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

Atsushi Inoue[†]
University of British Columbia/North Carolina State University

Lutz Kilian[‡]
University of Michigan

January 4, 2007

Abstract

This paper focuses on the widely studied question of whether the inclusion of indicators of real economic activity lowers the prediction mean-squared error of forecasting models of U.S. consumer price inflation. We propose three variants of the bagging algorithm specifically designed for this type of forecasting problem and evaluate their empirical performance. While bagging predictors in our application are clearly more accurate than equal-weighted forecasts, median forecasts, ARM forecasts, AFTER forecasts, or Bayesian forecast averages based on one extra predictor at a time, they are generally about as accurate as the Bayesian shrinkage estimator, the ridge regression predictor, the iterated LASSO predictor, or the Bayesian model average predictor based on random subsets of extra predictors. Our results show that bagging can achieve large reductions in prediction mean-squared errors even in challenging applications such as inflation forecasting. However, bagging is not the only method capable of achieving such gains.

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

---

*This paper has benefited from discussions with numerous people, most notably Tae-Hwy Lee and Mark Watson, and two anonymous referees. We thank seminar participants at Caltech, the CFS, the ECB, Johns Hopkins, Kobe, Michigan State, Maryland, Ohio State, Purdue, Rice, SMU, Texas A&M, Tokyo, UC Davis, UC Riverside, and UC Santa Barbara. The paper has also been presented at the 2003 Triangle Econometrics Conference, the 2004 Financial Econometrics Conference in Waterloo, the 2004 Forecasting Conference at Duke, the 2004 Econometric Society Meeting at Brown, the 2004 NBER-NSF Time Series Conference at SMU, the 2004 MEG Meeting at Northwestern, the 2004 CEPR Conference on Forecast Combinations in Brussels, the 2005 Conference on Inference in Finance in Québec City, the 2005 Conference on Forecasting in Montreal, and the 2006 Econometric Society Meeting in Boston.

†Department of Economics, University of British Columbia, 997-1873 East Mall, Vancouver, BC V6T 1Z1, Canada. E-mail: ainoue@interchange.ubc.ca

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

1 Introduction

A common situation in out-of-sample prediction is that there are many potentially useful predictors available to the forecaster, but few (if any) of these predictors have high predictive power and many of the potential predictors are correlated. This situation is particularly relevant for economic forecasting, because economic theory rarely puts tight restrictions on the set of potential predictors and usually many alternative proxies for the same economic variable are available to the forecaster. A case in point is the widely studied problem of forecasting consumer price inflation based on measures of real economic activity. There are many 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, all of which are believed to have some predictive power for consumer price inflation. It is well known that forecasts generated using only one of these indicators 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 for real activity is thought to lead to overfitting and poor out-of-sample forecast accuracy. Moreover, standard methods of ranking all possible combinations of predictors by means of an information criterion function and selecting the combination that minimizes the criterion, as discussed in Inoue and Kilian (2006), become computationally impractical when there are many potential predictors. This suggests that this forecasting problem is a natural candidate for the application of *bootstrap aggregation* or *bagging* methods, as discussed in Breiman (1996) and Bühlmann and Yu (2002).

Bagging is a statistical method designed to reduce the out-of-sample prediction mean-squared error (PMSE) of forecasting models selected by unstable decision rules such as pre-tests. Bagging involves generating a large number of bootstrap resamples of the original forecasting problem, applying a pre-test model selection rule to each of the resamples, and averaging the forecasts from the models selected by the pre-test on each bootstrap sample. While bagging has been found to work well in many statistical applications, to date little if any attention seems to have been devoted to the problem of bagging dynamic linear regression models with correlated regressors. The latter model is widely used in economic applications. In this paper, we propose three variants of the bagging algorithm specifically designed for this type of forecasting model: the BA method for models with possibly correlated regressors, the CBA method for models with orthogonalized regressors and the BA^F method for factor models. We compare the forecast accuracy of these bagging methods relative to one another as well as relative to some of the leading alternatives to bagging discussed in the literature.

One of those leading alternatives 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 (see, e.g., Stock and Watson 1999). More sophisticated methods of forecast averaging include adaptive regression by mixing or ARM (see Yang 2001, 2003) and aggregation of forecasts through exponential reweighting or $AFTER$ (see Yang 2004). Alternatively, one could weight individual forecasts by the posterior probabilities of each forecasting model (see, e.g., Min and Zellner 1993, Avramov 2002, Cremers 2002, Wright 2003a and Koop and Potter 2004 for applications in econometrics). This Bayesian model averaging (BMA) approach has been used successfully in forecasting inflation by Wright (2003b). A second leading alternative 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). Shrinkage estimation also could be implemented based on the well-known ridge estimator. Yet another variation on this idea is the least-absolute shrinkage and selection operator ($LASSO$) of Tibshirani (1996) that combines features of shrinkage and model selection. A third leading alternative 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 forecasting under quite general conditions (see, e.g., Stock and Watson 2002a, 2002b). A closely related approach to extracting common components has been developed by Forni et al. (2000, 2005) and applied in Forni et al. (2003).

The remainder of the paper is organized as follows. In section 2, we show how the bagging proposal may be adapted to applications involving dynamic linear multiple regression with possibly serially correlated and heteroskedastic errors that are typical of the inflation forecasting problem. We discuss applications of bagging in the correlated regressor model as well as in factor models.

In section 3, we investigate whether adding indicators of real economic activity to models involving lagged inflation rates improves the accuracy of one-month and twelve-month ahead forecasts of U.S. consumer price inflation, as measured by reductions in the prediction mean-squared error. This empirical application 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 none of the forecasting methods considered uniformly dominates the other methods in this application. Nevertheless, the results are fairly clear-cut in that some forecasting methods perform well at both one-month and one-year horizons, whereas other methods do not.

We show that there is no clear ranking between the standard bagging method for unorthogonalized predictors (BA) and the orthogonalized bagging method (CBA). Both perform about equally well. The bagging predictor in all cases is more accurate than the unrestricted model and typically at least as accurate as the pre-test predictor. While bagging predictors are clearly more accurate than equal-weighted forecasts, median forecasts, ARM forecasts, $AFTER$ forecasts, or Bayesian forecast averages based on one extra predictor at a time, they are generally about as accurate as the Bayesian shrinkage estimator, the ridge regression predictor, the iterated $LASSO$ predictor, or the Bayesian model average predictor based on random subsets of extra predictors. At the one-month horizon, all these methods achieve gains relative to the inflation-only model of 16-18 percent. At the one-year horizon, the gains increase to 35-40 percent. ARM and $AFTER$ predictors based on random subsets of extra predictors also perform well in some cases, but their performance is more uneven across horizons.

We find that the bagging predictors based on the regression model typically are more accurate than a factor model predictor of given rank r for the same data set. While bagging the r largest principal components (BA^F) tends to improve the forecast accuracy of a factor model for conventional choices of r, often the gains are not as large as for bagging methods designed for standard regression models.

We conclude in section 4 that bagging can achieve large reductions in prediction mean-squared errors, even in challenging applications such as inflation forecasting. The gains in accuracy compare favorably with the benchmark model and with results reported in previous studies. However, bagging is not the only method capable of achieving such gains. The high accuracy of the ridge regression predictor, in particular, suggests that similar accuracy is feasible at much lower computational cost than required for bagging or for Bayesian model averaging based on random subsets of predictors. Recently proposed asymptotic approximations to bagging methods for orthogonalized predictors, however, may eliminate the need for computer simulation in bagging. It will be of interest to see how these new approaches compare to the full-fledged bagging approach employed in this paper.

2 Three Proposals for Bagging Correlated Regressors in Dynamic Regressor Models

Most of the existing literature on bagging has focused on cross-sectional settings. Only recently, interest has grown in applications of bagging to time series models. For example, Lee and Yang (2006) in related work study the properties of bagging in binary prediction problems and quantile prediction of economic time series data. Whereas that paper focused on the ability of bagging to improve forecast accuracy under asymmetric loss functions, our paper studies the ability of the bagging predictor to reduce the PMSE. More specifically, we are concerned with the usefulness of bagging methods in forecasting economic time series from dynamic linear multiple regression models with correlated regressors. Such forecasting models are routinely used by practitioners, but no attempt has been made to utilize bagging methods in this context. Our application is motivated by the common problem of assessing the incremental predictive power of a vector of possibly mutually correlated predictors, the prototypical example of which is the inflation forecasting problem to be examined in section 3. We will consider two alternative frameworks: the correlated regressor model and the factor model and discuss how to implement bagging in each case.

2.1 Correlated Regressor Model

Consider the forecasting model:

$$y_{t+h} \;=\; \alpha' w_t + \beta' x_t + \varepsilon_{t+h}, \qquad h = 1, 2, 3, ... \tag{1}$$

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 possibly correlated predictors at time period t. We presume that y_t and x_t are jointly covariance stationary processes or have been suitably transformed to achieve covariance stationarity. w_t is a vector of predetermined variables such as deterministic regressors or lagged dependent variables with coefficient vector α. Model selection is applied only to the elements of x_t. For expository purposes we will suppress $\alpha' w_t$ in the description of the bagging algorithms.

2.1.1 Bagging Unorthogonalized Predictors

We begin by establishing some notation. Suppose we are interested in predicting the scalar y_{T+h} based on x_T, where T denotes the most recent observation available to the forecaster and $x_t \in \Re^M$. 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. Then 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 are

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

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 PT model forecast is obtained as follows: 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.

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 the best choice. Nevertheless, the simple t-test strategy studied here appears to work well in many cases. Throughout the paper we will use * to denote the bootstrap draws of the data and the bootstrap equivalents of previously defined estimators. For example, if $\hat{\beta}$ denotes the OLS estimator of β in model (1) conditional on the original data set $(y_{1+h}, x'_1), ..., (y_T, x'_{T-h})$, then $\hat{\beta}^*$ denotes the corresponding OLS estimator based on the bootstrap data set $(y^*_{1+h}, x'^*_1), ..., (y^*_T, x'^*_{T-h})$.

Proposal 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

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

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

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

$\varepsilon_{t+h}^* = y_{t+h}^* - \widehat{\beta}^{*\prime} x_t^*$, *and* b *is the integer part of* $(T-h)/m$ *(see, e.g., Inoue and Shintani 2006).*

(iii) *The bagged predictor is the expectation of the bootstrap pre-test predictor across bootstrap samples, conditional on* x_T*:*

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

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

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

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, we will consider a grid of alternative values of c.

2.1.2 Bagging Orthogonalized Predictors

One seeming drawback of Proposal 1 is that, when predictors are correlated, the effective size of the t-tests on individual predictors cannot be controlled. This fact suggests an alternative approach to bagging in which the predictors are orthogonalized prior to conducting the t-tests. When the regressor matrix is of full column rank, this may be accomplished by computing 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 unique $M \times M$ upper triangular matrix such that $P'P = E(x_t x_t')$. This allows us to express the forecasting model (1) equivalently as

$$y_{t+h} = \alpha' w_t + \beta' \widetilde{x}_t + \varepsilon_{t+h}, \qquad h = 1, 2, 3, ... \tag{2}$$

where $\alpha' w_t$ will be suppressed from now on for notational convenience. The unrestricted and fully restricted predictors will be unaffected by this transformation. We now introduce the pre-test estimator for this transformed model, which will be denoted CPT:

$$\hat{y}_{T+h}^{CPT}(\widetilde{x}_T) = 0, \text{ if } |t_j| < c \ \forall j \text{ and } \hat{y}_{T+h}^{CPT}(\widetilde{x}_T) = \widehat{\gamma}' S_T \widetilde{x}_T \text{ otherwise,}$$

where the notation is analogous to the correlated regressor case. The corresponding bagging predictor is described next:

Proposal 2. [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 $\widetilde{x}_T$

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

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

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

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

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

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

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:*

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

While the size of the pre-test is easier to control in this framework, it is unclear a priori whether the CBA method will select superior forecasting models. In the context of bagging the purpose of the pre-test is to select a forecasting model with lower PMSE, not to uncover the true relationships in the data. Notwithstanding the existence of size distortions, the BA method may lower the PMSE even in the presence of correlated regressors. An interesting question to be addressed in the empirical section is whether the performance of bagging may be improved by orthogonalizing the predictors.

2.2 Factor Models

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:

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

where f_t denotes a vector of the r 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 denote that estimator by $\hat{f}_t$. By construction, the estimated principal components or factors are orthogonal. If $N, T \to \infty$, then $\hat{f}_t$ is consistent for f_t (see Theorem 1, Stock and Watson 2002a, p. 1169). As before, w_t denotes a vector of predetermined regressors that we condition on in making forecasts. For notational convenience we again will suppress $\alpha' w_t$ in the description of the bagging algorithm.

2.2.1 Bagging Factor Predictors

Although principal components analysis generates N factors, in practice, researchers have typically focused on a small number of factors in generating forecasts from factor models, while ignoring information contained in the other factors. The obvious question is whether applying bagging to a suitably chosen subset of the principal components may improve forecast accuracy. We consider an r-dimensional subset of the N principal components that includes the r largest principal components where $r < T$. It is straightforward to adapt the bagging method to this situation. As before, we start by defining the pre-test estimator. The unrestricted predictor in the factor model will be equivalent to the standard factor model forecast. We refer to this predictor as UR^F. We denote the pre-test predictor in the factor model by PT^F. By analogy to the notation used for the correlated regressor model:

$$\hat{y}^{PT^F}_{T+h}(\hat{f}_T) = 0, \text{ if } |t_j| < c \;\forall j \text{ and } \hat{y}^{PT^F}_{T+h}(\hat{f}_T) = \widehat{\gamma}' S_T \hat{f}_T \text{ otherwise.}$$

The corresponding bagging predictor is described in Proposal 3.

Proposal 3. [BA^F method] The bagging predictor in the factor model framework is defined as follows:

(*i*) *Use principal components analysis to extract the* r largest common factors from the $T \times N$ matrix X of potential predictors. Denote the date t observation of these factor estimates by the $r \times 1$ 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* $(T-h) \times (r+1)$:

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

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* $\widehat{f}_T$

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

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* $(T-h)/m$.

(iv) The bagged predictor is the expectation of the bootstrap pre-test predictor across bootstrap samples, conditional on $\widehat{f}_T$*:*

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

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:*

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

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

In this section, we return to the problem that motivated the development of the bagging methods in the previous section and investigate whether one-month and twelve-months ahead U.S. consumer price inflation forecasts may be improved upon by adding indicators of real economic activity to models involving only lagged inflation rates. This empirical application 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.

We follow the common practice of choosing between competing forecasting methods based on the ranking of their recursive PMSEs in simulated out-of-sample forecasts. The idea is to apply each forecasting method in real time, using only information available at each point in time, and to compare the resulting forecasts to the actual realizations of inflation. As time progresses, the forecaster recursively updates his information set and generates a new forecast. Thus, a given forecasting method may select different models at each point in time. The recursive PMSE of each forecasting method is obtained by averaging the sequence of squared forecast errors.

The choice of the inflation-only benchmark model is conventional (see, e.g., Stock and Watson 2003, Forni et al. 2003) as is the focus on the PMSE. The measure of inflation is based on the seasonally adjusted urban consumer price index. The lag order of the benchmark model is determined by the AIC subject to an upper bound of 12 lags. Since the lag order of the benchmark model is selected recursively in real time, it may change as we move through the sample.

The benchmark model is compared to a number of alternative forecasting strategies that exploit in addition information about indicators of real economic activity. Since there is no universally agreed upon measure of real economic activity we consider 30 potential extra predictors that a priori can be expected to be correlated with real economic activity and that are available at monthly frequency. These predictors include production data, labor market data, monetary and financial data, and external data. A complete variable list and the data sources are provided at the end of the paper. Note that measures of wage cost and productivity are not available at monthly frequency for our sample period.

We use monthly data for 1971.4-2003.7. The starting point is dictated by data constraints. We set aside the last twenty years of data as our forecast evaluation 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 Bagging Forecasts in the Correlated Regressor Model

The alternative forecasting strategies under consideration in the first round of comparisons include the benchmark model involving only an intercept, the current value of inflation and lags of monthly inflation as well as five alternative methods that include in addition at least some indicators of economic activity. The unrestricted regression model (UR) includes current values of all 30 indicators of economic activity as separate regressors in addition to current and lagged inflation. The pre-test predictors (PT, CPT) use only a subset of these additional predictors. 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 $c\epsilon\{0.3853, 0.6745, 1.2816, 1.4395, 1.6449, 1.9600, 2.2414, 2.5758, 2.8070, 3.0233, 3.2905, 3.4808, 3.8906, 4.4172, 5.3267\}$. To conserve space, only results for the value of c that produced the lowest recursive PMSE are reported.

The bagging forecasts (BA, CBA) are computed as the average of the corresponding pre-test forecasts across 100 bootstrap replications with $M = 30$. We consider the same range of critical values as in the case of the pre-test predictors and report only the best result. 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 $m = 12$. To summarize, the forecasting methods under consideration are:

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

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}. Throughout the paper, all results are normalized relative to the recursive PMSE of the benchmark model, such that a ratio below 1 indicates that the method in question is more accurate than the benchmark model.

Table 1 summarizes the results for the unrestricted model, the two pre-test methods and the two bagging methods. It shows results for both the one-month ahead forecasts of U.S. consumer price inflation ($h = 1$) and the corresponding results for one-year ahead forecasts ($h = 12$). Table 1 shows that all forecasting methods under consideration beat the lagged inflation-only benchmark model. Even the UR model constitutes a clear improvement over the benchmark model with PMSE gains of 8 percent for $h = 1$ and 30 percent for $h = 12$. Pre-testing further improves forecasting accuracy relative to the unrestricted model. The PT predictor increases the accuracy gains to 14 percent relative to the benchmark model at $h = 1$, while being about as accurate as the UR model at $h = 12$. The CPT predictor only improves accuracy by 10 percent at $h = 1$, but lowers the PMSE by 36% at $h = 12$.

Table 1 suggests that the PT predictor can in turn be improved upon by bagging. The PMSE gains from using the BA predictor are 18 percent at $h = 1$, and 35 percent at $h = 12$. The additional reduction in PMSE is about 4 percentage points at both horizons. The corresponding gains for the CBA predictor relative to the benchmark model are 16 percent at $h = 1$, and 36 percent at $h = 12$, making the CBA predictor about equally accurate or more accurate than the CPT predictor. Table 1 also suggests that neither of the two bagging predictors dominates the other.

3.2 Unrestricted, Pre-Test and Bagging Forecasts in Factor Models

The empirical results in Table 1 have been obtained under the premise that the regressor matrix is of full column rank. An alternative view is that indicators of real economic activity are well approximated by factor models. That interpretation suggests that we impose the factor structure on the forecasting model and examine the implied UR^F, PT^F, and BA^F predictors. Each of this predictors is based on the first r of the 30 principal components that can be constructed from the set of indicators of real economic activity. The principal components are constructed as in Stock and Watson (2002a, 2002b). To summarize the methods under consideration are:

$$\begin{aligned}
UR^F &: \quad \pi^h_{t+h|t} = \widehat{\nu} + \sum\nolimits_{k=1}^{p} \widehat{\phi}_k \pi_{t-k+1} + \sum\nolimits_{j=1}^{r} \widehat{\beta}_j \widehat{f}_{j,t} \\
PT^F &: \quad \pi^h_{t+h|t} = \widehat{\nu} + \sum\nolimits_{k=1}^{p} \widehat{\phi}_k \pi_{t-k+1} + \sum\nolimits_{j=1}^{r} \widehat{\gamma}_j I(|t_j| > c) \widehat{f}_{j,t} \\
BA^F &: \quad \pi^h_{t+h|t} = \frac{1}{100} \sum_{i=1}^{100} \left(\widehat{\nu}^{*i} + \sum\nolimits_{k=1}^{p} \widehat{\phi}^{*i}_k \pi_{t-k+1} + \sum\nolimits_{j=1}^{r} \widehat{\gamma}^{*i}_j I(|t^{*i}_j| > c) \widehat{f}_{j,t} \right)
\end{aligned}$$

The benchmark model is the same as in previous subsection. We do not report results for models with more lags of the estimated factors to conserve space. We found that the performance of these models rarely improves with more than one lag of the principal components.

Table 2 presents results for $r \epsilon \{1, 2, 3, 4, 5, 6, 7, 8\}$, which covers most cases of practical interest. It is rare for users of factor models to use more than three principal components in generating forecasts. The UR^F results in Table 2 suggest that factor model forecasts are reasonably accurate in this application at the one-year horizon with gains ranging from 27 percent to 36 percent relative to the benchmark model, but less accurate at the one-month horizon with gains ranging from -2 percent to 8 percent, depending on the rank r.

Table 2 shows that, nevertheless, for $h = 1$, the UR^F forecast can usually be improved upon. With the exception of $r = 1$, the BA^F predictor is always more accurate than the UR^F and PT^F predictors at the one-month horizon. For $r = 1$, the BA^F predictor is marginally less accurate than the UR^F predictor, whereas the pre-test marginally improves on the standard factor model forecast. Overall, the gains from bagging for common choices of r are fairly systematic at $h = 1$, albeit rarely larger than 3 percentage points. For $h = 12$, the results are more mixed. Although the BA^F predictor improves on the UR^F predictor for all $r \leq 8$ and the gains can be as large as 6 percentage points, in some cases the PT^F predictor is even more accurate. Nevertheless, the results in Table 2 underscore the potential of bagging to improve the accuracy of factor model forecasts.

While there is clear evidence of gains in forecast accuracy relative to the UR^F predictor at both horizons, Table 2 also shows that for commonly used values of r the accuracy of the BA^F predictor, while roughly similar to that of the BA and CBA predictors at $h = 12$, is far inferior at $h = 1$. It may seem that this result is simply an artifact of the smaller information set used in constructing the BA^F predictor and could be overturned by allowing the information set to expand. This is not the case. When all 30 principal components are included, as in the last row of Table 2, the UR^F model reduces to the UR model. In the latter case, the BA^F predictor achieves large in gains in accuracy at $h = 1$ with a PMSE reduction of 18 percent relative to the benchmark model. This gain in accuracy is similar in magnitude to that for the other bagging predictors in Table 1. However, for $h = 12$ the he BA^F predictor not only is less accurate than the UR^F predictor, but more importantly it is far less accurate than the BA and CBA predictors in Table 1. This evidence suggests that the BA^F predictor is not as robust as the BA and CBA predictors.

3.3 Forecasts based on Shrinkage Estimation of the Correlated Regressor Model

Tables 1 and 2 provide compelling evidence that bagging predictors tend to improve forecast accuracy relative to unrestricted and pre-test predictors. These are not the only relevant alternatives, however. Since the bagging method involves features reminiscent of shrinkage estimation, it is only natural to compare the accuracy of bagging predictors to that of predictors based on alternative shrinkage methods. Here we consider three versions of shrinkage estimators that have been used by practitioners: (1) Bayesian shrinkage estimation with a Gaussian prior centered on zero; (2) ridge regression; and (3) the LASSO. The results for all three methods are summarized in Table 3.

3.3.1 Bayesian Shrinkage Estimation

Bayesian shrinkage estimation has a long tradition in econometric forecasting (see, e.g., Litterman 1986). 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. The Bayesian shrinkage estimator is applied to the model:

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

We postulate a diffuse Gaussian prior for $(\nu, \phi_1, ..., \phi_p)$. 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 OLS 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). We only report results for the value of λ that generated the lowest recursive PMSE.

Table 3 shows that this Bayesian shrinkage predictor is about as accurate as the BA and CBA methods. At $h = 1$, it lowers the PMSE by 18 percent relative to the benchmark model; at $h = 12$ the gains are 37 percent.

3.3.2 Ridge Regression

The Bayesian shrinkage estimator may be interpreted as a classical ridge regression estimator. For comparison, we also include the standard ridge regression estimator defined by

$$(\widehat{\nu}^R, \widehat{\phi}^R, \widehat{\beta}^R) = \arg\min \left\{ \sum_{t=1}^{T} \left(\pi^h_{t+h|t} - \nu - \sum_{k=1}^{p} \phi_k \pi_{t-k+1} + \sum_{j=1}^{M} \beta_j x_{j,t} \right)^2 + \lambda \sum_{j=1}^{M} \beta_j^2 \right\}$$

for $0 < \lambda < \infty$. This problem can be solved by standard numerical optimization methods. Since the estimator depends on λ, we considered a grid of values $\lambda \epsilon \{0.5, 1, 2, 3, 4, 5, 10, 20, 50, 100, 150, 200\}$. For $\lambda = 0$ the constraint is not binding and the ridge regression estimator reduces to OLS. Given this estimate, the ridge regression forecasts are generated from:

$$\pi^h_{t+h|t} = \widehat{\nu}^R + \sum_{k=1}^{p} \widehat{\phi}^R_k \pi_{t-k+1} + \sum_{j=1}^{M} \widehat{\beta}^R_j x_{j,t}$$

Only the result with the lowest recursive PMSE is reported in Table 3. Table 3 shows that the ridge regression predictor is marginally more accurate than the bagging predictor and the Bayesian shrinkage predictor. It achieves PMSE reductions of 19 percent relative to the benchmark model at $h = 1$ and 38 percent at $h = 12$.

3.3.3 LASSO

One of the potential disadvantages of ridge regression is that parameter estimates may be shrunk towards zero, but never will be zero exactly. This feature is an artifact of the imposition of a quadratic loss function in ridge estimation and may undermine forecast accuracy. To allow some parameters to be exactly zero requires the use of an absolute loss function, as in the least absolute shrinkage and selection operator (LASSO) of Tibshirani (1996). The LASSO is designed to shrink some regression coefficients towards zero, while setting others equal to zero. Thus it combines features of model selection with features of shrinkage estimation. The LASSO estimator is defined by:

$$(\widehat{\nu}^L, \widehat{\phi}^L, \widehat{\beta}^L) = \arg\min \left\{ \sum\nolimits_{t=1}^{T} \left(\pi_{t+h|t}^{h} - \nu - \sum\nolimits_{k=1}^{p} \phi_k \pi_{t-k+1} + \sum\nolimits_{j=1}^{M} \beta_j x_{j,t} \right)^2 + \lambda \sum\nolimits_{j=1}^{M} |\beta_j| \right\}$$

or equivalently

$$(\widehat{\nu}^L, \widehat{\phi}^L, \widehat{\beta}^L) = \arg\min \left\{ \sum\nolimits_{t=1}^{T} \left(\pi_{t+h|t}^{h} - \nu - \sum\nolimits_{k=1}^{p} \phi_k \pi_{t-k+1} + \sum\nolimits_{j=1}^{M} \beta_j x_{j,t} \right)^2 \right\} \quad \text{subject to } \sum\nolimits_{j=1}^{M} |\beta_j| \leq \tau.$$

In practice, we solve the latter problem by an iterative algorithm. First, we compute the LASSO estimator of β conditional on the OLS estimates of ν and ϕ in the UR model using the algorithm for models without predetermined regressors described in Tibshirani (1996, p. 278). Then we solve for the estimates of ν and ϕ conditional on the first-round LASSO estimate for β, and re-compute $\widehat{\beta}^L$ given the revised estimates of ν and ϕ. This process is iterated until convergence. Given the final estimate, the LASSO forecasts are generated from:

$$\pi_{t+h|t}^{h} = \widehat{\nu}^L + \sum\nolimits_{k=1}^{p} \widehat{\phi}_k^L \pi_{t-k+1} + \sum\nolimits_{j=1}^{M} \widehat{\beta}_j^L x_{j,t}$$

Since the estimator depends on τ, we considered a grid of values $\tau \epsilon \{1, 2, 3, 4, 5, 10, 20, 50, 100\}$. For sufficiently large τ the constraint is not binding and the LASSO estimator reduces to OLS. Only the result with the lowest recursive PMSE is reported in Table 3.

Table 3 shows that the iterated LASSO predictor is marginally less accurate than the BA and CBA methods at $h = 1$ with PMSE gains of 18 percent relative to the benchmark model, and somewhat more accurate at $h = 12$ with gains of 40 percent relative to the benchmark model. Interestingly, the LASSO predictor does not dominate the computationally less demanding ridge regression predictor.

3.4 Forecast Combinations

While one should not read too much into the (minor) differences in forecast accuracy between alternative shrinkage methods, the results in Table 3 underscore that some form of shrinkage, be it in the form of bagging or other methods, is highly advantageous in this application. Yet another branch of the economic forecasting literature has explored the use of forecast combinations as an alternative to shrinkage estimation. In this subsection we will consider several variations of this method, starting with the least sophisticated methods. The results are summarized in Tables 4 and 5.

3.4.1 Forecast Combination: One Extra Predictor at a Time

Equal-weighted Forecast Combinations 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. Each forecast receives weight $1/M$. Effectively this amounts to averaging across M forecasts. The idea of using averages of forecasts to improve forecast accuracy dates back to Bates and Granger (1969). Equal-weighted forecast combinations, despite their simplicity, have been found to produce highly accurate forecasts in a wide range of applications. Table 4 shows that this forecasting approach in our application does not work well. While equal-weighted forecasts beat the benchmark model, the PMSE reductions are limited to 3 percent at $h = 1$ and to 15 percent at $h = 12$.

One potential problem with using the arithmetic mean of forecasts is its sensitivity to outliers. For that reason some researchers prefer to use the median forecast or the trimmed mean of the M forecasts instead. Table 4 shows that this modification does little to improve forecast accuracy. In fact, the median forecast is distinctly less accurate than the equal-weighted forecast with gains relative to the benchmark model of only 1 percent at $h = 1$ and 6 percent at $h = 12$. Similar results are obtained for the trimmed mean.

ARM Forecasts A possible explanation of the relatively poor performance of mean forecasts, is the assumption of equal weights. In principle, one ought to be able to improve accuracy of forecast combination methods by letting the data choose the forecast weights. One such method is *adaptive regression by mixing* or *ARM*, as proposed by Yang (2001, 2003). There are two steps involved: In the first step, the first part of the sample available to the forecaster at a given point in time is used to estimate the forecasting models $i = 1..., M$, where each model includes one of the M extra predictors of interest. In the second step, each of these fitted models is used to generate predictions of the realizations of inflation in the remainder of the sample. Appropriate weights for the M forecasting models are constructed based on their predictive success in the second part of the sample. The lower the PMSE of model i in the second part of the sample, the higher its weight in the forecasting model. For a detailed description of this algorithm see Yang (2003, p. 786). Rather than splitting the sample in half, as described in Yang (2003), we conducted a grid search allowing between 30% and 90% of the sample to be allocated for initial estimation with increments of 10%. The results reported below are based on the most favorable results. Table 4 shows that *ARM* works moderately well in our context. The

PMSE gains at $h = 1$ are 7 percent; at $h = 12$ they rise to 32 percent, making ARM superior to the median forecast and the equal-weighted combination forecast.

AFTER Forecasts Since the ARM method involves splitting the sample, and both the estimation and the evaluation half of the sample tend to be short in our application, unreliable estimates of the model parameters and noisy measures of predictive performance may limit the success of this method. This suggests that we consider alternative methods that retain the idea of weighting forecasting models by some measure of predictive performance without relying on sample-splitting. One such method is aggregation of forecasts through exponential reweighting ($AFTER$) as proposed by Yang (2004, p. 186-188). Another such method is Bayesian model averaging of the M candidate models, as applied in Wright (2003), will be considered below.

The $AFTER$ method works as follows: The $i = 1..., M$ forecasting models are estimated on the full sample available to the forecaster. The first recursive forecast is obtained by assigning equal weights to all forecasting models. Subsequent recursive forecasts are based on recursively updated weights that reflect the predictive accuracy of each candidate model over the length of the recursive forecasting exercise up to that point in time. The weight function used in implementing this algorithm is given in equation (4) of Yang (2004). The conditional variance estimates used in equation (4) are obtained from equation (5) of Yang (2004). Table 4 shows that the $AFTER$ algorithm does not work well at the one-month horizon. In fact, it is the only competitor to perform worse than the inflation-only benchmark model. At the 12-month horizon there is a large improvement with gains of 28%, but even in that case $AFTER$ is still inferior to ARM. This result illustrates that methods that do not involve sample splitting are not necessarily superior to methods that do.

Forecasts based on Bayesian Model Averaging In related work, Wright (2003b) has shown that the accuracy of equal-weighted forecast combination methods may be improved upon further by weighting the individual forecast models based on the posterior probabilities associated with each forecasting model, which makes Wright's BMA method a natural competitor to consider. Whereas Wright imposes one lag of inflation in the benchmark model, we allow for potentially more than one lag of inflation in implementing his procedure. Otherwise our approaches are identical.

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 4 presents the outcome of a grid search over ϕ.

We find that, as in Wright (2003b), the BMA method is clearly superior to the equal-weighted forecast combination method. It also is superior to the median forecast and the $AFTER$ forecast, and about the equal of the ARM forecast. Nevertheless, with a ratio of 90 percent at $h = 1$ and 69 percent at $h = 12$, this BMA method in our application is less accurate than the BA and CBA methods or for that matter the other shrinkage methods.

3.4.2 Forecast Combination: Randomly Chosen Subsets of Extra Predictors

Forecasts Based on Bayesian Model Averaging As the previous discussion illustrated, 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 generate the forecast with the lowest possible PMSE, whether we use equal weights or data-based weights, because the search is restricted to a subset of the possible combinations of the extra predictors. A complete Bayesian solution to this problem 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 $2^{30} = 1073741824$ 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. Also see George and McCulloch (1993) for an alternative stochastic search variable selection algorithm.

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 demanding. 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 $\phi \epsilon \{0, 0.01, 0.05, 0.1, 0.2, 0.3, 0.4, 0.5, 1, 2, 5, 100\}$ in the notation used by Raftery et al., where ϕ measures the prior standard deviation of the standardized predictor data (see Raftery et al. for further details). We report the best empirical results in the last column of Table 5. 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 values of ϕ.

This version of BMA produces clearly more accurate results than the restricted version involving only one extra predictor at a time. Compared to BMA based on one extra predictor at a time, at the one-month horizon the PMSE ratio for the best BMA predictor falls from 90 percent to 82 percent and at the one-year horizon from 69 percent to 62 percent. Thus, the forecast accuracy of the Raftery et al. method is comparable to bagging methods and other shrinkage methods.

ARM Forecasts It is natural to broaden the set of candidate models underlying the ARM method along similar lines. We therefore also applied the ARM method based on 5000 randomly chosen subsets of the predictors. Table 5 shows that this greatly improves the accuracy of the one-month ahead forecasts, but not of the one-year ahead forecasts. In the former case, ARM is about as accurate as the Raftery et al. BMA method or the BA method, whereas in the latter case it is noticeably less accurate.

AFTER Forecasts The same modification may also be applied to the $AFTER$ method. Table 5 shows that the accuracy of the $AFTER$ forecasts improves at both horizons, but remains below that of the Raftery et al. method at the one-month horizon. Whereas the one-year forecast is comparable to that of the Raftery et al. method and even slightly better than BA and CBA, the one-month ahead forecast has a noticeably larger recursive PMSE than the many of the leading competitors.

4 Conclusion

In this paper, 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. consumer price inflation. We proposed three new variants of the bagging algorithm specifically designed for this type of forecasting problem. Over a twenty-year period, we compared the accuracy of simulated out-of-sample forecasts based on these bagging methods to that of alternative forecasting methods for U.S. consumer price 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 various shrinkage estimators, unweighted combination forecasts, median and trimmed mean forecasts, ARM forecasts, $AFTER$ forecasts, and finally forecasts obtained by Bayesian model averaging.

Consumer price inflation is not only one of the key macroeconomic variables of interest to businesses and policymakers, but it is also more difficult to predict than real economic growth, for example. This is especially true for the period since the 1980s, on which we focus; a period for which evidence that real economic indicators help predict inflation has become weaker. The empirical evidence presented in this paper showed that bagging can achieve large reductions in prediction mean-squared errors, even in challenging applications such as inflation forecasting. The gains in accuracy compare favorably with the benchmark model and with results reported in previous studies. At the one-month horizon, regression-based bagging methods achieve gains relative to the inflation-only model of 16-18 percent. At the one-year horizon, the gains increase to 35-36 percent.

However, bagging is not the only method capable of achieving such gains. We found that the Bayesian shrinkage predictor, the ridge regression predictor, the iterated $LASSO$ predictor and the Bayesian model average predictor based on random subsets of extra predictors are about as accurate as the bagging predictor. ARM and $AFTER$ predictors based on random subsets of extra predictors also performed well in some cases, but their performance was more uneven across horizons. The high accuracy of the ridge regression predictor in particular, suggests that similar accuracy is feasible at much lower computational cost than required for bagging or for Bayesian model averaging based on random subsets of predictors. Recently proposed asymptotic approximations to bagging methods for orthogonalized predictors such as CBA and BA^F, however, may eliminate the need for computer simulation in bagging (see Stock and Watson 2006). It will be of interest to see how these asymptotic approximations compare to the full-fledged bagging approach employed in this paper.

The fact that several methods are about equally accurate in this application does not mean that the choice of forecasting method does not matter. We showed that other methods including equal-weighted forecasts, median forecasts, ARM forecasts, $AFTER$ forecasts, and Bayesian forecast averages based on one extra predictor at a time, do not perform well in this application. We observed that forecast combination methods based on one extra predictor at a time may unduly constrain the predictor set and demonstrated that relaxing this constraint may greatly improve forecast accuracy.

A final question of interest is which of the three variants of bagging is preferred. We showed that there is no clear ranking between bagging methods based on untransformed regressors and bagging methods based on orthogonalized regressors. Both perform well. While bagging methods based on factor models also may perform well in some cases, their accuracy is sensitive to the choice of rank. In our application, the BA^F predictor was not as robust as the BA and CBA predictors. Nevertheless, we found that compared to the unrestricted factor model of rank r, the BA^F predictor tended to improve forecast accuracy for common choices of r.

Since in our application the set of potential predictors is only moderately large, the principal components may be estimated too imprecisely for the factor model forecast or its bagging equivalent to work well. An interesting avenue for future research will be the use of factor bagging methods on panels with larger cross-sections. Some preliminary evidence has been provided in Stock and Watson (2006) and in Edelstein (2006). Here as well the question arises whether simple and computationally inexpensive shrinkage methods may perform as well or better than bagging.

Data Sources

The indicators of real economic activity were obtained from http://www.economagic.com:

$INDPRO$	industrial production
$HOUST$	housing starts
$HSN1F$	house sales
ISM	ISM index of manufacturing activity
$HELPWANT$	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
$TOTASS\_AUSA$	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

References

1. Avramov, D. (2002), “Stock Return Predictability and Model Uncertainty,” *Journal of Financial Economics*, 64, 423-458.
2. Bates, J.M., and C.W.J. Granger (1969), “The Combination of Forecasts,” *Operations Research Quarterly*, 20, 451-468.
3. Bernanke, B.S., and J. Boivin (2003), “Monetary Policy in a Data-Rich Environment,” *Journal of Monetary Economics*, 50, 525-546.
4. Breiman, L. (1996), “Bagging Predictors,” *Machine Learning*, 36, 105-139.
5. Bühlmann, P. and B. Yu (2002), “Analyzing Bagging,” *Annals of Statistics*, 30, 927-961.
6. 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.
7. Cremers, K.J.M. (2002), “Stock Return Predictability: A Bayesian Model Selection Perspective,” *Review of Financial Studies*, 15, 1223-1249.
8. Edelstein, P. (2006), “Commodity Prices, Inflation Forecasts, and Monetary Policy,” mimeo, Department of Economics, University of Michigan.
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 (2005), “The Generalized Factor Model: One-Sided Estimation and Forecasting,” *Journal of the American Statistical Association*, 100, 830-840.
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 (2006), “On the Selection of Forecasting Models,” *Journal of Econometrics*, 130, 273-306.
17. Inoue, A. and M. Shintani (2006), “Bootstrapping GMM Estimators for Time Series,” *Journal of Econometrics*, 133, 531-555.
18. Koop, G., and S. Potter (2004), “Forecasting in Dynamic Factor Models Using Bayesian Model Averaging,” *Econometrics Journal*, 7, 550-565.

19. Lee, T.-H., and Y. Yang (2006), "Bagging Binary and Quantile Predictors for Time Series," *Journal of Econometrics*, 135, 465-497.

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. Stock, J.H., and M.W. Watson (1999), "Forecasting Inflation," *Journal of Monetary Economics*, 44, 293-335.

28. 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.

29. Stock, J.H., and M.W. Watson (2002b), "Macroeconomic Forecasting Using Diffusion Indexes," *Journal of Business and Economic Statistics*, 20, 147-162.

30. Stock, J.H., and M.W. Watson (2003), "Forecasting Output and Inflation: The Role of Asset Prices ," *Journal of Economic Literature*, 41, 788-829.

31. Stock, J.H., and M.W. Watson (2006), "Shrinkage Methods for Forecasting Using Many Predictors," mimeo, Department of Economics, Harvard University.

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

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.

37. Yang, Y. (2001), “Adaptive Regression by Mixing” *Journal of the American Statistical Association*, 96, 574-809.

38. Yang, Y. (2003), “Regression with Multiple Candidate Models: Selecting or Mixing?” *Statistica Sinica*, 13, 783-809.

39. Yang, Y. (2004), “Combining Forecasting Procedures: Some Theoretical Results,” *Econometric Theory*, 20, 176-222.

Table 1. U.S. Consumer Price Inflation Forecasts

Evaluation Period: 1983.8-2003.7

	Correlated Regressor Model Recursive PMSE Relative to Benchmark Model				
Horizon	UR	PT	BA	CPT	CBA
1 Month	**0.923**	**0.857**	**0.818**	**0.897**	**0.842**
12 Months	**0.703**	**0.697**	**0.653**	**0.636**	**0.637**

Table 2. U.S. Consumer Price Inflation Forecasts
Evaluation Period: 1983.8-2003.7

	Factor Model Recursive PMSE Relative to Benchmark Model					
	1-Month Horizon			**12-Month Horizon**		
Rank	UR^F	PT^F	BA^F	UR^F	PT^F	BA^F
1	**0.971**	**0.967**	**0.975**	**0.636**	**0.668**	**0.632**
2	**0.988**	**0.988**	**0.970**	**0.687**	**0.668**	**0.683**
3	**1.016**	**1.000**	**0.975**	**0.708**	**0.634**	**0.659**
4	**0.953**	**0.952**	**0.925**	**0.688**	**0.645**	**0.626**
5	**0.961**	**0.963**	**0.929**	**0.698**	**0.686**	**0.640**
6	**0.961**	**0.960**	**0.934**	**0.702**	**0.697**	**0.659**
7	**0.960**	**0.957**	**0.929**	**0.729**	**0.710**	**0.707**
8	**0.919**	**0.916**	**0.891**	**0.699**	**0.677**	**0.678**
30	**0.923**	**0.905**	**0.823**	**0.703**	**0.708**	**0.755**

Table 3. U.S. Consumer Price Inflation Forecasts
Evaluation Period: 1983.8-2003.7

	Shrinkage Methods Recursive PMSE Relative to Benchmark Model		
Horizon	**Bayesian shrinkage**	**Ridge regression**	**LASSO**
1 Month	0.817	0.807	0.825
12 Months	0.632	0.623	0.596

Table 4. U.S. Consumer Price Inflation Forecasts
Evaluation Period: 1983.8-2003.7

	Forecast Combination Methods One Extra Predictor at a Time Recursive PMSE Relative to Benchmark Model				
Horizon	**Equal-weighted**	**Median**	**ARM**	**AFTER**	**BMA Wright**
1 Month	0.968	0.990	0.933	1.018	0.904
12 Months	0.846	0.941	0.681	0.724	0.686

Table 5. U.S. Consumer Price Inflation Forecasts
Evaluation Period: 1983.8-2003.7

	Forecast Combination Methods Random Subsets of Predictors Recursive PMSE Relative to Benchmark Model		
Horizon	**ARM**	**AFTER**	**BMA Raftery et al.**
1 Month	0.815	0.882	0.821
12 Months	0.690	0.620	0.618