EUROPEAN JOURNAL OF PURE AND APPLIED MATHEMATICS
Vol. 3, No. 3, 2010, 478-501
ISSN 1307-5543 – www.ejpam.com

SPECIAL ISSUE ON
GRANGER ECONOMETRICS AND STATISTICAL MODELING
DEDICATED TO THE MEMORY OF PROF. SIR CLIVE W.J. GRANGER

Diffusion Index Models and Index Proxies: Recent Results and New Directions

Nii Ayi Armah[1] and Norman R. Swanson[2*†]

[1] *The Bank of Canada, 234 Wellington Street, Ottawa, ON K1A 0G9, Canada*
[2] *Department of Economics, Rutgers University, 75 Hamilton Street, New Brunswick, NJ 08901, USA*

Abstract. Diffusion index models have received considerable attention from both theoreticians and empirical econometricians in recent years. One reason for this is that datasets with many variables are increasingly becoming available and being utilized for economic modelling, and another is that common factors are often assumed to underlie the co-movements of a set of macroeconomic variables. In this paper we review some recent results in the study of diffusion index models, focusing primarily on advances due to [4, 5] and [1]. We discuss, for example, the construction of factors used in prediction models implemented using diffusion index methodology and approaches that are useful for assessing whether there are observable variables that adequately "proxy" for estimated factors.

2000 Mathematics Subject Classifications: 62M10,62-07

Key Words and Phrases: diffusion index, factor, forecast, macroeconometrics, parameter estimation error, proxy

1. Introduction

The basic premise for using diffusion indices to predict economic variables is that the information in large panel datasets can be condensed into a small set of estimated factors. This suggests that there is a small set of crucial latent factors which generate the co-movements in a large set of macroeconomic variables. In this paper we review recent results in the study of diffusion index models, focusing primarily on advances due to [4, 5] and [1]. We discuss, for

*Corresponding author.
†The views expressed in this paper are those of the authors. No responsibility for them should be attributed to the Bank of Canada.

Email addresses: narmah@bankofcanada.ca (N. Armah), nswanson@econ.rutgers.edu (N. Swanson)

example, the construction of factors used in prediction models implemented using diffusion index methodology and approaches that are useful for assessing whether there are observable variables that adequately "proxy" for estimated factors.

Following the approach of [19, 19], diffusion index forecasts involve a two-step procedure. First, the method of principal components is used to estimate the factors from a large panel of possible predictors, X. Second, the estimated factors are used to forecast the variable of interest, y_{t+1}. Stock and Watson [19] demonstrate that diffusion index forecasts yield encouraging results. Bai and Ng [4], 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, [5] examine whether observable economic variables can serve as proxies for the underlying unobserved factors. In particular, they use the $A(j)$ and $M(j)$ statistics to determine whether a group of observed variables yields precisely the same information as that contained in the latent factors. Stock and Watson [19] have also attempted to link the factors to observed variables. Armah and Swanson [1] 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 for prediction. Once the set of factor proxies is fixed, one effectively eliminates the incremental increase in forecast error variance (i.e., uncertainty) associated with the use of estimated factors. Along these lines, they consider "smoothed" versions of the $A(j)$ and $M(j)$ statistics that pre-select a set of factor proxies prior to the ex-ante construction of a sequence of predictions. Armah and Swanson [1] carry out a large variety of prediction experiments using the macroeconomic dataset of [22] in order to assess their new methodology, and we summarize many of their findings here, some of which 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, their "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.

The rest of the paper is organized as follows. In Section 2 we review certain important elements of the diffusion index literature, with some focus on the methods used by the above authors. In Section 3 we discuss the use of factor proxies, and Section 4 contains a summary of recent empirical findings on proxy construction. Section 5 briefly discuss open issues in the empirical literature, and concluding remarks are gathered in Section 5.

2. Diffusion Index Methodology

Summarizing the discussion in [1], 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 an $l \times 1$ vector of other observable variables with $l \ll N$, 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 non-negative powers of L. In general, dynamic factor models can be transformed into static factor models. In [19], 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 $j = 0, \ldots, q$ 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 [7]. 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 19]. In [10, 11, 18, 3, 2], it was 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 7]. Boivin and Ng [8] also 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. For further details, please refer to [1].

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 [17, 19, 18, 20, 21] and [3] for a detailed explanation of this procedure, and to [14, 15, 16], [12, 13] and [11] for further detailed discussion of diffusion models, in general.

As noted earlier F_t and λ_i are not separately identified, but rather identifiable only up to a square matrix. Stock and Watson [17] 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 dataset 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 [5].

Let k ($k < \min\{N, T\}$) be an arbitrary number of factors, Λ^k be the $N \times k$ matrix of factor loadings, $(\Lambda_1^k, \ldots, \Lambda_N^k)'$, and F^k be a $T \times k$ 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 $tr[\Lambda^{k\prime}X'X\Lambda^k]$, 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. 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 [3] 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 and Monte Carlo experiments, we use $g_2(N,T)$. Of note is that we tried the other penalty functions above, and our results were qualitatively the same. However, [3], as well as others, have shown that in certain contexts, results are sensitive to the choice of penalty function. 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 [3].

3. Recent Developments: Using Proxies in Place of Factors

Reconsider the general equation (3), $y_{t+h} = \alpha' F_t + \beta' W_t + \varepsilon_{t+h}$. As mentioned above, and shown in [18] and [3], 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 $z_t = (F_t', W_t')'$; $E(\varepsilon_{t+h}|y_t, z_t, y_{t-1}, z_{t-1}, \ldots) = 0$, for any $h > 0$; and let z_t and ε_t be independent of the idiosyncratic errors e_{is}, $\forall i,s$. 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:

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

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, $t = 1, \ldots, T-h$. We suppress the k superscript on $\widetilde{F}_t^k$ because we assume we have consistently estimated the number of factors underlying the dataset. 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 [4], since $y_{T+h} = y_{T+h|T} + \varepsilon_{T+h}$, it follows that the forecast error is:

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

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 [4], 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 [5]. In a "second stage," the observable proxies are used in the construction of a prediction model. In this way, all estimation error associated with the factor analysis and proxy selection is essentially "hidden" in the first stage, and does not directly manifest itself in the "second stage" prediction models and prediction errors. Put another way, we are trading-off "estimated factor uncertainty" for "variable selection uncertainty" (see introduction for further discussion). 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.

Now, suppose we observe G', a $(T \times m)$ matrix of observable economic variables that could potentially proxy the latent factors (i.e., G is an $m \times T$ matrix). 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 $G_{jt} = \gamma_j' \widetilde{F}_t + \rho_t$. 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

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

and $\widetilde{\Gamma}_t$ is defined below. The last step above is due to the normalization that $\widetilde{F}'\widetilde{F}/T = I_{\widehat{k}}$. Once again, $\widetilde{D}$ is a $k \times k$ diagonal matrix consisting of the k largest eigenvalues of XX'. Given the null hypothesis that $G_{jt} = \theta_j' F_t$ and that $\widehat{G}_{jt}$ converges to G_{jt} at rate $\sqrt{N}$, [5] 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:

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

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

and

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

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 $n = \min\{\sqrt{N}, \sqrt{T}\}$. 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 4, 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}, [4] 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 $\tau_t(j)$. A 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. 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 dataset, 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.

Another 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, 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. 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 $A(j)$ 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, we also consider a version of these methods where the sample period in our empirical analysis is broken into

[‡]Note that we also considered the confidence interval approach of [4]; but it did not perform better than the above methods.

three subsamples (R_1,R_2, and E, such that $T = R_1 + R_2 + E$). The first 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. Individual proxies are then 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 suppressed.

4. Empirics

4.1. Setup

In order to assess the performance of factor proxy based prediction models, [1] focus on direct multistep-ahead predictions. Forecasts are generated as h-step ahead predictions of y_t, say. Namely, they predict $y_{t+h} = \log\left(\frac{Y_{t+h}}{Y_{t+h-1}}\right)$, where Y_t is the variable of interest. In the appendix, we provide the specifications and brief descriptions of all of the forecast models examined.

They consider two classes of proxy forecast models. The first class of models, which are called "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:

$$S^A = \{G^A_{j_1}, \dots, G^A_{j_{\widehat{k}}}\} \tag{23}$$

where $\widehat{k} \leq m$ and $|A(j_1) - 2\xi| \leq |A(j_2) - 2\xi| \leq \dots \leq |A(j_{\widehat{k}}) - 2\xi|$. Here, S^A is the set of $\widehat{k}$ proxy 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. 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 are called "smoothed" proxy forecast models include Model *8* - Model *15*. Here, 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. The "smoothed" proxies are selected as the $\widehat{k}$ 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. The ex-ante prediction period, E, is the same for all models in our empirical experiments.

In order to evaluate forecast performance, the authors compare mean squared forecast errors (MSFEs) defined as $\frac{1}{E}\sum_{t=R-h+1}^{T-h}\left(\widehat{y}_{t+h} - y_{t+h}\right)^2$, where $R = R_1 + R_2$. They also carry out [9] 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 $\{\widehat{y}_{2,t}\}_{t=R-h+1}^{T-h}$. Since the "benchmark" contains estimated factors and the alternative models contain no estimated factors, the "benchmark" and alternative 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 $\widehat{d}_t = \widehat{\varepsilon}_{1,t}^2 - \widehat{\varepsilon}_{2,t}^2$ is the loss differential series. The DM test statistic is $DM = E^{-1/2}\frac{\overline{d}}{\sqrt{\widehat{\sigma}_d^2}}$, where $\overline{d} = \frac{1}{E}\sum_{t=R-h+1}^{T-h}\widehat{d}_t$, and $\widehat{\sigma}_d^2$ is a HAC standard error for $\widehat{d}_t$. Since the forecast models are non-nested, and assuming that parameter estimation error vanishes, the DM test statistic has a $N(0,1)$ limiting distribution. Finally, given this setup, a negative DM t-stat indicates that the factor model yields a lower point MSFE. For further discussion see [1].

4.2. Data

The dataset used to estimate the factors is the same as that used in [22], 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. For further discussion, see [1].

4.3. Findings

Predictions are constructed for the period 1989:5-2003:12, and results are gathered in Table 1 (frequency of selected factor proxies), Table 2 (CPI, PCED, and PPI forecasting competition results), and Table 3 (Industrial Production, Personal Income; Nonagricultural Employment, Manufacturing and Trade Sales). In Table 1, selection frequencies are reported, while in Tables 2-3 MSFEs and DM test statistics are reported. The 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 a 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 Table 2, 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 large number of DM test rejections in Panels A and B of Table 2. While the benchmark still yields the lower MSFE in many pairwise comparisons when examining PPI results (see Panel C of the table), 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 in the case of Models *12-15.* Recall that these are autoregressive models with exogenous variables (ARX). The lags of the ARX models are selected by the SIC and the exogenous variables are based on smoothed versions 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 3 panels and 3 forecast horizons (i.e., 9 variable/horizon combinations), it is interesting to note that one or many of Models 12-15 are "MSFE-best" 7 times. Furthermore, of these 7 "wins" it is Model 12 that yields the lowest MSFE in 4 instances. Thus, we have direct evidence that the parsimonious single proxy smoothed $A(j)$ model fares very well 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 pure autoregressive and other linear 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 [6], 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 lowest MSFE model.

Turning now to Table 3, the above conclusions still hold, with the exception that many other alternative models, and not just Models 12-15, are point MSFE "better" than the benchmark. Summarizing the results in Table 3, the benchmark model does yield the lowest MSFE for 3 of the 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 the lowest MSFE (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 1, 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 $M(j)$. 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.

These results, as was originally reported in [1], suggest that factor proxies are useful for prediction.

5. New Directions

Many issues remain open in the area of diffusion modeling. For example, while latent variables may underlie much of the systematic movements in economic variables, the relevance of particular latent variables may differ markedly across different market sectors. Also, there is clearly a large amount of excess information in large-scale macroeconomic datasets, and this may cause factor estimates to be very noisy. It remains to assess whether other data reduction techniques such as boosting, bagging, and application of the garrotte or various ridge regression techniques may prove useful for data reduction when used in conjunction with diffusion indices.

More flexible nonlinear methods for incorporating factors into prediction models also remains to be explored in detail, although [6] propose a more flexible structure that allows the relationship between the predictors and factors to be 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 general, though, little is known about the usefulness or relevance of implementing diffusion methodology in generic nonlinear settings, such as when specifying artificial neural networks. Moreover, issues of over-fitting associated with the use of large-scale datasets that are ever growing as new information is accumulated and new variables are measured are of some import. For example, when applying classical tests with fixed significance levels in diffusion index modeling, sequential test biases such as those that arise when comparing numerous different prediction models become relevant.

Other issues related to nonstationarity considerations also remain open. For example, are subsets of the latent variables estimated using factors cointegrated, and if so what are the implications for prediction? Does the data transformation undertaken to impose stationarity on variables in large-scale datasets have an impact on findings of stationarity and cointegration, and more generally on the extraction of factors?

6. Concluding Remarks

In this paper, we review some recent results on diffusion index modeling and suggest that these new methods appear promising, although many issues remain open when using diffusion indices in the specification of prediction models.

ACKNOWLEDGEMENTS Thanks are owed to Roberto Chang, Valentina Corradi, Roger Klein, Esfandia Maasoumi, Marcelo Medeiros, Serena Ng, Greg Tkacz, as well as to participants of the conference on "Real-Time Data" at the Philadelphia Federal Reserve Bank in 2009 and seminar participants at the Bank of Canada. All provided many useful comments on the research agenda that led to the writing of this paper. The authors also wish to thank the editor of the European Journal of Pure and Applied Mathematics, Eyup Cetin, for inviting our contribution to this special issue honoring the academic achievements of Sir Clive W.J. Granger, whose untimely recent passing has saddened the entire academic community.

References

[1] N Armah and N Swanson. Seeing inside the black box: Using diffusion index methodology to construct factor proxies in largescale macroeconomic time series environments. Economic Reviews, Forthcoming, 2010.

[2] J Bai. Inferential theory for factor models of large dimensions. *Econometrica*, 71(1):135–172, 2003.

[3] J Bai and S Ng. Determining the number of factors in approximate factor models. *Econometrica*, 70:192–221, 2002.

[4] J Bai and S Ng. Confidence intervals for diffusion index forecasts and inference for factor-augmented regressions. *Econometrica*, 74(4):1133–1150, 2006.

[5] J Bai and S Ng. Evaluating latent and observed factors in macroeconomics and finance. *Journal of Econometrics*, 113:507–537, 2006.

[6] J Bai and S Ng. Forecasting economic time series using targeted predictors. Working Paper, University of Michigan, Ann-Arbor, 2006.

[7] J Bai and S Ng. Determining the number of primitive shocks in factor models. *Journal of Business and Economic Statistics*, 25:52–60, 2007.

[8] J Boivin and S Ng. Understanding and comparing factor based macroeconomic forecasts. *International Journal of Central Banking*, 1(3):117–152, 2005.

[9] F Diebold and R Mariano. Comparing predictive accuracy. *Journal of Business and Economic Statistics*, 13:253–263, 1995.

[10] A Ding and J Hwang. Prediction intervals, factor analysis models, and high-dimensional empirical linear prediction. *Journal of the American Statistical Association*, 95:446–455, 1999.

[11] M Forni, M Hallin, M Lippi, and L Reichlin. The generalized dynamic factor model: Identification and estimation. *The Review of Economics and Statistics*, 82(4):540–552, 2000.

[12] M Forni and L Reichlin. Dynamic common factors in large cross-sections. *Empirical Economics*, 21:27–42, 1996.

[13] M Forni and L Reichlin. Lets get real: A dynamic factor analytical approach to disaggregated business cycle. *Review of Economic Studies*, 65:453–474, 1998.

[14] G Korajczyk and R Korajczyk. Performance measurement with the arbitrage pricing theory. *Journal of Financial Economics*, 15:373–394, 1986.

[15] G Korajczyk and R Korajczyk. Risk and return in an equilibrium apt: Application of a new test methodology. *Journal of Financial Economics*, 21:255–289, 1988.

[16] G Korajczyk and R Korajczyk. A test for the number of factors in an approximate factor model. *Journal of Finance*, 48(4), 1993.

[17] J Stock and M Watson. Diffusion indexes. Working Paper 6702, National Bureau of Economic Research, 1998.

[18] J Stock and M Watson. Forecasting using principal components from a large number of predictors. *Journal of American Statistical Association*, 97:1167–1179, 2002.

[19] J Stock and M Watson. Macroeconomic forecasting using diffusion indexes. *Journal of Business and Economic Statistics*, 20:147–161, 2002.

[20] J Stock and M Watson. An empirical comparison of methods for forecasting using many predictors. Working Paper Princeton University, 2004.

[21] J Stock and M Watson. Forecasting with many predictors. Handbook of Forecasting: A Forthcoming Book, 2004.

[22] J Stock and M Watson. Implications of Dynamic Factor Models for VAR Analysis. Working Paper 11467 National Bureau of Economic Research, 2004.

Appendix

Prediction Models Used in Empirical Experiments[§]

- Model 1 (Factor Model): This is the standard factor forecast model: $\widehat{y}_{T+h|T} = \widehat{a}_0 + \widehat{\alpha}'\widetilde{F}_T + \widehat{\beta} y_T$
- Model 2 (Autoregressive Model): This is an $AR(p)$ forecast model, with lags selected by 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 proxy regressor in the forecast model: $\widehat{y}_{T+h|T} = \widehat{a}_0 + \widehat{\alpha} G^A_{j_1 T} + \widehat{\beta} y_T$
- Model 5 (Ordinary $A(j)$ - $\widehat{k}$ Proxies Model): The "best" $\widehat{k}$ factor proxies selected by the $A(j)$ test are used: $\widehat{y}_{T+h|T} = \widehat{a}_0 + \widehat{\alpha}' S^A_T + \widehat{\beta} y_T$, where $S^A_T = \{G^A_{j_1 T}, \ldots, G^A_{j_{\widehat{k}} T}\}$.
- 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 lowest $M(j)$ -statistic) is used as the only proxy regressor in the forecast model: $\widehat{y}_{T+h|T} = \widehat{a}_0 + \widehat{\alpha} G^M_{jT} + \widehat{\beta} y_T$. Since it is possible for the $M(j)$ test to select no proxies at all, should that scenario occur, the model degenerates to: $\widehat{y}_{T+h|T} = \widehat{a}_0 + \widehat{\beta} y_T$.
- Model 7 (Ordinary $M(j)$ - $\widehat{k}$ Proxies Model): This forecast model is the same as Model 6, but $\widehat{k}$ factor proxies selected by the $M(j)$ test are used: $\widehat{y}_{T+h|T} = \widehat{a}_0 + \widehat{\alpha}' S^M_T + \widehat{\beta} y_T$.
- 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).

[§]See Sections 3.3 and 4 for further discussion of the factor proxy selection methodology used in the construction of the above models.

- 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 the lag of the autoregressive component is selected by the SIC rather than restricted to 1: $\widehat{y}_{T+h|T} = \widehat{a}_0 + \widehat{\alpha} G^{A*}_{j_1 T} + \sum_{j=1}^{p_x} \widehat{\beta}_j y_{T-j+1}$.

- Model 13 (Autoregressive plus Smoothed $A(j)$ - $\widehat{k}$ Proxies Model): This forecast model is the same as Model 9, except that the lag of the autoregressive component is selected by the SIC rather than restricted to 1.

- Model 14 (Autoregressive plus Smoothed $M(j)$ - 1 Proxy Model) : This forecast model is the same as Model 10, except that the lag of the autoregressive component is selected by the SIC rather than restricted to 1.

- Model 15 (Autoregressive plus Smoothed $M(j)$ - $\widehat{k}$ Proxies Model): This forecast model is the same as Model 8, except that the lag of the autoregressive component is selected by the SIC rather than restricted to 1.

Results

Table 1: Frequency of Selected Factor Proxies[1]

Selected Factor Proxy	Trans	A(j)	M(j)
fspin: S&P's Common Stock Price Index, Industrials	Δ log	1.000	1.000
fspcom: S&P's Common Stock Price Index, Composite	Δ log	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	log	1.000	0.949
hsbr: Housing Authorized, Total New Private Housing Units	log	0.989	0.455
ips10: Industrial Production Index, Total Index	Δ log	0.909	
exrus: United States, Effective Exchange Rate	Δ log	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

Table 2: Predictive Performance of Various Models for Price Variables[2]

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)

Table 2 Continued

Panel B: Consumption Deflator (PCE)				
Model 1	2.689	2.882	3.162	2.902
Model 2	**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**	**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**	**2.817**	3.271
	(0.175)	**(1.870)***	**(1.617)**	(-1.542)

Table 2 Continued

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	$\boxed{\mathbf{2.115}}$	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	$\boxed{\mathbf{2.200}}$	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	$\boxed{\mathbf{2.115}}$	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)

Table 3: Predictive Performance of Various Models for Output, Employment and Sales Variables[3]

Forecast Horizon (h)	1	3	12	24
Panel A: Industrial Production				
Model 1	$\boxed{2.226}$	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**	$\boxed{\mathbf{2.584}}$
	(-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**	$\boxed{\mathbf{2.654}}$	**2.609**
	(-1.911)†	**(1.268)**	**(1.784)***	**(1.852)***
Model 12	2.594	2.615	**2.737**	$\boxed{\mathbf{2.584}}$
	(-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	$\boxed{\mathbf{2.123}}$	**2.758**	**2.846**
	(-1.445)	**(2.445)***	**(1.565)**	**(0.232)**
Model 15	2.502	**2.240**	$\boxed{\mathbf{2.654}}$	**2.609**
	(-1.840)†	**(1.608)**	**(1.784)***	**(1.852)***

Table 3 Continued

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 3 Continued

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)***

Table 3 Continued

Panel D: Manufacturing and Trade Sales				
Model 1	$\boxed{7.001}$	8.243	8.603	8.187
Model 2	7.294	$\boxed{\mathbf{7.729}}$	**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**	$\boxed{\mathbf{7.881}}$
	(-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**	$\boxed{\mathbf{8.013}}$	**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)**

[1]Proxies that were frequently selected using the $A(j)$ and $M(j)$ tests, and the frequencies with which they were selected, are given in this table. The second column under "Trans" indicates the data transformation that was performed to induce stationarity, lv means no transformation; the series was left at level. Δlv means first difference of the level. log means the natural log function was applied to the data. Δ log means the series was first differenced after the natural log function was applied. Empty entries in the fourth column under $M(j)$ indicate that the respective variables were not selected at all by the $M(j)$ test.

[2]Primary entries in this table are mean square forecast errors (MSFEs) based upon recur-

sively 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 statistic. Boxed MSFE entries represent the lowest MSFE value among 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).

[3]See notes to Table 2.