GENERALIZED SHRINKAGE METHODS FOR FORECASTING USING MANY PREDICTORS

This revision: October 2009

James H. Stock

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

and

Mark W. Watson*

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

ABSTRACT

This paper provides a simple shrinkage representation that describes the operational characteristics of various forecasting methods designed for a large number of orthogonal predictors (such as principal components). These methods include pretest methods, Bayesian model averaging, empirical Bayes, and bagging. We compare empirically forecasts from these methods to dynamic factor model (DFM) forecasts using a U.S. macroeconomic data set with 143 quarterly variables spanning 1960-2008. For many series, including measures of real economic activity, the shrinkage forecasts are inferior to the DFM forecasts. For other series, however, the shrinkage methods improve upon the DFM forecasts, suggesting that for those series the DFM is overly restrictive.

We thank Jean Boivin, Domenico Giannone, Lutz Kilian, Serena Ng, Lucrezia Reichlin, Mark Steel, and Jonathan Wright for helpful discussions, and Anna Mikusheva for research assistance. The theoretical results in this paper were circulated earlier under the title "An Empirical Comparison of Methods for Forecasting Using Many Predictors." Replication files for the results in this paper can be downloaded from http://www.princeton.edu/~mwatson. This research was funded in part by NSF grants SBR-0214131 and SBR-0617811.

1. Introduction

Over the past ten years, the dynamic factor model (DFM) (Geweke (1977)) has been the predominant framework for research on macroeconomic forecasting using many predictors. The conceptual appeal of the DFM is twofold: methods for estimation of factors in a DFM can overcome the curse of dimensionality (Forni, Hallin, Lippi, and Reichlin (2000, 2004), Bai and Ng (2002, 2006), and Stock and Watson (1999, 2002a, b)), and the DFM arises naturally from log-linearized structural macroeconomic models including dynamic stochastic general equilibrium models (Sargent (1989), Bovin and Giannoni (2006)).[1] But the forecasting implications of the DFM – that the many predictors can be replaced by a small number of estimated factors – might not be justified in practice. Indeed, Eichmeier and Ziegler's (2008) meta-study finds mixed performance of DFM forecasts, providing a reason to consider other ways to handle many predictors. Accordingly, some recent papers have considered whether DFM macro forecasts can be improved upon using other many-predictor methods, including high-dimensional Bayesian Vector Autogression (e.g. De Mol, Giannone, and Reichlin (2008), Carriero, Kapetanios, and Marcellino (2009)), Bayesian model averaging (Koop and Potter (2004), Wright (2004)), bagging (Inoue and Kilian (2008)), LASSO (De Mol, Giannone, and Reichlin (2008), Bai and Ng (2007)), boosting (Bai and Ng (2007)), and forecast combination (multiple authors).

One difficulty in comparing these high-dimensional methods theoretically is that their derivations generally rely on specific modeling assumptions (for example, i.i.d. data and strictly exogenous predictors), and it is not clear from these derivations what the algorithms are actually doing when applied in settings in which the modeling assumptions do not hold. Moreover, although there have been empirical studies of the performance of many of these methods for macroeconomic forecasting, it is difficult to draw conclusions across methods because of differences in data sets and implementation across studies.

This paper therefore has two goals. The first is characterize the properties of some forecasting methods applied to many orthogonal predictors in a time series setting

[1] For a survey of econometric DFM research, see Bai and Ng (2008).

in which the predictors are predetermined but not strictly exogenous. The results cover pretest and information-criterion methods, Bayesian model averaging (BMA), empirical Bayes (EB) methods, and bagging. It is shown that asymptotically all these methods have the same "shrinkage" representation, in which the weight on a predictor is the OLS estimator times a shrinkage factor that depends on the *t*-statistic of that coefficient. These representations are a consequence of the algorithms and they hold under weak stationarity and moment assumptions about the actual statistical properties of the predictors; thus these methods can be compared directly using these shrinkage representations.

The second goal is to undertake an empirical comparison of these shrinkage methods DFM methods using a U.S. data set that includes 143 quarterly economic time series spanning 49 years. The DFM imposes a strong restriction – that only the first few principle components (ordered according to the eigenvalues of the original predictor matrix) are needed for efficient forecasting – and the shrinkage methods provide a way to assess the empirical validity of this restriction.

We find that, for many macroeconomic time series, among linear estimators the DFM forecasts make efficient use of the information in the many predictors by using only a small number of estimated factors. These series include measures of real economic activity and some other central macroeconomic series, including some interest rates. For these series, the shrinkage methods with estimated parameters fail to provide mean squared error improvements over the DFM.

For other macroeconomic series, however, the shrinkage methods can provide noteworthy improvements over DFMs. Series in this category include real wages and measures of regional residential construction activity. For these series, the DFM appears not to be an adequate approximation, at least for forecasting purposes, and the use of principle components beyond the first few, combined with shrinkage, reduces mean squared errors. Finally, none of the methods considered here help much for series that are notoriously difficult to forecast, such as exchange rates or price inflation.

The shrinkage representations for forecasts using orthogonal predictors are described in Section 2. Section 3 describes the data and the forecasting experiment. Section 4 presents the empirical results, and Section 5 offers some concluding remarks.

2. Shrinkage Representations of Forecasting Methods

We consider the multiple regression model with orthonormal regressors,

$$Y_t = \delta' P_{t-1} + \varepsilon_t, \quad t = 1,\dots, T, \quad P'P/T = I_n \tag{1}$$

where P_t is a n-dimensional predictor known at time t with i^{th} element P_{it}, Y_t is the variable to be forecast, and the error ε_t has variance σ^2. It is assumed that Y_t and P_t have sample mean zero. (Extensions to multi-step forecasting and including lagged values of Y are discussed below.) For the theoretical development it does not matter how the regressors are constructed; in our applications and in the recent empirical econometric literature they are constructed as the first n principal components, dynamic principal components, or a variant of these methods, using an original, potentially larger set of regressors, $\{X_t\}$.

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

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

where $\tilde{Y}_{T+1|T}$ is the forecast of Y_{T+1} made using data through time T, $\hat{\delta}_i = T^{-1}\sum_{t=1}^{T} P_{it-1}Y_t$ is the OLS estimator of δ_i (the i^{th} element of δ), $t_i = \sqrt{T}\hat{\delta}_i/s_e$, where $s_e^2 = \sum_{t=1}^{T}(Y_t - \hat{\delta}'P_{t-1})^2/(T-n)$, and ψ is a function specific to the forecasting method. We consider four classes of forecasting procedures: pretest and information criterion methods, Bayesian methods (including Bayesian model averaging), empirical Bayes, and bagging. The factor κ depends on the method. For pretest methods and bagging, $\kappa = 1$. For the Bayes methods, $\kappa = (s_e/\hat{\sigma})$, where $1/\hat{\sigma}^2$ is the Bayes method's posterior mean of $1/\sigma^2$. This factor arises because the posterior for σ may not concentrate around s_e^2.

Under general conditions, for Bayes, empirical Bayes, bagging and pre-test estimators, $0 \leq \psi(x) \leq 1$, so the operational effect of these methods is to produce linear combinations in which the weights are the OLS estimator, shrunk towards zero by the factor ψ. This is the reason for referring to (2) as the shrinkage representation of these forecasting methods.

A key feature of these results is that the proof that the remainder term in (2) is $o_p(1)$ for the different methods relies on much weaker assumptions on the true distribution of (Y, P) than the modeling assumptions used to derive the methods. As a result, the performance of these methods can be understood and analyzed even if they are applied in circumstances in which the original modeling assumptions clearly do not hold, for example when they are applied to multistep-ahead forecasting.

2.1 Pretest and Information Criterion Methods

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

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

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

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

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

2.2 Bayes Methods

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

(M1) $\{\varepsilon_t\} \perp \{P_t\}$ and ε_t is i.i.d. $N(0,\sigma^2)$.

We also adopt the usual modeling assumption of squared error loss. Bayes procedures constructed under assumption (M1) with squared error loss will be called "Normal Bayes" (NB) procedures. Note that we treat (M1) as a modeling tool, where the model is in general misspecified, that is, the true probability law for the data, or data generating process (DGP), is not assumed to satisfy (M1).

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

(M2) $\{\tau_i = \sqrt{T}\delta_i/\sigma\}$ ~ i.i.d G_τ, $\sigma^2 \sim G_{\sigma^2}$, and $\{\tau_i\}$, σ^2 are independent

If T is fixed, the only two restrictions in (M2) are that δ_i is i.i.d. and that σ^2 enters the conditional distribution of δ_i given σ^2 only as a scale factor.

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

$$\tilde{\delta}_i^{NB} = E_{\delta,\sigma^2}(\delta_i \mid Y, P), \tag{5}$$

where the subscript E_{δ,σ^2} indicates that the expectation is taken with respect to δ (which reduces to δ_i by independence under (M2)) and σ^2. Under (M1), ($\hat{\delta}$, s_e^2) are sufficient for ($\hat{\delta}$, σ^2). Moreover $\hat{\delta}_i$ and $\hat{\delta}_j$ are independently distributed for all $i \neq j$ conditional on (δ, σ^2), and $\hat{\delta}_i | \delta, \sigma^2$ is distributed N(δ_i, σ^2/T). Thus (M1) and (M2) imply that, conditional on σ^2, the posterior mean has the so-called simple Bayes form (Maritz and Lwin (1989)),

$$\tilde{\delta}_i^{NB} | \sigma^2 = \hat{\delta}_i + \frac{\sigma^2}{T} \ell_\delta(\hat{\delta}_i), \tag{6}$$

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

The shrinkage representation of the NB estimator follows from (6) by performing the change of variables $\tau_i = \sqrt{T}\,\delta_i/\sigma$. For priors satisfying (M2) and under conditions made precise below, the shrinkage function for the NB estimator is,

$$\psi^{NB}(u) = 1 + \ell(u)/u, \tag{7}$$

where $\ell(u)$ = dln$m(u)$/du, $m(u) = \int \phi(u-\tau) dG_\tau(\tau)$, and ϕ is the standard normal density. Integrating over the posterior distribution of σ^2 results in the posterior mean approaching its probability limit, which leads to ψ^{NB} being evaluated at $u = t_i \times \text{plim}(\sigma/\hat{\sigma})$.

It is shown in the Appendix that, if the prior density $g_\tau = dG_\tau(u)/du$ is symmetric around zero and is unimodal, then for all u,

$$\psi^{NB}(u) = \psi^{NB}(-u) \text{ and } 0 \le \psi^{NB}(u) \le 1. \tag{8}$$

2.3 Bayesian Model Averaging.

Our treatment of BMA with orthogonal regressors follows Clyde, Desimone, and Parmigiani (1996), Clyde(1999a,b), and Koop and Potter (2004). The Clyde, Desimone, and Parmigiani (1996) BMA setup adopts (M1) and a Bernoulli prior model for variable inclusion with a g-prior for δ conditional on inclusion. Specifically, with probability p let $\delta_i|\sigma \sim$ N(0, $\sigma^2/(gT)$) (so $\tau_i \sim$ N(0, $1/g$)), and with probability $1 - p$ let $\delta_i = 0$ (so $\tau_i = 0$). Note that this prior model satisfies (M2). Direct calculations show that, under these priors, the shrinkage representation (7) specializes to

$$\psi^{BMA}(u) = \frac{pb(g)\phi(b(g)u)}{(1+g)[pb(g)\phi(b(g)u)+(1-p)\phi(u)]} \tag{9}$$

where $b(g) = \sqrt{g/(1+g)}$ and ϕ is the standard normal density, and where ψ^{BMA} is evaluated at $u = \kappa t_i$, just as in the general case (7).

2.4 Empirical Bayes

Empirical Bayes (EB) estimation treats the prior G as an unknown distribution to be estimated. Under the stated assumptions, $\{\hat{\delta}_i\}$ constitute n i.i.d. draws from the marginal distribution m, which in turn depends on the prior G. Because the conditional distribution of $\hat{\delta}\,|\,\delta$ is known under (M1), this permits inference about G. In turn, the estimator of G can be used in (6) to compute the empirical Bayes estimator. The estimation of the prior can be done either parametrically or nonparametrically. We refer to the resulting empirical Bayes estimator generically as $\tilde{\delta}_i^{EB}$. The shrinkage function for the EB estimator is,

$$\psi^{EB}(u) = 1 + \hat{\ell}(u)/u, \tag{10}$$

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

2.5 Bagging

Bootstrap aggregation or "bagging" (Breiman (1996)) (BG) smoothes the hard threshold in pretest estimators by averaging over a bootstrap sample of pre-test estimators. Inoue and Kilian (2008) apply bagging to a forecasting situation like that considered in this paper and report some promising results; also see Lee and Yang (2006). Bühlmann and Yu (2002) considered bagging with a fixed number of strictly exogenous regressors and i.i.d. errors, and showed that asymptotically the bagging estimator can be represented in the form (2), where (for $u \neq 0$),

$$\psi^{BG}(u) = 1 - \Phi(u + c) + \Phi(u - c) + t^{-1}[\phi(u - c) - \phi(u + c)], \qquad (11)$$

where c is the pre-test critical value, ϕ is the standard normal density, and Φ the standard normal CDF. We consider a variant of bagging in which the bootstrap step is conducted using a parametric bootstrap under the exogeneity-normality assumption (M1). This algorithm delivers the Bühlmann-Yu expression (11), however the expression obtains under weaker assumptions on the number and properties of the regressors.

2.6 Theoretical results

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

Let P_T denote a vector of predictors used to construct the forecast and let $\{\tilde{\delta}_i\}$ denote the estimator of the coefficients for the method at hand. Then each method produces forecasts of the form $\tilde{Y}_{T+1|T} = \sum_{i=1}^{p} \tilde{\delta}_i P_{iT}$, with shrinkage approximation $\hat{Y}_{T+1|T} = \sum_{i=1}^{p} \psi(\kappa t_i)\hat{\delta}_i P_{iT}$ for appropriately chosen $\psi(.)$. This section shows that $\tilde{Y}_{T+1|T} - \hat{Y}_{T+1|T} \xrightarrow{m.s.} 0$ for the NB and BG forecasts. (It follows immediately from the definition of the pretest estimator that its shrinkage representation is $\tilde{Y}^{PT}_{T+1|T} = \sum_{i=1}^{n} \psi^{PT}(t_i)\hat{\delta}_i P_{iT}$, where $\psi^{PT}(u) = 1(|u| > c)$, is exact).

First consider the NB forecast described in section (2.2). If σ^2 was assumed known, then equation (7) implies that the shrinkage representation would hold exactly with $\kappa = s_e/\sigma$. The difference $\tilde{Y}^{NB}_{T+1|T} - \hat{Y}^{NB}_{T+1/T}$ is therefore associated with estimation of σ^2. The properties of the sampling error associated with estimation of σ^2 depend on the DGP and the modeling assumptions (likelihood and prior) underlying the construction of the Bayes forecast. Assumptions associated with the DGP and Bayes procedures are provided below. Several of these assumptions use the variable $\zeta = \hat{\sigma}^2/\sigma^2$, where, as described above, $1/\hat{\sigma}^2$ is the posterior mean of $1/\sigma^2$. The assumptions use the expectation operator E, which denotes expectation with respect to the true distribution of Y and P, and E^M, which denotes expectation with respect to the Bayes posterior distribution under the modeling assumptions (M1) and (M2).

The assumptions for the NB forecasts are:

(A1) $\max_i |P_{iT}| \leq P_{max}$, a finite constant.

(A2) $E\left(T^{-1}\sum_t Y_t^2\right)^2 \sim O(1)$.

(A3) $n/T \to \upsilon$, where $0 \leq \upsilon < 1$.

(A4) $E\{E^M[(\zeta - 1)^4|Y,P]\}^4 \sim O(T^{-4-\delta})$ for some $\delta > 0$.

(A5) $E\{E^M[\zeta^{-4}|Y,P]\}^4 \sim O(1)$.

(A6) $\sup_u|u^m d^m \psi^{NB}(u)/du^m)| \leq M$ for $m = 1, 2$.

Assumptions (A1)-(A2) are restriction on the DGP, while (A3) is the asymptotic nesting. Assumptions (A4)-(A5) involve both the DGP and the assumed model for the Bayes forecast, and these assumptions concern the rate at which the posterior for σ concentrates around $\hat{\sigma}$. To interpret these assumptions, consider the usual Normal-Gamma conjugate prior (i.e., $\tau_i \sim N(0,g^{-1})$ and $1/\sigma^2 \sim$ Gamma). A straightforward calculation shows that $E^M[(\zeta - 1)^4|Y,P] = 12(\nu+2)/\nu^3$ and $E^M[\zeta^{-4}|Y,P] = (\nu/2)^4/[(\nu/2-1)(\nu/2-2)(\nu/2-3)(\nu/2-4)]$ where ν denotes the posterior degrees of freedom. Because $\nu = O(T)$ under (A3), $E\{E^M[(\zeta - 1)^4|Y,P]\}^4 \sim O(T^{-8})$, and $E[E^M[\zeta^{-4}|Y,P]]^4 \sim O(1)$, so that assumptions (A4) and (A5) are satisfied in this case regardless of the DGP. Assumption

(A6) rules out priors that induce mass points in ψ^{NB} or for which $\psi^{NB}(u)$ approaches 1 very slowly as $u \to \infty$.

With these assumptions, the behavior of $\tilde{Y}^{NB}_{T+1|T} - \hat{Y}^{NB}_{T+1/T}$ is characterized in the following theorem:

Theorem 1: Under (A1)-(A6), $\tilde{Y}^{NB}_{T+1|T} - \hat{Y}^{NB}_{T+1/T} \overset{m.s.}{\to} 0$.

Proofs are given in the appendix.

An analogous result holds for the bagging forecast. To prove this result, we make two additional assumptions:

(A7) $n/B \to 0$.

(A8) $\max_i E(t_i^{12}) < \infty$.

In (A7), B denotes the number of bootstrap replications, and the finite twelfth moment assumption in (A8) simplifies the proof of the following theorem:

Theorem 2: Under (A1)-(A3) and (A7)-(A8), $\tilde{Y}^{BG}_{T+1|T} - \hat{Y}^{BG}_{T+1/T} \overset{m.s.}{\to} 0$.

Remarks

1. The theorems show that shrinkage factor representations hold under weaker assumptions than those upon which the estimators are derived: the shrinkage factor representations are consequences of the algorithm, not properties of the DGP.
2. Consider the (frequentist) MSE risk of an estimator $\tilde{\delta}$, $R(\tilde{\delta},\delta) = E(\tilde{\delta} - \delta)'(\tilde{\delta} - \delta)$, which is motivated by interest in the prediction problem with orthonormal regressors. Setting $\tilde{\delta} = \psi(\kappa t_i)\sqrt{T}\hat{\delta}_i$, this risk is $E(\tilde{\delta} - \delta)'(\tilde{\delta} - \delta) = \upsilon n^{-1}\sum_{i=1}^{n} E\left(\psi(\kappa t_i)\sqrt{T}\hat{\delta}_i - \sqrt{T}\delta_i\right)^2$. Suppose that $(\sqrt{T}(\hat{\delta}_i - \delta_i)/\sigma_\varepsilon,\ \hat{\sigma}^2/\sigma^2)$ are

identically distributed, $i = 1,\ldots, n$, and let $r_\psi(\tau_i) = E\left(\psi(\kappa t_i)\sqrt{T}\hat{\delta}_i / \sigma - \tau_i\right)^2$,

where $\tau_i = \sqrt{T}\ \delta_i/\sigma_\varepsilon$. Then $R(\tilde{\delta}, \delta) = \upsilon\sigma^2 \int r_\psi(\tau) d\tilde{G}_n(\tau)$, where $\tilde{G}_n$ is the empirical cdf of $\{\tau_i\}$. Thus the risk depends only on ψ, $\tilde{G}_n$ and the sampling distribution of $(\sqrt{T}(\hat{\delta}_i - \delta_i)/\sigma_\varepsilon,\ \hat{\sigma}_\varepsilon^2/\sigma_\varepsilon^2)$. Holding constant this sampling distribution, risk rankings of various estimators depend only on $\tilde{G}_n$. If $\sqrt{T}(\hat{\delta}_i - \delta_i)/\sigma_\varepsilon$ is asymptotically normally distributed, then the optimal choice of ψ is ψ^{NB}, with prior distribution equal to (the limit of) G_n (for details see Knox, Stock, and Watson (2004)). These considerations provide a justification for thinking that parametric empirical Bayes estimators will perform well even though the model assumption (M1) used to derive the parametric Bayes estimator does not hold in the time series context of interest here.

3. For empirical Bayes estimators, the shrinkage function depends on the estimated prior. Under suitable regularity conditions, if the empirical Bayes estimation step is consistent then the asymptotic empirical Bayes shrinkage representation ψ^{EB} is ψ^{NB} with the probability limit of the estimated prior replacing G_τ.
4. These representations permit the extension of these methods to direct multistep forecasting. In a multistep setting, the errors have a moving average structure. However the forecasting methods can be implemented by substituting HAC *t*-statistics into the shrinkage representations.
5. The shrinkage factor representation of bagging allows us to ascertain whether bagging is asymptotically admissible, a result that appears to be currently unavailable. Setting ψ^{BG} equal to ψ^{NB} yields the integral-differential equation,

$$\left.\frac{\mathrm{d}\ln\int\phi(z-s)dG_\tau(s)}{\mathrm{d}z}\right|_{z=u} = u[\Phi(u-c) - \Phi(u+c)] + \phi(u-c) - \phi(u+c). \qquad (12)$$

If there is a proper prior G_τ that satisfies (12), then this is the prior for which bagging is asymptotically Bayes, in which case bagging would be asymptotically admissible. Let G_τ have density g and characteristic function $h(s) = \int e^{ist} g(t)dt$.

Then g satisfies (12) if h satisfies the Fredholm equation of the second kind, $h(s) = \int K(s,t)h(t)dt$, where

$$K(s,t) = 2\frac{e^{-t^2+st}}{s}\left[\frac{\sin(c(s-t))}{(s-t)^2} - c\frac{\cos(c(s-t))}{s-t}\right]. \qquad (13)$$

3. Empirical Analysis: Data and Methods

The overall aim of the empirical analysis is to compare forecasts based on the dynamic factor model forecasts based on the shrinkage methods. This section first describes the data set, then describes the specifics of the methods used to conduct the analysis.

3.1 The Data

The data set consists of quarterly observations on 143 U.S. macroeconomic time series from 1960:I through 2008:IV, for a total of 196 quarterly observations, with earlier observations used for lagged values of regressors as necessary. We have grouped the series into thirteen categories, which are listed in Table 1. The series are transformed by taking logarithms and/or differencing. In general, first differences of logarithms (growth rates) are used for real quantity variables, first differences are used for nominal interest rates, and second differences of logarithms (changes in growth rates) for price series. The series and their transformations are listed in Appendix Table B.1. Table B.2 specifies the transformation used for the h-step ahead forecasted variable, Y_{t+h}^h, which depends on the transformation applied to the series. Generally speaking, for real activity variables, Y_{t+h}^h is the h-period growth at an annual rate; for interest rates, Y_{t+h}^h is the h-period change; and for nominal price and wage series, Y_{t+h}^h is h-period inflation at an annual rate, minus current 1-period inflation.

Of the 143 series in the data set, 34 are higher-level aggregates, which are related by an identity to subaggregates. Because including the higher-level aggregates does not add information, only the 109 disaggregated series were used to compute principle components; these 109 series are indicated in column "E" in Table B.1. However, all 143 series were used, one at a time, as the dependent variable to be forecasted.

3.2 Methods

This section describes the estimation of the model parameters and mean squared error (MSE). The standard procedure in the macro forecast comparison literature is to use pseudo out-of-sample forecasting methods, for example by estimating MSEs using recursive forecasts. Doing so here would introduce the undesirable complication that either the number of principal components would need to increase over the recursive sample, or the ratio n/T would change over the recursive sample, or fewer principle components could be considered if rolling methods were used. As described below, we therefore adopt a nonstandard approach and estimate the model parameters and MSE by cross-validation, moving through the sample with a "leave out" window and using observations on both sides for the OLS estimation of δ and σ_ε^2.

Forecasting methods. We examine six forecasting methods.

1. *DFM-5*. The DFM-5 forecast uses the first five principle components as predictors, with coefficients estimated by OLS and no shrinkage, and omitting the remaining principle components.
2. *Pretest*. The pretest shrinkage function is given by (4) and has one parameter, c.
3. *Bagging*. The bagging shrinkage function is given by (11) and has one parameter, c.
4. *BMA*. The BMA shrinkage function is given by (9) and has two parameters, p and g. Because the parameters are estimated, the BMA method as implemented here is in fact a parametric Empirical Bayes procedure.
5. *Logit*. In addition to the methods studied in Section 2, we considered a logit shrinkage function, chosen because it is a conveniently estimated flexible functional form with two parameters, β_0 and β_1:

$$\psi^{logit}(u) = \frac{\exp(\beta_0 + \beta_1|u|)}{1+\exp(\beta_0 + \beta_1|u|)}. \tag{14}$$

6. *OLS*. For comparison purposes we also report the OLS forecast based on all principle components (so $\psi^{OLS} = 1$).

We do not investigate nonparametric empirical Bayes methods because of the limited amount of data available.

Estimation. Consider the h-step ahead series to be predicted, Y_{t+h}^{h}, let X_t denote the vector of 109 time series (transformed as in Appendix B) and let $\psi(\tau;\theta)$ denote a candidate shrinkage function with parameter vector θ. Estimation of the parameters θ and δ and of the MSE for that forecasting method for that series proceeds in three steps.

1. Autoregressive dynamics are partialed out by initially regressing Y_{t+h}^{h} and X_t on 1, Y_t^1, Y_{t-1}^1, Y_{t-2}^1, and Y_{t-3}^1; let $\tilde{Y}_{t+h}^{h}$ and $\tilde{X}_t$ denote the residuals from this regression, standardized to have unit variance in the full sample. The principle components P_t of $\tilde{X}_t$ are computed using all observations (1960:I – 2008:IV) on the 109 series in the data set that are not higher-level aggregates. The principle components are ordered according to the magnitude of the eigenvalues with which they are associated, and the first 100 standardized principle components are retained as P_t.
2. Let $\Im_t = \{1,\ldots, t–2h–3, t+2h+3,\ldots, T\}$, that is, the full data set dropping the t^{th} observation and $2h+2$ observations on either side. At each date $t = 1,\ldots, T–h$, the OLS estimators of δ are computed by regressing $\tilde{Y}_{t+h}^{h}$ on P_t using observations $t \in \Im_t$; denote these OLS estimators as $\hat{\delta}_{j,t}^{h}$, $j = 1,.., n$. Let $\hat{\tau}_{j,t}^{h}$ denote the conventional OLS t-statistic corresponding to $\hat{\delta}_{j,t}^{h}$ (not adjusting for heteroskedasticity or serial correlation).
3. The parameter θ is then estimated by minimizing the sum of squared cross-validation prediction errors:

$$\hat{\theta}^h = \text{argmin}_\theta \text{MSE}(\theta), \text{ where } \text{MSE}(\theta) = \frac{1}{T-h}\sum_{t=1}^{T-h}\left(Y_{t+h}^{t} - \sum_{i=1}^{100}\psi(\hat{\tau}_{i,t}^{h};\theta)\hat{\delta}_{i,t}^{h}P_{i,t}\right)^2 \tag{15}$$

Because these are direct forecasts, the estimated value of θ differs by forecast horizon. The estimated shrinkage function for this dependent variable and horizon is $\psi(.,\hat{\theta}^h)$, and the corresponding empirical MSE is MSE($\hat{\theta}^h$).

Because four lags of Y_t^1 were partialed out in step 1 using full-sample regressions and the residuals were rescaled to have full-sample variance of 1, the MSE in (15) has the interpretation as being relative to a full-sample direct AR(4). To emphasize this, we will refer to MSE($\hat{\theta}^h$) as a relative MSEs, and to its square root as the relative root mean squared error (RMSE).

4. Empirical Results

We begin with results for one-step ahead forecasts, then turn to multi-step ahead forecasts and results for categories of series.

4.1 Results for one-step ahead forecasts

Table 2 presents percentiles of the distribution of one-step ahead RMSEs over the 143 series for the seven forecasting methods. Table 3 presents two measures of similarity of the one-step ahead forecasts, specifically the correlation among the forecasting methods, and the mean absolute difference of the RMSEs. Tables 4 and 5 report two summary measures of the shrinkage functions across series by method. Table 4 reports the distribution across series of the root mean square shrinkage function, $\left(\frac{1}{100}\sum_{i=1}^{100}\psi(\hat{\tau}_{i,t}^h;\hat{\theta}_j)^2\right)^{1/2}$, where $\hat{\theta}_j$ is the cross-validated estimated parameter for series *j* for the row method; because $\psi = 1$ for all principle components for OLS, for OLS this measure is 1.00 for all series. For DFM-5, $\psi = 1$ for the first five principle components and zero otherwise, so this measure is $\sqrt{5/100}$ = .224 for all series. Table 5 gives the distribution across series of the average fraction of the mean squared variation in the shrinkage attributable to the first five principle components, $\sum_{i=1}^{5}\psi(\hat{\tau}_{i,t}^h;\hat{\theta}_j)^2 \Big/ \sum_{i=1}^{100}\psi(\hat{\tau}_{i,t}^h;\hat{\theta}_j)^2$, among those series for which the root mean square shrinkage function considered in Table 4 is at least 0.05[2]. The final column of Table 5 reports the fraction of these series for which at least 90% of the mean-square weight, for the row model, is placed on the first five principle components.

Tables 2-5 suggests three conclusions for one-step ahead forecasts. First, for approximately one-quarter of the series, the DFM-5 and shrinkage methods provide substantial improvements over the AR(4), with all these methods having 25th percentiles of relative RMSEs less than 0.90. On the other hand, the forecasting improvements of any of these methods over the AR(4) forecast are quite small for at least a quarter of the 143 series: the smallest relative RMSE at the 75th percentile is 0.98.

Second, the DFM-5 and shrinkage methods provide improvements across the entire distribution of series, relative to OLS. The logit model has the smallest relative RMSE at each percentile and in this sense the logit dominates both DFM-5 and the other shrinkage methods. The bagging, BMA, pretest, and DFM-5 methods have similar performances, with different methods performing better at different quantiles, so that there is no clear ranking among these other methods. As indicated in Table 3, the shrinkage methods tend to produce very similar RMSEs, with correlations of RMSEs among the shrinkage methods all exceeding 0.98. The correlation of the RMSEs of the shrinkage methods with the DFM-5 method is considerably lower, ranging from 0.90 to 0.92. The correlation of the DFM-5 and shrinkage RMSEs with the OLS RMSE is lower yet.

Third, Table 5 shows that the fraction of mean-square weight that the shrinkage methods puts on the first five principle components varies considerably across series. For approximately one-fifth of the series (21%), the logit model places at least 90% of its mean-square weight on the first five principle components. For many other series, however, the shrinkage methods put considerable weight on principle components other than the first five: for one-quarter of the series the logit model places only 5.7% of its mean-square weight on the first five principle components.

4.2 Results for multi-step ahead forecasts

---

[2] A model with shrinkage functions equal to 0.5 for one principle component and equal to zero for the remaining 99 principle components has a root mean square ψ of .05.

Table 6 summarizes the distribution of RMSEs across series of the different forecasting methods at forecast horizons of 2 and 4 quarters (the results for 3-quarters ahead fall between those for 2- and 4-quarters ahead). In general, the conclusions drawn from Table 2 for $h = 1$ apply here. All methods improve upon OLS. Of the shrinkage methods, the logit direct forecasts have the lowest RMSEs at each percentile (as they do at the 1-quarter ahead horizon), and in this sense the logit forecasts dominate the other shrinkage method forecasts. When the shrinkage and DFM-5 methods offer gains over the AR(4) benchmark, those gains tend to be greater at forecast horizons of half a year or a year than at the one-quarter horizons.

As seen in Table 5, for some series the shrinkage methods place nearly all the mean-square weight on the first five principle components, while for other series the shrinkage methods place the same fraction (or less) of mean-square weight on the first five principle components as does OLS. This raises two related questions. First, for those series for which the shrinkage methods place most of the mean-square weight on the first five principle components, would one be better off simply using DFM-5 and not bothering with the empirical shrinkage function? Second, for series for which the shrinkae method places much weight on principle components other than the first five, does doing so in fact improve the RMSE relative to DFM-5? The answer to both questions is "yes," at least at the 1- and 2-quarter forecast horizons. Specifically, Table 7 shows the median RMSE by method among the series in the lower and upper quartiles of the distribution of fractions of mean-squared weight (for the logit model) on the first five principle components, among those series with root mean square ψ's of at least 0.05; for $h = 1$, these are the series in the lower and upper quartiles for the logit model in Table 5. For example, for $h = 1$ and for series in the upper quartile of mean-squared weights on the first five principle components, the median RMSE is smallest (.886) for the DFM-5 model. For series in the lower quartile of this distribution for $h = 1$, the median RMSE is smallest for the logit model (.977) and slightly exceeds 1 for DFM-5. Interestingly, the logit RMSEs are lower in both quartiles than the DFM-5 RMSEs for 4-quarter ahead forecasts. It is noteworthy that the logit RMSE is typically quite large for the series in the lower quartile, indicating that placing most of the weight on principle components beyond the first five provides a RMSE improvement, but only a very slight one relative to the AR(4).

4.3 Results by category of series

Because the foregoing results are for all series taken together, they place most weight on the categories of series that are most heavily populated in this data set, in particular prices (37 of 143 series), employment and hours (20 series), GDP components (16 series), and industrial production (14 series). Here, we turn to an analysis of the relative performance of the different procedures, broken down by category of series.

Tables 8 and 9 summarize results by category of series (the full category descriptions are given in Table 1). Table 8 presents the median relative RMSE by forecasting method, broken out by the category of series (NIPA aggregate, industrial production, etc.), at the 1-, 2-, and 4-quarter ahead forecast horizon; that is, Table 8 breaks down Table 2 by category. Table 9 presents the median fraction of the mean-squared variation in the ψ's associated with the first five principle components, among those series with root mean squared ψ's at least .05, broken down by category and method; that is, Table 9 breaks down Table 5 by category.

Generally speaking, the series fall into three groups. For the major measures of real economic activity, the DFM-5 method has the lowest, or nearly the lowest, relative RMSE among the various methods. For these series, RMSE improvements using the DFM-5 model are substantial relative to the AR(4) model, especially at longer horizons. Series in this group, for which DFM-5 typically has median RMSEs less than the shrinkage methods, include GDP components at all horizons, IP components at all horizons, employment series at all horizons. Other series in this group include some interest rates, money (at least at the 4-quarter horizon), inventories (at the 2- and 4-quarter horizons), and arguably unemployment rates, for which the shrinkage methods provide only very small improvements relative to DFM-5 forecasts at all horizons. Moreover, for series in this group, typically the fraction of the mean-square weight placed by the shrinkage methods on the first five principle components is large. Thus, for these series, the shrinkage methods are essentially approximating the DFM-5 model and the DFM-5 works as well or better than the shrinkage approximations to it.

Figure 1 presents estimated shrinkage functions for a series in this first group, total employment (results are shown for the 2-quarter horizon). The upper panel presents the estimated shrinkage functions, and the lower panel presents the weight placed by the various shrinkage functions on each of the 100 ordered principle components. At $h = 2$, the DFM-5 RMSE is .846, slightly less than the Logit RMSE of .860, and both RMSEs indicate a substantial improvement over the AR(4). The estimated shrinkage functions are broadly similar, in all cases placing substantial weight only for t-statistics in excess of approximately 3.5, and the logit and bagging shrinkage functions are quite close. The estimated shrinkage functions end up placing nearly all their weight on the first few principle components, and only a few higher principle components receive weight exceeding 0.1. For total employment, the shrinkage methods support the DFM-5 restrictions, and relaxing those restrictions results in higher RMSEs.

The second group consists of series for which the shrinkage methods, in particular the logit model, has RMSEs that are both less than the DFM-5 RMSE and substantially less than one. For these series, the shrinkage methods place most of the mean-square weight on principle components beyond the first five, and doing so reduces the relative RMSE. Series in this group include real wages, housing, and interest rate spreads. For example, for wages, the median relative RMSE for the logit model is .919 at the 2-quarter horizon, while for the DFM-5 it is .999, and at $h = 2$ the logit model places only 5% of its mean-square weight on the first five principle components. For these series, the principle components can be used to produce lower RMSEs, but shrinkage methods are more effective at doing so than the DFM-5 model.

Figure 2 presents estimated shrinkage functions and weights for a series in this second group, the spread between the 10-year and 90-day Treasury rates. In contrast to Figure 1, the estimated shrinkage functions place weight on principle components with small, even zero, t-statistics. The effect is that many principle components enter the forecast function, with most having substantial shrinkage, in the range of .1 - .3. Including all these principle components with substantial shrinkage reduces the RMSE from 0.947 for DFM-5 to 0.900 for the logit model.

The final group consists of series for which the principle components do not provide meaningful reductions in RMSE, using either the DFM-5 or shrinkage models.

Series in this group include exchange rates, stock returns, and consumer expectations. Figure 3 presents estimated shrinkage functions and weights for a series in this third group, the percentage change in the S&P 500 Index. The shrinkage function and estimated weights look broadly similar to those in Figure 2, with most principle component entering the forecast with a large amount of shrinkage (0.1-0.2). This shrinkage improves upon unweighted OLS estimates, reducing the RMSE from 1.071 to 0.984 for the logit, but (as one would expect for this series) does not substantially improve over the AR(4). It should be noted that for series in this group, the cross-validated objective functions are often very flat so the shrinkage parameters are imprecisely estimated and in some cases the estimated shrinkage functions are quite different from each other.

4.4 Additional results and sensitivity analysis

In addition to the four shrinkage models listed in Section 4, we estimated by cross-validation a logit model with a quadratic term to obtain a more flexible (but low dimensional) parametric specification. The shrinkage function is for the quadratic logit model is,

$$\psi^{logit\text{-}q}(u) = \frac{\exp(\beta_0 + \beta_1|u| + \beta_2 u^2)}{1+\exp(\beta_0 + \beta_1|u| + \beta_2 u^2)} \quad . \tag{16}$$

This model fits marginally better than the linear logit model (14), which is to be expected since it has one more parameter, however the RMSE improvement obtained using this model was slight.

As another sensitivity check, we repeated the analysis using Newey-West standard errors (with a window width of h+1), instead of the homoskedasticity-only OLS standard errors used above, to compute the cross-validated t-statistics that appear as arguments of the shrinkage functions, including reestimating the shrinkage parameters using the Newey-West t-statistics. In principle this could change the results for multiple step ahead forecasts, for which the errors will have (at least) a MA(h–1) autocorrelation structure. However, using Newey-West standard errors yielded no substantial changes in the findings discussed above.

5. Discussion

Three points should be borne in mind when interpreting these results. First, we have focused on whether the DFM provides a good framework for macro forecasting. This focus is related to, but different than, asking whether the DFM with a small number of factors provides a good contemporaneous or retrospective fit to many macro time series; for a discussion of this latter issue, see Giannone, Reichlin, and Sala (2004) and Watson (2004). Second, the DFM forecasting method used here was chosen so that it is nested within the shrinkage function framework (2). To the extent that other DFM forecasting methods, such as iterated forecasts based on a high-dimensional state space representation of the DFM (e.g. Doz, Giannone, and Reichlin (2006)), improve upon the first-five principle components forecasts used here, the results here understate the general forecasting applicability of the DFM. Third, these results are full-sample in the sense that the cross-validated RMSEs are estimated using all data outside the range of the observation left out at each cross-validation step. Thus this analysis focuses on average predictive content over the full 1960-2008 sample and abstracts from known time variation in the predictive regressions; see for example Stock and Watson (2002c, 2009) and D'Agostino, Giannone, and Surico (2006) for documentation of decreasing predictability over this period.

The facts that some of these shrinkage methods have an interpretation as an empirical Bayes method and that we have considered some relatively flexible functional forms gives us some confidence that it will be difficult to improve systematically upon these forecasts using forecasts that are time-invariant linear functions of the principle components of large macro data sets like the one considered here. This suggests that further forecast improvements over those presented here will need to come from models with nonlinearities and/or time variation. An important next step in this research program is incorporating those extensions into high-dimensional forecasting systems.

References

Bai, J., and S. Ng (2002), “Determining the Number of Factors in Approximate Factor Models,” *Econometrica*, 70, 191-221.

Bai, J., and S. Ng (2006), “Confidence Intervals for Diffusion Index Forecasts and Inference for Factor-Augmented Regressions,” *Econometrica*, 74,1133-1150.

Bai, J., and S. Ng (2007), “Boosting Diffusion Indexes,” manuscript, Columbia University.

Bai, J., and S. Ng (2008), “Large Dimensional Factor Models,” *Foundations and Trends in Econometrics,* 3 (2), 89-163.

Boivin, J., and M.P. Giannoni (2006), “DSGE Models in a Data-Rich Environment,” NBER WP12772.

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

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

Carriero, A., G. Kapetanios, and M. Marcellino (2009), “Forecasting Large Datasets with Bayesian Reduced Rank Multivariate Models”, EUI.

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

Clyde, M. (1999b), “Comment on ‘Bayesian model averaging: a tutorial’,” *Statistical Science*, 14, 401-404.

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

D’Agostino, A., D. Giannone, and P. Surico (2006), “(Un)Predictability and Macroeconomic Stability,” ECB Working Paper 605.

De Mol, C., D. Giannone, and Reichlin (2008), “Forecasting a Large Number of Predictors: Is Bayesian Regression a Valid Alternative to Principal Components?,” *Journal of Econometrics*, Vol. 146, No. 2, 318-328.

Doz, C., D. Giannone, and L. Reichlin (2006), "A Quasi Maximum Likelihood Approach for Large Approximate Dynamic Factor Models," ECB Working Paper 674.

Eickmeier, S., C. Ziegler (2008), "How successful are dynamic factor models at forecasting output and inflation? A meta-analytic approach," *Journal of Forecasting*, 27(3), 237-265.

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.

Forni, M., M. Hallin, M. Lippi, and L. Reichlin (2004), "The Generalized Factor Model: Consistency and Rates," *Journal of Econometrics*, 119, 231-255.

Geweke, J. (1977), "The Dynamic Factor Analysis of Economic Time Series," in *Latent Variables in Socio-Economic Models*, ed. by D.J. Aigner and A.S. Goldberger, Amsterdam: North-Holland.

Giannone, D., L. Reichlin, and L. Sala (2004), "Monetary Policy in Real Time," *NBER Macroeconomics Annual*, 2004, 161-200.

Inoue, Atsushi, and Lutz Kilian (2008), "How Useful Is Bagging in Forecasting Economic Time Series? A Case Study of U.S. CPI Inflation," *Journal of the American Statistical Association*, 103, 511 – 522

Knox, T., J.H. Stock, and M.W. Watson (2004), "Empirical Bayes Regression with Many Regressors," manuscript, Harvard University.

Koop, G., and Potter, S. (2004), "Forecasting in Dynamic Factor Models Using Bayesian Model Averaging," *The Econometrics Journal*, 7, 550-565.

Lee, T.-H. and Y. Yang (2006), "Bagging Binary and Quantile Predictors for Time Series," *Journal of Econometrics*, Volume 135, Issues 1-2, 465-497.

Maritz, J.S. and T. Lwin (1989), *Empirical Bayes Methods, Second Edition*. London: Chapman and Hall.

Sargent, T.J. (1989), "Two Models of Measurements and the Investment Accelerator," *Journal of Political Economy*, 97, 251-287.

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

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.

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

Stock, J.H. and M.W. Watson (2002c), "Has the Business Cycle Changed and Why?" *NBER Macroeconomics Annual* 2002, 159–218 (with M.W. Watson).

Stock, J.H. and M.W. Watson (2009), "Phillips Curve Inflation Forecasts," forthcoming in *Understanding Inflation and the Implications for Monetary Policy: A Phillips Curve Retrospective*, Federal Reserve Bank of Boston.

Watson, M. W. (2004), "Discussion of `Monetary Policy in Real Time'," *NBER Macroeconomics Annual*, 2004, 216-221.

Appendix A

Proofs of Results in Section 2

Proof of (8). First use (7) to write

$$\psi^{NB}(u) = 1 + \frac{1}{u}\frac{d}{dx}\ln\left\{\int \phi(x-\tau)dG_\tau(\tau)\right\}\Big|_{x=u}$$

$$= 1 - \frac{\int (u-\tau)\phi(u-\tau)g_\tau(\tau)d\tau}{u\int \phi(u-\tau)g_\tau(\tau)d\tau} = \frac{\int \tau\phi(u-\tau)g_\tau(\tau)d\tau}{u\int \phi(u-\tau)g_\tau(\tau)d\tau} \qquad (17)$$

The symmetry of $\psi^{NB}(u)$ follows from the final expression, the symmetry of the normal distribution, and the assumed symmetry of the prior density g_τ..

To show that $0 \le \psi^{NB}(u) \le 1$ is bounded, first note that, because of the symmetry of $\psi^{NB}(u)$, it suffices to consider $u \ge 0$. Also note that, for functions h and f where $h(x) > h(-x)$ for all $x \ge 0$, $f(x)$ is odd, and $f(x) \ge 0$ for $x \ge 0$, then $\int_{-\infty}^{\infty} f(x)h(x)dx \ge 0$. The result $\psi^{NB}(u) \ge 0$ follows from the final expression in (17) and this inequality by setting $f(x) = xg_\tau(x)$ and $h(x) = \phi(u-x)$ (note that, for $u \ge 0$, $\phi(u-x) \ge \phi(u+x)$). The result $\psi^{NB}(u) \le 1$ follows by using the inequality with $f(x) = x\phi(x)$ and $h(x) = g(u-x)$ (noting that $\phi(u-x) \ge \phi(u+x)$ for $u \ge 0$ because g is symmetric and unimodal) to show that $\int (u-\tau)\phi(u-\tau)g_\tau(\tau)d\tau \ge 0$.

Proof of Theorem 1: Let $\psi = \psi^{NB}$, and recall the notation $\tau_i = \sqrt{T}\delta_i/\sigma$, $\hat{\tau}_i = \sqrt{T}\hat{\delta}_i/\sigma$, $t_i = \sqrt{T}\hat{\delta}_i/s_e$, $\kappa = s_e/\hat{\sigma}$, and $\zeta = \hat{\sigma}^2/\sigma^2$. Let $\hat{t}_i = \kappa t_i$, so that $\hat{\tau}_i = \hat{t}_i\zeta^{1/2}$.

Note that $\tilde{Y}^{NB}_{T+1/T} - \sum_{i=1}^{n}\psi(\kappa t_i)\hat{\delta}_i P_{iT} = \sum_{i=1}^{n}\left[E^M\psi(\hat{\tau}_i) - \psi(\kappa t_i)\right]\hat{\delta}_i P_{iT} = \sum_{i=1}^{n}\rho_i P_{iT}$, where

$$\begin{aligned}\rho_i &= \left[E^M\psi(\hat{\tau}_i) - \psi(\kappa t_i)\right]\hat{\delta}_i \\ &= \left[E^M\psi^{NB}(\hat{t}_i\zeta^{1/2}) - \psi^{NB}(\hat{t}_i)\right]\hat{\delta}_i \\ &= \frac{1}{8}\hat{\delta}_i E^M\left[\left(\tilde{t}_i^2\psi^{NB''}(\tilde{t}_i) - \tilde{t}_i\,\psi^{NB'}(\tilde{t}_i)\right)\tilde{\zeta}^{-2}(\zeta-1)^2\right]\end{aligned} \qquad (18)$$

where the third equality follows from a second order mean value expansion of $\psi^{NB}(t\zeta^{1/2})$ around $\zeta = 1$, where $\psi^{NB\prime}$ and $\psi^{NB\prime\prime}$ are the first and second derivatives of ψ^{NB} (which exists by (A7)), $\tilde{\zeta} \in [1, \zeta]$, $\tilde{t}_i = \hat{t}_i\, \tilde{\zeta}^{1/2}$, and the first term in the mean value expansion vanishes because $E^M(\zeta - 1) = 0$. The theorem follows by showing that $\sum_i \rho_i P_{iT}$ converges to zero in mean square.

To show this, note that

$$\left(\sum_{i=1}^{n} \rho_i P_{iT}\right)^2 = \left(\frac{1}{8}\sum_{i=1}^{n} P_{iT}\hat{\delta}_i E^M\left[\left(\tilde{t}_i^2 \psi^{NB\prime\prime}(\tilde{t}_i) - \tilde{t}_i\, \psi^{NB\prime}(\tilde{t}_i)\right)\tilde{\zeta}^{-2}(\zeta-1)^2\right]\right)^2$$

$$\leq \frac{1}{64}\left(\sum_{i=1}^{n} P_{iT}^2\hat{\delta}_i^2\right)\sum_{i=1}^{n}\left\{E^M\left[\left(\tilde{t}_i^2 \psi^{NB\prime\prime}(\tilde{t}_i) - \tilde{t}_i\, \psi^{NB\prime}(\tilde{t}_i)\right)\tilde{\zeta}^{-2}(\zeta-1)^2\right]\right\}^2.$$

Now, $\sum_{i=1}^{n} P_{iT}^2\hat{\delta}_i^2 \leq P_{\max}^2 \sum_{i=1}^{n}\hat{\delta}_i^2 \leq P_{\max}^2\, T^{-1}\sum_{t=1}^{T} Y_t^2$, where the final inequality follows because $T\sum_{i=1}^{n}\hat{\delta}_i^2$ is the regression sum of squares from the regression of Y onto P. Also $\left\{E^M\left[\left(\tilde{t}_i^2 \psi^{NB\prime\prime}(\tilde{t}_i) - \tilde{t}_i\, \psi^{NB\prime}(\tilde{t}_i)\right)\tilde{\zeta}^{-2}(\zeta-1)^2\right]\right\}^2 \leq 4M^2\left\{E^M\left[\tilde{\zeta}^{-2}(\zeta-1)^2\right]\right\}^2$ (by A6). Thus, repeated application of the Cauchy-Schwarz inequality yields

$$E\left(\sum_{i=1}^{n} \rho_i P_{iT}\right)^2 \leq \frac{1}{16}P_{\max}^2 M^2 n E\left((T^{-1}\sum_{t=1}^{T} Y_t^2)\left\{E^M\left[\tilde{\zeta}^{-2}(\zeta-1)^2\right]\right\}^2\right)$$

$$\leq \frac{1}{16}P_{\max}^2 M^2 n\sqrt{E\left(T^{-1}\sum_{t=1}^{T} Y_t^2\right)^2}\sqrt{E\left\{E^M\left[\tilde{\zeta}^{-2}(\zeta-1)^2\right]\right\}^4}$$

$$\leq \frac{1}{16}P_{\max}^2 M^2 n\sqrt{E\left(T^{-1}\sum_{t=1}^{T} Y_t^2\right)^2}\sqrt{E\left\{\sqrt{E^M\tilde{\zeta}^{-4}}\sqrt{E^M(\zeta-1)^4}\right\}^4}$$

$$= \frac{1}{16}P_{\max}^2 M^2 n\sqrt{E\left(T^{-1}\sum_{t=1}^{T} Y_t^2\right)^2}\sqrt{E\left\{\left(E^M\tilde{\zeta}^{-4}\right)^2\left(E^M(\zeta-1)^4\right)^2\right\}}$$

$$\leq \frac{1}{16} P_{\max}^2 M^2 n \sqrt{E\left(T^{-1}\sum_{t=1}^{T} Y_t^2\right)^2} \sqrt{\sqrt{E\left(E^M \tilde{\zeta}^{-4}\right)^4} \sqrt{E\left(E^M (\zeta-1)^4\right)^4}}$$

$$\leq \frac{1}{16} P_{\max}^2 M^2 n \sqrt{E\left(T^{-1}\sum_{t=1}^{T} Y_t^2\right)^2} \sqrt{\sqrt{E\left(1+E^M \zeta^{-4}\right)^4} \sqrt{E\left(E^M (\zeta-1)^4\right)^4}} \sim o(1) \qquad (19)$$

Where the final inequality follows from $E^M(\tilde{\zeta}^{-4}) \leq E^M [\max(1,\zeta^{-4})] \leq 1 + E^M(\zeta^{-4})$, and the rate follows from (A2), (A4) and (A5) .

Proof of Theorem 2: The proof of Theorem 2 is facilitated by the following lemma:

***Lemma*:** Let $y \sim N(\mu,1)$ and let D be a random variable distributed independently of y. Then $\text{var}\left[y\times 1(|y|> D)\right] \leq 1 + \mu^2$.

Proof: $\text{var}\left[y\times 1(|y|> D)\right] \leq E\left[y\times 1(|y|> D)\right]^2 = E\left[y^2\times 1(|y|> D)\right] =$

$E\left\{E\left[y^2\times 1(|y|> D)\middle|D\right]\right\} \leq E\left\{E\left[y^2\middle|D\right]\right\} = Ey^2 = 1 + \mu^2$.

As discussed in Section 2.5, bagging is implemented using the parametric bootstrap based on the exogeneity-normality assumption (M1). Let the superscript * denote bootstrap realizations and let E^* denote expectations taken with respect to the bootstrap distribution conditional on the observed data (Y, P). Each parametric bootstrap realization draws T observations such that $P^{*\prime}P^*/T = I$ and $Y^*|P^* \sim N(P^*\hat{\delta},\ s_e^2 I)$. Let $\hat{\delta}_{ij}^*$ denote the j^{th} bootstrap draw of the OLS estimator of δ_i and let $s_{e,j}^{2*}$ denote the j^{th} bootstrap draw of the OLS estimator of σ^2, let $\xi^* = s_{e,j}^{2*}/s_e^2$, and let $t_{ij}^* = \sqrt{T}\ \hat{\delta}_{ij}^*/s_{e,j}^*$. The j^{th} bootstrap realization of the pretest estimator is $1(|t_{ij}^*|>c)\,\hat{\delta}_{ij}^*$. The bagging estimator is

$$\tilde{\delta}_i^{BG} = \frac{1}{B}\sum_{j=1}^{B} 1(|t_{ij}^*|> c)\hat{\delta}_{ij}^*, \qquad (20)$$

where B is the number of bootstrap draws.

By construction, under the * distribution, $\hat{\delta}_{ij}^* \sim$ i.i.d. N($\hat{\delta}_i, s_e^2/T$) so $\sqrt{T}\hat{\delta}_{ij}^*/s_e \sim$ i.i.d. N(t_i, 1), $\xi^* \sim$ i.i.d. $\chi^2_{T-n}/T-n$, and $\hat{\delta}_{ij}^*$ and ξ^* are independently distributed. It is useful to define $z_{ij}^* = \sqrt{T}\hat{\delta}_{ij}^*/s_e - t_i$, where $z_{ij}^* \sim$ i.i.d. N(0,1).

With this notation, $\tilde{Y}_{T+1/T}^{BG} - \hat{Y}_{T+1/T}^{BG} = \sum_{i=1}^{n}\rho_i P_{it}$, where $\rho_i = \tilde{\delta}_i^{BG} - \psi^{BG}(t_t)\hat{\delta}_i$. Thus $E(\tilde{Y}_{T+1/T}^{BG} - \hat{Y}_{T+1/T}^{BG})^2 = E(\sum_{i=1}^{n}\rho_i P_{iT})^2 \le P_{\max}^2 n[\sum_{i=1}^{n}E(\rho_i^2)] \le P_{\max}^2 n^2 \max_i E(\rho_i^2)$. The rest of the proof entails showing that $\max_i E(\rho_i^2) \sim o(n^{-2})$. Write $\rho_i = \rho_{1i} + \rho_{2i}$, where $\rho_{1i} = \tilde{\delta}_i^{BG} - E^*\tilde{\delta}_i^{BG}$ and $\rho_{2i} = E^*\tilde{\delta}_i^{BG} - \psi^{BG}(t_t)\hat{\delta}_i$, and note from Minkowski's theorem that $E(\rho_i^2) \le \left(\sqrt{E(\rho_{1i}^2)} + \sqrt{E(\rho_{2i)}^2}\right)^2$. The proof follows from showing $\max_i E(\rho_{1i}^2) \sim o(n^{-2})$ and $\max_i E(\rho_{2i}^2) \sim o(n^{-2})$.

<u>$E(\rho_{1i}^2)$</u> :

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

$$
\begin{aligned}
\mathrm{var}^*(\rho_{1i}) &= \mathrm{var}^*(\tilde{\delta}_i^{BG} - E^*\tilde{\delta}_i^{BG}) \\
&= \mathrm{var}^*\left\{\frac{1}{B}\sum_{j=1}^{B}\left[1(|t_{ij}^*|>c)\hat{\delta}_{ij}^* - E^*1(|t_{ij}^*|>c)\hat{\delta}_{ij}^*\right]\right\} \\
&= \frac{1}{B}\mathrm{var}^*\left[1(|t_{ij}^*|>c)\hat{\delta}_{ij}^*\right] \\
&= \frac{s_e^2}{T}\frac{1}{B}\mathrm{var}^*\left[1(|t_{ij}^*|>c)\frac{\sqrt{T}\hat{\delta}_{ij}^*}{s_e}\right] \\
&= \frac{s_e^2}{T}\frac{1}{B}\mathrm{var}^*\left[1\left(|t_i + z_{ij}^*|>c\sqrt{\xi_j^*}\right)\left(t_i + z_{ij}^*\right)\right] \\
&\le \frac{1}{T-n}\frac{1}{B}\left(T^{-1}\sum Y_t^2\right)\left(1+t_i^2\right),
\end{aligned}
$$

where the second equality follows by substituting (20), the third equality follows because the bootstrap draws are i.i.d., the fourth equality follows from multiplying and dividing by s_e^2/T, the fifth equality uses the notation introduces above, and the inequality follows from $(T-n)s_e^2 \le \sum_{t=1}^{T} Y_t^2$ and the lemma. Thus

$$\max_i E(\rho_{1i}^2) \le \frac{1}{(T-n)B}\left[E\left(T^{-1}\sum Y_t^2\right)^2\right]^{1/2} \max_i \left[E\left(1+t_i^2\right)^2\right]^{1/2} \sim o(n^{-2}),$$

where the rate follows from (A2), (A3), (A7), (A8), and the lemma.

$\underline{E(\rho_{2i}^2)}$:

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

Now $E^*\left[1\left(|\, t + z_{ij}^* | > d\right)\left(t + z_{ij}^*\right)\right] = \int\limits_{|t+z^*|>d} (t+z^*)\phi(z^*)dz^* = \psi^{BG}(t,d)t$, where $\psi^{BG}(t,d) \equiv 1 - \Phi(t+d) + \Phi(t-d) + t^{-1}[\phi(t-d) - \phi(t+d)]$ (cf. Bühlmann and Yu (2002)). Thus

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

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

$$\rho_{2i} = \frac{s_e}{\sqrt{T}} E^* \left\{ \frac{1}{8}\left[t_i \psi^{BG''}(t_i, \tilde{c})\tilde{c}^2 - t_i \psi^{BG'}(t_i, \tilde{c})\tilde{c} \right] \tilde{\xi}^{-2} (\xi_j^* - 1)^2 \right\} \quad (21)$$

where $\tilde{c} = c\sqrt{\tilde{\xi}}$, $\tilde{\xi} \in [1, \xi_j^*]$, and the first term in the mean value expansion vanishes because $E^*(\xi_j^*) = 1$. Thus

$$\begin{aligned} |\rho_{2i}| &= \frac{s_e}{\sqrt{T}} E^* \left\{ \frac{1}{8}\left[t_i \psi^{BG''}(t_i, \tilde{c})\tilde{c}^2 - t_i \psi^{BG'}(t_i, \tilde{c})\tilde{c} \right] \tilde{\xi}^{-2} (\xi_j^* - 1)^2 \right\} \\ &\le \frac{s_e}{\sqrt{T}} \frac{1}{8} \sup_u \left| t_i \psi^{BG''}(t_i, u)u^2 - t_i \psi^{BG'}(t_i, u)u \right| E^* \left[\tilde{\xi}^{-2} (\xi_j^* - 1)^2 \right] \\ &\le \frac{s_e}{\sqrt{T}} \frac{1}{8} \sup_u \left| t_i \psi^{BG''}(t_i, u)u^2 - t_i \psi^{BG'}(t_i, u)u \right| \sqrt{E^*(\xi_j^* - 1)^4} \sqrt{E^* \tilde{\xi}^{-4}}\,. \end{aligned} \quad (22)$$

Note, $E^*(\xi_j^* - 1)^4$ is the fourth central moment of a $\chi^2_{T-n} / T - n$ random variable, so

$$E^*(\xi_j^* - 1)^4 = 12(T-n)(T-n+4)/(T-n)^4 = a_{T-n} \quad (23)$$

where the final equality defines a_{T-n}. Next, because $\tilde{\xi} \in [1, \xi_j^*]$ and because the fourth moment of the reciprocal of a χ_r^2 random variable exists for $r > 8$ and is $[(r-2)(r-4)(r-6)(r-8)]^{-1}$, for $T - n \ge 8$ we have that

$$E^* \tilde{\xi}_j^{*-4} \le 1 + E^* \xi_j^{*-4} = 1 + \frac{(T-n)^4}{(T-n-2)(T-n-4)(T-n-6)(T-n-8)} = b_{T-n} \quad (24)$$

where the final equality defines b_{T-n}

Now turn to the sup term in (22). Direct evaluation of the derivatives using the definition of ψ^{BG} show that $t\psi^{BG\prime}(t,u)u = u^2[\phi(t+u) - \phi(t-u)]$ and $t\psi^{BG\prime\prime}(t,u)u^2 = u^2[\phi(t+u) - \phi(t-u)] - u^3[(t+u)\phi(t+u) + (t-u)\phi(t-u)]$. Thus

$$|t\psi^{BG\prime}(t,u)u| \le 2\sup_u u^2 \phi(t+u)$$

$$\leq 2\sup_v (v-t)^2\phi(v)$$
$$\leq 2[(\sup_v v^2\phi(v)) + 2t\,\sup_v|v\phi(v)| + t^2\sup_v\phi(v)]$$
$$= 2(h_2 + 2h_1 t + h_2 t^2) \tag{25}$$

where $h_m = m^{m/2}e^{-m/2}/\sqrt{2\pi}$. Similar calculations provide a bound on $|t\psi^{BG}{}''(t,u)u^2|$ which, combined with the bound in (25), yields

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

where the final equality uses $h_i < h_m$ for $i < m$ and $m > 1$. Substituting this bound, (23), and (24) into (22), squaring, taking expectations, and collecting terms yields

$$E(\rho_{2i}^2) \leq \frac{14^2}{64}a_{T-n}b_{T-n}h_4^2 E\left[\frac{s_e^2}{T}\left(\sum_{m=0}^{3}|t_i|^m\right)^2\right]$$
$$\leq \frac{14^2}{64}a_{T-n}b_{T-n}h_4^2\frac{1}{T-n}E\left[\left(T^{-1}\sum_{t=1}^{T}Y_t^2\right)\left(\sum_{m=0}^{3}|t_i|^m\right)^2\right]$$
$$\leq \frac{14^2}{64}a_{T-n}b_{T-n}h_4^2\frac{1}{T-n}\sqrt{E\left(T^{-1}\sum_{t=1}^{T}Y_t^2\right)^2}\sqrt{E\left(\sum_{m=0}^{3}|t_i|^m\right)^4}$$
$$\sim o(n^{-3}),$$

where the second inequality uses $(T-n)s_e^2 \leq \sum Y_t^2$, the third uses Cauchy-Schwarz, and the rate uses $a_{T-n} \sim o[(T-n)^{-2}]$, $b_{T-n} \sim O(1)$, (A2)-(A3), and (A8).

Appendix B

Data Sources and Transformations

Table B.1 lists all the series in the data set, the series mnemonic (label) in the source database, the transformation applied to the series (T, described in Table B.2), whether the series is used to compute the principle components (E; 1 = used), the category grouping of the series (C), and a brief data description. All series are from the Global Insight (formerly DRI) Basic Economics Database, except those that include TCB (which are from the Conference Board's Indicators Database) or AC (author's calculation).

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

Table B.1
Series Descriptions

Name	Label	T	E	C	Description
RGDP	GDP251	5	0	1	Real gross domestic product, quantity index (2000=100) , saar
Cons	GDP252	5	0	1	Real personal consumption expenditures, quantity index (2000=100) , saar
Cons-Dur	GDP253	5	1	1	Real personal consumption expenditures - durable goods , quantity index (2000=
Cons-NonDur	GDP254	5	1	1	Real personal consumption expenditures - nondurable goods, quantity index (200
Cons-Serv	GDP255	5	1	1	Real personal consumption expenditures - services, quantity index (2000=100) ,
GPDInv	GDP256	5	0	1	Real gross private domestic investment, quantity index (2000=100) , saar
FixedInv	GDP257	5	0	1	Real gross private domestic investment - fixed investment, quantity index (200
NonResInv	GDP258	5	0	1	Real gross private domestic investment - nonresidential , quantity index (2000
NonResInv-struct	GDP259	5	1	1	Real gross private domestic investment - nonresidential - structures, quantity
NonResInv-Bequip	GDP260	5	1	1	Real gross private domestic investment - nonresidential - equipment & software
Res.Inv	GDP261	5	1	1	Real gross private domestic investment - residential, quantity index (2000=100
Exports	GDP263	5	1	1	Real exports, quantity index (2000=100) , saar
Imports	GDP264	5	1	1	Real imports, quantity index (2000=100) , saar
Gov	GDP265	5	0	1	Real government consumption expenditures & gross investment, quantity index (2
Gov Fed	GDP266	5	1	1	Real government consumption expenditures & gross investment - federal, quantit
Gov State/Loc	GDP267	5	1	1	Real government consumption expenditures & gross investment - state & local, Q
IP: total	IPS10	5	0	2	Industrial production index - total index
IP: products	IPS11	5	0	2	Industrial production index - products, total
IP: final prod	IPS299	5	0	2	Industrial production index - final products
IP: cons gds	IPS12	5	0	2	Industrial production index - consumer goods
IP: cons dble	IPS13	5	1	2	Industrial production index - durable consumer goods
iIP:cons nondble	IPS18	5	1	2	Industrial production index - nondurable consumer goods
IP:bus eqpt	IPS25	5	1	2	Industrial production index - business equipment
IP: matls	IPS32	5	0	2	Industrial production index - materials
IP: dble mats	IPS34	5	1	2	Industrial production index - durable goods materials
IP:nondble mats	IPS38	5	1	2	Industrial production index - nondurable goods materials
IP: mfg	IPS43	5	1	2	Industrial production index - manufacturing (sic)
IP: fuels	IPS306	5	1	2	Industrial production index - fuels
NAPM prodn	PMP	1	1	2	NAPM production index (percent)
Capacity Util	UTL11	1	1	2	Capacity utilization - manufacturing (sic)
Emp: total	CES002	5	0	3	Employees, nonfarm - total private
Emp: gds prod	CES003	5	0	3	Employees, nonfarm - goods-producing
Emp: mining	CES006	5	1	3	Employees, nonfarm – mining
Emp: const	CES011	5	1	3	Employees, nonfarm – construction
Emp: mfg	CES015	5	0	3	Employees, nonfarm – mfg
Emp: dble gds	CES017	5	1	3	Employees, nonfarm - durable goods
Emp: nondbles	CES033	5	1	3	Employees, nonfarm - nondurable goods
Emp: services	CES046	5	1	3	Employees, nonfarm - service-providing

Emp: TTU	CES048	5	1	3	Employees, nonfarm - trade, transport, utilities
Emp: wholesale	CES049	5	1	3	Employees, nonfarm - wholesale trade
Emp: retail	CES053	5	1	3	Employees, nonfarm - retail trade
Emp: FIRE	CES088	5	1	3	Employees, nonfarm - financial activities
Emp: Govt	CES140	5	1	3	Employees, nonfarm – government
Help wanted indx	LHEL	2	1	3	Index of help-wanted advertising in newspapers (1967=100;sa)
Help wanted/emp	LHELX	2	1	3	Employment: ratio; help-wanted ads:no. Unemployed clf
Emp CPS total	LHEM	5	0	3	Civilian labor force: employed, total (thous.,sa)
Emp CPS nonag	LHNAG	5	1	3	Civilian labor force: employed, nonagric.industries (thous.,sa)
Emp. Hours	LBMNU	5	1	3	Hours of all persons: nonfarm business sec (1982=100,sa)
Avg hrs	CES151	1	1	3	Avg wkly hours, prod wrkrs, nonfarm - goods-producing
Overtime: mfg	CES155	2	1	3	Avg wkly overtime hours, prod wrkrs, nonfarm - mfg
U: all	LHUR	2	1	4	Unemployment rate: all workers, 16 years & over (%,sa)
U: mean duration	LHU680	2	1	4	Unemploy.by duration: average(mean)duration in weeks (sa)
U < 5 wks	LHU5	5	1	4	Unemploy.by duration: persons unempl.less than 5 wks (thous.,sa)
U 5-14 wks	LHU14	5	1	4	Unemploy.by duration: persons unempl.5 to 14 wks (thous.,sa)
U 15+ wks	LHU15	5	1	4	Unemploy.by duration: persons unempl.15 wks + (thous.,sa)
U 15-26 wks	LHU26	5	1	4	Unemploy.by duration: persons unempl.15 to 26 wks (thous.,sa)
U 27+ wks	LHU27	5	1	4	Unemploy.by duration: persons unempl.27 wks + (thous,sa)
HStarts: Total	HSFR	4	0	5	Housing starts:nonfarm(1947-58);total farm&nonfarm(1959-)(thous.,sa
HStarts: authorizations	HSBR	4	0	5	Housing authorized: total new priv housing units (thous.,saar)
HStarts: ne	HSNE	4	1	5	Housing starts:northeast (thous.u.)s.a.
HStarts: MW	HSMW	4	1	5	Housing starts:midwest(thous.u.)s.a.
HStarts: South	HSSOU	4	1	5	Housing starts:south (thous.u.)s.a.
HStarts: West	HSWST	4	1	5	Housing starts:west (thous.u.)s.a.
PMI	PMI	1	1	6	Purchasing managers' index (sa)
NAPM new ordrs	PMNO	1	1	6	NAPM new orders index (percent)
NAPM vendor del	PMDEL	1	1	6	Napm vendor deliveries index (percent)
NAPM Invent	PMNV	1	1	6	Napm inventories index (percent)
Orders (ConsGoods)	MOCMQ	5	1	6	New orders (net) - consumer goods & materials, 1996 dollars (bci)
Orders (NDCapGoods)	MSONDQ	5	1	6	New orders, nondefense capital goods, in 1996 dollars (bci)
PGDP	GDP272A	6	0	7	Gross domestic product Price Index
PCED	GDP273A	6	0	7	Personal consumption expenditures Price Index
CPI-ALL	CPIAUCSL	6	0	7	Cpi all items (sa) fred
PCED-Core	PCEPILFE	6	0	7	PCE Price Index Less Food and Energy (SA) Fred
CPI-Core	CPILFESL	6	0	7	CPI Less Food and Energy (SA) Fred
PCED-DUR	GDP274A	6	0	7	Durable goods Price Index
PCED-DUR-MOTORVEH	GDP274_1	6	1	7	Motor vehicles and parts Price Index
PCED-DUR-HHEQUIP	GDP274_2	6	1	7	Furniture and household equipment Price Index
PCED-DUR-OTH	GDP274_3	6	1	7	Other price index
PCED-NDUR	GDP275A	6	0	7	Nondurable goods Price Index
PCED-NDUR-FOOD	GDP275_1	6	1	7	Food price index
PCED-NDUR-CLTH	GDP275_2	6	1	7	Clothing and shoes Price Index
PCED-NDUR-ENERGY	GDP275_3	6	1	7	Gasoline, fuel oil, and other energy goods Price Index
PCED-NDUR-OTH	GDP275_4	6	1	7	Other price index
PCED-SERV	GDP276A	6	0	7	Services price index
PCED-SERV-HOUS	GDP276_1	6	1	7	Housing price index
PCED-SERV-HOUSOP	GDP276_2	6	0	7	Household operation Price Index
PCED-SERV-H0-ELGAS	GDP276_3	6	1	7	Electricity and gas Price Index
PCED-SERV-HO-OTH	GDP276_4	6	1	7	Other household operation Price Index
PCED-SERV-TRAN	GDP276_5	6	1	7	Transportation price index
PCED-SERV-MED	GDP276_6	6	1	7	Medical care Price Index
PCED-SERV-REC	GDP276_7	6	1	7	Recreation price index
PCED-SERV-OTH	GDP276_8	6	1	7	Other price index
PGPDI	GDP277A	6	0	7	Gross private domestic investment Price Index
PFI	GDP278A	6	0	7	Fixed investment Price Index
PFI-NRES	GDP279A	6	0	7	Nonresidential price index
PFI-NRES-STR Price Index	GDP280A	6	1	7	Structures
PFI-NRES-EQP	GDP281A	6	1	7	Equipment and software Price Index
PFI-RES	GDP282A	6	1	7	Residential price index
PEXP	GDP284A	6	1	7	Exports price index
PIMP	GDP285A	6	1	7	Imports price index
PGOV	GDP286A	6	0	7	Government consumption expenditures and gross investment Price Index
PGOV-FED	GDP287A	6	1	7	Federal price index
PGOV-SL	GDP288A	6	1	7	State and local Price Index
Com: spot price (real)	PSCCOMR	5	1	7	Real spot market price index:bls & crb: all commodities(1967=100) (psccom/pcepilfe)

OilPrice (Real)	PW561R	5	1	7	PPI crude (relative to core PCE) (pw561/pcepilfe)
NAPM com price	PMCP	1	1	7	Napm commodity prices index (percent)
Real AHE: goods	CES275R	5	0	8	Real avg hrly earnings, prod wrkrs, nonfarm - goods-producing (ces275/pi071)
Real AHE: const	CES277R	5	1	8	Real avg hrly earnings, prod wrkrs, nonfarm - construction (ces277/pi071)
Real AHE: mfg	CES278 R	5	1	8	Real avg hrly earnings, prod wrkrs, nonfarm - mfg (ces278/pi071)
Labor Prod	LBOUT	5	1	8	Output per hour all persons: business sec(1982=100,sa)
Real Comp/Hour	LBPUR7	5	1	8	Real compensation per hour,employees:nonfarm business(82=100,sa)
Unit Labor Cost	LBLCPU	5	1	8	Unit labor cost: nonfarm business sec (1982=100,sa)
FedFunds	FYFF	2	1	9	Interest rate: federal funds (effective) (% per annum,nsa)
3 mo T-bill	FYGM3	2	1	9	Interest rate: u.s.treasury bills,sec mkt,3-mo.(% per ann,nsa)
6 mo T-bill	FYGM6	2	0	9	Interest rate: u.s.treasury bills,sec mkt,6-mo.(% per ann,nsa)
1 yr T-bond	FYGT1	2	1	9	Interest rate: u.s.treasury const maturities,1-yr.(% per ann,nsa)
5 yr T-bond	FYGT5	2	0	9	Interest rate: u.s.treasury const maturities,5-yr.(% per ann,nsa)
10 yr T-bond	FYGT10	2	1	9	Interest rate: u.s.treasury const maturities,10-yr.(% per ann,nsa)
Aaabond	FYAAAC	2	0	9	Bond yield: moody's aaa corporate (% per annum)
Baa bond	FYBAAC	2	0	9	Bond yield: moody's baa corporate (% per annum)
fygm6-fygm3	SFYGM6	1	1	9	Fygm6-fygm3
fygt1-fygm3	SFYGT1	1	1	9	Fygt1-fygm3
fygt10-fygm3	SFYGT10	1	1	9	Fygt10-fygm3
FYAAAC-Fygt10	SFYAAAC	1	1	9	Fyaaac-fygt10
FYBAAC-Fygt10	SFYBAAC	1	1	9	Fybaac-fygt10
M1	FM1	6	1	10	Money stock: m1 (curr,trav.cks,dem dep,other ck'able dep)(bil$,sa)
MZM	MZMSL	6	1	10	Mzm (sa) frb st. Louis
M2	FM2	6	1	10	Money stock:m2(m1+o'nite rps,euro$,g/p&b/d mmmfs&sav&sm time dep(bil$,
MB	FMFBA	6	1	10	Monetary base, adj for reserve requirement changes(mil$,sa)
Reserves tot	FMRRA	6	1	10	Depository inst reserves:total,adj for reserve req chgs(mil$,sa)
BUSLOANS	BUSLOANS	6	1	10	Commercial and industrial loans at all commercial Banks (FRED) Billions $ (SA)
Cons credit	CCINRV	6	1	10	Consumer credit outstanding - nonrevolving(g19)
Ex rate: avg	EXRUS	5	1	11	United states;effective exchange rate(merm)(index no.)
Ex rate: Switz	EXRSW	5	1	11	Foreign exchange rate: switzerland (swiss franc per u.s.$)
Ex rate: Japan	EXRJAN	5	1	11	Foreign exchange rate: japan (yen per u.s.$)
Ex rate: UK	EXRUK	5	1	11	Foreign exchange rate: united kingdom (cents per pound)
EX rate: Canada	EXRCAN	5	1	11	Foreign exchange rate: canada (canadian $ per u.s.$)
S&P 500	FSPCOM	5	1	12	S&p's common stock price index: composite (1941-43=10)
S&P: indust	FSPIN	5	1	12	S&p's common stock price index: industrials (1941-43=10)
S&P div yield	FSDXP	2	1	12	S&p's composite common stock: dividend yield (% per annum)
S&P PE ratio	FSPXE	2	1	12	S&p's composite common stock: price-earnings ratio (%,nsa)
DJIA	FSDJ	5	1	12	Common stock prices: dow jones industrial average
Consumer expect	HHSNTN	2	1	13	U. Of mich. Index of consumer expectations(bcd-83)

Table B.2

Transformations

Transformation Code	X_t	Y_{t+h}^h
1	Z_t	Z_{t+h}
2	$Z_t - Z_{t-1}$	$Z_{t+h} - Z_t$
3	$(Z_t - Z_{t-1}) - (Z_{t-1} - Z_{t-2})$	$h^{-1}(Z_{t+h} - Z_t) - (Z_t - Z_{t-1})$
4	$\ln(Z_t)$	$\ln(Z_{t+h})$
5	$\ln(Z_t / Z_{t-1})$	$\ln(Z_{t+h}/Z_t)$
6	$\ln(Z_t/Z_{t-1}) - \ln(Z_{t-1}/Z_{t-2})$	$h^{-1}\{\ln(Z_t/Z_{t-1})\} - \ln(Z_t/Z_{t-1})$

Notes: This table defines the transformation codes (T) used in Table B.1. Z_t denotes the raw series, X_t denotes the transformed series used to compute the principal components, and Y_{t+h}^h denotes the series to be predicted.

Table 1

Categories of series in the data set

Group	Brief description	Examples of series	Number of series
1	GDP components	GDP, consumption, investment	16
2	IP	IP, capacity utilization	14
3	Employment	Sectoral & total employment and hours	20
4	Unempl. rate	unemployment rate, total and by duration	7
5	Housing	Housing starts, total and by region	6
6	Inventories	NAPM inventories, new orders	6
7	Prices	Price indexes, aggregate & disaggregate; commodity prices	37
8	Wages	Average hourly earnings, unit labor cost	6
9	Interest rates	Treasuries, corporate, term spreads, public-private spreads	13
10	Money	M1, M2, business loans, consumer credit	7
11	Exchange rates	average & selected trading partners	5
12	Stock prices	various stock price indexes	5
13	Cons. exp.	Michigan consumer expectations	1

Table 2

Distributions of Relative Root Mean Squared Errors (RMSE) by Forecasting Method, *h* = 1

Method	No. est'd shrinkage parameters	Percentiles				
		0.050	0.250	0.500	0.750	0.950
OLS	0	0.896	0.952	0.989	1.049	1.108
DFM-5	0	0.837	0.887	0.955	0.987	1.017
Pretest	1	**0.827**	0.885	0.935	0.993	1.000
Bagging	1	0.843	0.890	0.943	0.993	1.000
BMA	2	0.841	0.891	0.942	0.984	0.999
Logit	2	**0.827**	**0.878**	**0.929**	**0.977**	**0.995**

Notes: Entries in a given row are percentiles of the distribution of RMSEs, over the 143 series in the data set, of the forecasting method given in the first column. The RMSEs are relative to an AR(4). The forecasting methods, and the cross-validation method for computing the RMSE, are described in Section 3.2. The method with the lowest RMSE at each percentile appears in bold.

Table 3

Two measures of average similarity of forecast method, *h* = 1:

correlation (lower left) and mean absolute difference of forecasts (upper right)

	OLS	DFM-5	Pretest	Bagging	BMA	Logit
OLS		0.069	0.070	0.064	0.068	0.076
DFM-5	0.705		0.020	0.018	0.019	0.023
Pretest	0.803	0.906		0.008	0.009	0.006
Bagging	0.825	0.922	0.985		0.005	0.012
BMA	0.842	0.921	0.982	0.996		0.008
Logit	0.831	0.897	0.986	0.983	0.988	

Notes: Entries below the diagonal are the correlation between the RMSEs for the row/column forecasting methods, compute over the 143 series being forecasted. Entries above the diagonal are the mean absolute difference between the row/column method RMSEs, averaged across series.

Table 4

Distribution of Root Mean Square values of shrinkage function ψ , *h* = 1

Method	Percentiles				
	0.050	0.250	0.500	0.750	0.950
OLS	1.000	1.000	1.000	1.000	1.000
DFM-5	0.224	0.224	0.224	0.224	0.224
Pretest	0.000	0.100	0.141	0.300	0.812
Bagging	0.000	0.100	0.151	0.299	0.697
BMA	0.077	0.118	0.183	0.354	0.639
Logit	0.100	0.141	0.222	0.482	0.769

Table 5

Distribution of fraction of variance of ψ placed on the

first five principle components,

among series with root mean square shrinkage functions > 0.05, *h* = 1

Method	Number	Percentiles					Frac > 0.90
		0.050	0.250	0.500	0.750	0.950	
OLS	143	0.050	0.050	0.050	0.050	0.050	0.00
DFM-5	143	1.000	1.000	1.000	1.000	1.000	1.00
Pretest	112	0.000	0.121	0.429	1.000	1.000	0.38
Bagging	119	0.030	0.147	0.359	0.737	1.000	0.13
BMA	136	0.050	0.051	0.215	0.921	1.000	0.26
Logit	138	0.022	0.057	0.233	0.667	1.000	0.21

Table 6

Distributions of Forecast RMSE by Forecasting Method, *h* = 2, 4

(a) *h* = 2

Method	Percentiles				
	0.050	0.250	0.500	0.750	0.950
OLS	0.881	0.926	0.976	1.032	1.088
DFM-5	0.804	0.871	0.949	0.983	1.029
Pretest	**0.775**	0.869	**0.924**	0.968	1.000
Bagging	0.796	0.869	0.932	0.979	1.000
BMA	0.790	0.869	0.929	0.967	0.994
Logit	**0.776**	**0.861**	**0.924**	**0.957**	**0.991**

(b) *h* = 4

Method	Percentiles				
	0.050	0.250	0.500	0.750	0.950
OLS	0.892	0.934	0.965	1.013	1.068
DFM-5	0.781	**0.854**	0.936	0.982	1.046
Pretest	**0.778**	0.861	0.924	0.962	0.999
Bagging	0.781	0.855	0.926	0.966	1.000
BMA	0.780	0.858	0.922	0.961	0.989
Logit	**0.778**	**0.854**	**0.915**	**0.955**	**0.988**

Notes: Entries are percentiles of distributions of relative RMSEs over the 143 variables being forecasted, by series, at the 2- and 4-quarter ahead forecast horizon. All forecasts are direct. The method with the lowest RMSE at each percentile appears in bold. See the notes to Table 2.

Table 7

Median RMSE among series in the lower and upper quartiles of mean-square fraction of weight on the first five principle components in the logit model, *h* = 1, 2, and 4

Method	***h*** **= 1**		***h*** **= 2**		***h*** **= 4**	
	lower quartile	**upper quartile**	**lower quartile**	**upper quartile**	**lower quartile**	**upper quartile**
OLS	1.045	0.977	1.004	0.977	1.009	0.959
DFM-5	1.001	0.886	1.003	0.886	1.013	0.877
Pretest	0.999	0.891	0.982	0.904	0.982	0.849
Bagging	0.999	0.896	0.985	0.913	0.982	0.863
BMA	0.986	0.893	0.967	0.908	0.972	0.858
Logit	0.977	0.891	0.955	0.904	0.965	0.849

Table 8

Median RMSE by Forecasting Method and by Category of Series

(a) *h* = 1

Category	**Brief description**	OLS	DFM-5	Pretest	Bagging	BMA	Logit
1	GDP components	0.987	**0.905**	0.911	0.913	0.914	0.906
2	IP	0.954	**0.882**	0.890	0.888	0.887	0.884
3	Employment	0.968	**0.861**	0.871	0.878	0.878	0.871
4	Unempl. rate	0.929	0.800	**0.799**	**0.799**	**0.799**	**0.799**
5	Housing	0.940	0.936	**0.897**	0.911	0.912	**0.897**
6	Inventories	0.964	0.900	**0.886**	0.906	0.900	**0.886**
7	Prices	1.034	0.980	0.993	0.995	0.978	**0.970**
8	Wages	0.996	0.993	0.959	0.968	0.954	**0.938**
9	Interest rates	1.026	0.980	0.961	0.967	0.963	**0.946**
10	Money	0.987	0.953	**0.926**	0.948	0.944	**0.926**
11	Exchange rates	1.087	1.015	0.997	0.996	0.993	**0.981**
12	Stock prices	1.048	**0.983**	0.988	0.992	0.989	0.988
13	Cons. exp.	1.108	**0.977**	0.996	1.000	1.000	0.996

(b) *h* = 2

Category	**Brief description**	OLS	DFM-5	Pretest	Bagging	BMA	Logit
1	GDP components	0.945	0.907	0.882	0.892	0.889	**0.870**
2	IP	0.910	0.861	0.853	0.857	0.861	**0.852**
3	Employment	0.941	0.861	0.862	0.862	0.863	**0.859**
4	Unempl. rate	0.902	0.750	0.723	0.733	0.729	**0.723**
5	Housing	0.937	0.940	**0.902**	0.912	0.911	0.904
6	Inventories	0.944	**0.867**	0.879	0.879	0.878	0.876
7	Prices	1.042	0.977	0.968	0.979	0.973	**0.961**
8	Wages	0.942	0.999	0.937	0.942	0.933	**0.919**
9	Interest rates	0.945	0.952	0.934	0.943	0.938	**0.928**
10	Money	0.987	0.933	0.924	0.926	0.927	**0.921**
11	Exchange rates	1.036	1.015	1.000	1.000	0.986	**0.980**
12	Stock prices	1.013	0.977	0.968	0.975	0.971	**0.955**
13	Cons. exp.	1.149	0.963	**0.960**	0.987	0.977	**0.960**

Table 8, continued

(c) h = 4

Category	**Brief description**	OLS	DFM-5	Pretest	Bagging	BMA	Logit
1	GDP components	0.938	**0.906**	0.917	0.913	0.913	0.908
2	IP	0.944	**0.827**	0.837	0.847	0.845	0.836
3	Employment	0.940	0.844	0.846	0.846	0.847	**0.842**
4	Unempl. rate	0.903	0.762	**0.743**	0.750	0.747	**0.743**
5	Housing	0.916	0.926	0.889	0.887	0.888	**0.882**
6	Inventories	0.917	**0.856**	0.870	0.875	0.873	0.864
7	Prices	1.013	0.963	0.954	0.957	0.953	**0.948**
8	Wages	0.950	1.019	0.946	0.946	0.939	**0.931**
9	Interest rates	1.027	0.956	0.950	0.959	0.958	**0.949**
10	Money	0.998	**0.909**	0.939	0.937	0.940	0.937
11	Exchange rates	1.009	1.036	**0.965**	0.983	0.973	**0.965**
12	Stock prices	0.997	0.974	0.967	0.968	0.964	**0.961**
13	Cons. exp.	1.075	0.966	**0.955**	0.970	0.961	**0.955**

Notes: entries are the median relative RMSE among the relative RMSEs for the series among the row category, using the column forecasting method. For each series category, the smallest relative RMSE (i.e. the row minimum) appears in bold.

Table 9

Median fraction of variance of ψ placed on the first five principle components, among series with root mean square shrinkage functions > 0.05, by category of series

(a) h = 1

Category	Brief description	Pretest	Bagging	BMA	Logit
1	GDP components	0.750	0.549	0.611	0.378
2	IP	0.667	0.613	0.711	0.667
3	Employment	1.000	0.789	0.792	0.576
4	Unempl. rate	1.000	0.637	0.658	0.532
5	Housing	0.089	0.119	0.091	0.110
6	Inventories	1.000	0.771	0.993	0.788
7	Prices	0.222	0.180	0.056	0.057
8	Wages	0.052	0.057	0.050	0.043
9	Interest rates	1.000	0.199	0.130	0.199
10	Money	0.333	0.323	0.304	0.333
11	Exchange rates	0.000	0.010	0.037	0.019
12	Stock prices	0.250	0.235	0.117	0.114
13	Cons. exp.	.	.	.	.

(b) h = 2

Category	Brief description	Pretest	Bagging	BMA	Logit
1	GDP components	1.000	0.660	0.537	0.218
2	IP	0.500	0.315	0.409	0.357
3	Employment	1.000	0.559	0.799	0.438
4	Unempl. rate	1.000	0.883	1.000	1.000
5	Housing	0.218	0.185	0.189	0.198
6	Inventories	0.525	0.524	0.450	0.383
7	Prices	0.375	0.264	0.173	0.172
8	Wages	0.060	0.065	0.052	0.051
9	Interest rates	0.072	0.066	0.052	0.070
10	Money	1.000	0.139	0.994	0.396
11	Exchange rates	0.063	0.055	0.050	0.050
12	Stock prices	0.200	0.263	0.124	0.205
13	Cons. exp.	1.000	0.814	0.919	1.000

(c) h = 4

Category	Brief description	Pretest	Bagging	BMA	Logit
1	GDP components	0.233	0.182	0.170	0.143
2	IP	0.750	0.536	0.670	0.586
3	Employment	1.000	0.766	0.928	0.538
4	Unempl. rate	1.000	0.998	1.000	1.000
5	Housing	0.266	0.134	0.283	0.303
6	Inventories	0.500	0.348	0.398	0.355
7	Prices	0.400	0.346	0.247	0.237
8	Wages	0.093	0.080	0.063	0.047
9	Interest rates	0.294	0.268	0.092	0.102
10	Money	1.000	0.309	0.167	0.182
11	Exchange rates	0.029	0.063	0.053	0.054
12	Stock prices	0.085	0.103	0.079	0.077
13	Cons. exp.	1.000	0.385	0.521	1.000

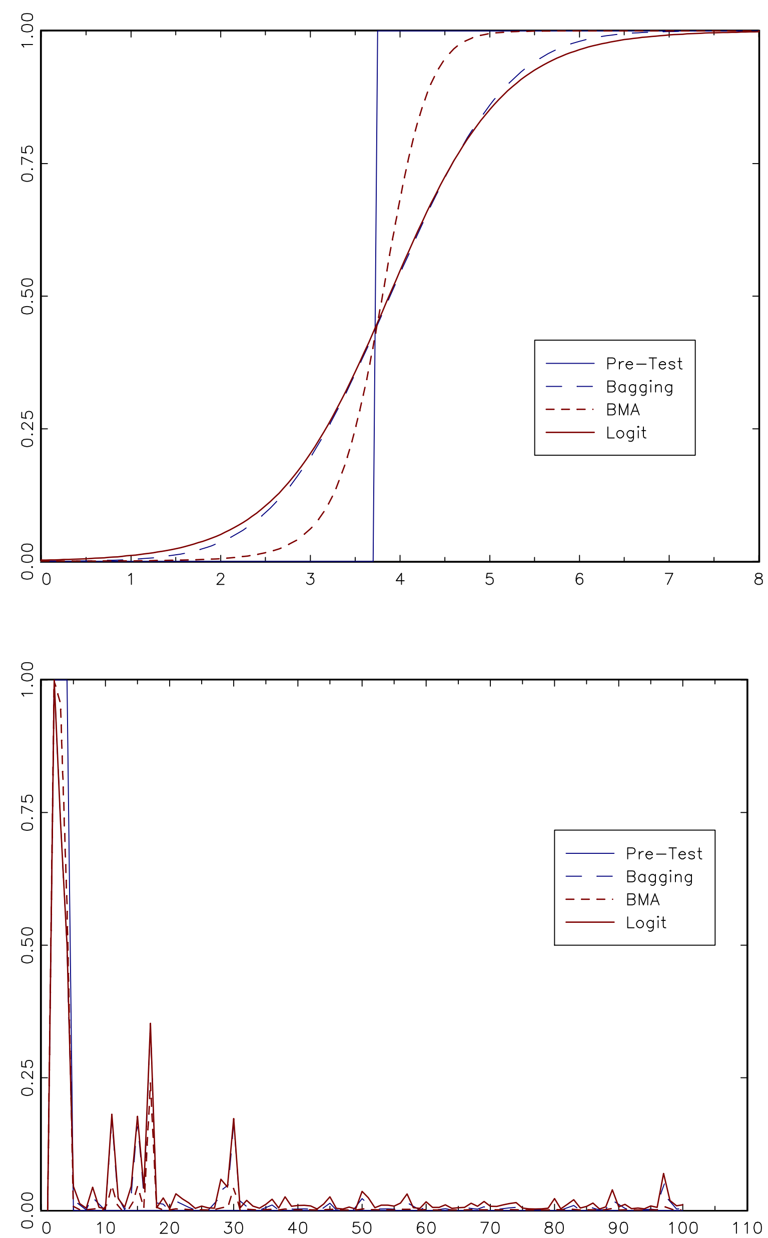

Figure 1

Estimated shrinkage functions (upper panel) and weights $\psi(t_i, \hat{\theta})$ on ordered principle components 1-100: Total employment growth, $h = 2$

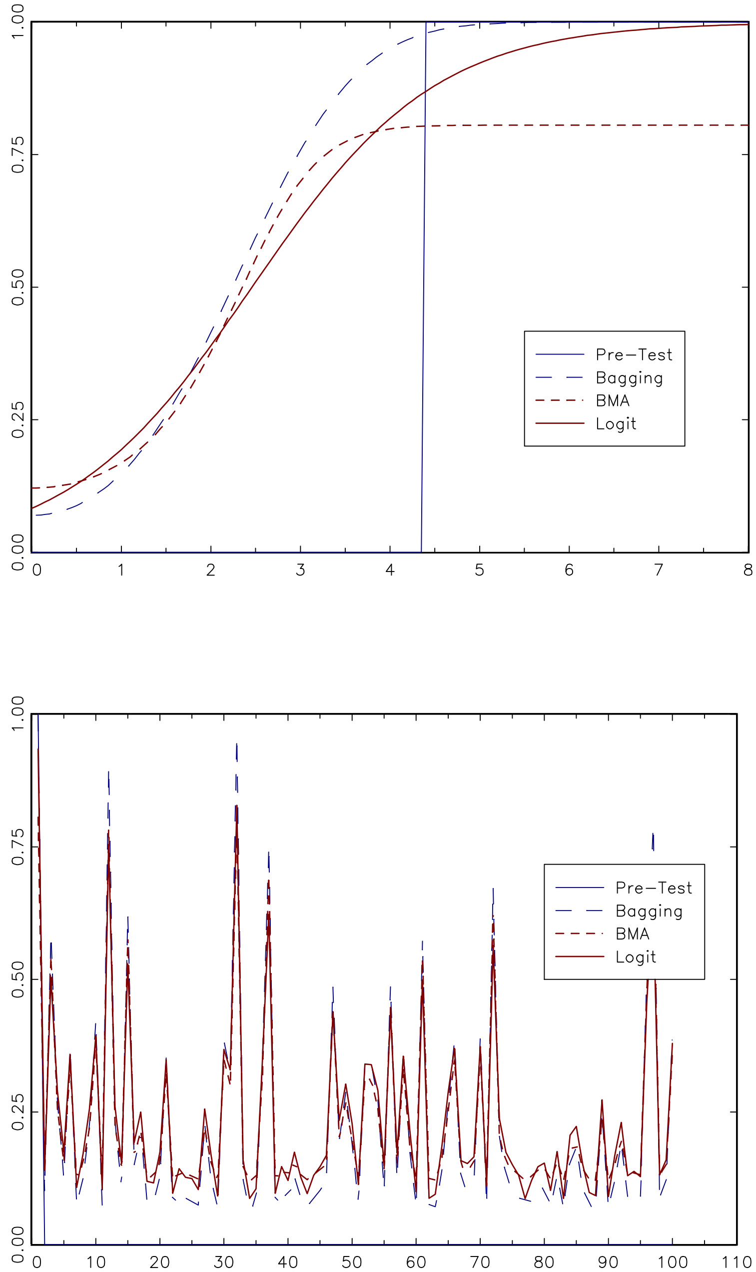

Figure 2

Estimated shrinkage functions (upper panel) and weights $\psi(t_i, \hat{\theta})$ on ordered principle components 1-100: Spread between 10-year and 90-day Treasury rates, $h = 1$

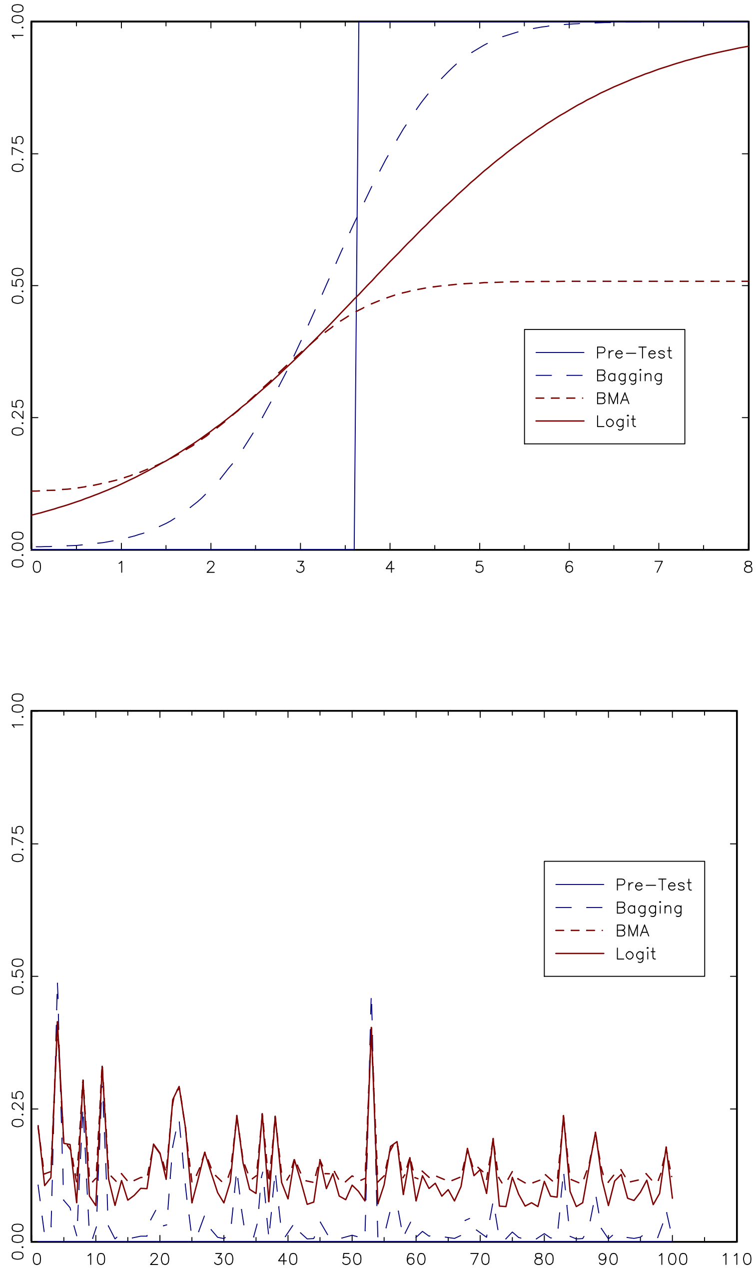

Figure 3
Estimated shrinkage functions (upper panel) and weights $\psi(t_i, \hat{\theta})$ on ordered principle components 1-100: Growth of S&P500 Index, h = 1