How Useful is Bagging in Forecasting Economic Time Series? A Case Study of U.S. CPI Inflation*

Atsushi Inoue[†]
North Carolina State University

Lutz Kilian[‡]
University of Michigan

March 26, 2005

Abstract

This article explores the usefulness of bagging methods in forecasting economic time series from linear multiple regression models. We focus on the widely studied question of whether the inclusion of indicators of real economic activity lowers the prediction mean-squared error of forecast models of U.S. consumer price inflation. We study bagging methods for linear regression models with correlated regressors and for factor models. We compare the accuracy of simulated out-of-sample forecasts of inflation based on these bagging methods to that of alternative forecast methods, including factor model forecasts, shrinkage estimator forecasts, combination forecasts and Bayesian model averaging. We find that bagging methods in this application are almost as acccurate or more accurate than the best alternatives. Our empirical analysis demonstrates that large reductions in the prediction mean squared error are possible relative to existing methods, a result that is also suggested by the asymptotic analysis of some stylized linear multiple regression examples.

KEYWORDS: Bootstrap aggregation; Bayesian model averaging; Forecast combination; Factor models; Shrinkage estimation; Forecast model selection; Pre-testing.

*We thank Bob Stine for stimulating our interest in bagging. We acknowledge helpful discussions with Todd Clark, Silvia Gonçalves, Peter R. Hansen, Kirstin Hubrich, Mike McCracken, Massimiliano Marcellino, Serena Ng, Barbara Rossi, Jim Stock, Mark Watson, Jonathan Wright and Arnold Zellner. We thank seminar participants at Caltech, the CFS, the ECB, Johns Hopkins, Michigan State, Maryland, Purdue, and Tokyo. We have also benefited from comments received at the 2003 Triangle Econometrics Conference, the 2004 Financial Econometrics Conference in Waterloo, the 2004 Forecasting Conference at Duke, the 2004 North American Summer Econometric Society Meeting at Brown, the 2004 NBER-NSF Time Series Conference at SMU, the 2004 Midwest Econometrics Group Meeting at Northwestern, and the 2004 CEPR Conference on Forecast Combinations in Brussels.

[†]Department of Agricultural and Resource Economics, Box 8109, North Carolina State University, Raleigh, NC 27695-8109. E-mail: atsushi_inoue@ncsu.edu.

[‡]Department of Economics, University of Michigan, 611 Tappan Street, Ann Arbor, MI 48109-1220. E-mail: lkilian@umich.edu.

1 Introduction

A common problem in out-of-sample prediction is that the researcher suspects that many predictors are potentially relevant, but few (if any) of these predictors individually are likely to have high predictive power. This problem is particularly relevant in economic forecasting, because economic theory rarely puts tight restrictions on the set of potential predictors. In addition, often alternative proxies of the same variable are available to the economic forecaster. A case in point are forecasts of consumer price inflation, which may involve a large number of alternative measures of real economic activity such as the unemployment rate, industrial production growth, housing starts, capacity utilization rates in manufacturing, or the number of help wanted postings, to name a few.

It is well known that forecasts generated using only one of these proxies tend to be unreliable and unstable (see, e.g., Cecchetti, Chu and Steindel 2000, Stock and Watson 2003). On the other hand, including all proxies (even if feasible) is thought to lead to overfitting and poor out-of-sample forecast accuracy. This fact suggests that we use formal statistical methods for selecting the best subset of these predictors. Standard methods of comparing all possible combinations of predictors by means of an information criterion function, however, become computationally infeasible when the number of potential predictors is moderately large.[1]

One strategy in this situation is to combine forecasts from many models with alternative subsets of predictors. For example, one could use the mean, median or trimmed mean of these forecasts as the final forecast or one could use regression-based weights for forecast combination (see Bates and Granger 1969, Stock and Watson 2003). There is no reason, however, for simple averages to be optimal, and the latter approach of regression-based weights tends to perform poorly in practice, unless some form of shrinkage estimation is used (see, e.g., Stock and Watson 1999). More sophisticated methods of forecast model averaging weight individual forecasts by the posterior probabilities of each forecast model (see, e.g., Min and Zellner 1993, Avramov 2002, Cremers 2002, Wright 2003a and Koop and Potter 2003 for applications in econometrics). This Bayesian model averaging (BMA) approach has been used successfully in forecasting inflation by Wright (2003b). An alternative strategy involves shrinkage estimation of the unrestricted model that includes all potentially relevant predictors. Such methods are routinely used for example in the literature on Bayesian vector autoregressive models (see Litterman 1986). A third strategy is to reduce the dimensionality of the regressor set by extracting the principal components from the set of potential predictors. If the data are generated by an approximate factor model, then factors estimated by principal components analysis can be used for efficient forecasting under quite general conditions (see, e.g., Stock and Watson 2002a, 2002b; Bai and Ng 2004).[2]

A fourth strategy is to rely on a testing procedure for deciding which predictors to include in the forecast model and which to drop. For example, we may fit a model including all potentially

[1] See Inoue and Kilian (2004) for a discussion of this and related approaches to ranking competing forecast models. The difficulty in using information criteria when the number of potential predictors, M, is large is that the criterion must be evaluated for 2^M combinations of predictors. For $M > 20$ this task tends to become computationally prohibitive.

[2] A closely related approach to extracting common components has been developed by Forni et al. (2000, 2001) and applied in Forni et al. (2003).

relevant predictors, conduct a two-sided t-test for each predictor and discard all insignificant predictors prior to forecasting. Such pre-tests lead to inherently unstable decision rules in that small alterations in the data set may cause a predictor to be added or to be dropped. This instability tends to inflate the variance of the forecasts and may undermine the accuracy of pre-test forecasts in applied work. The predictive accuracy of simple pre-test strategies, however, may be greatly enhanced by application of the bagging technique, leading to a fifth strategy that will be the focus of this paper.

Bagging is a statistical method designed to reduce the out-of-sample prediction mean-squared error of forecast models selected by unstable decision rules such as pre-tests. The term *bagging* is short for *bootstrap aggregation* (see Breiman 1996). In essence, bagging involves fitting the unrestricted model including all potential predictors to the original sample, generating a large number of bootstrap resamples from this approximation of the data, applying the pre-test rule to each of the resamples, and averaging the forecasts from the models selected by the pre-test on each bootstrap sample.

By averaging across resamples, bagging effectively removes the instability of the decision rule. Hence, one would expect the variance of the bagged prediction model to be smaller than that of the model that would be selected based on the original data. Especially when the decision rule is unstable, this variance reduction may be substantial. In contrast, the forecast bias of the prediction model is likely to be of similar magnitude, with or without bagging. This heuristic argument suggests that bagging will reduce the prediction mean squared error of the regression model after variable selection. Indeed, there is substantial evidence of such reductions in practice. There are some counterexamples, however, in which this intuition fails and bagging does not improve forecast accuracy. This fact has prompted increased interest in the theoretical properties of bagging. Bühlmann and Yu (2002) recently have investigated the ability of bagging to lower the asymptotic prediction mean-squared error (PMSE) of regressions with a single regressor when the data are i.i.d. They show that bagging does not always improve on pre-testing, but nevertheless has the potential of achieving dramatic reductions in asymptotic forecast mean squared errors.

In this article, we explore the usefulness of bagging methods in forecasting economic time series from linear multiple regression models. Such forecasting models are routinely used by practitioners, but no attempt has been made to utilize bagging methods in this context.[3] In section 2, we show how the bagging proposal may be adapted to applications involving multiple regression models with possibly serially correlated and heteroskedastic errors. We briefly review the theory behind bagging, and - drawing on the analysis of the single-regressor model in Bühlmann and Yu (2002) - provide some intuition for how and when bagging works in the single-regressor model with iid data. We then investigate the asymptotic properties of bagging in the multiple regressor model. We discuss applications of bagging in the correlated regressor model as well as in factor models. Our analysis of some stylized examples shows that bagging has the potential of reducing the asymptotic prediction mean-squared error in the multiple regressor model. This result holds when we apply the bagging method to the M largest estimated factors of a factor model, where M is treated as fixed, as well as when bagging is applied to

[3]In related work, Lee and Yang (2004) study the properties of bagging in binary prediction problems and quantile prediction of economic time series data.

the regressors of a correlated regressor model. In the latter case, the potential for asymptotic gains arises, whether the regressors have been orthogonalized or not, except when the degree of correlation is very high.

While these theoretical results are encouraging, they are not dispositive. First, the extent to which the asymptotic gains in accuracy suggested by our theory translate into PMSE reductions in finite samples is unclear. Second, our asymptotic analysis shows that the relative performance of bagging will depend on unknown features of the data generating process, so the performance of bagging must be assessed case by case. Third, our asymptotic results treat the regressors as exogenously given. This simplifying assumption facilitates the derivation of asymptotic results. When regressors are possibly endogenous, as seems plausible in many applications in economics and finance, the asymptotic theory for the bagging predictor becomes intractable.[4]

We therefore recommend that, in practice, researchers choose between the alternative forecasting methods based on the ranking of their recursive PMSEs in simulated out-of-sample forecasts. In section 3, we illustrate this approach for a typical forecasting problem in economics. Specifically, we investigate whether one-month and twelve-month ahead CPI inflation forecasts for the United States may be improved upon by adding indicators of real economic activity to models involving only lagged inflation rates. This empirical example is in the spirit of recent work by Stock and Watson (1999, 2003), Marcellino et al. (2003), Bernanke and Boivin (2003), Forni et al. (2003) and Wright (2003b), among others.

We show that bagging is a very accurate forecasting procedure in this empirical application. Bagging outperforms the benchmark model involving only lags of inflation, the unrestricted model and factor models with rank 1, 2, 3, or 4 and different lag structures. Given that bagging may be viewed as a shrinkage estimator, we also compare its performance to Bayesian shrinkage estimators. We find that bagging forecasts in some cases are almost as accurate as the forecast from the best Bayesian shrinkage estimator and in the others more accurate. Bagging also is more accurate than forecast combination methods such as equal-weighted forecasts of models including one indicator of real economic activity at a time or the type of BMA studied by Wright (2003b). Finally, we show that bagging forecasts - depending on the horizon - are almost as accurate as or somewhat more accurate than BMA forecasts generated using the method of Raftery, Madigan and Hoeting (1997) that is based on randomly selected subsets of the predictors. The superior performance of bagging methods in this application is robust to increasing the number of potential predictors by 25% and to decreasing it by 25%.

We also contrast the relative performance of alternative methods of bagging in this context. While all bagging methods perform well in this application, bagging predictors based on the orthogonalized regressors are slightly more accurate than those based on the untransformed regressors. Bagging predictors based on the M largest principal components also worked well. This finding is surprising as the cross-sectional dimension of our problem is relatively small, casting doubt on the applicability of standard asymptotic arguments for bagging factor models. An interesting avenue for future research will be the use of bagging methods on panels with large cross-sections that are commonly used in other forecasting applications. We conclude in section 4.

[4] A similar exogeneity assumption has also been used in the literature to facilitate the derivation of the PMSE of factor model forecasts (see, e.g., Bai and Ng 2004, p.4).

2 How Does Bagging Work?

Consider the forecasting model:

$$y_{t+h} = \beta' x_t + \varepsilon_{t+h}, \quad h = 1, 2, 3, ... \tag{1}$$

where ε_{t+h} denotes the h-step ahead linear forecast error, β is an M-dimensional column vector of parameters and x_t is a column vector of M predictors at time period t. We presume that y_t and x_t are stationary processes or have been suitably transformed to achieve stationarity.

Let $\hat{\beta}$ denote the ordinary least-squares (OLS) estimator of β in (1) and let t_j denote the t-statistic for the null that β_j is zero in the unrestricted model, where β_j is the jth element of β. Further, let $\widehat{\gamma}$ denote the OLS estimator of the forecast model after variable selection. Note that - unlike Bühlmann and Yu (2002) - we re-estimate the model after variable selection. For $x_t \in \Re^M$, we define the predictor from the unrestricted model (UR), the predictor from the fully restricted model (FR), and the pre-test (PT) predictor conditional on x_{T-h+1} by

$$\begin{aligned}
\hat{y}^{UR}(x_{T-h+1}) &= \hat{\beta}' x_{T-h+1}, \\
\hat{y}^{FR}(x_{T-h+1}) &= 0, \\
\hat{y}^{PT}(x_{T-h+1}) &= 0, \text{ if } |t_j| < c \; \forall j \text{ and } \hat{y}^{PT}(x_{T-h+1}) = \widehat{\gamma}' S_T x_{T-h+1} \text{ otherwise,}
\end{aligned}$$

where S_T is the stochastic selection matrix obtained from the $M \times M$ diagonal matrix with $(i,i)th$ element $I(|t_i| > c)$ by deleting rows of zeros, and c is the critical value of the pre-test.

The UR model forecast is based on the fitted values of a regression including all M potential predictors. The FR model forecast emerges when all predictors are dropped, as in the well-known no-change forecast model of asset returns. The latter forecast sometimes is also referred to as a random walk forecast in the literature.

The pre-test strategy that we analyze is particularly simple. We first fit the unrestricted model that includes all potential predictors. We then conduct two-sided t-tests on each slope parameter based on a pre-specified critical value c. We discard the insignificant predictors and re-estimate the final model, before generating the PT forecast. In constructing the t-statistic we use appropriate standard errors that allow for serial correlation and/or conditional heteroskedasticity. Specifically, when the error term follows an MA($h-1$) process, the pre-test strategy may be implemented based on White (1980) robust standard errors for $h = 1$ or West (1997) robust standard errors for $h > 1$. For more general error structures, nonparametric robust standard errors such as the HAC estimator proposed by Newey and West (1987) would be appropriate.

2.1 Algorithm for Bagging Dynamic Regression Models

The bootstrap aggregated or bagging predictor is obtained by averaging the pre-test predictor across bootstrap replications. Bagging can in principle be applied to any pre-testing strategy, not just to the specific pre-testing strategy discussed here, and there is no reason to believe that our t-test strategy is optimal. Nevertheless, the simple t-test strategy studied here appears to work well in many cases.

Definition 1. [BA method] The bagging predictor in the standard regression framework is defined as follows:

(i) Arrange the set of tuples $\{(y_{t+h}, x_t')\}$, $t = 1, ..., T-h$, *in the form of a matrix of dimension* $(T-h) \times (M+1)$:

$$\begin{matrix} y_{1+h} & x_1' \\ \vdots & \vdots \\ y_T & x_{T-h}' \end{matrix} .$$

Construct bootstrap samples $(y_{1+h}^*, x_1'^*), ..., (y_T^*, x_{T-h}'^*)$ *by drawing with replacement blocks of* m rows of this matrix, *where the block size* m is chosen to capture the dependence in the error term *(see, e.g., Hall and Horowitz 1996, Gonçalves and White 2004).*

(ii) For each bootstrap sample, compute the bootstrap pre-test predictor conditional on x_{T-h+1}

$$\hat{y}^{*PT}(x_{T-h+1}) = 0, \text{ if } |t_j^*| < c \; \forall j \text{ and } \hat{y}^{*PT}(x_{T-h+1}) = \hat{\gamma}^{*\prime} S_T^* x_{T-h+1} \text{ otherwise,}$$

where $\hat{\gamma}^*$ *and* S_T^* *are the bootstrap analogues of* $\hat{\gamma}$ *and* S_T*, respectively. In constructing* $|t_j^*|$ *we compute the variance of* $\sqrt{T}\hat{\beta}^*$ *as* $\hat{H}^{*-1}\hat{V}^*\hat{H}^{*-1}$ *where*

$$\hat{V}^* = \frac{1}{bm}\sum_{k=1}^{b}\sum_{i=1}^{m}\sum_{j=1}^{m}(x_{(k-1)m+i}^*\varepsilon_{(k-1)m+i+h}^*)(x_{(k-1)m+j}^*\varepsilon_{(k-1)m+j+h}^*)',$$

$$\hat{H}^* = \frac{1}{bm}\sum_{k=1}^{b}\sum_{i=1}^{m}(x_{(k-1)m+i}^* x_{(k-1)m+i}^{*\prime}),$$

$\varepsilon_{t+h}^* = y_{t+h}^* - \hat{\beta}^{*\prime} x_t^*$, *and* b *is the integer part of* T/m *(see, e.g., Inoue and Shintani 2003).*

(iii) The bagged predictor is the expectation of the bootstrap pre-test predictor across bootstrap samples, conditional on x_{T-h+1}*:*

$$\hat{y}^{BA}(x_{T-h+1}) = E^*[\hat{\gamma}^{*\prime} S_T^* x_{T-h+1}],$$

where E^* *denotes the expectation with respect to the bootstrap probability measure. The bootstrap expectation in (iii) may be evaluated by simulation:*

$$\hat{y}^{BA}(x_{T-h+1}) = \frac{1}{B}\sum_{i=1}^{B}\hat{\gamma}^{*i\prime} S_T^{*i} x_{T-h+1},$$

where $B = \infty$ *in theory. In practice,* $B = 100$ *tends to provide a reasonable approximation.*

An important design parameter in applying bagging is the block size m. If the forecast model at horizon h is correctly specified in that $E(\varepsilon_{t+h}|\Omega_t) = 0$, where Ω_t denotes the date t

information set, then $m = h$ (see, e.g., Gonçalves and Kilian 2004). Otherwise $m > h$. In the latter case, data-dependent rules such as calibration may be used to determine m (see, e.g., Politis, Romano and Wolf 1999).

The performance of bagging will in general depend on the critical value chosen for pre-testing not unlike the way in which shrinkage estimators depend on the degree of shrinkage. In practice, c may be chosen by comparing the accuracy of the bagging forecast method for alternative values of c in simulated out-of-sample forecasts. This question will be taken up in section 3 when we discuss the empirical application.

2.2 Asymptotic Properties of the Bagging Predictor in the Single-Regressor Model

Bühlmann and Yu (2002) have analyzed the asymptotic properties of the bagging algorithm in Definition 1 for the special case of a linear model with only a single regressor when the data are i.i.d. They showed that the BA predictor in many (but not all) cases has lower asymptotic PMSE than the PT predictor. It is instructive to review this evidence. Let $\beta = \delta T^{-1/2}$, and, for expository purposes, suppose that $x_t = 1 \ \forall t$, ε_t is distributed $iid(0, \sigma^2)$, $\sigma^2 = 1$, and $h = 1$. In that case, the forecasts from the unrestricted (UR) model, the fully restricted (FR) model and the pre-test (PT) model, and the bagging (BA) forecast can be written as

$$\begin{aligned}
\hat{y}^{UR} &= \hat{\beta}, \\
\hat{y}^{FR} &= 0, \\
\hat{y}^{PT} &= \widehat{\beta} I(|T^{1/2}\hat{\beta}| > c), \\
\hat{y}^{BA} &= \frac{1}{B}\sum_{i=1}^{B} \widehat{\beta}^{*i} I(|T^{1/2}\hat{\beta}^{*i}| > c).
\end{aligned}$$

We are interested in comparing the asymptotic PMSE of these predictors.

Definition 2. [APMSE] The asymptotic PMSE (or APMSE) is defined as the second-order term of the asymptotic approximation of the prediction mean squared error:

$$E[(\hat{y}(x) - y(x))^2] = \sigma^2 + \frac{1}{T} APMSE(\hat{y}(x)) + o\left(\frac{1}{T}\right).$$

Following Bühlmann and Yu (2002) it can be shown that:

$$\begin{aligned}
APMSE(\hat{y}^{UR}(x)) &= 1, \\
APMSE(\hat{y}^{FR}(x)) &= \delta^2, \\
APMSE(\hat{y}^{PT}(x)) &= E[(\xi - \delta) I(|\xi| > c) + \delta I(|\xi| \leq c)]^2, \\
APMSE(\hat{y}^{BA}(x)) &= E[\delta - \xi + \xi\Phi(c - \xi) - \phi(c - \xi) \\
&\quad -\xi\Phi(-c - \xi) + \phi(-c - \xi)]^2.
\end{aligned}$$

where $\xi \sim N(\delta, 1)$.

How does the $APMSE$ of the BA predictor compare to that of the PT predictor? Note that the $APMSE$ expression for the BA predictor does not depend on the indicator function, reflecting the smoothing implied by bootstrap aggregation. Although this smoothing should typically help to reduce the forecast variance relative to the PT predictor, it is not obvious a priori whether bagging the pre-test predictor will also improve the $APMSE$. Figure 2 investigates this question. We set $c = 1.96$ for expository purposes. The upper panel shows the squared asymptotic bias of the two predictors. Although bagging does reduce the asymptotic bias somewhat for most values of δ, the gains are small. The second panel, in contrast, shows dramatic reductions in variance relative to the PT predictor for most δ, which, as shown in the third panel, result in substantial improvements in the overall accuracy measured by the $APMSE$. Figure 2 illustrates the potential of the bagging principle to improve forecast accuracy relative to the pre-test. As noted by Bühlmann and Yu (2002), although this improvement does not occur for all values of δ, it does for a wide range of δ.

More importantly, from our point of view, it can be shown that for some values of δ bagging will have lower $APMSE$ than the FR and UR predictors as well. We illustrate this fact in Figure 2, which plots the $APMSEs$ of the UR, FR, PT and BA predictors as a function of δ. For $\delta > 1$ the UR predictor has lower $APMSE$ than the FR predictor, for $\delta = 1$ both models are tied and for $\delta < 1$ the FR predictor is asymptotically more accurate. Although the PT predictor protects the user from choosing the UR forecast when δ is close to zero and the FR forecast when δ is large, the PT predictor for any given choice of δ is always dominated by either the UR or the FR predictor.[5] In contrast, the BA predictor not only dominates the PT predictor for most values of δ, but for values of δ near one, it has the lowest $APMSE$ of all predictors shown in Figure 2.

This stylized example based on Bühlmann and Yu (2002) conveys two valuable insights: First, bagging under certain conditions *can* yield asymptotic improvements in the PMSE relative to the UR, FR, and PT predictors. This fact suggests that it deserves further study. Second, the extent of these asymptotic improvements depends very much on unobservable features of the data. Under some conditions bagging may actually result in a higher asymptotic PMSE than alternative methods. This seems especially likely when the signal-to-noise ratio in the data is very weak, as in forecasting asset returns for example. This is not a limitation of the bagging method alone, of course, but simply a reflection of the bias-variance trade-off in forecasting. The same type of problem would arise with any other forecasting method in the literature.

2.3 Asymptotic Properties of the Bagging Predictor in the Correlated-Regressor Model

The example of Bühlmann and Yu (2002), while instructive, is of limited relevance for applications of bagging to the multiple linear regression model. In practice, predictors will inevitably be correlated to various degrees and this correlation will affect the PMSE and potentially the ranking of the forecasting methods. In this subsection, we will establish that the qualitative findings of Bühlmann and Yu (2002) for the single-regressor model continue to hold in the multiple regressor model. We do so in the simplest possible setting when the regressors are iid.

[5] For a related discussion of the MSE of inequality constrained estimators see Thomson and Schmidt (1982).

2.3.1 Case 1: Bagging the Untransformed Predictors

We begin by deriving the $APMSE$ for the UR, FR, PT, and BA predictors in the correlated regressor model with iid regressors.

Assumption 1.

(a) x_t and ε_t are iid over time with finite fourth moments and x_t and ε_t are independent of one another.

(b) $y_t = \beta' x_t + \varepsilon_t$ where $\beta = T^{-1/2}\delta$.

(c) $T^{-1/2}\sum_{t=1}^{T} x_t\varepsilon_t \xrightarrow{d} N(0, \sigma^2 E(x_t x_t'))$ where $\sigma^2 > 0$ and $E(x_t x_t')$ is positive definite.

Proposition 1. Under Assumption 1

$$
\begin{aligned}
APMSE(\hat{y}^{UR}(x)) &= E[(\xi-\delta)'x]^2, \\
APMSE(\hat{y}^{FR}(x)) &= (\delta' x)^2, \\
APMSE(\hat{y}^{PT}(x)) &= E[\xi' E(x_t x_t') S'(SE(x_t x_t')S')^{-1} SxI(\exists j \text{ s.t. } |\xi_j| > c\sqrt{\sigma^2[(E(x_t x_t'))^{-1}]_{jj}}) - \delta' x]^2, \\
APMSE(\hat{y}^{BA}(x)) &= E\{E[\xi^{*\prime} E(x_t x_t') S^{*\prime}(S^* E(x_t x_t') S^{*\prime})^{-1} S^* xI(\exists j \text{ s.t.} |\xi_j^*| > c\sqrt{\sigma^2[(E(x_t x_t'))^{-1}]_{jj}})|\xi] - \delta' x\}^2
\end{aligned}
$$

where ξ and ξ^* are M-dimensional random vectors such that $\xi \sim N(\delta, \sigma^2[E(x_t x_t')]^{-1})$ and $\xi^*|\xi \sim N(\xi, \sigma^2(E(x_t x_t'))^{-1})$, S is the stochastic selection matrix obtained from the $M \times M$ diagonal matrix with $(i,i)th$ element $I(|\xi_i| > c\sqrt{\sigma^2[(E(x_t x_t'))^{-1}]_{ii}})$ by deleting rows of zeros, and S^* is defined as S with ξ replaced by ξ^*.

The proof of Proposition 1 is in the Appendix.

The results of Proposition 1 may be used to study the relative merits of these forecasting strategies for a given data generating process. Although the APMSE has no closed form solution it is straightforward to evaluate the APMSE given by Proposition 1 by simulation. In the examples below, we evaluate all expectations based on 5000 random draws. To facilitate graphical representations, we focus on the simplest possible correlated regressor model. We postulate that $M = 2$. Let $\rho\epsilon\{0, 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9\}$ denote the correlation between these regressors. Furthermore, let $\delta = (\delta_1, \delta_2)'$where δ_1 varies in the range between 0 and 2, as in the single-regressor model, while δ_2 is arbitrarily fixed at 0.55 for expository purposes. Similar results are obtained for other choices of δ_2.To facilitate comparisons with the single-regressor example, we postulate that ε_t is distributed $iid(0,1)$, and we evaluate all forecasts at $x_T = \mathbf{1}$.

Figure 3 plots the relative $APMSE$ of the PT and BA methods as a function of δ_1 and ρ. Ratios above unity indicate that the BA predictor has lower $APMSE$. In this and the subsequent figures we cut off ratios below unity by imposing a plane of unit height on the graph. This plane highlights the contours of the region, in which the BA predictor works better than the alternative(s). The results in Figure 3 are stronger than for the single-regressor case in that the PT predictor has lower APMSE for all combinations of δ_1 and ρ. More importantly,

Figure 4 establishes that there is a set of pairs of δ_1 and ρ, for which the BA predictor has lower asymptotic PMSE than any of the other methods under consideration. As expected the BA predictor works best for intermediate ranges of δ_1. When δ_1 is very small or very large, the FR and UR predictors, respectively, will be more accurate than the BA predictor. The range of δ_1, for which the BA predictor.has the lowest $APMSE$, shrinks somewhat, as ρ increases.

The results in Figures 3 and 4 of course are specific to the stylized example. Here our aim has only been to establish that there are potential asymptotic gains in accuracy from bagging even in the correlated regressor model. How large these gains from bagging will be in practice, will depend on unknown features of the data generating process and is an empirical question.

2.3.2 Case 2: Bagging the Orthogonalized Predictors

One seeming drawback of the bagging proposal in Definition 1 is that when predictors are correlated the effective size of the t-tests on individual predictors will be distorted. This fact suggests an alternative approach to bagging in which the predictors are orthogonalized prior to conducting the t-tests. This may be accomplished as follows:

Definition 3. [CBA method] The bagging predictor for the orthogonalized regressors may be obtained via a Cholesky decomposition as follows:

(i) Arrange the set of tuples $\{(y_{t+h}, x_t')\}$, $t = 1, ..., T-h$, *in the form of a matrix of dimension* $(T-h) \times (M+1)$:

$$\begin{matrix} y_{1+h} & x_1' \\ \vdots & \vdots \\ y_T & x_{T-h}' \end{matrix} .$$

Construct bootstrap samples $(y_{1+h}^*, x_1'^*), ..., (y_T^*, x_{T-h}'^*)$ *by drawing with replacement blocks of* m rows of this matrix*, where the block size* m is chosen to capture the dependence in the error term.

(ii) Compute the orthogonalized predictor $\widetilde{x}_t = P'^{-1} x_t$, where P is the Cholesky decomposition of $E(x_t x_t')$, i.e., the $M \times M$ upper triangular matrix such that $P'P = E(x_t x_t')$. *For each bootstrap sample, compute the bootstrap pre-test predictor conditional on* $\widetilde{x}_{T-h+1}$

$$\hat{y}^{*PT}(\widetilde{x}_{T-h+1}) = 0, \text{ if } |t_j^*| < c \ \forall j \text{ and } \hat{y}^{*PT}(\widetilde{x}_{T-h+1}) = \widehat{\gamma}^{*\prime} S_T^* \widetilde{x}_{T-h+1} \text{ otherwise,}$$

where $\widehat{\gamma}^*$ *and* S_T^* *are the bootstrap analogues of* $\widehat{\gamma}$ *and* S_T*, respectively, applied to the orthogonalized predictor model. In constructing* $|t_j^*|$ *we compute the variance of* $\sqrt{T}\widehat{\beta}^*$ *as* $\widehat{H}^{*-1}\widehat{V}^*\widehat{H}^{*-1}$ *where*

$$\widehat{V}^* = \frac{1}{bm}\sum_{k=1}^{b}\sum_{i=1}^{m}\sum_{j=1}^{m}(\widetilde{x}^*_{(k-1)m+i}\widetilde{\varepsilon}^*_{(k-1)m+i+h})(\widetilde{x}^*_{(k-1)m+j}\widetilde{\varepsilon}^*_{(k-1)m+j+h})',$$

$$\widehat{H}^* = \frac{1}{bm}\sum_{k=1}^{b}\sum_{i=1}^{m}(\widetilde{x}^*_{(k-1)m+i}\widetilde{x}^{*\prime}_{(k-1)m+i}),$$

$\widetilde{\varepsilon}^*_{t+h} = y^*_{t+h} - \widehat{\beta}^{*\prime}\widetilde{x}^*_t$, *and* b *is the integer part of* T/m.

(*iii*) *The bagged predictor is the expectation of the bootstrap pre-test predictor across bootstrap samples, conditional on* $\widetilde{x}_{T-h+1}$*:*

$$\hat{y}^{CBA}(\widetilde{x}_{T-h+1}) = E^*[\widehat{\gamma}^{*\prime}S^*_T\widetilde{x}_{T-h+1}],$$

where E^* *denotes the expectation with respect to the bootstrap probability measure. The bootstrap expectation in* (*iii*) *may be evaluated by simulation based on* B bootstrap replications*:*

$$\hat{y}^{CBA}(\widetilde{x}_{T-h+1}) = \frac{1}{B}\sum_{i=1}^{B}\widehat{\gamma}^{*i\prime}S^{*i}_T\widetilde{x}_{T-h+1}.$$

It is unclear a priori whether the CBA method will select superior forecast models. In the context of bagging the purpose of the pre-test is to select a forecast model with lower PMSE, not to uncover the true relationship in the data. As we have already seen, notwithstanding the existence of size distortions, the BA method may lower the APMSE even in the presence of correlated regressors. An interesting question is whether the performance of bagging may be improved by orthogonalizing the predictors. For this purpose, we now derive the APMSE for the CBA predictor in the correlated regressor model with i.i.d. regressors, given by Assumption 1. We also derive the APMSE of the corresponding pre-test predictor based on the orthogonalized regressors (CPT).

Proposition 2. Under Assumption 1

$$APMSE(\hat{y}^{CPT}(\tilde{x})) = E(\tilde{\xi}'\tilde{S}'\tilde{S}\tilde{x}I(\exists\, j \text{ s.t. } |\tilde{\xi}_j| > c\sigma) - \delta' x]^2,$$
$$APMSE(\hat{y}^{CBA}(\tilde{x})) = E[E(\tilde{\xi}^{*\prime}\tilde{S}^{*\prime}\tilde{S}^*\tilde{x}I(\exists\, j \text{ s.t. } |\tilde{\xi}^*_j| > c\sigma)|\xi) - \delta' x]^2$$

where $\tilde{x}, \tilde{\xi}$ and $\tilde{\xi}^*$ are M-dimensional random vectors such that $\tilde{x} = P'^{-1}x$, $\tilde{\xi} \sim N(P\delta, \sigma^2 I_M)$, and $\tilde{\xi}^*|\tilde{\xi} \sim N(\tilde{\xi}, \sigma^2 I_M)$, $\tilde{S}$ is the stochastic selection matrix obtained from the $M \times M$ diagonal matrix with $(i,i)th$ element $I(|\tilde{\xi}_i| > c\sigma)$ by deleting rows of zeros, $\tilde{S}^*$ is defined as $\tilde{S}$ with $\tilde{\xi}$ replaced by $\tilde{\xi}^*$, and P is the Cholesky decomposition of $E(x_t x_t')$, that is, the $M \times M$ upper triangular matrix, P, such that $P'P = E(x_t x_t')$.

The proof of Proposition 2 is in the Appendix.

Returning to the stylized example with $M = 2$, we now investigate the relative $APMSE$ of the CBA predictor. Figure 5 plots the $APMSE$ ratios of the CBA predictor relative to the best alternative model. The underlying model is the same as that for Figures 3 and 4. As in Figure 4, there is a range of pairs of δ_1 and ρ, for which the CBA predictor has lower $APMSE$ than any other method under consideration. This range is slightly smaller than for the BA predictor when ρ is large. There is no clear ranking of these two bagging methods in terms of the $APMSE$, however.

Obviously, there is no reason for this stylized example to be representative. Actual gains in accuracy may be higher or lower than shown here. Nevertheless, this example establishes that bagging predictors, whether based on untransformed or orthogonalized predictors, may have desirable asymptotic properties at least under some circumstances.

2.4 Asymptotic Properties of the Bagging Predictor in the Factor Model

Bagging methods for the correlated regressor model are not designed to handle situations when the regressor matrix is of reduced rank. A leading example of a reduced rank structure is a factor model. In that case the forecasting model reduces to:

$$y_{t+h} = \beta' f_t + \varepsilon_{t+h}, \qquad h = 1, 2, 3, ... \tag{2}$$

where f_t denotes a vector of the M largest factors which may be extracted from the set of N potential predictors by principal components analysis (see, e.g., Stock and Watson 2002a,2002b). We treat M as fixed with respect to T. By construction we require $M < T$. It is straightforward to adapt the bagging method to this situation.

Definition 4. [BA^F method] The bagging predictor in the factor model framework is defined as follows:

(*i*) *Use principal components analysis to extract the* M largest common factors from the $T \times N$ matrix X of potential predictors. Denote the date t observation of these factor estimates by the $M \times 1$ vector $\widehat{f}_t$.

(*ii*) *Arrange the set of tuples* $\{(y_{t+h}, \widehat{f}_t')\}$, $t = 1, ..., T-h$, *in the form of a matrix of dimension* $(T-h) \times (M+1)$:

$$\begin{matrix} y_{1+h} & \widehat{f}_1' \\ \vdots & \vdots \\ y_T & \widehat{f}_{T-h}' \end{matrix} .$$

Construct bootstrap samples $(y^*_{1+h}, \widehat{f}_1'^*), ..., \{(y^*_T, \widehat{f}'^*_{T-h})$ *by drawing with replacement blocks of* m rows of this matrix*, where the block size* m is chosen to capture the dependence in the error term*, and subsequently orthogonalizing the bootstrap factor draws via principal components.*

(*iii*) *For each bootstrap sample, compute the bootstrap pre-test predictor conditional on* $\widehat{f}_{T-h+1}$

$$\hat{y}^{*PT}(\widehat{f}_{T-h+1}) = 0, \text{ if } |t^*_j| < c \ \forall j \text{ and } \hat{y}^{*PT}(\widehat{f}_{T-h+1}) = \widehat{\gamma}^{*\prime} S^*_T \widehat{f}_{T-h+1} \text{ otherwise,}$$

where $\widehat{\gamma}^*$ and S_T^* are the bootstrap analogues of $\widehat{\gamma}$ and S_T, respectively, applied to the factor model.

In constructing $|t_j^*|$, compute the variance of $\sqrt{T}\widehat{\beta}^*$ as $\widehat{H}^{*-1}\widehat{V}^*\widehat{H}^{*-1}$ where

$$\begin{aligned}
\widehat{V}^* &= \frac{1}{bm}\sum_{k=1}^{b}\sum_{i=1}^{m}\sum_{j=1}^{m}(\widehat{f}^*_{(k-1)m+i}\widehat{\varepsilon}^*_{(k-1)m+i+h})(\widehat{f}^*_{(k-1)m+j}\widehat{\varepsilon}^*_{(k-1)m+j+h})', \\
\widehat{H}^* &= \frac{1}{bm}\sum_{k=1}^{b}\sum_{i=1}^{m}(\widehat{f}^*_{(k-1)m+i}\widehat{f}^{*\prime}_{(k-1)m+i}),
\end{aligned}$$

$\widehat{\varepsilon}^*_{t+h} = y^*_{t+h} - \widehat{\beta}^{*\prime}\widehat{f}^*_t$, and b is the integer part of T/m.

(iv) The bagged predictor is the expectation of the bootstrap pre-test predictor across bootstrap samples, conditional on $\widehat{f}_{T-h+1}$:

$$\hat{y}^{BA^F}(\widehat{f}_{T-h+1}) = E^*[\widehat{\gamma}^{*\prime}S_T^*\widehat{f}_{T-h+1}],$$

where E^* denotes the expectation with respect to the bootstrap probability measure. The bootstrap expectation in (iv) may be evaluated by simulation based on B bootstrap replications*:*

$$\hat{y}^{BA^F}(\widehat{f}_{T-h+1}) = \frac{1}{B}\sum_{i=1}^{B}\widehat{\gamma}^{*i\prime}S_T^{*i}\widehat{f}_{T-h+1}.$$

We now derive the $APMSE$ of the UR, FR, PT and BA predictors applied to the factor model:

Assumption 2.

(a) f_t and ε_t are iid over time with finite fourth moments and f_t and ε_t are independent of one another.

(b) $y_t = \beta' f_t + \varepsilon_t$ where $\beta = T^{-1/2}\delta$.

(c) $T^{-1/2}\sum_{t=1}^{T} f_t\varepsilon_t \xrightarrow{d} N(0, \sigma^2 I_M))$ where $\sigma^2 > 0$.

(d) $\hat{f}_t - f_t = O_p(T^{-1/2})$ uniformly in t and $\text{plim}(1/T)\sum_{t=1}^{T}(\hat{f}_t - f_t)\varepsilon_t = 0$.

Proposition 3. Under Assumption 2

$$\begin{aligned}
APMSE(\hat{y}^{UR^F}(x)) &= E(\xi' x - \delta' x)^2, \\
APMSE(\hat{y}^{FR^F}(x)) &= (\delta' x)^2, \\
APMSE(\hat{y}^{PT^F}(x)) &= E[\xi' S' S x I(\exists\, j \text{ s.t. } |\xi_j| > c\sigma) - \delta' x]^2, \\
APMSE(\hat{y}^{BA^F}(x)) &= E[E(\xi^{*\prime} S^{*\prime} S^* x I(\exists\, j \text{ s.t. } |\xi_j^*| > c\sigma)|\xi) - \delta' x]^2,
\end{aligned}$$

where ξ and ξ^* are M-dimensional random vectors such that $\xi \sim N(\delta, \sigma^2 I_M)$, $\xi^*|\xi \sim N(\xi, \sigma^2 I_M)$, S is the stochastic selection matrix obtained from the $M \times M$ diagonal matrix with $(i,i)th$ element $I(|\xi_i| > c\sigma$ by deleting rows of zeros, and S^* is defined as S with ξ replaced by ξ^*.

The proof of Proposition 3 is in the Appendix.

Assumption 2(d) is satisfied if $\sqrt{T}/N \to 0$ as $N, T \to \infty$ and if the factors are exogenous (see Theorem 2 of Bai and Ng, 2004).[6] The assumption that $\sqrt{T}/N \to 0$ is appealing when $N > T$, as is the case in many applications of factor models. Since the estimation uncertainty about the factors is negligible asymptotically under these assumptions, and the factors are orthogonal by construction, the asymptotic properties of the bagging predictor based on $M < T < N$ estimated factors will be the same as for the bagging predictor in the standard regression model with M orthogonal predictors. From Figure 4 it is immediately apparent that the bagging predictor may have lower asymptotic PMSE than the UR, FR or PT models for certain combinations of δ and $\rho = 0$.

When $T > N$, as in our empirical application in the next section, one would not expect the usual asymptotic approximation to work well. Thus, we have no reason to expect the bagging predictor based on the M largest estimated factors to be any more accurate than other methods. We nevertheless will include the BA^F method in the empirical section as an additional competitor.

3 Application: Do Indicators of Real Economic Activity Improve the Accuracy of U.S. Inflation Forecasts?

There are two main limitations of the asymptotic analysis in the preceding section. First, in the theoretical analysis we have treated the regressors as exogenous, which rarely will be appropriate in applied work, and we have focused on a very stylized example. In more general settings, it is difficult to work out analytical solutions for the asymptotic PMSE of the bagging method in multiple regression, and indeed not particularly informative since we do not know the properties of the data generating process and cannot consistently estimate the relevant parameter δ (or its multiple regression analogue). Second, nothing ensures that the finite-sample properties of bagging are similar to its asymptotic properties. We therefore have no way of knowing a priori whether the data generating process in a given empirical application will favor bagging or some other forecasting method. It is also unclear which of the three alternative bagging methods discussed in section 2 will work best in a given application.

The question of whether bagging works better than the alternatives must be resolved on a case-by-case basis. We recommend that, in practice, researchers choose between competing forecasting methods based on the ranking of their recursive PMSE in simulated out-of-sample forecasts. The model with the lower recursive PMSE up to date $T - h$ will be chosen for forecasting y_{T+1}. We will illustrate this approach in this section for a typical forecasting problem

[6] When N diverges at a slower rate than T, the analysis becomes less tractable. We do not pursue this question here.

in economics.

We investigate whether one-month and twelve-months ahead U.S. CPI inflation forecasts may be improved upon by adding indicators of real economic activity to models involving only lagged inflation rates. This empirical example is in the spirit of recent work by Stock and Watson (1999), Bernanke and Boivin (2003), Forni et al. (2003), and Wright (2003b), among others. The choice of the benchmark model is conventional (see, e.g., Stock and Watson 2003, Forni et al. 2003) as is the focus on the PMSE. The lag order of the benchmark model is determined by the AIC subject to an upper bound of 12 lags. The optimal model is determined recursively in real time, so the lag order may change as we move through the sample.

Since there is no universally agreed upon measure of real economic activity we consider 26 potential predictors that can be expected to be correlated with real economic activity. A complete variable list is provided at the end of the paper. We obtain monthly data for the United States from the Federal Reserve Bank of St. Louis data base (FRED) and the Federal Reserve Board. Note that measures of wage cost and productivity are not available at monthly frequency for our sample period. We convert all data with the exception of the interest rates into annualized percentage growth rates. Interest rates are expressed in percent. Data are used in seasonally adjusted form where appropriate. All predictor data are standardized (i.e., demeaned and scaled to have unit variance and zero mean), as is customary in the factor model literature. We do not attempt to identify and remove outliers.

3.1 Unrestricted, Pre-Test and Dynamic Factor Model Forecasts

The alternative forecasting strategies under consideration in the first round of comparisons include the benchmark model involving only an intercept and lags of monthly inflation and eleven models that include in addition at least some indicators of economic activity. The unrestricted regression model (UR) includes one or more lags of all 26 indicators of economic activity as separate regressors in addition to lagged inflation. The pre-test predictor (PT, CPT) uses only a subset of these additional predictors. Similarly, the pre-test predictor based on the factor model (PT^F) uses only a subset of the 26 principal components of the set of indicators. The subsets for the pre-test strategy are selected using 2-sided t-tests for each predictor. We experimented with a range of critical values c.

The bagging forecast (BA, CBA, BA^F) is the average of the corresponding pre-test forecasts across 100 bootstrap replications with $M = 26$. For the one-month ahead forecast model there is no evidence of serial correlation in the unrestricted model, so we use White (1980) robust standard errors for the pre-tests and the pairwise bootstrap. For the twelve-month ahead-forecast we use West (1997) standard errors with a truncation lag of 11 and the block bootstrap with $m = 12$. Throughout this section we set $B = 100$.

Finally, we also fit factor models with rank $r \epsilon \{1, 2, 3, 4\}$ to the 26 potential predictors and generate forecasts by adding one or more lagged values of this factor to the benchmark model (FM). We estimate the factors by principal components analysis as in Stock and Watson (2002a, 2002b). To summarize, the forecast methods under consideration are:

$$
\begin{aligned}
Benchmark &: \pi^h_{t+h|t} = \widehat{\alpha} + \sum\nolimits_{k=1}^{p} \widehat{\phi}_k \pi_{t-k} \\
UR &: \pi^h_{t+h|t} = \widehat{\alpha} + \sum\nolimits_{k=1}^{p} \widehat{\phi}_k \pi_{t-k} + \sum\nolimits_{l=1}^{q} \sum\nolimits_{j=1}^{M} \widehat{\beta}_{jl} x_{j,t-l+1} \\
FM &: \pi^h_{t+h|t} = \widehat{\alpha} + \sum\nolimits_{k=1}^{p} \widehat{\phi}_k \pi_{t-k} + \sum\nolimits_{l=1}^{q} \widehat{\theta}_l \, \widehat{f}_{t-l+1} \\
PT &: \pi^h_{t+h|t} = \widehat{\alpha} + \sum\nolimits_{k=1}^{p} \widehat{\phi}_k \pi_{t-k} + \sum\nolimits_{l=1}^{q} \sum\nolimits_{j=1}^{M} \widehat{\gamma}_{jl} I(|t_{jl}| > c) x_{j,t-l+1} \\
CPT &: \pi^h_{t+h|t} = \widehat{\alpha} + \sum\nolimits_{k=1}^{p} \widehat{\phi}_k \pi_{t-k} + \sum\nolimits_{l=1}^{q} \sum\nolimits_{j=1}^{M} \widehat{\gamma}_{jl} I(|t_{jl}| > c) \widetilde{x}_{j,t-l+1} \\
PT^F &: \pi^h_{t+h|t} = \widehat{\alpha} + \sum\nolimits_{k=1}^{p} \widehat{\phi}_k \pi_{t-k} + \sum\nolimits_{l=1}^{q} \sum\nolimits_{j=1}^{M} \widehat{\gamma}_{jl} I(|t_{jl}| > c) \widehat{f}_{j,t-l+1} \\
BA &: \pi^h_{t+h|t} = \frac{1}{100} \sum_{i=1}^{100} \left(\widehat{\alpha}^{*i} + \sum\nolimits_{k=1}^{p} \widehat{\phi}^{*i}_k \pi_{t-k} + \sum\nolimits_{l=1}^{q} \sum\nolimits_{j=1}^{M} \widehat{\gamma}^{*i}_{jl} I(|t^{*i}_{jl}| > c) x_{j,t-l+1} \right) \\
CBA &: \pi^h_{t+h|t} = \frac{1}{100} \sum_{i=1}^{100} \left(\widehat{\alpha}^{*i} + \sum\nolimits_{k=1}^{p} \widehat{\phi}^{*i}_k \pi_{t-k} + \sum\nolimits_{l=1}^{q} \sum\nolimits_{j=1}^{M} \widehat{\gamma}^{*i}_{jl} I(|t^{*i}_{jl}| > c) \widetilde{x}_{j,t-l+1} \right) \\
BA^F &: \pi^h_{t+h|t} = \frac{1}{100} \sum_{i=1}^{100} \left(\widehat{\alpha}^{*i} + \sum\nolimits_{k=1}^{p} \widehat{\phi}^{*i}_k \pi_{t-k} + \sum\nolimits_{l=1}^{q} \sum\nolimits_{j=1}^{M} \widehat{\gamma}^{*i}_{jl} I(|t^{*i}_{jl}| > c) \widehat{f}_{j,t-l+1} \right)
\end{aligned}
$$

where π^h_{t+h} denotes the rate of inflation over the period t to $t+h$ and superscript i the denotes parameter estimates for the ith bootstrap replication.

The accuracy of each forecasting method is measured by the average of the squared forecast errors obtained by recursively re-estimating the model at each point in time t and forecasting π^h_{t+h}. Note that we also re-estimate the lag orders at each point in time. The evaluation period consists of 240 observations covering the most recent twenty years in the sample. Table 1 summarizes the results for the unrestricted model, the three pre-test methods and the four factor models. Table 1a shows the results for one-month ahead forecasts of U.S. CPI inflation ($h = 1$); Table 1b the corresponding results for one-year ahead forecasts ($h = 12$). The best results for each method are shown in bold face.

We compute results for each of these methods for up to three lags of the block of indicator variables in the unrestricted model. Note that adding more lags tends to result in near-singularity problems, when the estimation window is short. In some cases, even with only two lags of the 26 indicator variables there are near-singularity problems at the beginning of the recursive sample. When such problems arise, the corresponding entry in the table has been left blank. We also show results based on the SIC with an upper bound of 2 lags. For larger upper bounds, again near-singularity problems tend to arise at the beginning of the sample. In contrast, factor models are more parsimonious and hence allow for richer dynamics. We show results for models including up to five additional lags of the estimated factor. We also allow the lag order q to be selected by the SIC. The SIC generally produced more accurate forecasts than the AIC. The results are robust to the upper bound on the lag order.

Table 1a shows that somewhat surprisingly the UR model with one lag is the most accurate forecasting procedure. At $h = 1$, factor models, in contrast, outperform the benchmark at best by 3 percentage points. These results are robust to extending or shortening the evaluation period of 240 observations. One would expect that imposing the factor structure becomes more useful at longer forecast horizons. Table 1b shows the corresponding results for a horizon of twelve months ($h = 12$). In that case, the benchmark model no longer is an autoregression. At this longer horizon, factor models achieve PMSE gains of up to 33 percentage points relative to the benchmark model. Although the factor models are still outperformed by the unrestricted predictor when the number of lags of the extra predictors is fixed at one, allowing the factor models more flexibility allows it to beat the unrestricted predictor by almost 3 percentage points. Using the SIC for selecting the lag order q at each point in time does not necessarily improve the accuracy of the forecast model forecasts relative to fixed lag structures for the factor models, but it helps to keep down the PMSE of the UR and pre-test predictors that are particularly sensitive to overfitting.

Tables 1a and 1b also suggest that pre-testing usually does not improve forecasting accuracy relative to the unrestricted model. Both the PT and CPT predictors do worse than the best UR predictor. Similarly, more often than not, the PT^F strategy performs worse than including a small fixed number of principal components.[7] The poor performance of pre-test based strategies is not unexpected, given the motivation for bagging. The performance of the corresponding bagging strategies is summarized in Table 2. For the bagging methods we do not report results for more lags than one to conserve space. We note, however, that the performance of bagging rarely improves with more than one lag of the extra predictors.

An important question in implementing bagging is which critical value to use. Table 2 presents the results of a grid search for each of the three bagging methods that helps answer that question. We considered $c \epsilon \{2.575, 2.241, 1.96, 1.645, 1.440, 1.282, 0.675\}$. Table 2 shows that the performance of the bagging methods is remarkably insensitive to the choice of c over this range. Table 2 also suggests that $c = 1.96$ results in the highest accuracy for the BA predictor, whereas a somewhat higher value of $c = 2.575$ works best for the CBA and BA^F predictors. The results holds whether we focus on $h = 1$ or $h = 12$.

Compared to the methods in Table 1, all three bagging predictors are far more accurate. The bagging forecasts outperform the benchmark autoregressive model, the unrestricted model, the factor models with rank 1, 2, 3, or 4 (regardless of lag structure) and the three pre-test predictors. The gains in accuracy relative to the benchmark model are substantial. The CBA predictor is the most accurate bagging procedure at the one-month horizon with a PMSE ratio of 82% relative to the benchmark model, closely followed by the BA^F and BA predictors with 83% each. At the one-year horizon, again the CBA predictor is most accurate with 56%, followed by the BA^F predictor with 57% and the BA predictor with 58%. The strong performance of the BA^F predictor is somewhat surprising, given the relatively small cross-sectional dimension in this application.

It is also worth noting the gains in accuracy relative to the best factor model. They amount to up to 15 percentage points at $h = 1$ and up to 11 percentage points at $h = 12$. It is particularly interesting to compare the BA^F predictor that selects a subset of the first M factors and the

[7] Clearly, the performance of the pre-test predictors will depend on the choice of critical value. To conserve space we only report the results for $c = 1.96$. Qualitatively similar results are obtained for other values of c.

factor model forecasts based on a small fixed number of factors. Whereas the performance of the PT^F predictor in Table 1 was disappointing, its bagged version is a resounding success. The substantial gains in accuracy from bagging are even more surprising in that the dynamics allowed for in bagging are much more restrictive than for factor models.

3.2 Sensitivity Analysis

Given the strong performance of bagging it is important to note that we did not in any way select our predictors based on the results of previous studies. We simply focused on series in the public domain that on a priori grounds would be expected to be correlated with real economic activity. Nevertheless, it is possible that by accident we selected a set of predictors that is not representative *and* unduly favors the bagging method. To address this concern, in Table 3 we present the results of a sensitivity analysis. First, we deleted at random 25% of our predictor set and recomputed the PMSE ratio for the three bagging methods relative to the benchmark model that only includes lags of inflation. We report median results for 30 such draws including $M = 20$ predictors each. Second, we added six more series to the baseline predictor set of $M = 26$ predictors (representing a 25% increase in the predictor set) and recomputed the PMSE ratio. The additional variables are listed in the Appendix. With the exception of the ISM index and the interest rate spread all variables are expressed in percentage changes.

As Table 3 shows, our results for $h = 1$ are remarkably robust to changes in the data set. For $h = 12$, the choice of data set becomes more important. For $M = 32$, the PMSE ratio relative to the benchmark model drops to 43 percent for all three bagging methods. For $M = 20$, the PMSE ratio rises to 59-62 percent. Even the worst results in Table 3, however, are better than the best results shown in Table 1. While it is possible that bagging may not perform as well in other applications, there is no evidence that the results in Table 2 are not representative for the application considered here.

3.3 Bayesian Shrinkage Estimators

The bagging method also has similarities with shrinkage estimators such as the Stein-type estimator or the Bayesian shrinkage estimator used by Litterman (1986) in a different context. Thus, it is natural to compare the accuracy of bagging to that of the shrinkage estimator. A Bayesian approach is convenient in this context because it allows us to treat the parameters of the benchmark model differently from the parameters of the real economic indicators. Note that the use of prior distributions in this context does not reflect subjectively held beliefs, but simply is a device for controlling the degree of shrinkage. To facilitate the exposition and to preserve consistency with related studies, in the remainder of the paper we will include at most one lag of each indicator of real economic activity. The Bayesian shrinkage estimator is applied to the model:

$$\pi^h_{t+h|t} = \widehat{\alpha} + \sum_{k=1}^{p} \widehat{\phi}_k \pi_{t-k} + \sum_{j=1}^{M} \widehat{\beta}_j x_{j,t}$$

We postulate a diffuse Gaussian prior for $(\alpha, \phi_1, ..., \phi_p)$. The prior mean is based on the fitted values of a regression of inflation on lagged inflation and the intercept over the pre-sample period, as proposed by Wright (2003b). In our case, the pre-sample period includes 1947.1-1971.3. The prior variance is infinity. We use a different prior mean for each combination of h and p used in the benchmark model. For the remaining parameters we postulate a Gaussian prior with mean zero and standard deviation $\lambda\epsilon\{0.01, 0.05, 0.1,\ 0.2, 0.3, 0.4, 0.5, 1, 2, 5, 100\}$ for the standardized data. For $\lambda = \infty$, the shrinkage estimator reduces to the least-squares estimator of the unrestricted model. All prior covariances are set to zero. For further details on the implementation of this estimator see Lütkepohl (1993, ch. 5.4).

Table 4 shows selected results of the grid search over λ. We find that for $h = 1$ a moderate degree of shrinkage helps reduce the PMSE. The optimal degree of shrinkage is near $\lambda = 0.5$; as λ declines further, the PMSE ratio quickly starts deteriorating. The best shrinkage estimator is slightly more accurate than the bagging estimator at the one-month horizon with a ratio of 81 percent compared with 82-83 percent for bagging, depending on the method chosen. In contrast, at the one-year horizon, the unrestricted model with a ratio of 70 percent is more accurate than any shrinkage estimator, and bagging is even more accurate than with a ratio of 56-58 percent. We conclude that bagging in this application performs almost as well or better than Bayesian shrinkage estimators, depending on the horizon.

3.4 Bayesian Model Averaging: One Extra Predictor at a Time

Recently, there has been mounting evidence that forecast combination methods are a promising approach to improving forecast accuracy. For example, Stock and Watson (2003) have shown that simple methods of forecast combination such as using the median forecast from a large set of models may effectively reduce the instability of inflation forecasts and lower their prediction mean-squared errors. In its simplest form, forecast combination methods assign equal weight to all possible combinations of the benchmark model and one extra predictor at a time. More recently, Wright (2003b) has shown that the accuracy of forecast combination methods may be improved upon further by weighting the individual forecast models based on the posterior probabilities associated with each forecast model. in this subsection, we will expand the list of competitors of bagging to include Wright's BMA method. A key difference between our papers is that Wright imposes one lag of inflation only in the benchmark model, whereas we allow for potentially more than one lag of inflation. Otherwise our approaches are identical.

As before, for the benchmark model we follow Wright (2003b) in postulating a diffuse Gaussian prior with the prior mean based on the fitted values of a regression of inflation on lagged inflation and the intercept over the pre-sample period. For the remaining parameters we postulate a Gaussian prior with mean zero and a prior standard deviation of $\phi\epsilon\{0, 0.01, 0.05, 0.1, 0.2, 0.3, 0.4, 0.5, 1, 2, 5, 100\}$ for the standardized data. Again the prior treats the predictors as independent. The prior probability for each forecast model is $1/M$, as in the equal-weighted forecast combination. For $\phi = 0$, the BMA method of forecast combination reduces to the equal-weighted method. Table 5a presents selected results of the grid search over ϕ.

We find that, as in Wright (2003b), the BMA method is clearly superior to the equal-weighted forecast combination method. Table 5a also shows the PMSE ratio of the median forecast. This alternative combination forecast was inferior to both the BMA forecast and the equal-weighted forecast. The best results for the BMA method at the one-month horizon are achieved with $\phi = 0.1$. At the one-year horizon an even tighter prior of $\phi = 0.05$ works best. These results are of course problem-specific. For example, for Wright's (2003b) quarterly data set much larger prior standard deviations appear to work best.

With a ratio of 90 percent the BMA method in our application is more accurate than the factor model forecast at the one-month horizon, but somewhat less accurate than the three bagging forecasts. At the one-year horizon, the best BMA forecast with a ratio of 84 percent is inferior to the factor model forecast and much less accurate than the bagging forecasts. We conclude that in this application bagging clearly outperforms the BMA method.

3.5 Bayesian Model Averaging: Randomly Chosen Subsets of Extra Predictors

Papers on forecast combination methods for inflation typically restrict the forecast models under consideration to include only one indicator of real economic activity at a time. There is no reason for this approach to be optimal, whether we use equal weights or posterior probability weights. In fact, a complete Bayesian solution to this problem that provides optimal predictive ability would involve averaging over all possible forecast model combinations (see Madigan and Raftery 1994). The problem is that such a systematic comparison of all possible subsets of such indicators would be computationally prohibitive in realistic situations. In our example, there are $2^{26} = 67,108,864$ possible combinations of predictors to be considered. In response to this problem, Raftery, Madigan and Hoeting (1997) proposed an alternative method of BMA for linear regression models based on a randomly selected subsets of predictors that approximates the Bayesian solution to searching over all models.[8] The random selection is based on a Markov Chain Monte Carlo (MCMC) algorithm that moves through the forecast model space. Unlike Wright's method, this algorithm involves simulation of the posterior distribution and is quite computationally intensive. Our results are based on 5000 draws from the posterior distribution at each point in time.

MATLAB code for the Raftery et al. algorithm is publicly available at http://www.spatial-econometrics.com. We modified the Raftery et al. approach to ensure that the benchmark model including only lags of inflation and the intercept is retained in each random selection. For the models of the benchmark model we use a diffuse Gaussian prior identical to the priors used for the Wright (2003b) method. For the remaining parameters of the forecast prior the algorithm involves a Gaussian prior with mean zero and hyperparameters $\nu = 2.58$, $\lambda = 0.28$, and $\phi \epsilon \{0, 0.01, 0.05, 0.1, 0.2, 0.3, 0.4, 0.5, 1, 2, 5, 100\}$, where ϕ measures the prior standard deviation of the standardized predictor data (see Raftery et al. for further details). We report a subset of the empirical results in Table 5b. We also experimented with $\phi = 2.85$, the value recommended by Raftery et al. for a generic linear model, but the results were clearly worse than for our preferred value of ϕ below.

We find that a value of about $\phi = 0.01$ works best for $h = 1$ and $\phi = 0$ for $h = 12$. This version of BMA produces clearly more accurate results than the restricted version involving only one extra predictor at a time. Compared to Table 5a, at the one-month horizon the PMSE ratio for the best BMA predictor falls from 90 percent to 80 percent and at the one-year horizon from

[8] See Sala-i-Martin, Doppelhofer and Miller (2004) for a similar approach to BMA in a different context. Also see George and McCulloch (1993) for an alternative stochastic search variable selection algorithm.

84 percent to 62 percent. Thus, for $h = 1$, this BMA method is somewhat more accurate than the best bagging predictor; for $h = 12$, however, the best bagging predictor promises somewhat higher accuracy with a ratio of 56 percent.

4 Conclusion

Recently, there has been increased interest in forecasting methods that allow the user to extract the relevant information from a large set of potentially relevant predictors. One such method is bootstrap aggregation of forecasts (or *bagging* for short). Bagging is intended to reduce the out-of-sample prediction mean-squared error of forecast models selected by unstable decision rules such as pre-tests. This article explored the usefulness of bagging methods in forecasting economic time series. We first described how to implement the bagging idea in the context of multiple regression models with possibly serially correlated and heteroskedastic errors. We discussed three different algorithms for implementing the bagging idea: two based on the correlated regressor model and one based on the factor model. Using asymptotic theory and empirical evidence we showed that bagging, while no panacea, is a promising alternative to existing forecasting methods in many cases.

Whether bagging is likely to improve out-of-sample forecast accuracy in a given application may be assessed based on a simulated out-of-sample forecast exercise. For illustrative purposes, we considered the widely studied question of whether the inclusion of indicators of real economic activity lowers the prediction mean-squared error of forecast models of U.S. CPI inflation. Over a twenty-year period, we compared the accuracy of simulated out-of-sample forecasts based on the bagging method to that of alternative forecast methods for U.S. inflation, including forecasts from a benchmark model that includes only lags of inflation, forecasts from the unrestricted model that includes all potentially relevant predictors, forecasts from models with a subset of these predictors selected by pre-tests, forecasts from estimated factor models, forecasts from models estimated by shrinkage estimators, standard combination forecasts and finally forecasts obtained by state-of-the-art methods of Bayesian model averaging.

We found that all three bagging methods under consideration greatly reduce the prediction mean squared error of forecasts of U.S. CPI inflation at horizons of one month and one year relative to the unrestricted, fully restricted and pre-test model forecasts. Bagging forecasts in this application also were more accurate than forecasts from estimated factor models. Particularly striking were the improvements from applying bagging to a larger subset of the estimated factors relative to forecasts from a small fixed number of factors. We showed that the superior performance of bagging methods in this application is robust to alterations of the data set and we addressed the important practical question of how to choose an appropriate critical value for the bagging method.

We also compared bagging methods to other methods of forecast combination. Bagging performed better than equal-weighted or median forecasts. In addition, in this application, bagging performed better than the method of Bayesian model averaging recently proposed by Wright (2003b), and - depending on the horizon - almost as well as or somewhat better than forecasts based on Bayesian shrinkage estimators or on the method of Bayesian model averaging proposed by Raftery et al. (1997).

Our analysis demonstrated that significant improvements in forecasting accuracy can be obtained over existing methods, and we illustrated how researchers can determine whether such gains are likely in a given application. We provided an empirical example in which bagging achieved substantial gains in forecasting accuracy. Whether bagging will perform equally well in other applications is an open question that calls for more research. For example, our asymptotic analysis of a stylized regression model suggests that bagging may not work well when the regressors are highly correlated. Our asymptotic analysis also suggested that, regardless of the bagging method adopted, bagging is unlikely to work as well when the degree of predictability is very low, as would be the case in forecasting asset returns, for example. More research is needed before bagging can be considered a standard tool for applied forecasters using multiple linear regression models.

We also note that the analysis of bagging presented in this paper assumes a covariance stationary environment and abstracts from the possibility of structural change. The same is true of the standard theory of forecast combination, which relies on information pooling in a stationary environment. An interesting avenue for future research would be the development of bagging methods that allow for smooth structural change. Another interesting avenue for future research will be to compare the properties of the bagging predictor in the factor model to factor model forecasts based on a small fixed number of factors, when the cross-sectional dimension is relatively large.

Appendix

Proof of Propositions 1, 2 and 3. It follows from applications of the law of large numbers and the central limit theorem that

$$
\begin{aligned}
T^{1/2}\hat{y}^{UR}(x) &\xrightarrow{d} \xi' x,\\
T^{1/2}\hat{y}^{PT}(x) &\xrightarrow{d} \xi' E(x_t x_t')S'[SE(x_t x_t')S']^{-1}SxI(|\xi_j| > c\sqrt{\sigma^2[(E(x_t x_t'))^{-1}]_{jj}} \text{ for some } j),\\
T^{1/2}\hat{y}^{CPT}(\tilde{x}) &\xrightarrow{d} \tilde{\xi}'\tilde{S}'\tilde{S}\tilde{x}I(\exists\, j \text{ s.t. } |\tilde{\xi}_j| > c\sigma),\\
T^{1/2}\hat{y}^{BA}(x) &\xrightarrow{d} E\{\xi^{*\prime} E(x_t x_t')S^{*'}[S^* E(x_t x_t')S^{*'}]^{-1}S^* xI(|\xi_j^*| > c\sqrt{\sigma^2[(E(x_t x_t'))^{-1}]_{jj}} \text{ for some } j)|\xi\},\\
T^{1/2}\hat{y}^{CBA}(\tilde{x}) &\xrightarrow{d} E\{\tilde{\xi}^{*\prime}\tilde{S}^{*\prime}\tilde{S}^*\tilde{x}I(|\tilde{\xi}_j^*| > c\sigma \text{ for some } j)|\xi\},
\end{aligned}
$$

and

$$
\begin{aligned}
T^{1/2}\hat{y}^{UR}(x) &\xrightarrow{d} \xi' x,\\
T^{1/2}\hat{y}^{PT}(x) &\xrightarrow{d} \xi' S' SxI(|\xi_j| > c\sigma \text{ for some } j),\\
T^{1/2}\hat{y}^{BA}(x) &\xrightarrow{d} E\{\xi^{*\prime} S^{*'} S^* xI(|\xi_j^*| > c\sigma \text{ for some } j)|\xi\}.
\end{aligned}
$$

Thus, Propositions 1 and 2 follow. Proposition 3 can be proved analogously to Proposition 2 with suitable changes in notation.

Data Sources

All data are for the United States. The sample period for the raw data is 1971.4-2003.7. This choice is dictated by data constraints. The data are from the Federal Reserve Board and the database of the Federal Reserve Bank of St. Louis (FRED). They are available at http://www.economagic.com:

$INDPRO$	industrial production
$HOUST$	housing starts
$HSN1F$	house sales
$NAPM$	purchasing managers index
$HELPWANT$	help wanted index
TCU	capacity utilization
$UNRATE$	unemployment rate
$PAYEMS$	nonfarm payroll employment
$CIVPART$	civilian participation rate
$AWHI$	aggregate weekly hours, private nonfarm payrolls
$MORTG$	mortgage rate
$MPRIME$	prime rate
$CD1M$	1-month CD rate
$FEDFUND$	Federal funds rate
$M1SL$	M1
$M2SL$	M2
$M3SL$	M3
$BUSLOANS$	business loans
$CONSUMER$	consumer loans
$REALN$	real estate loans
$EXGEUS$	DM/USD rate (extrapolated using the Euro/USD rate)
$EXJPUS$	Yen/USD rate
$EXCAUS$	Canadian Dollar/USD rate
$EXUSUK$	USD/British Pound rate
$OILPRICE$	WTI crude oil spot price
$TRSP500$	SP500 stock returns

The additional data used in the sensitivity analysis are:

$NAPM$	ISM index of manufacturing activity
$TOTASS\_AUSA$	total number of motor vehicle assemblies
$TCM20Y - TBSM3M$	spread of 10-year T-bond rate over 3-month T-bill rate
$UEMP15OV$	number of civilians unemployed for more than 15 weeks
$UEMPLT5$	number of civilians unemployed for less than 5 weeks
$AWHNONAG$	average weekly hours, private nonagricultural establishments

References

1. Avramov, D. (2002), "Stock Return Predictability and Model Uncertainty," *Journal of Financial Economics*, 64, 423-458.
2. Bai, J., and S. Ng (2004), "Confidence Intervals for Diffusion Index Forecasts with a Large Number of Predictors" mimeo, Department of Economics, University of Michigan.
3. Bates, J.M., and C.W.J. Granger (1969), "The Combination of Forecasts," *Operations Research Quarterly*, 20, 451-468.
4. Bernanke, B.S., and J. Boivin (2003), "Monetary Policy in a Data-Rich Environment," *Journal of Monetary Economics*, 50, 525-546.
5. Breiman, L. (1996), "Bagging Predictors," *Machine Learning*, 36, 105-139.
6. Bühlmann, P. and B. Yu (2002), "Analyzing Bagging," *Annals of Statistics*, 30, 927-961.
7. Cecchetti, S., R. Chu, and C. Steindel (2000), "The Unreliability of Inflation Indicators," *Federal Reserve Bank of New York Current Issues in Economics and Finance*, 6, 1-6.
8. Cremers, K.J.M. (2002), "Stock Return Predictability: A Bayesian Model Selection Perspective," *Review of Financial Studies*, 15, 1223-1249.
9. Forni, M., M. Hallin, M. Lippi, and L. Reichlin (2000), "The Generalized Factor Model: Identification and Estimation," *Review of Economics and Statistics*, 82, 540-554.
10. Forni, M., M. Hallin, M. Lippi, and L. Reichlin (2001), "The Generalized Factor Model: One-Sided Estimation and Forecasting," mimeo, ECARES, Free University of Brussels.
11. Forni, M., M. Hallin, M. Lippi, and L. Reichlin (2003), "Do Financial Variables Help Forecasting Inflation and Real Activity in the Euro Area," *Journal of Monetary Economics*, 50, 1243-1255.
12. George, E.I., and R.E.McCulloch (1993), "Variable Selection via Gibbs Sampling," *Journal of the American Statistical Association*, 88, 881-890.
13. Gonçalves, S. and L. Kilian (2004), "Bootstrapping Autoregressions with Conditional Heteroskedasticity of Unknown Form," *Journal of Econometrics*, 123, 89-120.
14. Gonçalves, S. and H. White (2004), "Maximum Likelihood and the Bootstrap for Nonlinear Dynamic Models," *Journal of Econometrics*, 119, 199-220.
15. Hall, P. and J.L. Horowitz (1996), "Bootstrap critical values for tests based on generalized method of moments estimators," *Econometrica,* 64, 891–916.
16. Inoue, A., and L. Kilian (2004), "On the Selection of Forecasting Models," forthcoming: *Journal of Econometrics*.
17. Inoue, A. and M. Shintani (2003), "Bootstrapping GMM Estimators for Time Series," forthcoming: *Journal of Econometrics*.
18. Koop, G., and S. Potter (2003), "Forecasting in Large Macroeconomic Panels Using Bayesian Model Averaging," *Federal Reserve Bank of New York Staff Report*, 163.
19. Lee, T.-H., and Y. Yang (2004), "Bagging Binary and Quantile Predictors for Time Series," mimeo, Department of Economics, UC Riverside.

20. Litterman, R.B. (1986), "Forecasting with Bayesian Vector Autoregressions - Five Years of Experience," *Journal of Business and Economic Statistics*, 4, 25-38.
21. Lütkepohl, H. (1993), *Introduction to Multiple Time Series Analysis*, Springer-Verlag: Berlin.
22. Madigan, D., and A.E. Raftery (1994), "Model Selection and Accounting for Model Uncertainty in Graphical Models Using Occam's Window," *Journal of the American Statistical Association*, 89, 1535-1546.
23. Marcellino, M., J.H. Stock and M.W. Watson (2003), "Macroeconomic Forecasting in the Euro Area: Country-Specific versus Area-Wide Information," *European Economic Review*, 47, 1-18.
24. Newey, W., and K. West (1987), "A Simple Positive Semi-Definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix," *Econometrica*, 55, 703-708.
25. Politis, D.N., J.P. Romano and M. Wolf (1999), *Subsampling*, Springer-Verlag: New York.
26. Raftery, A.E., D. Madigan, and J.A. Hoeting (1997), "Bayesian Model Averaging for Linear Regression Models," *Journal of the American Statistical Association*, 92, 179-191.
27. Sala-i-Martin, G. Doppelhoffer, and R.I. Miller (2004), "Determinants of Long-Term Growth: A Bayesian Averaging of Classical Estimates (BACE) Approach," *American Economic Review*, 94, 813-835.
28. Stock, J.H., and M.W. Watson (1999), "Forecasting Inflation," *Journal of Monetary Economics*, 44, 293-335.
29. Stock, J.H., and M.W. Watson (2002a), "Forecasting Using Principal Components from a Large Number of Predictors," *Journal of the American Statistical Association*, 97, 1167-1179.
30. Stock, J.H., and M.W. Watson (2002b), "Macroeconomic Forecasting Using Diffusion Indexes," *Journal of Business and Economic Statistics*, 20, 147-162.
31. Stock, J.H., and M.W. Watson (2003), "Forecasting Output and Inflation: The Role of Asset Prices ," *Journal of Economic Literature*, 41, 788-829.
32. Thomson, M., and P. Schmidt (1982), "A Note on the Comparison of the Mean Square Error of Inequality Constrained Least-Squares and Other Related Estimators ," *Review of Economics and Statistics*, 64, 174-176.
33. West, K. (1997), "Another Heteroskedasticity and Autocorrelation Consistent Covariance Matrix Estimator," *Journal of Econometrics*, 76, 171-191
34. White, H. (1980), "A Heteroskedasticity-Consistent Covariance Matrix Estimator and a Direct Test of Heterogeneity," *Econometrica*, 48, 817-838.
35. Wright, J.H. (2003a), "Bayesian Model Averaging and Exchange Rate Forecasts," *International Finance Discussion Papers*, No. 779, Board of Governors of the Federal Reserve System.
36. Wright, J.H. (2003b), "Forecasting U.S. Inflation by Bayesian Model Averaging," *International Finance Discussion Papers*, No. 780, Board of Governors of the Federal Reserve System.

Table 1a. Out-of-Sample Forecast Accuracy:
U.S. Inflation Forecasts: 1 Month Ahead
Evaluation Period: 1983.8-2003.7

	Models with Indicators of Economic Activity							
	PMSE Relative to Benchmark at h=1							
Lags of					**FM**			
Indicators	**UR**	**PT**	**CPT**	**PT^F**	**rank 1**	**rank 2**	**rank 3**	**rank 4**
1	**0.885**	**0.899**	**0.896**	**0.937**	0.985	0.991	**1.036**	**0.978**
2	1.168	0.925	0.993	1.104	**0.969**	**0.983**	1.049	1.021
3	1.668	1.017	-	1.086	0.984	1.000	1.055	1.049
4	-	-	-	-	0.990	1.013	1.094	1.089
5	-	-	-	-	0.993	1.019	1.123	1.142
6	-	-	-	-	0.998	1.012	1.168	1.185
SIC	0.885	0.899	0.896	0.937	0.984	1.014	1.135	1.066

Table 1b. Out-of-Sample Forecast Accuracy:
U.S. Inflation Forecasts: 12 Months Ahead
Evaluation Period: 1983.8-2003.7

	Models with Indicators of Economic Activity							
	PMSE Relative to Benchmark at h=12							
Lags of					**FM**			
Indicators	**UR**	**PT**	**CPT**	**PT^F**	**rank 1**	**rank 2**	**rank 3**	**rank 4**
1	**0.695**	1.190	**0.860**	**0.820**	0.720	0.739	0.785	0.731
2	0.838	**1.046**	-	1.045	0.674	0.691	0.746	**0.704**
3	1.207	1.061	-	0.997	**0.668**	**0.685**	0.755	0.743
4	-	-	-	-	0.673	0.687	0.774	0.790
5	-	-	-	-	0.686	0.703	0.784	0.829
6	-	-	-	-	0.708	0.732	0.803	0.884
SIC	0.695	1.190	0.867	0.907	0.776	0.700	**0.738**	0.830

SOURCE: The sample period of the raw data is 1971.4-2003.7. The PMSE is based on the average of the squared recursive forecasts errors. The pre-test forecasts are all based on $c = 1.96$. The pre-test results for other values of c are qualitatively similar.

Table 2. Out-of-Sample Forecast Accuracy:
U.S. Inflation Forecasts: 1 Month and 12 Months Ahead
Evaluation Period: 1983.8-2003.7

		Alternative Bagging Predictors with Critical Value c						
		PMSE Relative to Benchmark						
		$c = 2.575$	$c = 2.241$	$c = 1.96$	$c = 1.645$	$c = 1.440$	$c = 1.282$	$c = 0.675$
$h = 1$	BA	0.835	0.834	**0.833**	0.839	0.844	0.854	0.860
	CBA	**0.817**	0.828	0.835	0.841	0.845	0.847	0.857
	BA^F	**0.828**	0.832	0.841	0.847	0.851	0.854	0.861
$h = 12$	BA	0.617	0.597	**0.582**	0.587	0.590	0.591	0.606
	CBA	**0.555**	0.557	0.562	0.574	0.587	0.594	0.610
	BA^F	**0.567**	0.578	0.588	0.600	0.605	0.608	0.614

SOURCE: See Table 1. All results based on one lag of the extra predictors only. For $h = 1$, all pre-tests are based on White (1980) robust standard errors. The bagging results are based on the pairwise bootstrap. For $h = 12$, all pre-tests are based on West (1997) robust standard errors. The bagging results are based on blocks of length $m = 12$.

Table 3. Out-of-Sample Forecast Accuracy:
U.S. Inflation Forecasts: 1 Month and 12 Months Ahead
Evaluation Period: 1983.8-2003.7

		Sensitivity of Performance of Bagging Predictors to M		
		PMSE Relative to Benchmark		
		-25%*	Baseline Data Set	+25%
		$M = 20$	$M = 26$	$M = 32$
$h = 1$	BA	0.812	0.833	0.826
	CBA	0.803	0.817	0.836
	BA^F	0.821	0.828	0.847
$h = 12$	BA	0.621	0.582	0.430
	CBA	0.592	0.555	0.432
	BA^F	0.601	0.567	0.426

SOURCE: See Table 2. All results based on optimal value of c in Table 2. * Median result based on 30 random draws of 20 predictors from baseline predictor set.

Table 4. Out-of-Sample Forecast Accuracy:
U.S. Inflation Forecasts: 1 Month and 12 Months Ahead
Evaluation Period: 1983.8-2003.7

	Shrinkage Estimator of Unrestricted Model						
	PMSE Relative to Benchmark						
	Bayesian shrinkage estimator					**UR**	**CBA**
	$\lambda = 0.5$	$\lambda = 1$	$\lambda = 2$	$\lambda = 5$	$\lambda = 100$	$\lambda = \infty$	$c = 2.575$
$h = 1$	**0.809**	0.826	0.843	0.865	0.885	0.885	0.817
$h = 12$	0.710	0.703	0.696	0.695	0.695	**0.695**	**0.555**

SOURCE: See Table 1.

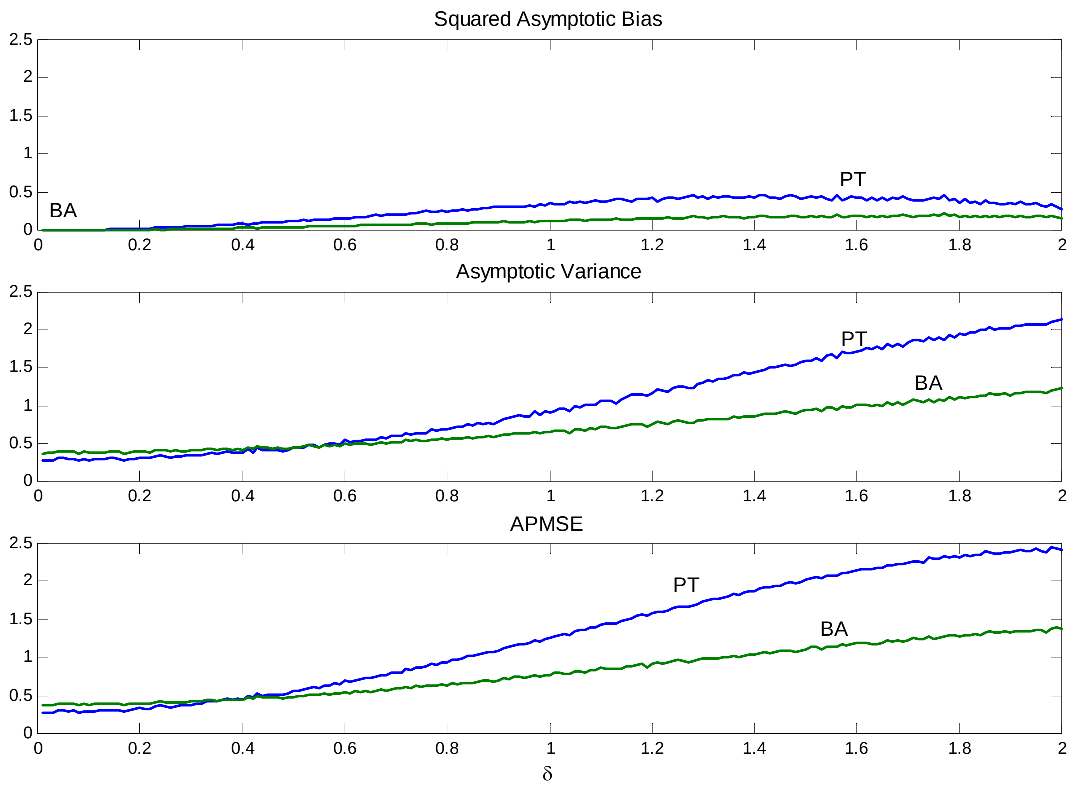
Table 5. Out-of-Sample Forecast Accuracy:
U.S. Inflation Forecasts: 1 Month and 12 Months Ahead
Evaluation Period: 1983.8-2003.7

(a)	**Bayesian Model Averaging: One Extra Predictor at a Time**							
	PMSE Relative to Benchmark							
	Median	**Equal-weighted**	**BMA**					**CBA**
		$\phi = 0$	$\phi = 0.01$	$\phi = 0.05$	$\phi = 0.1$	$\phi = 0.2$	$\phi = 0.3$	$c = 2.575$
$h = 1$	0.993	0.974	0.970	0.910	**0.904**	0.915	0.919	**0.817**
$h = 12$	0.947	0.871	0.852	**0.843**	0.885	0.948	0.958	**0.555**

(b)	**Bayesian Model Averaging: Random Sets of Extra Predictors**								
	PMSE Relative to Benchmark								
	Equal-weighted	**BMA**							**CBA**
	$\phi = 0$	$\phi = 0.01$	$\phi = 0.05$	$\phi = 0.1$	$\phi = 0.2$	$\phi = 0.5$	$\phi = 1$	$\phi = 2$	$c = 2.575$
$h = 1$	0.820	**0.804**	0.817	0.819	0.827	0.828	0.832	0.839	**0.817**
$h = 12$	**0.622**	0.643	0.676	0.681	0.666	0.656	0.645	0.646	**0.555**

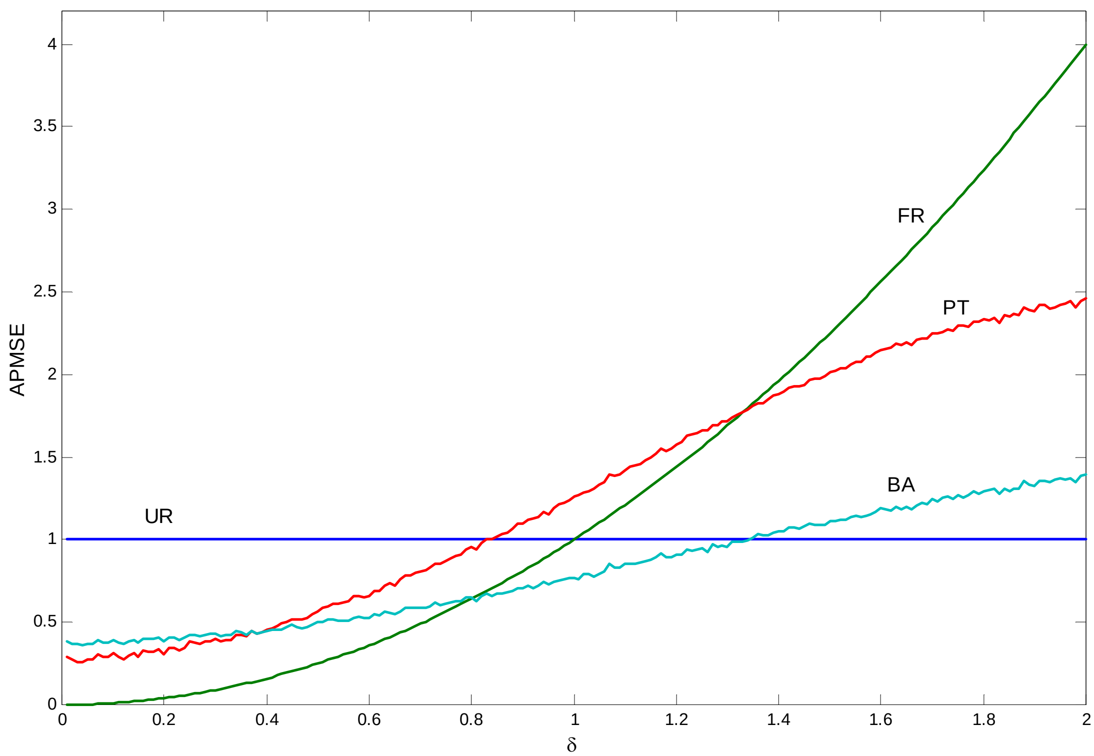
SOURCE: See Table 1.

Figure 1: Asymptotic Properties of PT and BA Predictors in Single-Regressor Model



NOTES: PT=Pre-test predictor. BA=Bagging predictor.

Figure 2: APMSE of Alternative Predictors in Single-Regressor Model



NOTES: PT=Pre-test predictor. BA=Bagging predictor. UR=Unrestricted predictor. FR=Fully restricted predictor

Figure 3: APMSE Gains of the BA Predictor Relative to the PT Predictor in Correlated Regressor Model

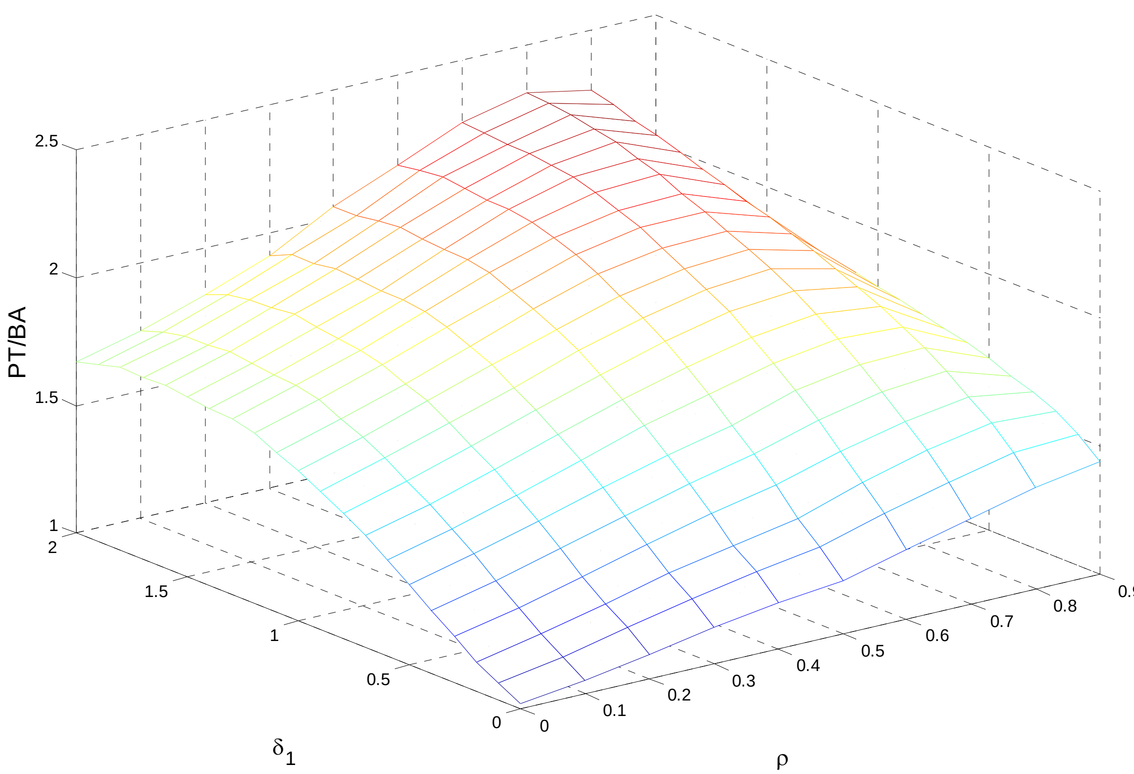


Figure 4: APMSE Gains of the BA Predictor Relative to Best Alternative Predictor in Correlated Regressor Model

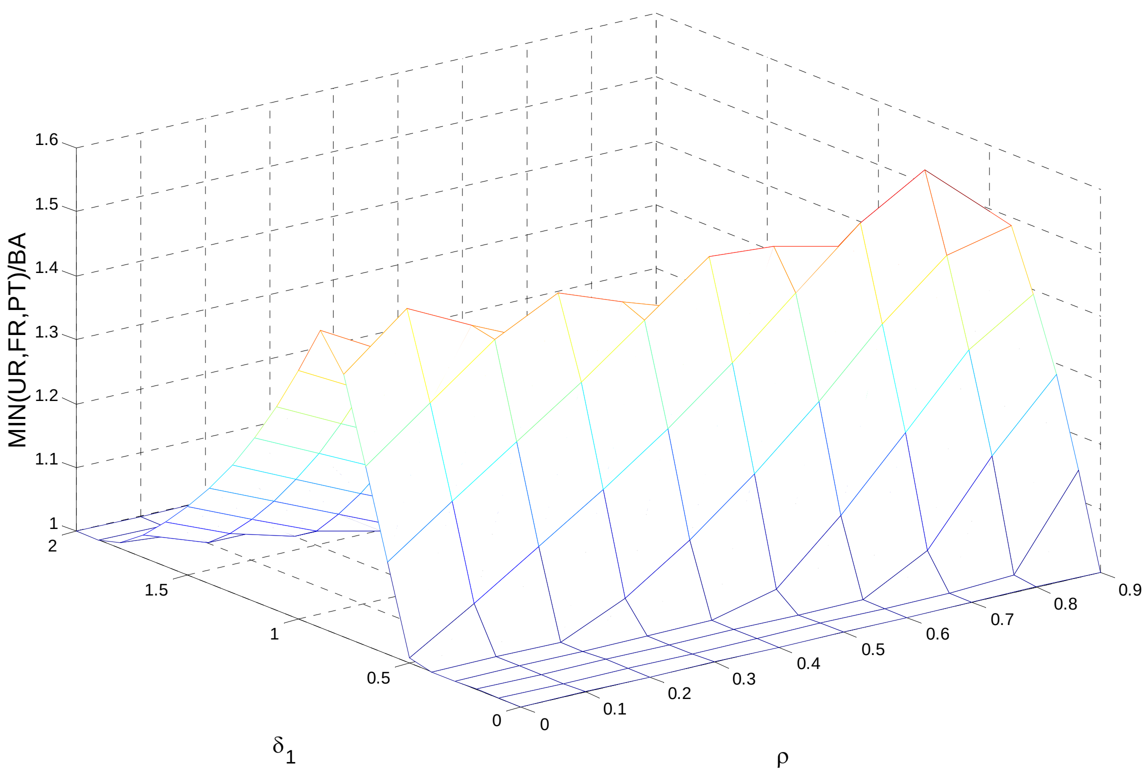


Figure 5: APMSE Gains of CBA Predictor Relative to Best Alternative Predictor in Correlated Regressor Model