Journal of Econometrics 178 (2014) 352–367

Contents lists available at ScienceDirect

Journal of Econometrics

journal homepage: www.elsevier.com/locate/jeconom

# Forecasting financial and macroeconomic variables using data reduction methods: New empirical evidence☆

Hyun Hak Kim [a,*], Norman R. Swanson [b]

[a] The Bank of Korea, 55 Namdaemunno, Jung-Gu, Seoul 100-794, Republic of Korea

[b] Department of Economics, Rutgers University, 75 Hamilton Street, New Brunswick, NJ 08901, USA

ARTICLE INFO

*Article history:*
Available online 5 September 2013

*JEL classification:*
C1
C22
C52
C58

*Keywords:*
Bagging
Bayesian model averaging
Boosting
Diffusion index
Elastic net
Forecasting
Least angle regression
Non-negative garotte
Prediction
Reality check
Ridge regression

ABSTRACT

In this paper, we empirically assess the predictive accuracy of a large group of models that are specified using principle components and other shrinkage techniques, 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 "hybrid" combination factor/shrinkage methods often yields superior predictions. More specifically, when using recursive estimation windows, which dominate other "windowing" approaches, "hybrid" models are mean square forecast error "best" around 1/3 of the time, when used to predict 11 key macroeconomic indicators at various forecast horizons. Baseline linear (factor) models also "win" around 1/3 of the time, as do model averaging methods. Interestingly, these broad findings change noticeably when considering different sub-samples. For example, when used to predict only recessionary periods, "hybrid" models "win" in 7 of 11 cases, when condensing findings across all "windowing" approaches, estimation methods, and models, while model averaging does not "win" in a single case. However, in expansions, and during the 1990s, model averaging wins almost 1/2 of the time. Overall, combination factor/shrinkage methods "win" approximately 1/2 of the time in 4 of 6 different sample periods. 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 model specification.



## 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 (big) 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 $(T \times 1)$, 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 $\Lambda$ 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 (AR) forecasting models augmented by the inclusion of common factors.

In this paper, we use the forecasting models of Stock and Watson (2002a,b) and Bai and Ng (2006a) as a starting point. In particular, we first estimate unobserved factors, say $\hat{F}_t$, and then forecast a scalar target variable, $Y_{t+h}$, using observed variables

☆ 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 Harvey, for hosting this special event. We have benefited from numerous useful comments made on earlier versions of this paper by the editor, Graham Elliott, and by an anonymous referee. In addition to thanking the editor and referee, the authors wish to thank the seminar participants from the conference, as well as at the Bank of Canada and Rutgers University, for useful comments. Additional thanks are owed to Nii Armah, Dongjun Chung, 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 views stated herein are those of the authors and not necessarily those of the Bank of Korea.

* Corresponding author.
*E-mail addresses:* khdouble@bok.or.kr (H.H. Kim), nswanson@econ.rutgers.edu (N.R. Swanson).

0304-4076/$ – see front matter © 2013 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.jeconom.2013.08.033

and $\hat{F}_t$. We then draw on the fact that even though factor models are now widely used, several 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 implement prediction models where the numbers and functions of factors are selected using a variety of shrinkage methods. In this sense, we add to the recent work of Stock and Watson (2012) as well as Bai and Ng (2008, 2009), who survey several methods for shrinkage in the context of 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 also 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 models; principal components models where the factors are constructed using subsets of variables first selected based on shrinkage techniques; principle components models where the factors are first constructed, and are then refined using shrinkage methods; 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 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 "wins" as often as is usually found to be the case in empirical investigations of this sort.

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, non-farm 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 has been 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, for a number of our target variables, we find that various sophisticated shrinkage methods, such as component-wise boosting, bagging, ridge regression, least angle regression, the elastic net, and the non-negative garotte yield predictions with lower mean square forecast errors (MSFEs) than a variety of benchmark linear autoregressive forecasting models constructed using only observable variables. Moreover, these shrinkage methods, when used in conjunction with diffusion indexes, yield a surprising number of MSFE "best" models, hence suggesting that "hybrid" models that combine diffusion index methodology with other shrinkage techniques offer a convenient way to filter the information contained in large-scale economic datasets, particularly if they are specified using sophisticated shrinkage techniques. More specifically, when using recursive estimation windows, which dominate other "windowing" approaches, "hybrid" models are MSFE "best" around 1/3 of the time, when used to predict 11 key macroeconomic indicators at various forecast horizons. Baseline linear (factor) models also "win" around 1/3 of the time, as do model averaging methods. Interestingly, these broad findings change noticeably when considering different sub-samples. For example, when used to predict only recessionary periods, "hybrid" models "win" in 7 of 11 cases, when condensing findings across all "windowing" approaches, estimation methods, and models, while model averaging does not "win" in a single case. However, in expansions, and during the 1990s, model averaging wins almost 1/2 of the time. Overall, combination factor/shrinkage methods "win" approximately 1/2 of the time in 4 of 6 different sample periods. Ancillary findings based on our forecasting experiments underscore the advantages of using recursive estimation strategies,[1] and provide new evidence of the usefulness of yield and yield-spread variables in nonlinear prediction model 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 Sir 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 diffusion index models. In Section 3, we briefly survey the robust shrinkage estimation methods used in our prediction experiments. Data, forecasting methods, and benchmark forecasting models are discussed in Sections 4 and 5, while empirical results are presented in Section 6. Concluding remarks are gathered in Section 7.

## 2. Diffusion index models

Recent forecasting studies using large-scale datasets and pseudo out-of-sample forecasting include: Artis et al. (2005), Boivin and Ng (2005, 2006), Forni et al. (2005), and Stock and Watson (1999, 2002a,b, 2005, 2006, 2012). 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 (2002a).

Let $X_{tj}$ be the observed datum for the $j$-th cross-sectional unit at time $t$, for $t = 1, \ldots, T$ and $j = 1, \ldots, N$. 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, $\Lambda_j$ 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 $r \ll N$, as is usually assumed to be the case in the empirical literature. Following Bai and Ng (2002), the whole panel of data $X = (X_1, \ldots, X_N)$, where $X_i$, $i = 1, \ldots, N$, is a $T \times 1$ vector of observations on a single variable, can be represented as in (1). Connor and Korajczyk (1986,

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

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 (2002a) 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 $e_t = (e_{t1}, \ldots, e_{tN})$ does not have to be a diagonal matrix. We will also assume that $\{F_t\}$ and $\{e_{tj}\}$ are two groups of mutually independent stochastic variables. Furthermore, it is well known that if $\Lambda = (\Lambda_1, \ldots, \Lambda_N)'$ for $F_t\Lambda' = F_tQQ^{-1}\Lambda'$, a normalization is needed in order to identify 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 (2002a) 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 criteria to consistently estimate the number of factors. Namely, 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):

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

A consistent estimate of the true number of factors is $\hat{r} = \arg\min_{0\le r\le r_{\max}} \mathrm{SIC}(r)$. In a number of our models, we use this criterion 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 in prediction models. In addition, shrinkage methods are also directly implemented, yielding "factor-free" prediction models. Finally, shrinkage methods are used to "pre-select" subsets of $X$ for subsequent use in the construction of factors.

The basic structure of the forecasting models examined is the same as that examined in Artis et al. (2005), Bai and Ng (2002, 2006a,b, 2008, 2009), Boivin and Ng (2005) and Stock and Watson (2002b, 2005, 2006, 2012). 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 $Y_t$, $\varepsilon_t$ is a disturbance term, and the $\beta$'s are parameters estimated using least squares. 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). As discussed above, one aspect of our empirical analysis is that we use various shrinkage methods for estimating $\hat{\beta}_F$ and then compare the predictive accuracy of the resulting forecasting models.[2]

[2] We refer the reader to Stock and Watson (1999, 2002a, 2005, 2012) and Bai and Ng (2002, 2009, 2008) 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 factor augmented autoregression models. Finally, note that Ding and Hwang (1999) also analyze the properties of forecasts constructed using principal components when $N$ and $T$ are large, although they carry out their analysis under the assumption that the error processes $\{e_{tj}, \varepsilon_{t+h}\}$ are cross-sectionally and serially i.i.d.

## 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, the 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 principal components 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 key papers that we cite as well as discussed in detail in an earlier working paper version of the current paper (see Kim and Swanson, 2013).

### 3.1. Statistical learning (bagging and boosting)

#### 3.1.1. Bagging

Bagging, which is 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. If a bootstrap sample based predictor is defined as $\hat{Y}_b^* = \hat{\beta}_b^* X_b^*$, where $b = 1, \ldots, B$ denotes the $b$-th bootstrap sample drawn from the original dataset, then the bagging predictor is $\hat{Y}^{\mathrm{Bagging}} = \frac{1}{B}\sum_{b=1}^{B}\hat{Y}_b^*$. In this paper, we follow Bühlmann and Yu (2002) and Stock and Watson (2012) who note that, asymptotically, the bagging estimator can be represented in shrinkage form. Namely:

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

where $\hat{Y}_{t+h}^{\mathrm{Bagging}}$ 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 $\psi$ is a function specific to the forecasting method. In the current context we set:

$$\begin{aligned}\psi(t) &= 1 - \Phi(t+c) + \Phi(t-c)\\ &\quad + t^{-1}[\phi(t-c) - \phi(t+c)],\end{aligned} \tag{5}$$

where $c$ is the pretest critical value, $\phi$ is the standard normal density and $\Phi$ is the standard normal CDF. In this paper, we follow Stock and Watson (2012), and set the pretest critical value for bagging, $c$ to be 1.96.

*3.1.2. Boosting*

Boosting (see e.g., 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 $\kappa_m$ can be interpreted as weights, and $f\left(X; \beta_m\right)$ are functions of the panel dataset, $X$. Friedman (2001) introduce "$L_2$ Boosting", which takes the simple approach of refitting "base learners" to residuals from previous iterations.[3] Bühlmann and Yu (2003) develop a boosting algorithm fitting "learners" using one predictor at a time, in contexts where large numbers of predictors are available, and data are i.i.d. Bai and Ng (2009) modify this algorithm to handle time-series. We use their "Component-Wise $L_2$ Boosting" 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 scalar output of the boosting algorithm as $\hat{\mu}^M\left(\hat{F}_t\right)$. Then, predictions are constructed using the following model:

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

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 e.g., Kim and Swanson, 2013 for further discussion). Ridge regression is characterized by an $L_2$ penalty function, while several recent advances in penalized regression have centered to some extent on $L_1$ penalty functions. For example, there has been much recent 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 include 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 stage-wise regression approaches, we 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 using $M$ iterations, say. Possible explanatory variables are incrementally examined, and they are added to the estimator function, $\widehat{\mu}$, according to their explanatory power. Following the same notation as that used above, in the case where shrinkage is done solely on common factors, the objective is to construct predictions,

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

*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" (see 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 $\eta_1$ and $\eta_2$, the elastic net criterion is defined as:

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

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

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

where $\hat{\beta}^{\text{LS}}$ is the least square estimator of $\beta$ and sign $(\cdot)$ equals $\pm 1$. 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; and so that they propose a version of the elastic net in which this double shrinkage is corrected. In this context, the elastic net estimator, $\hat{\beta}^{\text{EN}}$, is defined as:

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

where $\eta_2$ is a constant, usually "optimized" via cross validation methods. Zou and Hastie (2005) propose an algorithm called "LAR-EN" to estimate $\hat{\beta}^{\text{EN}}$.[4] In the current context, $\hat{\beta}^{\text{EN}}$ is either the coefficient vector associated with the $\hat{F}_t$ in a forecasting model of the variety given in (3), assuming that $\psi(\cdot) = 1$, or is a coefficient vector constructed by operating directly on the panel dataset, $X$.

*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. We follow Yuan and Lin (2007), who develop an efficient garotte algorithm and prove consistency in variable selection. As far as we know, this method has previously not been used in the time series econometrics literature. As usual, we begin by considering standardized $X$ and $Y$. Assume that the following shrinkage factor is given: $q(\zeta) = (q_1(\zeta), q_2(\zeta), \ldots, q_N(\zeta))'$, where $\zeta > 0$ is a tuning parameter. The objective is to choose the shrinkage factor in order to minimize:

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

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

[4] We use their algorithm, which is discussed in more detail in Kim and Swanson (2013).

where $G = (G_1, \ldots, G_N)'$, $G_j = X_j\widehat{\beta}_j^{LS}$, and $\widehat{\beta}^{LS}$ is the least squares estimator. The NNG estimator of the regression coefficient vector is defined as $\hat{\beta}_j^{NNG} = q_j(\zeta)\,\hat{\beta}_j^{LS}$, and the estimate of $Y$ is defined as $\widehat{\mu} = X\hat{\beta}^{NNG}(\zeta)$, so that predictions can be formed in a manner that is analogous to that discussed in the previous subsections. Assuming, for example, that $X'X = I$, the minimizer of expression (10) has the following explicit form: $q_j(\zeta) = \left(1 - \frac{\zeta}{(\hat{\beta}_j^{LS})^2}\right)_+$, for $j = 1, \ldots, N$. 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 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).

### 3.3. Bayesian model averaging

In recent years, Bayesian model averaging (BMA) has been applied to many forecasting problems, and has frequently been 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 (2013).

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 $M_1, \ldots, M_Q$, 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, and is different from 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, see Kim and Swanson (2013).

## 4. Data

Following a long tradition in the diffusion index literature, we examine monthly data observations on 144 US 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 production, 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 monetary policy, as noted in Armah 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, 2002b, 2012 for complete details). Note that the full list of predictor variables is provided in the appendix to an earlier working paper version of the current paper, and is available upon request from the authors.

[5] This is an updated and expanded version of the Stock and Watson (2005, 2012) dataset.

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

**Table 1**
Target forecasting variables.[a]

| 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 | TB | $Z_{t+1} - Z_t$ |
| Consumer price index | CPI | $\ln(Z_{t+1}/Z_t)$ |
| Producer price index | PPI | $\ln(Z_{t+1}/Z_t)$ |
| Nonfarm payroll employment | NPE | $\ln(Z_{t+1}/Z_t)$ |
| 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 | $\ln(Z_{t+1}/Z_t)$ |

[a] Notes: Data used in model estimation and prediction construction are monthly US figures for the period 1960:1–2009:5. Data transformations used in prediction experiments are given in the last column of the table. See Section 4 for further details.

## 5. Forecasting methodology

Using the transformed dataset, factors are estimated as discussed in Section 2. After estimating factors, the alternative methods outlined in the previous sections are used to form forecasting models and predictions. Summarizing, we consider four "specification types":

*Specification Type* 1 *without lags* (*SP*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* 1 *with lags* (*SP*1*L*): This is the same as SP1, except that lags are included in the datasets used to construct principal components.

*Specification Type* 2 (*SP*2): Principal component models of the type given in (3) are constructed using subsets of variables from the large scale 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 in $X$.

*Specification Type* 3 (*SP*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 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 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. Various out-of-sample periods are examined, including periods from 1992:10–2009:5 (i.e., $P = 200$), 1984:6–2009:5 (i.e., $P = 300$), and 1972:2–2009:5 (i.e., $P = 400$), where $P$ denotes the length of the out-of-sample period, and in subsequent discussion, $R$ is the in-sample estimation period. In-sample estimation periods are adjusted so that $P$ remains fixed, regardless of forecast horizon. 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, and 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 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:

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

where $\widehat{Y}_{i,t+h}$ is the forecast at horizon $h$ for the $i$-th model. Forecast accuracy is evaluated using point MSFEs as well as the so-called DM predictive accuracy test of Diebold and Mariano (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 $H_0 : E\left[l\left(\varepsilon^1_{t+h|t}\right)\right] - E\left[l\left(\varepsilon^2_{t+h|t}\right)\right] = 0$, where $\varepsilon^i_{t+h|t}$ is $i$-th model's prediction error and $l(\cdot)$ is the quadratic loss function. The actual statistic in this case is constructed as: $\text{DM} = P^{-1}\sum_{i=1}^{P} d_t/\hat{\sigma}_{\bar{d}}$, where $d_t = \left(\widehat{\varepsilon^1_{t+h|t}}\right)^2 - \left(\widehat{\varepsilon^2_{t+h|t}}\right)^2$, $\bar{d}$ is the mean of $d_t$, $\hat{\sigma}_{\bar{d}}$ is a heteroskedasticity and autocorrelation robust estimator of the standard deviation of $\bar{d}$, and $\widehat{\varepsilon^1_{t+h|t}}$ and $\widehat{\varepsilon^2_{t+h|t}}$ are estimates of the true prediction errors $\varepsilon^1_{t+h|t}$ and $\varepsilon^2_{t+h|t}$. Thus, if the statistic is negative and significantly different from zero, then Model 1 is preferred over Model 2. Related reality check tests due to White (2000) are also constructed (see below for further discussion).

In addition to 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 $Y^{\text{ARX}}_{t+h} = \hat{\alpha} + \hat{\beta}Z_t + \hat{\phi}(L)Y_t$, 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.

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

*Factor augmented autoregression*: Based on Eq. (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 (2012), 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 $\hat{Y}^{\text{ADL}}_{t+h} = \hat{\alpha} + \hat{\beta}_i(L)X_{i,t} + \hat{\phi}_i(L)Y_t$. The combined forecast is $\hat{Y}^{\text{Comb},h}_{T+h|T} = \sum_{t=1}^{n} w_i\hat{Y}^{\text{ADL},h}_{T+h|T}$. Here, we set $w_i = 1/N$. 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 (2006); and in the literature on factor models, Stock and Watson (2004, 2006, 2012), 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 forecasting models summarized in Table 2.

**Table 2**
Models and methods used in real-time forecasting experiments.[a]

| Method | Description |
|---|---|
| AR(SIC) | Autoregressive model with lags selected by the SIC |
| ARX | Autoregressive model with exogenous regressors |
| CADL | Combined autoregressive distributed lag model |
| FAAR | Factor augmented autoregressive model |
| PCR | Principal components regression |
| Bagging | Bagging with shrinkage, $c = 1.96$ |
| Boosting | Component boosting, $M = 50$ |
| BMA1 | Bayesian model averaging with $g$-prior $= 1/T$ |
| BMA2 | Bayesian model averaging with $g$-prior $= 1/N^2$ |
| Ridge | Ridge regression |
| LAR | Least angle regression |
| EN | Elastic net |
| NNG | Non-negative garotte |
| Mean | Arithmetic mean |

[a] Notes: This table summarizes the model specification methods used in the construction of prediction models. In addition to directly estimating the above pure linear and factor models (i.e., AR, ARX, CADL, FAAR, PCR), three different combined factor and shrinkage type prediction "specification methods" are used in our forecasting experiments, including: Specification Type 1—Principal components are first constructed, and then prediction models are formed using the above shrinkage methods (including Bagging, Boosting, Ridge, LAR, EN, and NNG) to select functions of and weights for the factors to be used in our prediction models. 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 (including Bagging, Boosting, Ridge, LAR, EN, and NNG), without use of factor analysis at any stage. See Sections 3 and 4 for complete details.

## 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, detailed results are gathered in Tables 3–6. Analogous results based on rolling estimation are omitted for the sake of brevity, although they are available upon request from the authors. Summary results are contained in Tables 7–11.

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). Results in these tables are for the out-of-sample period 1984:6–2009:5 ($P = 300$). Other out-of-sample results ($P = 200$ and 400) are available upon request from the authors. Panels A–C in these tables contain results for $h = 1$, 3 and 12 month ahead predictions, respectively. In each panel, the first row of entries reports MSFEs for benchmark AR(SIC) models, and all other rows report MSFEs relative to those of the AR(SIC) model. 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 principal component regression (PCR), with a relative MSFE of 0.974. 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 using a single star (denoting rejection at the 10% level), and a double star (denoting rejection at the 5% level).

There are no models that uniformly yield the lowest MSFEs, across both forecast horizon and variable. However, various models perform quite well, including in particular our benchmark 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. Various other results are also apparent, upon inspection of tables.

**Table 3**
Relative mean square forecast errors of Specification Type 1 without lags.[a]

| Method | UR | PI | TB | CPI | PPI | NPE | HS | IPX | M2 | SNP | GDP |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Panel A: Recursive, $h = 1$ | | | | | | | | | | | |
| AR(SIC) | 6.257 | 0.007 | 18.650 | 0.002 | 0.010 | 0.000 | 1.316 | 0.010 | 0.003 | **0.400** | 0.006 |
| ARX(SIC) | 1.011 | 0.941 | 1.136 | 0.966 | 0.992 | 1.347** | 1.088 | 0.913 | 1.181 | 1.135 | **0.836** |
| CADL | 0.976** | 1.031 | 0.991 | 1.026 | 1.018 | **0.920*** | 0.988 | 0.975 | 1.060 | 1.009 | 1.075 |
| FAAR | **0.889** | 0.907 | 1.029 | 0.897 | 0.941 | 1.014 | 1.054 | 0.865* | **0.953** | 1.046 | 0.964 |
| PCR | 0.935 | **0.856** | 1.046 | **0.858** | **0.906** | 1.276** | 2.084** | 0.864* | 1.472** | 1.046 | 0.875 |
| Bagging | 0.917 | 1.039 | 0.962 | 1.129 | 1.039 | 1.373** | 1.147 | 1.126 | 0.963 | 1.010 | 1.001 |
| Boosting | 0.972 | 0.994 | 0.929 | 0.961 | 0.990 | 0.965 | 1.003 | 0.882** | 0.997 | 1.008 | 1.032 |
| BMA1 | 0.974 | 0.984 | 0.941 | 0.961 | 0.994 | 0.974 | 1.019 | 0.875** | 1.014 | 1.027 | 1.047 |
| BMA2 | 0.980 | 0.990 | 0.952 | 0.955 | 0.991 | 1.002 | 1.012 | 0.870** | 1.009 | 1.028 | 1.040 |
| Ridge | 0.984 | 0.990 | 0.959 | 0.969 | 0.988 | 1.052 | 1.000 | **0.864**** | 1.003 | 1.035 | 1.013 |
| LAR | 0.963* | 0.984 | 0.957** | 0.991 | 0.991 | 0.970** | 1.003 | 0.950** | 0.987 | 1.005 | 1.021 |
| EN | 0.963* | 0.984 | 0.957** | 0.991 | 0.991 | 0.970** | 1.003 | 0.950** | 0.987 | 1.005 | 1.021 |
| NNG | 0.987* | 0.992 | 0.995 | 0.991 | 0.993 | 0.978** | 0.999 | 0.986** | 0.995 | 1.003 | 1.005 |
| Mean | 0.924** | 0.952* | **0.924** | 0.927 | 0.953 | 0.939 | **0.974** | 0.884** | 0.969 | 1.006 | 0.939** |
| Panel B: Recursive, $h = 3$ | | | | | | | | | | | |
| AR(SIC) | 6.288 | 0.006 | 22.471 | 0.002 | 0.011 | 0.000 | 2.540 | 0.010 | 0.003 | 0.432 | 0.005 |
| ARX(SIC) | 1.088 | **0.953** | 1.039 | 0.976 | 0.945 | 1.149 | 1.144 | 1.030 | 1.028 | 1.042 | 1.135** |
| CADL | 0.994 | 0.999 | 0.986** | 1.049 | 1.038 | **0.873**** | 0.979* | 0.960 | 1.066 | 1.014 | 1.019 |
| FAAR | 1.013 | 0.975 | 1.024 | 0.985 | 0.962 | 0.998 | 1.174 | 1.014 | 1.019 | 1.069** | 1.193** |
| PCR | **0.975** | 0.964 | 0.992 | 0.989 | **0.931** | 1.107 | 1.757** | 1.013 | 1.295** | 1.069 | 1.178** |
| Bagging | 1.129** | 1.144** | **0.981** | 1.121 | 0.987 | 1.640** | **0.905** | 0.982 | 1.009 | **0.991** | 1.056** |
| Boosting | 1.020 | 1.000 | 1.009 | 0.992 | 0.989 | 1.069 | 1.039 | 1.015 | 1.010 | 1.014 | 1.042 |
| BMA1 | 1.062 | 1.001 | 1.012 | 1.012 | 0.991 | 1.101 | 1.039 | 1.047 | 1.017 | 1.013 | 1.074 |
| BMA2 | 1.034 | 0.987 | 1.020 | 1.009 | 0.994 | 1.097 | 1.056 | 1.023 | 1.015 | 1.021 | 1.080 |
| Ridge | 1.011 | 0.965 | 1.013 | 0.994 | 1.003 | 1.086 | 1.100 | 0.986 | 1.005 | 1.032 | 1.099** |
| LAR | 1.035 | 1.006 | 1.012 | 0.998 | 0.996 | 1.054 | 1.033 | 1.043 | 1.002 | 1.005 | 1.012 |
| EN | 1.035 | 1.006 | 1.012 | 0.998 | 0.996 | 1.054 | 1.033 | 1.043 | 1.002 | 1.005 | 1.012 |
| NNG | 1.004 | 1.002 | 1.014 | 0.997 | 0.999 | 1.014 | 1.000 | 1.002 | 0.998 | 1.013 | **0.998** |
| Mean | 1.003 | 0.967 | 0.993 | **0.963** | 0.956 | 0.992 | 0.990 | **0.951** | **0.988** | 1.012 | 1.031 |
| Panel C: Recursive, $h = 12$ | | | | | | | | | | | |
| AR(SIC) | 7.579 | **0.007** | 21.740 | 0.002 | 0.011 | 0.001 | 12.047 | 0.012 | 0.004 | 0.437 | 0.005 |
| ARX(SIC) | 1.026 | 1.022 | 1.043 | 0.984 | 1.008 | 0.992 | 0.933 | 1.037 | **0.927** | 1.030 | 1.023 |
| CADL | 0.985** | 1.012 | 0.983** | 0.995 | 1.029 | **0.786**** | 0.964** | **0.973** | 0.974 | 1.016 | 1.061 |
| FAAR | 0.982 | 1.079** | 0.992 | 0.976 | 1.000 | 1.071 | 0.950 | 1.100** | 0.976 | 1.017 | 1.039 |
| PCR | **0.974** | 1.052 | **0.973** | 1.075 | 1.028 | 1.079 | 1.220** | 1.093** | 1.110** | 1.008 | 1.013 |
| Bagging | 1.051 | 1.022 | 1.015 | 0.973 | 1.010 | 1.366** | **0.843*** | 1.016 | 0.983 | 1.024 | 1.010 |
| Boosting | 0.996 | 1.046 | 0.999 | 0.963 | 0.988 | 1.095** | 1.007 | 1.051 | 1.011 | 1.004 | 1.002 |
| BMA1 | 1.004 | 1.058** | 0.996 | 0.968 | 0.992 | 1.126** | 1.032 | 1.065 | 1.018 | 1.000 | 1.003 |
| BMA2 | 0.997 | 1.060** | 0.993 | 0.969 | 0.992 | 1.120** | 1.033 | 1.066 | 1.014 | 0.999 | 1.004 |
| Ridge | 0.975 | 1.058** | 0.985 | 0.966 | 0.990 | 1.079 | 1.039 | 1.070 | 1.001 | 1.003 | 1.018 |
| LAR | 0.990 | 1.020 | 0.997 | 0.999 | 0.998 | 1.064** | 0.964 | 1.020 | 1.000 | 1.001 | 0.999 |
| EN | 0.990 | 1.020 | 0.997 | 0.999 | 0.998 | 1.064** | 0.964 | 1.020 | 1.000 | 1.001 | 1.000 |
| NNG | 0.990 | 1.004 | 0.995 | 0.999 | 0.998 | 1.006 | 0.993** | 0.998 | 0.999 | 1.001 | 1.001 |
| Mean | 0.977 | 1.008 | 0.986 | **0.954** | **0.980** | 0.996 | 0.933** | 1.004 | 0.957** | **0.998** | **0.992** |

[a] 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 1984:6–2009:5 ($P = 300$). Models and target variables are given 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. 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 using a * and **. See Sections 4 and 5 for complete details.

* Denoting rejection at the 10% level.

** Denoting rejection at the 5% level.

For example, turning to Table 3, which summarizes results for Specification Type 1 without lags (SP1), notice that in Panel A, AR(SIC) and ARX(SIC) models "win" for only 2 of 11 variables (SNP and GDP). A similar results holds for $h = 3$ and $h = 12$ (see Panels B and C). Notice also that model averaging (BMA and "Mean") wins only 27% of the time, across all variable/horizon permutations.

In Table 4, we report the results for Specification Type 1 with lags (SP1L). In Panel A ($h = 1$), benchmark models including AR(SIC), ARX(SIC), and CADL yields lowest point MSFEs for 7 of 11 variables. This suggests that including lags (of factors) does not appreciably improve model forecast performance, given that the structure of these benchmark models do not change when lags are added to our factor specification methods, and that they "win" less frequently under SP1 than under SP1L.[7] Indeed, 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 (elsewhere called our "hybrid" model) employed when constructing forecasts. Note however, that there are (possibly) *some* lagged variables used in the construction of diffusion indexes

[7] Recall that the addition of lags when moving from SP1 to SP1L only involves including lags of explanatory variables used in diffusion index construction. As AR, ARX, and CADL models do not include diffusion indexes, they are thus unchanged when constructed under either SP1 or SP1L.

**Table 4**
Relative mean square forecast errors of Specification Type 1 with lags.[a]

| Method | UR | PI | TB | CPI | PPI | NPE | HS | IPX | M2 | SNP | GDP |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Panel A: Recursive, $h = 1$ | | | | | | | | | | | |
| AR(SIC) | 6.257 | 0.007 | 18.650 | 0.002 | 0.010 | 0.000 | 1.316 | 0.010 | **0.003** | **0.400** | 0.006 |
| ARX(SIC) | 1.011 | 0.941 | 1.136 | **0.966** | 0.992 | 1.347** | 1.088 | **0.913** | 1.181 | 1.135 | 0.836 |
| CADL | 0.976** | 1.031 | **0.991** | 1.026 | 1.018 | **0.920*** | **0.988** | 0.975 | 1.060 | 1.009 | 1.075 |
| FAAR | 1.069 | 0.882 | 1.851** | 1.392 | 1.121 | 1.985** | 3.631** | 1.209 | 1.454 | 1.390** | 0.799 |
| PCR | 1.013 | 0.881 | 1.676** | 1.359 | 1.120 | 1.837** | 4.869** | 1.163 | 1.700** | 1.371 | 0.889 |
| Bagging | 1.310** | 0.984 | 1.762** | 1.477** | 1.120 | 3.494** | 3.905** | 1.225 | 1.229 | 1.131 | 0.804 |
| Boosting | 1.038 | 0.897 | 1.465** | 1.310 | 1.101 | 1.758** | 3.714** | 1.104 | 1.419 | 1.224** | 0.795 |
| BMA1 | 1.042 | 0.890 | 1.544** | 1.328 | 1.097 | 1.758** | 3.711** | 1.124 | 1.446 | 1.253** | 0.802 |
| BMA2 | 1.028 | 0.888 | 1.538** | 1.344 | 1.103 | 1.762** | 3.720** | 1.111 | 1.430 | 1.242** | 0.805 |
| Ridge | 1.033 | 0.891 | 1.636** | 1.272 | 1.086 | 1.735** | 3.664** | 1.125 | 1.478** | 1.270** | 0.784 |
| LAR | 1.068 | 0.901 | 1.395** | 1.276 | 1.094 | 1.782** | 3.823** | 1.072 | 1.360 | 1.210** | 0.791 |
| EN | 1.069 | 0.902 | 1.396** | 1.275 | 1.094 | 1.777** | 3.825** | 1.073 | 1.360 | 1.210** | 0.788 |
| NNG | 1.070 | 0.907 | 1.408** | 1.275 | 1.097 | 1.779** | 3.844** | 1.072 | 1.362 | 1.207** | 0.784 |
| Mean | **0.966** | **0.876** | 1.259** | 1.046 | **0.973** | 1.441** | 2.681** | 0.986 | 1.143 | 1.150** | **0.745**** |
| Panel B: Recursive, $h = 3$ | | | | | | | | | | | |
| AR(SIC) | 6.288 | 0.006 | 22.471 | 0.002 | 0.011 | 0.000 | 2.540 | 0.010 | 0.003 | **0.432** | **0.005** |
| ARX(SIC) | 1.088 | **0.953** | 1.039 | 0.976 | 0.945 | 1.149 | 1.144 | 1.030 | 1.028 | 1.042 | 1.135** |
| CADL | **0.994** | 0.999 | 0.986** | 1.049 | 1.038 | **0.873**** | **0.979*** | **0.960** | 1.066 | 1.014 | 1.019 |
| FAAR | 1.024 | 0.986 | 1.175** | 0.806** | 0.900 | 1.543** | 2.490** | 1.228** | 1.170 | 1.115 | 1.170** |
| PCR | 1.021 | 0.974 | 1.151 | 0.822* | 0.906 | 1.456** | 3.539** | 1.201** | 1.444** | 1.088 | 1.126 |
| Bagging | 1.122 | 1.061** | 1.102 | 1.041 | 1.034 | 2.088** | 2.465** | 1.146 | 0.989 | 1.031 | 1.101 |
| Boosting | 1.064 | 1.024 | 1.001 | 0.803** | 0.907 | 1.572** | 2.638** | 1.148** | 1.065 | 1.023 | 1.131 |
| BMA1 | 1.053 | 1.032 | 1.009 | 0.788** | 0.894 | 1.565** | 2.641** | 1.161** | 1.097 | 1.032 | 1.141 |
| BMA2 | 1.059 | 1.028 | 0.990 | 0.790** | 0.900 | 1.570** | 2.647** | 1.173** | 1.091 | 1.030 | 1.142 |
| Ridge | 1.035 | 1.010 | 1.094 | 0.788** | 0.886 | 1.518** | 2.617** | 1.154 | 1.121 | 1.081 | 1.127 |
| LAR | 1.078 | 1.036 | 0.979 | 0.857* | 0.907 | 1.610** | 2.690** | 1.151** | 1.007 | 1.024 | 1.104 |
| EN | 1.081 | 1.036 | 0.979 | 0.858* | 0.906 | 1.598** | 2.691** | 1.137 | 1.006 | 1.025 | 1.099 |
| NNG | 1.075 | 1.035 | 0.973 | 0.857* | 0.913 | 1.580** | 2.687** | 1.114 | 1.003 | 1.019 | 1.093 |
| Mean | 0.998 | 0.974 | **0.969** | **0.786**** | **0.871**** | 1.312** | 1.984** | 1.045 | **0.983** | 1.020 | 1.046 |
| Panel C: Recursive, $h = 12$ | | | | | | | | | | | |
| AR(SIC) | 7.579 | 0.007 | 21.740 | 0.002 | 0.011 | 0.001 | 12.047 | 0.012 | 0.004 | **0.437** | 0.005 |
| ARX(SIC) | 1.026 | 1.022 | 1.043 | 0.984 | 1.008 | 0.992 | **0.933** | 1.037 | **0.927** | 1.030 | 1.023 |
| CADL | 0.985** | 1.012 | **0.983**** | 0.995 | 1.029 | **0.786**** | 0.964** | **0.973** | 0.974 | 1.016 | 1.061 |
| FAAR | 1.155 | 1.083 | 1.148 | 0.956 | 1.025 | 1.115 | 1.135 | 1.204** | 1.049 | 1.117 | 1.093 |
| PCR | 1.170** | 1.065 | 1.163 | 0.973 | 1.030 | 1.128 | 1.410** | 1.190** | 1.177** | 1.115** | 1.089 |
| Bagging | 1.089 | 1.025 | 1.105 | 1.001 | 0.986 | 1.582** | 1.413** | 1.094** | 0.968 | 1.072 | 0.997 |
| Boosting | **0.963** | 1.033 | 1.020 | 0.926 | 1.012 | 1.070 | 1.210** | 1.063 | 0.992 | 1.011 | 0.987 |
| BMA1 | 0.981 | 1.037 | 1.015 | 0.927 | 1.012 | 1.075 | 1.229** | 1.085** | 0.996 | 1.011 | 0.998 |
| BMA2 | 0.979 | 1.034 | 1.021 | 0.924 | 1.012 | 1.067 | 1.231** | 1.088** | 0.986 | 1.011 | 0.992 |
| Ridge | 1.057 | 1.038 | 1.090 | 0.925 | 1.009 | 1.055 | 1.205** | 1.122** | 1.037 | 1.064 | 1.043 |
| LAR | 0.981 | 1.013 | 1.017 | 0.933 | 0.988 | 1.060** | 1.196** | 1.011 | 0.976 | 1.006 | 0.984 |
| EN | 0.982 | 1.012 | 1.016 | 0.934 | 0.987 | 1.052** | 1.199** | 1.009 | 0.978 | 1.005 | 0.984 |
| NNG | 0.979 | 1.021 | 1.010 | 0.930 | 0.988 | 1.044** | 1.216** | 1.003 | 1.011 | 1.003 | 0.985 |
| Mean | 0.971 | **0.991** | 1.007 | **0.885**** | **0.963** | 0.977 | 1.052 | 1.013 | 0.940* | 1.011 | **0.981** |

[a] Notes: See notes to Table 3. Note that numerical entries for AR(SIC), ARX(SIC) and CADL models are identical to those given for Specification Type 1 without lags (see Table 1), since those benchmark models do not involve factors or shrinkage. See Section 5 for further details.

under *all* specification types, including SP1 and SP1L. In particular, all datasets used to construct factors may contain lagged values of the *target* variable. An exception to the above finding concerning the usefulness of SP1L as a specification method is as follows. Under SP1L, all models (except CADL) yield lower MSFEs than AR(SIC), when used to forecast GDP, for $h = 1$. However, under SP1, few models yield lower MSFEs than AR(SIC). This result may be due to the interpolation approach used to construct our monthly GDP figures.

For Specification Type 2 (SP2), results are reported in Table 5. In this table, FAAR and PCR models are omitted since all diffusion indexes used in the prediction models involve using variables first selected via shrinkage, and our benchmark "pure" factor models use no shrinkage. In Specification Type 3 (SP3), FAAR and PCR are also omitted (since no SP3 specifications use diffusion indices). In Table 5, our "hybrid" SP2 type models (that use bagging, boosting, ridge, LAR, EN, and NNG) are MSFE "best" for 3, 3, and 2 out of 11 variables, for $h = 1, 3$ and 12, respectively. On the other hand, linear benchmark models (AR, ARX and CADL) yield the lowest MSFEs for 6, 5, 4 out of 11 variables, for $h = 1, 3$ and 12, respectively. Thus, SP2 type factor models "win" roughly 1/3 of the time. In Table 6, our "pure shrinkage" (SP3) type models that use bagging, boosting, ridge, LAR, EN, or NNG are MSFE "best" for 5, 5, 4 out of 11 target variables, for $h = 1, 3$ and 12, respectively. Of note is that bagging performs very poorly in SP3, so that "Mean" does not yield reasonable MSFEs in most cases. We also constructed "Mean" predictions without bagging, and although results improved, our overall model rankings did not change.

Entries in Tables 3–6 only summarize a subset of our experiments results, and are hence not as informative as we would like them to be. For this reason, we also provide a series of "summary" tables, denoted Tables 7–11, from which a number of further interesting findings can be drawn.

Table 7 summarizes MSFE "best" forecast models *by* specification type (Panels A–D), as well as *across* all specification types (Panel E). Entries in the Panels A–D correspond to entries in bold in Tables 3–6, respectively. For example, when forecasting unemployment (UR), FAAR under Specification Type 1 without lags yields the lowest MSFE, for $h = 1$. When comparing results for individual specification types, note that benchmark AR and ARX models sometimes "win". However, when the MSFE "best" model across all four specifications is selected, these benchmark models

**Table 5**
Relative mean square forecast errors of Specification Type 2.[a]

| Method | UR | PI | TB | CPI | PPI | NPE | HS | IPX | M2 | SNP | GDP |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Panel A: Recursive, $h = 1$ | | | | | | | | | | | |
| AR(SIC) | 6.257 | 0.007 | 18.650 | 0.002 | 0.010 | 0.000 | 1.316 | 0.010 | 0.003 | **0.400** | 0.006 |
| ARX(SIC) | 1.011 | 0.941 | 1.136 | **0.966** | 0.992 | 1.347** | 1.088 | **0.913** | 1.181 | 1.135 | **0.836** |
| CADL | 0.976** | 1.031 | **0.991** | 1.026 | 1.018 | 0.920* | **0.988** | 0.975 | 1.060 | 1.009 | 1.075 |
| Bagging | 1.492** | 1.194 | 1.282** | 1.266 | **0.952** | 1.823** | 9.795** | 1.316** | 1.655 | 1.081 | 1.089 |
| Boosting | 1.465** | 0.940 | 1.545** | 1.606 | 1.206 | 1.196 | 1.282 | 1.287 | 1.157 | 1.119 | 1.257 |
| BMA1 | 1.544** | 1.012 | 1.595** | 1.707 | 1.057 | 2.558** | 1.337** | 1.529** | 1.246 | 1.049 | 1.793** |
| BMA2 | 1.021 | 0.934 | 1.603** | 1.416 | 1.444 | 1.219 | 1.229 | 1.570** | 1.237 | 1.638 | 1.093 |
| Ridge | 1.089 | 0.996 | 1.654** | 1.447 | 1.477 | 1.329** | 1.206 | 1.363** | 1.157 | 1.734 | 1.077 |
| LAR | **0.891**** | 0.890 | 1.388** | 1.402** | 1.119 | 1.173** | 1.368** | 0.961 | 1.108 | 1.083 | 1.332** |
| EN | 0.893** | **0.885** | 1.441** | 1.402** | 1.100 | 1.172** | 1.367** | 0.961 | 1.115 | 1.123 | 1.289** |
| NNG | 0.959 | 0.925 | 1.394** | 1.404** | 1.118 | 1.114** | 1.383** | 0.943 | 1.142 | 1.976 | 1.083 |
| Mean | 0.965 | 0.917 | 1.131 | 1.142 | 0.973 | **0.864**** | 1.229 | 1.008 | **0.996** | 1.021 | 1.000 |
| Panel B: Recursive, $h = 3$ | | | | | | | | | | | |
| AR(SIC) | 6.288 | 0.006 | 22.471 | 0.002 | 0.011 | 0.000 | 2.540 | 0.010 | **0.003** | **0.432** | **0.005** |
| ARX(SIC) | 1.088 | 0.953 | 1.039 | 0.976 | 0.945 | 1.149 | 1.144 | 1.030 | 1.028 | 1.042 | 1.135** |
| CADL | 0.994 | 0.999 | 0.986** | 1.049 | 1.038 | **0.873**** | **0.979*** | 0.960 | 1.066 | 1.014 | 1.019 |
| Bagging | 1.056 | 1.016 | 1.042 | 1.001 | 0.960 | 1.312** | 2.267** | 1.124 | 1.369** | 1.086 | 1.164** |
| Boosting | 0.941 | 0.962 | 0.998 | 1.147 | 1.050 | 1.005 | 1.172 | 1.084 | 1.157 | 1.019 | 1.153** |
| BMA1 | 0.992 | 0.958 | 0.989 | 1.090 | 1.139 | 1.151 | 1.178 | 0.971 | 1.144 | 1.010 | 1.168 |
| BMA2 | **0.937** | 0.974 | 1.014 | 1.021 | 0.970 | 0.908 | 1.312 | 1.013 | 1.199 | 1.051 | 1.223** |
| Ridge | 0.943 | 0.975 | 0.993 | 1.086 | 0.978 | 0.961 | 1.276** | 1.053 | 1.183 | 1.055 | 1.216** |
| LAR | 0.980 | 1.016 | 1.009 | 0.938 | 0.934 | 1.179** | 1.371** | 0.958 | 1.121 | 1.029 | 1.175** |
| EN | 0.982 | 1.004 | 0.998 | 0.937 | **0.934** | 1.201** | 1.379** | 0.985 | 1.118 | 1.026 | 1.147 |
| NNG | 1.001 | 0.990 | 1.004 | **0.936** | 0.953 | 1.107** | 1.356** | 1.054 | 1.174 | 1.019 | 1.111 |
| Mean | 0.949 | **0.952*** | **0.970** | 0.941 | 0.936* | 0.949 | 1.106 | **0.936** | 1.060 | 1.011 | 1.066 |
| Panel C: Recursive, $h = 12$ | | | | | | | | | | | |
| AR(SIC) | 7.579 | 0.007 | 21.740 | 0.002 | **0.011** | 0.001 | 12.047 | 0.012 | 0.004 | 0.437 | 0.005 |
| ARX(SIC) | 1.026 | 1.022 | 1.043 | **0.984** | 1.008 | 0.992 | 0.933 | 1.037 | 0.927 | 1.030 | 1.023 |
| CADL | 0.985** | 1.012 | **0.983**** | 0.995 | 1.029 | **0.786**** | 0.964** | 0.973 | 0.974 | 1.016 | 1.061 |
| Bagging | 0.971 | 1.051 | 1.021 | 2.434** | 1.503** | 1.004 | 1.193** | 1.030 | 1.151 | 1.010 | 1.049 |
| Boosting | 0.968 | 0.999 | 1.025 | 1.039 | 1.061 | 1.059 | 0.952 | 1.115** | 1.006 | 1.002 | 0.994 |
| BMA1 | 1.065 | 0.986 | 0.999 | 1.050 | 1.095 | 1.009 | 1.002 | 1.009 | 0.919 | **0.981** | 0.984 |
| BMA2 | 0.980 | 1.025 | 1.027 | 1.079 | 1.092** | 1.021 | 1.013 | 1.046 | 1.055 | 1.007 | 1.026 |
| Ridge | 0.981 | 1.017 | 1.017 | 1.064 | 1.105** | 0.951 | 0.992 | 1.036 | 1.041 | 1.009 | 1.019 |
| LAR | 0.984 | 0.997 | 1.011 | 0.995 | 1.007 | 1.083 | 1.005 | **0.972** | 0.974 | 1.003 | 0.999 |
| EN | 0.985 | 1.000 | 1.010 | 1.023 | 1.006 | 1.083 | 1.002 | 1.014 | 0.977 | 1.003 | 1.000 |
| NNG | 0.992 | 1.002 | 1.004 | 1.005 | 1.024 | 1.011 | 1.036 | 0.993 | 1.004 | 1.004 | 0.988 |
| Mean | **0.950** | **0.976** | 0.994 | 0.987 | 1.014 | 0.900** | **0.930**** | 0.974 | **0.914**** | 0.996 | **0.970** |

[a] Notes: See notes to Tables 3–4. Note that FAAR and PCR model results are not included in this table, since these models are not constructed under Specification Type 2.

never win (see Panel E). This is interesting, as it suggests that the oft noted dominance of simple AR type models in forecasting experiments may, to some extent, be due to a lack of suitable alternative "model/specification method" combinations against which to compare the AR model. Turning again to Panel E, note that, for $h = 1$, a model that incorporates either shrinkage (including bagging, boosting, ridge, LAR, EN, NNG, and "Mean"),[8] or is a "hybrid" model, "wins" for 6 of 11 variables (models not in this subset include FAAR, PCR, and CADL). Notice, also, that this number increases from 6 out of 11 to 10 out of 11 ($h = 3$) and 9 out of 11 ($h = 12$), when longer forecast horizons are used. This is strong evidence in favor of using shrinkage and "hybrid" modeling approaches. These summary finding remain qualitatively the same for out-of-sample periods not reported here (i.e., $P = 200$ and 400).

Table 8 (Panel A), summarizes the number of MSFE "wins" for each model, across *all* specification types, for two cases. In the first case ("Recursive Estimation Window"), results are reported only for recursively estimated models. In the second case ("Recursive and Rolling Estimation Windows"), results are summarized *across* both "windowing" methods. In Panel B, the number of "wins" are reported for the two windowing approaches ("Winners by Estimation Window") and by specification type ("Winners by Specification Type"). From Panel A, we see that our findings based on results reported in Table 7, which are based only on recursive estimation, remain essentially the same when rolling estimation is also considered (i.e., compare the first 4 columns of entries in the table with the second 4 columns of entries). This is not surprising, given that inspection of Panel B of the table indicates that recursive estimation often "dominates" rolling estimation. However, it is noteworthy that recursive estimation do not *always* lead to MSFE "best" models, particularly for $h = 12$.

In Table 8, notice also that SP1 clearly dominates for $h = 1$, while SP2 dominates for $h = 12$. This suggests that longer horizon forecasts, which are generally know to be more difficult to accurately construct, are best tackled using more parsimonious model specification methods (i.e., pre-select the variables to use in diffusion index construction via shrinkage methods, as is done under SP2). Also, note that as the forecast horizon increases, model averaging methods "win" more frequently; which is, again, not surprising (see Panel A).

In order to shed light on the importance of sub-samples in our empirical analysis, we report summary results for a variety of NBER business-cycle related sample periods in Table 9, for $h = 1$. Although conclusions based upon inspection of this table are largely in accord with those reported above, two additional noteworthy findings are worth stressing. First, in Panel A of the table, note that

[8] BMA and simple arithmetic model averaging (collectively called "Mean") are included in this subset of models because these model averaging methods involve the combination of various shrinkage type models.

**Table 6**
Relative mean square forecast errors of Specification Type 3.[a]

| Method | UR | PI | TB | CPI | PPI | NPE | HS | IPX | M2 | SNP | GDP |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Panel A: Recursive, $h = 1$ | | | | | | | | | | | |
| AR(SIC) | 6.257 | 0.007 | 18.650 | 0.002 | 0.010 | 0.000 | 1.316 | 0.010 | 0.003 | 0.400 | 0.006 |
| ARX(SIC) | 1.011 | **0.941** | 1.136 | 0.966 | 0.992 | 1.347** | 1.088 | 0.913 | 1.181 | 1.135 | 0.836 |
| CADL | **0.976**** | 1.031 | 0.991 | 1.026 | 1.018 | **0.920*** | **0.988** | 0.975 | 1.060 | 1.009 | 1.075 |
| Bagging | 6.843** | 1.142 | 5.235** | 5.337 | 2.120 | 13.78** | 10.14** | 3.521** | 4.217 | 11.14 | 3.637 |
| Boosting | 1.010** | 0.985 | **0.991**** | 0.947 | 0.983 | 1.143 | 1.011 | **0.893** | 1.118 | 1.024 | **0.826** |
| BMA1 | 1.137** | 1.010 | 1.062** | 0.989 | 1.069 | 2.150** | 1.409** | 0.966** | 1.080 | 1.162 | 0.842** |
| BMA2 | 0.980 | 0.998 | 1.081** | 0.996 | 1.027 | 1.440 | 1.301 | 0.905** | 1.115 | 1.121 | 0.834 |
| Ridge | 1.564 | 1.274 | 1.354** | 1.051 | 1.186 | 2.258** | 1.231 | 1.322** | 1.334 | 1.417 | 1.079 |
| LAR | 0.990** | 0.998 | 1.000** | 1.017** | 1.004 | 0.981** | 1.003** | 1.000 | **0.992** | **0.998** | 0.847** |
| EN | 0.990** | 0.996 | 1.000** | 1.016** | 1.004 | 0.982** | 1.001** | 1.000 | 0.993 | 0.998 | 0.850** |
| NNG | 0.996 | 0.998 | 0.998** | 0.999** | 1.000 | 0.989** | 1.003** | 0.995 | 1.001 | 1.001 | 1.000 |
| Mean | 1.012 | 0.954 | 1.365 | **0.928** | **0.937** | 1.822** | 1.013 | 0.928 | 1.001 | 1.102 | 0.865 |
| Panel B: Recursive, $h = 3$ | | | | | | | | | | | |
| AR(SIC) | 6.288 | 0.006 | 22.471 | 0.002 | 0.011 | 0.000 | 2.540 | 0.010 | 0.003 | 0.432 | 0.005 |
| ARX(SIC) | 1.088 | 0.953 | 1.039 | **0.976** | **0.945** | 1.149 | 1.144 | 1.030 | 1.028 | 1.042 | 1.135** |
| CADL | **0.994** | 0.999 | **0.986**** | 1.049 | 1.038 | **0.873**** | 0.979* | 0.960 | 1.066 | 1.014 | 1.019 |
| Bagging | 13.34 | 2.042 | 5.843 | 7.926 | 3.300 | 19.95** | 12.30** | 3.530 | 8.547** | 13.84 | 19.75** |
| Boosting | 1.012 | **0.939** | 1.023 | 0.991 | 1.061 | 0.966 | **0.934** | 0.983 | 1.034 | 1.008 | 1.033** |
| BMA1 | 1.033 | 0.956 | 1.258 | 1.125 | 1.098 | 1.072 | 1.301 | 1.014 | 1.074 | 1.088 | 1.116 |
| BMA2 | 1.035 | 0.968 | 1.027 | 1.062 | 1.051 | 1.030 | 1.219 | 1.044 | 0.991 | 1.050 | 1.090** |
| Ridge | 1.471 | 1.239 | 1.974 | 1.261 | 1.350 | 1.647 | 1.148** | 1.460 | 1.148 | 1.261 | 1.662** |
| LAR | 0.996 | 0.988 | 1.002 | 1.006 | 1.005 | 0.981** | 0.994** | 0.962 | 0.986 | 0.999 | 1.002** |
| EN | 0.994 | 0.988 | 1.002 | 1.006 | 1.005 | 0.982** | 0.995** | **0.955** | 0.990 | **0.999** | **0.999** |
| NNG | 0.999 | 0.998 | 1.000 | 0.999 | 0.999 | 1.001** | 1.006** | 1.000 | 1.001 | 1.003 | 1.002 |
| Mean | 1.130 | 0.968* | 1.536 | 1.035 | 1.010* | 2.469 | 1.012 | 1.208 | **0.936** | 1.059 | 1.202 |
| Panel C: Recursive, $h = 12$ | | | | | | | | | | | |
| AR(SIC) | 7.579 | 0.007 | 21.740 | 0.002 | 0.011 | 0.001 | 12.047 | 0.012 | 0.004 | 0.437 | 0.005 |
| ARX(SIC) | 1.026 | 1.022 | 1.043 | 0.984 | 1.008 | 0.992 | 0.933 | 1.037 | **0.927** | 1.030 | 1.023 |
| CADL | 0.985** | 1.012 | **0.983**** | 0.995 | 1.029 | **0.786**** | 0.964** | **0.973** | 0.974 | 1.016 | 1.061 |
| Bagging | 2.681 | 2.070 | 9.818 | 3.783** | 7.000** | 10.45 | 6.306** | 9.277 | 9.428 | 12.94 | 8.207 |
| Boosting | **0.958** | **0.998** | 1.022 | 0.953 | 1.002 | 0.915 | 0.899 | 1.026** | 0.973 | 1.006 | 1.002 |
| BMA1 | 1.209 | 1.035 | 1.162 | 1.579 | 1.253 | 1.669 | 2.078 | 1.112 | 1.045 | 1.038 | 1.203 |
| BMA2 | 1.133 | 1.011 | 1.054 | 1.383 | 1.126** | 1.365 | 1.783 | 1.063 | 1.004 | 1.021 | 1.036 |
| Ridge | 1.321 | 1.178 | 1.863 | 1.210 | 1.306** | 1.124 | 1.129 | 1.598 | 1.083 | 1.358 | 1.692 |
| LAR | 0.983 | 1.002 | 1.001 | 1.004 | 0.997 | 0.968 | 0.928 | 1.006 | 0.983 | 1.000 | 1.001 |
| EN | 0.985 | 1.002 | 1.001 | 0.992 | **0.995** | 0.982 | 0.984 | 0.999 | 0.981 | **0.999** | 1.000 |
| NNG | 0.996 | 1.002 | 1.000 | 0.999 | 1.000 | 0.999 | 0.998 | 1.001 | 1.001 | 1.001 | **0.998** |
| Mean | 1.118 | 1.154 | 1.076 | **0.950** | 1.020 | 1.185** | **0.897**** | 1.526 | 0.996** | 1.062 | 1.080 |

[a] Notes: See notes to Tables 3–5.

when MSFE "best" models are tabulated by specification type, our model averaging methods perform quite well, particularly for SP2 and SP3. This finding can be confirmed by carefully examining the results reported in Tables 3–6 (i.e., compare the MSFE "best" models, denoted by bold entries). However, notice that when results are summarized across *all* specification types (see Panel B of the table), then the model averaging methods yield MSFE "best" predictions in far fewer cases. This is because SP1, where model averaging clearly "wins" the least, is the predominant winner when comparing results across *all* specification types, as mentioned previously. Moreover, it is clear that our "hybrid" 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 (i.e., SP1) is the dominant specification type. In summary, when more complicated specification methods are used, model averaging methods fare worse and "hybrid" methods fare better. Still, pure factor type models sometimes perform well, particularly for the long expansion period from 1982 to 1990. Second, careful examination of Panel B of this table indicated that our so-called "Nonlinear Factor" models (i.e., all shrinkage/factor combination models—see footnote to Table 9) exhibit business cycle dependent performance. In particular, the "Nonlinear Factor" models are MSFE "best" for 7 of 11 variables in recessionary periods (in recessions, "Mean" – or model averaging – "wins" for 0 of 11 variables, while "Linear Factor" models and simple AR type models each win twice). On the other hand, "Nonlinear Factor" models "win" for only 4 of 11 variable (or approximately 1/3 of the time) during expansionary periods (in expansions, "Mean" models "win" for 5 of 11 variables, while "Linear Factor" models and simple AR type models each "win" once). Additionally, for all but 1 other sub-sample, "Nonlinear Factor" models win for 4 or 5 of 11 variables. Indeed, it is only during the 01:11–07:11 period that "Nonlinear Factor" models do not win more than 1/3 of the time, suggesting that during extremely stable periods there is less to be gained by using more sophisticated nonlinear models.

Table 10 reports reality check statistics based on White (2000) and Corradi and Swanson (2007). In order to facilitate construction of these statistics as well as simulation of critical values (following CS (2007)), we re-ran all experiments, fixing the number of lags and the number of factors used in our recursive and rolling window based forecast model construction. The reality check tests the null hypothesis that a given benchmark model, say Model 1 yields equal or better predictive performance than all competitors, say Models $2, \ldots, n$. The alternative is that at least one models among Models $2, \ldots, n$ outperforms the benchmark. Formally, we test:

$$H_0 : \max_{i=2,\ldots,n} E\left(g\left(u_{1,t+1}\right) - g\left(u_{i,t+1}\right)\right) \leq 0$$

$$H_A : \max_{i=2,\ldots,n} E\left(g\left(u_{1,t+1}\right) - g\left(u_{i,t+1}\right)\right) > 0,$$

where $u_{i,t+1} = y_{t+1} - \kappa_i\left(Z_t, \theta_i\right)$, $\kappa_i$ is $i$-th conditional mean function, $\theta_i$ are model parameters, $Z_t$ is the set of predictors, and

**Table 7**
Forecast experiment summary results.[a] Best forecasting methods by the target variable and specification type.

| $h$ | UR | PI | TB | CPI | PPI | NPE | HS | IPX | M2 | SNP | GDP |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Panel A: Specification Type 1 without lags | | | | | | | | | | | |
| 1 | FAAR | PCR | Mean | PCR | PCR | CADL | Mean | Ridge | FAAR | AR | ARX |
| 3 | PCR | ARX | Bag | Mean | PCR | CADL | Bag | Mean | Mean | Bag | NNG |
| 12 | PCR | AR | PCR | Mean | Mean | CADL | Bag | CADL | ARX | Mean | Mean |
| Panel B: Specification Type 1 with lags | | | | | | | | | | | |
| 1 | Mean | Mean | CADL | ARX | Mean | CADL | CADL | ARX | AR | AR | Mean |
| 3 | CADL | ARX | Mean | Mean | Mean | CADL | CADL | CADL | Mean | AR | AR |
| 12 | Boost | Mean | CADL | Mean | Mean | CADL | ARX | CADL | ARX | AR | Mean |
| Panel C: Specification Type 2 | | | | | | | | | | | |
| 1 | LAR | EN | CADL | ARX | Bag | Mean | CADL | ARX | Mean | AR | ARX |
| 3 | BMA2 | Mean | Mean | NNG | EN | CADL | CADL | Mean | AR | AR | AR |
| 12 | Mean | Mean | CADL | ARX | AR | CADL | Mean | LAR | Mean | BMA1 | Mean |
| Panel D: Specification Type 3 | | | | | | | | | | | |
| 1 | CADL | ARX | Boost | Mean | Mean | CADL | CADL | Boost | LAR | LAR | Boost |
| 3 | CADL | Boost | CADL | ARX | ARX | CADL | Boost | EN | Mean | EN | EN |
| 12 | Boost | Boost | CADL | Mean | EN | CADL | Mean | CADL | ARX | EN | NNG |
| Panel E: All Specification Types | | | | | | | | | | | |
| 1 | SP1<br>FAAR | SP1<br>PCR | SP1<br>Mean | SP1<br>PCR | SP1<br>PCR | SP2<br>Mean | SP1<br>Mean | SP1<br>Ridge | SP1<br>FAAR | SP3<br>LAR | SP1L<br>Mean |
| 3 | SP2<br>BMA2 | SP3<br>Boost | SP1L<br>Mean | SP1L<br>Mean | SP1L<br>Mean | –<br>CADL | SP1<br>Bag | SP2<br>Mean | SP3<br>Mean | SP1<br>Bag | SP1<br>NNG |
| 12 | SP2<br>Mean | SP2<br>Mean | SP1<br>PCR | SP1L<br>Mean | SP1L<br>Mean | –<br>CADL | SP1<br>Bag | SP2<br>LAR | SP2<br>Mean | SP2<br>BMA1 | SP2<br>Mean |

[a] Notes: See notes to Tables 1–5. Entries in first four panels denote the method yielding the lowest point MSFE for given target variable. These entries correspond to entries in bold in Tables 3–6. Each pair of entries in Panel E denotes the MSFE "best" method and specification type for given target variables and forecast horizon, using recursive estimation. The way to read entries in Panels E is as follows. When forecasting unemployment (UR), the FAAR model under Specification Type 1 without lags yields the lowest MSFE, for $h = 1$, across all specification types.

$g$ is a given loss function, here assumed to be quadratic. Following White (2000), define the statistic $S_p = \max_{i=2,\ldots,n} S_P(1, i)$ with

$$S_P(1, i) = \frac{1}{\sqrt{P}} \sum_{t=R}^{T-1} \left( g\left(\hat{u}_{1,t+1}\right) - g\left(\hat{u}_{i,t+1}\right) \right), \quad i = 2, \ldots, n$$

where $\hat{u}_{i,t+1} = y_{t+1} - \kappa_i\left(Z_t, \hat{\theta}_{i,t}\right)$ and $\kappa_i\left(Z_t, \hat{\theta}_{i,t}\right)$ is the estimated conditional mean under model $i$. See CS (2007) for technical details about computing critical values. We executed all tests using three different block lengths ($l = 2, 5$, and 10), although we report results only for $l = 5$, since our findings were not affected by block length. Double and single starred entries denote rejection of the null hypothesis at 5% and 10% significance levels, respectively. We consider various test scenarios, including: (I) set the AR(SIC) model as the benchmark, and compare this model against a variety of alternative models, including, for example, all models under SP1, SP2, or SP3; (II) set our factor augmented autoregressive model (FAAR) as the benchmark; and (III) set the point MSFE "best" model from each specification type as the benchmark (see Table 7 for summary of "best" models).

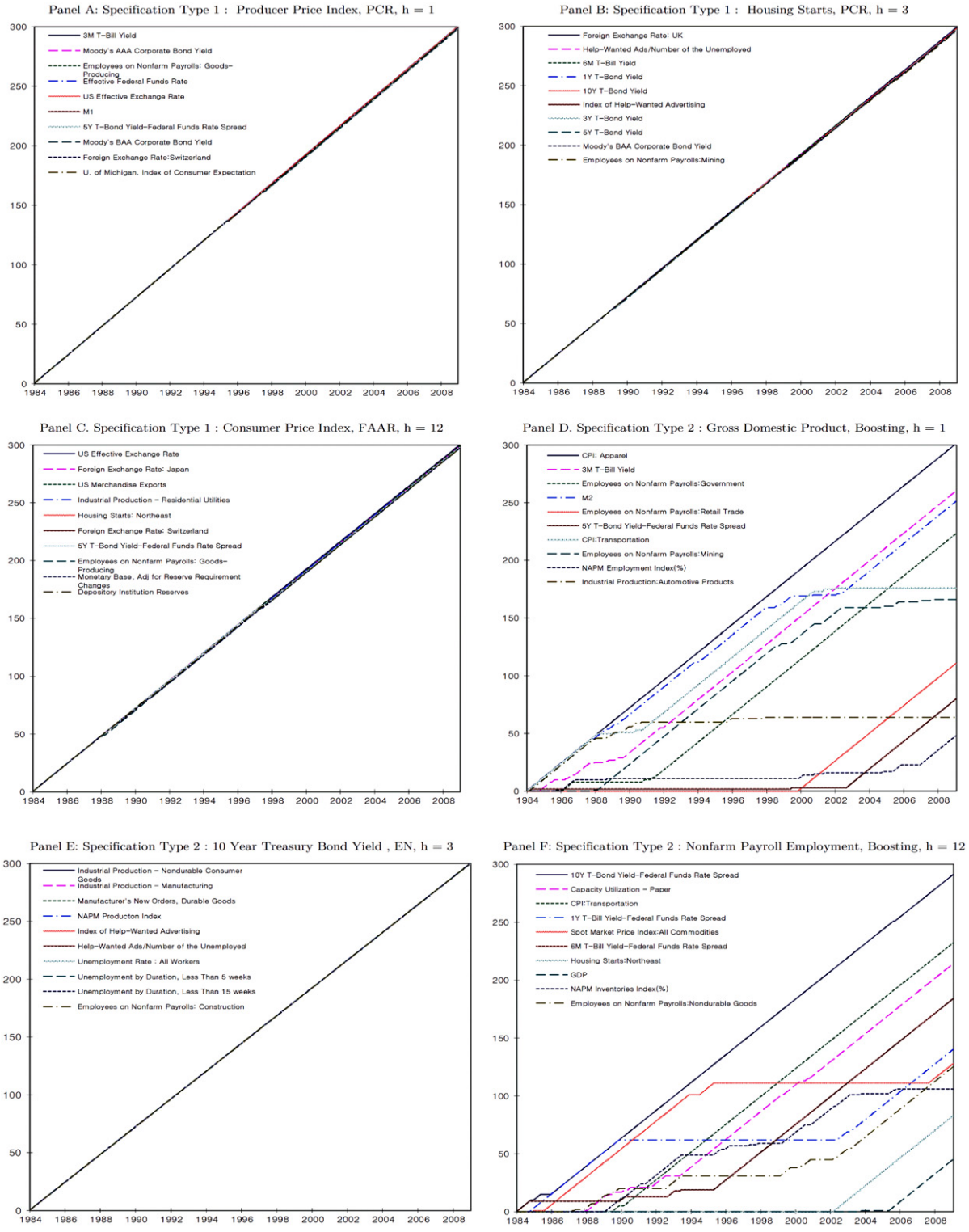
As evident from inspection of the tabulated results, under scenario (I), the null is almost always rejected, particularly under SP1, as expected given our findings that are discussed above. Under scenario (II), the null hypothesis that the FAAR model is at least as "MSFE-accurate" as any other model is rejected for approximately 9 out of 11 variables, at all forecast horizons. This supports our earlier finding concerning the effectiveness of our "hybrid" models. Namely, pure factor models alone do not usually "win" when compared with either shrinkage and/or "hybrid" models. Indeed, when FAAR is compared with pure shrinkage models with no diffusion indices (i.e., SP3) the null is rejected for 6 out of 11 variables when $h = 1$, whereas for $h = 12$ the null is rejected for all variables. Thus, pure shrinkage models outperform pure linear factor models in many cases. However, it is only when "hybrid" models are used that pure linear factor models are generally "dominated". In summary, our reality check results are largely in agreement with our point MSFE results.

As a final measure of the relevance of our findings, we constructed a simple measure of forecast accuracy "percentage differences" between various AR, "Mean", and MSFE "best" models discussed above, as reported in Table 11. The "percentage differences" were calculated using the following formula:

$$\frac{1}{T} \sum_{t} \left( \left| \frac{\hat{Y}_{1,t} - Y_t}{Y_t} \right| - \left| \frac{\hat{Y}_{2,t} - Y_t}{Y_t} \right| \right) \times 100\%,$$

where $Y_t$ is the actual value of the target variable at time $t$. Additionally, $\hat{Y}_{1,t}$ and $\hat{Y}_{2,t}$ are the forecasts from the two models being compared. In rare occurrences where $Y_t$ is zero, it is replaced by $Y_{t-1}$. Turning to the tabulated results, in the first row of entries, "AR vs. SP1Best" denotes the "percentage difference" between AR(SIC) forecasts and those of the MSFE "best" model under SP1. The value of 11.07% in the first numerical entry of the table (upper left), thus indicates a marked improvement in predictive accuracy when using "SP1best" (i.e., FAAR—see Table 3) instead of AR(SIC). This "percentage difference" might be assumed to have significant impact when predictions are used to calibrate macroeconomic policy decisions, for example. Although results vary somewhat, it is clear upon inspection of the tabulated entries that "percentage differences" often range from around 5% to 15%, and for some variables, such as unemployment, the 10-year Treasury bond rate, CPI, PPI, and housing starts, most "percentage differences" lie in this range.

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 Fig. 1, where we report the ten most frequently selected variables for a variety of MSFE "best" models and forecast horizons. Keeping




*Notes: Panels in this figure depict the 10 most commonly selected variables for use in factor construction, across the entire prediction period from 1984:6-2009:5, 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 every point in time.

**Fig. 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 1984:6–2009:5, where factors are re-estimated at each point in time, prior to each new prediction being constructed. 45° lines denote cases for which a particular variable 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 every point in time.

**Table 8**
Forecast experiment summary results.[a]

| Panel A: Summary of MSFE "best" models across all specification types | | | | | | |
|---|---|---|---|---|---|---|
| | Recursive estimation windows | | | Recursive and rolling estimation windows | | |
| | $h = 1$ | $h = 3$ | $h = 12$ | $h = 1$ | $h = 3$ | $h = 12$ |
| AR(SIC) | 0 | 0 | 0 | 0 | 0 | 0 |
| ARX(SIC) | 0 | 0 | 0 | 0 | 0 | 0 |
| CADL | 0 | 1 | 1 | 0 | 1 | 1 |
| FAAR | 2 | 0 | 0 | 2 | 0 | 0 |
| PCR | 3 | 0 | 1 | 3 | 0 | 1 |
| Bagging | 0 | 2 | 1 | 0 | 2 | 0 |
| Boosting | 0 | 1 | 0 | 0 | 1 | 1 |
| BMA1 | 0 | 0 | 1 | 0 | 0 | 3 |
| BMA2 | 0 | 1 | 0 | 0 | 1 | 0 |
| Ridge | 1 | 0 | 0 | 1 | 0 | 0 |
| LAR | 1 | 0 | 1 | 0 | 0 | 1 |
| EN | 0 | 0 | 0 | 0 | 0 | 0 |
| NNG | 0 | 1 | 0 | 2 | 1 | 0 |
| Mean | 4 | 5 | 6 | 3 | 5 | 4 |
| Panel B: Summary of MSFE-"best" models | | | | | | |
| | Winners by estimation window type | | | Winners by specification type | | |
| | $h = 1$ | $h = 3$ | $h = 12$ | $h = 1$ | $h = 3$ | $h = 12$ |
| Specification Type 1 without lags | | | | | | |
| Recursive | 6 | 10 | 4 | 8 | 4 | 3 |
| Rolling | 5 | 1 | 7 | | | |
| Specification Type 1 with lags | | | | | | |
| Recursive | 8 | 10 | 7 | 1 | 3 | 2 |
| Rolling | 3 | 1 | 4 | | | |
| Specification Type 2 | | | | | | |
| Recursive | 7 | 7 | 6 | 1 | 2 | 6 |
| Rolling | 4 | 4 | 5 | | | |
| Specification Type 3 | | | | | | |
| Recursive | 6 | 10 | 7 | 1 | 2 | 0 |
| Rolling | 5 | 1 | 4 | | | |

[a] Notes: See notes to Tables 2, 3 and 7. Numerical entries in Panel A denote the number of "wins" of each forecasting method, at a given forecasting horizon. The left-hand set of 3 columns summarizes results for recursive estimation, and correspond to entries in Panel B of Table 7. The right-hand set of 3 columns summarizes results based on both recursive and rolling estimation methods.

**Table 9**
Forecast experiment summary results.[a]

| Subsample | Specification Type 1 without lags | | | | Specification Type 1 with lags | | | | Specification Type 2 | | | | Specification Type 3 | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Mean | Linear factor | Nonlinear factor | Other | Mean | Linear factor | Nonlinear factor | Other | Mean | Linear factor | Nonlinear factor | Other | Mean | Linear factor | Nonlinear factor | Other |
| Panel A: Wins by specification types, $h = 1$, recursive estimation | | | | | | | | | | | | | | | | |
| 84:06–90:06 | 3 | 3 | 4 | 1 | 4 | 0 | 1 | 6 | 0 | 0 | 5 | 6 | 4 | 1 | 4 | 2 |
| 91:03–01:02 | 5 | 2 | 4 | 0 | 4 | 0 | 1 | 6 | 3 | 0 | 1 | 7 | 4 | 0 | 4 | 3 |
| 01:11–07:11 | 0 | 4 | 4 | 2 | 2 | 1 | 1 | 7 | 1 | 0 | 1 | 9 | 5 | 1 | 1 | 4 |
| Expansion | 4 | 4 | 2 | 1 | 4 | 1 | 0 | 6 | 2 | 0 | 3 | 6 | 5 | 1 | 4 | 1 |
| Recession | 1 | 4 | 2 | 4 | 1 | 1 | 2 | 7 | 0 | 2 | 3 | 6 | 0 | 2 | 7 | 2 |
| Panel B: Wins across all specification types, $h = 1$, recursive estimation | | | | | | | | | | | | | | | | |
| 84:06–90:06 | 0 | 1 | 1 | 0 | 0 | 0 | 1 | 3 | 0 | 0 | 3 | 1 | 1 | 0 | 0 | 0 |
| 91:03–01:02 | 1 | 1 | 3 | 0 | 2 | 0 | 0 | 1 | 1 | 0 | 1 | 0 | 1 | 0 | 0 | 0 |
| 01:11–07:11 | 0 | 2 | 1 | 1 | 0 | 1 | 0 | 1 | 0 | 0 | 0 | 3 | 2 | 0 | 0 | 0 |
| Expansion | 1 | 0 | 1 | 0 | 0 | 0 | 3 | 1 | 2 | 1 | 0 | 0 | 2 | 0 | 0 | 0 |
| Recession | 0 | 0 | 1 | 2 | 0 | 1 | 1 | 0 | 0 | 1 | 2 | 0 | 0 | 0 | 3 | 0 |

[a] Notes: See notes to Tables 3 and 8. Entries in Panel A enumerate how many times each "group" of forecasting methods yields the lowest MSFE for each target variable and for each specification type. The "groups" are defined as follows. "Mean" includes: BMA, Combined-ADL and Mean. "Linear Factor" includes: FAAR and PCR. "Nonlinear Factor" includes: all shrinkage/factor combination models (i.e., see Specification Types 1 and 2 in Table 2). Finally, "Other" includes our linear AR(SIC) and ARX(SIC) models. See Section 3 for further details. Entries in Panel B enumerate how many times each set of forecasting methods yields lowest MSFE across all specification types. Subsamples follow business cycles dated by the NBER. The first relevant NBER dated expansion started in 82:11, but our forecasting experiments start in 84:6, so that the start date of first subsample is set at 84:6. "Expansion" includes all of expansion dates in the forecasting period and "Recession" includes all of contraction dates in the forecasting period.

in mind that factors are re-estimated at each point in time, prior to each new prediction being constructed, a 45° line denotes cases for which a particular variable is selected each time a forecast is constructed. For example, in Panel A, note that the BAA Bond Yield—Federal Funds Rate spread is always selected as a predictor when constructing factors to forecast the Producer Price Index for $h = 1$. For SP1, variables are selected based on the $A(j)$ and $M(j)$ statistics of Bai and Ng (2006a) and Armah and Swanson (2010), and for SP2, we directly observe variables that are selected via shrinkage, and subsequently used to construct diffusion indexes. The list of selected variables does not vary much, for SP1. On the other hand, in Panels D and F, we see that frequently

**Table 10**
Reality check statistics for various forecast experiments.[a]

| Models | UR | PI | TB | CPI | PPI | NPE | HS | IPX | M2 | SNP | GDP |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Panel A: Recursive estimation, $h = 1$ | | | | | | | | | | | |
| AR vs. SP1s | 0.402* | 5.800** | 0.809* | 1.550* | 5.429* | 0.132* | 0.019* | 0.818* | 0.859* | −0.069 | 5.982* |
| AR vs. SP1FAARs | 0.402* | 5.800** | 0.809* | 1.550* | 5.429* | 0.101* | 0.019* | 0.818* | 0.859* | −0.069 | 4.563* |
| AR vs. SP1LFAARs | 0.122* | 4.977** | 0.096* | 0.375* | 1.582* | 0.132* | 0.009* | 0.520* | −1.090 | −0.208 | 9.301* |
| AR vs. SP2s | 0.395** | 3.986** | −1.251 | −1.265 | 1.845* | 0.149* | −0.168 | 0.345* | 1.716* | −1.146 | 1.984** |
| AR vs. SP3s | 0.168* | 2.245** | 0.620** | 1.072** | 4.217* | 0.031* | −0.001 | 0.640** | 0.924** | 0.048* | 7.098* |
| SP1FAAR vs. SP1Shrinks | −1.029 | 2.061* | 1.126* | 0.425* | 1.993** | 1.246* | 0.603* | 0.007* | −0.189 | 0.984* | 3.251* |
| SP1FAAR vs. SP1Ls | −0.279 | 1.237* | −2.469 | −1.624 | −1.855 | −0.705 | −1.236 | −0.726 | −3.463 | −1.980 | 7.989* |
| SP1FAAR vs. SP2s | −0.006 | 0.246* | −0.934 | −2.390 | −1.591 | 0.173* | −0.126 | −0.465 | 0.857* | −0.093 | 0.673* |
| SP1FAAR vs. SP3s | −0.234 | −1.495 | 0.937* | −0.053 | 0.780** | 0.054* | 0.040** | −0.170 | 0.065** | 1.101* | 5.786* |
| SP1Best vs. SP1s | −0.103 | −2.061 | −0.046 | −0.425 | −1.993 | −0.031 | −0.010 | −0.001 | −0.189 | −0.069 | −1.419 |
| SP1Best vs. SP1Ls | −0.279 | −0.082 | −3.595 | −2.049 | −0.385 | −0.862 | −0.130 | −0.733 | −0.346 | −3.033 | 3.319* |
| SP1Best vs. SP2s | −0.006 | −1.814 | −2.060 | −2.815 | −3.584 | 0.016* | −0.187 | −0.472 | 0.857* | −1.146 | −3.998 |
| SP1Best vs. SP3s | −0.234 | −3.555 | −0.188 | −0.478 | −1.213 | −0.102 | −0.020 | −0.178 | 0.065* | 0.048* | 1.116* |
| SP1LBest vs. SP1Ls | −0.352 | −0.193 | −0.096 | −0.375 | −1.095 | −1.323 | −0.089 | −0.368 | −1.090 | −0.208 | −1.404 |
| SP2Best vs. SP2s | 0.000 | −0.003 | −0.096 | −0.375 | −0.136 | −0.016 | −0.001 | −0.284 | −0.172 | −0.208 | −5.982 |
| SP3Best vs. SP3s | −0.081 | −0.119 | −0.520 | −0.496 | −3.236 | −0.102 | −0.009 | −0.040 | −0.775 | −0.009 | −0.750 |
| SP3Best vs. SP1s | 0.234 | 0.344 | 0.188 | 0.478 | 0.121 | −0.031 | 0.001 | 0.178 | −0.006 | −0.117 | −2.535 |
| SP3Best vs. SP1Ls | −0.045 | 0.261 | −3.406 | −1.571 | −0.263 | −0.862 | −0.129 | −0.555 | −0.353 | −3.081 | 2.203 |
| SP3Best vs. SP2s | 0.227 | 0.162 | −1.872 | −2.337 | −0.237 | 0.016 | −0.018 | −0.295 | 0.079 | −1.194 | −5.113 |
| Panel B: Recursive estimation, $h = 3$ | | | | | | | | | | | |
| AR vs. SP1s | 0.092* | 1.723** | 0.241* | 0.519* | 4.492* | 0.284* | 0.139* | 0.293* | 0.240* | 0.221** | 0.059* |
| AR vs. SP1FAARs | 0.092* | 1.338** | 0.241* | 0.519* | 4.492* | 0.017* | 0.139* | 0.293* | 0.240* | 0.221** | 0.059* |
| AR vs. SP1LFAARs | 0.023* | 1.723** | 0.404** | 3.072* | 8.413* | 0.284* | 0.030* | 0.241** | 0.345* | −0.340 | −0.539 |
| AR vs. SP2s | 0.286* | 2.075** | 0.301** | 1.432** | 5.018* | 0.237* | −0.423 | 0.410** | 0.604* | −0.290 | −0.980 |
| AR vs. SP3s | 0.063* | 3.273** | 0.009* | 1.237** | 1.685* | 0.288** | 0.096* | 0.417** | 2.432** | 0.111** | 0.037* |
| SP1FAAR vs. SP1Shrinks | 1.407* | 0.415* | 0.546* | 0.310* | 1.986* | 0.114* | 3.936* | 0.377* | 0.622* | 1.954* | 5.530* |
| SP1FAAR vs. SP1Ls | 0.057* | 0.038* | 0.709* | 2.863* | 5.907* | −0.700 | −1.188 | −0.184 | 0.727* | 1.252* | 4.175* |
| SP1FAAR vs. SP2s | 0.335* | 1.151* | 0.606* | 1.222* | 2.512* | 0.231** | −0.168 | 0.493* | 0.986** | 1.442* | 4.491** |
| SP1FAAR vs. SP3s | 0.111* | 2.350* | 0.314* | 1.027* | −0.821 | 0.282** | 0.351* | 0.501* | 2.813** | 1.844* | 5.508* |
| SP1Best vs. SP1s | −0.070 | −0.385 | −0.055 | −0.168 | −0.889 | −0.267 | −0.109 | −0.052 | −0.210 | −0.221 | −0.059 |
| SP1Best vs. SP1Ls | −0.084 | −0.076 | 0.164* | 2.553* | 0.392* | −0.979 | −0.158 | −0.561 | 0.011* | −0.702 | −1.355 |
| SP1Best vs. SP2s | 0.194* | 0.352* | 0.060* | 0.913* | 0.526* | −0.047 | −0.562 | 0.116* | 0.364* | −0.512 | −1.039 |
| SP1Best vs. SP3s | −0.030 | 1.550* | −0.232 | 0.717* | −2.807 | 0.004* | −0.043 | 0.124* | 2.192* | −0.110 | −0.022 |
| SP1LBest vs. SP1Ls | −0.140 | −0.761 | −0.061 | −0.023 | −0.991 | −2.838 | −0.303 | −0.241 | −0.134 | −0.340 | −0.539 |
| SP2Best vs. SP2s | −0.008 | −0.035 | −0.115 | −0.437 | −0.142 | −0.047 | −0.003 | −0.169 | −0.060 | −0.340 | −0.539 |
| SP3Best vs. SP3s | −0.040 | −0.892 | −0.177 | −0.885 | −1.918 | −0.004 | −0.066 | −0.147 | −2.158 | −0.088 | −0.037 |
| SP3Best vs. SP1s | 0.030 | −0.194 | 0.055 | −0.717 | 0.089 | −0.271 | 0.004 | −0.124 | −0.219 | 0.110 | 0.022 |
| SP3Best vs. SP1Ls | −0.054 | −0.231 | 0.219 | 1.835 | 0.481 | −0.983 | −0.154 | −0.684 | −0.209 | −0.592 | −1.333 |
| SP3Best vs. SP2s | 0.224 | −0.120 | 0.115 | 0.195 | 0.142 | −0.051 | −0.052 | −0.007 | −0.183 | −0.402 | −1.017 |
| Panel C: Recursive estimation, $h = 12$ | | | | | | | | | | | |
| AR vs. SP1s | 0.114* | −0.159 | 0.344* | 0.663* | 1.242** | 1.135* | 1.090* | 0.185** | 1.646** | 0.040* | 0.236* |
| AR vs. SP1FAARs | 0.114* | −0.159 | 0.344* | 0.663* | 1.242** | 0.024* | 1.090* | 0.015* | 0.956** | 0.040* | 0.236* |
| AR vs. SP1LFAARs | 0.163* | 0.354* | 0.217** | 1.641** | 2.325** | 1.135** | 0.469* | 0.185** | 1.646* | −0.078 | 0.606* |
| AR vs. SP2s | 0.215* | 1.360* | 0.142* | 1.072** | 0.906* | 0.915* | 0.202* | 0.324* | 2.350** | −0.007 | 1.044* |
| AR vs. SP3s | 0.183* | 1.378** | 0.007* | 1.423** | 2.383** | 0.825* | 0.704* | 0.298** | 3.192** | 0.241* | 0.367* |
| SP1FAAR vs. SP1Shrinks | 0.334* | 2.923* | 0.247* | 0.319* | 1.240* | 4.014* | 7.455* | 0.711* | 0.419* | 0.480* | 1.436* |
| SP1FAAR vs. SP1Ls | 0.082* | 3.436* | −0.183 | 1.298* | 2.324* | 0.501* | −0.703 | 0.671* | 0.813* | 0.362* | 1.805* |
| SP1FAAR vs. SP2s | 0.134* | 4.442* | 0.046* | 0.729* | 0.905* | 1.292* | −0.143 | 1.020* | 1.813* | 0.433* | 2.243* |
| SP1FAAR vs. SP3s | 0.102* | 4.460* | −0.089 | 1.079* | 2.382* | 1.202* | 0.359* | 0.993* | 2.655* | 0.680* | 1.566* |
| SP1Best vs. SP1s | −0.003 | −0.159 | −0.126 | −0.135 | −0.516 | −1.094 | −0.621 | −0.170 | −0.690 | −0.021 | −0.196 |
| SP1Best vs. SP1Ls | 0.049* | 0.035* | −0.431 | 0.979* | 0.108* | −1.011 | −0.145 | −0.209 | −0.030 | −0.118 | 0.369* |
| SP1Best vs. SP2s | 0.101* | 1.360* | −0.202 | 0.410* | −0.335 | −0.220 | −0.888 | 0.139* | 0.705* | −0.047 | 0.808* |
| SP1Best vs. SP3s | 0.069* | 1.378* | −0.336 | 0.760* | 1.141* | −0.309 | −0.386 | 0.112* | 1.546* | 0.200* | 0.131* |
| SP1LBest vs. SP1Ls | −0.378 | −0.354 | −0.217 | −0.560 | −1.448 | −10.113 | −2.209 | −0.185 | −0.296 | −0.078 | −0.095 |
| SP2Best vs. SP2s | −0.005 | −0.136 | −0.217 | −0.842 | −0.091 | −0.662 | −0.022 | −0.139 | −0.070 | −0.080 | −0.508 |
| SP3Best vs. SP3s | −0.057 | −0.613 | −0.210 | −0.744 | −2.084 | −0.309 | −0.203 | −0.112 | −1.546 | −0.046 | −0.303 |
| SP3Best vs. SP1s | −0.069 | −0.154 | 0.126 | −0.760 | −0.114 | −1.111 | 0.039 | −0.282 | −0.224 | −0.200 | −0.131 |
| SP3Best vs. SP1Ls | −0.020 | −0.102 | −0.305 | 0.218 | −0.006 | −1.011 | −0.106 | −0.322 | −0.184 | −0.319 | 0.239 |
| SP3Best vs. SP2s | 0.032 | −0.002 | −0.075 | −0.351 | −0.148 | −0.220 | −0.050 | 0.027 | −0.084 | −0.248 | 0.677 |

[a] Notes: See notes to Tables 1, 2 and 7. Numerical entries are reality check statistics due to White (2000) and Corradi and Swanson (2007), calculated for the forecast period 1984:6–2009:05, and summarize various results from Tables 3–6. The null hypothesis of the statistic is that no alternative model "outperforms" the benchmark model. See Section 6 for further details. The first entry under the header "Model" denotes the benchmark model, and the second entry denotes the group of alternative models against which the benchmark is compared. Here, AR denotes our autoregressive benchmark model, SP1s denotes all SP1 models, excluding the benchmark, SP2s and SP3s are analogously defined. A mnemonic with "Best" appended onto it denotes the point MSFE "best" model for a particular specification type. The notion "FAARs" and "Shrinks" refers to all pure factor models, and all pure shrinkage (i.e., Bagging, Boosting, Ridge, LAR, EN, and NNG) models, respectively. Finally, SP1FAAR, when used as a benchmark, refer to the FAAR model estimation under SP1.

* Denote rejection of the null hypothesis at 10% significance level.

** Denote rejection of the null hypothesis at 5% significance level.

**Table 11**
Predictive accuracy percentage differences based on the comparison of various pairs of forecasting models.[a]

| Models | UR | PI | TB | CPI | PPI | NPE | HS | IPX | M2 | SNP | GDP |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Panel A: Recursive estimation, $h = 1$ | | | | | | | | | | | |
| AR vs. SP1Best | 11.07% | 0.10% | 4.68% | 1.44% | 4.23% | 0.02% | 1.25% | 0.14% | 0.06% | 0.00% | 0.23% |
| AR vs. SP1LBest | 5.04% | 0.04% | 1.84% | 0.75% | 1.22% | 0.02% | 4.97% | 0.11% | 0.04% | 0.47% | 0.04% |
| AR vs. SP2Best | 2.21% | 0.03% | 3.99% | 0.14% | 0.46% | 0.07% | 2.31% | 0.00% | 0.06% | 0.58% | 0.00% |
| AR vs. SP3Best | 4.00% | 0.12% | 0.58% | 0.00% | 0.01% | 0.01% | 1.60% | 0.36% | 0.49% | 0.89% | 0.27% |
| AR vs. SP1Mean | 5.36% | 0.03% | 4.68% | 0.78% | 1.20% | 0.02% | 1.25% | 0.09% | 0.03% | 0.34% | 0.06% |
| AR vs. SP1LMean | 5.04% | 0.04% | 4.61% | 0.75% | 1.22% | 0.02% | 1.20% | 0.11% | 0.04% | 0.38% | 0.04% |
| AR vs. SP2Mean | 1.70% | 0.13% | 3.99% | 0.14% | 0.46% | 0.07% | 7.16% | 0.14% | 0.22% | 0.51% | 0.14% |
| AR vs. SP3Mean | 1.77% | 0.10% | 4.33% | 0.12% | 0.16% | 0.07% | 6.32% | 0.13% | 0.19% | 0.41% | 0.13% |
| SP1Mean vs. SP1Best | 7.90% | 0.08% | 0.00% | 0.91% | 3.22% | 0.03% | 0.00% | 0.07% | 0.03% | 0.34% | 0.19% |
| SP1LMean vs. SP1LBest | 0.00% | 0.00% | 3.36% | 0.00% | 0.00% | 0.03% | 4.58% | 0.00% | 0.05% | 0.49% | 0.00% |
| SP2Mean vs. SP2Best | 2.71% | 0.13% | 0.00% | 0.00% | 0.00% | 0.00% | 8.64% | 0.14% | 0.23% | 0.53% | 0.14% |
| SP3Mean vs. SP3Best | 3.63% | 0.12% | 4.22% | 0.12% | 0.17% | 0.07% | 6.70% | 0.32% | 0.30% | 0.81% | 0.22% |
| SP1Best vs. SP2Mean | 18.04% | 0.17% | 15.44% | 6.26% | 6.00% | 0.09% | 8.98% | 0.21% | 0.27% | 1.18% | 0.30% |
| SP1Best vs. SP2Best | 13.66% | 0.12% | 11.17% | 5.44% | 5.88% | 0.06% | 3.23% | 0.18% | 0.16% | 1.06% | 0.26% |
| SP1Best vs. SP3Mean | 17.45% | 0.14% | 10.26% | 6.23% | 5.88% | 0.08% | 7.47% | 0.17% | 0.21% | 0.87% | 0.22% |
| SP1Best vs. SP3Best | 13.55% | 0.14% | 4.68% | 5.36% | 5.88% | 0.02% | 1.37% | 0.34% | 0.49% | 0.89% | 0.21% |
| SP2Best vs. SP3Mean | 2.77% | 0.12% | 11.53% | 0.13% | 0.38% | 0.08% | 6.46% | 0.15% | 0.22% | 0.98% | 0.17% |
| SP2Best vs. SP3Best | 4.47% | 0.14% | 10.22% | 0.14% | 0.28% | 0.07% | 3.48% | 0.35% | 0.49% | 1.38% | 0.28% |
| Panel B: Recursive estimation, $h = 3$ | | | | | | | | | | | |
| AR vs. SP1Best | 8.56% | 0.12% | 3.97% | 0.73% | 3.94% | 0.03% | 1.80% | 0.11% | 0.03% | 0.18% | 0.02% |
| AR vs. SP1LBest | 10.27% | 0.10% | 0.76% | 1.52% | 2.07% | 0.03% | 4.14% | 0.13% | 0.04% | 0.07% | 0.00% |
| AR vs. SP2Best | 2.73% | 0.07% | 2.27% | 0.15% | 0.04% | 0.05% | 6.54% | 0.10% | 0.07% | 0.00% | 0.00% |
| AR vs. SP3Best | 3.52% | 0.07% | 4.05% | 0.01% | 0.66% | 0.16% | 0.28% | 0.23% | 0.53% | 0.62% | 0.00% |
| AR vs. SP1Mean | 6.05% | 0.04% | 1.76% | 0.73% | 1.26% | 0.02% | 2.05% | 0.11% | 0.03% | 0.18% | 0.05% |
| AR vs. SP1LMean | 4.88% | 0.04% | 3.34% | 0.73% | 2.07% | 0.02% | 2.00% | 0.09% | 0.04% | 0.22% | 0.05% |
| AR vs. SP2Mean | 1.46% | 0.11% | 2.01% | 0.15% | 0.30% | 0.07% | 6.54% | 0.11% | 0.24% | 0.23% | 0.10% |
| AR vs. SP3Mean | 1.77% | 0.11% | 2.73% | 0.13% | 0.34% | 0.06% | 5.72% | 0.11% | 0.22% | 0.27% | 0.10% |
| SP1Mean vs. SP1Best | 6.22% | 0.09% | 3.98% | 0.00% | 2.79% | 0.05% | 2.43% | 0.00% | 0.00% | 0.19% | 0.03% |
| SP1LMean vs. SP1LBest | 7.89% | 0.07% | 3.00% | 0.84% | 0.00% | 0.04% | 4.25% | 0.17% | 0.04% | 0.19% | 0.05% |
| SP2Mean vs. SP2Best | 3.12% | 0.14% | 2.16% | 0.00% | 0.27% | 0.08% | 0.00% | 0.13% | 0.26% | 0.23% | 0.10% |
| SP3Mean vs. SP3Best | 2.78% | 0.13% | 3.25% | 0.13% | 0.42% | 0.10% | 5.78% | 0.20% | 0.31% | 0.50% | 0.10% |
| SP1Best vs. SP2Mean | 9.08% | 0.12% | 4.41% | 8.52% | 7.55% | 0.07% | 7.34% | 0.14% | 0.24% | 0.40% | 0.13% |
| SP1Best vs. SP2Best | 9.24% | 0.11% | 6.84% | 8.23% | 4.84% | 0.05% | 7.62% | 0.18% | 0.09% | 0.45% | 0.11% |
| SP1Best vs. SP3Mean | 8.12% | 0.11% | 3.10% | 8.50% | 7.22% | 0.06% | 5.72% | 0.11% | 0.23% | 0.31% | 0.11% |
| SP1Best vs. SP3Best | 11.42% | 0.08% | 5.41% | 8.12% | 4.96% | 0.14% | 1.80% | 0.20% | 0.54% | 0.70% | 0.02% |
| SP2Best vs. SP3Mean | 2.82% | 0.11% | 4.90% | 0.13% | 0.29% | 0.06% | 6.37% | 0.14% | 0.23% | 0.44% | 0.13% |
| SP2Best vs. SP3Best | 5.18% | 0.08% | 6.55% | 0.15% | 0.30% | 0.17% | 7.37% | 0.25% | 0.56% | 0.62% | 0.11% |
| Panel C: Recursive estimation, $h = 12$ | | | | | | | | | | | |
| AR vs. SP1Best | 8.95% | 0.00% | 5.02% | 1.05% | 2.24% | 0.08% | 8.61% | 0.19% | 0.11% | 0.13% | 0.02% |
| AR vs. SP1LBest | 7.59% | 0.04% | 0.86% | 0.63% | 2.67% | 0.08% | 4.75% | 0.19% | 0.19% | 0.00% | 0.03% |
| AR vs. SP2Best | 3.70% | 0.01% | 0.00% | 0.00% | 0.02% | 0.06% | 3.39% | 0.12% | 0.00% | 0.00% | 0.05% |
| AR vs. SP3Best | 1.67% | 0.19% | 3.75% | 0.17% | 0.66% | 0.18% | 3.85% | 0.23% | 0.56% | 0.74% | 0.00% |
| AR vs. SP1Mean | 5.78% | 0.04% | 1.21% | 1.05% | 2.24% | 0.03% | 4.29% | 0.09% | 0.04% | 0.13% | 0.02% |
| AR vs. SP1LMean | 5.94% | 0.04% | 1.92% | 0.63% | 2.67% | 0.03% | 4.75% | 0.09% | 0.05% | 0.24% | 0.03% |
| AR vs. SP2Mean | 1.39% | 0.12% | 2.00% | 0.15% | 0.27% | 0.07% | 3.39% | 0.09% | 0.26% | 0.23% | 0.13% |
| AR vs. SP3Mean | 1.83% | 0.11% | 2.86% | 0.11% | 0.12% | 0.05% | 3.78% | 0.10% | 0.23% | 0.32% | 0.11% |
| SP1Mean vs. SP1Best | 5.45% | 0.04% | 4.39% | 0.00% | 0.00% | 0.10% | 5.58% | 0.25% | 0.08% | 0.00% | 0.00% |
| SP1LMean vs. SP1LBest | 4.14% | 0.00% | 1.74% | 0.00% | 0.00% | 0.10% | 0.00% | 0.25% | 0.16% | 0.24% | 0.00% |
| SP2Mean vs. SP2Best | 2.85% | 0.11% | 2.00% | 0.15% | 0.29% | 0.10% | 0.00% | 0.17% | 0.26% | 0.23% | 0.11% |
| SP3Mean vs. SP3Best | 1.89% | 0.18% | 3.19% | 0.14% | 0.64% | 0.13% | 4.14% | 0.21% | 0.34% | 0.64% | 0.11% |
| SP1Best vs. SP2Mean | 7.61% | 0.15% | 3.89% | 7.93% | 9.25% | 0.09% | 5.35% | 0.15% | 0.23% | 0.31% | 0.13% |
| SP1Best vs. SP2Best | 8.70% | 0.06% | 4.95% | 7.86% | 9.12% | 0.11% | 9.82% | 0.27% | 0.11% | 0.26% | 0.07% |
| SP1Best vs. SP3Mean | 6.80% | 0.14% | 2.79% | 7.89% | 9.13% | 0.07% | 3.56% | 0.15% | 0.19% | 0.31% | 0.10% |
| SP1Best vs. SP3Best | 10.63% | 0.19% | 3.24% | 7.94% | 9.58% | 0.10% | 6.44% | 0.30% | 0.51% | 0.69% | 0.02% |
| SP2Best vs. SP3Mean | 3.18% | 0.15% | 4.42% | 0.10% | 0.53% | 0.08% | 5.56% | 0.17% | 0.21% | 0.38% | 0.09% |
| SP2Best vs. SP3Best | 3.93% | 0.20% | 3.75% | 0.17% | 0.60% | 0.20% | 5.55% | 0.22% | 0.56% | 0.77% | 0.08% |

[a] Notes: See notes to Table 10. Mnemonics used under the header "Models" are defined in Table 10. Numerical entries are average percentage differences when comparing actual and predicted values of a given target variable, defined as follows: $T^{-1}\sum_t \left|\left|\frac{\hat{Y}_{1,t}-Y_t}{Y_t}\right| - \left|\frac{\hat{Y}_{2,t}-Y_t}{Y_t}\right|\right| \times 100\%$, where $Y_t$ is the actual value of the target variable at time $t$. Additionally, $\hat{Y}_{1,t}$ and $\hat{Y}_{2,t}$ are the forecasts from the two models being compared. In rare occurrences where $Y_t$ is zero, it is replaced by $Y_{t-1}$.

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, but only 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 modeling contexts.

## 7. Concluding remarks

In this paper we empirically examine approaches to combining factor modeling methods and robust (shrinkage based) estimation techniques, including bagging, boosting, ridge regression, least angle regression, the elastic net, and the non-negative garotte. In particular, we present the results of a "horse-race" in which mean-square-forecast-error (MSFE) "best" models are found, using reality check and related predictive accuracy tests, for a variety of

forecast horizons, estimation "windowing" schemes and sample periods. Our empirical models include "pure" common factor models, which are estimated using principal components methods, simple linear models, and sophisticated "hybrid" models that combine latent factor modeling approaches with the various shrinkage based techniques outlined above. For the majority of the target variables that we forecast, we find that "hybrid" forecasting models outperform our benchmark and pure factor type models. This suggests that diffusion index methodology is particularly useful when combined with other shrinkage methods (see also, Bai and Ng, 2008, 2009, and Stock and Watson, 2012). We also find that model averaging methods perform surprisingly poorly. 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 modeling approaches. Finally, we find that model performance is rather sensitive to the state of the business cycle. Our "hybrid" models, for example, are most accurate during recessions, while model averaging methods are never chosen as "best" during these times.

## References


Armah, N.A., 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., 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., Marcellino, M., 2005. Factor forecasts for the UK. Jounral of Foreacasting 24, 278–298.

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

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

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

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

Bai, J., Ng, S., 2009. Boosting diffusion indices. Journal of Applied Econometrics 24, 607–629.

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

Boivin, J., Ng, S., 2006. Are more data always better for factor analysis? Journal of Econometrics 132 (1), 169–194.

Breiman, L., 1995. Better subset regression using the nonnegative garrote. Technometrics 37, 373–384.

Breiman, L., 1996. Bagging predictors. Machine Learning 24 (2), 123–140.

Bühlmann, P., Yu, B., 2002. Analyzing bagging. Annals of Statistics 30, 927–961.

Bühlmann, P., 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., Mcculloch, R.E., 2001. The practical implementation of Bayesian model selection. Institute of Mathematical Statistics, 65–134.

Chow, G.C., Lin, A.-L., 1971. Best linear unbiased interpolation, distribution, and extrapolation of time series by related series. The Review of Economics and Statistics 53, 372–375.

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

Connor, G., 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., 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., Korajczyk, R.A., 1993. A test for the number of factors in an approximate factor model. Journal of Finance 48, 1263–1291.

Corradi, V., Swanson, N.R., 2007. Nonparametric bootstrap procedures for predictive inference based on recursive estimation schemes. International Economic Review 48, 67–109.

Diebold, F.X., Lopez, J.A., 1996. Forecast evaluation and combination. In: NBER Technical Working Papers 0192. National Bureau of Economic Research, Inc.

Diebold, F.X., Mariano, R.S., 1995. Comparing predictive accuracy. Journal of Business & Economic Statistics 13, 253–263.

Ding, A.A., 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., Tibshirani, R., 2004. Least angle regression. Annals of Statistics 32, 407–499.

Fernandez, C., Ley, E., Steel, M.F.J., 2001. Benchmark priors for Bayesian model averaging. Journal of Econometrics 100, 381–427.

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

Freund, Y., 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., Croux, C., 2008. Least angle regression for time series forecasting with many predictors, Working Paper. Technical Report. Katholieke Universiteit Leuven.

Hoerl, A., Kennard, R., 1970. Ridge regression: biased estimation for nonorthogonal problems. Technometrics 12, 55–67.

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

Kim, H.H., Swanson, N.R., 2013. Large dataset mining using parsimonious factor and shrinkage methods. Working Paper, Rutgers University.

Koop, G., Potter, S., 2004. Forecasting in dynamic factor models using Bayesian model averaging. Econometrics Journal 7, 550–565.

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

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

Ridgeway, G., Madigan, D., Richardson, T., 1999. Boosting methodology for regression problems. In: The Seventh International Workshop on Artificial Intelligence and Statistics, Uncertainty'99. Morgan Kaufmann, pp. 152–161.

Shrestha, D.L., Solomatine, D.P., 2006. Experiments with adaboost.rt, an improved boosting scheme for regression. Neural Computation 18, 1678–1710.

Stock, J.H., Watson, M.W., 1999. Forecasting inflation. Journal of Monetary Economics 44, 293–335.

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

Stock, J.H., Watson, M.W., 2002b. Macroeconomic forecasting using diffusion indexes. Journal of Business and Economic Statistics.

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

Stock, J.H., Watson, M.W., 2005. Implications of dynamic factor models for var analysis. In: NBER Working Papers 11467 National Bureau of Economic Research, Inc.

Stock, J.H., Watson, M.W., 2006. Forecasting with many predictors. In: Elliott, G., Granger, C., Timmermann, A. (Eds.), Handbook of Economic Forecasting, Vol. 1. Elsevier, pp. 515–554 (Chapter 10).

Stock, J.H., Watson, M.W., 2012. Generalized shrinkage methods for forecasting using many predictors. Journal of Business and Economic Statistics 30, 481–493.

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

Timmermann, A.G., 2006. Forecast combinations. In: Elliott, G., Granger, C., Timmermann, A. (Eds.), Handbook of Economic Forecasting, Vol. 1. Elsevier, pp. 135–196 (Chapter 4).

White, H., 2000. A reality check for data snooping. Econometrica 68, 1097–1126.

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

Wright, J.H., 2009. Forecasting US 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, JMLR Workshop and Conference Proceedings, pp. 660–666.

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