AN EMPIRICAL COMPARISON OF METHODS FOR FORECASTING USING MANY PREDICTORS

December 2004
(revised August 2005)

James H. Stock

Department of Economics, Harvard University
and the National Bureau of Economic Research

and

Mark W. Watson*

Woodrow Wilson School and Department of Economics, Princeton University
and the National Bureau of Economic Research

ABSTRACT

This paper provides a simple shrinkage representation that describes the operational characteristics of various forecasting methods that are applicable when there are a large number of orthogonal predictors (such as principal components). These methods include pretest methods, Bayesian model averaging, empirical Bayes, and bagging. We then compare these and other many-predictor forecasting methods in the context of macroeconomic forecasting (real activity and inflation) using 131 monthly predictors with monthly U.S. economic time series data, 1959:1 - 2003:12. The theoretical shrinkage representations serve to inform our empirical comparison of these forecasting methods.

We thank Jean Boivin, Domenico Giannone, Lutz Kilian, Serena Ng, Lucrezia Reichlin, Mark Steele, and Jonathan Wright for helpful discussions, and Anna Mikusheva for research assistance. This research was funded in part by NSF grant SBR-0214131.

1. Introduction

Over the past ten years, there has been a great deal of research on forecasting using many predictors. Currently available methods include forecast combination, model selection, dynamic factor model forecasts, Bayesian model averaging, empirical Bayes methods, and bagging; for a recent survey, see Stock and Watson (2004a). This wealth of methods raises the question of whether recommendations can be made about which of these methods might be the best starting point in an empirical forecasting exercise, both from a theoretical perspective and based on empirical forecast performance. One difficulty in comparing these methods theoretically is that their derivations generally rely on different modeling assumptions, and it is not clear from these derivations what the algorithms are actually doing when applied in settings in which the modeling assumptions do not hold. Also, although there have been empirical studies of the performance of many of these methods for macroeconomic forecasting, it is difficult to draw conclusions across methods because of differences in data sets and implementation across studies.

This paper therefore has two objectives. The first is to set out some preliminary results on a unified framework for characterizing the properties of forecasting methods applied to many orthogonal predictors (such as principal components of an original set of predictors). The results cover pretest and information-criterion methods, Bayesian model averaging (BMA), empirical Bayes (EB) methods, and bagging. It is shown that asymptotically all these methods have the same "shrinkage" representation, in which the weights in the forecasts are the OLS estimator times a shrinkage factor that depends on the *t*-statistic of that coefficient. These representations are a consequence of the algorithms and they hold under weak stationarity and moment assumptions about the actual statistical properties of the predictors; thus these methods can be compared directly using these shrinkage representations.

The second objective is to report the results of an extensive empirical comparison of these and other methods for forecasting with many predictors. Recent research has focused on three classes of methods to improve forecast accuracy with many predictors. The first class of methods is based on combining a large number of forecasts computed

from relatively simple models. The second class uses a small number of latent factors to summarize the information in predictors, and the third class includes shrinkage, model averaging and model selection methods that reduce sampling error in estimated regression coefficients. Each of these methods has both theoretical and empirical support. The classic papers of Bates and Granger (1969) and Granger and Ramanathan (1984) provide the theoretical case for forecast combining, and this method has proven very successful in practice (see Stock and Watson (2004a) and Timmerman (2004) for references). Principal components and factor analysis are standard tools for descriptive analysis (see Anderson (1984), for example), and have been applied to the large-*n* forecasting problem by Boivin and Ng (2003, 2005), Stock and Watson (2002a,b), Forni, Hallin, Lippi and Reichlin (2003) among others. Shrinkage methods are well known from the classic work of Stein (1955) and James and Stein (1960), and flexible shrinkage methods for the regression model have been developed using Bayesian model averaging methods originally discussed in Leamer (1978) and surveyed in Hoeting, Madigan, Raftery, and Volinsky (1999). Fernandez, Ley and Steele (2001) provide important practical developments so the method can be routinely used in large-*n* regression forecasting with nonorthogonal regressors; Clyde, Desimone, and Parmigiani (1996), Clyde(1999a,b) study the case of orthogonal regressors, and Koop and Potter (2003) and Wright (2004) use versions of large-*n* Bayesian model averaging for macroeconomic forecasting. Finally, "bagging" (Breiman (1996)) is another model averaging scheme, and Inoue and Kilian (2004) report promising results using this method for large-*n* forecasting of inflation.

This paper compares the empirical accuracy of these methods for forecasting U.S. macroeconomic time series over a roughly 30-year out-of-sample period. Using a dataset consisting of 131 monthly macroeconomic economic time series over 1960-2003, nine of the series are forecast using the other series as predictors. Pseudo out-of-sample forecasts are constructed over 1974-2003 for horizons ranging from 1 to 12 months ahead using 52 different large-*n* forecasting models.

The empirical comparison focuses on three questions. First, which models perform best, for which horizons and for which series? Second, do the methods produce very similar forecasts or do the forecasts differ in important ways? Third, if they are

different, how and why to they differ, and can the differences and similarities be explained in light of the shrinkage representations of these methods?

The shrinkage representations for forecasts using orthogonal predictors are described in Section 2. Section 3 provides additional details on the various models used in the empirical comparison, and Section 4 describes the data and the forecasting experiment. As it turns out, the results for the 52 models are well summarized by 14 of the models, and Section 5 presents detailed empirical results for this subset. (Stock and Watson (2004b) contains detailed results for all 52 models.) Section 6 offers some concluding remarks.

2. Shrinkage Representations of Forecasting Methods

We consider the multiple regression model with orthonormal regressors,

$$Y_{t+1} = \delta' P_t + \varepsilon_{t+1}, \quad t = 1,\ldots, T, P'P/T = I_n \tag{1}$$

where P_t is a n-dimensional predictor known at time t with i^{th} element P_{it}, Y_{t+1} is the variable to be forecast, and the error ε_{t+1} has variance σ^2. It is assumed that Y_{t+1} and P_t have sample mean zero. (The extension to multi-step forecasting is discussed below, where we also augment the model with a constant and lagged values of Y.) For the theoretical development it does not matter how the regressors are constructed, however in our applications and in the recent empirical econometric literature they are constructed as the first n principal components, dynamic principal components, or a variant of these methods, using an original, potentially larger set of regressors, $\{X_t\}$. In practice we might imagine constructing $n = 100$ principal components from 200 series and using them to forecast monthly data using 20 years of history for $T = 240$.

With so many regressors, OLS will work poorly so we consider forecasting methods that impose and exploit additional structure on the coefficients in (1). We will show that all these methods have a shrinkage representation, that is, the forecasts from these methods can all be written as,

$$\tilde{Y}_{T+1|T} = \sum_{i=1}^{n} \psi(\kappa t_i)\hat{\delta}_i P_{T+1} + o_p(1), \tag{2}$$

where $\tilde{Y}_{T+1|T}$ is the forecast of Y_{T+1} made using data through time T, $\hat{\delta}_i = T^{-1}\sum_{t=1}^{T} P_{it}Y_{t+1}$ is the OLS estimator of δ_i (the i^{th} element of δ), $t_i = \sqrt{T}\hat{\delta}_i / s_e$, where $s_e^2 = \sum_{t=1}^{T}(Y_{t+1} - \hat{\delta}' P_t)^2 / (T-n)$, and ψ is a function specific to the forecasting method. The factor κ depends on the method. For pretest methods and bagging, $\kappa = 1$. For the Bayes methods, $\kappa = (1 - n/T)^{-1/2}$; this factor arises because the posterior distribution for σ^2 concentrates around the MLE instead of around the degrees-of-freedom adjusted estimator s_e^2.

We refer to (2) as the shrinkage representation of these forecasting methods because, if $0 \leq \psi(x) \leq 1$, the operational effect of these methods is to produce linear combinations in which the weights are the OLS estimator, shrunk towards zero by the factor ψ.

We consider four classes of forecasting procedures: pretest and information criterion methods; Bayesian methods, including Bayesian model averaging, empirical Bayes, and bagging. A key feature is that the proof that the remainder term in (2) is $o_p(1)$ for these different methods relies on far weaker assumptions on the true distribution of (Y, P) than the modeling assumptions used to derive the methods. As a result, the performance of these methods can be understood and analyzed even if they are applied in circumstances in which the original modeling assumptions clearly do not hold, for example when they are applied to multistep-ahead forecasting.

2.1 Pretest and Information Criterion Methods

Because the regressors are orthogonal, a hard threshold pretest for model selection in (2) corresponds to including those regressors with t-statistics exceeding some threshold c. For pretest (PT) methods, the i^{th} coefficient, $\tilde{\delta}_i^{PT}$, is estimated by the OLS estimator if $|t_i| > c$, and is zero otherwise, that is,

$$\tilde{\delta}_i^{PT} = \mathbf{1}(|t_i| > c)\,\hat{\delta}_i . \tag{3}$$

Expressed in terms of (2), the pretest ψ function is,

$$\psi^{PT}(t) = \mathbf{1}(|t| > c). \tag{4}$$

Under some additional conditions, the pretest methods correspond to information criteria methods, at least asymptotically. For example, consider AIC applied sequentially to the increasing sequence of models constructed by sorting the regressors by the decreasing magnitude of their *t*-statistics. If n is fixed and if some of the δ coefficients are fixed while others are in a $n^{-1/2}$ neighborhood of zero, then asymptotically the same regressors will be selected by AIC as by applying the pretest (4) with $c = \sqrt{2}$.

2.2 Bayes Methods

For tractability, Bayes methods in the linear model have focused almost exclusively on the case of strictly exogenous regressors and independently distributed homoskedastic normal errors. For our purposes, the leading case in which these assumptions are used are the Bayesian model averaging (BMA) methods discussed in the next subsection. This modeling assumption is,

$$\{\varepsilon_t\} \perp \{P_t\} \text{ and } \varepsilon_t \text{ is i.i.d. } N(0,\sigma^2) \text{ (strict exogeneity + normality).} \tag{5}$$

We also adopt the usual modeling assumption of squared error loss. Bayes procedures constructed under assumption (5) with squared error loss will be called "Normal Bayes" (NB) procedures. Note that we treat (5) as a modeling tool, where the model is in general misspecified, that is, the actual DGP is not assumed to satisfy (5).

We suppose that the prior distribution specifies that the coefficients $\{\delta_i\}$ are i.i.d., that the prior distribution on δ_i given σ^2 can written in terms of $\tau_i = \sqrt{T}\delta_i / \sigma$, and that $\{\tau_i\}$ and σ^2 have independent prior distributions:

$$\{\tau_i = \sqrt{T}\delta_i / \sigma\} \sim \text{i.i.d } G_\tau(\tau),\ \sigma^2 \sim G_\sigma(\sigma), \text{ and } \{\tau_i\},\ \sigma^2 \text{ are independent (prior)} \qquad (6)$$

If T is fixed, the only two restrictions in (6) are that δ_i is i.i.d. and that σ^2 enters the conditional distribution of δ_i given σ^2 only as a scale factor.

Under squared error loss, the normal Bayes estimator $\tilde{\delta}_i^{NB}$ is the posterior mean,

$$\tilde{\delta}_i^{NB} = E_{\delta,\sigma}(\delta_i | Y, P), \qquad (7)$$

where the subscript $E_{\delta,\sigma}$ indicates that the expectation is taken with respect to δ (which reduces to δ_i by independence under (6)) and σ^2. Under the normality-exogeneity assumption (5), $(\hat{\delta},\ s_e^2)$ are sufficient for $(\hat{\delta},\ \sigma^2\}$. Moreover $\hat{\delta}_i$ and $\hat{\delta}_j$ are independently distributed for all $i \neq j$ conditional on $(\delta,\ \sigma^2)$, and $\hat{\delta}_i | \delta,\ \sigma^2$ is distributed $N(\delta_i,\ \sigma^2/T)$. Thus, conditional on σ, under the normality-exogeneity assumption the posterior mean has the so-called simple Bayes form (cf. Maritz and Lwin (1989)),

$$\tilde{\delta}_i^{NB} | \sigma = \hat{\delta}_i + \frac{\sigma^2}{T}\, \ell_\delta(\hat{\delta}_i), \qquad (8)$$

where $\ell_\delta(x) = \mathrm{d}\ln(m_\delta(x))/\mathrm{d}x$, where $m_\delta(x) = \int \phi_{\sigma/\sqrt{T}}(x-\delta) dG_{\delta|\sigma}(\delta | \sigma)$ is the marginal distribution of an element of $\hat{\delta}$, $G_{\delta|\sigma}$ is the conditional prior of an element of δ given σ, and ϕ_ω is the pdf of a $N(0,\omega^2)$ random variable.

The shrinkage representation of the NB estimator follows from (8) by performing the change of variables $\tau_i = \sqrt{T}\ \delta_i/\sigma$. For priors satisfying (6) and under conditions made precise below, the shrinkage function for the NB estimator is,

$$\psi^{NB}(u) = 1 + \ell(u)/u, \qquad (9)$$

where $\ell(u) = \mathrm{d}\ln m(u)/\mathrm{d}u$, $m(u) = \int \phi(u-\tau)dG_\tau(\tau)$, and ϕ is the standard normal density. Integrating over the posterior distribution of σ^2 results in the posterior mean approaching the MLE for σ^2, which leads to ψ^{NB} being evaluated at $u = \kappa t_i$ in the shrinkage representation (2), with $\kappa = (1 - n/T)^{-1/2}$.

2.3 Bayesian Model Averaging.

Our treatment of BMA with orthogonal regressors follows Clyde, Desimone, and Parmigiani (1996), Clyde(1999a,b), and Koop and Potter (2003). The Clyde, Desimone, and Parmigiani (1996) BMA setup adopts the exogeneity-normality assumption (5) and a Bernoulli prior model for variable inclusion with a *g*-prior for δ conditional on inclusion. Specifically, with probability p let $\delta_i|\sigma \sim \mathrm{N}(0, \sigma^2/g)$ (so $\tau_i \sim \mathrm{N}(0, T/g)$), and with probability $1 - p$ let $\delta_i = 0$ (so $\tau_i = 0$). Note that this prior model satisfies (6). Direct calculations show that, under these priors, the shrinkage representation (9) specializes to

$$\psi^{BMA}(u) = \frac{pb(g)\phi(b(g)u)}{(1+g)[pb(g)\phi(b(g)u) + (1-p)\phi(u)]} \tag{10}$$

where $b(g) = \sqrt{g/(1+g)}$ and ϕ is the standard normal density, and where ψ^{BMA} is evaluated at $u = \kappa t_i$, just as in the general case (9).

2.4 Empirical Bayes

Empirical Bayes (EB) estimation treats the prior *G* as an unknown distribution to be estimated. Under the stated assumptions, $\{\hat{\delta}_i\}$ constitute *n* i.i.d. draws from the marginal distribution *m*, which in turn depends on the prior *G*. Because the conditional distribution of $\hat{\delta}|\delta$ is known in the exogenous-normal model, this permits inference about *G*. In turn, the estimator of *G* can be used in (8) to compute the empirical Bayes estimator. The estimation of the prior can be done either parametrically or nonparametrically. We refer to the resulting empirical Bayes estimator generically as $\tilde{\delta}_i^{EB}$. The shrinkage function for the EB estimator is,

$$\psi^{EB}(u) = 1 + \hat{\ell}(u)/u, \tag{11}$$

where $\hat{\ell}(u)$ is the estimate of the score of the marginal distribution of $\{t_i\}$. This score can be estimated directly or alternatively can be computed using an estimated prior $\hat{G}_\tau$, in which case $\hat{\ell}(t) = \mathrm{d}\ln\hat{m}(t)/\mathrm{d}t$, where $\hat{m}(t) = \int \phi(t-\tau)d\hat{G}_\tau(\tau)$.

2.5 Bagging

Bootstrap aggregation or "bagging" (Breiman (1996)) (BG) smoothes the hard threshold in pretest estimators by averaging over a bootstrap sample of pre-test estimators. Inoue and Kilian (2004) apply bagging to a forecasting situation like that considered in this paper and report some promising results; also see Lee and Yang (2004). Bühlmann and Yu (2002) considered bagging with a fixed number of strictly exogenous regressors and i.i.d. errors, and showed that asymptotically the bagging estimator can be represented in the form (2), where (for $t_i \neq 0$),

$$\psi^{BG}(t) = 1 - \Phi(t+c) + \Phi(t-c) + t^{-1}[\phi(t-c) - \phi(t+c)], \tag{12}$$

where c is the pre-test critical value, ϕ is the standard normal density, and Φ the standard normal CDF. We consider a variant of bagging in which the bootstrap step is conducted using a parametric bootstrap under the exogeneity-normality assumption (5). This algorithm delivers the Bühlmann-Yu expression (12), however the expression obtains under weaker assumptions on the number and properties of the regressors.

2.6 Formal results

We now turn to a formal statement of the validity of the shrinkage representations of the foregoing forecasting methods.

Let p_T denote a vector of predictors used to construct the forecast and let $\{\tilde{\delta}_i\}$ denote the estimator of the coefficients for the method at hand. Then each method produces linear (in p_{iT}) forecasts of the form,

$$\tilde{Y}_{T+1|T} = \sum_{i=1}^{n} \tilde{\delta}_i p_{iT} \tag{13}$$

Formally we treat the predictor vector p_T as nonrandom with finite elements $|p_{iT}| \leq p_{\max}$. Equivalently we could treat p_T as random and distributed independently of the estimation sample, in which case the results would be interpreted as conditional upon its realized value.

The first theorem bounds the difference between the forecast made using the exact forecasting algorithm and its shrinkage representation for the NB and BG forecasts. (It follows immediately from the definition of the pretest estimator that its shrinkage representation $\hat{Y}^{PT}_{T+1|T} = \sum_{i=1}^{n} \psi^{PT}(t_i)\hat{\delta}_i p_{iT}$, where $\psi^{PT}(t) = \mathbf{1}(|t| > c)$, is exact, so this representation is not listed in the theorem.) The bounds in Theorem 1 depend only on the algorithm and modeling assumptions and make no assumptions about the true DGP, so Theorem 1 characterizes the behavior of the forecasting algorithm. The bounds hold for all T, n (subject to some conditions stated in the theorem).

First, we introduce some notation. Let $E()$ without any subscript denote expectation over the true unknown (frequentist) sampling distribution given $\{\delta_i\}$, σ^2, which in general is not the exogenous-normal distribution (5) that is used to compute the estimator. For the Bayes procedure, let $E_\sigma(.|Y,P)$ denote expectations taken over the posterior distribution of σ, calculated using normal likelihood and exogenous regressors and priors that satisfy (6). Finally, let $\hat{\sigma}^2_Y = Y'Y/T$ and let $\hat{\sigma}^2_\varepsilon = Y'[I - P(P'P)^{-1}P']Y/T$ be the MLE of σ^2, so $s^2_\varepsilon = \kappa^2\hat{\sigma}^2_\varepsilon$.

***Theorem* 1.**

(a) *Normal Bayes*. Let $\zeta = \hat{\sigma}^2_\varepsilon / \sigma^2$. Suppose that G_τ and G_σ are such that, for all

$T - n > r_0$,

(i) $|E_\sigma[(\zeta - 1)|Y,P]| \le K_1/(T-n)$,

$E_\sigma[(\zeta - 1)^4|Y,P] \le K_2/(T-n)^2$,

$E_\sigma[\zeta^{-4}|Y,P] \le K_3 - 1$;

(ii) $\sup_u|u^r d^m \psi^{NB}(u)/du^m)| \le M$ for $r, m = 1, 2$.

Then, for all (T, n) such that $T - n > r_0$,

$$E\left[\hat{Y}^{NB}_{T+1|T} - \sum_{i=1}^{n}\psi^{NB}(\kappa t_i)\hat{\delta}_i p_{iT}\right]^2 \le p^2_{\max}\,\frac{1}{n}\sum_{i=1}^{n}(a_i^{NB})^2\,,$$

where $\kappa = (1 - n/T)^{-1/2}$,

$$\psi^{NB}(t) = 1 + \frac{\ell(t)}{t},\ \ell(t) = \frac{\mathrm{d}\ln m(t)}{\mathrm{d}t},\ m(t) = \int \phi(t-\tau)dG_\tau(\tau)\,,\ \text{and}$$

$$\left|a_i^{NB}\right| \le \frac{n}{(T-n)^{3/2}}\left[\frac{1}{2\kappa}\sqrt{E\left(K_1^2M^2\hat{\sigma}_Y^2\right)} + \frac{1}{4\sqrt{2}}\sqrt{E\left(K_2K_3M^2t_i^2\hat{\sigma}_Y^2\right)}\right].$$

(b) *Bagging*. Let B be the number of parametric bootstrap draws used to construct the bagging estimator. Then for all (T, n) such that $T - n > 8$,

$$E\left[\hat{Y}^{BG}_{T+1|T} - \sum_{i=1}^{n}\psi^{BG}(t_i)\hat{\delta}_i p_{iT}\right]^2 \le p^2_{\max}\,\sqrt{E\hat{\sigma}_Y^4}\,\frac{1}{n}\sum_{i=1}^{n}(a_i^{BG})^2\,,$$

where

$$\psi^{BG}(t) = 1 - \Phi(t+c) + \Phi(t-c) + t^{-1}[\phi(t-c) - \phi(t+c)],$$

$$|a_i^{BG}| \le \sqrt{\frac{n^2}{B(T-n)}}\sqrt{m_{T-n}(c)} + \sqrt{\frac{n^2}{(T-n)^3}}K_4\mu_{T-n}\left[E\left(\sum_{m=0}^{3}|t_i|^m\right)^4\right]^{1/4},$$

$m_r(c) = \max_\mu \mathrm{var}[(z+\mu)\mathbf{1}(|z+\mu| > c\sqrt{\xi}\,)] < \infty$, where $z \sim \mathrm{N}(0,1)$,

$\xi \sim \chi^2_r / r$, and z and ξ are independent,

$$\mu_r = \left[\frac{r+4}{r} + \frac{r^3(r+4)}{(r-2)(r-4)(r-6)(r-8)}\right]^{1/2},\ \text{and}\ K_4 = 28e^{-2}\sqrt{3/\pi}\,.$$

Proofs are given in the Appendix.

The rates conditions in (i) in Theorem 1(a) specify rates at which the posterior distribution of σ^2 concentrates around the MLE $\hat{\sigma}^2_\varepsilon$. These rates are consistent with those

arising in a Bernstein-von Mises theorem and specify rate. Conditions (*ii*) bound the behavior of the derivatives of ψ^{NB}. In general the terms M, K_1, K_2, and K_3 are constants conditional on the data but depend on $(\hat{\delta}, \hat{\sigma}_\varepsilon^2)$, where the specific form of the dependence is determined by the priors. For example, consider the usual conjugate priors, that is, τ_i i.i.d. N(0, T/g) and $1/\sigma^2 \sim \Gamma(\alpha, 1/\beta)$. In this case, condition (*ii*) is satisfied because ψ^{NB} is constant (the amount of shrinkage does not depend on t_i), so $M = 0$. The posterior for $\zeta = \hat{\sigma}_\varepsilon^2/\sigma^2$ is $\Gamma(a, 1/b)$, where $a = \alpha + T/2$ and $b = \beta/\hat{\sigma}_\varepsilon^2 + T/2 + [2\kappa^{-2}(1/T + 1/g)]^{-1} T^{-1}\sum_i t_i^2$. In the case of diffuse priors $g = \beta = 0$, the posterior for ζ is $\Gamma(\alpha + T/2, 2/T]$, so r_0, K_1, K_2, and K_3 are constants that do not depend on $(\hat{\delta}, \hat{\sigma}_\varepsilon^2)$.[1]

The next result provides high-level conditions on the DGP (that is, on the true distribution of $(Y, P)|(\delta, \sigma)$), and on the rate at which the number of regressors grows as T increases, under which the shrinkage representations provide asymptotically valid approximations to the forecasting algorithm.

***Theorem* 2.** Suppose that the conditions of Theorem 1 are satisfied and additionally that (*i*) $n/T \to \nu$, $0 \le \nu < 1$. Then:

(a) *Normal Bayes*. Suppose, in addition to (*i*) that (*ii*) $E\left(K_1^2 M^2 \hat{\sigma}_Y^2\right) < \infty$ and (*iii*) $E\left(K_2 K_3 M^2 t_i^2 \hat{\sigma}_Y^2\right) < \infty$. Then as $T \to \infty$,

$$E\left[\hat{Y}_{T+1|T}^{NB} - \sum_{i=1}^{n} \psi^{NB}(\kappa t_i)\hat{\delta}_i p_{iT}\right]^2 \to 0.$$

(b) *Bagging*. Suppose, in addition to (*i*), that (*iv*) $B/n \to \infty$ and (*v*) $E\hat{\sigma}_Y^4 < \infty$, and (*vi*) $\max_i E t_i^{12} < \infty$. Then as $T \to \infty$,

$$E\left[\hat{Y}_{T+1|T}^{BG} - \sum_{i=1}^{n} \psi^{BG}(t_i)\hat{\delta}_i p_{iT}\right]^2 \to 0.$$

We make several comments.

[1] Specifically, for $g = \beta = 0$, let $r_0 = 8$ to ensure that $E_\sigma(\sigma^2|Y,P)^{-4}$ exists. Then $K_1 = 2\alpha$, $K_2 = (2 + \alpha/2)^2\{3 + [5 + 4\alpha + (1 - \alpha)^4]/(\alpha + 4)\}$, and $K_3 = \max[1,(4/\alpha)^4]$.

1. Theorem 2 states that the validity of the shrinkage factor representations require much weaker assumptions than those upon which the models are based, specifically Y and the t-statistic have sufficiently many moments. In this sense, the shrinkage factor representations are consequences of the algorithm, not properties of the true DGP. In particular, the frequentist risk (under squared error loss) of the forecasts is the same whether the regressors are strictly or weakly exogenous, and whether the errors have a normal distribution or not, as long as the conditions of Theorem 2 are satisfied. This observation serves to generalize the narrow optimality results for BMA and NB to cases of interest in time series, in particular with predetermined by not strictly exogenous regressors.
2. The results hold when the number of regressors is fixed or when n increases proportionally to the sample size. If $n/T \to 0$, then $\kappa \to 1$ and this factor drops out of the representation (2). However, if n increases proportionately to T, then $\kappa \to (1 - \nu)^{-1/2} > 1$ for the Bayes representations. This arises because the posterior concentrates around the MLE, which equals $\kappa^2 s_e^2$, but s_e appears in the OLS t-statistic. This same factor would arise in the expression for bagging if bagging were implemented with the t-statistic computed with $\hat{\sigma}_\varepsilon$ in the denominator, not s_e, however that is not the way bagging has been implemented in the literature.
3. As discussed following the statement of Theorem 1, the specific moment conditions under part (*a*) depend on the priors G_τ and G_σ.
4. Condition (*vi*) in Theorem 2, that t_i has twelve moments, might be stronger than necessary arises in the proof through a series of convenient inequalities.
5. The results presented for the general normal Bayes estimator will extend to an empirical Bayes estimator under suitable conditions. The results in Theorem 1 are conditional on the data (no expectations are taken over (Y, P)) so the results in part (a) apply directly to an empirical Bayes estimator, where G_τ now is interpreted as involving estimated parameters. The results in Theorem 2 have expectations over the data and now the shrinkage function ψ^{NB} is a sequence of shrinkage functions, which depend on the estimated prior. Under suitable regularity conditions, if the empirical Bayes estimation step is consistent then the

asymptotic empirical Bayes shrinkage representation ψ^{EB} is the ψ^{NB} with the probability limit of the estimated prior replacing G_τ.

6. These representations permit the extension of these methods to direct multistep forecasting. In a multistep setting, the errors have a moving average structure. However the forecasting methods can be implemented by substituting HAC *t*-statistics, OLS coefficient estimators into the shrinkage representations.
7. The shrinkage factor representation of bagging allows us to ascertain whether bagging is asymptotically admissible, a result that is currently unavailable. Setting ψ^{BA} equal to ψ^{NB} yields the integral-differential equation,

$$\left.\frac{\mathrm{d}\ln\int \phi(z-s)dG_\tau(s)}{\mathrm{d}z}\right|_{z=t} = t[\Phi(t-c)-\Phi(t+c)] + \phi(t-c)-\phi(t+c). \qquad (14)$$

If there is a proper prior G_τ that satisfies (14), then this is the prior for which bagging is asymptotically Bayes, in which case bagging would be asymptotically admissible. Let G_τ have density *g* and characteristic function $h(s) = \int e^{ist} g(t)dt$. Then *g* satisfies (14) if it is a density and if *h* satisfies the Fredholm equation of the second kind,

$$h(s) = \int K(s,t)h(t)dt, \qquad (15)$$

where

$$K(s,t) = 2\frac{e^{-t^2+st}}{s}\left[\frac{\sin(c(s-t))}{(s-t)^2} - c\frac{\cos(c(s-t))}{s-t}\right]. \qquad (16)$$

3. Empirical Comparisons: Details of Forecasting Methods

We now turn to an empirical comparison of various many-predictor forecasting methods. This section describes the specifics of the forecasting methods, including the specifics of the transformations from the original predictors (X) to the orthonormal predictors (P). We also describe some benchmark methods that will be used in the univariate analysis, as well as two large-n methods that are implemented directly to the original X predictors (forecasting combination and non-orthonormal g-prior BMA). This section also explains multistep forecasting and the treatment of the intercept and lags of Y included as predictors, issues that were suppressed in the development of Section 2. The extant theory for many of the forecasting methods assumes that X and Y are I(0) variables, so that many of the series are first or second differences of the raw data series (or the logarithm of the raw series); the specific transformation for each series is given in the data appendix.

Let h denote the forecast horizon and Y_{T+h}^h denote the I(0) horizon-specific transformation of Y to be forecast. (For example, when forecasting the monthly index of industrial production, IP, $Y_{T+h}^h = (1200/h)\times\ln(\text{IP}_{t+h}/\text{IP}_t)$, the h-period growth rate expressed in percentage points at an annual rate.) Forecasts of Y_{T+h}^h are constructed by estimating the projections of Y_{T+h}^h onto X_T and lags of Y_T, so that the forecast of Y_{T+h}^h at time T, denoted $Y_{T+h/T}^h$, is given by $Y_{T+h/T}^h = \alpha + \beta' X_T + \phi(\text{L})Y_T$, where $(\alpha, \beta, \phi(\text{L}))$ contain parameters to be estimated by the competing methods.

The remainder of this section discusses the specific forecasting methods used in the empirical comparison. The discussion focuses on methods for forecasting Y_{T+h}^h using data through period T; the next section discusses the recursive forecasting experiment in which T varies through the pseudo out-of-sample period. Discussion of selection of AR lag length, number of factors in the factor model, and computational details are also postponed until the next section.

3.1 Two Benchmarks

Two benchmark method are included in the forecast comparision.

Univariate autoregressions. Forecasts from the univariate AR(p) model are computed as $\hat{Y}_{T+h/T}^{AR,h} = \hat{\alpha} + \hat{\phi}(\mathrm{L})Y_T$ where the coefficients are estimated by the OLS regression of Y_{t+h}^h onto (1, Y_t, Y_{t-1}, … , Y_{t+1-p}) for t = 1,…, $T-h$, where pre-sample values of Y_0, Y_{-1},…, Y_{1-p} are used to initialize the regression. This sample period will be used for all of the methods described below.

OLS using all predictors. The forecast is constructed as $\hat{Y}_{T+h,T}^{OLS,h} = \hat{\alpha} + \hat{\beta}' X_T + \hat{\phi}(\mathrm{L})Y_T$, where the parameters are estimated by the OLS regression of Y_{t+h}^h onto (1, X_t, Y_t, Y_{t-1},…, Y_{t+1-p}).

3.2 Combined Bivariate ADL models

These forecasts are constructed by combining forecasts computed from bivariate autoregressive distributed lag (ADL) models. The i^{th} ADL model includes $p_{i,x}$ lags of $X_{i,t}$, $p_{i,y}$ lags of Y_t, and has the form $\hat{Y}_{T+h/T}^{ADL_i,h} = \hat{\alpha}_i + \hat{\beta}_i(\mathrm{L})X_{i,T} + \hat{\phi}_i(\mathrm{L})Y_T$, where the parameters are estimated by the OLS regression of Y_{t+h}^h onto (1, $X_{i,t}$, $X_{i,t-1}$,…, $X_{i,t+1-p_{i,x}}$, Y_t, Y_{t-1},…, $Y_{t+1-p_{i,y}}$). The combined forecast is $\hat{Y}_{T+h/T}^{Comb,h} = \sum_{i=1}^{n} w_i \hat{Y}_{T+h/T}^{ADL_i,h}$ where w_i denotes the weight given to the i'th forecast. The weights are chosen in one of two ways.

Simple Averages. These methods use simple averages such as the mean (w_i = $1/n$), median, or trimmed mean for the combined forecasts.

Weights based on in-sample fit. It seems natural to choose weights in a way that reflects, at least in part, the relative in-sample fit of the corresponding ADL models. There are several methods to do this; Wright (2003) suggests a method based on a particular Bayesian Model Averaging (BMA) model that has some theoretical appeal, and this method will be used in the comparison below. When $p_{i,x} = p_x$ and $p_{i,y} = p_y$ for all i, this yields weights

$$w_i \propto \left[\frac{1}{\lambda SSR_{AR} + (1-\lambda) SSR_{ADL,i}} \right]^{\left(\frac{T-1-p_y}{2}\right)} \qquad (17)$$

where SSR_{AR} and $SSR_{ADL,i}$ denote the sum of squared residuals from the AR and i'th ADL model, $1+p_y$ denotes the number of estimated coefficients in the AR model, and λ is a parameter that is between 0 and 1. When $\lambda = 1$, each model receives the same weight, and as λ decreases better fitting models receive greater weight.

3.3 Factor Models

The factor model forecasts are based on the model

$$X_t = \Lambda F_t + e_t \tag{18}$$

$$Y_{t+h}^h = \alpha + \gamma' F_t + \phi(\mathrm{L}) Y_t + u_{t+h}. \tag{19}$$

Equation (18) describes the $n\times 1$ vector X_t using a $k\times 1$ vector of unobserved factors F_t and an error term e_t. Empirical content is given to this relation by assuming that k is much smaller than n and the elements of e_t are only weakly correlated; this implies that covariances between the (many) elements of X_t are explained, in large part, by the (few) factors in F_t. The model is described in detail in Stock and Watson (2004b) and the references cited there. Equation (19) describes Y_{t+h}^h using the factors, autoregressive lags, and an forecast error u_{t+h} that is assumed to satisfy $E(u_{t+h} | \{X_\tau, Y_\tau\}_{\tau \le t}) = 0$. Because X_t does not enter (19) directly, the elements of X_t are useful for predicting Y_{t+h}^h only because they contain information about F_t. Equation (19) is a factor-augmented autoregression (FAAR), and forecasts are computed as

$$\hat{Y}_{T+h/T}^{FAAR,h} = \hat{\alpha} + \hat{\gamma}' \hat{F}_T + \hat{\phi}_T(L) Y_T \tag{20}$$

where $\hat{F}_t$ denotes the estimate of F_t constructed from $\{X_j\}_{j=1}^T$ and $\hat{\alpha}$, $\hat{\gamma}_T$ and $\hat{\phi}_T(L)$ are obtained from the OLS regression of Y_{t+h}^h onto $(1, \hat{F}_t, Y_t, \ldots, Y_{t-p+1})$. Three estimators of F_t are considered in the forecasting comparison.

FAAR-OLS. Stock and Watson (2002a,b) proposed forecasts as in (20) using the principal components estimator of F_t. The principal components estimator solves the least squares problem:

$$\min_{F,\Lambda,F'F=I_p} \sum_{t=1}^{T} (X_t - \Lambda F_t)'(X_t - \Lambda F_t) . \tag{21}$$

Because of the form of the objective function, forecasts based on these estimates of F_t are labeled as FAAR-OLS.

FAAR-GLS. Forni, Hallin, Lippi, and Reichlin (2003) proposed a generalized principal component for F_t that solves the GLS problem

$$\min_{F,\Lambda,F'F=I_p} \sum_{t=1}^{T} (X_t - \Lambda F_t)'\hat{\Omega}^{-1}(X_t - \Lambda F_t) \tag{22}$$

where $\hat{\Omega}$ is an estimator of $\Omega = \text{var}(e_t)$ constructed from frequency domain principal component analysis. Forecasts using this estimator will be labeled FAAR-GLS.

FAAR-WLS. Boivin and Ng (2004) and Forni et al (2003) suggest an alternative to FAAR-GLS that instead solves (22) using a diagonal version of $\hat{\Omega}$. That is, the estimator solves the weighted least squares version of (22). This estimator will be labeled FAAR-WLS.

3.4 Bayesian Model Averaging, Variable Selection, and Other Shrinkage Methods

Bayesian model averaging with non-orthogonal predictors. To begin, consider the forecasting problem for $h = 1$, and for notational ease write $Y_{t+1}^1 = Y_{t+1}$. The Bayesian model averaging (BMA) methods used here begin with the regression model

$$Y_{t+1} = \lambda' W_t + \theta' Z_t + u_{t+1} \tag{23}$$

where $W_t = (1\ Y_t \ldots Y_{t+1-p})'$ includes the constant and AR components and Z_t contains the elements of X, transformed so that they are orthogonal to W. (Using obvious notation $Z =$

$(I - W(W'W)^{-1}W')X$.) The formal Bayesian analysis uses a likelihood derived from the assumption that u_t is iid N(0, σ^2) and priors for the parameters (λ, θ, σ) to construct the predictive density of Y_{T+1}; the mean of the predictive density is the minimum mean square error forecast of Y_{T+1}.

Fernandez, Ley, and Steele (2001a) survey BMA methods, and propose useful computational methods and benchmark priors for the BMA regression model. Uninformative (flat) priors are used for (λ, σ): $\pi(\lambda, \sigma) \propto 1/\sigma$, where π denotes the prior density. The prior for θ is more complicated and captures the notion that many of the values of θ_i are likely to be zero, or equivalently, many of the regressors can be excluded from the model. It is useful to represent θ_i as the product $\theta_i = \rho_i\gamma_i$, where γ_i is a 0-1 binary variable. Let γ denote the $n\times1$ vector of γ_i's. The vector γ can take on $m = 2^n$ possible values, which can be listed as γ^1, γ^2,…, γ^m. Let S_j denote the selection matrix that selects the non-zero elements of γ^j from γ, and let $\rho^j = S_j\rho$ and $Z_t^j = S_jZ_t$, so that the vector ρ^j contains the non-zero regression coefficients and Z_t^j are the corresponding regressors. Conditional on $\gamma = \gamma^j$, the model is $Y_{t+1} = \lambda'W_t + \rho^{j\prime} Z_t^j + u_{t+1}$, and a standard g-prior (Zellner (1986)) is used for ρ^j: specifically, $\rho^j|\gamma = \gamma^j, \sigma \sim N(0, \sigma^2[gZ^{j\prime}Z^j]^{-1})$.

Forecasts for the conditional model are then straightforward to derive: conditional on $\gamma = \gamma^j$, the forecast is given by $\hat{Y}_{T+1/T}^j = \hat{\lambda}'W_T + \frac{1}{1+g}\hat{\rho}^{j\prime}Z_T^j$, where $(\hat{\lambda}, \hat{\rho}^j)$ is the OLS estimator from the regression of Y_{t+1} onto (W_t, Z_T^j). The BMA forecast is the weighted average of these forecasts: $\hat{Y}_{T+1/T}^{BMA} = \sum_{j=1}^{m} w_j\hat{Y}_{T+1/T}^j$, where $w_j = \text{Prob}(\gamma = \gamma^j| Y)$ is the posterior probability that $\gamma = \gamma^j$, which, from Bayes theorem, satisfies $w_j \propto L(Y|\gamma=\gamma^j)P(\gamma = \gamma^j)$ where $L(Y|\gamma=\gamma^j)$ denotes is the likelihood of the data conditional on $\gamma = \gamma^j$, marginalized with respect to (λ, ρ^j, σ). Following Fernandez, Ley, and Steele (2001a),

$$L(Y|\gamma = \gamma^j) \propto a(g)^{k_j/2}[a(g)SSR^R + (1-a(g))SSR_j^U]^{-\frac{p+1}{2}} \qquad (24)$$

where $a(g) = g/(1+g)$, SSR^R is the sum of squared residuals from the regression of Y_{t+1} onto W_t, SSR_j^U is the sum of squared residuals from the regression of Y_{t+1} onto $(W_t,\ Z_t^j)$, and k_j is the number of regressor in Z_t^j. The prior $P(\gamma = \gamma^j)$ follows from an assumption that γ_i are iid with $P(\gamma_i = 1) = p$. When $p = ½$, each model is equally likely, so that the w_j is proportional to $L(Y|\gamma=\gamma^j)$.

Evidently, calculation of $\hat{Y}_{T+1/T}^{BMA}$ requires estimation of each of the $m = 2^n$ models, and this is infeasible when n is large.[2] However, $\hat{Y}_{T+1/T}^{BMA}$ can be very well approximated using simulation methods to select models with high posterior probability. As suggested by Fernandez, Ley, and Steele (2001) the analysis here uses the MC^3 algorithm of Madigan and York (1995) for the simulations. Details are provided in the next section.

Bayesian model averaging using orthogonal predictors. As discussed in Clyde, Desimone, and Parmigiani (1996), Clyde(1999a,b), and Koop and Potter (2003), BMA estimation is greatly simplified when the regressors are orthogonal. Thus, let P_t denote the standardized principal components of Z; that is, $P_t = HZ_t$, where $T^{-1}\sum_t P_t P_t' = I_n$. The regression model (23) can then be written as

$$Y_{t+1} = \lambda' W_t + \delta' P_t + u_{t+1}. \tag{25}$$

where and let $\delta' = \theta' H^{-1}$. The BMA forecast is the extension of the method discussed in Section 2.2, specifically,

$$\hat{Y}_{T+1/T}^{BMA-PC} = \hat{\lambda}' W_T + \sum_{i=1}^{n} \psi^{BMA}(t_i)\hat{\delta}_i P_{i,T}\ , \tag{26}$$

where ψ^{BMA} is given in (10) and $\hat{\lambda}$ is the OLS estimator of λ. (Note that this formulation uses the plug-in approach for σ^2, rather than integrating over the posterior.)

[2] With $n = 130$, estimation of all of the models would take approximately 4×10^{28} years using a computer that could estimate 1000 models per second.

Empirical Bayes model averaging using principal components. The empirical Bayes forecast is,

$$\hat{Y}_{T+1/T}^{EB-PC} = \hat{\lambda}' W_T + \sum_{i=1}^{n} \psi^{EB}(t_i)\hat{\delta}_i P_{i,T}\,, \tag{27}$$

where ψ^{EB} is given by (7). The EB forecast was implemented for the BMA mixture distribution, which implies the marginal distribution, $\sqrt{T}\hat{\delta}_i \sim \text{N}(0, \sigma^2+g^{-1})$ with probability p, and $\sqrt{T}\hat{\delta}_i \sim \text{N}(0, \sigma^2)$ with probability $(1-p)$. The empirical Bayes model averaging estimator constructs the mixed normal MLE's of σ^2, g, and p from $\{\hat{\delta}_i\}$.

Model selection using principal components. The pretest/information criterion forecasts have the form,

$$\hat{Y}_{T+1/T}^{IC-PC} = \hat{\lambda}' W_T + \sum_{i=1}^{n} \psi^{IC}(t_i)\hat{\delta}_i P_{i,T}\,, \tag{28}$$

where $\psi^{IC}(\tau) = \mathbf{1}(\tau^2 > 2)$ for AIC and $\psi^{IC}(\tau) = \mathbf{1}(\tau^2 > \ln(T))$ for BIC.

Bagging using principal components. The bagging forecast is

$$\hat{Y}_{T+1/T}^{Bagging-PC} = \hat{\lambda}' W_T + \sum_{i=1}^{n} \psi^{Bagging}(t_i)\hat{\delta}_i P_{i,T} \tag{29}$$

where $\psi^{Bagging}$ is given by (12). Our empirical results use the asymptotic representation (29) and thus avoid the need for the bootstrap calculations that normally are part of the bagging computation.

Multistep forecasts. The likelihood function underlying the BMA methods is based on the assumption that the regression error u in (25) is iid $\text{N}(0,\sigma^2)$, but when $h > 1$, u is serially correlated, and eliminating the serial correlation using a GLS-like transformation is inappropriate because the regressors are not strictly exogenous. Thus, when $h > 1$, the BMA forecasts will not have their usual theoretical justification as the

minimum mean square error predictor associated with the relevant prior distribution. However, an alternative and less formal interpretation of the BMA methods is that they are methods for generating sensible shrinkage estimators for the regressions coefficients, which in turn may produce accurate forecasts. Under this interpretation, it is interesting to see how the compare to other forecasting methods even when $h > 1$, and this interpretation motivates the empirical analysis reported below.

4. Description of the Forecasting Experiment

4.1 Data and Sample Period

The data are monthly observations on 131 U.S. macroeconomic time series from 1959:1 through 2003:12. Forecasts are constructed for nine of these series (four measures of real activity, two interest rates, and three price indices), and all 131 series are used as predictors. As discussed above, the series are transformed by taking logarithms and/or differencing. In general, first differences of logarithms (growth rates) are used for real quantity variables, first differences are used for nominal interest rates, and second differences of logarithms (changes in growth rates) for price series. The forecasted variable, Y_{t+h}^h, is the rate of growth of the real activity variables, the change in the interest rates, and the change in the rate of price inflation over the forecast period. Thus, $Y_{t+h}^h = (1200/h)\times\ln(IP_{t+h}/IP_t)$ for the index of industrial production (IP), $Y_{t+h}^h = TBILL_{t+h} - TBILL_t$ for the interest rate on three month Treasury bills (TBILL), and $Y_{t+h}^h = 1200[(1/h)\ln(CPI_{t+h}/CPI_t) - \Delta\ln(CPI_t)]$ for the consumer price index (CPI),[3] and so forth. Table 1 shows the definition of Y_{t+h}^h and Y_t for each of the nine variables that are forecast.

[3] Several of the models were also computed using an alternative transformation of the price series: $Y_{t+h}^h = (p_{t+h} - p_{t+h-12}) - (p_t - p_{t-12})$, where p_t denotes the logarithm of the price index. This transformation is used in Giannone, Reichlin, and Sala (2004), and, with $h = 12$ and no AR lags, is the model used by Atkeson and Ohanian (2001). Averaging across series and over the entire pseudo out-of-sample period, the AR forecasts that used this transformation performed slightly worse than the forecasts that used the transformation discussed in the text for horizons $h = 1$, 3, 6, and 12 when lag lengths were chosen by AIC; they also performed worse for $h = 6$ and 12 when BIC was used. When the out-of-sample period was restricted to post 1983, and again averaging across the four price series, the alternative transformation performed marginally better for $h = 1$ and 3 when AIC was used, but marginally worse for $h = 6$ and 12. Similar results were found for the other forecasting models.

Psuedo out-of-sample forecasts are calculated for each variable and method and for horizons h = 1, 3, 6, and 12 months. The pseudo out-of-sample forecasting period begins in T = 1974:7 and continues through T = 2003:12−h. Forecasts constructed at date T are based on models that are estimated using data dated T and earlier; the estimation period for recursive forecasts begins in1960:1 with previous values used to initialize the autogressions. Models were also estimated using rolling samples containing the most recent 120 periods. In this case, the OLS, BMA, BIC, Bagging and FAAR models were constructed using 60 predictors that were randomly selected from the original set of 131.

4.2 Estimation Details

AR lags. All of the models include AR lags of Y and this requires a choice of the lag length. For most of the models the lag length was chosen by several different methods: recursively computed AIC and BIC (with $0 \le p \le 12$), and pre-specified values p=4 and p=12. For the BMA methods it was computationally convenient to used fixed values, and these models were estimated using 4 and 12 lags.

FAAR models. The number of factors for the FAAR forecasts was determined three different methods: recursive AIC and BIC ($0 \le k \le 10$) applied to the forecasting equation, and with a pre-specified value of k = 3. The GLS principal component estimator used the estimator of $\hat{\Omega}$ described in Forni, et al (2003) and their accompanying software. Specifically, let x_t denote the standardized values of X_t. The estimated spectrum of x, $S_{xx}(\omega)$, is computed at 101 equally spaced ordinates using a Bartlett kernel applied to $p = T^{1/2}$ sample autocovariances. The estimated spectrum of the dynamic factor components, $S_{cc}(\omega)$, is computed for each of the 101 frequencies using k^{dyn} dynamic principal components of $S_{xx}(\omega)$; as suggested by Gianone, Reichlin, and Sala (2004), the empirical analysis uses $k^{dyn} = 2$. The estimated value of Ω is computed as $\hat{\Omega} = S_{xx} - S_{cc}$, where S_{xx} is the sample second moment matrix of x and S_{cc} is the inverse fourier transform of $S_{cc}(\omega)$. The diagonal elements of this estimator were used for the WLS estimator.

Combined ADL models. Each bivariated ADL model was estimated with β_i(L) and ϕ(L) lag lengths chosen by recursively computed AIC and BIC (with $0 \le p_x \le 12$ and

$0 \le p_y \le 12$) and with fixed values $p_x = p_y = 4$. The simple combining methods were computed using the mean, median and 2-percent-symmetrically-trimmed mean of the ADL forecasts. The combined forecasts that used weights given in (17) used the ADL models with fixed lag lengths, and were computed using a grid of fixed values of λ between 0 and 1, and with λ estimated recursively by predictive least squares (PLS), where the first PLS out-of-sample forecast period is T = 1970:12.

BMA. The priors for the BMA models required specification of the parameters p and g (see the discussion following (23)), and these were chosen based on suggestions in Fernandez, Ley and Steele (2002). The parameter p was set to 0.5, so that all constituent models in BMA had an equal prior probability; g was chosen so that the priors were very diffuse ($g = 1/T$ or $g = 1/n^2$) or informative ($g = 1.0$). The BMA forecasts were computed using a relatively small number of MC^3 simulations. For each series and horizon, 10^7 initial "burn-in" simulations were carried out before the first forecast period. Then, for each forecast period, 10^6 simulations were performed, where the final value for forecast period T served as the initial value for period T+1. This process was carried out using two different randomly chosen initial values. The resulting forecasts were, for the purpose of the analysis carried out in this paper, essentially identical. Stock and Watson (2004b) addresses this is more detail.

Bagging The t-statistics used for the shrinkage factors were computed using the usual OLS formula and with Newey-West standard errors (with a lag length parameter equal to the forecast horizon). The pretest critical value for bagging was set at $c = 1.96$ (5% pretest) and $c = 2.58$ (1% pretest).

Forecasts reported. The combination of different models with procedures for estimating lag lengths, factors, and so forth led to 52 distinct forecasts for each date, series, and horizon. Examination of these forecasts suggested that their key features were well represented by the 14 forecasts described in Table 2. Results for this subset of the forecasts are presented in the next section; detailed results for all of the 52 forecasts can be found in Stock and Watson (2004b).

5. Relative Performance of the Forecasts

5.1 Forecast Performance and Stability

Table 3 summarizes the pseudo out-of-sample forecasting performance for the 14 methods and for each of the 9 series forecast at the 4 forecast horizons. Panel A shows the results for $h = 1$, panel B for $h = 3$, and so forth. The first row of each panel shows the root mean square forecast error (root-MSFE) for the combined ADL forecast using simple average ($w_i = 1/n$) combining weights; the forecast is labeled Combined-Mean in the table. This forecast will serve as a benchmark in what follows, so that the other forecasts will be evaluated relative to this forecast. As documented in the survey paper by Timmerman (2004), this is formidable benchmark that typically outperforms univariate autoregressions and many multivariate methods.

The other rows in each of the panels of Table 3 show the MSFE of the various forecasting methods relative to the MSFE for the benchmark. For example, in Panel A (h =1), the value of 1.04 in the row labeled AR and the column labeled PI indicates that the ratio of MSFE for the AR forecast relative to the MSFE for the Combined-Mean forecast was 1.04 for the series PI (Personal Income). Thus, the MSFE for the AR forecast was 4% higher than the corresponding MSFE for the Combined-Mean forecast.

Several results can be gleaned from Table 3. For example, the Combined-Mean forecast uniformly dominates the AR, OLS and Bagging forecasts across all series and horizons. The OLS forecasts are particularly poor, often with MSFE's 2 or even 3 times larger than the benchmark. The Bagging forecasts also perform poorly, with relative MSFE's 14%-60% larger than the benchmark for $h = 1$, and 54%-142% larger for $h = 12$. The FAAR forecasts generally outperform the benchmark, however there are a few exceptions (for example, the relative MSFE for FAAR-OLS and FAAR-WLS are 1.34 and 1.38 for IP at $h = 12$, and the relative MSFE for FAAR-GLS is 1.23 for EMP at h = 1). As a group, the BMA forecasts outperform the benchmark for $h = 1$, but their relative forecasting performance deteriorates as the forecast horizon increases. Indeed, as the horizon increases there is a general deterioration of all of the forecasts relative to the benchmark: in Panel A ($h = 1$) 51 of the 117 of the reported relative MSFE's are less than 1.0; this falls to 17 in Panel D ($h = 12$).

Table 4 summarizes the detailed results from Table 3 and investigates the stability of the conclusions across sub-samples of the out-of-sample period. The second column

of Table 4 shows the relative MSFE for each method averaged across the 9 series and 4 horizons; it also shows the fraction of the 36 series/horizon relative MSFE's that are less than 1.0. For example, the average relative MSFE for the AR forecast was 1.10 and none of the 36 entries in Table 3 for this forecast were less than 1.0. The final two columns repeat these calculations for the first and second half of the out-of-sample period. Averaged across all series and horizons, the FAAR methods performed best over the full out-of-sample period. For example, the average relative MSFE for FAAR-OLS was 0.96, and this method outperformed the benchmark in 81% of the series/horizon forecasts. The next-best performing forecasts are the BMA methods using principal components and the SSR-combined ADL forecasts. These methods performed slightly worse than the benchmark, but slightly better than the AR forecast. The other forecasts (OLS, the BMA methods applied the untransformed regressors, and BIC and Bagging applied to principal components) all performed worse than the AR forecast. The last two columns of the table show a general increase in the relative MSFE for many of the methods in the second period indicating an improvement in the relative performance of the benchmark forecast. The relative performance of the AR model also improved in the second period. However, the relative ranking of methods is generally robust across the two sample periods.

When there are large and persistent changes in model parameters, estimates computed from rolling samples are more accurate than estimates computed from recursive samples. Table 5 takes up this issue by comparing the recursively computed forecasts examined in Tables 3 and 4 with forecasts computed from models estimated using 120-month rolling samples. The results in the table are based on the MSFE of the recursively estimated model relative to the MSFE for the rolling estimates. (A relative MSFE less than 1.0 implies that the recursive estimates produced more accurate forecasts than the rolling estimates.) The first panel of Table 5 summarizes the results for the various methods; it shows the fraction of the relative MSFE for 36 series/horizon forecasts that are less than 1.0 (column 2) and the mean of these 36 relative MSFE's (column 3). Panel B shows analogous results for each series averaged across the 56 method/horizon forecasts. The mean relative MSFE is less 1.0 for all methods (Panel A) and for all series (Panel B). For all but two methods (OLS and Combined-SSR), the recursive forecasts are more accurate than the rolling forecasts for over 75% of the series

and horizons (panel A, column 2); for all of the series, the recursive forecasts are more accurate than the rolling forecasts for over 70% of the methods and horizons (panel B, column 2). Thus, as a general rule the recursive forecasts are more accurate than the rolling forecasts.

Results for other variants of the forecast methods. Stock and Watson (2004b) reports detailed results for 52 variants of the basic forecasting models that include variations on AR lags, numbers of factors, and priors for the BMA models. It is useful to highlight a few of the results from these additional models. First, for several of the models it was computationally convenient to fix the number of AR lags, and the results reported in Tables 2-4 use 4 lags. Forecasts for the real variables and interest rates were generally more accurate using 4 rather than 12 lags, but this was not true for the inflation series. However, the ranking of the various methods and the results on stability are the same for the 12 lag models. Second, the bagging forecasts reported here are based on HAC robust *t*-statistics, but the results are essentially identical using *t*-statistics computed used standard OLS formula. Apparently, the serial correlation in the principal components is sufficiently small that the overlapping sample for $h > 1$ is not important. Third, the performance of the bagging forecast improved when the pre-test critical value was increased. For example, the average full sample relative MSFE reported in Table 4 falls from 1.54 for $c = 1.96$, to 1.28 for $c = 2.58$. (The reason for this is discussed below). Finally, the uninformative prior for the BMA forecasts reported here uses $g = 1/n^2$; there is slight decrease in forecasting performance using $g = 1/T$, a result that is consistent with the simulation results reported in Fernandez, Ley, and Steele (2002).

5.2 Properties of the Forecasts

The average correlations across series and horizon for the forecasts constructed by each method are shown in Table 6. Values greater that 0.80 are shown in bold and values greater than 0.90 are shown in italics. Evidently, there are relatively few large correlations: only 27 out of the 91 correlations are greater than 0.80, only 5 are greater than 0.90, and the average correlation is 0.72. Apparently, the forecasts are subject to markedly different sampling error and/or capture different features of the predictors. Forecasts within groups are more highly correlated. For example, the FAAR forecasts

have an average correlation of 0.90 (and the FAAR-OLS and FAAR-WLS correlation is 0.98); the average correlation within the BMA forecasts is 0.85; the average correlation of methods using all of the principal components is 0.86. Finally, the AR forecast is highly correlated with the combined mean forecast (correlation = 0.94), and the bagging forecast is highly correlated with the OLS forecast (correlation = 0.96).

The next set of results compare the variance of the forecasts across methods and series, and shows the decomposition of this variance into components associated with the AR lags and with the other predictors. The variance decomposition is based on the decomposition of the forecast

$$\hat{Y}_{t+h}^{h} = \hat{Y}_{t+h}^{h,AR} + \hat{Y}_{t+h}^{h,Z} + \hat{Y}_{t+h}^{h,Residual}, \qquad (30)$$

where $\hat{Y}_{t+h}^{h,AR}$ is constructed as the projection of $\hat{Y}_{t+h}^{h}$ onto $W_t = (1, Y_t, Y_{t-1}, \ldots, Y_{t-12})$, $\hat{Y}_{t+h}^{h,Z}$ is the projection of $\hat{Y}_{t+h}^{h}$ onto Z_t, where Z_t is the residual from the projection of X_t onto W_t, and $\hat{Y}_{t+h}^{h,Residual}$ is the residual from the projection of $\hat{Y}_{t+h}^{h}$ onto W_t and Z_t. By construction, the three components are mutually orthogonal. With the exception of the ADL forecasts, each forecast is constructed as linear combination of the elements of W_t and Z_t, and indeed, this motivates the decomposition. The projections defining the components in (30) are computed by the OLS regression of the recursively computed values of $\hat{Y}_{t+h}^{h}$ onto W_t and Z_t over the entire pseudo out-of-sample period; because of time variation in the coefficients in the recursively estimated models the regression of $\hat{Y}_{t+h}^{h}$ onto W_t and Z_t will not fit perfectly, and the component $\hat{Y}_{t+h}^{h,Residual}$ captures this effect.

Table 7 shows the value of the variance of the forecast $\hat{Y}_{t+h}^{h}$ and its components $\hat{Y}_{t+h}^{h,AR}$, and $\hat{Y}_{t+h}^{h,Z}$, relative to series being forecast Y_{t+h}. The first panel shows the average values of these relative variances across series and horizons for each of the forecasting methods, and the second panel shows the averages across methods and horizon for each series.

Looking first at the forecasting methods summarized in Panel A, the Combined-Mean and AR forecasts are very similar; both have essentially the same variance, with a small fraction of this variance associated with the *Z*'s. This is not surprising for the AR

forecast in which the *Z*'s do not enter the forecast; here the non-zero component $\hat{Y}_{t+h}^{h,Z}$ is associated with time variation in the estimated AR coefficients which is correlated with Z_t. The result is more surprising for the Combined-Mean forecast: it implies that this forecast is essentially the AR forecast with a small, but apparently useful, *Z* component. The variance associated with $\hat{Y}_{t+h}^{h,AR}$ is similar for each of the other forecasts, and these methods differ only in the importance of $\hat{Y}_{t+h}^{h,Z}$. For the OLS forecast, this component has a very large variance, so large in fact that the variance of the forecast $\hat{Y}_{t+h}^{h}$ exceeds the variance of the series forecast Y_{t+h}^{h}. The bagging forecast also has a large $\hat{Y}_{t+h}^{h,Z}$ component. Interestingly, the FAAR-GLS forecast puts far less weight on the predictors than the FAAR-OLS or FAAR-WLS forecasts.

Looking at the series summarized Panel B, three results stand out. First, the real variables (PI, IP, UR, and EMP) have relatively small AR contributions and high contributions associated with the *Z*'s. In contrast, the AR contributions associated with the inflation series (CPI, PPI, PCE) are larger, and the relative contributions associated with *Z* are smaller. The two interest rates are apparently less predictable than the other series; both the AR and *Z* components are relatively small.

5.3 Comparison of the Principal Component Forecasts

As shown in section 2, the various forecasts using principal components all have the form $\hat{Y}_{T+1/T} = \hat{\lambda}'W_T + \sum_{i=1}^{n} \psi(t_i)\hat{\delta}_i P_{i,T}$ where $\hat{\lambda}$ and $\hat{\delta}$ are the OLS estimators of λ and δ from (25) (see (26). (28), and (29)), and where $\psi_i = \psi(|t_i|)$, where t_i is the *t*-statistic for testing $\delta_i = 0$. The methods differ only in that each uses a different function ψ, and this suggests that insight into their relative performance might gained by studying these functions.

For the BMA forecasts the function *s* depends on the prior parameters *g* and *p*; the formal Bayes procedures used $p = 0.5$ and two values of *g* (1.0 and $1/n^2$); the empirical Bayes forecast used values of *p* and *g* estimated from the sample values of $\hat{\delta}_i$. Table 8 shows the values of *p* and *g* estimated for each series and horizon for $T = 2003{:}12$. Evidently, for most series and horizons the empirical Bayes forecasts use values of *p* that

are much less than 0.5. When $p = 0.5$, each of the constituent models in the BMA procedure receives equal prior weight; in contrast, when $p < 0.5$ more prior weight is placed on models with fewer regressors. Evidently, the marginal distribution of the $\hat{\delta}_i$'s is more consistent with a model with few regressors. The empirical Bayes estimates of g are small; conditional on the choice of regressors with non-zero coefficients, there is little shrinkage of the OLS estimates. The series TBOND is an exception: the estimated values of p and g are relatively large for this series.

Figure 1 compares the shrinkage ψ functions for the BMA and bagging models, together with the BIC for $T = 360$ (for this figure we set $\kappa = 1$ for all methods). The BMA function is plotted for three values of (p, g): (0.5, 1.0), (0.5, $1/n^2$), with $n = 130$, and (0.03, .0.03), the estimated values for the unemployment rate series for $h = 6$ from Table 8. The s function for BIC is a 0-1 step function at $|t| = 2.42$, while all of the other functions increase smoothly as $|t|$ increases. The $\psi^{BMA}_{p,g}$ functions are quite different from ψ^{Bagging} and from one another. $\psi^{BMA}_{0.5,1.0}$ puts significant weight on all of the OLS estimates but has a maximum value of 0.5. (A simple calculations shows that the maximum value of $\psi^{BMA}_{p,g}$ is $1/(1+g)$.) In contrast, $\psi^{BMA}_{0.5,1/n^2}$ and $\psi^{BMA}_{0.03,0.03}$ put little weight on OLS estimates with t-statistics less than 2.5 in absolute value, and nearly full weight on OLS estimates with absolute t-statistics greater than 4.0. Bagging puts relatively large weight on all of the OLS estimates; $\psi^{Bagging}_{c=1.96}$ has a minimum value of 0.28, and is greater than 0.5 for values of $|t| > 1.3$. When c increases to 2.58, $\psi^{Bagging}_{c=2.58}$ (not shown in the figure) has a minimum value of 0.08, and is greater than 0.5 for values of $t > 2.1$. Thus, when $c = 2.58$, $\psi^{Bagging}_{c=2.58}$ looks more like $\psi^{BMA}_{0.5,1/n^2}$ and $\psi^{BMA}_{0.03,0.03}$, functions that led to more accurate forecasts than bagging using $c = 1.96$.

To see what these functions means for the forecasts, Figure 2 shows the values of ψ^{BIC}, $\psi^{Bagging}_{c=1.96}$, $\psi^{BMA}_{0.5,1.0}$, $\psi^{BMA}_{0.5,1/n^2}$ and $\psi^{BMA}_{\hat{p},\hat{g}}$ for each principal component for UR, CPI and TBOND at $T = 2003{:}12$. The results for UR and CPI are similar. The empirical Bayes forecasts put near unit weight on the first few principal components, and very little weight on the other principal components. For UR and CPI, the empirical Bayes forecasts are essentially the same as FAAR-OLS forecast. The BMA forecast with p=0.5

and $g = 1/n^2$ behaves similarly, although it places somewhat more weight on handful of other principal components. In contrast, the p=0.5 and g=1.0 BMA and the bagging forecast put large, but variable, weight on all of factors. The results for TBOND are different, as now the empirical Bayes forecast uses $\hat{p} = 0.48$ and $\hat{g} = 0.42$, and thus puts substantial weight on all of the principal components.

6. Conclusions

The common structure of a large class of shrinkage estimators in the orthogonal regressor model facilitates comparison of the estimators and analysis of their properties in settings more general (and more relevant) than those used to develop the estimators. For example Bayesian model averaging, empirical Bayes, and bagging can be studied in the multi-step prediction problem with non-exogenous regressors.

The empirical analysis suggests the following answers to the questions raised in the introduction. First, the most accurate forecasts are produced by the factor models. The OLS and WLS factor models performed marginally better than the GLS model, but the performance of all three models was similar, and these models outperformed the other models. As expected, OLS performed poorly, but so did bagging (which placed substantial weight on all of the regressors). These results are generally robust to sample period, and recursively estimated models generally outperformed rolling estimates.

The second empirical conclusion is that the forecasts are different, but it is useful to consider three distinct groups. The first group includes the AR model and the combined-mean of the bivariate ADL models. These forecasts are almost entirely a function of lagged values of the series being forecast. The second group includes OLS and bagging. These models put substantial weight on all of the predictors, and the resulting large sampling error leads to inaccurate forecasts. The third group includes the FAAR models and the principal component BMA models with small values of g. These models put weight on only a few of the principal components, resulting in less sampling error, and more accurate forecasts.

Taken together, these empirical results suggest that the large-n predictor information set is usefully summarized by a small number of factors, at least for the

purposes of short-run macroeconomic forecasting. Future research might usefully be focused on refining the factor models, perhaps by improving the ψ functions that underlie the various principal component forecasts.

Appendix: Proofs of Theorems in Section 2

***Proof of Theorem* 1**. Consider the linear forecast $\tilde{Y}_{T+1|T} = \sum_{i=1}^{n} \tilde{\delta}_i p_{iT}$ with possible shrinkage function $\psi(\kappa t_i)$, where $\kappa = (1 - n/T)^{-1/2}$ for NB and $\kappa = 1$ for BG. Then

$$\tilde{Y}_{T+1|T} - \sum_{i=1}^{n} \psi(\kappa t_i)\hat{\delta}_i p_{iT} = \sum_{i=1}^{n} \left[\tilde{\delta}_i - \psi(\kappa t_i)\hat{\delta}_i \right] p_{iT} = \sum_{i=1}^{n} \rho_i p_{iT}$$

where $\rho_i = \tilde{\delta}_i - \psi(\kappa t_i)\hat{\delta}_i$. Thus, by the Cauchy-Schwartz inequality and the assumption $|p_{iT}| \le p_{\max}$,

$$\left| \tilde{Y}_{T+1|T} - \sum_{i=1}^{n} \psi(\kappa t_i)\hat{\delta}_i p_{iT} \right| \le \sum_{i=1}^{n} |\rho_i p_{iT}| \le \left(n p_{\max}^2 \right)^{1/2} \left(\sum_{i=1}^{n} \rho_i^2 \right)^{1/2}$$

so that

$$E\left[\tilde{Y}_{T+1|T} - \sum_{i=1}^{n} \psi(\kappa t_i)\hat{\delta}_i p_{iT} \right]^2 \le p_{\max}^2 \frac{1}{n} \sum_{i=1}^{n} n^2 E\rho_i^2 . \qquad (31)$$

The remainder of the proof bounds $n^2 E\rho_i^2$.

(a) Let $\hat{t}_i = \sqrt{T}\hat{\delta}_i / \hat{\sigma}_\varepsilon$, so $\hat{t}_i = \kappa t_i$. With this notation, $\rho_i = \tilde{\delta}_i - \psi(\hat{t}_i)\hat{\delta}_i$. By (7), (8), the sufficiency of $(\hat{\delta}_i, s_e^2)$ for (δ_i, σ^2), and the lack of dependence of the prior G_τ on σ, the NB estimator is,

$$\begin{aligned} \tilde{\delta}_i^{NB} &= E_\sigma[\, \hat{\delta}_i + (\sigma^2/T)\, \ell_\delta(\hat{\delta}_i) \,|\, \hat{\delta}, s_e^2 \,] \\ &= E_\sigma[1 + \ell(\hat{\tau}_i)/\hat{\tau}_i \,|\, \hat{\delta}, s_e^2 \,]\hat{\delta}_i \\ &= E_\sigma[\psi^{NB}(\hat{\tau}_i) | \hat{\delta}, s_e^2 \,]\hat{\delta}_i , \end{aligned}$$

where the second line obtains by the change of variables $\tau_i = \sqrt{T}\ \delta_i/\sigma$ and $\hat{\tau}_i = \sqrt{T}\hat{\delta}_i/\sigma$ and the third line follows from the definition of ψ^{NB}. Let $\zeta = \hat{\sigma}_\varepsilon^2/\sigma^2$ so $\hat{\tau}_i = \hat{t}_i\,\zeta^{1/2}$. Then,

$$
\begin{aligned}
\rho_i &= \tilde{\delta}_i^{NB} - \psi^{NB}(\hat{t}_i)\,\hat{\delta}_i \\
&= E_\sigma\left[\psi^{NB}(\hat{\tau}_i) - \psi^{NB}(\hat{t}_i) \mid \hat{\delta}, s_e^2\right]\hat{\delta}_i \\
&= \frac{s_e}{\sqrt{T}} t_i E_\sigma\left[\psi^{NB}(\hat{t}_i\zeta^{1/2}) - \psi^{NB}(\hat{t}_i) \mid \hat{\delta}, s_e^2\right] \\
&= \frac{1}{2}\frac{s_e}{\kappa\sqrt{T}}\hat{t}_i^2\psi^{NB\prime}(\hat{t}_i)E_\sigma\left[(\zeta-1) \mid \hat{\delta}, s_e^2\right] \\
&\qquad + \frac{1}{8}\frac{s_e}{\sqrt{T}} t_i E_\sigma\left[\left(\tilde{t}_i^2\psi^{NB\prime\prime}(\tilde{t}_i) - \tilde{t}_i\,\psi^{NB\prime}(\tilde{t}_i)\right)\tilde{\zeta}^{-2}(\zeta-1)^2 \mid \hat{\delta}, s_e^2\right] \\
&= \rho_{1i} + \rho_{2i}, \qquad (32)
\end{aligned}
$$

where the penultimate equality follows from a second order mean value expansion of $\psi^{NB}(t\zeta^{1/2})$ around $\zeta = 1$, where $\psi^{NB\prime}$ and $\psi^{NB\prime\prime}$ are the first and second derivatives of ψ^{NB}, $\tilde{\zeta} \in [1, \zeta]$, $\tilde{t}_i = \hat{t}_i\,\tilde{\zeta}^{1/2}$, and in the first expression setting $t_i = \hat{t}_i/\kappa$; and where, in the final line, ρ_{1i} and ρ_{2i} are the two respective major terms in the preceding line. (The derivative exists by condition (*ii*) in the statement of the theorem.) Note that by Minkowski's inequality,

$$
E\rho_i^2 \le \left(\sqrt{E\rho_{1i}^2} + \sqrt{E\rho_{2i}^2}\right)^2. \qquad (33)
$$

Now

$$
\begin{aligned}
|\rho_{1i}| &= \frac{1}{2}\frac{s_e}{\kappa\sqrt{T}}\left|\hat{t}_i^2\psi^{NB\prime}(\hat{t}_i)\right|\left|E_\sigma\left[(\zeta-1) \mid \hat{\delta}, s_e^2\right]\right| \\
&\le \frac{1}{2}\frac{s_e}{\kappa\sqrt{T}} M \frac{K_1}{T-n}
\end{aligned}
$$

$$\leq \frac{1}{2}\frac{1}{\kappa(T-n)^{3/2}} MK_1\hat{\sigma}_Y \tag{34}$$

where the first inequality follows from assumptions (*i*) and (*ii*) and the second inequality follows from

$$s_e^2/T = Y'[I - P(P'P)P']Y/[(T-n)T] \leq Y'Y/[(T-n)T] = \hat{\sigma}_Y^2/(T-n), \tag{35}$$

where $\hat{\sigma}_Y^2 = Y'Y/T$. Also,

$$\begin{aligned}
|\rho_{2i}| &= \frac{1}{8}\frac{s_e}{\sqrt{T}}\left| t_i E_\sigma\left[\left(\tilde{t}_i^2\psi^{NB''}(\tilde{t}_i) - \tilde{t}_i\,\psi^{NB''}(\tilde{t}_i)\right)\tilde{\zeta}^{-2}(\zeta-1)^2 \mid \hat{\delta}, s_e^2\right]\right| \\
&\leq \frac{1}{8}\frac{s_e}{\sqrt{T}}|t_i|\sqrt{E_\sigma\left[\left(\tilde{t}_i^2\psi^{NB''}(\tilde{t}_i) - \tilde{t}_i\,\psi^{NB''}(\tilde{t}_i)\right)^2\tilde{\zeta}^{-4} \mid \hat{\delta}, s_e^2\right]}\sqrt{E_\sigma\left[(\zeta-1)^4 \mid \hat{\delta}, s_e^2\right]} \\
&\leq \frac{1}{8}\frac{s_e}{\sqrt{T}}|t_i| M\sqrt{2E_\sigma\left[\tilde{\zeta}^{-4} \mid \hat{\delta}, s_e^2\right]}\sqrt{E_\sigma\left[(\zeta-1)^4 \mid \hat{\delta}, s_e^2\right]},
\end{aligned} \tag{36}$$

where the first equality follows from the definition of ρ_{2i} in (32), the second line uses the Cauchy-Schwartz inequality, and the third inequality uses condition (*ii*). Because $\tilde{\zeta} \in [1, \zeta]$ and $\zeta \geq 0$,

$$\begin{aligned}
E_\sigma(\tilde{\zeta}^{-4} \mid \hat{\delta}, s_e^2) &\leq E_\sigma[\max(1, \zeta^{-4}) \mid \hat{\delta}, s_e^2)] \\
&= \int_0^1 f_{\zeta^{-4}|\hat{\delta}_i, s_e^2}(v \mid \hat{\delta}, s_e^2)dv + \int_1^\infty v f_{\zeta^{-4}|\hat{\delta}_i, s_e^2}(v \mid \hat{\delta}, s_e^2)dv \\
&\leq 1 + E_\sigma(\zeta^{-4} \mid \hat{\delta}, s_e^2) \\
&\leq K_3 \text{ for } T - n > r_0,
\end{aligned} \tag{37}$$

where the second inequality follows by extending the range of integration of both integrals to $[0,\infty]$ (the integrands are nonnegative), and the final inequality follows from

condition (*i*). Substituting (37) into (36), using (35) and assumptions (*i*) and (*ii*), and collecting terms yields,

$$|\rho_{2i}| \leq \frac{1}{8}\frac{1}{(T-n)^{3/2}}\sqrt{2K_2K_3M^2}\,|t_i|\hat{\sigma}_Y\,.$$

Squaring the final bounds for $|\rho_{1i}|$ and $|\rho_{2i}|$, taking expectations over the distribution of $(\hat{\delta}\,,\ s_e^2)$ given (δ, σ^2), substituting the result into (33), multiplying by n^2, and simplifying yields,

$$n^2 E\,\rho_i^2 \leq \frac{n^2}{(T-n)^3}\left[\frac{1}{2\kappa}\sqrt{E\left(K_1^2M^2\hat{\sigma}_Y^2\right)}+\frac{1}{4\sqrt{2}}\sqrt{E\left(K_2K_3M^2t_i^2\hat{\sigma}_Y^2\right)}\right]^2, \tag{38}$$

which, upon substitution into (31), yields the result stated in the theorem.

(b) This proof applies to bagging implemented using the parametric bootstrap based on the exogeneity-normality assumption (5). Let the superscript * denote bootstrap realizations and let E^* denote expectations taken with respect to the bootstrap distribution conditional on the observed data (Y, P). Each parametric bootstrap realization draws T observations such that $P^{*\prime}P^*/T = I$ and $Y^*|P^* \sim \mathrm{N}(P^*\hat{\delta}\,,\ s_e^2 I)$. Let $\hat{\delta}_{ij}^*$ denote the j^{th} bootstrap draw of the OLS estimator of δ_i and let $s_{e,j}^{2*}$ denote the j^{th} bootstrap draw of the OLS estimator of σ^2, let $\xi^* = s_{e,j}^{2*}/s_e^2$, and let $t_{ij}^* = \sqrt{T}\ \hat{\delta}_{ij}^*/s_{e,j}^*$. The j^{th} bootstrap realization of the pretest estimator is $\mathbf{1}(|t_{ij}^*|>c)\,\hat{\delta}_{ij}^*$. The bagging estimator is

$$\tilde{\delta}_i^{BG} = \frac{1}{B}\sum_{j=1}^{B}1(|t_{ij}^*|>c)\hat{\delta}_{ij}^*\,, \tag{39}$$

where B is the number of bootstrap draws.

By construction, under the * distribution, $\hat{\delta}_{ij}^* \sim$ i.i.d. N($\hat{\delta}_i, s_e^2/T$) so $\sqrt{T}\hat{\delta}_{ij}^*/s_e \sim$ i.i.d. N(t_i, 1), $\xi^* \sim$ i.i.d. $\chi^2_{T-n}/T-n$, and $\hat{\delta}_{ij}^*$ and ξ^* are independently distributed. It is useful to introduce the representation $\sqrt{T}\hat{\delta}_{ij}^*/s_e = t_i + z_{ij}^*$, where $z_{ij}^* \sim$ i.i.d. N(0,1) (the equality in this equation is equality in distribution if the bootstrap draws are of (P^*, Y^*) but the equation can alternatively be taken as the primitive parametric bootstrap sampler in which case it is an equality.)

We now turn to the evaluation of $E\rho_i^2$. Now

$$\begin{aligned}\rho_i &= \tilde{\delta}_i^{BG} - \psi^{BG}(t_t)\hat{\delta}_i \\ &= (\tilde{\delta}_i^{BG} - E^*\tilde{\delta}_i^{BG}) + (E^*\tilde{\delta}_i^{BG} - \psi^{BG}(t_t)\hat{\delta}_i) \\ &= \rho_{1i} + \rho_{2i},\end{aligned}$$

where $\rho_{1i} = \tilde{\delta}_i^{BG} - E^*\tilde{\delta}_i^{BG}$ and $\rho_{2i} = E^*\tilde{\delta}_i^{BG} - \psi^{BG}(t_t)\hat{\delta}_i$. Note that, by Minkowski's inequality,

$$E\rho_i^2 \le \left(\sqrt{E\rho_{1i}^2} + \sqrt{E\rho_{2i}^2}\right)^2. \tag{40}$$

First consider $E\rho_{1i}^2$. Now $E\rho_{1i}^2 = E[E^*\rho_{1i}^2] = E\text{var}^*(\rho_{1i})$ since the * distribution is conditional on (Y,P) and $E^*\rho_{1i} = 0$. Now

$$\begin{aligned}\text{var}^*(\rho_{1i}) &= \text{var}^*(\tilde{\delta}_i^{BG} - E^*\tilde{\delta}_i^{BG}) \\ &= \text{var}^*\left\{\frac{1}{B}\sum_{j=1}^{B}\left[1(|t_{ij}^*| > c)\hat{\delta}_{ij}^* - E^*1(|t_{ij}^*| > c)\hat{\delta}_{ij}^*\right]\right\} \\ &= \frac{1}{B}\text{var}^*\left[1(|t_{ij}^*| > c)\hat{\delta}_{ij}^*\right] \\ &= \frac{s_e^2}{T}\frac{1}{B}\text{var}^*\left[1(|t_{ij}^*| > c)\frac{\sqrt{T}\hat{\delta}_{ij}^*}{s_e}\right]\end{aligned}$$

$$= \frac{s_e^2}{T}\frac{1}{B}\text{var}^*\left[1\left(|t_i + z_{ij}^*| > c\sqrt{\xi_j^*}\right)\left(t_i + z_{ij}^*\right)\right],$$

where the second equality follows by substituting (39), the third equality follows because the bootstrap draws are i.i.d., and the fourth equality adopts the (z_{ij}^*, ξ_j^*) notation introduced above. Now $1\left(|t + z_{ij}^*| > d\right)\left(t + z_{ij}^*\right)$ has a N(t,1) distribution if the variable exceeds d and equals zero otherwise. This random variable has a variance that is bounded as a function of t and the maximum variance (over t) is continuous in d. Because ξ_j^* is distributed $\chi_{T-n}^2/T-n$ and is independent of z_{ij}^*, it follows that $\text{var}^*\left[1\left(|t_i + z_{ij}^*| > c\sqrt{\xi_j^*}\right)\left(t_i + z_{ij}^*\right)\right]$ is bounded as a function of t_i. Using the notation introduced in the statement of the theorem, we have that

$$\text{var}^*(\rho_{1i}) \leq \frac{s_e^2}{T}\frac{1}{B}m_{T-n}(c) \leq \frac{\hat{\sigma}_Y^2}{B(T-n)}m_{T-n}(c),$$

where the second inequality uses the inequality (35); thus

$$E\rho_{1i}^2 \leq \frac{E\hat{\sigma}_Y^2}{B(T-n)}m_{T-n}(c). \tag{41}$$

Now turn to $E\rho_{2i}^2$. We have that,

$$\begin{aligned}
\rho_{2i} &= E^*\tilde{\delta}_i^{BG} - \psi^{BG}(t_t)\hat{\delta}_i \\
&= E^*1(|t_{ij}^*| > c)\hat{\delta}_{ij}^* - \psi^{BG}(t_t)\hat{\delta}_i \\
&= \frac{s_e}{\sqrt{T}}\left[E^*1(|t_{ij}^*| > c)\frac{\sqrt{T}\hat{\delta}_{ij}^*}{s_e} - \psi^{BG}(t_i)\frac{\sqrt{T}\hat{\delta}_i}{s_e}\right] \\
&= \frac{s_e}{\sqrt{T}}\left[E^*1\left(|t_i + z_{ij}^*| > c\sqrt{\xi_j^*}\right)\left(t_i + z_{ij}^*\right) - \psi^{BG}(t_i)t_i\right].
\end{aligned}$$

Now

$$E^*\left[1\left(|t+z_{ij}^*|>d\right)\left(t+z_{ij}^*\right)\right] = \int_{|t+z^*|>d} (t+z^*)\phi(z^*)dz^*$$

$$= t\int_{|t+z^*|>d} \phi(z^*)dz^* + \int_{|t+z^*|>d} z^*\phi(z^*)dz^*$$

$$= \psi^{BA}(t,d)t$$

where $\psi^{BG}(t,d) \equiv 1 - \Phi(t+d) + \Phi(t-d) + t^{-1}[\phi(t-d) - \phi(t+d)]$ (cf. Bühlmann and Yu (2002)). Thus

$$\rho_{2i} = \frac{s_e}{\sqrt{T}} t_i E^*\left[\psi^{BG}(t_i, c\sqrt{\xi_j^*}) - \psi^{BG}(t_i, c)\right].$$

Let $\psi^{BG\prime}$ and $\psi^{BG\prime\prime}$ denote the first two derivatives of ψ^{BG} with respect to its second argument (direct calculation show that $t\psi^{BG\prime}(t,c)$ and $t\psi^{BG\prime\prime}(t,c)$ exist). By the extendend mean value theorem, the second order expansion of $\psi^{BG}(t_i, c\sqrt{\xi_j^*})$ around $\xi_j^* = 1$ yields,

$$\rho_{2i} = \frac{s_e}{\sqrt{T}} E^*\left\{\frac{1}{2}t_i\psi^{BG\prime}(t_i,c)c(\xi_j^*-1) + \frac{1}{8}\left[t_i\psi^{BG\prime\prime}(t_i,\tilde{c})\tilde{c}^2 - t_i\psi^{BG\prime}(t_i,\tilde{c})\tilde{c}\right]\frac{(\xi_j^*-1)^2}{\tilde{\xi}^2}\right\} \quad (42)$$

where $\tilde{c} = c\sqrt{\tilde{\xi}}$, where $\tilde{\xi} \in [1, \xi_j^*]$. The * expectation of the first term in large brackets in (42) is

$$E^*\left[\frac{1}{2}t_i\psi^{BG\prime}(t_i,c)c(\xi_j^*-1)\right] = \frac{1}{2}t_i\psi^{BG\prime}(t_i,c)cE^*(\xi_j^*-1) = 0$$

because ξ_j^* is distributed $\chi^2_{T-n}/T-n$. Thus from (42) we have,

$$|\rho_{2i}| = \frac{s_e}{\sqrt{T}} E^* \left\{ \frac{1}{8} \left[t_i \psi^{BG''}(t_i, \tilde{c}) \tilde{c}^2 - t_i \psi^{BG'}(t_i, \tilde{c}) \tilde{c} \right] \frac{(\xi_j^* - 1)^2}{\tilde{\xi}^2} \right\}$$

$$\leq \frac{s_e}{\sqrt{T}} \frac{1}{8} \sup_u \left| t_i \psi^{BG''}(t_i, u) u^2 - t_i \psi^{BG'}(t_i, u) u \right| E^* \left[\frac{(\xi_j^* - 1)^2}{\tilde{\xi}^2} \right]$$

$$\leq \frac{s_e}{\sqrt{T}} \frac{1}{8} \sup_u \left| t_i \psi^{BG''}(t_i, u) u^2 - t_i \psi^{BG'}(t_i, u) u \right| \sqrt{E^*(\xi_j^* - 1)^4} \sqrt{E^* \tilde{\xi}^{-4}} \,. \tag{43}$$

We now calculate or bound the three final terms in (43). First, $E^*(\xi_j^* - 1)^4$ is the fourth central moment of a $\chi^2_{T-n} / T - n$ random variable, so

$$E^*(\xi_j^* - 1)^4 = 12(T-n)(T-n+4)/(T-n)^4. \tag{44}$$

Next, because $\tilde{\xi} \in [1, \xi_j^*]$ and because the fourth moment of the reciprocal of a χ^2_r random variable exists for $r > 8$ and is $[(r-2)(r-4)(r-6)(r-8)]^{-1}$, for $T - n \geq 8$ we have that

$$E^* \tilde{\xi}_j^{*-4} \leq 1 + E^* \xi_j^{*-4}$$

$$= 1 + \frac{(T-n)^4}{(T-n-2)(T-n-4)(T-n-6)(T-n-8)} \tag{45}$$

where the first inequality obtains using the argument leading to the second inequality in (37).

Now turn to the sup term in (43). Direct evaluation of the derivatives using the definition of ψ^{BG} show that $t\psi^{BG'}(t,u)u = u^2[\phi(t+u) - \phi(t-u)]$ and $t\psi^{BG''}(t,u)u^2 = u^2[\phi(t+u) - \phi(t-u)] - u^3[(t+u)\phi(t+u) + (t-u)\phi(t-u)]$. Thus

$$|t\psi^{BG'}(t,u)u| \leq 2\sup_u u^2 \phi(t+u)$$

$$\leq 2\sup(v-t)^2 \phi(v)$$

$$\leq 2[(\sup_v v^2\phi(v)) + 2t\sup_v|v\phi(v)| + t^2\sup_v\phi(v)]$$
$$= 2(h_2 + 2h_1 t + h_2 t^2) \qquad (46)$$

where the second line follows by the change of variables $v = t + u$ and where $h_m = m^{m/2}e^{-m/2}/\sqrt{2\pi}$. Similar calculations provide a bound on $|t\psi^{BG\prime\prime}(t,u)u^2|$ which, combined with the bound in (46), yields

$$\sup_u \left| t_i\psi^{BG\prime\prime}(t_i,u)u^2 - t_i\psi^{BG\prime}(t_i,u)u \right|$$
$$\leq 2[(2h_2 + h_4) + (4h_1 + 3h_3)|t_i| + (2h_0 + 3h_2)t_i^2 + h_1|t_i|^3]$$
$$\leq 14h_4\sum_{m=0}^{3}|t_i|^m ,$$

where the final equality uses $h_i < h_m$ for $i < m$ and $m > 1$. Substituting this bound, (44), and (45) into (43), squaring, taking expectations, and collecting terms yields

$$E\rho_{2i}^2 \leq \frac{K_4^2\mu_{T-n}^2}{(T-n)^2}E\left[\frac{s_e^2}{T}\left(\sum_{m=0}^{3}|t_i|^m\right)^2\right]$$
$$\leq \frac{K_4^2\mu_{T-n}^2}{(T-n)^3}E\left[\hat{\sigma}_Y^2\left(\sum_{m=0}^{3}|t_i|^m\right)^2\right]$$
$$\leq \frac{K_4^2\mu_{T-n}^2}{(T-n)^3}\sqrt{E\hat{\sigma}_Y^4}\sqrt{E\left(\sum_{m=0}^{3}|t_i|^m\right)^4} \qquad (47)$$

where the second inequality uses the inequality (35), the third inequality is Cauchy-Schwartz, direct evaluation shows that $K_4 = 28e^{-2}\sqrt{3/\pi} \approx 3.703\ldots$, and

$$\mu_r = \left[\frac{r+4}{r} + \frac{r^3(r+4)}{(r-2)(r-4)(r-6)(r-8)}\right]^{1/2}.$$

Substituting (41) and (47) into (40), multiplying by n^2, and using $E\hat{\sigma}_Y^2 \le \sqrt{E\hat{\sigma}_Y^4}$ yields,

$$n^2 E\rho_i^2 \le \sqrt{E\hat{\sigma}_Y^4}\left\{\sqrt{\frac{n^2}{B(T-n)}}\sqrt{m_{T-n}(c)} + \sqrt{\frac{n^2}{(T-n)^3}}K_4\mu_{T-n}\left[E\left(\sum_{m=0}^{3}|t_i|^m\right)^4\right]^{1/4}\right\}^2,$$

which upon substitution into (31) yields the result stated in the theorem.

***Proof of Theorem* 2**. Theorem 2 follows directly from Theorem 1.

Data Appendix

All of the data series used as predictors are listed in Table A.1, which lists the series name (which is the series label used in the source database), the transformation applied to the series, and a brief data description. All series are from the Global Insights (formerly DRI) Basic Economics Database, except those that include TCB (which are from the Conference Board's Indicators Database) or AC (author's calculation).

Before using the series as predictors they were screened for outliers. Observations of the transformed series with absolute median deviations larger than 6 times the inter quartile range were replaced with the median value of the preceding 5 observations.

Table A.1

Series	Tran	Description
a0m052	Δln	Personal Income (AR, Bil. Chain 2000 $) (TCB)
a0m051	Δln	Personal Income Less Transfer Payments (AR, Bil. Chain 2000 $) (TCB)
a0m224_r	Δln	Real Consumption (AC) a0m224/gmdc (a0m224 is from TCB)
a0m057	Δln	Manufacturing And Trade Sales (Mil. Chain 1996 $) (TCB)
a0m059	Δln	Sales Of Retail Stores (Mil. Chain 2000 $) (TCB)
ips10	Δln	Industrial Production Index - Total Index
ips11	Δln	Industrial Production Index - Products, Total
ips299	Δln	Industrial Production Index - Final Products
ips12	Δln	Industrial Production Index - Consumer Goods
ips13	Δln	Industrial Production Index - Durable Consumer Goods
ips18	Δln	Industrial Production Index - Nondurable Consumer Goods
ips25	Δln	Industrial Production Index - Business Equipment
ips32	Δln	Industrial Production Index - Materials
ips34	Δln	Industrial Production Index - Durable Goods Materials
ips38	Δln	Industrial Production Index - Nondurable Goods Materials
ips43	Δln	Industrial Production Index - Manufacturing (Sic)
ips307	Δln	Industrial Production Index - Residential Utilities
ips306	Δln	Industrial Production Index - Fuels
pmp	lv	Napm Production Index (Percent)
a0m082	Δlv	Capacity Utilization (Mfg) (TCB)
lhel	Δlv	Index Of Help-Wanted Advertising In Newspapers (1967=100;Sa)
lhelx	Δlv	Employment: Ratio; Help-Wanted Ads:No. Unemployed Clf
lhem	Δln	Civilian Labor Force: Employed, Total (Thous.,Sa)
lhnag	Δln	Civilian Labor Force: Employed, Nonagric.Industries (Thous.,Sa)
lhur	Δlv	Unemployment Rate: All Workers, 16 Years & Over (%,Sa)
lhu680	Δlv	Unemploy.By Duration: Average(Mean)Duration In Weeks (Sa)
lhu5	Δln	Unemploy.By Duration: Persons Unempl.Less Than 5 Wks (Thous.,Sa)
lhu14	Δln	Unemploy.By Duration: Persons Unempl.5 To 14 Wks (Thous.,Sa)
lhu15	Δln	Unemploy.By Duration: Persons Unempl.15 Wks + (Thous.,Sa)
lhu26	Δln	Unemploy.By Duration: Persons Unempl.15 To 26 Wks (Thous.,Sa)
lhu27	Δln	Unemploy.By Duration: Persons Unempl.27 Wks + (Thous,Sa)
a0m005	Δln	Average Weekly Initial Claims, Unemploy. Insurance (Thous.) (TCB)
ces002	Δln	Employees On Nonfarm Payrolls - Total Private
ces003	Δln	Employees On Nonfarm Payrolls - Goods-Producing
ces006	Δln	Employees On Nonfarm Payrolls - Mining
ces011	Δln	Employees On Nonfarm Payrolls - Construction
ces015	Δln	Employees On Nonfarm Payrolls - Manufacturing
ces017	Δln	Employees On Nonfarm Payrolls - Durable Goods
ces033	Δln	Employees On Nonfarm Payrolls - Nondurable Goods
ces046	Δln	Employees On Nonfarm Payrolls - Service-Providing
ces048	Δln	Employees On Nonfarm Payrolls - Trade, Transportation, And Utilities
ces049	Δln	Employees On Nonfarm Payrolls - Wholesale Trade
ces053	Δln	Employees On Nonfarm Payrolls - Retail Trade
ces088	Δln	Employees On Nonfarm Payrolls - Financial Activities
ces140	Δln	Employees On Nonfarm Payrolls - Government
a0m048	Δln	Employee Hours In Nonag. Establishments (AR, Bil. Hours) (TCB)
ces151	lv	Avg Weekly Hrs of Prod or Nonsup Workers On Private Nonfarm Payrolls - Goods-Producing

ces155	Δlv	Avg Weekly Hrs of Prod or Nonsup Workers On Private Nonfarm Payrolls - Mfg Overtime Hours
aom001	lv	Average Weekly Hours, Mfg. (Hours) (TCB)
pmemp	lv	Napm Employment Index (Percent)
hsfr	ln	Housing Starts:Nonfarm(1947-58);Total Farm&Nonfarm(1959-)(Thous.,Saar)
hsne	ln	Housing Starts:Northeast (Thous.U.)S.A.
hsmw	ln	Housing Starts:Midwest(Thous.U.)S.A.
hssou	ln	Housing Starts:South (Thous.U.)S.A.
hswst	ln	Housing Starts:West (Thous.U.)S.A.
hsbr	ln	Housing Authorized: Total New Priv Housing Units (Thous.,Saar)
hsbne*	ln	Houses Authorized By Build. Permits:Northeast(Thou.U.)S.A
hsbmw*	ln	Houses Authorized By Build. Permits:Midwest(Thou.U.)S.A.
hsbsou*	ln	Houses Authorized By Build. Permits:South(Thou.U.)S.A.
hsbwst*	ln	Houses Authorized By Build. Permits:West(Thou.U.)S.A.
pmi	lv	Purchasing Managers' Index (Sa)
pmno	lv	Napm New Orders Index (Percent)
pmdel	lv	Napm Vendor Deliveries Index (Percent)
pmnv	lv	Napm Inventories Index (Percent)
a0m008	Δln	Mfrs' New Orders, Consumer Goods And Materials (Bil. Chain 1982 $) (TCB)
a0m007	Δln	Mfrs' New Orders, Durable Goods Industries (Bil. Chain 2000 $) (TCB)
a0m027	Δln	Mfrs' New Orders, Nondefense Capital Goods (Mil. Chain 1982 $) (TCB)
a1m092	Δln	Mfrs' Unfilled Orders, Durable Goods Indus. (Bil. Chain 2000 $) (TCB)
a0m070	Δln	Manufacturing And Trade Inventories (Bil. Chain 2000 $) (TCB)
a0m077	Δlv	Ratio, Mfg. And Trade Inventories To Sales (Based On Chain 2000 $) (TCB)
fm1	Δ^2ln	Money Stock: M1(Curr,Trav.Cks,Dem Dep,Other Ck'able Dep)(Bil$,Sa)
fm2	Δ^2ln	Money Stock:M2(M1+O'nite Rps,Euro$,G/P&B/D Mmmfs&Sav&Sm Time Dep(Bil$,Sa)
fm3	Δ^2ln	Money Stock: M3(M2+Lg Time Dep,Term Rp's&Inst Only Mmmfs)(Bil$,Sa)
fm2dq	Δln	Money Supply - M2 In 1996 Dollars (Bci)
fmfba	Δ^2ln	Monetary Base, Adj For Reserve Requirement Changes(Mil$,Sa)
fmrra	Δ^2ln	Depository Inst Reserves:Total,Adj For Reserve Req Chgs(Mil$,Sa)
fmrnba	Δ^2ln	Depository Inst Reserves:Nonborrowed,Adj Res Req Chgs(Mil$,Sa)
fclnq	Δ^2ln	Commercial & Industrial Loans Oustanding In 1996 Dollars (Bci)
fclbmc	lv	Wkly Rp Lg Com'l Banks:Net Change Com'l & Indus Loans(Bil$,Saar)
ccinrv	Δ^2ln	Consumer Credit Outstanding - Nonrevolving(G19)
a0m095	Δlv	Ratio, Consumer Installment Credit To Personal Income (Pct.) (TCB)
fspcom	Δln	S&P's Common Stock Price Index: Composite (1941-43=10)
fspin	Δln	S&P's Common Stock Price Index: Industrials (1941-43=10)
fsdxp	Δlv	S&P's Composite Common Stock: Dividend Yield (% Per Annum)
fspxe	Δln	S&P's Composite Common Stock: Price-Earnings Ratio (%,Nsa)
fyff	Δlv	Interest Rate: Federal Funds (Effective) (% Per Annum,Nsa)
cp90	Δlv	Cmmercial Paper Rate (AC)
fygm3	Δlv	Interest Rate: U.S.Treasury Bills,Sec Mkt,3-Mo.(% Per Ann,Nsa)
fygm6	Δlv	Interest Rate: U.S.Treasury Bills,Sec Mkt,6-Mo.(% Per Ann,Nsa)
fygt1	Δlv	Interest Rate: U.S.Treasury Const Maturities,1-Yr.(% Per Ann,Nsa)
fygt5	Δlv	Interest Rate: U.S.Treasury Const Maturities,5-Yr.(% Per Ann,Nsa)
fygt10	Δlv	Interest Rate: U.S.Treasury Const Maturities,10-Yr.(% Per Ann,Nsa)
fyaaac	Δlv	Bond Yield: Moody's Aaa Corporate (% Per Annum)
fybaac	Δlv	Bond Yield: Moody's Baa Corporate (% Per Annum)
scp90	lv	cp90-fyff
sfygm3	lv	fygm3-fyff
sfygm6	lv	fygm6-fyff
sfygt1	lv	fygt1-fyff
sfygt5	lv	fygt5-fyff
sfygt10	lv	fygt10-fyff
sfyaaac	lv	fyaaac-fyff
sfybaac	lv	fybaac-fyff
exrus	Δln	United States;Effective Exchange Rate(Merm)(Index No.)
exrsw	Δln	Foreign Exchange Rate: Switzerland (Swiss Franc Per U.S.$)
exrjan	Δln	Foreign Exchange Rate: Japan (Yen Per U.S.$)
exruk	Δln	Foreign Exchange Rate: United Kingdom (Cents Per Pound)
exrcan	Δln	Foreign Exchange Rate: Canada (Canadian $ Per U.S.$)
pwfsa	Δ^2ln	Producer Price Index: Finished Goods (82=100,Sa)
pwfcsa	Δ^2ln	Producer Price Index:Finished Consumer Goods (82=100,Sa)
pwimsa	Δ^2ln	Producer Price Index:Intermed Mat.Supplies & Components(82=100,Sa)
pwcmsa	Δ^2ln	Producer Price Index:Crude Materials (82=100,Sa)
psm99q	Δ^2ln	Index Of Sensitive Materials Prices (1990=100)(Bci-99a)
pmcp	lv	Napm Commodity Prices Index (Percent)

punew	Δ^2ln	Cpi-U: All Items (82-84=100,Sa)
pu83	Δ^2ln	Cpi-U: Apparel & Upkeep (82-84=100,Sa)
pu84	Δ^2ln	Cpi-U: Transportation (82-84=100,Sa)
pu85	Δ^2ln	Cpi-U: Medical Care (82-84=100,Sa)
puc	Δ^2ln	Cpi-U: Commodities (82-84=100,Sa)
pucd	Δ^2ln	Cpi-U: Durables (82-84=100,Sa)
pus	Δ^2ln	Cpi-U: Services (82-84=100,Sa)
puxf	Δ^2ln	Cpi-U: All Items Less Food (82-84=100,Sa)
puxhs	Δ^2ln	Cpi-U: All Items Less Shelter (82-84=100,Sa)
puxm	Δ^2ln	Cpi-U: All Items Less Midical Care (82-84=100,Sa)
gmdc	Δ^2ln	Pce,Impl Pr Defl:Pce (1987=100)
gmdcd	Δ^2ln	Pce,Impl Pr Defl:Pce; Durables (1987=100)
gmdcn	Δ^2ln	Pce,Impl Pr Defl:Pce; Nondurables (1996=100)
gmdcs	Δ^2ln	Pce,Impl Pr Defl:Pce; Services (1987=100)
ces275	Δ^2ln	Avg Hourly Earnings of Prod or Nonsup Workers On Private Nonfarm Payrolls - Goods-Producing
ces277	Δ^2ln	Avg Hourly Earnings of Prod or Nonsup Workers On Private Nonfarm Payrolls - Construction
ces278	Δ^2ln	Avg Hourly Earnings of Prod or Nonsup Workers On Private Nonfarm Payrolls - Manufacturing
hhsntn	Δlv	U. Of Mich. Index Of Consumer Expectations(Bcd-83)

References

Anderson, T.W. (1984), *An Introduction to Multivariate Statistical Analysis,* (2nd Edition), New York: Wiley.

Atkeson, A., and Ohanian, L. (2001), "Are Phillips curves useful for inflation forecasting," *Quarterly Review*, Federal Reserve Bank of Minneapolis, Winter, 2-11.

Bates, J.M., and Granger, C.W.J. (1969), "The combination of forecasts," *Operations Research Quarterly*, 20, 451–468.

Boivin, J. and Ng, S. (2003), "Are more data always better for factor analysis?" *Journal of Econometrics*, forthcoming.

Boivin, J. and S. Ng (2005), "Understanding and Comparing Factor-Based Forecasts," manuscript, University of Michigan.

Breiman, L. (1996), "Bagging predictors," *Machine Learning*, 36, 105-139.

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

Clyde, M. (1999a), "Bayesian model averaging and model search strategies (with discussion)," in: J.M. Bernardo, A.P. Dawid, J.O. Berger, and A.F.M. Smith, eds., *Bayesian Statistics* 6, Oxford: Oxford University Press.

Clyde, M., Desimone, H., and Parmigiani, G. (1996), "Prediction via orthogonalized model mixing," *Journal of the American Statistical Association*, 91, 1197-1208.

Clyde, M. (1999b), "Comment on 'Bayesian model averaging: a tutorial'," *Statistical Science*, 14, 401-404.

Fernandez, C., Ley, E., and Steele, M.F.J. (2001), "Benchmark priors for Bayesian model averaging," *Journal of Econometrics*, 100, 381-427.

Forni, M., Hallin, M., Lippi, M., and Reichlin, L. (2003), "The generalized dynamic factor model: one-sided estimation and forecasting", manuscript.

Giannoni, D., Reichlin, L., and Sala, L. (2004), "Monetary policy in real time," forthcoming, NBER *Macroeconomics Annual*, 2004.

Granger, C.W.J., and Ramanathan, R. (1984), "Improved methods of combining forecasting," *Journal of Forecasting*, 3, 197–204.

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

Inoue, A., and Kilian, L. (2004), "How useful in bagging in forecasting economic time seires? A case study of U.S. CPI inflation," manuscript, University of Michigan.

James, W., and Stein, C. (1960), "Estimation with quadratic loss," *Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability*, Vol. 1, 361-379.

Koop, G., and Potter, S. (2003), "Forecasting in large macroeconomic panels using Bayesian model averaging," manuscript (University of Leicester).

Knox, T., Stock, J.H., and Watson, M.W. (2001), "Empirical Bayes forecasts of one time series using many regressors," Technical Working Paper No. 269 (NBER).

Leamer, E.E. (1978), *Specification Searches*, New York: Wiley.

Lee, T.-H. and Y. Yang (2004), "Bagging Binary and Quantile Predictors for Time Series," manuscript, UC-Riverside.

Madigan, D., and York, J. (1995), "Bayesian graphical models for discrete data," *International Statistical Review*, 63, 215-232.

Stein, C. (1955), "Inadmissibility of the usual estimator for the mean of the multivariate normal distribution," *Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability*, 1, 197-206.

Stock, J.H., and Watson, M.W. (2002a), "Macroeconomic forecasting using diffusion indexes", *Journal of Business and Economic Statistics*, 20, 147-162.

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

Stock, J.H., and Watson, M.W. (2004a), "Forecasting with many predictors," manuscript.

Stock, J.H., and Watson, M.W. (2004b), "Forecasting using many predictors: detailed results," manuscript available at http://www.wws.princeton.edu/~mwatson.

Timmerman, A. (2004), "Forecast combinations," manuscript UCSD.

Wright, J.H. (2004), "Forecasting inflation by Bayesian model averaging," manuscript, Board of Governors of the Federal Reserve System.

Zellner, A. (1986), "On assessing prior distributions and Bayesian regression analysis with g-prior distributions," in *Bayesian Inference and Decision Techniques: Essays in Honor of Bruno de Finietti*, eds. P.K. Goel and A. Zellner, Amsterdam: North Holland, pp. 233-243.

Table 1

Forecasted Series

Series	Abbreviation	Y_{t+h}^{h}	Y_t
Real Personal Income	PI	$(1200/h)\ln(Z_{t+h}/Z_t)$	$\Delta\ln(Z_t)$
Industrial Production	IP	$(1200/h)\ln(Z_{t+h}/Z_t)$	$\Delta\ln(Z_t)$
Unemployment Rate	UR	$(Z_{t+h} - Z_t)$	ΔZ_t
Employment	EMP	$(1200/h)\ln(Z_{t+h}/Z_t)$	$\Delta\ln(Z_t)$
3-Mth Tbill Rate	TBILL	$(Z_{t+h} - Z_t)$	ΔZ_t
10-Yr TBond Rate	TBOND	$(Z_{t+h} - Z_t)$	ΔZ_t
Producer Price Index	PPI	$1200[(1/h)\ln(Z_{t+h}/Z_t) - \Delta\ln(Z_t)]$	$\Delta^2\ln(Z_t)$
Consumer Price Index	CPI	$1200[(1/h)\ln(Z_{t+h}/Z_t) - \Delta\ln(Z_t)]$	$\Delta^2\ln(Z_t)$
PCE Deflator	PCED	$1200[(1/h)\ln(Z_{t+h}/Z_t) - \Delta\ln(Z_t)]$	$\Delta^2\ln(Z_t)$

Notes: This table lists the nine series that are forecast. The first column gives the description of the series, the second lists the abbreviation used in the results tables, the next two columns shows the transformations that define the variable forecast, Y_{t+h}^{h}, and the autoregressive lags, Y.

Table 2
Forecasting Models

Method	Description
Combined-Mean	Combined ADL Models, AIC Lag Selection, sample mean ($w_i = 1/n$). This serves as the benchmark for many of the comparisons.
AR	AR Model, AIC Lag Selection
OLS	All X Variables, $p_Y = 4$, all coefficients estimated by OLS
Combined-SSR	Combined ADL Models, $p_Y = 4$, w_i from (17) with λ chosen by PLS
FAAR-OLS	Factor Augmented AR model, OLS estimation of Factors (PC), AIC selection of factors and AR lags
FAAR-GLS	Factor Augmented AR model, GLS estimation of Factors (PC), AIC selection of factors and AR lags
FAAR-WLS	Factor Augmented AR model, WLS estimation of Factors (PC), AIC selection of factors and AR lags
BMA($1/n^2$,0.5)	BMA using X, $p_Y = 4$, $g = 1/n^2$, $p = 0.5$
BMA(1,0.5)	BMA using X, $p_Y = 4$, $g = 1$, $p = 0.5$
BMA-PC($1/n^2$,0.5)	BMA using Principal Components of X, $p_Y = 4$, $g = 1/n^2$, $p = 1/2$
BMA-PC(1,0.5)	BMA using Principal Components of X, $p_Y = 4$, $g = 1$, $p = 1/2$
PEB-PC	BMA using Principal Components of X, $p_Y = 4$, empirical Bayes estimates of g and p
BIC-PC	Principal Components of X with BIC Selection, $p_Y = 4$
Bagging-PC	Bagging using Principal Components of X, $c = 1.96$ with Newey-West t-statistics, $p_Y = 4$

Notes: The table describes the forecasting models that form the basis for the empirical results shown in Section 4. The first column shows the abbreviation for the method (this abbreviation is used in all of the tables below), and the second column describes the method.

Table 3
Mean Square Forecast Errors

a. $h = 1$

	PI	IP	UR	EMP	TBILL	TBOND	PPI	CPI	PCED
Combined-Mean Root-MSFE	6.51	7.54	0.17	2.16	0.55	0.34	5.48	2.52	1.93
MSFE Relative to Combined-Mean									
AR	1.04	1.09	1.07	1.07	1.03	1.02	1.04	1.06	1.03
OLS	1.87	1.94	1.83	2.41	1.73	1.43	2.70	2.33	2.04
Combined-SSR	0.88	0.91	0.90	0.93	1.03	0.93	1.06	1.07	1.02
FAAR-OLS	0.96	0.91	0.86	0.92	0.86	0.92	1.00	0.97	0.98
FAA-GLS	1.00	1.05	0.95	1.23	0.93	0.96	1.04	1.01	1.05
FAAR-WLS	0.96	0.90	0.86	0.95	0.86	0.93	1.01	0.95	0.98
BMA(1/n^2,0.5)	0.92	0.89	0.88	1.02	1.08	0.94	1.07	1.10	1.02
BMA(1,0.5)	0.94	0.83	0.87	1.10	1.05	0.99	1.21	1.21	1.22
BMA-PC(1/n^2,0.5)	0.95	0.90	0.87	0.91	0.89	0.90	1.08	1.13	1.02
BMA-PC(1,0.5)	0.99	0.92	0.97	0.96	0.99	0.95	1.14	1.15	1.09
PEB-PC	0.97	0.99	0.86	0.99	0.91	0.96	1.29	1.18	1.11
BIC-PC	1.05	0.96	0.97	1.02	1.07	1.02	1.28	1.26	1.22
Bagging-PC	1.22	1.15	1.16	1.44	1.28	1.14	1.60	1.54	1.42

b. $h = 3$

	PI	IP	UR	EMP	TBILL	TBOND	PPI	CPI	PCED
Combined-Mean Root-MSFE	3.40	5.56	0.32	1.76	1.26	0.75	3.92	1.97	1.43
MSFE Relative to Combined-Mean									
AR	1.07	1.14	1.13	1.11	1.03	1.02	1.07	1.14	1.08
OLS	2.00	1.57	1.53	1.36	1.33	1.26	2.71	2.72	2.35
Combined-SSR	1.04	1.00	0.90	0.93	0.87	0.92	1.22	1.24	1.07
FAAR-OLS	0.98	0.85	0.85	0.91	0.91	0.96	0.98	0.91	0.95
FAA-GLS	0.99	0.96	0.84	1.05	0.88	0.94	1.01	0.92	0.98
FAAR-WLS	0.99	0.84	0.86	0.92	0.89	0.93	1.01	0.92	0.97
BMA(1/n^2,0.5)	0.96	0.96	0.88	0.98	0.88	0.95	1.30	1.31	1.09
BMA(1,0.5)	0.97	0.82	0.84	0.92	0.94	1.02	1.56	1.41	1.23
BMA-PC(1/n^2,0.5)	1.01	0.86	0.83	0.92	0.81	0.87	1.30	1.26	1.13
BMA-PC(1,0.5)	0.99	0.83	0.91	0.82	0.84	0.88	1.38	1.39	1.18
PEB-PC	0.96	0.83	0.82	0.89	0.82	0.91	1.42	1.30	1.19
BIC-PC	1.23	1.01	0.96	1.10	0.99	0.98	1.56	1.48	1.40
Bagging-PC	1.40	1.04	1.07	1.00	1.04	1.03	1.92	1.89	1.57

Table 3 (Continued)

c. h = 6

	PI	IP	UR	EMP	TBILL	TBOND	PPI	CPI	PCED
Combined-Mean Root-MSFE	2.39	4.15	0.50	1.64	1.66	1.06	3.08	1.71	1.19
MSFE Relative to Combined-Mean									
AR	1.08	1.18	1.20	1.06	1.06	1.02	1.07	1.17	1.10
OLS	2.58	2.64	1.75	2.07	1.35	1.47	3.07	2.32	2.69
Combined-SSR	1.16	1.12	1.01	0.90	0.90	0.98	1.24	1.24	1.14
FAAR-OLS	1.10	1.21	0.93	1.04	0.86	0.99	0.99	0.85	0.97
FAA-GLS	0.97	1.00	0.84	1.09	0.88	0.96	1.01	0.86	0.97
FAAR-WLS	1.13	1.20	0.92	1.04	0.80	0.99	0.99	0.84	0.97
BMA($1/n^2$,0.5)	1.22	1.46	1.19	1.38	0.86	1.08	1.35	1.24	1.19
BMA(1,0.5)	1.19	1.39	1.01	1.35	0.87	1.13	1.69	1.41	1.43
BMA-PC($1/n^2$,0.5)	1.12	1.09	0.97	1.12	0.79	0.99	1.41	1.20	1.17
BMA-PC(1,0.5)	1.07	0.99	0.95	1.03	0.86	0.98	1.47	1.27	1.26
PEB-PC	1.04	1.06	0.93	1.04	0.80	1.03	1.35	1.15	1.16
BIC-PC	1.46	1.30	1.15	1.57	0.98	1.15	1.93	1.38	1.68
Bagging-PC	1.71	1.71	1.23	1.64	1.07	1.25	2.19	1.71	1.94

d. h = 12

	PI	IP	UR	EMP	TBILL	TBOND	PPI	CPI	PCED
Combined-Mean Root-MSFE	1.86	3.40	0.84	1.66	2.12	1.57	2.73	1.59	1.12
MSFE Relative to Combined-Mean									
AR	1.05	1.25	1.21	1.15	1.12	1.02	1.14	1.28	1.19
OLS	2.28	2.83	2.07	2.30	2.15	1.75	3.16	2.62	3.44
Combined-SSR	1.14	1.16	1.05	1.07	1.11	0.97	1.28	1.17	1.11
FAAR-OLS	1.10	1.34	0.99	0.99	0.90	1.02	0.97	0.76	0.96
FAA-GLS	1.03	1.04	0.93	1.08	0.97	1.03	1.03	0.84	0.98
FAAR-WLS	1.14	1.38	0.92	1.01	0.93	1.01	0.95	0.74	0.99
BMA($1/n^2$,0.5)	1.41	1.80	1.47	1.61	1.21	1.33	1.41	1.29	1.17
BMA(1,0.5)	1.41	1.64	1.27	1.55	1.50	1.33	1.86	1.51	1.67
BMA-PC($1/n^2$,0.5)	1.19	1.22	1.05	1.21	1.22	1.15	1.47	1.23	1.40
BMA-PC(1,0.5)	1.04	1.12	1.04	1.15	1.25	1.05	1.50	1.28	1.36
PEB-PC	1.11	1.14	0.99	1.12	1.02	1.12	1.43	1.09	1.28
BIC-PC	1.57	1.88	1.45	1.58	1.61	1.52	1.93	1.60	2.02
Bagging-PC	1.76	2.14	1.54	1.83	1.85	1.57	2.28	1.80	2.42

Notes: The first row of the table shows the root mean square forecast error (MSFE) for the combined-mean forecast over the pseudo out-of-sample period. The other entries are the MSFE relative the MSFE for the combined-mean.

Table 4
Summary of Relative MSFE's for all Series and Horizons

Method	Average Rel. MSFE (Fraction Rel. MSFE < 1)		
	Full Out-of-Sample Period	Split Out-of-Sample Period	
		First Half	Second Half
AR	1.10 (0.00)	1.12 (0.00)	1.07 (0.03)
OLS	2.16 (0.00)	2.44 (0.00)	2.02 (0.00)
Combined-SSR	1.05 (0.39)	1.01 (0.50)	1.14 (0.22)
FAAR-OLS	0.96 (0.81)	0.96 (0.67)	1.00 (0.69)
FAAR-GLS	0.98 (0.61)	0.94 (0.67)	1.14 (0.44)
FAAR-WLS	0.96 (0.75)	0.95 (0.64)	1.02 (0.67)
BMA(1/n^2,0.5)	1.16 (0.31)	1.13 (0.33)	1.31 (0.17)
BMA(1,0.5)	1.23 (0.28)	1.17 (0.31)	1.49 (0.17)
BMA-PC(1/n^2,0.5)	1.07 (0.39)	1.01 (0.53)	1.24 (0.22)
BMA-PC(1,0.5)	1.08 (0.44)	1.07 (0.47)	1.16 (0.31)
PEB-PC	1.06 (0.42)	1.04 (0.42)	1.15 (0.33)
BIC-PC	1.34 (0.17)	1.33 (0.25)	1.51 (0.06)
Bagging-PC	1.54 (0.00)	1.61 (0.11)	1.63 (0.03)

Notes: The entries show the relative MSFE averaged across all 9 series and 4 horizons. The number in parentheses is the fraction of the 36 series/horizon entries for which the relative MSFE was less than 1.0. The relative MSFE shows the forecast MSFE relative to the MSFE for the combined -mean) model.

Table 5
Recursive Forecast MSFE Relative to Rolling Forecast MSFE

A. Summary by method across series and horizon

Method	Fraction <1	Mean
Combined-Mean	0.81	0.95
AR	0.92	0.93
OLS	0.64	0.95
Combined-SSR	0.69	0.99
FAAR-OLS	0.94	0.88
FAAR-GLS	0.94	0.91
FAAR-WLS	0.97	0.85
BMA(1/n^2,0.5)	0.75	0.96
BMA(1,0.5)	0.75	0.92
BMA-PC(1/n^2,0.5)	0.83	0.93
BMA-PC(1,0.5)	0.75	0.95
PEB-PC	0.92	0.89
BIC-PC	0.92	0.86
Bagging-PC	0.75	0.92

B. Summary by series across method and horizon

Series	Fraction < 1	Mean
PI	0.79	0.93
IP	0.73	0.97
UR	0.82	0.91
EMP	0.73	0.96
TBILL	0.89	0.88
TBOND	0.95	0.86
PPI	0.96	0.88
CPI	0.75	0.95
PCED	0.82	0.95

Notes: The entries in panel A are based on the recursive relative to rolling MSFE's for the 36 series/horizon forecasts. The second column (labeled Fraction < 1) shows the fraction of the 36 values that are less than 1.0 (that is, the fraction of series/horizons for which the MSFE of the recursive forecast was smaller than rolling forecast). The next columns shows the mean of these 36 values. Panel B is similar, except it is based on the 56 method/horizons forecasts for each series.

Table 6
Correlation of Forecasts
Averages across Series and Horizon

	Method	1	2	3	4	5	6	7	8	9	10	11	12	13	14
1	Combined-Mean	1.00	.	.	.	.	.	.	.	.	.	.	.	.	.
2	AR	***0.94***	1.00	.	.	.	.	.	.	.	.	.	.	.	.
3	OLS	0.43	0.36	1.00	.	.	.	.	.	.	.	.	.	.	.
4	Combined-SSR	0.75	0.65	0.47	1.00	.	.	.	.	.	.	.	.	.	.
5	FAAR-OLS	0.77	0.65	0.50	0.77	1.00	.	.	.	.	.	.	.	.	.
6	FAAR-GLS	0.73	0.61	0.53	0.73	**0.86**	1.00	.	.	.	.	.	.	.	.
7	FAAR-WLS	0.77	0.65	0.50	0.78	***0.98***	**0.86**	1.00	.	.	.	.	.	.	.
8	BMA(1/n^2,0.5)	0.65	0.56	0.59	0.79	0.77	0.73	0.77	1.00	.	.	.	.	.	.
9	BMA(1,0.5)	0.60	0.50	**0.80**	0.71	0.73	0.73	0.73	**0.86**	1.00	.	.	.	.	.
10	BMA-PC(1/n^2,0.5)	0.68	0.57	0.63	0.78	**0.82**	0.77	**0.82**	**0.82**	**0.83**	1.00	.	.	.	.
11	BMA-PC(1,0.5)	0.71	0.65	**0.87**	0.74	0.72	0.71	0.72	0.79	**0.88**	**0.87**	1.00	.	.	.
12	PEB-PC	0.67	0.57	0.66	0.77	**0.80**	0.75	**0.80**	**0.80**	**0.82**	***0.94***	**0.87**	1.00	.	.
13	BIC-PC	0.57	0.48	0.70	0.66	0.70	0.66	0.69	0.74	**0.80**	**0.88**	**0.85**	**0.84**	1.00	.
14	Bagging-PC	0.52	0.44	***0.96***	0.59	0.62	0.63	0.62	0.70	**0.87**	0.78	***0.94***	0.78	**0.84**	1.00

Notes: The entries are the average correlation (over 36 series×horizons) of the pseudo out-of-sample forecasts.

Table 7
Variance Decomposition of $\hat{Y}^h_{t+h}$

A. Average Across Series and Horizon

Method	Total	Components	
	$\frac{Var(\hat{Y}^h_{t+h})}{Var(Y^h_{t+h})}$	$\frac{Var(\hat{Y}^{h,AR}_{t+h})}{Var(Y^h_{t+h})}$	$\frac{Var(\hat{Y}^{h,Z}_{t+h})}{Var(Y^h_{t+h})}$
Combined-Mean	0.32	0.29	0.02
AR	0.31	0.28	0.02
OLS	1.23	0.21	0.78
Combined-SSR	0.40	0.21	0.17
FAAR-OLS	0.58	0.24	0.32
FAAR-GLS	0.39	0.22	0.15
FAAR-WLS	0.58	0.26	0.31
BMA(1/N^2,0.5)	0.53	0.23	0.27
BMA(1,0.5)	0.60	0.20	0.35
BMA-PC(1/N^2,0.5)	0.47	0.21	0.23
BMA-PC(1,0.5)	0.37	0.22	0.12
PEB-PC	0.45	0.22	0.21
BIC-PC	0.68	0.21	0.37
Bagging-PC	0.80	0.21	0.48

B. Average Across Method and Horizon

Series	Total	Components	
	$\frac{Var(\hat{Y}^h_{t+h})}{Var(Y^h_{t+h})}$	$\frac{Var(\hat{Y}^{h,AR}_{t+h})}{Var(Y^h_{t+h})}$	$\frac{Var(\hat{Y}^{h,Z}_{t+h})}{Var(Y^h_{t+h})}$
PI	0.47	0.03	0.37
IP	0.71	0.12	0.53
UR	0.52	0.07	0.40
EMP	0.57	0.24	0.30
TBILL	0.37	0.07	0.23
TBOND	0.27	0.06	0.16
PPI	0.73	0.56	0.13
CPI	0.70	0.49	0.17
PCED	0.61	0.43	0.15

Notes: Results in Panel A were computed using the recursively computed pseudo out-of-sample forecasts, and are average values across the 36 series×horizon forecasts. The second column shows the ratio of the variance of the forecast to the variance of the series' actual value. The last two columns decomposes the ratio in column (2) by expressing the forecast in terms of a component associated the AR lags and a component associated with the predictors X that is orthogonal to the AR lags. Column (2) is not the exact sum of columns (3) and (4) because of time variation in the recursively estimated model parameters. Panel B is similar, except it is based on the 56 method×horizons forecasts for each series.

Table 8
Full-Sample Parameter Estimators for the Empirical Bayes Forecasting Model

	Forecast Horizon							
Series	**1**		**3**		**6**		**12**	
	$\hat{p}$	$\hat{g}$	$\hat{p}$	$\hat{g}$	$\hat{p}$	$\hat{g}$	$\hat{p}$	$\hat{g}$
PI	0.01	0.06	0.01	0.05	0.08	0.14	0.09	0.15
IP	0.19	0.15	0.13	0.09	0.10	0.05	0.07	0.04
UR	0.02	0.04	0.04	0.04	0.03	0.03	0.27	0.04
EMP	0.01	0.03	0.10	0.08	0.13	0.09	0.12	0.06
TBILL	0.07	0.11	0.05	0.07	0.08	0.07	0.07	0.08
TBOND	0.37	1.00	0.41	0.63	0.48	0.42	0.24	0.22
PPI	0.60	1.36	0.04	0.13	0.01	0.04	0.06	0.10
CPI	0.46	0.28	0.01	0.03	0.01	0.02	0.04	0.04
PCED	0.22	0.46	0.02	0.09	0.01	0.04	0.05	0.08

Notes: This table shows the estimates of p and g for the empirical Bayes forecasts for T = 2003:12.

Figure 1

Shrinkage Factors for PC Forecasting Models

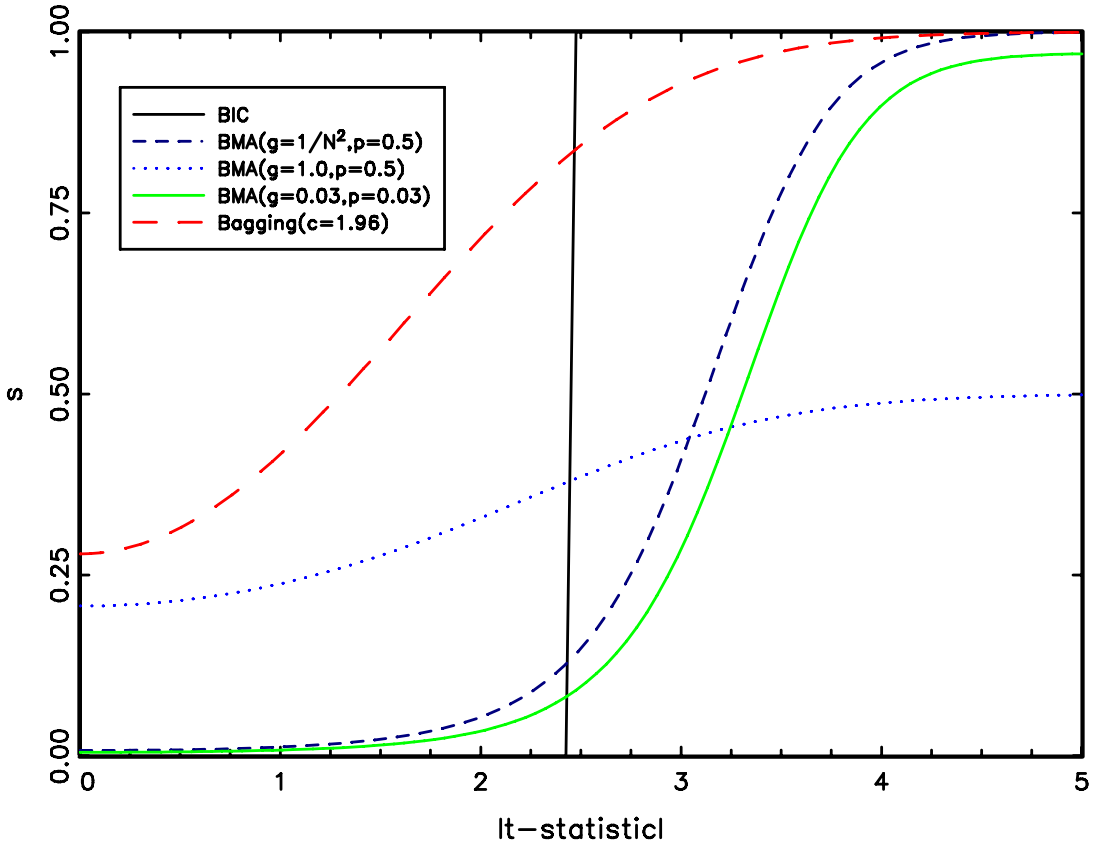
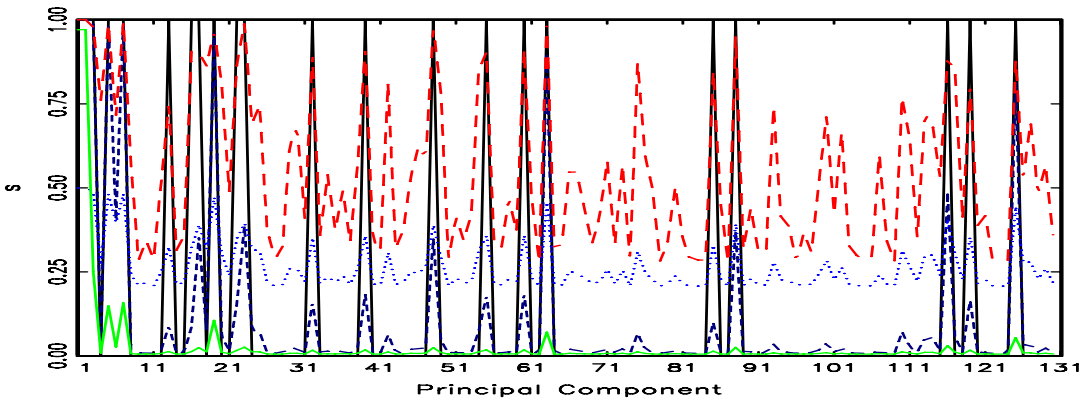
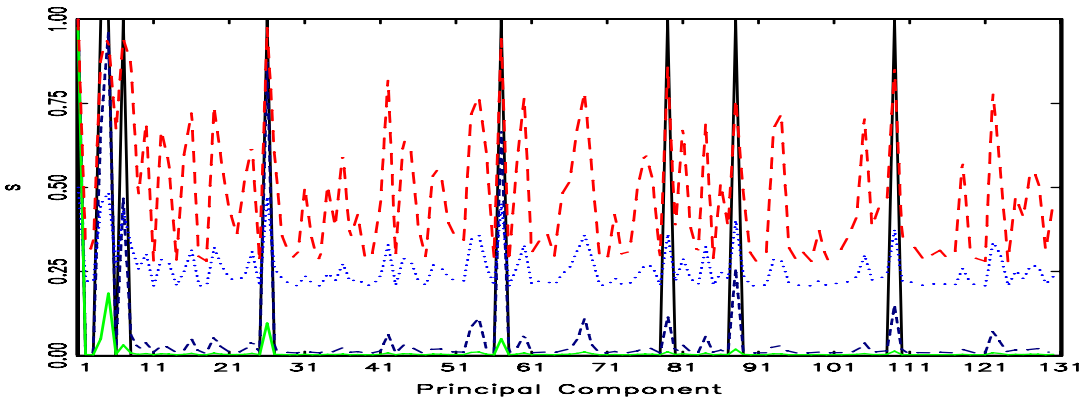


Figure 2
Shrinkage Factors for Principal Components
***h* = 6, *T* = 2003:12**

a. Unemployment Rate (UR)



b. CPI Inflation Rate (CPI)



C. 10-Year Treasury Bond Rate (TBOND)