Seeing Inside the Black Box: Using Diffusion Index Methodology to Construct Factor Proxies in Largescale Macroeconomic Time Series Environments*

Nii Ayi Armah and Norman R. Swanson
Rutgers University

February 2007

Abstract

In economics, common factors are often assumed to underlie the co-movements of a set of macroeconomic variables. For this reason, many authors have used estimated factors in the construction of prediction models. In this paper, we begin by surveying the extant literature on diffusion indexes. We then outline a number of approaches to the construction of factor proxies using the statistics developed in Bai and Ng (2006a,b). Our approach to factor proxy selection is examined via a small Monte Carlo experiment, where good size and power properties are reported, and via a large set of prediction experiments using the panel dataset of Stock and Watson (2005). One of our main findings is that our "smoothed" approaches to factor proxy selection appear to yield predictions that are often superior not only to a benchmark factor model, but also to simple linear time series models which are generally difficult to beat in forecasting competitions. In some sense, by using our approaches to predictive factor proxy selection, one is able to open up the "black box" often associated with factor analysis, and to identify actual variables that can serve as primitive building blocks for (prediction) models of a host of macroeconomic variables. Our findings suggest that important observable varibles include: various S&P500 variables, including stock price indices and dividend series; a 1-year bond rate; various housing activity variables; industrial production; and an exchange rate variable.

JEL classification: C22, C33, C51.
Keywords: diffusion index, factor, forecast, macroeconometrics, parameter estimation error, proxy.

----

* Department of Economics, Rutgers University, 75 Hamilton Street, New Brunswick, NJ 08901, USA, armah@econ.rutgers.edu and nswanson@econ.rutgers.edu. A great many thanks are owed to Mark Watson for making the data used in this paper available for public consumption. Thanks are also owed to Valentina Corradi, Roger Klein, Serena Ng, and participants of the conference on "Empirical Methods in Macroeconomics and Finance" at Bocconi University in October 2003, where many interesting discussions led to the idea for this paper. Finally, the authors would like to thank the editor, Marcelo C. Medeiros, whose many emails on the subject served to spur this research forward. Swanson thanks the Rutgers University Research Council for financial support.

1 Introduction

The idea that individual economic variables can be forecast with some precision by refining the information from a large panel of data into a small set of estimated factors to be used in subsequent prediction model specification is intriguing, as it suggests that there is a small set of crucial latent factors that can be used to measure common comovements in a large set of macroeconomic variables. Indeed, this sort of idea is quite consistent with the notion that a small set of underlying shocks are responsible for the dynamic behavior implicit in stochastic dynamic general equilibrium models. Moreover, if one adheres to the notion that certain of these factors can be proxied using observable economic variables, this idea is also consistent with the notion espoused on the Federal Reserve Bank of New York's website, where it is stated that: *"In formulating the nation's monetary policy, the Federal Reserve considers a number of factors, including the economic and financial indicators which follow, as well as the anecdotal reports compiled in the Beige Book. Real Gross Domestic Product (GDP); Consumer Price Index (CPI); Nonfarm Payroll Employment Housing Starts; Industrial Production/Capacity Utilization; Retail Sales; Business Sales and Inventories; Advance Durable Goods Shipments, New Orders and Unfilled Orders; Lightweight Vehicle Sales; Yield on 10-year Treasury Bond; S&P 500 Stock Index M2"* (see *http://www.newyorkfed.org/education/bythe.html).* It is thus not surprising that there have been numerous interesting papers in the are of diffusion index models in recent years. Indeed, the recent literature on this subject is rich and diverse, and a very few of the most important papers include: Bai (2003), Bai and Ng (2002,2005,2006a,b,c,d), Boivin and Ng (2005), Connor and Korajczyk (1993), Ding and Hwang (1999), Forni, Hallin, Lippi, and Reichlin (2000,2005), Forni and Reichlin (1996,1998), Geweke (1977), and Stock and Watson (1996,1998,1999,2002a,b,2004a,b,2005).

In this paper, our purpose is twofold. first, we provide a review of the extant literature, with careful emphasis on the implementation of factor estimation and prediction using the methods of Bai and Ng as well as Stock and Watson. We then outline a simple methodology for the construction of factor proxies for use in prediction models, where our proxies are observable economic variables. In this sense, we attempt to look inside the "black box". Thus, our main contribution is to add to the literature on prediction using factor models. The methodology that we outline is very straightforward, and is based upon application of the $A(j)$ and $M(j)$ statistics developed in Bai and Ng (2006a,b).

Using the approach of Stock and Watson (2002a,b), forecasting using so-called diffusion indices follows a two-step procedure. First, the factors are estimated from a large panel of predictors, X, using the method of principal components. Second, the estimated factors are used to forecast the variable of interest, y_{t+1}. Stock and Watson (2002a) demonstrate that diffusion index forecasts based on the use of factors instead of large panels yield encouraging results. Bai and Ng (2006a) however point out that the regressors (factors) in the diffusion index model are estimated, hence substantially increasing the forecast error variance. In a related paper, Bai and Ng (2006b), examine whether observable economic variables can serve as proxies for the underlying unobserved factors. In particular, they develop a methodology based upon the use of their $A(j)$ and $M(j)$ statistics in order to determine whether a group of observed variables yields precisely the same information as that contained in the latent factors. Thus, in some sense, Bai and Ng already look inside the "black box". Our approach is to take their argument one step further, and to argue that if observable economic variables are indeed good proxies of the unobserved factors, then these proxies can be used in place of the factors in the diffusion index model, which is in turn used for prediction. Once the set of factor proxies is fixed, we effectively eliminate the incremental increase in forecast error variance associated with the use of estimated factors. Along these lines, we consider various versions of the $A(j)$ and $M(j)$ statistics, including "smoothed" versions that are used to pre-select a set of factor proxies, prior to the ex-ante construction of a sequence of predictions based upon recursively estimated prediction models.

In a Monte Carlo experiment, we show that the $A(j)$ and $M(j)$ statistics have good finite sample properties. We additionally carry out a large variety of prediction experiments using the macroeconomic dataset of Stock and Watson (2005). In these experiments, we predict a number of price and income variables, including industrial production, real personal income less transfers, real manufacturing and trade sales, the number of employees on nonagricultural payrolls, the consumer price index, the personal consumption expenditure implicit price deflator, and the producer price index for finished goods. Using recursively estimated models, we construct $h = 1, 3, 12$, and 24 step ahead forecasts. We show that the $A(j)$ and $M(j)$ statistics appear to offer an interesting means by which factor proxies for later use in prediction models can be chosen. Indeed, our "smoothed" approaches to factor proxy selection appear to yield predictions that are often mean square forecast error "superior" not only relative to a benchmark factor model, but also to simple linear time series models which are often difficult to beat in forecasting competitions. Furthermore, our methods based on the use of the $A(j)$ statistic appear to perfrom better than those based on the $M(j)$ statistic. Finally, we provide evidence that: (i) versions of our factor proxy selection method that use only a single factor proxy are preferred to those based on the use of $\widehat{k}$ proxies, where $\widehat{k}$ is a consistent estimate of the true number of factors; and (ii) while our "smoothed" proxy selection method is clearly superior for $h = 1, 3$, and 12, the method breaks down at the longest forecast horizon that we consider (i.e. $h = 24$). For the longest horizons, the usual factor approach to prediction (e.g. that used by Stock and Watson (2002a,b) dominates.

By using our approach to predictive factor proxy selection, one is able to further open up the "black box" often associated with factor analysis, and to identify actual variables that can serve as primitive building blocks for (prediction) models of a host of macroeconomic variables. Along these lines, our empirical analysis suggest that important observable varibles include: various S&P500 variables, including stock price indices and dividend series; a 1-year bond rate; various housing activity variables; industrical production; and an exchange rate variable.

We leave the analysis of our methodology to contexts where factor analysis methods other than those of Stock and Watson (2002a,b) are used to future research; although it should perhaps be noted that Boivin and Ng (2005) compare alternative factor based forecast methodologies, and conclude that when the dynamic structure is unknown and the model is characterized by complex dynamics, the approach of Stock and Watson performs favorably.

The rest of the paper is organized as follows. In Section 2 we review the diffusion index literature, with some focus on the methods that are used in our Monte Carlo and empirical experiments. In Section 3 we discuss the use of factor proxies, including a discussion of the Bai and Ng (2006a,b) tests, and a discussion of the methodological approach to the construction and use of factor proxies for prediction. Section 4 contains a summary of the empirical methodology used in the paper, and Section 5 summarizes the data used. In Section 6, the results of a small Monte Carlo experiment studying the finite sample properties of the Bai and Ng (2006a,b) tests are presented, and in Section 7 we summarize our empirical findings. Finally, in Section 8 we briefly discuss the most recent advances in the diffusion index methodology; and concluding remarks are gathered in Section 9.

2 Review: Diffusion Index Models and the Principle Components Approach to Estimation

2.1 The diffusion index model

Following Stock and Watson (2002a,b), let y_{t+1}be the series we wish to forecast and X_t be an N-dimensional vector of predictor variables, for $t = 1, \ldots, T$. Assume that (y_{t+1}, X_t) has a dynamic factor model representation with $\overline{r}$ common dynamic factors, f_t. Hence, f_t is an $\overline{r} \times 1$ vector. The dynamic factor model is written as:

$$y_{t+h} = \alpha(L) f_t + \beta' W_t + \varepsilon_{t+h} \tag{1}$$

and

$$x_{it} = \lambda_i(L) f_t + e_{it}, \tag{2}$$

for $i = 1, 2, \ldots, N$, where W_t is a vector l of other observable variables with $l << N$, such as contemporaneous and lagged values of y_t; $h > 0$ is the lead time between information available and the dependent variable; x_{it} is a single datum for a particular predictor variable; e_{it} is the idiosyncratic shock component of x_{it}; and $\alpha(L)$ and $\lambda_i(L)$ are lag polynomials in nonnegative powers of L. In general, dynamic factor models can be transformed into static factor models. In Stock and Watson (2002a), the lag polynomials $\alpha(L)$ and $\lambda_i(L)$ are modeled as $\alpha(L) = \sum_{j=0}^{q} \alpha_j L^j$ and $\lambda_i(L) = \sum_{j=0}^{q} \lambda_{ij} L^j$. The finite order of the lag polynomials allows us to rewrite (1) and (2) as:

$$y_{t+h} = \alpha' F_t + \beta' W_t + \varepsilon_{t+h} \tag{3}$$

and

$$x_{it} = \Lambda_i' F_t + e_{it}, \tag{4}$$

where $F_t = (f_t', \ldots, f_{t-q}')'$ is an $r \times 1$ vector, with $r = (q+1)\overline{r}$ and α is an $r \times 1$ vector. Here, r is the number of static factors (i.e. the number of elements in F_t). Additionally, $\Lambda_i = (\lambda_{i0}', \ldots, \lambda_{iq}')'$ is a vector of factor loadings on the r static factors, where λ_{ij} is an $\overline{r} \times 1$ vector for $j = 1, \ldots, q$ and $\beta = (\beta_1, \ldots, \beta_l)'$. Alternatively, from (2), the dynamic factor model can be represented as:

$$\begin{aligned} x_{it} &= \lambda_{i0}' f_t + \lambda_{i1}' f_{t-1} + \ldots + \lambda_{iq}' f_{t-q} + e_{it} && (5) \\ &= \lambda_i'(L) f_t + e_{it} && (6) \end{aligned}$$

and:

$$\lambda_i(L) = \lambda_{i0} + \lambda_{i1} L^1 + \ldots + \lambda_{iq} L^q.$$

For complete details, see Bai and Ng (2005). Now, (6) can be written in the static form (4) where F_t and Λ_i are defined as above. The static factor model refers to the contemporaneous relationship between x_{it} and F_t. One major advantage of the static representation of the dynamic factor model is it enables us to use principal components to estimate the factors. This involves estimating F_t using an eigenvalue-eigenvector decomposition of the sample covariance matrix of the data. It is worth noting that the use of principal components to estimate the factors cannot be done with infinitely distributed lags of the factors (see Stock and Watson (2002a)). Ding and Hwang (1999), Forni et al. (2000), Stock and Watson (2002b), Bai and Ng (2002) and Bai (2003) showed that the space spanned by both the static and dynamic factors can be consistently estimated when N and T are both large. For forecasting purposes, little is gained from a clear distinction between the static and the dynamic factors. However, many economic analyses hinge on the ability to isolate the primitive shocks or the number of dynamic factors (see Bai and Ng (2005)). If the idiosyncratic errors $e_t = (e_{1t}, \ldots, e_{Nt})'$, are cross-sectionally independent and iid over time, then (4) is the classical factor analysis model. The literature on principal components and factor models is quite extensive. Geweke (1977) and Sargent and Sims (1977) were among the first to extend the classical factor analysis model to dynamic models.

Following Bai and Ng (2002), let $\underline{X}_i$ be a $T \times 1$ vector of observations for the ith time-series variable. For a given cross-section i, we have:

$$\underset{(T\times 1)}{\underline{X}_i} = \underset{(T\times r)}{F^0}\underset{(r\times 1)}{\Lambda_i} + \underset{(T\times 1)}{\underline{e}_i}$$

where $\underline{X}_i = (X_{i1}, \ldots, X_{iT})'$, $F^0 = (F_1, \ldots, F_T)'$ and $\underline{e}_i = (e_{i1}, \ldots, e_{iT})'$. The whole panel of data $X = (\underline{X}_1, \ldots, \underline{X}_N)$ can consequently be represented as:

$$\underset{(T\times N)}{X} = \underset{(T\times r)}{F^0}\underset{(r\times N)}{\Lambda'} + \underset{(T\times N)}{e},$$

where $\Lambda = (\Lambda_1, \ldots, \Lambda_N)'$ and $e = (\underline{e}_1, \ldots, \underline{e}_N)$. In most recent applications of the classical factor model and its dynamic generalization, the dimension of X is small, so the issue of the consistent estimation of the factors is not relevant. However, Connor and Korajczyk (1986, 1988, 1993) note that in static models, the factors can be consistently estimated by principal components as

$N \to \infty$ even if e_{it} is weakly cross-sectionally correlated. Similarly, for dynamic factor models, Forni and Reichlin (1996,1998) and Forni, Hallin, Lippi and Reichlin (2000) discuss consistent estimation of the factors when $N, T \to \infty$. In a predictive context, Ding and Hwang (1999) analyze the properties of forecasts constructed from principal components when N and T are large. They perform their analysis under the assumption that the error processes $\{e_{it}, \varepsilon_{t+h}\}$ are cross-sectionally and serially iid. This assumption is inappropriate for macroeconomic models, particularly when multiperiod forecasts are involved. Stock and Watson (2002b) emphasize that most macroeconomic variables are serially correlated and some like alternative measures of the money supply may be cross-correlated even after the aggregate factors are controlled for. We work with high-dimensional factor models that allow both N and T to tend to infinity, and in which e_{it} may be serially and cross-sectionally correlated so that the covariance matrix of $e_t = (e_{1t}, \ldots, e_{Nt})$ does not have to be a diagonal matrix. We will also assume $\{F_t\}$ and $\{e_{it}\}$ are two groups of mutually independent stochastic variables. Furthermore, it is well known that for $\Lambda F_t = \Lambda Q Q^{-1} F_t$, a normalization is needed in order to uniquely define the factors, where Q is a nonsingular matrix. Now, assuming that $(\Lambda'\Lambda/N) \to I_r$, we restrict Q to be orthonormal, for example. This assumption, together with others noted in Stock and Watson (2002b), enables us to identify the factors up to a change of sign and consistently estimate them up to an orthonormal transformation. Forecasts of y_{T+h} based on (3) and (4) involve a two step procedure because both the regressors and coefficients in the forecasting equations are unknown. The data sample $\{X_t\}_{t=1}^T$ are first used to estimate the factors, $\{\widetilde{F}_t\}_{t=1}^T$ by means of principal components. With the estimated factors in hand, we obtain the estimators $\widehat{\alpha}$ and $\widehat{\beta}$ by regressing y_{t+1} onto $\widetilde{F}_t$ and the observable variables in W_t. Of note is that if $\sqrt{T}/N \to 0$, then the generated regressor problem does not arise, in the sense that least squares estimates of $\widehat{\alpha}$ and $\widehat{\beta}$ are $\sqrt{T}$ consistent and asymptotically normal (see Bai and Ng (2005)).

2.2 Common factor estimation using principal components

The problem of obtaining the necessary estimates in (4) would be simplified if we knew F^0. Then Λ_i could be estimated via least squares by setting $\{x_{it}\}_{t=1}^T$ to be the dependent variable and $\{F_t\}_{t=1}^T$ to be the explanatory variable. On the other hand, if Λ were known, F_t could be estimated by regressing $\{x_{it}\}_{i=1}^N$ on $\{\Lambda_i\}_{i=1}^N$. Since the common factors are not observed, in the regression analysis of (4), we replace F_t by $\widetilde{F}_t$, estimates that span the same space as F_t when $N, T \to \infty$. Estimation of these common factors from large panel data sets of macroeconomic variables can be carried out using principal component analysis. We refer the reader to Stock and Watson (1998, 2002a, 2002b, 2004a, 2004b) and Bai and Ng (2002) for a detailed explanation of this procedure. Note further that predictions can be constructed using either the Stock and Watson approach discussed in this paper, or using the dynamic approach of Forni et al. (2005). The reader is referred to Boivin and Ng (2005) for a comparison of alternative factor based forecast methodologies. As mentioned above, in their paper they conclude that when the dynamic structure is unknown and the model is characterized by complex dynamics, the approach of Stock and Watson performs favorably.

As noted earlier F_t and λ_i are not separately identified, but rather identifiable only up to a square matrix. Stock and Watson (1998) further demonstrate that when principal components is used, the factors remain consistent even when there is some time variation in Λ and small amounts of data contamination, so long as the number of variables in the panel data set or the number of predictors is very large (i.e. $N >> T$). In this paper, we only give an outline of how principal component analysis is carried out, with particular emphasis on those features of the analysis that allow us to carry out our prediction experiments using the $A(j)$ and $M(j)$ statistics of Bai and Ng (2006b). Let k ($k < \min\{N, T\}$) be an arbitrary number of factors, Λ^k be the $N \times r$ matrix of factor loadings, $(\Lambda_1^k, \ldots, \Lambda_N^k)'$, and F^k be a $T \times r$ matrix of factors $(F_1^k, \ldots, F_T^k)'$. From (4), estimates of Λ_i^k and F_t^k are obtained by solving the optimization problem:

$$V(k) = \min_{\Lambda^k, F^k} (NT)^{-1} \sum_{i=1}^{N} \sum_{t=1}^{T} (x_{it} - \Lambda_i^{k\prime} F_t^k)^2 \tag{7}$$

Let $\widetilde{F}^k$ and $\widetilde{\Lambda}^k$ be the minimizers of equation (7). Since Λ^k and F^k are not separately identifiable, if $N > T$, a computationally expedient approach would be to concentrate out $\widetilde{\Lambda}^k$ and minimize (7) subject to the normalization $F^{k\prime}F^k/T = I_k$. Minimizing (7) is equivalent to maximizing $tr[F^{k\prime}(XX')F^k]$. This optimization is solved by setting $\widetilde{F}^k$ to be the matrix of the k eigenvectors of XX' that correspond to the k largest eigenvalues of XX'. Note that $tr[\cdot]$ represents the matrix trace. The superscript in Λ^k and F^k signifies the use of k factors in the estimation and the fact that the estimates will depend on k. Let $\widetilde{D}$ be a $k \times k$ diagonal matrix consisting of the k largest eigenvalues of XX'. The estimated factor matrix, denoted by $\widetilde{F}^k$, is $\sqrt{T}$ times the eigenvectors corresponding to the k largest eigenvalues of the $T \times T$ matrix XX'. Given $\widetilde{F}^k$ and the normalization $F^{k\prime}F^k/T = I_k$, $\widetilde{\Lambda}^{k\prime} = (\widetilde{F}^{k\prime}\widetilde{F}^k)^{-1}\widetilde{F}^{k\prime}X = \widetilde{F}^{k\prime}X/T$ is the corresponding factor loadings matrix. The solution to the optimization problem in (7) is not unique. If $N < T$, it becomes computationally advantageous to concentrate out $\overline{F}^k$ and minimize (7) subject to $\overline{\Lambda}^{k\prime}\overline{\Lambda}^k/N = I_k$.

This minimization is the same as maximizing $tr[\Lambda^{k\prime}X'X\Lambda]$, the solution of which is to set $\overline{\Lambda}^k$ equal to the eigenvectors of the $N \times N$ matrix $X'X$ that correspond to its k largest eigenvalues. One can consequently estimate the factors as $\overline{F}^k = X'\overline{\Lambda}^k/N$. $\widetilde{F}^k$ and $\overline{F}^k$ span the same column spaces, hence for forecasting purposes, they can be used interchangeably depending on which one is more computationally efficient.

We now focus on how to consistently estimate the true number of factors, r, that underlie the panel data set. The reader is referred to Bai and Ng (2002) for a detailed exposition on the theoretical underpinnings of this procedure. Given $\widetilde{F}^k$ and $\widetilde{\Lambda}^k$, let:

$$\widehat{V}(k) = (NT)^{-1}\sum_{i=1}^{N}\sum_{t=1}^{T}(x_{it} - \widetilde{\Lambda}_i^{k\prime}\widetilde{F}_t^k)^2$$

be the sum of squared residuals from regressions of X_i on the k factors, $\forall i$. A penalty function for over fitting, $g(N,T)$, is chosen such that the loss function

$$IC(k) = \log(\widehat{V}(k)) + kg(N,T) \tag{8}$$

can consistently estimate r. Let $kmax$ be a bounded integer such that $r \leq k\max$. Bai and Ng (2002) propose three versions of the penalty function $g(N,T)$. Namely:

$$g_1(N,T) = \left(\frac{N+T}{NT}\right)\log\left(\frac{NT}{N+T}\right),$$

$$g_2(N,T) = \left(\frac{N+T}{NT}\right)\log C_{NT}^2,$$

and

$$g_3(N,T) = \left(\frac{\log(C_{NT}^2}{C_{NT}^2}\right),$$

all of which lead to consistent estimation of r. In our empirical implementation, we use $g_2(N,T)$, hence (8) becomes:

$$IC(k) = \log(\widehat{V}(k)) + k(\frac{N+T}{NT})\log C_{NT}^2$$

where $C_{NT} = \min\{\sqrt{N}, \sqrt{T}\}$. The consistent estimate of the true number of factors is then:

$$\widehat{k} = \arg\min_{0\leq k\leq k\max} IC(k), \tag{9}$$

and $\lim_{N,T\to\infty}\text{Prob}[\widehat{k} = r] = 1$ if $g(N,T) \to 0$ and $C_{NT}^2 \cdot g(N,T) \to \infty$ as $N,T \to \infty$, as shown in Bai and Ng (2002).

3 Using Proxies In Place of Factors for Prediction

In this section, we discuss various methods for replacing estimated factors with actual economic variables that can be thought of as proxies for the factors. The motivation behind this approach is that in the context of forecasting, estimation error associated with factor construction can be avoided to some extent by using directly observable economic variables as factor proxies. In contexts where N is very large, the gains to using proxies might be substantial. Of note is that Stock and Watson (2002b) show that the difference between feasible and unfeasible factor based forecasts converge in probability to zero as $N, T \to \infty$. Thus, our approach is meant to address finite sample prediction performance, as feasible factor based forecasts are asymptotically efficient.

3.1 Prediction using factors

Reconsider the general equation (3):

$$y_{t+h} = \alpha' F_t + \beta' W_t + \varepsilon_{t+h}$$

As mentioned above, and shown in Stock and Watson (200b) and Bai and Ng (2005), under a set of moment conditions on (ε, e, F^0) and an asymptotic rank condition on Λ, if the space spanned by F_t can be consistently estimated, then $\sqrt{T}$ consistent estimates of α and β are obtainable. Under a similar set of conditions, it is also possible to obtain $\min[\sqrt{N}, \sqrt{T}]$ consistent forecasts if $\sqrt{T/N} \to 0$ as $N, T \to \infty$. Let $z_t = (F_t', W_t')'$; $E(\varepsilon_{t+h}|y_t, z_t, y_{t-1}, z_{t-1}, \dots) = 0$, for any $h > 0$; and let z_t and ε_t be independent of the idiosyncratic errors e_{is}, $\forall i, s$. At this point, the k superscript on $\widetilde{F}_t^k$ is suppressed, because we assume we have consistently estimated the number of factors underlying the dataset. If F_t is observable and α and β are known, based on the above assumption that the mean of ε_{t+h} conditional on past information is zero, the conditional mean and minimum mean square error forecast of y_{T+h} is given by:

$$y_{T+h|T} = E(y_{T+h}|z_T, z_{T-1}, \dots) = \alpha' F_T + \beta' W_T \equiv \delta' z_T$$

Such a prediction is not feasible, however, since α, β and F_t are all unobserved. The feasible prediction that replaces the unknown objects by their estimates is:

$$\widehat{y}_{T+h|T} = \widehat{\alpha}' \widetilde{F}_T + \widehat{\beta}' W_T = \widehat{\delta}' \widehat{z}_T, \tag{10}$$

where $\widehat{z}_t = (\widetilde{F}_t', W_t')'$. Here, $\widehat{\alpha}$ and $\widehat{\beta}$ are the least squares estimates obtained from regressing y_{t+h} on $\widetilde{F}_t$ and W_t, $t = 1, \ldots, T-h$. The factors, F_t, are estimated from x_{it} by the method of principal components, as discussed above. As the objective is to forecast y_{T+h}, a crucial aspect of our analysis is the distribution of the forecast error. As explained in detail in Bai and Ng (2006a), since $y_{T+h} = y_{T+h|T} + \varepsilon_{T+h}$, it follows that the forecast error is:

$$\widehat{\varepsilon}_{T+h} = \widehat{y}_{T+h|T} - y_{T+h} = (\widehat{y}_{T+h|T} - y_{T+h|T}) + \varepsilon_{T+h}$$

If $\varepsilon_t \sim N(0, \sigma_\varepsilon^2)$, then:

$$\widehat{\varepsilon}_{T+h} \sim N(0, \sigma_\varepsilon^2 + var(\widehat{y}_{T+h|T})) \tag{11}$$

where

$$var(\widehat{y}_{T+h|T}) = \frac{1}{T}\widehat{z}_T' Avar(\widehat{\delta})\widehat{z}_T + \frac{1}{N}\widehat{\alpha}' Avar(\widetilde{F}_T)\widehat{\alpha}. \tag{12}$$

Here, $var(\widehat{y}_{T+h|T})$ reflects both parameter uncertainty and regressor uncertainty. In large samples, $var(\widehat{\varepsilon}_{T+h})$ is dominated by σ_ε^2. If we ignore $var(\widehat{y}_{T+h|T})$, σ_ε^2 alone will under-estimate the true forecast uncertainty for finite T and N. Let us now assume for a moment that F_t is observable. The feasible prediction of y_{T+h} would then be $\overline{y}_{T+h|T} = \overline{\alpha}' F_T + \overline{\beta}' W_T = \overline{\delta}' z_T$, where $\overline{\alpha}$ and $\overline{\beta}$ are the least squares estimates obtained from regressing y_{t+h} on F_t and W_t. Once again, since $y_{T+h} = y_{T+h|T} + \varepsilon_{T+h}$, the forecast error is:

$$\overline{\varepsilon}_{T+h} = \overline{y}_{T+h|T} - y_{T+h} = (\overline{y}_{T+h|T} - y_{T+h|T}) + \varepsilon_{T+h}$$

If $\varepsilon_t \sim N(0, \sigma_\varepsilon^2)$, then

$$\overline{\varepsilon}_{T+h} \sim N(0, \sigma_\varepsilon^2 + var(\overline{y}_{T+h|T})), \tag{13}$$

where

$$var(\overline{y}_{T+h|T}) = \frac{1}{T} z_T' Avar(\overline{\delta}) z_T. \tag{14}$$

Thus, and as discussed by Bai and Ng (2006a), when comparing $var(\overline{y}_{T+h|T})$ with $var(\widehat{y}_{T+h|T})$ it is clear that estimating the factors increases the forecast error variance, $var(\widehat{y}_{T+h|T})$, by $\frac{1}{N}\widehat{\alpha}' Avar(\widetilde{F}_T)\widehat{\alpha}$. Of course, if we could observe the factors instead of estimating them, we would reduce the forecast error variance from (11) to (13). In finite samples, this may yield important prediction error variance reduction. It is for this reason that we consider replacing the factors in (10) with observable variables that closely proxy the factors. The approach taken in order to do this involves implementing a "first stage" factor analysis in which proxies are formed using the $A(j)$ and $M(j)$ statistics of Bai and Ng (2006b). In a "second stage", the observable proxies are used in the construction of a prediction model. In this way, all estimation error associated with our factor analysis is essentially "hidden" in the first stage, and does not directly manifest itself in the "second stage" prediction models, predictions and prediction errors. Of course, issues related to "pre-testing" and sequential testing bias still arise. Nevertheless, in our prediction experiments we attempt to quantify through finite sample experiments the potential gains to using the "proxy" approach.

3.2 Using the $A(j)$ and $M(j)$ tests of Bai and Ng (2006b) to uncover factor proxies

For a detailed theoretical discussion of the results presented in this subsection, the reader is referred to Bai and Ng (2006b). Here, we draw heavily on aspects of that paper that are relevant to our empirical implementation. Note that while Bai and Ng (2006b) suggest using the $A(j)$ and $M(j)$ statistics to assess whether key business cycle indicators approximate factors, we use the $A(j)$ and $M(j)$ statistics to select factor proxies for subsequent use in prediction models.

We considered three methods for selecting the proxies. One, the $A(j)$ statistic depends on the actual size of a t-test. Another, the $M(j)$ test is based on a measure of the distance between observed variables and factor estimates thereof. The third is based on the confidence interval of an estimation error.

Suppose we observe G', a $(T \times m)$ matrix of observable economic variables that could potentially proxy the latent factors. At any given time t, any of the m elements of G_t $(m \times 1)$ will be a good proxy if it is a linear combination of the $r \times 1$ latent factors, F_t. Let G_{jt} be an element of the m vector G_t. The null hypothesis is that G_{jt} is an exact proxy, or more precisely, $\exists\, \theta_j$ $(r \times 1)$ such that $G_{jt} = \theta_j' F_t$. In order to implement all of the methods, consider the regression $G_{jt} = \gamma_j' \widetilde{F}_t + \rho$. Let $\widehat{\gamma}_j$ be the least squares estimate of γ_j and let $\widehat{G}_{jt} = \widehat{\gamma}_j' \widetilde{F}_t$. The test is carried out by constructing the following t-statistic:

$$\tau_t(j) = \frac{(\widehat{G}_{jt} - G_{jt})}{(\widehat{var}(\widehat{G}_{jt}))^{1/2}} \tag{15}$$

where

$$
\begin{aligned}
\widehat{var}(\widehat{G}_{jt}) &= \frac{1}{N}\widehat{\gamma}_j'\widetilde{D}^{-1}\left(\frac{\widetilde{F}'\widetilde{F}}{T}\right)\widetilde{\Gamma}_t\left(\frac{\widetilde{F}'\widetilde{F}}{T}\right)\widetilde{D}^{-1}\widehat{\gamma}_j \\
&= \frac{1}{N}\widehat{\gamma}_j'\widetilde{D}^{-1}\widetilde{\Gamma}_t\widetilde{D}^{-1}\widehat{\gamma}_j, \qquad (16)
\end{aligned}
$$

where $\widetilde{\Gamma}_t$ is defined below. The last step above is due to the normalization that $\widetilde{F}'\widetilde{F}/T = I_{\widehat{k}}$. Given the null hypothesis that $G_{jt} = \theta_j' F_t$ and that $\widehat{G}_{jt}$ converges to G_{jt} at rate $\sqrt{N}$, Bai and Ng (2006b) show that the limiting distribution of $\sqrt{N}(\widehat{G}_{jt} - G_{jt})$ is asymptotically normal and hence $\tau_t(j)$ has a standard normal limiting distribution. Consistent choices for the the $\widehat{k} \times \widehat{k}$ matrix $\widetilde{\Gamma}_t$ include the following:

$$
\widetilde{\Gamma}_t^1 = \frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{n}\widetilde{\Lambda}_i\widetilde{\Lambda}_j'\frac{1}{T}\sum_{t=1}^{T}\widetilde{e}_{it}\widetilde{e}_{jt}\ ,\ \ \forall\ t, \qquad (17)
$$

$$
\widetilde{\Gamma}_t^2 = \frac{1}{N}\sum_{i=1}^{N}\widetilde{e}_{it}^2\widetilde{\Lambda}_i\widetilde{\Lambda}_i', \qquad (18)
$$

and

$$
\widetilde{\Gamma}^3 = \widehat{\sigma}_e^2\frac{\widetilde{\Lambda}'\widetilde{\Lambda}}{N}, \qquad (19)
$$

where $\widehat{\sigma}_e^2 = \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\widetilde{e}_{it}^2$, $\widetilde{e}_{it} = x_{it} - \widetilde{\Lambda}_i'\widetilde{F}_t$ and $\frac{n}{\min[N,T]} \to 0$ as $N, T \to \infty$. In our Monte Carlo simulation and our empirical analysis, we choose $n = \min\{\sqrt{N}, \sqrt{T}\}$. Note that equation (17) allows cross-section correlation but assumes time-series stationarity of e_{it}. This covariance estimator is a HAC type estimator because it is robust to cross-correlation (see Bai and Ng (2006a) for complete details). Equation (18) allows for time-series heteroskedasticity, but assumes no cross-sectional correlation of e_{it}. Equation (19) assumes no cross-sectional correlation and constant variance, $\forall i$ and $\forall t$. For small cross-sectional correlation in e_{it}, Bai and Ng (2006a) found that constraining the correlations to be zero could sometimes be desirable. In this regard, they make the point that (18) and (19) are useful even if residual cross-correlation is genuinely present.

As mentioned earlier, $\tau_t(j)$ in (15) has a standard normal limiting distribution. Let Φ_ξ^τ be the ξ percentage point of the limiting distribution of $\tau_t(j)$. The hypothesis test based on the t-statistic in (15) enables us to determine whether an observed value of a candidate variable is a good proxy at a specific time t. For our purposes however, given information up to time T, whatever methods or procedures we use to select the proxies ought to select whole time series G_j, for which G_{jt} satisfies the null hypothesis, $\forall t$. In this regard, our first proxy selection method is based upon the following statistic:

$$A(j) = \frac{1}{T}\sum_{t=1}^{T} 1(|\tau_t(j)| > \Phi_\xi^\tau). \tag{20}$$

The $A(j)$ statistic is the actual size of the test (i.e. the probability of Type I error given the sample size). Since $\tau_t(j)$ is asymptotically standard normal and the test is a two-tailed test, the actual size, $A(j)$, of the t-test should converge to the nominal size (the desired significance level is 2ξ) as $T \to \infty$. This means that if a candidate variable is a good proxy of the underlying factors of a data set, the $A(j)$ statistic calculated from its sample time series should approach 2ξ as the sample size increases. This is the basis on which we use the $A(j)$ statistic to select proxies. It should be noted that the $A(j)$ statistic does not constitute a test in the strict sense since we do not compare a test statistic to a critical value to determine whether or not to reject a null hypothesis. Rather, this procedure gives a ranking of the proxies with the best proxy having an $A(j)$ statistic value closest to 2ξ. In our implementation, the candidate set of proxies, G', is the same as the the panel data set X from which we estimate the factors. Based on the choice of the significance level 2ξ, the $A(j)$ statistic incorporates some degree of robustness by allowing G_{jt} to deviate from $\widehat{G}_{jt}$ for a specified number of time points.

The second method for selecting the proxies considers the statistic:

$$M(j) = \max_{1 \le t \le T} |\tau_t(j)|, \tag{21}$$

which is based on a measure of how far the $\widehat{G}_{jt}$ curve is from G_{jt}. If e_{it} is serially uncorrelated, then:

$$P(M(j) \le x) \approx [2\Phi(x) - 1]^T, \tag{22}$$

where $\Phi(x)$ is the cdf of a standard normal random variable. From (21) and (22), we can perform a test to determine whether the time series of a candidate variable is a good proxy for the latent factors. For instance, suppose we are given a significance level 2ξ and a sample of size T from a particular candidate variable time series, G_j. From the right hand side of (22), we can calculate the corresponding critical value, x, for the test. For the same sample, we calculate $M(j)$ from (21) and conclude that G_j is a good proxy if $M(j) \le x$, and a bad proxy otherwise. The test based on the $M(j)$ statistic is thus stronger than the selection method based on the $A(j)$ statistic, as the $M(j)$ test gives a sharp decision rule. However, the $M(j)$ test has at least one disadvantage. Namely, it requires e_{it} to be serially uncorrelated. We ignore this requirement in our experimental analysis. It should be noted that x increases with the sample size, T. Depending on the nature of the observed sample, this fact could either preserve or reduce the power of the $M(j)$ test.

The requirement that G_{jt} be an exact linear combination of the factors $\forall t$ is quite strong. Due to noise, an observed time series could be a good proxy of the factors without having each of its components be an exact linear combination of the factors. Breeden, Gibbons and Litzenberger (1989) point out that measurement error and time aggregation could be responsible for deviations between the observed variables and the latent factors. Additionally, Bai and Ng (2006b) find that allowing for nonlinearity may lead to improved predictions in some contexts. In this light, let us consider the third factor proxy selection method also found in Bai and Ng (2006b). Suppose that:

$$G_{jt} = \mu_j' F_t + \eta_{jt}$$

with $\eta_{jt} \sim N(0, \sigma_\eta^2(j))$. Let $\widehat{G}_{jt} = \widehat{\gamma}_j' \widetilde{F}_t$ where $\widehat{\gamma}_j$ is as defined earlier and $\widehat{\eta}_{jt} = G_{jt} - \widehat{G}_{jt}$. As $N, T \to \infty$, $\widehat{\sigma}_\eta^2(j) = \frac{1}{T}\sum_{t=1}^{T} \widehat{\eta}_{jt}^2 \xrightarrow{p} \sigma_\eta^2(j)$ and

$$\frac{(\widehat{\eta}_{jt} - \eta_{jt})}{\widehat{s}_{jt}} \xrightarrow{d} N(0,1), \tag{23}$$

where

$$\begin{aligned} \widehat{s}_{jt}^2 &= \frac{1}{T}\widetilde{F}_t' Avar(\widehat{\mu}_j)\widetilde{F}_t + \widehat{var}(\widehat{G}_{jt}) \\ &= \frac{1}{T}\widetilde{F}_t' \Big(\frac{1}{T}\sum_{s=1}^{T} \widetilde{F}_s \widetilde{F}_s' \widehat{\eta}_{js}^2\Big)\widetilde{F}_t + \widehat{var}(\widehat{G}_{jt}). \end{aligned} \tag{24}$$

The first term of (24) is the standard error obtained from White(1980) and $\widehat{var}(\widehat{G}_{jt})$ is the same as in (16). From (23), at the 95% level, the confidence interval for η_{jt} is $(\widehat{\eta}_{jt} - 1.96\widehat{s}_{jt}, \widehat{\eta}_{jt} + 1.96\widehat{s}_{jt})$. It is this confidence interval that we considered using for selecting the proxies; because if G_j is an exact proxy, zero should lie in the confidence interval of η_{jt}, for each t. Once again, given a sample of size T of a candidate proxy variable G_j, the requirement that the confidence interval for η_{jt} should contain zero for all $1 \le t \le T$ in order for G_j to be a good proxy is a very strong one. G_j could still be a good proxy even if there are a few points, G_{jt}, for which the confidence interval of η_{jt} does not contain zero. It is important to note that the proxies selected by all three methods depend on the structure of the $\widehat{k} \times \widehat{k}$ matrix $\widetilde{\Gamma}_t$ that we use in (16). For a given proxy selection method, if the choice of $\widetilde{\Gamma}_t^1, \widetilde{\Gamma}_t^2, \widetilde{\Gamma}^3$ used in calculating (16) all produce the same proxies, it could mean that the respective assumptions associated with the use of $\widetilde{\Gamma}_t^1, \widetilde{\Gamma}_t^2, \widetilde{\Gamma}^3$ might not be very relevant, empirically. Of note is that our experiments resulted in no evidence that the confidence interval method is ever preferred to the other two methods discussed above. Hence, we do not report results for the confidence interval method. Additionally, we found no gains, in our experimental set-up, to using $\widetilde{\Gamma}_t^1$ and $\widetilde{\Gamma}^3$, and hence all reported results are for the case where we use $\widetilde{\Gamma}_t^2$

Finally, it should be noted that Shanken (1992) points out that it is theoretically crucial for the observed selected proxies to span the same space as the r latent factors. We nevertheless consider versions of the above methods where the number of factors is greater than the number of proxies, given the principle of parsimony.

3.3 Smoothed $A(j)$ and $M(j)$ tests for selecting factor proxies

Note that the $A(j)$ and $M(j)$ statistics discussed above may yield a different set of proxies at each point in time when used to construct a sequence of recursive forecasts. Namely, if the information set used in the parameterization of a prediction model is updated prior to the construction of each new forecast for some sequence of E ex ante predictions, then the "first stage" factor analysis discussed above may yield a sequence of E different vectors of factor proxies. Thus, in addition to the $A(j)$ and $M(j)$ proxy selection methods discussed above, we also consider **a** version of these methods where the sample period in our empirical analysis is broken into three subsamples (R_1, R_2, and E, where $T = R_1 + R_2 + E$). The fist subsample is used to estimate proxies. Thereafter, one observation from R_2 is added, and this new larger sample is used to recursively select a second set of factor proxies. This is continued until the second subsample is exhausted, yielding a sequence of R_2 different vectors of factor proxies. Then, individual proxies are ranked according to their selection frequency, and those occurring the most frequently are selected and fixed for further use in constructing E ex ante predictions. As some of our models (such as the autoregressive model) select the number of lags and re-estimate all parameters prior to the formation of each new prediction, this smoothed approach is at a disadvantage, in the sense that it is static (i.e. the set of proxies is fixed throughout the forecast experiment). However, loading parameters for the proxies are still re-estimated prior to the formation of each new recursive prediction. Of course, the potential advantage to this approach is that noise across the proxy selection process is potentially suppressed.

4 Empirical Methodology

In this section, we discuss the methodology used in our subsequent empirical application.

In order to assess the performance of factor proxy based prediction models, we focus our attention on direct multistep-ahead prediction. Forecasts are generated as h-step ahead predictions of y_t, say. Namely, we predict $y_{t+h} = \log\left(\frac{Y_{t+h}}{Y_{t+h-1}}\right)$, where Y_t is the variable of interest. Our approach is to compare the performance of factor based predictions with a host of proxy based predictions as well as various "strawman" predictions. For the "strawman" forecast models, we use an autoregressive ($AR(p)$) model (with lags selected using the Schwarz Information Criterion (SIC)), and a random walk model. The "strawman" models are included because they serve as parsimonious benchmarks that are often difficult to outperform. In Table 1, we provide the specifications and brief descriptions of all of the forecast models examined.

We consider two classes of proxy forecasts models.

The first class of models, which we call "ordinary" proxy forecast models, include *Model 4* - *Model 7*. With these models, proxies are re-selected recursively, prior to the construction of each h-step ahead prediction. Let $\{A(j)\}_{j=1}^{m}$, be a set of $A(j)$ statistics calculated for each candidate proxy variable j. As suggested above, in this particular paper, we set $m = N$; but this need not always be the case. Define:

$$S^A = \{G^A_{j_1}, \dots, G^A_{j_{\widehat{k}}}\} \tag{25}$$

where $\widehat{k} \leq m$ and

$$|A(j_1) - 2\xi| \leq |A(j_2) - 2\xi| \leq \dots \leq |A\left(j_{\widehat{k}}\right) - 2\xi|.$$

Here, S^A is the set of $\widehat{k}$ proxy time series variables selected via implementation of the $A(j)$ test. Further, define $G^A_{j_1}$ as the "best" possible proxy as determined by the $A(j)$ while $G^A_{j_2}$ is the next "best" proxy, and so on. Recall that G_j is an observable time series variable, such as the CPI or the Federal Funds Rate.

Turning next to proxies selected via implementation of the $M(j)$ test, define:

$$S^M = \{G_j \in G \mid M(j) \leq x\}, \quad j = 1, \dots, m.$$

Here, S^M is a set of proxies selected by the $M(j)$ test. The number of proxy variables selected at each recursive stage is indeterminate (and may be zero, in which case the resultant forecasting model is simply a constant). Furthermore, the selected proxies are not ranked. For *Model 6*, where the $M(j)$ test is used to select a single proxy, our approach is to select the proxy in the set S^M that is associated with the smallest value of $M(j)$.

The second class of models, which we call "smoothed" proxy forecast models, are discussed in Section 3.3, and include *Model 8 - Model 15.* The proxies used in these models are still based on implementing the $A(j)$ and $M(j)$ statistics as discussed above. However, whereas in *Models 1, 4-7*; $\widehat{k}$, the factors, and the proxies are re-estimated recursively prior to construction of each ex-ante prediction; in the "smoothed" models, $\widehat{k}$ and the proxies are fixed prior to the construction of any predictions. In order to implement the "smoothed" models, the sample is split into 3 parts such that $T = R_1 + R_2 + E$. Then, $\widehat{k}$, the factors and the proxies are estimated recursively, just as in *Models 1, 4-7,* but this is done starting with R_1 observations and ending with R_1+R_2 observations. For the ex-ante forecast evaluation period E**,** $\widehat{k}$ is held constant at its last value estimated from R_1+R_2 observations. The "smoothed" proxies are then selected as the $\widehat{k}$ ($\widehat{k}$ held constant as above) proxies that are "most frequently" picked by the $A(j)$ and $M(j)$ tests. Thereafter, all proxies are fixed, although their "weights" in the prediction models are still re-estimated recursively, prior to the construction of each of the E ex-ante forecasts. To differentiate between proxies picked using the "ordinary" and "smoothed" versions of the tests, we define S^{A*} and S^{M*} to be the "smoothed" versions of S^A and S^M. Of further note is that the ex-ante prediction period, E, is the same for all models in our empirical experiments.

In order to evaluate forecast performance, we compare mean squared forecast errors (MSFEs) defined as $\frac{1}{E}\sum_{t=R-h+1}^{T-h}(\widehat{y}_{t+h} - y_{t+h})^2$, where $R = R_1+R_2$. We also carry out Diebold and Mariano (DM: 1995) predictive accuracy tests. Let $\{\widehat{y}_{1,t}\}_{t=R-h+1}^{T-h}$ and $\{\widehat{y}_{2,t}\}_{t=R-h+1}^{T-h}$ be two forecasts of the time series $\{y_t\}_{t=R-h+1}^{T-h}$. The "benchmark" is *Model 1* (i.e. the factor model), and is used to generate $\{\widehat{y}_{1,t}\}_{t=R-h+1}^{T-h}$, while *Models 2-15* are used to generate $\{\widehat{y}_{2,t}\}_{t=R-h+1}^{T-h}$. For simplicity, we assume that the forecast models are non-nested. The corresponding out-of-sample forecast errors are $\{\widehat{\varepsilon}_{1,t}\}_{t=R-h+1}^{T-h}$ and $\{\widehat{\varepsilon}_{2,t}\}_{t=R-h+1}^{T-h}$. The null hypothesis of equal forecast accuracy for two forecasts is given by

$$H_0 : E[\widehat{\varepsilon}_{1,t}^2] = E[\widehat{\varepsilon}_{2,t}^2]$$

or

$$H_0 : E[\widehat{d}_t] = 0,$$

where $d_t = \widehat{\varepsilon}^2_{1,t} - \widehat{\varepsilon}^2_{2,t}$ is the loss differential series. The DM test statistic is:

$$DM = E^{-1/2} \frac{\overline{d}}{\sqrt{\widehat{\sigma}^2_d}}$$

where $\overline{d} = \frac{1}{E} \sum_{t=R-h+1}^{T-h} \widehat{d}_t$, and $\widehat{\sigma}^2_d$ is a HAC standard error for d_t. Since the forecast models are assumed to be non-nested, and assuming that parameter estimation error vanishes, the DM test statistic has a $N(0,1)$ limiting distribution. Finally, given this setup, note that a negative DM t-stat indicates that the factor model yields a lower point MSFE. For further discussion of parameter estimation error and nestedness issues in the context of predictive accuracy tests, the reader is referred to Corradi and Swanson (2002, 2006a, 2006b).

5 Data

The dataset used to estimate the factors is the same as that used in Stock and Watson (2005), which can be obtained at *http://www.princeton.edu/˜mwatson*. This dataset contains 132 monthly time series for the United States for the entire period from 1960:1 to 2003:12, hence $N = 132$ and $T = 528$ observations. The series were selected to represent the following categories of macroeconomic time series: real output and income; employment, manufacturing and trade sales; consumption; housing starts and sales; real inventories and inventory-sales ratios; orders and unfilled orders; stock price indices; exchange rates; interest rate spreads; money and credit quantity aggregates, and price indexes. Most of the series were taken from the Global Insights Basic Economic Database or The Conference Board's Indicators Database (TCB). Others were calculated by Stock and Watson with information from either Global Insights or TCB or both. The theory outlined assumes that the panel dataset used to estimate the factors is $I(0)$. To achieve this, some of the 132 series were subjected to transformations by taking logarithms and/or first differencing. In general, logarithms were taken for all nonnegative series that were not already in rates or percentage units (see Stock and Watson (2002a,2005) for complete details). After these transformations were carried out, all series were further standardized to have sample mean zero and unit sample variance. Using the transformed data set, denoted above by X, the factors are estimated by the method of principal components. As mentioned earlier, in our implementation, the set of candidate proxies for the factors G', will be the same as X. Although this need not be the case, it is done mainly because X represents the biggest set of (standardized and stationary) observable time series variables available to us. We perform real-time forecasts of 7 of the 8 major monthly macroeconomic time series studied in Stock and Watson (2002a). The four real variables we concentrate on are total industrial production (IP), real personal income less transfers, real manufacturing and trade sales and the number of employees on nonagricultural payrolls. The three price series considered are the consumer price index (CPI), the personal consumption expenditure implicit price deflator (PCED) and the producer price index for finished goods (PPI). All of these variables are expressed in log-differences (i.e. monthly growth rates).[1]

6 Monte Carlo Experiment

Table 2 contains results from a small Monte Carlo experiment used to assess the finite sample performance of the $A(j)$ and $M(j)$ tests when they are used to select factor proxies.

In the empirical panel dataset spanning 1960:1 to 2003:12, $\widehat{k} = 13$ factors were consistently estimated using the methodology of Bai and Ng (2002). For this reason, we assume there are 13 factors underlying our simulated dataset and set $r = 13$. We further assume $F_{kt} \sim N(0,1)$, $\Lambda_{ik} \sim N(0,1)$ and $e_{it} \sim N(0, \sigma^2_{e_i})$ where e_{it} is uncorrelated with e_{jt} for $i \neq j$, $i,j = 1,\dots,N$; $t = 1,\dots,T$; $k = 1,\dots,r$. The simulated panel dataset is consequently generated as

$$x_{it} = \Lambda_i' F_t + e_{it}$$

The observed variables that serve as the set of candidate proxies in Table 2 are simulated through the basic data generating process (DGP)

$$G_{mt} = g_m(F_t) + \varepsilon_{mt}$$

where $g_m(\cdot)$ is a function (see Table 2 for details of the different functions tried), $\varepsilon_{mt} \sim N(0,1)$, and $m = 1,\dots,9$. From Table 2, note that with the exception of G_5, G_7 and G_8, $g_m(\cdot)$ is a

[1]Note that Stock and Watson (1999, 2002a) model some of our price variables as $I(2)$ in logarithms. However, they find little discrepancy in performance under $I(1)$ and $I(2)$ assumptions for factor forecasts of our three target price variables. For this reason, we limit our analysis by assuming that our price variables as well as other variables in X are $I(1)$ in logarithms (see Section 10 for further details). In all other respects, our dataset is the same as that used by Stock and Watson (2005).

linear combination of the underlying factors. Without loss of generality, some of the factor loadings are set equal to zero. At each iteration, before estimating the factors by the method of principal components and consistently estimating k, the simulated panel dataset is first standardized to have mean zero and unit variance. The candidate proxies (i.e. G_m) are also standardized to have mean zero as well as unit variance. The standardized proxies, $G_1, \ldots, G_9$, are in turn evaluated using the $A(j)$ and $M(j)$ tests to determine whether they are good proxies of the estimated underlying factors, $\widetilde{F}_t$. All results are based on 1000 simulations with samples of $T = 528$ and $N = 132$ observations. All numerical entries are test rejection frequencies. G_1 and G_2 represent data generated under the null that proxies are perfect linear combinations of the factors. G_1 and G_2 consequently correspond to empirical size experiments. G_3 - G_9 are DGPs specified under the alternative; and thus corresponding rejection frequencies are empirical power. At the 5% nominal size, note that the $A(j)$ test is fairly well sized with empirical size of 0.068. However, under G_1, the $M(j)$ test is slightly oversized with an empirical size of 0.080; while $M(j)$ is quite well sized under G_2, as the empirical size in this case is 0.064. Thus, based upon our limited experiment, it appears both $A(j)$ and $M(j)$ have good empirical size properties in finite samples. With a rejection frequency of 1.000 for DGPs G_3 - G_9 generated under the alternative, the $M(j)$ test has extremely good power. With the exception of G_3, $A(j)$ also has reasonable power. However, unlike $M(j)$, the power of $A(j)$ diminishes considerably as the DGP approaches a perfect linear combination of the factors. For example, G_3 and G_6, which are not too far from perfect linear combinations of the factors, have the lowest rejection rates.

7 Empirical Findings

In this section, we discuss the results of a series of prediction experiments using the dataset discussed above, and applying the models outlined in Table 1 to construct sequences of recursive ex-ante h-step ahead predictions. The dataset consists of 132 variables (see Section 5), and data are available for the period 1960:1-2003:12. Furthermore, predictions are constructed for the period 1989:5-2003:12. Please see Section 4 for complete details concerning the estimation strategy used in order to specify and estimate the prediction models prior to forecast construction. For models in which proxies were selected using the $M(j)$ and $A(j)$ tests, we set $2\xi = 0.05$. Hence we carry out the tests at a 5% significance level.

We estimate two versions of *Models 1-15.* In one version, a single lag of the target variable is always included as an explanatory variable in the prediction model, while in the other version of the models, the models specified in Table 1 are directly estimated (i.e. with no additional lag included). The reasons why we consider these two versions of the models are as follows: (i) The importance of autoregressive lags in prediction is well established, and considering versions of the basic factor model with additional autoregressive terms involving the target variable is a good way to give the basic factor model a fair chance to "win" our forecasting competition. (ii) Our early empirical investigations suggested that price variable prediction models generally benefited when autoregressive lags were included, while the opposite was the case for our income and output variables. Roughly speaking, adaptive expectations provides one reason why the first autoregressive lag might be important for price variables. Namely, if $E[\pi_t|past] = E[\pi_{t-1}] + \iota(\pi_{t-1} - E[\pi_{t-1}])$, where π_t is inflation, then π_{t-1}clearly appears in the "best" forecasting model. Notice that by considering the above two versions of *Models 1-15*, we create some "overlap", in the sense that a small number of models under the first version of the models is identical to other models under the second version of the models. Equivalent models can be seen by comparing the Models in Table 1 under both versions, or by comparing numerical entries in either Tables 4 and 5 or Tables 6 and 7. However, for the sake of completeness, and in order to facilitate easy comparisons across tables, we report on all 15 models in all tables, and ignore cases of "overlap".

Results of our empirical experiments are gathered in Table 3 (frequency of selected factor proxies), Tables 4-5 (CPI, PCED, and PPI forecasting competition results), and Tables 6-7 (Industrial Production, Personal Income; Nonagricultural Employment, Manufacturing and Trade Sales). Note that in Table 3, selection frequencies are reports, while in Tables 4-7 MSFEs and DM test statistics are reported. Further, results for the case where all prediction models are constrained to include one autoregressive lag are reported in Tables 4 and 6, while results for the case where this constraint is not imposed are collected in Table 5 and 7. Finally, note that MSFE values reported for CPI, PCED and Nonagricultural Employment are multiplied by 100,000 and those reported for Producer Price Index, Industrial Production, Manufacturing and Trade Sales and Personal Income are multiplied by 10,000. For the benchmark *Model 1* (i.e. the factor model), the only tabular entry for all forecast horizons is the MSFE. With all of the other models (i.e. our *alternative* models), there are two entries: The top entry is the MSFE and the bottom entry in parenthesis is the DM t-statistic. As mentioned earlier, a positive DM statistic value indicates that the alternative model has MSFE that is lower than the benchmark, while a negative statistic value indicates the reverse. Entries in bold signify instances where the alternative model outperforms the factor model as determined by a point MSFE comparison. Boxed MSFE entries represent the lowest MSFE value among all the models for a particular forecast horizon. DM statistic entries with a $*$ indicate instances where the respective alternative model significantly outperforms the factor model at a 10% significance level, whereas for entries with a † sign, the factor model significantly outperforms the alternative model at a 10% significance level. We now provide a number of conclusions based on the tables.

Upon inspection of Tables 4 and 5, it is clear that the benchmark factor model (i.e. *Model 1)* significantly outperforms most of the alternative models in the forecast of CPI and PCED. This point is supported by the overwhelming number of DM test rejections in Panels A and B of both Tables 4 and 5. While the benchmark still yields the lower MSFE in many pairwise comparisons when examining PPI results (see Panel C of the tables), the DM test null of equal predictive accuracy is not frequently rejected.

A key exception to the above conclusion that the benchmark model yields superior predictions is the case of *Models 12-15.* From Table 1, recall that these are the models that select factor proxies based upon a smoothed version of the $A(j)$ and $M(j)$ tests. For $h = 1, 3, 12$, these models not only frequently yield lower point MSFEs than the benchmark, but the difference in performance is often significant. Across all 6 panels and 4 forecast horizons (i.e. 24 variable/model type/horizon combinations), it is interesting to note that one or many of $Models$ 12-15 are "MSFE-best" 16 times. Furthermore, of these 16 "wins" it is $Model$ 12 that yields the lowest MSFE in 11 instances. Thus, we have direct evidence that the parsimonious single proxy smoothed $A(j)$ model fares very well indeed, when compared not only to the benchmark, but also to other models which yield lower MSFEs than the benchmark. This suggests that while the factor approach is very useful, often beating the autoregressive and other linear time series models when used for predicting price variables, a parsimonious version of the smoothed $A(j)$ factor proxy approach performs the best, overall. Thus, as pointed out by Bai and Ng (2006c), parsimony is still important. This is even true in the context of ordinary proxy models (*Models 4-7*), as choosing one proxy rather than $\widehat{k}$ proxies often yields the MSFE-best model. Furthermore, factor proxies appear to be quite useful, perhaps in part because their use avoids, to a certain degree, parameter estimation error and model stability issues associated with recursive prediction and model estimation. This argument is consistent with the findings of Stock and Watson (1996), who find evidence of structural instability in the majority of the time series that they examine. Other authors have also noted similar problems. For example, Pesaran and Timmermann (2002) propose a two-stage approach whereby in the first stage, one monitors and tests for structural breaks, and in the second stage one uses an estimation window that accounts for the timing of a possible break when generating forecasts.

Interestingly, in Tables *4* and *5*, the above conclusions hold for $h = 1, 3, 12$ and not for $h = 24$. Indeed, it appears that all models perform quite poorly for $h = 24$, with the notable exception of the benchmark, which clearly outperforms virtually all competitors in all price variable cases when $h = 24$. Thus, at the longest forecast horizons, we have evidence that our simple factor proxy approaches are not faring well at all.

Note that in Table 5 and Table 7, the effect of imposing one autoregressive lag in the models only has an effect on *Models 1* and *Models 4-11*. The other models are the same in both cases. Now, recall that we claimed that when forecasting CPI or PCED, constraining the forecast model to always include one autoregressive lag might make a difference because these variables have a very important autoregressive component. To support this point, note that the MSFEs of *Models 4-11* in Panel A and B of Table 5 are in general considerably bigger than their counterparts in Table 4. However, for PPI, the inclusion constraint does not make a considerable difference. Furthermore, the inclusion constraint does not substantially change the forecast performance of *Model 1* for many forecast horizons. For instance, for CPI, when $h = 1$ (see Table 4), the MSFE for *Model 1* is 3.496. The corresponding value in Table 5 is higher at 3.508. But for $h = 3$, the analogous MSFEs are 3.464 and 3.358. Thus, the inclusion constraint may not be needed.

Turning now to Tables 6 and 7, note that the above conclusions still hold, with the exception of the finding that many alternative models, and not just $Models$ 12-15, are point MSFE "better" than the benchmark, and indeed are also significantly more accurate, based upon application of the DM test. Summarizing results across both tables, it should be noted that the benchmark model does yield the lowest MSFE for 2 of 4 variables when $h = 1$ and for 1 variable when $h = 3$, although the DM test null is not rejected in any of these cases. Furthermore, for all remaining horizon/variable combinations, the benchmark does not yield the lowest MSFE. Indeed, in all but one of these other cases, factor proxy approaches yield superior predictions (the sole exception is a random walk "win" for Manufacturing and Trade Sales when $h = 3$.

Given the above results, it is of interest to tabulate which factor proxies were used in our prediction experiments. This is done in Table 4, where factor proxies that are (most frequently) selected using the $A(j)$ and $M(j)$ test and the frequencies with which they are selected are reported. The second column under "Trans" indicates the data transformation that was performed to induce data stationarity. As is evident, S&P's Common Stock Price Index, Industrials; S&P's Common Stock Price Index, Composite; Dividend Yields, a 1-Year Bond Rate; and Housing Starts are the five most common proxies selected by both $A(j)$ and $M(j)$. Of note is that structural change could account for some of the proxies being selected less frequently than the five above proxies. Clearly, the importance of proxies may in some cases depend on the period in history represented by the data. However, it is interesting that a variety of factor proxies are "picked" across our entire ex-ante prediction period.

In closing, we note that factor proxies appear useful for prediction. Additionally, since factors are unobserved, analyzing and studying them on their own can be quite difficult. For instance, in our context is not clear how relevant it is to study the evolution of the individual factors over time because prior to each new prediction, the factors are re-estimated. Creating a clearly defined historical path for a factor is consequently complicated. The ability to proxy the unobserved factors with observed variables enables us to identify actual variables that can serve as primitive building blocks for (prediction) models of a host of macroeconomic variables.

8 Recent Advances in the Construction of Diffusion Indices

In this section, we briefly highlight some of the most recent work relating to Diffusion Index (factor) models. Some of the concerns raised in this paper such as the use of the same factors and consequently the same proxies to forecast *any* variable are addressed in a number of the papers. For example, Bai and Ng (2006c) offer two refinements to the method of factor forecasting. The current framework is confined to a linear relation between the predictors and the forecasted series. Bai and Ng (2006c) propose a more flexible structure. Their so-called squared principal components approach allows the relationship between the predictors and the factors to be a non-linear. They use a non-linear "link" function that involves expanding the set of predictors to include non-linear functions of the observed variables. In this regard, (4) can be modified as follows:

$$h(x_{it}) = \vartheta_i' J_t + e_{it},$$

where $h(\cdot)$ is a non-linear link function, J_t are the common factors, and ϑ_i is the vector of factor loadings. The second order factor model is consequently:

$$x_t^* = \Omega J_t + e_t \tag{26}$$

where $x_t^* = \{x_{it}, x_{it}^2\} \forall i$ is an $N^* \times 1$ vector and $N^* = 2N$. Estimation of J_t from (26) is done using the usual method of principal components. The forecasting equation in (10) remains linear regardless of the form of $h(\cdot)$. The second refinement proposed by Bai and Ng (2006c) takes explicit account of the fact that the ultimate aim is to forecast a specific time series variable, say y_t. The authors propose using principal components analysis with a "targeted" subset of the predictors in X, which have been tested to have predictive power for y. This implies that the set of predictors used to extract the factors change with y, the targeted forecast variable. "Hard" and so-called "soft" thresholding is used to determine which subset of X the factors are to be extracted from. Under "hard" thresholding, a test with a sharp decision rule determines which variables are "in" or "out". With "soft" thresholding, the top variables are kept in the subset of predictors used to extract the factors. The ordering of the predictors is based on the particular soft-thresholding rule. The "soft" thresholding approach is thus related to our "smoothed" test statistic approach to factor proxy selection.

As a reminder, the use of factor models (diffusion indices) involves a two-step approach in which the factors are first estimated from a large panel dataset. The estimated factors are then used as predictors in the forecast models. Although the estimated factors in the first stage are capable of parsimoniously capturing almost all the information in a large dataset, standard tools for specifying the forecast model in the second stage remain unsastisfactory in certain contexts. The specified prediction models are still susceptible to overfitting or underfitting, for example. In this light, Bai and Ng (2006d) suggest a stopping rule for "boosting" that prevents a model from being overfitted with estimated factors or other predictors. Boosting is a procedure that estimates the conditional mean using M stagewise regressions (Bai and Ng (2006d)). The authors also propose two ways to handle lagged predictors: a component-wise approach that treats each lag as a separate variable, and a block-wise approach that treats lags of the same variable jointly. Some important papers on boosting include Schapire (1990), Freund (1995), Friedman (2001) and Buhlmann and Hothorn (2006).

9 Concluding Remarks

Using Monte Carlo and empirical analysis, we have shown that the $A(j)$ and $M(j)$ statistics of Bai and Ng (2006b) appear to offer an interesting means by which factor proxies for later use in prediction models can be chosen. Indeed, our "smoothed" approaches to factor proxy selection appear to yield predictions that are often superior not only to a benchmark factor model, but also to simple linear time series models which are often difficult to beat in forecasting competitions. In some sense, by using our approaches to predictive factor proxy selection, one is able to open up the "black box" often associated with factor analysis, and to identify actual variables that can serve as primitive building blocks for (prediction) models of a host of macroeconomic variables. This approach in some cases leads to improved prediction, and may also possibly lead to improved policy analysis.

10 Appendix

Table A.1: Data Appendix*

Variable #	Mnemonic	Description	Trans. Code
1	A0m052	personal income (ar, bil. chain 2000 $)	4
2	A0M051	personal income less transfer payments (ar, bil. chain 2000 $)	4
3	A0M224_R	real consumption (ac) a0m224/gmdc	4
4	A0M057	manufacturing and trade sales (mil. chain 1996 $)	4
5	A0M059	sales of retail stores (mil. chain 2000 $)	4
6	IPS10	industrial production index - total index	4
7	IPS11	industrial production index - products, total	4
8	IPS299	industrial production index - final products	4
9	IPS12	industrial production index - consumer goods	4
10	IPS13	industrial production index - durable consumer goods	4
11	IPS18	industrial production index - nondurable consumer goods	4
12	IPS25	industrial production index - business equipment	4
13	IPS32	industrial production index - materials	4
14	IPS34	industrial production index - durable goods materials	4
15	IPS38	industrial production index - nondurable goods materials	4
16	IPS43	industrial production index - manufacturing (sic)	4
17	IPS307	industrial production index - residential utilities	4
18	IPS306	industrial production index - fuels	4
19	PMP	napm production index (percent)	1
20	A0M082	capacity utilization (mfg)	2
21	LHEL	index of help-wanted advertising in newspapers (1967=100;sa)	2
22	LHELX	employment: ratio; help-wanted ads:no. unemployed clf	2
23	LHEM	civilian labor force: employed, total (thous.,sa)	4
24	LHNAG	civilian labor force: employed, nonagric.industries (thous.,sa)	4
25	LHUR	unemployment rate: all workers, 16 years & over (%,sa)	2
26	LHU680	unemploy.by duration: average(mean)duration in weeks (sa)	2
27	LHU5	unemploy.by duration: persons unempl.less than 5 wks (thous.,sa)	4
28	LHU14	unemploy.by duration: persons unempl.5 to 14 wks (thous.,sa)	4
29	LHU15	unemploy.by duration: persons unempl.15 wks + (thous.,sa)	4
30	LHU26	unemploy.by duration: persons unempl.15 to 26 wks (thous.,sa)	4
31	LHU27	unemploy.by duration: persons unempl.27 wks + (thous,sa)	4
32	A0M005	average weekly initial claims, unemploy. insurance (thous.)	4
33	CES002	employees on nonfarm payrolls - total private	4
34	CES003	employees on nonfarm payrolls - goods-producing	4
35	CES006	employees on nonfarm payrolls - mining	4
36	CES011	employees on nonfarm payrolls - construction	4
37	CES015	employees on nonfarm payrolls - manufacturing	4
38	CES017	employees on nonfarm payrolls - durable goods	4
39	CES033	employees on nonfarm payrolls - nondurable goods	4
40	CES046	employees on nonfarm payrolls - service-providing	4
41	CES048	employees on nonfarm payrolls - trade, transportation, and utilities	4
42	CES049	employees on nonfarm payrolls - wholesale trade	4
43	CES053	employees on nonfarm payrolls - retail trade	4
44	CES088	employees on nonfarm payrolls - financial activities	4
45	CES140	employees on nonfarm payrolls - government	4
46	A0M048	employee hours in nonag. establishments (ar, bil. hours)	4
47	CES151	average weekly hours of production	1
48	CES155	average weekly hours of overtime production: manufacturing	2
49	A0M001	average weekly hours, mfg. (hours)	1
50	PMEMP	napm employment index (percent)	1
51	HSFR	housing starts:nonfarm(1947-58);total farm&nonfarm(1959-)(thous.),sa	3
52	HSNE	housing starts:northeast (thous.u.)s.a.	3
53	HSMW	housing starts:midwest(thous.u.)s.a.	3
54	HSSOU	housing starts:south (thous.u.)s.a.	3
55	HSWST	housing starts:west (thous.u.)s.a.	3
56	HSBR	housing authorized: total new priv housing units (thous.,saar)	3
57	HSBNE	houses authorized by build. permits:northeast(thou.u.)s.a	3
58	HSBMW	houses authorized by build. permits:midwest(thou.u.)s.a.	3
59	HSBSOU	houses authorized by build. permits:south(thou.u.)s.a.	3
60	HSBWST	houses authorized by build. permits:west(thou.u.)s.a.	3

Table A.1 cont.

Variable #	Mnemonic	Description	Trans. Code
61	PMI	purchasing managers' index (sa)	1
62	PMNO	napm new orders index (percent)	1
63	PMDEL	napm vendor deliveries index (percent)	1
64	PMNV	napm inventories index (percent)	1
65	A0M008	mfrs' new orders, consumer goods and materials (bil. chain 1982 $)	4
66	A0M007	mfrs' new orders, durable goods industries (bil. chain 2000 $)	4
67	A0M027	mfrs' new orders, nondefense capital goods (mil. chain 1982 $)	4
68	A1M092	mfrs' unfilled orders, durable goods indus. (bil. chain 2000 $)	4
69	A0M070	manufacturing and trade inventories (bil. chain 2000 $)	4
70	A0M077	ratio, mfg. and trade inventories to sales (based on chain 2000 $)	2
71	FM1	money stock: m1(curr,trav.cks,dem dep,other ck'able dep)(bil$,sa)	4
72	FM2	money stock:m2(m1+o'nite rps,euro$,g/p&b/d mmmfs&sav&sm time dep(bil$)	4
73	FM3	money stock: m3(m2+lg time dep,term rp's&inst only mmmfs)(bil$,sa)	4
74	FM2DQ	money supply - m2 in 1996 dollars (bci)	4
75	FMFBA	monetary base, adj for reserve requirement changes(mil$,sa)	4
76	FMRRA	depository inst reserves:total,adj for reserve req chgs(mil$,sa)	4
77	FMRNBA	depository inst reserves:nonborrowed,adj res req chgs(mil$,sa)	4
78	FCLNQ	commercial & industrial loans oustanding in 1996 dollars (bci)	4
79	FCLBMC	wkly rp lg com'l banks:net change com'l & indus loans(bil$,saar)	1
80	CCINRV	consumer credit outstanding - nonrevolving(g19)	4
81	A0M095	ratio, consumer installment credit to personal income (pct.)	2
82	FSPCOM	s&p's common stock price index: composite (1941-43=10)	4
83	FSPIN	s&p's common stock price index: industrials (1941-43=10)	4
84	FSDXP	s&p's composite common stock: dividend yield (% per annum)	2
85	FSPXE	s&p's composite common stock: price-earnings ratio (%,nsa)	4
86	FYFF	interest rate: federal funds (effective) (% per annum,nsa)	2
87	CP90	cmmercial paper rate (ac)	2
88	FYGM3	interest rate: u.s.treasury bills,sec mkt,3-mo.(% per ann,nsa)	2
89	FYGM6	interest rate: u.s.treasury bills,sec mkt,6-mo.(% per ann,nsa)	2
90	FYGT1	interest rate: u.s.treasury const maturities,1-yr.(% per ann,nsa)	2
91	FYGT5	interest rate: u.s.treasury const maturities,5-yr.(% per ann,nsa)	2
92	FYGT10	interest rate: u.s.treasury const maturities,10-yr.(% per ann,nsa)	2
93	FYAAAC	bond yield: moody's aaa corporate (% per annum)	2
94	FYBAAC	bond yield: moody's baa corporate (% per annum)	2
95	SCP90	cp90-fyff	1
96	sFYGM3	fygm3-fyff	1
97	sFYGM6	fygm6-fyff	1
98	sFYGT1	fygt1-fyff	1
99	sFYGT5	fygt5-fyff	1
100	sFYGT10	fygt10-fyff	1
101	sFYAAAC	fyaaac-fyff	1
102	sFYBAAC	fybaac-fyff	1
103	EXRUS	united states; effective exchange rate(merm)(index no.)	4
104	EXRSW	foreign exchange rate: switzerland (swiss franc per u.s.$)	4
105	EXRJAN	foreign exchange rate: japan (yen per u.s.$)	4
106	EXRUK	foreign exchange rate: united kingdom (cents per pound)	4
107	EXRCAN	foreign exchange rate: canada (canadian $ per u.s.$)	4
108	PWFSA	producer price index: finished goods (82=100,sa)	4
109	PWFCSA	producer price index:finished consumer goods (82=100,sa)	4
110	PWIMSA	producer price index:intermed mat.supplies & components(82=100,sa)	4
111	PWCMSA	producer price index:crude materials (82=100,sa)	4
112	PSCCOM	spot market price index:bls & crb: all commodities(1967=100)	4
113	PSM99Q	index of sensitive materials prices (1990=100)(bci-99a)	4
114	PMCP	napm commodity prices index (percent)	1
115	PUNEW	cpi-u: all items (82-84=100,sa)	4
116	PU83	cpi-u: apparel & upkeep (82-84=100,sa)	4
117	PU84	cpi-u: transportation (82-84=100,sa)	4
118	PU85	cpi-u: medical care (82-84=100,sa)	4
119	PUC	cpi-u: commodities (82-84=100,sa)	4
120	PUCD	cpi-u: durables (82-84=100,sa)	4

Table A.1 cont.

Variable #	Mnemonic	Description	Trans. Code
121	PUS	cpi-u: services (82-84=100,sa)	4
122	PUXF	cpi-u: all items less food (82-84=100,sa)	4
123	PUXHS	cpi-u: all items less shelter (82-84=100,sa)	4
124	PUXM	cpi-u: all items less medical care (82-84=100,sa)	4
125	GMDC	pce, impl pr defl:pce (1987=100)	4
126	GMDCD	pce, mpl pr defl:pce; durables (1987=100)	4
127	GMDCN	pce, impl pr defl:pce; nondurables (1996=100)	4
128	GMDCS	pce, impl pr defl:pce; services (1987=100)	4
129	CES275	average hourly earnings of production: goods	4
130	CES277	average hourly earnings of production: construction	4
131	CES278	average hourly earnings of production: manufacturing	4
132	HHSNTN	u. of mich. index of consumer expectations(bcd-83)	2

* Notes: Table A.1 is a list of the time series variables contained in the panel dataset used in Stock and Watson (2005). This data is available at *http://www.princeton.edu/˜mwatson*. "Mnemonic" denotes the label by which the series is identified in the source database. Entries under "Trans. Code" signify how the particular time series is transformed to $I(0)$: 1 = no transformation, 2 = first difference, 3 = logarithm, 4 = first difference of logarithms. The variables Stock and Watson (2005) treat as $I(2)$ in logarithms, we treat as $I(1)$ in logarithms. All series are from the Global Insights Economics Database, The Conference Board's Indicators Database (TCB) or calculated by Stock and Watson with information from the two databases. The Stock and Watson calculated variables have (AC) next to them under "Description".

11 References

Bai, J. (2003). Inferential Theory for Factor Models of Large Dimensions. *Econometrica* 71(1), 135-172.

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

Bai, J. and Ng, S. (2005). Determining the Number of Primitive Shocks in Factor Models. *Working Paper*, University of Michigan, Ann-Arbor.

Bai, J. and Ng, S. (2006a). Confidence Intervals for Diffusion Index Forecasts and Inference for Factor-Augmented Regressions. *Econometrica* 74(4), 1133-1150.

Bai, J. and Ng, S. (2006b). Evaluating Latent and Observed Factors in Macroeconomics and Finance. *Journal of Econometrics* 113, 507-537.

Bai, J. and Ng, S. (2006c). Forecasting Economic Time Series Using Targeted Predictors. *Working Paper*, University of Michigan, Ann-Arbor.

Bai, J. and Ng, S. (2006d). Boosting Diffusion Indices. *Working Paper*, University of Michigan, Ann-Arbor.

Battini, J. and Haldane, A. (1999). Forward-looking Rules for Monetary Policy. In Taylor, J., (ed.), *Monetary Policy Rules*. University of Chicago Press for NBER, Chicago.

Bernanke, B. and Boivin, J. (2003). Monetary Policy in a Data-Rich Environment. *Journal of Monetary Economics* 50, 525-546.

Boivin, J. and Ng, S. (2005). Understanding and Comparing Factor Based Macroeconomic Forecasts. *International Journal of Central Banking* 1(3), 117-152.

Breeden, D., Gibbons, M. and Litzenberger, R. (1989). Empirical Tests of the Consumption-Oriented CAPM. *Journal of Finance* XLIV (2), 231-262.

Buhlmann, P. and Hothorn, T. (2006). Boosting Algorithms: Regularization, Prediction and Model Fitting. *Working Paper*, ETH Zürich.

Clarida, R, Gali, G. and Gertler, M. (1999). The Science of Monetary Policy: A new Keynesian Perspective. *Journal of Economic Literature* 37, 1661-1707.

Clarida, R, Gali, G. and Gertler, M. (2000). Monetary Policy Rules and Macroeconomic Stability: Evidence and Some Theory. *Quarterly Journal of Economics* 115, 147-180.

Connor, G., and Korajczyk, R. (1986). Performance Measurement With the Arbitrage Pricing Theory. *Journal of Financial Economics* 15, 373-394.

Connor, G., and Korajczyk, R. (1988). Risk and Return in an Equilibrium APT: Application of a New Test Methodology. *Journal of Financial Economics* 21, 255-289.

Connor, G., and Korajczyk, R. (1993). A Test for the Number of Factors in an Approximate Factor Model. *Journal of Finance* 48(4) 1263-1291.

Corradi, V., and Swanson, N.R. (2002). A Consistent Test for Out of Sample Nonlinear Predictive Ability. *Journal of Econometrics* 110, 353-381.

Corradi, V., and Swanson, N.R. (2006a). Bootstrap Conditional Distribution Tests in the Presence of Dynamic Misspecification. *Journal of Econometrics* 133, 779-806.

Corradi, V., and Swanson, N.R. (2006b). Predictive Density Evaluation. In Granger, C., Elliot, G. and Timmerman, A. (eds.), *Handbook of Economic Forecasting*. Elsevier, Amsterdam, 197-284.

Diebold, F. and Mariano, R. (1995). Comparing Predictive Accuracy. *Journal of Business and Economic Statistics*. 13, 253-263.

Ding, A. and Hwang, J. (1999). Prediction Intervals, Factor Analysis Models, and High-Dimensional Empirical Linear Prediction. *Journal of the American Statistical Association* 94, 446-455.

Forni, M., Hallin, M., Lippi, M., and Reichlin, L. (2000). The Generalized Dynamic Factor Model: Identification and Estimation. *The Review of Economics and Statistics* 82(4), 540-552.

Forni, M., Hallin, M., Lippi, M., and Reichlin, L. (2005). The Generalized Dynamic Factor Model: One-Sided Estimation and Forecasting. *Journal of the American Statistical Association* 100(471), 830-840.

Forni, M.and Reichlin, L. (1996). Dynamic Common Factors in Large Cross-Sections. *Empirical Economics* 21, 27-42.

Forni, M.and Reichlin, L. (1998). Lets Get Real: A Dynamic Factor Analytical Approach to Disaggregated Business Cycle. *Review of Economic Studies* 65, 453-474.

Freund, Y. (1995). Boosting a Weak Learning Algorithm by Majority. *Information and Computation* 121, 256-285.

Friedman, J. (2001). Greedy Function Approximation: A Gradient Boosting Machine. *The Annals of Statistics* 29, 1189-1232.

Geweke, J. (1977). The Dynamic Factor Analysis of Economic Time Series. In Aigner, D. and Goldberger, A. (eds.), *Latent Variables in Socio-Economic Models*. Amsterdam.

Hayashi, F. (2000). *Econometrics*. Princeton University Press, Princeton.

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

Pesaran, M. and Timmermann, A. (2002). Market Timing and Return Prediction Under Model Instability. *Journal of Empirical Finance* 9, 495-510.

Sargent, T., and Sims, C. (1977). Business Cycle Modeling without Pretending to Have Too Much A-Priori Economic Theory. In Sims, C. et al. (eds.), *New Methods in Business Cycle Research*. Minneapolis: Federal Reserve Bank of Minneapolis.

Schapire, R. (1990). The Strength of Weak Learnability. *Machine Learning* 5, 197-227.

Shanken, J. (1992). On the Estimation of Beta-Pricing Models. *Review of Financial Studies* 5 (1), 1-33.

Stock, J. and Watson, M. (1996). Evidence on Structural Instability in Macroeconomic Time Series Relations. *Journal of Business and Economic Statistics* 14, 11-30.

Stock, J. and Watson, M. (1998). Diffusion Indexes. *Working Paper* 6702, National Bureau of Economic Research.

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

Stock, J. and Watson, M. (2002a). Macroeconomic Forecasting Using Diffusion Indexes. *Journal of Business and Economic Statistics* 20, 147-161.

Stock, J. and Watson, M. (2002b). Forecasting Using Principal Components from a Large Number of Predictors. *Journal of American Statistical Association* 97, 1167-1179.

Stock, J. and Watson, M. (2004a). An Empirical Comparison of Methods for Forecasting Using Many Predictors. *Working Paper*, Princeton University.

Stock, J. and Watson, M. (2004b). Forecasting with Many Predictors. *Handbook of Forecasting*, forthcoming.

Stock, J. and Watson, M. (2005). Implications of Dynamic Factor Models for VAR Analysis. *Working Paper* 11467, National Bureau of Economic Research.

Taylor, J. (1993). Discretion Versus Policy Rules in Practice. *Carnegie Rochester Conference Series on Public Policy* 39, 195-214.

White, H. (1980). A Heteroskedasticity-Consistent Covariance Matrix Estimator and a Direct Test for Heteroskedasticity. *Econometrica* 48, 817-838.

Table 1: Prediction Models Used in Empirical Experiments*

Model 1 (Factor Model): This is the standard factor forecast model: $\widehat{y}_{T+h|T} = \widehat{a}_0 + \widehat{\alpha}'\widetilde{F}_T$

Model 2 (Autoregressive Model): This is an $AR(p)$ forecast model, with lags selected using the SIC: $\widehat{y}_{T+h|T} = \widehat{a}_0 + \sum_{j=1}^{p} \widehat{\alpha}_j y_{T-j+1}$

Model 3 (Random Walk Model): This is a random walk forecast model: $\widehat{y}_{T+h|T} = y_T$

Model 4 (Ordinary $A(j)$ - 1 Proxy Model): In this forecast model, the single "best" proxy selected by the $A(j)$ test (i.e. the proxy associated with the $A(j)$ statistic value closest to 2ξ in absolute value) is used as the only regressor in the forecast model: $\widehat{y}_{T+h|T} = \widehat{a}_0 + \widehat{\alpha}G^A_{j_1T}$

Model 5 (Ordinary $A(j)$ - $\widehat{k}$ Proxies Model): As Model 4, but the "best" $\widehat{k}$ factor proxies selected by the $A(j)$ test are used: $\widehat{y}_{T+h|T} = \widehat{a}_0 + \widehat{\alpha}'S^A_T$, where $S^A_T = \{G^A_{j_1T}, \ldots, G^A_{j_{\widehat{k}}T}\}$.

Model 6 (Ordinary $M(j)$ - 1 Proxy Model): In this forecast model, the single "best" factor proxy selected by the $M(j)$ test (i.e. the proxy associated with the highest t-statistic) is used as the only regressor in the forecast model. Since it is possible for the $M(j)$ test to select no proxies at all, should that scenario occur, the model degenerates to a constant: $\widehat{y}_{T+h|T} = \widehat{a}_0 + \widehat{\alpha}G^M_{jT}$.

Model 7 (Ordinary $M(j)$ - $\widehat{k}$ Proxies Model): As Model 6, but the "best" $\widehat{k}$ factor proxies selected by the $M(j)$ test are used: $\widehat{y}_{T+h|T} = \widehat{a}_0 + \widehat{\alpha}'S^M_T$.

Model 8 (Smoothed $A(j)$ - 1 Proxy Model): This forecast model is the same as Model 4, except that the smoothed version of the $A(j)$ test is used (see Section 3.3 for further discussion).

Model 9 (Smoothed $A(j)$ - $\widehat{k}$ Proxies Model): This forecast model is the same as Model 5, except that the smoothed version of the $A(j)$ test is used (see Section 3.3 for further discussion).

Model 10 (Smoothed $M(j)$ - 1 Proxy Model): This forecast model is the same as Model 6, except that the smoothed version of the $M(j\ )$ test is used (see Section 3.3 for further discussion).

Model 11 (Smoothed $M(j)$ - $\widehat{k}$ Proxies Model): This forecast model is the same as Model 7, except that the smoothed version of the $M(j\ )$ test is used (see Section 3.3 for further discussion).

Model 12 (Autoregressive plus Smoothed $A(j)$ - 1 Proxy Model): This forecast model is the same as Model 8, except that an autoregressive component with lags selected by the SIC is added: $\widehat{y}_{T+h|T} = \widehat{a}_0 + \widehat{\alpha}G^{A*}_{j_1T} + \sum_{j=1}^{p_x} \widehat{\beta}_j y_{T-j+1}$.

Model 13 (Autoregressive plus Smoothed $A(j)$ - $\widehat{k}$ Proxies Model): This forecast model is the same as Model 12, except that $\widehat{k}$ proxies instead of 1 factor proxy are used.

Model 14 (Autoregressive plus Smoothed $M(j)$ - 1 Proxy Model) : This forecast model is the same as Model 12, except that $M(j)$ is used instead of $A(j)$ for selecting a single factor proxy.

Model 15 (Autoregressive plus Smoothed $M(j)$ - $\widehat{k}$ Proxies Model): This forecast model is the same as Model 13, except that $M(j)$ is used instead of $A(j)$ for selecting smoothed factor proxies.

* Notes: See Sections 3.3 and 4 for further discussion of the factor proxy selection methodology used in the construction of the above models.

Table 2: Monte Carlo Experiment Results*

DGP	A(j)	M(j)
$G_1 = F_1 + F_2$	0.068	0.080
$G_2 = 3F_3$	0.068	0.064
$G_3 = F_4 + 0.2\varepsilon$	0.271	1.000
$G_4 = 2F_5 + F_2 + 9\varepsilon$	0.934	1.000
$G_5 = F_3 + F_1^2 + \varepsilon$	0.841	1.000
$G_6 = F_4 + 5 + \varepsilon$	0.779	1.000
$G_7 = \tan^{-1}(F_5) + \varepsilon$	0.852	1.000
$G_8 = \exp(F_3) + \varepsilon$	0.722	1.000
$G_9 = \varepsilon$	0.964	1.000

* Notes: All results are based on 1000 simulations with samples of $T = 528$, corresponding to the number of observations in the datset used in the empirical application. All numerical entries are test rejection frequencies. G_1 and G_2 are models where data are generated under the null that proxies are perfect linear combinations of the factors. G_1 and G_2 consequently correspond to empirical size experiments. G_3 to G_9 are models where data are generated under the alternative; and thus the corresponding rejection frequencies indicate empirical power.

Table 3: Frequency of Selected Factor Proxies*

Selected Factor Proxy	Trans	A(j)	M(j)
fspin: S&P's Common Stock Price Index, Industrials	Δ ln	1.000	1.000
fspcom: S&P's Common Stock Price Index, Composite	Δ ln	1.000	1.000
fsdxp: S&P's Composite Common Stock: Dividend Yield	Δlv	1.000	
fygt1: Interest Rate: U.S. Treasury Const Maturities, 1-Yr	Δlv	1.000	
hsfr: Housing Starts, Nonfarm	ln	1.000	0.949
hsbr: Housing Authorized, Total New Private Housing Units	ln	0.989	0.455
ips10: Industrial Production Index, Total Index	Δ ln	0.909	
exrus: United States, Effective Exchange Rate	Δ ln	0.835	0.370
sfygm6: 6 month Treasury Bills - Federal Funds, spread	lv	0.813	
sfygt5: 5 yr Treasury Bond Const. Maturities - Federal Funds, spread	lv	0.750	
sfygt10: 10 yr Treasury Bond Const. Maturities - Federal Funds, spread	lv	0.659	0.420
fygm6: Interest Rate, U.S. Treasury Bills, Sec Mkt, 6-Mo.	Δlv	0.460	
a0m077: Ratio, Mfg. and Trade Inventories to Sales	Δlv	0.341	0.261

* Notes: In this table we report proxies that were frequently selected using the $A(j)$ and $M(j)$ tests, and the frequencies with which they were selected, in our recursive forecasting experiments. The second column under "Trans" indicates the data transformation that was performed to induce stationarity.

Table 4: Predictive Performance of Various Models for Price Variables[*]

All prediction models include one autoregressive lag

Forecast Horizon (h)	1	3	12	24
		Panel A: CPI		
Model 1	3.496	3.464	4.299	$\boxed{4.089}$
Model 2	**3.457**	**3.330**	4.357	5.069
	(0.136)	**(0.375)**	(-0.155)	(-2.270)†
Model 3	4.785	5.270	6.347	6.129
	(-3.788)†	(-3.795)†	(-3.768)†	(-3.087)†
Model 4	3.809	4.075	4.792	5.305
	(-1.164)	(-1.873)†	(-1.336)	(-2.737)†
Model 5	4.079	4.592	5.255	5.337
	(-1.125)	(-1.775)†	(-1.650)†	(-1.878)†
Model 6	3.802	4.107	4.757	4.891
	(-1.139)	(-2.011)†	(-1.347)	(-1.770)†
Model 7	4.516	4.747	5.095	5.103
	(-1.479)	(-2.223)†	(-1.480)	(-1.600)
Model 8	3.810	4.111	4.759	4.960
	(-1.169)	(-2.048)†	(-1.382)	(-2.014)†
Model 9	3.677	3.921	4.472	4.665
	(-0.775)	(-1.798)†	(-0.618)	(-1.645)†
Model 10	3.819	4.101	4.769	5.208
	(-1.212)	(-2.040)†	(-1.304)	(-2.576)†
Model 11	3.720	4.050	4.563	4.740
	(-0.935)	(-2.022)†	(-0.881)	(-1.659)†
Model 12	$\boxed{\mathbf{3.340}}$	$\boxed{\mathbf{3.158}}$	**4.020**	4.448
	(0.549)	**(0.995)**	**(0.921)**	(-0.981)
Model 13	3.519	**3.296**	**4.097**	4.259
	(-0.086)	**(0.539)**	**(0.606)**	(-0.537)
Model 14	**3.486**	**3.381**	4.351	5.124
	(0.035)	**(0.232)**	(-0.145)	(-2.379)†
Model 15	**3.351**	**3.331**	$\boxed{\mathbf{3.999}}$	4.297
	(0.527)	**(0.411)**	**(0.938)**	(-0.634)
		Panel B: Consumption Deflator (PCE)		
Model 1	2.689	2.882	3.162	$\boxed{2.902}$
Model 2	$\boxed{\mathbf{2.613}}$	**2.540**	**3.097**	3.918
	(0.245)	**(1.598)**	**(0.275)**	(-2.985)†
Model 3	4.318	3.956	4.521	4.823
	(-2.312)†	(-3.275)†	(-3.082)†	(-3.373)†
Model 4	3.561	3.214	3.608	4.114
	(-1.911)†	(-1.525)	(-1.983)†	(-3.754)†
Model 5	2.900	3.488	3.557	3.663
	(-1.106)	(-2.348)†	(-1.990)†	(-2.308)†
Model 6	3.542	3.220	3.587	3.835
	(-1.871)†	(-1.593)	(-2.118)†	(-2.933)†
Model 7	3.123	3.386	3.501	3.648
	(-1.865)†	(-2.486)†	(-1.834)†	(-2.349)†
Model 8	3.562	3.283	3.921	4.412
	(-1.910)†	(-1.847)†	(-3.021)†	(-4.066)†
Model 9	3.375	3.233	3.491	3.826
	(-1.687)†	(-1.948)†	(-1.729)†	(-2.957)†
Model 10	3.593	3.227	3.673	4.207
	(-1.887)†	(-1.614)	(-1.969)†	(-3.925)†
Model 11	3.548	3.196	3.496	3.781
	(-1.717)†	(-1.504)	(-1.769)†	(-2.905)†
Model 12	**2.619**	$\boxed{\mathbf{2.485}}$	**3.118**	3.846
	(0.237)	**(2.005)***	**(0.191)**	(-2.904)†
Model 13	**2.669**	**2.554**	**2.874**	3.294
	(0.066)	**(1.669)***	**(1.360)**	(-1.562)
Model 14	**2.637**	**2.558**	**3.123**	3.978
	(0.163)	**(1.544)**	**(0.160)**	(-3.229)†
Model 15	**2.633**	**2.525**	$\boxed{\mathbf{2.817}}$	3.271
	(0.175)	**(1.870)***	**(1.617)**	(-1.542)

Table 4 (cont.): Predictive Performance of Various Models for Price Variables*

All prediction models include one autoregressive lag

Forecast Horizon (h)	1	3	12	24
Panel C: Producer Price Index (PPI)				
Model 1	2.142	$\boxed{2.152}$	2.351	$\boxed{2.198}$
Model 2	2.445	2.360	2.433	2.385
	(-1.813)†	(-1.349)	(-0.660)	(-1.232)
Model 3	3.140	4.070	3.625	3.737
	(-3.026)†	(-3.407)†	(-3.214)†	(-3.404)†
Model 4	2.201	2.413	**2.300**	2.421
	(-0.387)	(-1.424)	**(0.370)**	(-1.599)
Model 5	2.282	2.391	2.370	2.536
	(-1.143)	(-1.339)	(-0.152)	(-1.576)
Model 6	2.203	2.392	**2.256**	2.303
	(-0.402)	(-1.320)	**(0.729)**	(-0.743)
Model 7	2.332	2.480	**2.273**	2.420
	(-1.205)	(-1.828)†	**(0.632)**	(-1.110)
Model 8	2.206	2.397	**2.257**	2.332
	(-0.420)	(-1.351)	**(0.730)**	(-1.021)
Model 9	$\boxed{\textbf{2.115}}$	2.192	**2.245**	2.238
	(0.394)	(-0.769)	**(1.369)**	(-0.352)
Model 10	2.217	2.474	**2.345**	2.407
	(-0.465)	(-1.806)†	**(0.043)**	(-1.350)
Model 11	2.199	2.409	$\boxed{\textbf{2.200}}$	2.313
	(-0.385)	(-1.569)	**(1.449)**	(-0.938)
Model 12	2.396	2.299	2.356	2.332
	(-1.654)†	(-0.888)	(-0.054)	(-1.021)
Model 13	$\boxed{\textbf{2.115}}$	2.344	**2.245**	2.238
	(0.394)	(-1.512)	**(1.369)**	(-0.352)
Model 14	2.447	2.401	2.465	2.407
	(-1.784)†	(-1.558)	(-0.912)	(-1.350)
Model 15	2.406	2.387	2.383	2.313
	(-1.650)†	(-1.337)	(-0.327)	(-0.938)

* Notes: Primary entries in this table are mean square forecast errors (MSFEs) based upon recursively constructed ex ante predictions for the period 1960:01-2003:12, using Models 1-15 (see Table 1 for an explanation of the different models). Bracketed entries are MSFE type Diebold and Mariano (DM: 1995) predictive accuracy test statistics, where Model 1 is compared with each of the other models). Entries in bold indicate instances where the alternative model (i.e. each of Models 2-15) outperforms the factor model (i.e. Model 1), as indicated by both a lower MSFE and a positive DM test stistic. Boxed MSFE entries represent the lowest MSFE value amongst all models, for a particular forecast horizon, h. DM statistic entries with a $*$ sign indicate instances where the respective alternative model significantly outperforms the factor model at a 10% significance level, whereas for entries with a † sign, the factor model significantly outperforms the alternative model at a 10% significance level, under the assumption that the DM test statistic has a standard normal limiting distribution (see above for further discussion).

Table 5: Predictive Performance of Various Models for Price Variables*

Prediction models not constained to include one autoregressive lag

Forecast Horizon (h)	1	3	12	24
Panel A: CPI				
Model 1	3.508	3.358	4.308	$\boxed{4.059}$
Model 2	**3.457**	**3.329**	4.357	5.069
	(0.197)	**(0.085)**	(-0.133)	(-2.400)†
Model 3	4.785	5.270	6.347	6.129
	(-3.967)†	(-3.969)†	(-3.754)†	(-3.157)†
Model 4	5.336	5.382	5.532	6.088
	(-3.918)†	(-4.210)†	(-2.199)†	(-3.955)†
Model 5	4.923	5.480	5.662	5.730
	(-2.452)†	(-2.417)†	(-1.599)	(-2.470)†
Model 6	5.452	5.458	5.519	5.574
	(-4.182)†	(-4.276)†	(-2.188)†	(-2.965)†
Model 7	6.360	6.234	5.697	5.551
	(-3.108)†	(-3.574)†	(-1.917)†	(-2.158)†
Model 8	5.448	5.461	5.501	5.636
	(-4.170)†	(-4.286)†	(-2.184)†	(-3.169)†
Model 9	3.782	3.919	4.490	4.627
	(-0.990)	(-2.341)†	(-0.626)	(-1.729)†
Model 10	5.417	5.645	5.660	5.904
	(-3.788)†	(-4.621)†	(-2.552)†	(-3.959)†
Model 11	4.950	5.201	5.024	5.200
	(-3.184)†	(-4.096)†	(-1.444)	(-2.627)†
Model 12	$\boxed{\mathbf{3.340}}$	$\boxed{\mathbf{3.158}}$	**4.020**	4.448
	(0.659)	**(0.689)**	**(0.958)**	(-1.082)
Model 13	3.519	**3.296**	**4.097**	4.259
	(-0.044)	**(0.210)**	**(0.636)**	(-0.647)
Model 14	**3.486**	3.381	4.351	5.124
	(0.085)	(-0.069)	(-0.122)	(-2.512)†
Model 15	**3.351**	**3.331**	$\boxed{\mathbf{3.999}}$	4.297
	(0.636)	**(0.088)**	**(0.973)**	(-0.736)
Panel B: Consumption Deflator (PCE)				
Model 1	2.668	2.959	3.208	$\boxed{2.891}$
Model 2	$\boxed{\mathbf{2.613}}$	**2.540**	**3.097**	3.918
	(0.182)	**(1.795)**	**(0.468)**	(-3.179)†
Model 3	4.318	3.956	4.521	4.823
	(-2.355)†	(-3.006)†	(-2.908)†	(-3.131)†
Model 4	4.788	4.805	4.915	5.313
	(-5.373)†	(-5.245)†	(-4.547)†	(-6.110)†
Model 5	3.694	5.109	4.771	4.659
	(-3.057)†	(-2.060)†	(-1.998)†	(-4.614)†
Model 6	4.889	4.896	4.940	4.902
	(-5.644)†	(-5.300)†	(-4.547)†	(-5.118)†
Model 7	4.782	5.539	5.131	4.802
	(-5.075)†	(-3.228)†	(-2.916)†	(-4.015)†
Model 8	4.905	5.038	5.567	5.745
	(-5.696)†	(-5.622)†	(-5.506)†	(-6.324)†
Model 9	4.410	4.494	4.464	4.661
	(-4.510)†	(-4.035)†	(-3.189)†	(-4.729)†
Model 10	4.800	4.991	5.052	5.294
	(-5.377)†	(-5.695)†	(-4.732)†	(-6.288)†
Model 11	4.254	4.445	4.447	4.625
	(-4.324)†	(-4.318)†	(-3.540)†	(-4.800)†
Model 12	**2.619**	$\boxed{\mathbf{2.485}}$	**3.118**	3.846
	(0.171)	**(2.170)***	**(0.398)**	(-3.254)†
Model 13	2.669	**2.554**	**2.874**	3.294
	(-0.004)	**(1.823)***	**(1.566)**	(-1.733)†
Model 14	**2.637**	**2.558**	**3.123**	3.978
	(0.100)	**(1.753)***	**(0.350)**	(-3.422)†
Model 15	**2.633**	**2.525**	$\boxed{\mathbf{2.817}}$	3.271
	(0.112)	**(2.053)***	**(1.817)***	(-1.752)†

Table 5 (cont.): Predictive Performance of Various Models for Price Variables*

Prediction models not constained to include one autoregressive lag

Forecast Horizon (h)	1	3	12	24
Panel C: Producer Price Index (PPI)				
Model 1	2.135	$\boxed{2.150}$	2.358	$\boxed{2.194}$
Model 2	2.445	2.360	2.433	2.385
	(-1.815)†	(-1.363)	(-0.602)	(-1.270)
Model 3	3.410	4.070	3.625	3.737
	(-3.004)†	(-3.420)†	(-3.148)†	(-3.415)†
Model 4	2.314	2.343	**2.299**	2.452
	(-1.244)†	(-1.008)	**(0.375)**	(-1.808)†
Model 5	2.385	2.328	**2.353**	2.509
	(-1.658)†	(-0.961)	**(0.037)**	(-1.565)
Model 6	2.335	2.318	**2.283**	2.322
	(-1.367)	(-0.876)	**(0.484)**	(-0.905)
Model 7	2.469	2.408	**2.270**	2.423
	(-1.691)†	(-1.383)	**(0.660)**	(-1.179)
Model 8	2.340	2.327	**2.279**	2.353
	(-1.395)	(-0.922)	**(0.523)**	(-1.182)
Model 9	**2.119**	2.173	**2.239**	2.238
	(0.168)	(-0.399)	**(1.535)**	(-0.386)
Model 10	2.340	2.441	2.385	2.431
	(-1.217)	(-1.475)	(-0.171)	(-1.529)
Model 11	2.263	2.302	$\boxed{\mathbf{2.170}}$	2.314
	(-0.889)	(-0.850)	**(1.552)**	(-0.995)
Model 12	2.396	2.299	**2.356**	2.332
	(-1.664)†	(-0.900)	**(0.019)**	(-1.057)
Model 13	$\boxed{\mathbf{2.115}}$	2.344	**2.245**	2.238
	(0.297)	(-1.525)	**(1.496)**	(-0.392)
Model 14	2.447	2.401	2.465	2.407
	(-1.790)†	(-1.571)	(-0.854)	(-1.388)
Model 15	2.406	2.387	2.383	2.313
	(-1.661)†	(-1.349)	(-0.262)	(-0.977)

* Notes: See notes to Table 4.

Table 6: Predictive Performance of Various Models for Output, Employment and Sales Variables*

All prediction models include one autoregressive lag

Forecast Horizon (h)	1	3	12	24
Panel A: Industrial Production				
Model 1	2.226	2.459	3.114	2.871
Model 2	2.471	2.490	**2.811**	**2.797**
	(-1.529)	(-0.192)	**(1.343)**	**(0.673)**
Model 3	4.267	3.931	4.541	5.528
	(-4.910)†	(-3.142)†	(-3.165)†	(-4.884)†
Model 4	2.804	2.655	**2.785**	**2.708**
	(-3.270)†	(-1.093)	**(1.436)**	**(1.417)**
Model 5	2.284	2.478	**3.100**	**2.747**
	(-0.419)	(-0.147)	**(0.081)**	**(0.560)**
Model 6	2.682	2.623	**2.795**	**2.688**
	(-2.613)†	(-1.039)	**(1.383)**	**(1.584)**
Model 7	2.678	**2.352**	**2.708**	**2.620**
	(-2.563)†	**(0.948)**	**(1.752)***	**(1.853)***
Model 8	2.719	2.652	**2.737**	**2.584**
	(-2.542)†	(-1.210)	**(1.598)**	**(2.195)***
Model 9	2.445	**2.406**	**2.912**	**2.681**
	(-1.542)	**(0.447)**	**(0.803)**	**(1.504)**
Model 10	2.666	**2.164**	**2.758**	**2.846**
	(-2.474)†	**(2.155)***	**(1.565)**	**(0.232)**
Model 11	2.512	**2.291**	**2.654**	**2.609**
	(-1.911)†	**(1.268)**	**(1.784)***	**(1.852)***
Model 12	2.594	2.615	**2.737**	**2.584**
	(-2.009)†	(-0.976)	**(1.598)**	**(2.195)***
Model 13	2.445	**2.402**	**2.912**	**2.681**
	(-1.542)	**(0.490)**	**(0.803)**	**(1.504)**
Model 14	2.453	**2.123**	**2.758**	**2.846**
	(-1.445)	**(2.445)***	**(1.565)**	**(0.232)**
Model 15	2.502	**2.240**	**2.654**	**2.609**
	(-1.840)†	**(1.608)**	**(1.784)***	**(1.852)***
Panel B: Personal Income Less Transfers				
Model 1	5.919	5.841	5.660	6.235
Model 2	7.167	6.811	**5.576**	**5.994**
	(-1.444)	(-1.522)	**(0.293)**	**(1.841)***
Model 3	15.316	12.858	6.533	10.327
	(-2.046)†	(-1.697)†	(-0.534)	(-1.459)
Model 4	6.408	6.028	**5.225**	**6.083**
	(-0.725)	(-0.927)	**(1.627)**	**(1.587)**
Model 5	6.030	6.028	**5.642**	**6.148**
	(-0.292)	(-1.125)	**(0.118)**	**(1.117)**
Model 6	6.373	5.996	**5.298**	**6.071**
	(-0.674)	(-0.790)	**(1.513)**	**(1.889)***
Model 7	6.570	6.249	**5.518**	**6.027**
	(-0.941)	(-1.328)	**(0.418)**	**(2.272)***
Model 8	6.368	5.991	**5.300**	**6.075**
	(-0.666)	(-0.764)	**(1.505)**	**(1.840)***
Model 9	6.334	6.147	5.690	**6.132**
	(-0.741)	(-2.102)†	(-0.074)	**(0.969)**
Model 10	6.569	6.077	**5.363**	**6.026**
	(-0.734)	(-0.834)	**(0.940)**	**(1.581)**
Model 11	6.336	6.057	**5.358**	**6.042**
	(-0.610)	(-0.782)	**(1.347)**	**(1.887)***
Model 12	6.766	6.674	**5.490**	**6.075**
	(-1.268)	(-1.327)	**(0.767)**	**(1.840)***
Model 13	6.659	6.791	5.920	**6.150**
	(-1.220)	(-1.589)	(-0.676)	**(1.004)**
Model 14	7.164	6.809	**5.587**	**6.007**
	(-1.440)	(-1.491)	**(0.269)**	**(1.548)**
Model 15	6.649	6.796	**5.482**	**6.042**
	(-1.022)	(-1.417)	**(0.936)**	**(1.887)***

Table 6 (cont.): Predictive Performance of Various Models for Output, Employment and Sales Variables*

All prediction models include one autoregressive lag

Forecast Horizon (h)	1	3	12	24
Panel C: Nonagricultural Employment				
Model 1	1.893	1.693	3.587	3.279
Model 2	**1.135**	**1.471**	**3.446**	3.626
	(4.013)*	**(1.323)**	**(0.561)**	(-1.836)†
Model 3	**1.655**	**1.571**	3.685	6.021
	(0.991)	**(0.542)**	(-0.239)	(-5.224)†
Model 4	2.203	2.134	3.607	3.424
	(-1.460)	(-2.614)†	(-0.079)	(-0.970)
Model 5	2.360	2.441	**3.345**	$\boxed{\textbf{2.726}}$
	(-2.191)†	(-3.580)†	**(0.977)**	**(3.068)***
Model 6	2.102	2.032	**3.566**	3.408
	(-0.982)	(-2.115)†	**(0.090)**	(-0.866)
Model 7	2.235	2.102	**3.177**	**2.992**
	(-1.323)	(-2.570)†	**(1.569)**	**(2.170)***
Model 8	2.090	2.024	**3.547**	3.426
	(-0.929)	(-2.073)†	**(0.170)**	(-0.986)
Model 9	2.223	2.219	**3.385**	**2.772**
	(-1.635)	(-3.206)†	**(0.786)**	**(2.767)***
Model 10	**1.772**	**1.632**	**3.311**	3.657
	(0.574)	**(0.333)**	**(1.066)**	(-2.064)†
Model 11	2.084	2.009	$\boxed{\textbf{3.029}}$	**2.784**
	(-0.935)	(-2.081)†	**(2.256)***	**(3.210)***
Model 12	**1.275**	1.719	**3.547**	3.426
	(3.526)*	(-0.187)	**(0.170)**	(-0.986)
Model 13	**1.327**	1.744	**3.385**	**2.772**
	(3.691)*	(-0.428)	**(0.786)**	**(2.767)***
Model 14	$\boxed{\textbf{1.128}}$	$\boxed{\textbf{1.406}}$	**3.311**	3.657
	(4.087)*	**(1.546)**	**(1.066)**	(-2.064)†
Model 15	**1.257**	1.695	$\boxed{\textbf{3.029}}$	**2.784**
	(3.825)*	(-0.015)	**(2.256)***	**(3.210)***
Panel D: Manufacturing and Trade Sales				
Model 1	$\boxed{7.001}$	8.243	8.603	8.187
Model 2	7.294	$\boxed{\textbf{7.729}}$	**8.075**	**7.920**
	(-0.639)	**(1.802)***	**(1.494)**	**(0.912)**
Model 3	21.172	12.915	15.844	18.207
	(-5.572)†	(-3.449)†	(-4.636)†	(-5.484)†
Model 4	7.811	**8.132**	**8.076**	$\boxed{\textbf{7.881}}$
	(-1.696)†	**(0.447)**	**(1.461)**	**(1.073)**
Model 5	7.885	**7.787**	**8.292**	8.425
	(-1.239)	**(2.022)***	**(0.734)**	(-0.914)
Model 6	7.541	**7.808**	**8.074**	**7.925**
	(-1.197)	**(1.895)***	**(1.451)**	**(0.915)**
Model 7	7.706	**7.890**	**8.183**	8.420
	(-1.359)	**(1.643)**	**(1.083)**	(-0.907)
Model 8	7.429	**7.795**	**8.079**	**7.926**
	(-0.959)	**(1.955)***	**(1.447)**	**(0.910)**
Model 9	7.199	**7.836**	**8.148**	**8.033**
	(-0.458)	**(1.589)**	**(1.128)**	**(0.602)**
Model 10	7.571	**7.895**	**8.091**	**7.964**
	(-1.109)	**(1.546)**	**(1.424)**	**(0.763)**
Model 11	7.465	**7.917**	**8.092**	**7.984**
	(-1.019)	**(1.585)**	**(1.237)**	**(0.687)**
Model 12	7.429	**7.795**	**8.079**	**7.926**
	(-0.959)	**(1.955)***	**(1.447)**	**(0.910)**
Model 13	7.199	**7.836**	$\boxed{\textbf{8.013}}$	**8.033**
	(-0.458)	**(1.589)**	**(1.422)**	**(0.602)**
Model 14	7.195	**7.895**	**8.091**	**7.964**
	(-0.398)	**(1.546)**	**(1.424)**	**(0.763)**
Model 15	7.465	**7.917**	**8.092**	**7.984**
	(-1.019)	**(1.585)**	**(1.237)**	**(0.687)**

* Notes: See notes to Table 4.

Table 7: Predictive Performance of Various Models for Output, Employment and Sales Variables*

Prediction models not constained to include one autoregressive lag

Forecast Horizon (h)	1	3	12	24
Panel A: Industrial Production				
Model 1	2.200	2.455	3.082	2.890
Model 2	2.471	2.490	2.811	**2.797**
	(-1.735)†	(-0.220)	(-1.197)	**(0.841)**
Model 3	4.267	3.931	4.541	5.528
	(-4.990)†	(-3.159)†	(-3.248)†	(-4.855)†
Model 4	2.898	2.905	**2.728**	**2.647**
	(-2.898)†	(-2.332)†	**(1.535)**	**(1.937)***
Model 5	$\boxed{\mathbf{2.194}}$	**2.430**	3.149	**2.739**
	(0.045)	**(0.220)**	(-0.360)	**(0.730)**
Model 6	2.842	2.885	**2.760**	**2.659**
	(-2.748)†	(-2.507)†	**(1.385)**	**(1.670)***
Model 7	2.797	2.580	**2.649**	**2.608**
	(-2.526)†	(-0.962)	**(1.782)***	**(1.792)***
Model 8	2.926	2.939	**2.708**	$\boxed{\mathbf{2.535}}$
	(-2.746)†	(-2.621)†	**(1.581)**	**(2.356)***
Model 9	2.450	2.572	**2.811**	**2.602**
	(-1.381)	(-0.885)	**(1.078)**	**(2.101)***
Model 10	2.702	**2.375**	**2.695**	**2.785**
	(-2.078)†	**(0.467)**	**(1.676)***	**(0.866)**
Model 11	2.537	2.542	$\boxed{\mathbf{2.632}}$	**2.568**
	(-1.804)†	(-0.603)	**(1.749)***	**(2.035)***
Model 12	2.594	2.615	**2.737**	**2.584**
	(-2.208)†	(-1.012)	**(1.459)**	**(2.325)***
Model 13	2.445	**2.402**	**2.912**	**2.681**
	(-1.789)†	**(0.454)**	**(0.677)**	**(1.665)***
Model 14	2.453	$\boxed{\mathbf{2.123}}$	**2.758**	**2.846**
	(-1.656)†	**(2.436)***	**(1.420)**	**(0.400)**
Model 15	2.502	**2.240**	**2.654**	**2.609**
	(-2.085)†	**(1.593)**	**(1.661)***	**(1.978)***
Panel B: Personal Income Less Transfers				
Model 1	$\boxed{5.810}$	5.905	6.185	6.265
Model 2	7.167	6.811	**5.576**	**5.994**
	(-2.001)†	(-1.483)	**(0.730)**	**(1.544)**
Model 3	15.316	12.858	6.533	10.327
	(-2.136)†	(-1.712)†	(-0.140)	(-1.417)
Model 4	5.927	5.913	**5.911**	**6.069**
	(-0.785)	(-0.070)	**(1.166)**	**(2.566)***
Model 5	5.857	6.006	6.267	**6.119**
	(-0.433)	(-0.975)	(-0.263)	**(1.637)**
Model 6	5.870	**5.876**	**5.936**	**6.053**
	(-0.407)	**(0.274)**	**(1.058)**	**(2.675)***
Model 7	6.057	6.111	**6.098**	**6.002**
	(-1.579)	(-0.930)	**(0.264)**	**(2.476)***
Model 8	5.858	$\boxed{\mathbf{5.869}}$	**5.936**	**6.058**
	(-0.327)	**(0.314)**	**(1.057)**	**(2.602)***
Model 9	6.040	6.101	**6.128**	**6.077**
	(-1.535)	(-2.440)†	**(0.193)**	**(1.767)***
Model 10	5.894	**5.882**	**5.909**	$\boxed{\mathbf{5.925}}$
	(-0.751)	**(0.225)**	**(1.274)**	**(1.643)**
Model 11	5.872	5.930	**5.978**	**6.031**
	(-0.473)	(-0.155)	**(0.859)**	**(2.080)***
Model 12	6.766	6.674	**5.490**	**6.075**
	(-1.908)†	(-1.257)	**(0.764)**	**(1.904)***
Model 13	6.659	6.791	**5.920**	**6.150**
	(-1.851)†	(-1.515)	**(0.367)**	**(0.923)**
Model 14	7.164	6.809	**5.587**	**6.007**
	(-1.997)†	(-1.451)	**(0.723)**	**(1.312)**
Model 15	6.649	6.796	$\boxed{\mathbf{5.482}}$	**6.042**
	(-1.629)	(-1.374)	**(0.796)**	**(1.816)***

Table 7 (cont.): Predictive Performance of Various Models for Output, Employment and Sales Variables*

Prediction models not constained to include one autoregressive lag

Forecast Horizon (h)	1	3	12	24
Panel C: Nonagricultural Employment				
Model 1	1.837	1.698	3.586	3.227
Model 2	**1.135**	**1.471**	**3.446**	3.626
	(4.056)*	**(1.341)**	**(0.568)**	(-2.164)†
Model 3	**1.655**	**1.571**	3.685	6.021
	(0.788)	**(0.568)**	(-0.243)	(-5.331)†
Model 4	3.850	3.870	3.602	3.399
	(-5.598)†	(-5.180)†	(-0.067)	(-1.123)
Model 5	2.381	2.410	**3.311**	**2.793**
	(-2.259)†	(-3.376)†	**(1.090)**	**(2.662)***
Model 6	3.472	3.542	**3.540**	3.382
	(-4.658)†	(-4.716)†	**(0.204)**	(-1.012)
Model 7	3.703	3.563	**3.115**	**2.973**
	(-3.368)†	(-4.341)†	**(1.754)***	**(1.724)***
Model 8	3.470	3.525	**3.525**	3.396
	(-4.667)†	(-4.677)†	**(0.269)**	(-1.100)
Model 9	3.330	3.570	**3.273**	$\boxed{\mathbf{2.733}}$
	(-4.402)†	(-4.359)†	**(1.176)**	**(2.699)***
Model 10	3.437	3.270	**3.392**	3.528
	(-4.437)†	(-4.022)†	**(0.800)**	(-1.542)
Model 11	3.225	3.345	$\boxed{\mathbf{3.001}}$	**2.761**
	(-4.444)†	(-4.732)†	**(2.415)***	**(3.100)***
Model 12	**1.275**	1.719	**3.547**	3.426
	(3.500)*	(-0.146)	**(0.168)**	(-1.335)
Model 13	**1.327**	1.744	**3.385**	**2.772**
	(3.629)*	(-0.377)	**(0.771)**	**(2.558)***
Model 14	$\boxed{\mathbf{1.128}}$	$\boxed{\mathbf{1.406}}$	**3.311**	3.657
	(4.135)*	**(1.567)**	**(1.081)**	(-2.409)†
Model 15	**1.257**	**1.695**	**3.029**	**2.784**
	(3.823)*	**(0.019)**	**(2.268)***	**(2.971)***
Panel D: Manufacturing and Trade Sales				
Model 1	7.268	8.466	8.512	8.563
Model 2	7.294	$\boxed{\mathbf{7.729}}$	**8.075**	$\boxed{\mathbf{7.920}}$
	(-0.047)	**(2.197)***	**(1.395)**	**(1.395)**
Model 3	21.172	12.915	15.844	18.207
	(-5.647)†	(-3.238)†	(-4.760)†	(-5.615)†
Model 4	8.302	8.536	**8.041**	**8.013**
	(-1.672)†	(-0.228)	**(1.449)**	**(1.292)**
Model 5	8.269	**7.886**	**8.234**	**8.470**
	(-1.215)	**(2.252)***	**(0.699)**	**(0.377)**
Model 6	8.126	**8.117**	**8.054**	**8.096**
	(-1.408)	**(1.206)**	**(1.391)**	**(1.079)**
Model 7	8.058	**8.098**	**8.111**	**8.512**
	(-1.268)	**(1.483)**	**(1.109)**	**(0.170)**
Model 8	8.035	**8.098**	**8.061**	**8.094**
	(-1.264)	**(1.285)**	**(1.381)**	**(1.088)**
Model 9	8.039	**8.034**	**8.113**	**8.137**
	(-1.303)	**(1.450)**	**(1.068)**	**(0.975)**
Model 10	8.053	**8.173**	**8.070**	**8.103**
	(-1.209)	**(0.996)**	**(1.364)**	**(1.028)**
Model 11	8.160	**8.175**	**8.077**	**8.147**
	(-1.403)	**(1.129)**	**(1.136)**	**(0.943)**
Model 12	7.429	**7.795**	**8.079**	**7.926**
	(-0.288)	**(2.428)***	**(1.350)**	**(1.428)**
Model 13	**7.199**	**7.836**	$\boxed{\mathbf{8.013}}$	**8.033**
	(0.123)	**(2.082)***	**(1.314)**	**(1.252)**
Model 14	$\boxed{\mathbf{7.195}}$	**7.895**	**8.091**	**7.964**
	(0.120)	**(2.069)***	**(1.316)**	**(1.309)**
Model 15	7.465	**7.917**	**8.092**	**7.984**
	(-0.330)	**(2.220)***	**(1.109)**	**(1.278)**

* Notes: See notes to Table 4.