Generalized Shrinkage Methods for Forecasting using Many Predictors

This revision: February 2011

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

AUTHOR'S FOOTNOTE

James H. Stock is the Harold Hitchings Burbank Professor of Economics, Harvard University (e-mail: james_stock@harvard.edu). Mark W. Watson is the Howard Harrison and Gabrielle Snyder Beck Professor of Economics and Public Affairs, Princeton University (mwatson@princeton.edu). The authors thank Jean Boivin, Domenico Giannone, Lutz Kilian, Serena Ng, Lucrezia Reichlin, Mark Steel, and Jonathan Wright for helpful discussions, Anna Mikusheva for research assistance, and the referees for helpful suggestions. An earlier version of the theoretical results in this paper was 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.

Generalized Shrinkage Methods for Forecasting using Many Predictors

JBES09-255 (first revision)

This revision: February 2011

[BLINDED COVER PAGE]

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 most series, including measures of real economic activity, the shrinkage forecasts are inferior to the DFM forecasts.

Key Words: high dimensional model; empirical Bayes; dynamic factor models

JEL Code: C53

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 turn the curse of dimensionality into a blessing (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). Bai and Ng (2008) and Stock and Watson (2011) survey econometric research on DFMs over this period. 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, which suggests considering 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 (Andersson and Karlsson 2008, Bańbura, Giannone, and Reichlin 2010, Korobilis 2008, Carriero, Kapetanios, and Marcellino 2009, and De Mol, Giannone, and Reichlin 2008), Bayesian model averaging (Koop and Potter 2004, Wright 2004, Jacobson and Karlsson 2004, and Eklund and Karlsson 2007), 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 those derivations what the algorithms are actually doing when they are 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 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 using a quarterly U.S. macro data set that includes 143 quarterly economic time series spanning 49 years. The DFM imposes a strong restriction: that there are only a few factors and these factors can supplant the full large data set for the purpose of forecasting. There are now a number of ways to estimate factors in large data sets, and a commonly used estimator is the first few principle components of the many predictors (ordered by their eigenvalues). The empirical question, then, is whether information in the full data set, beyond the first few principle components, makes a significant marginal forecasting contribution. There are various ways to approach this question. One could, for example, retain the predictors in their original form, then (by appealing to Frisch-Waugh) consider the marginal predictive power of the part of those predictors orthogonal to the factors. Algorithms for averaging or selecting models using the original predictors, which have been used for macro forecasting or closely related problems, include BMA and large VARs. However, we share De Mol, Giannone, and Reichlin's (2008) skepticism about the reliability of any resulting economic interpretation because of the colinearity of the data and the resulting instability of the weights and variable/model selection. Moreover, any economic interpretation that might have been facilitated by using the original series would be obscured by using instead their orthogonal projection on the first few factors. A different approach, the one we adopt, is to retain the perspective of a factor model but to imagine that the number of selected factors is simply smaller than it should be, that is, that the conventional wisdom that a few factors suffices to describe the postwar U.S. data is wrong. Because the principle components are estimates of the factors, this approach leads us consider forecasts that potentially place nonzero weight on principle components beyond the first few. Because the principle components are orthogonal, shrinkage procedures for orthogonal regressors provide a theoretically well-grounded way to assess the empirical validity of the DFM forecasting restrictions.

We find that, for most 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 and monetary variables. For these series, the shrinkage methods with estimated parameters fail to provide mean squared error improvements over the DFM. For a small number of series, the shrinkage forecasts improve upon DFM forecasts, at least at some horizons and by some measures, and for these few series, the DFM might not be an adequate approximation. Finally, none of the methods considered here help much for series that are notoriously difficult to forecast, such as exchange rates, stock prices, 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,

$$Y_t = \delta' P_{t-1} + \varepsilon_t, \quad t = 1,\ldots, T, \quad P'P/T = I_n \qquad (1)$$

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, $\{X_t\}$.

When n is large, there are many regressors and OLS will work poorly. Therefore we consider forecasting methods that impose and exploit additional structure on the coefficients in (1). We 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 $s_e^2 = \sum_{t=1}^{T}(Y_t - \hat{\delta}'P_{t-1})^2 \big/ (T-n)$, 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, these methods can be applied and their performance understood 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 (PT) 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,

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

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 Normal Bayes (NB) 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, respectively G_τ and G_{σ^2} (where G denotes a generic prior):

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

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 (BMA).

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 (Zellner 1986) for δ conditional on inclusion. Specifically, with probability p let $\delta_i|\sigma \sim \mathrm{N}(0, \sigma^2/(gT))$ (so $\tau_i \sim \mathrm{N}(0, 1/g)$), and with probability $1 - p$ let δ_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 (EB)

Empirical Bayes 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 (BG)

Bootstrap aggregation or "bagging" (Breiman 1996) smooths 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 $u \neq 0$),

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

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 (2002) expression (11) under weaker assumptions on the number and properties of the regressors than in Bühlmann and Yu (2002).

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 $\psi(.)$. 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. This section shows that $\tilde{Y}_{T+1|T} - \hat{Y}_{T+1|T} \xrightarrow{m.s.} 0$ for the NB and BG forecasts.

First consider the NB forecast described in section (2.2). If σ^2 were 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 $1/\hat{\sigma}^2$ 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) $\max_i |P_{iT}| \le P_{max}$, a finite constant.

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

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

(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) $\sup_u|u^m d^m \psi^{NB}(u)/du^m)| \le M$ 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 representation of bagging allows us to obtain a condition which, if satisfied, implies that bagging is asymptotically admissible; this result appears to be unavailable elsewhere. 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 $\tilde{g}(s) = \int e^{ist}g(t)dt$. Then g satisfies (12) if $\tilde{g}$ satisfies the Fredholm equation of the second kind, $\tilde{g}(s) = \int K(s,t)\tilde{g}(t)dt$, where

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

6. Tibshirani (1996, Section 2.2) provides a soft-thresholding or shrinkage representation for the Lasso estimator with orthonormal regressors, derived for strictly exogenous regressors.

3. Empirical Analysis: Data and Methods

The empirical analysis examines whether the shrinkage methods improve upon dynamic factor model forecasts that use only the first few principle components.

3.1 The Data

The data set consists of quarterly observations on 143 U.S. macroeconomic time series from 1960:II through 2008:IV, for a total of 195 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 rates of inflation) 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}. 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*-quarter inflation minus current 1-quarter inflation (both at annual rates).

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

3.2 Methods

This section summarizes the forecasting procedures and the estimation of their parameters and mean square forecast error (MSE). The MSE is estimated in two complementary ways: the first uses "leave *m* out" cross-validation using the full sample, and the second uses a rolling pseudo out-of-sample forecast method. The cross-validation approach has the advantage of using more observations for estimation and forecasting than the rolling approach, and it provides an estimate of the average performance of the forecasting methods over the full data set. The rolling approach is more conventional in the macro forecasting literature and allows for time variation in predictive relations and volatility, however because of the need for a large startup sample the rolling results pertain only to the post-1984 "Great Moderation" period.

Forecasting procedures. We examine six forecasting procedures.

1. *DFM-5*. The DFM-5 forecast uses the first five principle components as predictors, with coefficients estimated by OLS without shrinkage; the remaining principle components are omitted.
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 $\psi^{OLS} = 1$).

Preliminary investigation showed considerable instability in nonparametric empirical Bayes estimators, perhaps because the number of observations is too small for nonparametrics, so those methods are not pursued here.

MSE estimation by cross-validation. 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 series/horizon/forecasting method 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 $\tilde{Y}_{t+h}^{h,cv}$ and $\tilde{X}_t^{cv}$ denote the residuals from these regressions, standardized to have unit variance in the full sample. The principle components P_t^{cv} of $\tilde{X}_t^{cv}$ 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^{cv}.

2. Let $\Im_t^{cv}$ = {1,…, t–2h–3, t+2h+3,…, T}, that is, the full data set dropping the t^{th} observation and 2h+2 observations on either side. At each date t = 1,…, T–h, the OLS estimators of δ are computed by regressing $\tilde{Y}_{t+h}^{h,cv}$ on P_t^{cv} using observations $t \in \Im_t^{cv}$. Denote these estimators as $\hat{\delta}_{j,t}^{h,cv}$, j = 1,.., n. Let $\hat{\tau}_{j,t}^{h,cv}$ denote the conventional OLS t-statistic corresponding to $\hat{\delta}_{j,t}^{h,cv}$ (not adjusting for heteroskedasticity or serial correlation).

3. The parameter θ is then estimated by minimizing the sum of squared cross-validated prediction errors:

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

Because these are direct forecasts, the estimator $\hat{\theta}^h$ differs by forecast horizon. The estimated shrinkage function for this dependent variable and horizon is $\psi(.,\hat{\theta}^h)$. The cross-validation estimate of the MSE is MSE^{cv}($\hat{\theta}^h$).

All regressions involving P (over all sample periods) impose the moment condition that $P'P$/rows(P) = I. 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).

MSE estimation by rolling pseudo out-of-sample forecasting. In the rolling calculation, the forecaster, standing at date *t*, applies the cross-validation algorithm (described above) to the most recent 100 observations to estimate θ for a series/horizon/forecasting method, then uses this estimate of θ to forecast Y_{t+h}^{h}; this is repeated for the 96 – *h* rolling forecast dates *t* = 1985:I,…, 2008:IV–*h*. Because rolling sample is roughly half as large as the cross-validation full sample (100 versus 196 observations), *n* = 50 principal components are used in rolling sample forecasts (versus 100 in the full-sample forecasts). This produces a sequence of rolling pseudo out-of-sample forecasts, $\hat{Y}_{t+h|t}^{h,rolling}$, computed using the rolling window of length 100-*h*. The rolling estimate of the MSE for a candidate forecast is $\text{MSE}^{rolling} = (96-h)^{-1}\sum_{t=1985:\text{I}}^{2008:\text{IV}-h}\left(Y_{t+h}^{h}-\hat{Y}_{t+h|t}^{h,rolling}\right)^{2}$.

4. Empirical Results

We begin with results for all series combined, then break the results down by category of series.

4.1 Results for all series

Tables 2-5 report the distributions of relative root mean squared errors (RMSEs) for various forecasting methods. Table 2 presents percentiles of the distribution of one-step ahead cross-validation RMSEs over the 143 series for the seven forecasting methods, where the RMSEs are relative to the AR(4) benchmark. The AR forecast is a conventional benchmark in the macro forecasting literature and the results in Table 2 are in line with results in that literature for U.S. data over this full-sample period. According to these estimates, the DFM-5 forecasts improve upon the AR(4) in more than 75% of series, and in 25% of series the improvement is substantial, with relative RMSEs less than 0.887. The shrinkage methods provide improvements over the AR(4) model that are in the same range as the DFM-5 forecasts.

Because our primary interest is in whether the use of additional principal components improves upon conventional low-dimensional factor model forecasts, Tables 3-5 consider RMSEs relative to DFM-5. Specifically, the distributions in Tables 3-5 are of the ratio (by series) of the candidate forecast RMSE to the DFM-5 forecast RMSE, for horizons h = 1, 2, and 4. Table 3 summarizes full-sample relative RMSEs estimated by cross-validation. Tables 4 and 5 respectively report the cross-validated and rolling estimates of distributions of RMSEs over the rolling pseudo out-of-sample forecast period, 1985:I to 2008:IV-h (for Table 4, the forecasts are computed using full-sample cross-validation method of Section 3.2, but the MSEs are computed using the subsample of forecasts from 1985:I to 2008:IV-h).

Comparing Tables 3-5 requires bearing in mind two considerations. First, the rolling forecast period coincides with the Great Moderation, during which these series experienced reduced volatility (Kim and Nelson 1999, McConnell and Perez-Quiros 2000, Stock and Watson 2002) and reduced predictability (Stock and Watson 2002, D'Agostino, Giannone, and Surico 2006). Second, the cross-validation MSEs for the shrinkage methods, and hence for the RMSEs relative to DFM-5, are presumably biased down because θ was estimated by minimizing MSE^{cv} (see (15)) and the MSE does not have a "degrees-of-freedom" adjustment. The rolling shrinkage MSEs should not have this in-sample downward bias because they are computed from pseudo out-of-sample forecasts. Because we do not have a degrees-of-freedom adjustment for the cross-validated RMSEs – what that adjustment should be is not clear given the serial correlation and overlapping data structure of the multiperiod forecasts – we would expect the cross-validated RMSE to be less than the rolling RMSE, with a bias that is larger for the shrinkage methods in which dim(θ) = 2 (BMA and logit) than for methods for which dim(θ) = 1 (pretest and bagging).

With these considerations in mind, three features of Tables 3-5 are noteworthy. First, based on the full-sample cross-validated results (Table 3), the median improvements of the shrinkage methods over the DFM-5 benchmark are at best modest: the logit model provides the greatest improvement at $h = 1$, but the relative RMSE improvement is only 1.1%. Some series show larger improvements, especially at more distant horizons: at $h = 4$, for 25% of the series the RMSE of the logit forecast, relative to DFM-5, is 0.947 or less.

Second, comparing Tables 3 and 4 reveals that the distributions of RMSEs of the shrinkage methods, relative to DFM-5, are quite similar in the full sample and in the 1985-2008 subsample. Although the post-85 distributions are more dispersed than the full-sample distributions, this increase in dispersion is roughly consistent with the full-sample RMSEs in Table 3 being based on twice as many observations as the subsample RMSEs in Table 4. This finding of substantial stability in these distributions of MSEs, relative to DFM-5, is rather surprising given the large documented shifts in the time series properties of these series across the pre-85 and post-85 samples.

Third, the rolling estimates of the relative RMSEs in Table 5 for the shrinkage methods are larger than the corresponding cross-validation estimates in Table 4. For example, for BMA the median relative RMSE at $h = 1$ increases by .023 (from 0.991 in

Table 4 to 1.014 in Table 5); for pretest, this increase is .058 (from 0.990 to 1.048). An increase is expected because the cross-validated MSE has no “degrees-of-freedom” adjustment, but absent a formula for this adjustment it is hard to calibrate the magnitude of this increase. This increase is larger at longer horizons, which (again appealing to the degrees-of-freedom intuition) is not surprising because of the overlapping observations used to estimate θ in the multi-step direct forecasts. At the median, the rolling relative RMSEs all exceed 1. The improvements at the 25th percentile in Table 5 are quite modest. It appears that, for a few series, substantial improvements are possible, at least at longer horizons, but there are also series for which using the shrinkage methods results in a substantial deterioration of forecast performance. Indeed, the distributions of relative RMSEs in Tables 3-5 are roughly symmetric, which is consistent with the view that their dispersion simply arises from sampling variability.

One question, not addressed in Tables 3-5, is whether the shrinkage methods offer improvements for those series in which they add value relative to the AR benchmark; that is, among series for which multivariate methods are useful, are the shrinkage methods better than DFM-5? Table 6 examines this question by reporting the median rolling pseudo out-of-sample RMSE, relative to DFM-5, conditional on the candidate method improving upon the AR forecast for that series/horizon combination. For nearly all shrinkage methods and horizons, these medians are quite close to 1, indeed they exceed 1 in 8 of the 12 method/horizon combinations. Even for those series for which the shrinkage method outperforms the AR(4), the forecaster is typically better off just using DFM-5.

Tables 7-9 explore the extent to the cross-validation shrinkage forecasts differ from each other and from the DFM-5 forecast. Table 7 presents two measures of similarity of the performance one-step ahead forecasts: the correlation (over series) among the cross-validation RMSEs, relative the AR(4) forecasts, and the mean absolute difference of these relative RMSEs. Table 8 reports the distribution across series of the root mean square shrinkage function, $\left(\sum_{i=1}^{100}\psi(\hat{\tau}_{i,t}^{h,cv};\hat{\theta}_j^{h,cv})^2 \Big/ 100\right)^{1/2}$, where $\hat{\theta}_j^{h,cv}$ 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 9 gives the distribution across series of the average fraction of the mean squared variation in the ψ's attributable to the first five principle components, $\sum_{i=1}^{5}\psi(\hat{\tau}_{i,t}^{h,cv};\hat{\theta}_j^{h,cv})^2 \Big/ \sum_{i=1}^{100}\psi(\hat{\tau}_{i,t}^{h,cv};\hat{\theta}_j^{h,cv})^2$, among those series for which the root mean square shrinkage function considered in Table 8 is at least 0.05. (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.) The final column of Table 9 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.

As indicated in Table 7 the shrinkage methods tend to produce forecasts with similar performance, with RMSE correlations all exceeding 0.98. The correlations between the shrinkage and DFM-5 RMSEs are distinctly lower, ranging from 0.897 to 0.922, and their correlations with the OLS RMSE are lower yet. According to Table 8, the typical weight ψ for the shrinkage methods is somewhat less than for DFM-5, but these weights differ substantially across series. Table 9 shows that the fraction of mean-square weight that the shrinkage methods put on the first five principle components also 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, but for one-quarter of the series the logit model places only 5.7% of its mean-square weight on the first five principle components.

Taken together, these results suggest that the shrinkage methods seem to offer little or no improvements over DFM-5, at least on average over all these series. The median cross-validation relative RMSEs are somewhat less than 1 and the median rolling RMSEs are somewhat greater than 1. It is plausible to think that these two estimates bracket the true RMSE: the cross-validation estimates are biased down because they do not include an adjustment for the estimation of θ, while the rolling estimates arguably understate performance because the rolling sample of 100 observations is smaller than the full sample size, which increases estimation uncertainty for the shrinkage parameters. This bracketing argument suggests that, for this full sample of series, the typical relative RMSE of a shrinkage method to the DFM-5 is quite close to 1 at all horizons considered.

4.2 Results by category of series

Tables 10 and 11 break down the results by the 13 categories of series in Table 1. Tables 10 and 11 report the median cross-validated RMSE by category, respectively relative to AR(4) and relative to DFM-5. To save space, we only report results here for MSEs estimated over the full sample by cross validation.

Generally speaking, the categories fall into three groups. For the major measures of real economic activity (GDP components, IP, employment, unemployment rates, and inventories), the DFM-5 method has the lowest, or nearly the lowest, RMSE among the various methods, relative to the AR. For these series, RMSE improvements using the DFM-5 model are substantial relative to the AR(4) model, especially at longer horizons. For these categories, the shrinkage methods typically do not improve upon the DFM-5, especially once one takes into account the downward bias of the cross-validated shrinkage MSEs. Series in this group also include some interest rates and monetary series (at least at the 4-quarter horizon). 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 (results not tabulated here). 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, at $h = 1$. The upper panel presents the estimated shrinkage functions, and the lower panel plots the weight placed by the various shrinkage functions on each of the 100 ordered principle components. At $h = 1$, the DFM-5 RMSE, relative to AR, is 0.852, slightly less than the Logit relative RMSE of 0.870, and both RMSEs indicate a substantial improvement over the AR (the respective rolling relative RMSEs are 0.910 and 0.935). All the estimated shrinkage functions are similar, placing substantial weight only for t-statistics in excess of approximately 3.5, and the estimated logit and pretest shrinkage functions are nearly identical. The shrinkage functions end up placing nearly all the 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 increases the RMSE.

There is some evidence of a second, smaller group of series for which one or more shrinkage forecast improves on both the AR and DFM-5 forecast, but that evidence is delicate and mixed over horizons, among series within categories, and over cross-validation versus rolling RMSEs. Series in this group include real wages, some housing variables, and some interest rates/spreads. For example, for wages, the median cross-validation RMSE, relative to AR, for the logit model is .919 at the 2-quarter horizon, whereas the corresponding relative RMSE for DFM-5 is .999. Relative to DFM-5, the median RMSE for all four shrinkage methods for wages at $h = 2$ is between .927 and .934. These improvements for wages, however, are not found using the rolling RMSEs. One possibility is that these improvements are restricted to the pre-1985 period and thus do not show up in the rolling relative RMSEs, but another possibility is that they are simply a statistical artifact of cross-validated fitting.

The final group consists of hard-to-forecast series for which the principle components do not provide meaningful reductions in cross-validation or rolling RMSEs, relative to AR, using either the DFM-5 or shrinkage forecasts. This group includes price inflation, exchange rates, stock returns, and consumer expectations. The shrinkage parameter objective function (15) is quite flat for many of these series. Figure 2 presents estimated shrinkage functions and weights for a series in this third group, the percentage change in the S&P 500 Index. For all but the pretest forecast, most shrinkage methods place a weight of 0.1 to 0.2 on most of the principle components. For the S&P 500, the cross-validation RMSE of DFM-5, relative to AR, is 0.989, and these relative RMSEs range from 0.994 to 1.000 for the shrinkage methods; the respective rolling relative RMSEs are 0.994 and, for shrinkage methods, from 1.013 to 1.027.

4.3 Additional results and sensitivity checks

We also estimated by cross-validation a logit model with a quadratic term to obtain a more flexible parametric specification. The shrinkage function is for the quadratic logit model is,

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

The cross-validated fit of (16) is only marginally better than the linear logit model (14), which we interpret as yielding essentially no improvement accounting for the additional parameter.

We also repeated the analysis using Newey-West (1987) standard errors (with a window width of h+1), instead of the homoskedasticity-only OLS standard errors used above, including reestimating (by full-sample cross-validation) the shrinkage parameters using the Newey-West t-statistics. There were no substantial changes in the findings discussed above.

5. Discussion

Two 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 explains most of the variation in 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 (the first five principle components) 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 forecasting potential of improved DFM variants.

The facts that some of these shrinkage methods have an interpretation as an empirical Bayes method and that we have considered a number of flexible functional forms leads us to conclude that it will be difficult to improve systematically upon DFM forecasts using time-invariant linear functions of the principle components of large macro data sets like the one considered here. This conclusion complements Bańbura, Giannone, and Reichlin (2010) and De Mol, Giannone, and Reichlin (2008), who reached a similar conclusion concerning many-predictor models specified in terms of the original variables instead of the factors. This suggests that further forecast improvements over those presented here will need to come from models with nonlinearities and/or time variation, and work in this direction has already begun (e.g. Banerjee, Marcellino, and Masten 2009, Del Negro and Otrok 2008, Stock and Watson 2009, and Stevanović 2010a, b).

References

Andersson, M.K. and S. Karlsson (2008), "Bayesian forecast combination for VAR models (2008), "Bayesian Forecast Combination for VAR Models," *Advances in Econometrics*, 23, 501-524.

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.

Bańbura, M., D. Giannone, and L. Reichlin (2010), "Large Bayesian Vector Auto Regressions," *Journal of Applied Econometrics* 25, 71-92.

Banerjee, A., M. Marcellino, and I. Masten (2009), "Forecasting Macroeconomic Variables Using Diffusion Indexes in Short Samples With Structural Change," in H. Beladi and E. Kwan Choi (eds.), *Forecasting in the Presence of Structural Breaks and Model Uncertainty (Frontiers of Economics and Globalization, Volume 3)*, Emerald Group Publishing Limited, pp.149-194

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.

Del Negro, M. and C. Otrok (2008), "Dynamic Factor Models with Time-Varying Parameters: Measuring Changes in International Business Cycles," Federal Reserve Bank of New York Staff Reports No. 326.

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.

Eklund, J. and S. Karlsson (2007), "Forecast Combination and Model Averaging using Predictive Measures," *Econometric Reviews* 26, 329-363.

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.

Jacobson, T. and S. Karlsson (2004), "Finding Good Predictors for Inflation: A Bayesian Model Averaging Approach," *Journal of Forecasting* 23, 479-496.

Kim, C.-J. and C.R. Nelson (1999), "Has the U.S. economy become more stable? A Bayesian approach based on a Markov-switching model of the business cycle," *The Review of Economics and Statistics* 81, 608 – 616.

McConnell, M.M and G. Perez-Quiros, (2000), "Output fluctuations in the United States: What has changed since the early 1980's" *American Economic Review*, Vol. 90, No. 5, 1464-1476.

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

Korobilis, D. (2008), "Forecasting in VARs with Many Predictors," *Advances in Econometrics*, 23, 403-431.

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.

Newey, W. K. and West, K. D. (1987). "A Simple Positive Semi-Definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix," *Econometrica* 55, 703-708.

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

Stevanović, D. (2010a) "Factor Time Varying Parameter Models," manuscript, University of Montreal.

Stevanović, D. (2010b) "Common Sources of Parameter Instability in Macroeconomic Models: A Factor-TVP Approach," manuscript, University of Montreal.

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.

Stock, J.H. and M.W. Watson (2009), "Forecasting in Dynamic Factor Models Subject to Structural Instability," Ch. 7. in N. Shephard and J. Castle (eds), *The Methodology and Practice of Econometrics: Festschrift in Honor of D.F. Hendry*. Oxford: Oxford University Press.

Stock, J.H. and M.W. Watson (2011), "Dynamic Factor Models," forthcoming in *Oxford Handbook of Economic Forecasting*, Michael P. Clements and David F. Hendry (eds), Oxford University Press.

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

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

Zellner A. (1986), “On Assessing Prior Distributions and Bayesian Regression Analysis with g-Prior Distributions,” in *Bayesian Inference and Decision Techniques: Essays in Honour of Bruno de Finetti*, P.K. Goel and A. Zellner (eds). North-Holland: Amsterdam, 233–243.

Appendix A

Proofs of Results in Section 2

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

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

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

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

$$\begin{aligned} \rho_i &= \left[E^M \psi(\hat{\tau}_i) - \psi(\kappa t_i) \right] \hat{\delta}_i \\ &= \left[E^M \psi^{NB}(\hat{t}_i \zeta^{1/2}) - \psi^{NB}(\hat{t}_i) \right] \hat{\delta}_i \\ &= \frac{1}{8} \hat{\delta}_i E^M \left[\left(\tilde{t}_i^2 \psi^{NB\prime\prime}(\tilde{t}_i) - \tilde{t}_i\, \psi^{NB\prime}(\tilde{t}_i) \right) \tilde{\zeta}^{-2} (\zeta - 1)^2 \right] \end{aligned} \tag{18}$$

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

$$\begin{aligned} \left(\sum_{i=1}^{n} \rho_i P_{iT} \right)^2 &= \left(\frac{1}{8} \sum_{i=1}^{n} P_{iT} \hat{\delta}_i E^M \left[\left(\tilde{t}_i^2 \psi^{NB\prime\prime}(\tilde{t}_i) - \tilde{t}_i\, \psi^{NB\prime}(\tilde{t}_i) \right) \tilde{\zeta}^{-2} (\zeta - 1)^2 \right] \right)^2 \\ &\le \frac{1}{64} \left(\sum_{i=1}^{n} P_{iT}^2 \hat{\delta}_i^2 \right) \sum_{i=1}^{n} \left\{ E^M \left[\left(\tilde{t}_i^2 \psi^{NB\prime\prime}(\tilde{t}_i) - \tilde{t}_i\, \psi^{NB\prime}(\tilde{t}_i) \right) \tilde{\zeta}^{-2} (\zeta - 1)^2 \right] \right\}^2 . \end{aligned}$$

Now, $\sum_{i=1}^{n} P_{iT}^2 \hat{\delta}_i^2 \le P_{\max}^2 \sum_{i=1}^{n} \hat{\delta}_i^2 \le P_{\max}^2\, T^{-1} \sum_{t=1}^{T} Y_t^2$, 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

$\left\{E^M\left[\left(\tilde{t}_i^2\psi^{NB''}(\tilde{t}_i)-\tilde{t}_i\,\psi^{NB'}(\tilde{t}_i)\right)\tilde{\zeta}^{-2}(\zeta-1)^2\right]\right\}^2 \le 4M^2\left\{E^M\left[\tilde{\zeta}^{-2}(\zeta-1)^2\right]\right\}^2$ (by A6). Thus,

repeated application of the Cauchy-Schwarz inequality yields

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

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

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

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

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

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

Where the final inequality follows from $E^M(\tilde{\zeta}^{-4}) \le E^M[\max(1,\zeta^{-4})] \le 1 + E^M(\zeta^{-4})$, 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 $\text{var}\left[y\times 1(|y|>D)\right] \le 1 + \mu^2$.

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

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

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

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

where B is the number of bootstrap draws.

By construction, under the * distribution, $\hat{\delta}_{ij}^{*}$ ~ i.i.d. N($\hat{\delta}_i$, s_e^2/T) so $\sqrt{T}\hat{\delta}_{ij}^{*}/s_e$ ~ i.i.d. N(t_i, 1), ξ^* ~ i.i.d. $\chi_{T-n}^2/T-n$, 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}^{*} ~ 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 $\rho_i = \tilde{\delta}_i^{BG} - \psi^{BG}(t_i)\hat{\delta}_i$. 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 $\rho_{2i} = E^*\tilde{\delta}_i^{BG} - \psi^{BG}(t_i)\hat{\delta}_i$, 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>:

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

$$\text{var}^*(\rho_{1i}) = \text{var}^*(\tilde{\delta}_i^{BG} - E^*\tilde{\delta}_i^{BG})$$

$$= \text{var}^*\left\{\frac{1}{B}\sum_{j=1}^{B}\left[1(|t_{ij}^*|>c)\hat{\delta}_{ij}^* - E^*1(|t_{ij}^*|>c)\hat{\delta}_{ij}^*\right]\right\}$$

$$= \frac{1}{B}\text{var}^*\left[1(|t_{ij}^*|>c)\hat{\delta}_{ij}^*\right]$$

$$= \frac{s_e^2}{T}\frac{1}{B}\text{var}^*\left[1(|t_{ij}^*|>c)\frac{\sqrt{T}\hat{\delta}_{ij}^*}{s_e}\right]$$

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

$$\leq \frac{1}{T-n}\frac{1}{B}\left(T^{-1}\sum Y_t^2\right)\left(1+t_i^2\right) ,$$

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 $(T-n)s_e^2 \leq \sum_{t=1}^{T} Y_t^2$ and the lemma. Thus

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

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

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

$$\rho_{2i} = E^*\tilde{\delta}_i^{BG} - \psi^{BG}(t_t)\,\hat{\delta}_i$$

$$= E^*[1(|\, t_{ij}^* \,|> c)\hat{\delta}_{ij}^*] - \psi^{BG}(t_t)\,\hat{\delta}_i$$

$$= \frac{s_e}{\sqrt{T}}\left[E^*\left\{1(|\, t_{ij}^* \,|> c)\frac{\sqrt{T}\hat{\delta}_{ij}^*}{s_e}\right\} - \psi^{BG}(t_i)\frac{\sqrt{T}\hat{\delta}_i}{s_e}\right]$$

$$= \frac{s_e}{\sqrt{T}}\left[E^*\left\{1\left(|\, t_i + z_{ij}^* \,|> c\sqrt{\xi_j^*}\right)\left(t_i + z_{ij}^*\right)\right\} - \psi^{BG}(t_i)t_i\right].$$

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

$$\rho_{2i} = \frac{s_e}{\sqrt{T}} E^*\left\{\frac{1}{8}\left[t_i\psi^{BG\prime\prime}(t_i,\tilde{c})\tilde{c}^2 - t_i\psi^{BG\prime}(t_i,\tilde{c})\tilde{c}\right]\tilde{\xi}^{-2}(\xi_j^* - 1)^2\right\} \quad (21)$$

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

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

$$\leq \frac{s_e}{\sqrt{T}}\frac{1}{8}\sup_u \left|t_i\psi^{BG\prime\prime}(t_i,u)u^2 - t_i\psi^{BG\prime}(t_i,u)u\right| E^*\left[\tilde{\xi}^{-2}(\xi_j^* - 1)^2\right]$$

$$\leq \frac{s_e}{\sqrt{T}}\frac{1}{8}\sup_u \left|t_i\psi^{BG\prime\prime}(t_i,u)u^2 - t_i\psi^{BG\prime}(t_i,u)u\right| \sqrt{E^*(\xi_j^* - 1)^4}\sqrt{E^*\tilde{\xi}^{-4}}\,. \quad (22)$$

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 $T - n \geq 8$ we have that

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

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

$$\begin{aligned} |t\psi^{BG\prime}(t,u)u| &\leq 2\sup_u u^2\phi(t+u) \\ &\leq 2\sup_v (v-t)^2\phi(v) \\ &\leq 2[(\sup_v v^2\phi(v)) + 2t\sup_v|v\phi(v)| + t^2\sup_v\phi(v)] \\ &= 2(h_2 + 2h_1 t + h_2 t^2) \qquad (25) \end{aligned}$$

where $h_m = m^{m/2}e^{-m/2}/\sqrt{2\pi}$. Similar calculations provide a bound on $|t\psi^{BG\prime\prime}(t,u)u^2|$ which, combined with the bound in (25), yields

$$\sup_u \left| t_i\psi^{BG\prime\prime}(t_i,u)u^2 - t_i\psi^{BG\prime}(t_i,u)u \right|$$

$$\leq 2[(2h_2 + h_4) + (4h_1 + 3h_3)|t_i| + (2h_0 + 3h_2)\, t_i^2 + h_1|t_i|^3]$$

$$\leq 14h_4 \sum_{m=0}^{3} |\, t_i \,|^m ,$$

where the final equality uses $h_i < h_m$ for $i < m$ and $m > 1$. Substituting this bound, (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 $(T\text{–}n)\, s_e^2 \leq \sum Y_t^2$, the third uses Cauchy-Schwarz, and the rate uses $a_{T-n} \sim o[(T\text{–}n)^{-2}]$, $b_{T-n} \sim O(1)$, (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	$\ln(Z_t/Z_{t-1}) - \ln(Z_{t-1}/Z_{t-2})$	$h^{-1}\{\ln(Z_{t+\mathrm{h}}/Z_t)\} - \ln(Z_t/Z_{t-1})$

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), Relative to the AR(4) Forecast, Estimated by Cross-Validation, by Forecasting Method, *h* = 1

Method	No. est'd shrinkage parameters	Percentiles				
		0.050	0.250	0.500	0.750	0.950
AR(4)	0	1.000	1.000	1.000	1.000	1.000
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 relative RMSEs, relative to the AR(4), over the 143 series in the data set, for the forecasting method given in the first column (that is, the tabulated distribution is the distribution over series of the ratio of the candidate forecast RMSE relative to the AR(4) forecast RMSE). Forecasts and RMSEs were computed by cross-validation as described in Section 3.2.

Table 3
Distributions of Relative RMSE by Forecasting Method, Relative to DFM-5, Estimated by Cross-Validation, *h* = 1, 2, and 4

(a) *h* = 1

Method	Percentiles				
	0.050	0.250	0.500	0.750	0.950
AR(4)	0.983	1.013	1.047	1.127	1.194
OLS	0.972	1.035	1.072	1.102	1.154
DFM-5	1.000	1.000	1.000	1.000	1.000
Pretest	0.941	0.975	0.996	1.011	1.023
Bagging	0.952	0.985	1.001	1.013	1.029
BMA	0.951	0.980	0.996	1.010	1.027
Logit	0.934	0.968	0.989	1.005	1.021

(b) *h* = 2

Method	Percentiles				
	0.050	0.250	0.500	0.750	0.950
AR(4)	0.971	1.017	1.054	1.149	1.244
OLS	0.962	1.014	1.060	1.093	1.153
DFM-5	1.000	1.000	1.000	1.000	1.000
Pretest	0.932	0.972	0.990	1.006	1.023
Bagging	0.932	0.975	0.991	1.012	1.028
BMA	0.928	0.970	0.989	1.008	1.026
Logit	0.923	0.962	0.984	1.002	1.019

(c) *h* = 4

Method	Percentiles				
	0.050	0.250	0.500	0.750	0.950
AR(4)	0.956	1.018	1.069	1.171	1.280
OLS	0.933	1.007	1.067	1.119	1.175
DFM-5	1.000	1.000	1.000	1.000	1.000
Pretest	0.913	0.955	0.987	1.017	1.043
Bagging	0.926	0.959	0.992	1.021	1.047
BMA	0.920	0.955	0.986	1.013	1.049
Logit	0.909	0.947	0.985	1.008	1.037

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. RMSEs are relative to the DFM-5 forecast RMSE. All forecasts are direct.

Table 4

Distributions of Relative RMSE for 1985-2008 by Forecasting Method, Relative to DFM-5, Estimated by Cross-Validation, *h* = 1, 2, and 4

(a) *h* = 1

Method	Percentiles				
	0.050	0.250	0.500	0.750	0.950
AR(4)	0.962	0.991	1.019	1.076	1.186
OLS	0.966	1.038	1.084	1.159	1.244
DFM-5	1.000	1.000	1.000	1.000	1.000
Pretest	0.922	0.965	0.990	1.019	1.059
Bagging	0.915	0.966	0.995	1.019	1.047
BMA	0.909	0.965	0.991	1.013	1.047
Logit	0.908	0.957	0.987	1.012	1.050

(b) *h* = 2

Method	Percentiles				
	0.050	0.250	0.500	0.750	0.950
AR(4)	0.926	0.979	1.005	1.053	1.183
OLS	0.948	1.009	1.080	1.138	1.218
DFM-5	1.000	1.000	1.000	1.000	1.000
Pretest	0.920	0.963	0.987	1.019	1.070
Bagging	0.913	0.960	0.983	1.016	1.049
BMA	0.910	0.953	0.983	1.006	1.052
Logit	0.907	0.949	0.976	1.010	1.046

(c) *h* = 4

Method	Percentiles				
	0.050	0.250	0.500	0.750	0.950
AR(4)	0.880	0.957	0.998	1.037	1.152
OLS	0.897	0.964	1.051	1.113	1.161
DFM-5	1.000	1.000	1.000	1.000	1.000
Pretest	0.871	0.937	0.977	1.010	1.078
Bagging	0.877	0.938	0.968	1.007	1.051
BMA	0.868	0.924	0.964	0.999	1.044
Logit	0.859	0.922	0.964	0.998	1.038

Notes: Entries are computed as in Table 3, except that the RMSEs are calculated only for forecasts from 1985 – 2008. See the notes to Table 3.

Table 5

Distributions of Relative RMSE for 1985-2008 by Forecasting Method, Relative to DFM-5, Estimated by Rolling Forecasts, *h* = 1, 2, and 4

(a) *h* = 1

Method	Percentiles				
	0.050	0.250	0.500	0.750	0.950
AR(4)	0.918	0.979	1.007	1.041	1.144
OLS	0.968	1.061	1.110	1.179	1.281
DFM-5	1.000	1.000	1.000	1.000	1.000
Pretest	0.966	1.007	1.048	1.091	1.144
Bagging	0.938	0.996	1.022	1.060	1.104
BMA	0.921	0.993	1.014	1.053	1.103
Logit	0.941	0.999	1.027	1.071	1.120

(b) *h* = 2

Method	Percentiles				
	0.050	0.250	0.500	0.750	0.950
AR(4)	0.889	0.958	0.990	1.025	1.134
OLS	0.963	1.024	1.087	1.135	1.231
DFM-5	1.000	1.000	1.000	1.000	1.000
Pretest	0.957	1.003	1.030	1.082	1.156
Bagging	0.931	0.982	1.011	1.043	1.106
BMA	0.918	0.976	1.009	1.038	1.106
Logit	0.937	0.988	1.019	1.052	1.116

(c) *h* = 4

Method	Percentiles				
	0.050	0.250	0.500	0.750	0.950
AR(4)	0.879	0.945	0.980	1.020	1.107
OLS	0.942	1.015	1.066	1.113	1.194
DFM-5	1.000	1.000	1.000	1.000	1.000
Pretest	0.934	1.011	1.048	1.084	1.128
Bagging	0.924	0.984	1.016	1.052	1.094
BMA	0.898	0.979	1.014	1.047	1.086
Logit	0.924	0.982	1.022	1.064	1.120

Notes: Entries are computed as in Table 4, except that the RMSEs are calculated for rolling pseudo out-of-sample forecasts. See the notes to Table 4.

Table 6

Median Relative RMSE, Relative to DFM-5, Conditional on the Forecasting Method Improving on AR(4), Estimated by Rolling Forecasts, *h* = 1, 2, and 4

h	OLS	DFM-5	PT	Bag	BMA	Logit
h = 1	1.087 (13)	1.000 (85)	1.015 (33)	1.007 (45)	1.009 (51)	1.014 (43)
h = 2	1.002 (17)	1.000 (59)	1.008 (32)	0.986 (49)	0.989 (46)	0.998 (39)
h = 4	1.000 (18)	1.000 (53)	1.012 (29)	1.007 (40)	1.007 (39)	0.997 (36)

Notes: Entries are the relative RMSE of the column forecasting method, relative to DFM-5, computed for those series for which the column forecasting method has an RMSE less than the AR(4) forecast. The number of such series appears in parentheses below the relative RMSE. All RMSEs are computed by cross validation.

Table 7

Two Measures of Similarity of Forecast Performance, Cross-Validation, *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 cross-validation 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 8

Distribution of Root Mean Square Values of Shrinkage Function ψ, Cross-Validation, 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 9

Distribution of Fraction of Mean-Squared Variation of ψ Placed on the First Five Principle Components among Series with Root Mean Square Shrinkage Functions > 0.05, Cross-Validation, 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 10

Median RMSE by Forecasting Method and by Category of Series, Relative to AR(4), Cross-Validation Estimates

(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 10, 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.

Table 11
Median RMSE by Forecasting Method and by Category of Series, Relative to DFM-5, Cross-Validation Estimates

(a) *h* = 1

Category	**Brief description**	AR(4)	OLS	DFM-5	Pretest	Bagging	BMA	Logit
1	GDP components	1.093	1.056	1.000	0.989	0.995	0.992	0.987
2	IP	1.129	1.076	1.000	1.002	1.006	1.001	0.998
3	Employment	1.158	1.092	1.000	1.000	1.007	1.005	0.999
4	Unempl. rate	1.250	1.108	1.000	0.998	0.998	0.994	0.985
5	Housing	1.062	1.011	1.000	0.972	0.972	0.970	0.960
6	Inventories	1.112	1.053	1.000	1.010	1.016	1.009	0.994
7	Prices	1.020	1.049	1.000	0.995	1.002	0.992	0.981
8	Wages	1.001	0.990	1.000	0.968	0.987	0.970	0.961
9	Interest rates	1.021	1.068	1.000	0.996	1.008	0.999	0.996
10	Money	1.049	1.044	1.000	0.975	0.988	0.986	0.961
11	Exchange rates	0.985	1.074	1.000	0.983	0.982	0.979	0.966
12	Stock prices	1.017	1.075	1.000	1.011	1.010	1.006	1.007
13	Cons. exp.	1.023	1.133	1.000	1.019	1.023	1.023	1.019

(b) *h* = 2

Category	**Brief description**	AR(4)	OLS	DFM-5	Pretest	Bagging	BMA	Logit
1	GDP components	1.078	1.055	1.000	0.977	0.989	0.983	0.980
2	IP	1.159	1.064	1.000	1.000	1.000	1.005	0.993
3	Employment	1.161	1.080	1.000	0.997	0.998	0.999	0.989
4	Unempl. rate	1.333	1.156	1.000	0.970	0.973	0.971	0.970
5	Housing	1.053	0.991	1.000	0.980	0.980	0.978	0.979
6	Inventories	1.148	1.087	1.000	1.020	1.017	1.021	1.017
7	Prices	1.024	1.073	1.000	0.999	1.007	1.001	0.995
8	Wages	0.993	0.937	1.000	0.933	0.934	0.927	0.929
9	Interest rates	1.051	0.985	1.000	0.976	0.981	0.974	0.969
10	Money	1.072	1.057	1.000	0.993	0.991	0.993	0.987
11	Exchange rates	0.985	1.025	1.000	0.973	0.973	0.963	0.952
12	Stock prices	1.024	1.038	1.000	0.990	0.989	0.984	0.980
13	Cons. exp.	1.038	1.193	1.000	0.996	1.024	1.014	0.996

Table 11, continued

(c) *h* = 4

Category	**Brief description**	AR(4)	OLS	DFM-5	Pretest	Bagging	BMA	Logit
1	GDP components	1.103	1.044	1.000	0.987	0.990	0.978	0.976
2	IP	1.202	1.139	1.000	1.019	1.034	1.031	1.019
3	Employment	1.180	1.112	1.000	0.988	0.996	1.000	0.989
4	Unempl. rate	1.312	1.170	1.000	0.970	0.976	0.973	0.970
5	Housing	1.061	0.971	1.000	0.933	0.933	0.933	0.929
6	Inventories	1.149	1.068	1.000	1.001	1.017	1.020	1.003
7	Prices	1.038	1.061	1.000	0.987	0.992	0.987	0.985
8	Wages	0.979	0.963	1.000	0.951	0.954	0.954	0.942
9	Interest rates	1.046	1.044	1.000	0.992	0.991	0.985	0.985
10	Money	1.100	1.061	1.000	0.998	1.004	0.992	0.992
11	Exchange rates	0.966	1.004	1.000	0.929	0.942	0.935	0.926
12	Stock prices	1.027	1.004	1.000	0.993	0.991	0.984	0.987
13	Cons. exp.	1.035	1.112	1.000	0.989	1.004	0.994	0.989

Notes: entries are the median relative RMSE among the relative RMSEs for the series among the row category, using the column forecasting method.

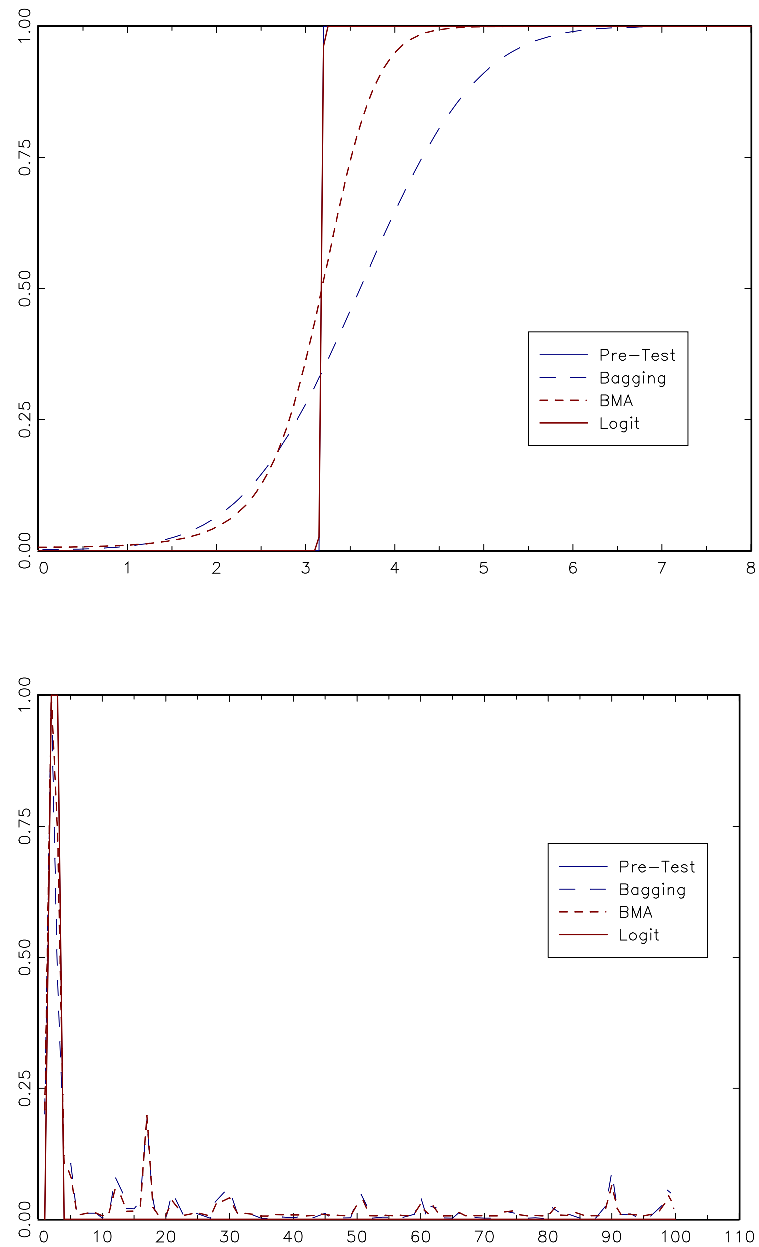
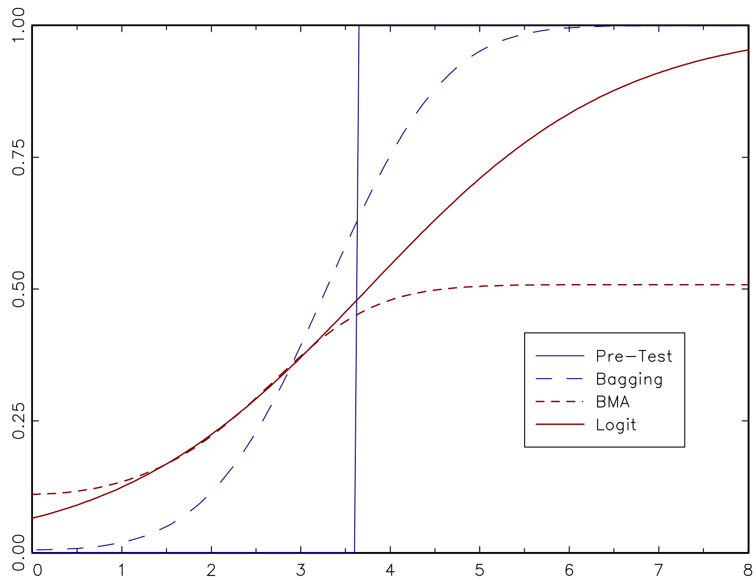


Figure 1
Estimated shrinkage functions (upper panel) and weights $\psi(t_i, \hat{\theta})$ on ordered principle components 1-100: Total employment growth, $h = 1$



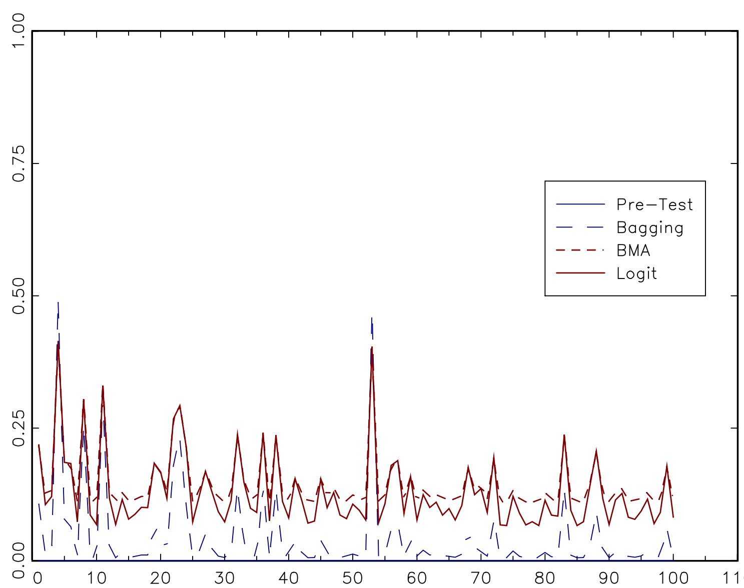


Figure 2
Estimated shrinkage functions (upper panel) and weights $\psi(t_i, \hat{\theta})$ on ordered principle components 1-100: percentage change of S&P500 Index, $h = 1$