	2.804**	11.041**	1.186**	4.340**	1.092**	1.308**
NNG	0.993**	0.996*	0.997	0.999	1.000	0.991**	1.001*	0.997*	1.000	1.001	1.000
Mean	0.907**	0.963**	**0.968**	0.960	**0.979**	0.886**	0.953**	0.902**	0.951*	**0.984**	0.93**

Panel B: Recursive, h = 3

Method	UR	PI	TB10Y	CPI	PPI	NPE	HS	IPX	M2	SNP	GDP
AR(SIC)	12.857	0.009	47.642	0.004	0.014	0.001	5.173	0.023	0.005	0.620	0.009
C-Boosting	**0.934**	0.902	1.028	0.946	1.020	0.847**	0.780	0.819*	1.016	1.017	**0.985**
BMA(1/T)	0.959	**0.920**	1.011	0.996	1.023	**0.903**	0.882	**0.902**	0.994	1.009	0.991
BMA(1/N^2)	0.946	0.937	1.006	**1.005**	1.011	0.912	**0.871**	0.890**	**1.001**	1.010	1.027
LARS	0.983	0.982**	1.000	0.996	1.005	0.968**	0.937	0.960*	0.990	0.998	0.994
EN	1.136**	1.206**	0.961	2.678**	1.280**	2.166**	5.287**	1.103*	3.488**	1.010	**1.240
NNG	0.997**	0.996**	1.000	0.997	0.998	0.995**	1.000	0.999	0.999	1.001	0.998**
Mean	0.943	0.922**	**1.005**	0.966	**0.994**	0.887**	0.827**	0.871**	0.976	**0.997**	0.966

Panel C: Recursive, h = 12

Method	UR	PI	TB10Y	CPI	PPI	NPE	HS	IPX	M2	SNP	GDP
AR(SIC)	14.951	0.009	46.773	0.004	0.014	0.002	20.916	0.026	0.006	0.620	0.009
C-Boosting	**0.936**	0.976	1.031	0.907	0.972	0.845*	0.786**	0.940	0.962	1.016	**1.006**
BMA(1/T)	0.947	**1.000**	1.003	0.902**	0.991	**0.930**	0.887*	**0.959**	0.997	1.004	1.011
BMA(1/N^2)	0.938	1.007	1.003	**0.917***	0.975	0.920	**0.881****	0.993	**0.981**	1.007	1.024
LARS	0.957	0.979	1.002	0.970**	0.979**	0.966**	0.910**	0.912**	0.959**	1.006	0.981
EN	0.977	1.19**	0.979	2.497**	1.251**	1.242**	1.307**	0.977	3.206**	1.010	1.226**
NNG	0.997**	0.999	1.001	0.997**	0.997**	0.995**	0.997**	0.995**	0.999	1.002	0.999
Mean	0.933	0.965	**1.004**	0.913**	**0.966****	0.892**	0.846**	0.925*	0.961*	**1.004**	0.994

*Notes: See notes to Table 3. Dot-circled entries denote cases for which the Specification Type 2 MSFE-best model using recursive estimation yields lower MSFE than using rolling estimation. Circled entries denote models that are MSFE-best across all specification types and estimation types (i.e. rolling and recursive). Boxed entries denote cases where models are "winners" across all specification types, when only viewing recursively estimated models.

Table 6: Relative Mean Square Forecast Errors: Recursive Estimation, Specification Type 3*

Panel A: Recursive, h = 1

Method	UR	PI	TB10Y	CPI	PPI	NPE	HS	IPX	M2	SNP	GDP
AR(SIC)	12.713	0.009	0.000	0.003	0.012	0.001	2.477	0.021	0.004	0.573	0.008
ARX(SIC)	**0.897**	0.974	1.038	0.939	1.031	0.989	**0.900**	**0.874**	1.120	1.104	0.916
Combined-ADL	0.957**	1.052	**0.987**	1.030	1.019	0.938	0.977**	0.944	1.101*	1.002	1.093**
C-Boosting	0.944	0.965*	0.992	0.962	0.975	0.910	0.924*	0.936	1.010	**0.988**	0.915**
BMA(1/T)	1.012	1.137	1.059	1.541	1.223**	1.685**	1.250	0.980	1.193	1.231**	0.933
BMA(1/N^2)	0.933	0.985	1.028	1.018	1.089	1.042	1.066	0.891	1.131	1.077	0.911
Ridge	1.668**	1.575**	1.424**	1.547**	1.643**	1.743**	1.795**	1.789**	1.430**	1.688**	1.388**
LARS	1.952**	0.993	1.797**	0.998	1.008	0.914**	2.02**	1.008	0.978**	1.975**	0.875**
EN	1.057	0.994	1.116	0.998	1.008	0.916**	1.082	0.996	0.982**	1.258**	0.876**
NNG	0.993**	0.996*	0.997	0.999	1.000	0.991**	1.001*	0.997*	1.000	1.001	1.000
Mean	0.924	**0.943***	0.995	**0.933**	**0.956**	**0.826****	0.910	0.875**	**0.977**	1.045	**0.873****

Panel B: Recursive, h = 3

Method	UR	PI	TB10Y	CPI	PPI	NPE	HS	IPX	M2	SNP	GDP
AR(SIC)	12.857	0.009	47.642	0.004	0.014	0.001	5.173	0.023	0.005	0.62	0.009
ARX(SIC)	**0.988**	0.902	1.016*	0.981	0.945	0.940	**1.000**	**0.895**	1.000	1.028	1.032
Combined-ADL	0.977**	1.058	**0.998**	1.059*	1.045	0.948	0.955**	0.948	1.233**	1.010	1.109
C-Boosting	0.943	0.951*	1.010	0.999	1.016	0.899**	0.820**	0.886**	0.980	**1.014**	0.974
BMA(1/T)	1.154	1.022	1.241**	1.092	1.094	1.109	1.041	1.076	1.089	1.158*	1.168**
BMA(1/N^2)	0.969	0.922	1.025	1.047	1.034	0.877	0.941	0.881	1.063	1.034	1.011
Ridge	1.873**	1.517**	1.743**	1.362**	1.479**	1.675**	1.133	1.811**	1.447**	1.813**	1.95**
LARS	2.183**	0.977**	1.923**	0.997	1.006	0.962**	1.299	0.958**	0.989	2.099**	1.255**
EN	1.169	0.982**	1.319**	0.996	1.005	0.968**	0.828	0.96**	0.990	1.243**	0.994
NNG	0.997**	0.996**	1.000	0.997	0.998	0.995**	1.001	0.999	0.999	1.000	0.998**
Mean	0.991	**0.911****	1.070	**0.926***	**0.953**	**0.859****	0.723**	0.881**	**0.938***	1.033	**0.992**

Panel C: Recursive, h = 12

Method	UR	PI	TB10Y	CPI	PPI	NPE	HS	IPX	M2	SNP	GDP
AR(SIC)	14.951	0.009	46.773	0.004	0.014	0.002	20.916	0.026	0.006	0.62	0.009
ARX(SIC)	**1.014**	0.993	1.001	1.004	1.006	0.991	**1.000**	**0.995**	1.000	1.046	1.000
Combined-ADL	0.980**	1.064	**0.996**	1.043	1.037	0.966	0.952**	0.952	1.212**	1.010	1.172**
C-Boosting	0.926	0.961	1.015	0.934*	0.971	0.862**	0.874**	0.934	0.969	**1.007**	0.995
BMA(1/T)	1.233	1.073	1.152**	1.298**	1.199	1.760	**1.760	1.164	1.366**	1.082	1.254**
BMA(1/N^2)	1.019	1.009	1.039	1.127	1.106	1.447	1.618**	0.958	1.163*	1.017	1.074
Ridge	1.555**	1.807**	1.752**	1.382**	1.677*	1.859*	1.087	1.936**	1.316**	1.794**	1.925**
LARS	1.858**	0.979	1.983**	0.975*	0.979*	1.123	1.312	2.212**	0.957**	2.226**	0.983
EN	1.207	0.978	1.327**	0.97**	0.979**	0.966**	0.803**	0.889	0.959**	1.283**	0.981
NNG	0.997**	0.999	1.001	0.997**	0.997**	0.995**	0.997**	0.995**	0.999	1.002	0.999
Mean	0.960	**0.966**	1.076*	**0.899****	**0.953**	**0.885**	0.840**	0.925	****0.910**	1.047	**1.011**

*Notes: See notes to Table 3. Dot-circled entries denote cases for which the Specification Type 3 MSFE-best model using recursive estimation yields lower MSFE than using rolling estimation. Circled entries denote models that are MSFE-best across all specification types and estimation types (i.e. rolling and recursive). Boxed entries denote cases where models are "winners" across all specification types, when only viewing recursively estimated models.

Table 7: Forecast Experiment Summary Results*

Panel A: Summary of MSFE-"best" Models Across All Specification Types

	Recursive Estimation Window			Recursive and Rolling Estimation Windows		
	h = 1	h = 3	h = 12	h = 1	h = 3	h = 12
AR(SIC)	0	0	0	1	0	0
ARX(SIC)	1	0	0	1	0	0
Combined-ADL	0	0	0	0	0	0
FAAR	2	0	1	2	0	1
PCR	4	3	0	3	2	0
Bagging	0	0	0	0	0	0
C-Boosting	2	1	2	3	2	3
BMA(1/T)	0	1	0	0	0	0
BMA($1/N^2$)	0	0	0	0	2	0
Ridge	1	0	0	1	0	0
LARS	0	0	1	0	0	1
EN	0	1	3	0	1	3
NNG	0	1	1	0	1	0
Mean	1	4	3	0	3	3

Panel B: Summary of MSFE-"best" Models

	Winners by Estimaton Window Type			Winners by Specification Type		
	h = 1	h = 3	h = 12	h = 1	h = 3	h = 12
Specification Type 1						
Rolling	2	2	3	7	4	5
Recursive	9	9	8			
Specification Type 2						
Rolling	5	9	4	4	6	5
Recursive	6	2	7			
Specification Type 3						
Rolling	3	2	2	0	1	1
Recursive	8	9	9			

*Notes: See notes to Table 3. Specification types are defined as follows. Specification Type1 - Principal components are first constructed, and then prediction models are formed using the above shrinkage methods (ranging from bagging to NNG) to select functions of and weights for the factors to be used in our prediction model. Specification Type 2 - Principal component models are constructed using subsets of variables from the largescale dataset that are first selected via application of the above shrinkage methods (ranging from bagging to NNG). This is different from the above approach of estimatiing factors using all of the variables. Specification Type 3 - Prediction models are constructed using only the above shrinkage methods (ranging from bagging to NNG), without use of factor analysis at any stage.

Table 8: Forecast Experiment Summary Results: Various Subsamples*

Panel A: Wins by Specification Type

h = 1, Recursive Estimation

	Specification Type 1				Specification Type 2				Specification Type 3			
Subsample	Mean	Linear Factor	Nonlinear Factor	Other	Mean	Linear Factor	Nonlinear Factor	Other	Mean	Linear Factor	Nonlinear Factor	Other
75:03 ~ 79:12	3	1	5	2	4	0	6	1	3	0	5	3
80:07 ~ 81:06	1	4	2	4	5	0	5	1	6	0	2	3
82:11 ~ 90:06	1	8	2	0	8	0	3	0	4	0	4	3
91:03 ~ 01:02	5	2	2	2	6	0	5	0	8	0	1	2
01:11 ~ 07:11	5	0	4	2	6	0	5	0	5	0	2	4
Non Recession	1	6	2	2	8	0	3	0	7	0	3	1
Recession	3	5	1	2	5	0	6	0	7	0	2	2

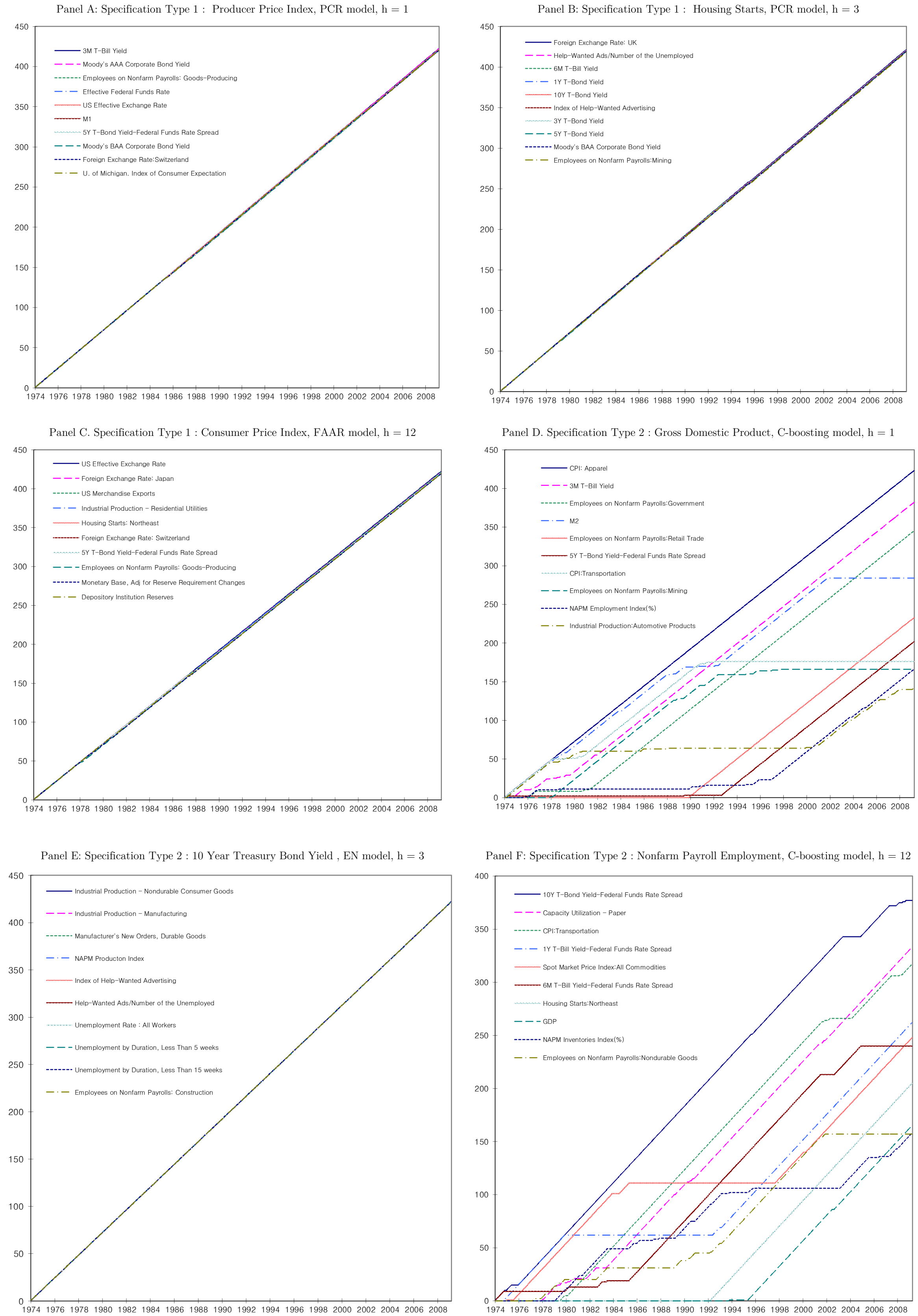
Panel B: Wins Across All Specification Types

h = 1, Recursive Estimation

	Specification Type 1				Specification Type 2				Specification Type 3			
Subsample	Mean	Linear Factor	Nonlinear Factor	Other	Mean	Linear Factor	Nonlinear Factor	Other	Mean	Linear Factor	Nonlinear Factor	Other
75:03 ~ 79:12	2	1	3	1	0	0	2	0	1	0	1	0
80:07 ~ 81:06	0	4	0	2	1	0	3	0	1	0	0	0
82:11 ~ 90:06	1	8	0	0	1	0	0	0	1	0	0	0
91:03 ~ 01:02	3	1	2	0	0	0	0	0	4	0	0	1
01:11 ~ 07:11	0	4	3	1	1	0	0	0	2	0	0	0
Non Recession	1	5	2	0	0	0	1	0	2	0	0	0
Recession	1	3	0	1	0	0	1	0	4	0	1	0

*Notes: See notes to Tables 3 and 7. In the above table, "Mean" includes the following models: BMA, Combined-ADL and Mean. "Linear Factor" includes the following models: FAAR and PCR. "Nonlinear Factor" includes the following models: all shrinkage/factor combination models (i.e. Specification Types 1 and 2). Finally, "Other" includes our linear AR(SIC) and ARX(SIC) models. See Section 4.3 for further details.

Figure 1: Most Frequently Selected Variables by Various Specification Types*



*Notes: Panels in this figure depict the 10 most commonly selected variables for use in factor construction, across the entire prediction period from 1974:3-2009:5, where factors are re-estimated at each point in time, prior to each new prediction being constructed. 45 degree lines denote cases for which a particular variables is selected every time. All models reported on are MSFE-best models, across Specification Types 1 and 2, and estimation window types. For example, in Panels A and B, the BAA Bond Yield - Federal Funds Rate spread is the most frequently selected predictor when constructing factors to forecast the Producer Price Index and Housing Starts, respectively. Note that in Panel E, the 10 most commonly selected variables by EN are picked at