Understanding and Comparing Factor-Based Forecasts

Boivin, Jean and Ng, Serena

18 May 2005

Understanding and Comparing Factor-Based Forecasts*

Jean Boivin[a] and Serena Ng[b]
[a]Columbia University and NBER
[b]University of Michigan

Forecasting using "diffusion indices" has received a good deal of attention in recent years. The idea is to use the common factors estimated from a large panel of data to help forecast the series of interest. This paper assesses the extent to which the forecasts are influenced by (i) how the factors are estimated and/or (ii) how the forecasts are formulated. We find that for simple data-generating processes and when the dynamic structure of the data is known, no one method stands out to be systematically good or bad. All five methods considered have rather similar properties, though some methods are better in long-horizon forecasts, especially when the number of time series observations is small. However, when the dynamic structure is unknown and for more complex dynamics and error structures such as the ones encountered in practice, one method stands out to have smaller forecast errors. This method forecasts the series of interest directly, rather than the common and idiosyncratic components separately, and it leaves the dynamics of the factors unspecified. By imposing fewer constraints, and having to estimate a smaller number of auxiliary parameters, the method appears to be less vulnerable to misspecification, leading to improved forecasts.

JEL Codes: E37, E47, C3, C53.

*We thank Domenico Giannone, Mario Forni, Marco Lippi, and Lucrezia Reichlin for useful comments and for sharing their computer codes for the one-sided estimation of the generalized dynamic factor model. We also thank James Stock and Mark Watson for many helpful discussions. The authors acknowledge financial support from the National Science Foundation under grant SES-0214104 (Boivin) and SES-0136923 (Ng). Correspondence: Boivin: Columbia Business School, 821 Uris Hall, 3022 Broadway, New York, NY 10027; e-mail: jb903@columbia.edu; www.columbia.edu/˜jb903. Ng: Department of Economics, University of Michigan, 317 Lorch Hall, Ann Arbor, MI 48109; e-mail: Serena.Ng@umich.edu.

Many economic decisions, whether made by policymakers, firms, investors, or consumers, are often based on the forecasts of relevant macroeconomic variables. The accuracy of these forecasts can thus have important repercussions. In theory, the optimal forecast of a variable under quadratic loss is its expectation conditional on information available. In practice, the relevant information set might be very large. For instance, central banks are known to monitor hundreds of macroeconomic indicators, each potentially carrying useful additional information. Forecasting using "diffusion indices" has provided a formal way to systematically handle this information. The idea is to use factors estimated from a large panel of data to help forecast the series of interest, so that information in a large number of variables can be used while keeping the dimension of the forecasting model small. Stock and Watson (2004b) provide a survey of the factor approach and alternative methods that exploit information in a large number of predictors in forecasting.

Various authors have provided convincing evidence in support of the diffusion index forecast methodology. Stock and Watson (2002b), Stock and Watson (1999), Stock, Watson, and Marcellino (2003), Forni et al. (2001), and Forni and Reichlin (2001), among others, all find that diffusion index forecasts have smaller mean-squared errors than forecasts based upon simple autoregressions and more elaborate structural models. Diffusion index forecasts are considered not just by academic economists. Various institutions, including the Federal Reserve of Chicago, the U.S. Treasury, the European Central Bank, the European Commission, and the Center for Economic Policy Research (CEPR) are all investigating the potential of the factor forecasts.[1]

Although using factors estimated from large panels for forecasting has generally been viewed as a sound idea, diffusion index forecasts can be implemented in a variety of ways. The two leading methods in the literature are the "dynamic" method of Forni et al. (2005) (hereafter FHLR), and the "static" method of Stock and Watson (2002a) (hereafter SW). For example, the CEPR coincident indicator of the euro-business cycle (EUROCOIN) is based on FHLR, while

[1]See Grenouilleau (2004) and references therein.

the Federal Reserve Bank of Chicago's Activity Index (CFNAI) as well as the model of Kitchen and Monaco (2003) at the U.S. Treasury are based on SW, and all these forecasts exploit the factors to summarize information from a large panel of data. It is generally thought that the methods differ primarily because of the methodology used to estimate the factors, though whether this is the main reason why the forecasts differ remains to be confirmed. Monte Carlo experiments designed to shed light on the finite sample properties of the procedures tend to be counterfactually simple, and thus have not been too useful in guiding practitioners as to whether and when one method works better than the other. There is thus a good deal of confusion as to which is the best implementation of diffusion index forecasts, and why.

To make some progress toward a better understanding of this issue, we take as a starting point that there are two steps to diffusion index forecasting. Step E estimates the factors from a large panel of data, and step F uses the factor estimates to forecast the series of interest. Two researchers can arrive at different forecasts because the factors are estimated differently and/or the forecasting equations are specified differently. Accordingly, we seek to understand the sensitivity of steps E and F to (i) the dynamics of the factors and (ii) the specification of the forecasting equation.

We evaluate the out-of-sample forecast errors of five methods that incorporate factors into the forecasts. We use simple and calibrated experiments to assess the sensitivity of the forecast errors in a variety of data-generating processes. We then apply the methods to real data and find that their performance is more in line with simulations that assume complex error structures. Simple data-generating processes appear not to give a good guide to the properties of the different methods in practice. Our main finding is that the choice of step E holding step F fixed does not generate significant discrepancies in forecast errors. However, how step E is used in conjunction with step F can be important both in simulations and in applications. Not imposing the factor structure on step F tends to give more robust forecasts when the data-generating process is unknown. This suggests unconstrained modeling of the series to be forecasted, instead of careful modeling of the components underlying the series. The diffusion index forecasting methodology proposed by Stock and Watson apparently has these properties.

1. Preliminaries

The precise environment we consider is the following. We have T time series observations for N cross-section units, which we denote by x_{it} $(i = 1, \ldots, N,\ t = 1, \ldots T)$. We are interested in x_{iT+h}, the h-step-ahead out-of-sample forecast of a series in the panel. As a matter of notation, we let X be the $T \times N$ matrix of observations; $x_{t\,\bullet}$ is a row vector denoting all N observations at time t, while $x_{\bullet\, i}$ is a column vector denoting all T observations for unit i.

We consider two factor representations of the data. The *static factor model* is

$$x_{it} = \lambda_{i1}F_{1t} \ldots + \lambda_{ir}F_{rt} + e_{it} = \lambda_i' F_t + e_{it}, \tag{1}$$

where F_t is a vector of r common factors, λ_i is the corresponding vector of loadings for unit i, and e_{it} is an idiosyncratic error. We assume $\frac{1}{N}\sum_{i=1}^{N} \lambda_i\lambda_i' \xrightarrow{p} \Sigma_\Lambda$ as $N \to \infty$, and $\frac{1}{T}\sum_{t=1}^{T} F_t F_t' \xrightarrow{p} \Sigma_F$ as $T \to \infty$, where Σ_F and Σ_Λ are $r \times r$ positive definite matrices. As we cannot separately identify the factors and loadings, Σ_Λ is normalized to an identity matrix of dimension r. The model is said to have r static factors because the N dimensional population covariance matrix of x_{it} has r nonzero eigenvalues that diverge with N. Weak cross-section correlation in e_{it} is allowed so long as $\frac{1}{N}\sum_{i=1}^{N}\sum_{j=1}^{N} |E(e_{it}e_{jt})|$ is bounded. The factor model is thus an "approximate factor model" in the sense of Chamberlain and Rothschild (1983). Dynamics are entertained by allowing both the factors and the errors to be serially correlated. If

$$\Big(I_r - A(L)L\Big)F_t \;=\; u_t, \tag{2}$$

$$\Big(1 - \rho_i(L)L\Big)e_{it} \;=\; v_{it} \quad i = 1, \ldots N, \tag{3}$$

where $A(L) = A_1 + A_2L + \ldots$, and $\rho_i(L) = \rho_{i1} + \rho_{i2}L + \ldots$, we assume that the characteristic roots of $|I - A(z)z| = 0$ and $1 - \rho_i(z)z = 0$ are inside the unit circle. Furthermore, F_t and e_{is} are assumed to be mutually uncorrelated at all t and s.

The static factor model is to be distinguished from a *dynamic factor model*

$$x_{it} = b_{i1}(L)f_{1t} + b_{i2}(L)f_{2t} + \ldots + b_{iq}(L)f_{qt} + e_{it} = b_i'(L)f_t + e_{it}, \tag{4}$$

where $f_t = (f_{1t}, \ldots, f_{qt})'$ is a q-vector of dynamic factors with $a(L)f_t = u_t$, u_t being a vector of q orthonormal white noise processes. We suppose that $b_{ij}(L)$ is of order s, for every $j = 1, \ldots q$. Note that some coefficients in $b_{ij}(L)$ can be zero, since s is the maximum lag order (over all i and j) of $b_{ij}(L)$. The data generated under (4) are said to have q dynamic factors since the N dimensional spectral density matrix of x_{it} has rank q. Hereafter, we will refer to

$$\chi_{it} = x_{it} - e_{it}$$

as the common component. Under the static model, $\chi_{it} = \lambda_i' F_t$, and under the dynamic model, $\chi_{it} = b_i'(L)f_t$.

Clearly, if we let $F_t = (f_t', f_{t-1}', \ldots f_{t-s}')'$, the dynamic factors also have a static representation with $\lambda_i' F_t = b_i(L)' f_t$. A model with q dynamic factors thus has $r = q(s+1)$ static factors. We can likewise represent data generated by (1) using a dynamic model upon specifying both q and s. For example, if $x_{it} = \lambda_{i1} F_t + \lambda_{i2} F_{t-1} + e_{it}$, the corresponding dynamic model is defined by $f_t = F_t$, $q = 1$, and $s = 1$. An important distinction between the static and the dynamic model is that r, the total number of static factors, completely characterizes the static model. With the dynamic model, separate specifications of q and s are required. Yet given r, we cannot separately identify q and s without additional assumptions.

Because the dynamic model always has a static representation, it is useful to use the latter to discuss some general issues. Predictability of x_{it} requires that F_t and/or e_{it} are serially correlated. To understand the difference between diffusion index and autoregressive forecasts, consider $h = 1$, and assume $\lambda_i \neq 0$. We have

$$x_{iT+1|T} = \rho_i(L) x_{iT} + \lambda_i'(F_{T+1|T} - \rho_i(L) F_T). \tag{5}$$

Equation (5) makes it apparent that an autoregressive forecast is a special case of a diffusion index forecast that imposes the restriction that $F_{T+1|T} - \rho_i(L)F_T$ is unpredictable. The factors should contribute to forecast error reduction if the restriction is false. This occurs when the factors and e_{it} have different dynamics.

The result that an autoregressive forecast is a special case of a diffusion index forecast implies that the factors can be used to improve forecasts without adopting the factor model as the forecasting model. This observation is important and it is worth considering the

case of one factor. Suppose $F_t = \alpha F_{t-1} + u_t$ and $e_{it} = \rho_i e_{it-1} + v_{it}$. We have

$$\begin{aligned} x_{it} &= \lambda_i(\alpha F_{t-1} + u_t) + (\rho_i e_{it-1} + v_{it}) \\ &= \rho_i x_{it-1} + \lambda_i u_t + v_{it} + \lambda_i(\alpha - \rho_i) F_{t-1}. \end{aligned} \tag{6}$$

When the factors and the parameters are known, the diffusion index forecast is $x_{iT+1|T} = \rho_i x_{iT} + \lambda_i(\alpha - \rho_i)F_T$. When $\alpha \neq \rho_i$ and $\lambda_i \neq 0$, the factor forecast will have smaller errors than an AR(1) forecast. To achieve this forecast error reduction, separate forecasts of the common and idiosyncratic components are not necessary. One only needs to augment an autoregression with the factors. Note, however, that the effectiveness of diffusion index forecasts depends on ρ_i, λ_i, and the dynamics of the factors, α_i. It is thus series specific.

If the parameters, the factors, and the loadings (and thus the components) were observed, the following three forecasts

$$x_{iT+h|T} = \lambda_i' F_{T+h|T} + e_{iT+h|T} \tag{7}$$

$$= \chi_{iT+h|T} + e_{iT+h|T} \tag{8}$$

$$= \rho_i(L) x_{iT+h-1|T} + (1 - \rho_i(L)L)\lambda_i' F_{T+h|T} \tag{9}$$

are mathematically equivalent. In other words, forecasting the components separately should be the same as forecasting the sum plus one of the two components separately. But when the parameters and the factors are unknown and have to be estimated, the equivalence of (7), (8), and (9) breaks down. The sampling error of the estimates might dominate the information gain in the factors. An autoregressive forecast might well have a smaller mean-squared forecast error in finite samples. We will now consider the feasible variants of (7)–(9).

2. Step E: Estimation

In classical factor analysis, e_{it} is serially uncorrelated and iid across i. Under the assumption that N is fixed and T is large, the maximum likelihood estimator yields $\sqrt{T}$ consistent estimates of the loadings. As shown in Anderson and Rubin (1956), the estimator relies on convergence of the $N \times N$ sample to the population covariance matrix of x. Brillinger (1981) showed that the sample spectral density matrices can also be used to consistently estimate the dynamic

factors. From an empirical perspective, the fixed N assumption is unappealing because the number of time series available for economic analysis is by no means small. Connor and Korajzcyk (1986) showed that the method of "asymptotic principal components" can be used to consistently estimate the factors when N is large. Stock and Watson (2002a) and Bai and Ng (2002) formalized the conditions under which the factor space can be consistently estimated by the static estimator when N and T are both large, with no restriction in the relation between N and T. Bai (2003) further showed that the convergence rate of the estimated factors is $\sqrt{N}$. Forni et al. (2000) showed that the method of dynamic principal components provides pointwise consistent estimates of the common component when $N, T \to \infty$. Conditions for achieving convergence to the dynamic space spanned by the common shocks was further developed in Forni et al. (2004).

Static [S]. Let V be the eigenvectors corresponding to the r largest eigenvalues of the $N \times N$ matrix $\hat{\Gamma}_X(0) = \frac{1}{T}\sum_{t=1}^{T} x_{t\,\bullet} x'_{t\,\bullet}$. The static principal components estimator yields

$$\hat{F} = XV \qquad \hat{\Lambda} = V \qquad \hat{\chi} = \hat{F}\hat{\Lambda}' = XVV'.$$

Given $\hat{\chi}$, $\hat{e} = X - \hat{\chi}$.

Dynamic [D]. (i) Construct the sample autocovariances $\hat{\Gamma}_X(k) = \frac{1}{T}\sum_{t=k+1}^{T} x'_{t\,\bullet} x_{t-k\,\bullet}$, $k = 1, \ldots M$. (ii) For each frequency $\omega_h = \frac{2\pi h}{2H}$, $h = -H, \ldots H$, compute the eigenvalues of $\hat{\Sigma}_X(\omega_h) = \frac{1}{2\pi}\sum_{k=-M}^{M} w_k \hat{\Gamma}_X(k)\exp(-i\omega_h k)$, $w_k = 1 - \frac{|k|}{M+1}$; (iii) let $D_q(\omega_h)$ be a diagonal matrix with the q largest eigenvalues of $\hat{\Sigma}_X(\omega_h)$ on the diagonal, and let $U_q(\omega_h)$ be the corresponding matrix of eigenvectors. Inverse Fourier transform $\hat{\Sigma}_\chi(\omega_h) = U_q(\omega_h)D_q(\omega_h)U_q(\omega_h)'$ to obtain $\tilde{\Gamma}_\chi(k) = \frac{2\pi}{2H+1}\sum_{h=-H}^{H} \hat{\Sigma}_\chi(\omega_h)\exp(i\omega_h k)$; (iv) repeat step (iii) using the $q+1$ to N ordered eigenvalues values to obtain $\tilde{\Gamma}_e(k)$; (v) let Z be the r generalized eigenvectors (with eigenvalues in descending order) of $\tilde{\Gamma}_\chi(0)$ with respect to $\tilde{\Gamma}_e(0)$ under the normalization that $Z_j\tilde{\Gamma}_e(0)Z'_i = 1$ if $i = j$ and zero otherwise.[2] The estimated dynamic factors are:

$$\tilde{F} = XZ.$$

[2]In practice, FHLR only used the diagonal elements of $\tilde{\Gamma}_e(0)$ in this step.

The in-sample estimate of the common component is obtained from an artificial projection of (the unobserved) χ_{it} on $\tilde{F}_t$:

$$\tilde{\chi} = XZ\Big(Z'\hat{\Gamma}_X(0)Z\Big)^{-1}Z'\tilde{\Gamma}_\chi(0).$$

Given $\tilde{\chi}$, $\tilde{e} = x - \tilde{\chi}$ can be defined residually.

While the static and dynamic estimators can consistently estimate the static and dynamic factor space respectively, there are notable differences in terms of implementation. First, the static method requires only the specification of r. The dynamic method requires input of four parameters q, M, H, and s (or r since $r = q(s+1)$). Second, the dynamic estimates are obtained from an eigenvalue decomposition of the spectrum smoothed over different frequencies, while the static estimates are obtained from the sample covariance matrix. Evidently, $\hat{F}$ obtains as a special case of $\tilde{F}$ with $M = 0$. Third, the dynamic approach performs a generalized eigenvalue decomposition, while the static approach performs a simple eigenvalue decomposition of the covariance matrix. The former effectively scales the data by the standard deviation of the idiosyncratic components, while the latter works with data standardized to have unit variances. The generalized principal components produce linear combinations of x_{it} that have the smallest ratio of the variance of the idiosyncratic to common component. Whether this normalization is more efficient for forecasting is an open question.

A drawback of the static estimator is that it does not take into account the dynamics among the factors, if they exist. For example, the model $x_{it} = F_t + F_{t-1} + e_{it}$ is viewed as having two static factors, even though there is only one common source of variation. Such "shifted" relation between F_t and x_{it} is dealt with by the dynamic estimator via evaluation of the peridogram at different frequencies. On the other hand, if such shifted relation between F_t and x_{it} is not present in the data, unnecessary estimation of the spectral density matrices could induce efficiency loss. Therefore, neither estimator necessarily dominates the other. Which is more desirable ultimately depends on the data on hand.

3. Step F: Forecasting

The object of interest is the h-step-ahead forecast $x_{iT+h|T}$. It follows from (7)–(9) that to form a feasible forecast, we need

$\hat{F}_{T+h|T}$, $\hat{e}_{iT+h|T}$, and/or $\hat{\chi}_{iT+h|T}$. Different possibilities for forecasting these components arise because when the parameters are not observed, $\hat{\chi}_{iT+h|T} = \lambda_i' \widehat{F_{T+h|T}} \neq \hat{\lambda}_i' \hat{F}_{T+h|T}$. Furthermore, an h-step-ahead forecast can be obtained as a sequence of one-step-ahead forecasts, or directly from a long-horizon forecasting equation.

Sequential One-Step Forecasts [S]. Obtain $\hat{\rho}_i(L)$ from a regression of $\hat{e}_{it}$ on p_e^S of its lags. Then, starting with $\hat{e}_{iT+1|T}$, form a sequence of one-step-ahead forecasts to yield

$$\hat{e}_{iT+h|T} = \hat{\rho}_i(L)\hat{e}_{iT+h-1|T}. \tag{10}$$

Direct h-step Forecasts [D]. Let $\hat{\varphi}_i(L)$ be the coefficients from a projection of $\hat{e}_{it+h}$ on $\hat{e}_{it}$ and p_e^D of its lags. Then

$$\hat{e}_{iT+h|T} = \hat{\varphi}_i(L)\hat{e}_{iT}. \tag{11}$$

As an example, if e_{it} is an AR(1) with $\rho_i(L) = \rho_i$, then $e_{iT+h|T} = \rho_i^h e_{iT}$. The sequential forecast is $(\hat{\rho}_i)^h e_{iT}$, while the direct forecast is $\hat{\rho_i^h} e_{iT}$.

Analogously, two forecasts of F_{T+h} are available:

$$\begin{aligned} \hat{F}_{T+h|T} &= \hat{A}(L)\hat{F}_{T+h-1|T} \\ \hat{F}_{T+h|T} &= \hat{\mathcal{A}}(L)\hat{F}_T, \end{aligned}$$

where $\hat{A}(L)$ and $\hat{\mathcal{A}}(L)$ are polynomials of order p_F^D and p_F^S, respectively. Marcellino, Stock, and Watson (2004) find in univariate and bivariate models that the sequential approach typically outperforms the direct approach, if the lag length is appropriately chosen. The present context is somewhat different as it involves $\hat{F}_t$, which has estimation errors.

It should be made clear that when there is more than one factor, vector-autoregressive forecasts of the factors should be considered, not univariate autoregressive forecasts. The reason is that we can only estimate the space spanned by the factors. The dynamics of an estimated factor need not coincide with the dynamics of any underlying factor. Thus, the information set is the history of all *estimated* factors.

Unrestricted Forecasts [U]. Consider the forecast

$$x_{iT+h|T} = \hat{\beta}_i(L)'\hat{F}_T + \hat{\varphi}_i(L)x_T, \tag{12}$$

where $\hat{\beta}_i(L)$ and $\hat{\varphi}_i(L)$ obtained when x_{it+h} is regressed on p_F lags of $\hat{F}_t$ and p_x lags of x_{it}. We refer to this as an unrestricted forecast because $\hat{\beta}_i'(L)$ is not constrained to equal $(1-\hat{\rho}_i(L)L)\hat{\lambda}_i'$, and no restriction is placed between the coefficients on the p_F lags of $\hat{F}_t$ and the p_x lags of x_{it}.

Nonparametric Forecasts [N]. A forecast of χ_{iT+h} can be obtained by artificially projecting each χ_{it+h} on $\tilde{F}_t$ and then replacing the population matrices by sample estimates. This yields

$$\tilde{\chi}_{T+h|T} = x_T \bullet Z\Big(Z'\hat{\Gamma}_X(0)Z\Big)^{-1} Z'\tilde{\Gamma}_\chi(h). \tag{13}$$

Notice that parametric estimation of time series models for $\tilde{\chi}_{it}$ is not necessary, nor are explicit estimates of F_t. One only needs Z and $\tilde{\Gamma}_\chi(k)$, which are provided by step E. For this reason, we refer to these as nonparametric forecasts (denoted with a tilde). In contrast, the other three methods are based on parametric regression models with estimates of F_t as regressors.

4. Five Diffusion Forecasts Compared

Let XY be a diffusion forecast that uses method X in step E and Y in step F. Given the two alternatives ([S]tatic or [D]ynamic) for step E and the four alternatives ([S]equential, [D]irect, [U]nrestricted, [N]onparametric) for step F, we have the following:

$$\begin{aligned}
\text{SU: } \hat{x}_{iT+h|T} &= \hat{\beta}_i'(L)\hat{F}_T + \hat{\gamma}_i(L)x_{iT} \\
\text{DU: } \hat{x}_{iT+h|T} &= \tilde{\beta}_i'(L)\tilde{F}_T + \tilde{\gamma}_i(L)x_{iT} \\
\text{SS: } \hat{x}_{iT+h|T} &= \hat{\lambda}_i'\hat{A}(L)\hat{F}_{T+h-1|T} + \hat{\rho}_i(L)\hat{e}_{iT+h-1|T} \\
\text{SD: } \hat{x}_{iT+h|T} &= \hat{\lambda}_i'\hat{\mathcal{A}}(L)\hat{F}_T + \hat{\varphi}_i(L)\hat{e}_{iT} \\
\text{DN: } \hat{x}_{iT+h|T} &= \tilde{\chi}_{iT+h|T} + \hat{\varphi}_i(L)\tilde{e}_{iT}.
\end{aligned}$$

We have not considered the sequential forecasts

$$\hat{x}_{iT+h|T} = \hat{\theta}_i(L)\hat{\chi}_{iT+h-1|T} + \hat{\rho}_i(L)\hat{e}_{it+h-1|T}$$

or the direct forecast

$$\hat{x}_{iT+h|T} = \hat{\Theta}_i(L)\hat{\chi}_{iT} + \hat{\varphi}_i(L)\hat{e}_{iT},$$

even though we could have obtained $\hat{\theta}_i(L)$ and $\hat{\Theta}_i(L)$ from least-squares regression of $\hat{\chi}_{it}$ on its lags. As explained above, there is less information in the history of $\hat{\chi}_{it}$ than the history of the r estimated factors separately. Simulations confirm that forecasts using lags of $\hat{\chi}_{it}$ are inferior to forecasts using lags of $\hat{F}_t$.

The five methods above all "plug" forecasts of $\hat{e}_{iT+h}$, $\hat{F}_{T+h}$, or $\hat{\chi}_{iT+h}$ into (7)–(9), and in this sense are all diffusion index forecasts. They differ in the implementation of step E and/or step F. For example, SS and SD should yield identical forecasts if $\hat{A}(L)\hat{F}_{T+h-1} = \hat{\mathcal{A}}(L)F_T$. If F_t is a scalar AR(1), this holds if $\hat{\mathcal{A}} = \hat{A}^h$. Whereas the parameters of SU and DU are estimated directly from the forecasting equation, the other three are two-step procedures that forecast the factors and the components separately. The factor structure is thus maintained in step F of SS, SD, and DN.

To get a sense of what it means to impose the factor structure on step F, consider the SD. A regression of $\hat{F}_t$ on $\hat{F}_{t-h}$ yields $\bar{\beta} = (\hat{F}'_{-h}\hat{F}_{-h})^{-1}\hat{F}'_{-h}\hat{F}$ and thus $\hat{F}_{T+h|T} = \bar{\beta}\hat{F}_T$. The SD forecast, being $\hat{\lambda}_i'\hat{F}_{T+h|T} = \hat{\lambda}_i'\bar{\beta}\hat{F}_T$, is

$$\hat{x}_{iT+h|T} = V_{i\,\bullet}(V'\Gamma_X(0)V)^{-1}V'\Gamma_X(h)VV'x'_{T\,\bullet}.$$

The SU (see below) imposes $V'\hat{\Gamma}_X(h)V$ to be an identity matrix. By not imposing this constraint, the SD allows the estimated factor loadings to enter the forecast.

It has often been thought that the difference between SW and FHLR is how the factors are estimated. If step E was the only difference, a comparison of SY with DY would have been appropriate, where $Y \in (S, D, N, U)$. But in fact, what is implemented by Stock and Watson is SU, while FHLR adopt DN. The important difference is that FHLR exploit the factor structure in both steps E and F, while Stock and Watson do not impose the factor structure in step F.

The contrast between the two methods can be made more transparent if we let e_{it} be iid so $x_{iT+h|T} = \chi_{iT+h|T}$. The Stock and Watson forecast (which in our notation is SU) begins with $\hat{\beta} = (\hat{F}'_{-h}\hat{F}_{-h})^{-1}\hat{F}'_{-h}x_{\bullet i}$, or more precisely

$$\hat{\beta} = (V'\hat{\Gamma}_X(0)V)^{-1}V'X'_h x_{\bullet i} = (V'\hat{\Gamma}_X(0)V)^{-1}V'[\hat{\Gamma}_X(h)]_{\bullet\, i}.$$

The forecast is $\hat{x}_{iT+h|T} = \hat{\beta}'\hat{F}_T$. With $\hat{F}'_T = x_{T\,\bullet}V$, we have

$$\hat{x}_{iT+h|T} = \left[\hat{\Gamma}_X(h)\right]_{i\,\bullet} V\left(V'\hat{\Gamma}_X(0)V\right)^{-1}V'x'_{T\,\bullet},$$

where V are the eigenvectors of $\hat{\Gamma}_X(0)$. On the other hand, the FHLR forecast (which in our notation is DN), can be shown to be

$$\tilde{x}_{iT+h|T} = \left[\tilde{\Gamma}_\chi(h)\right]_{i\,\bullet} Z\left(Z'\hat{\Gamma}_X(0)Z\right)^{-1}Z'x'_{T\,\bullet},$$

where Z are the generalized eigenvectors associated with $\tilde{\Gamma}_\chi(0)$.

Clearly, the difference is not just V versus Z. The SU forecast involves the matrix $[\hat{\Gamma}_X(h)]_{\bullet\, i}$ while the DN forecast involves the matrix $[\tilde{\Gamma}_\chi(h)]_{\bullet\, i}$. Essentially, the SU treats $\hat{F}_T$ like any other conditioning variable, without insisting that $x_{iT+h|T} = \lambda'_i F_{T+h|T}$. The DN makes full use of the factor structure to forecast $\tilde{\chi}_{iT+h}$ explicitly.

If a forecast of χ_{iT+h} is the objective, the SU method cannot be expected to perform well because x_{iT+h} is χ_{iT+h} measured with error. But the objective is to forecast x_{iT+h}, not χ_{iT+h}. The SU can be effective as it produces a forecast of $x_{iT+h|T}$ directly. The DN, on the other hand, forecasts the *estimated* components of x_{iT+h} separately. Thus, precise estimation of the factor space plays a more important role under DN than SU.

A fundamental distinction between SU and DN is thus whether the factor structure is imposed on step F. The question is relevant only because the true parameters are unobserved, and to the extent that the parameters and factor space can be consistently estimated, both methods should yield forecasts that converge to the true conditional mean. Stock and Watson (2002a) showed that the SU will consistently forecast the conditional mean when $N, T \to \infty$ with no restriction on the relation between N and T. The DN also provides a consistent forecast, but appears to require that $N/T \to 0$. Furthermore, as is well known, forecasting the components even when they are observed will generally yield results that are different from forecasting the series directly. Thus, even if step E was held fixed, the SU and DN can be expected to yield different forecasts in finite samples.

5. Simulations

We consider two Monte Carlo experiments. In the first, we specify the dynamics of F_t to better understand the sensitivity of the forecast methods to the true factor processes. The second experiment takes the dynamics of $\hat{F}_t$ as given by the data to shift focus to the idiosyncratic errors. Specifically, the error structure is fully calibrated to the data. To our knowledge, this is the first assessment of diffusion index forecasts in a calibrated environment.

5.1 Static Versus Dynamic Factors

For $i = 1, \dots N$, $t = 1, \dots T$, $j = 1, 2$, $u_{jt} \sim N(0,1)$, and $\lambda_{ij} \sim N(0,1)$, we consider:

DGP (1). 2 Static Factors, $r = q = 2$, $\alpha_1 = .8, \alpha_2 = .4$

$$\begin{aligned} x_{it} &= \lambda_{i1} F_{1t} + \lambda_{i2} F_{2t} + e_{it} \\ F_{jt} &= \alpha_j F_{jt-1} + u_{jt}. \end{aligned}$$

DGP (2). 2 Dynamic Factors, $s = 3$, $q = 2$, $\alpha_1 = .8, \alpha_2 = .4$

$$\begin{aligned} x_{it} &= \sum_{k=0}^{s} \lambda_{i1k} f_{1t-k} + \sum_{k=0}^{s} \lambda_{i2k} f_{2t-k} + e_{it} \\ f_{jt} &= \alpha_j f_{jt-1} + u_{jt}. \end{aligned}$$

DGP (3). 1 Static + 1 Dynamic Factors, $s = 3$, $q = 2$, $\alpha_1 = .8$, $\alpha_2 = .7$

$$\begin{aligned} x_{it} &= \lambda_{i1} F_{1t} + \sum_{k=0}^{s} \lambda_{i2k} f_{2t-k} + e_{it} \\ F_{1t} &= \alpha_1 F_{1t-1} + u_{1t} \\ f_{2t} &= \alpha_2 f_{2t-1} + u_{2t} \end{aligned}$$

DGP (4). 2 Static Factors, $r = q = 2$, $\alpha_1 = .8, \alpha_2 = .5, \theta_1 = .5$

$$\begin{aligned} x_{it} &= \lambda_{i1} F_{1t} + \lambda_{i2} F_{2t} + e_{it} \\ F_{1t} &= \alpha_1 F_{1t-1} + u_{1t} + \theta_1 u_{1t-1} \\ F_{2t} &= \alpha_2 F_{2t-1} + u_{2t}. \end{aligned}$$

DGP (5). 2 Static Factors, $r = q = 2$, $\alpha_1 = .8, \alpha_2 = .5, \theta_1 = -.5$

$$\begin{aligned} x_{it} &= \lambda_{i1}F_{1t} + \lambda_{i2}F_{2t} + e_{it} \\ F_{1t} &= \alpha_1 F_{1t-1} + u_{1t} + \theta_1 u_{1t-1} \\ F_{2t} &= \alpha_2 F_{2t-1} + u_{2t}. \end{aligned}$$

DGP (6). 2 Dynamic Factors, $s = 3$, $q = 2, \alpha_2 = .8$

$$\begin{aligned} x_{it} &= \sum_{k=0}^{s} \lambda_{i1k} f_{1t-k} + \sum_{k=0}^{s} \lambda_{i2k} f_{2t-k} + e_{it}, \\ f_{1t} &= u_{1t} \\ f_{2t} &= \alpha_2 f_{2t-1} + u_{2t}. \end{aligned}$$

DGP (7). 2 Dynamic Factors, $s = 3$, $q = 2$

$$\begin{aligned} x_{it} &= \sum_{k=0}^{s} \lambda_{i1k} f_{1t-k} + \sum_{k=0}^{s} \lambda_{i2k} f_{2t-k} + e_{it} \\ f_{jt} &= u_{jt}. \end{aligned}$$

Throughout, N is fixed at 147, the number of variables in the empirical application we consider below. For all i,

$$e_{it} = \kappa v_{it} \qquad v_{it} \sim N\big(0, \sigma_{vi}^2\big),$$

with κ chosen so that on average, the common component explains a fraction ϑ of the variance of x_{it}. That is,

$$\kappa = \frac{1-\vartheta}{\vartheta} \left(\frac{\frac{1}{N}\sum_{i=1}^{N} \operatorname{var}(\lambda_i' F_t)}{\frac{1}{N}\sum_{i=1}^{N} \operatorname{var}(e_{it})} \right).$$

We use $\vartheta = 0.5$, so that on average, 50 percent of the variation in x_{it} is explained by the common component. Following Forni et al. (2005), $\sigma_{vi}^2 \sim U(.1, 1.1)$. This means that even though ϑ is .5 on average, there is a good deal of variation in the size of the common component. Moreover, the variable to be forecasted is the first series in the panel. For this series, i.e., $i = 1$, $var(\lambda_1' F_t)/var(x_{1t}) = 0.75$.

The DGPs are designed to evaluate the sensitivity of the methods to the dynamics of the factor processes. Static and dynamic factors

are considered, as are mixtures of these. Recall that $r = q(s + 1)$. Thus, the DGPs encompass factor models with r as small as two, as in DGP 1, and as large as eight, as in DGP 7. It should be remarked that Stock and Watson's simulations are generally based upon DGP 1, while FHLR emphasize on DGP 7, which is a special case of DGP 2 with iid dynamic factors.

We use the AR(1) forecasts as benchmarks. Our criterion is the mean-squared error of out-of-sample forecasts. Kapetanios and Marcellino (2002) found that the static estimates provide better in-sample fit of the components, while the dynamic method is less able to distinguish the common and the idiosyncratic components. Forni et al. (2005) also reported huge discrepancies between in- and out-of-sample performance. It should be made clear that our objective here is not precise estimation of the components, but precise forecasts of a series that is the sum of two components. Any finding in favor of one method to be reported below should be interpreted with this objective in mind.[3]

We will refer to the ratio of the MSE for a given method to the MSE of an AR(1) as RMSE (relative mean-squared error). An entry less than one indicates that the diffusion index forecast is superior to the naïve AR(1) forecast. We report results for $h = 1, 2, 4, 6, 8$, and 12. Hereafter, we will refer to $h = 1, 2, 4$ as short-horizon forecasts, and $h = 6, 8, 12$ as long-horizon forecasts. In the simulations, we assume that r, s, q, and the lag order of the factors are known, expect for the ARMA factors, which will be discussed below. For estimation of the dynamic factors, we use the programs provided by FHLR. Still, we have to determine M and H. We use $M = H = 20$ throughout.[4]

We simulate data for $T = 300$ and consider the forecasts with $T = 100$, and then with $T = 300$. These correspond roughly to the number of observations for the sample 1950:1–1969:12 and 1950:1–1996:12. These are the beginning and end dates of the forecasting exercise in Stock and Watson. Given that $N = 147$, these two parameterizations

[3]The results for DU and DN here differ from the working version of this paper. The earlier results were based on demeaned instead of standardized data. The DU and DN tend to be larger than the ones reported here.

[4]The computer program distributed by FHLR seems to impose $M = H$, a restriction that is not necessary on theoretical grounds. Forni et al. (2005) set $M = \sqrt{T}$, but this rule is not used systematically in all papers implementing DN. Forni et al. (2000) use $M = \sqrt{T}/4$, EUROCOIN is based on $M = 18$, and Kapetanios and Marcellino (2002) use $M = 3$. In the empirical application below, we check the robustness of the results to alternative choices of M and H.

also shed some light on how the forecasts behave when N exceeds T, and vice versa.

Tables 1a and 1b report results for $T = 100$ and 300, respectively.[5] With respect to static versus dynamic factor estimates, SU and DU are similar, showing that the estimator per se is not a choice of first-order importance. With respect to sequential versus direct forecasts, SS tends to yield smaller errors than SD, especially when $T = 100$, but the differences are smaller when $T = 300$. This suggests that the choice of the forecasting equation can be especially important when the sample size is small.

Other results are noteworthy. First, the RMSEs are all below unity except for DGPs 4 and 5, which are ARMA factors with a moving average component. For these two DGPs, the DN is significantly better than the other four methods, especially when $T = 100$. Even with $T = 300$, the DN continues to be the better method for these DGPs. One reason could be because the true lag order of the first factor is infinity. In the simulations, we use an AR(3) approximation. As the autoregressive coefficients corresponding to an ARMA(.8, –.5) die off slowly, the AR(3) approximation is likely inadequate. It appears that dynamic misspecification can hamper the properties of the factor forecasts.

Second, while the differences across methods are small at $T = 300$, they are much larger at $T = 100$, where the DN and SS are clearly better at long-horizon forecasts. The advantage of the DN at long horizons is most evident under DGPs 2 and 7, which, incidentally, are the DGPs often considered by FHLR. However, the gain is smaller when we mix dynamic with static factors, as in DGP 3. Even though DN has smaller errors at long horizons than SU at $T = 100$, observe also that SS often has smaller errors than DN. This means the dynamic factor estimates need not always outperform the static estimates in forecasting. Attributing differences in forecast errors to the choice estimator would be misguided.

Taken together, this set of simulation results suggests that the choice of estimator per se does not seem to make a difference of first-order importance to the forecast errors. What seems important is how one combines steps E and F, especially when the time span of the data is not too long.

[5] In results not reported, we find that the higher α is, the lower the RMSE, though the relative rankings of the methods do not change.

Table 1a. RMSE (MSE Relative AR(4)) from 1,000 Simulations of DGP 1–7

Horizon	T	Forecasting Methods				
h	T	SU/AR	SD/AR	SS/AR	DN/AR	DU/AR
DGP1: Static AR(1) Factors						
1	100	0.87	0.86	0.86	0.87	0.86
2	100	0.85	0.85	0.83	0.85	0.85
4	100	0.92	0.89	0.87	0.91	0.92
6	100	0.96	0.94	0.87	0.94	0.96
8	100	0.94	0.91	0.87	0.91	0.94
12	100	0.98	0.92	0.87	0.91	0.98
DGP2: Dynamic AR(1) with $\alpha = 0.8$						
1	100	0.54	0.53	0.53	0.55	0.52
2	100	0.54	0.51	0.51	0.53	0.53
4	100	0.59	0.56	0.55	0.59	0.58
6	100	0.63	0.61	0.59	0.65	0.63
8	100	0.76	0.72	0.67	0.73	0.75
12	100	0.96	0.90	0.84	0.81	0.94
DGP3: Factor 1: Static AR(1); Factor 2: Dynamic AR(1)						
1	100	0.62	0.61	0.61	0.64	0.62
2	100	0.53	0.53	0.52	0.57	0.55
4	100	0.55	0.54	0.53	0.62	0.56
6	100	0.73	0.69	0.65	0.72	0.73
8	100	0.84	0.80	0.75	0.81	0.84
12	100	0.87	0.83	0.78	0.82	0.88
DGP4: Static, Factor 1: ARMA(0.8,0.5); Factor 2: AR(0.5)						
1	100	0.86	0.79	0.79	0.82	0.85
2	100	0.91	0.79	0.79	0.79	0.90
4	100	0.97	0.86	0.81	0.88	0.96
6	100	1.16	0.98	0.92	0.93	1.15
8	100	1.18	0.97	0.92	0.91	1.17
12	100	1.10	0.96	0.85	0.90	1.10

Table 1a (continued). RMSE (MSE Relative AR(4)) from 1,000 Simulations of DGP 1–7

Horizon	T	Forecasting Methods				
h	T	SU/AR	SD/AR	SS/AR	DN/AR	DU/AR
DGP5: Static, Factor 1: ARMA(0.8,–0.5); Factor 2: AR(0.5)						
1	100	1.12	1.01	1.01	1.00	1.12
2	100	1.08	0.97	0.97	0.98	1.07
4	100	1.09	0.98	0.98	0.98	1.09
6	100	1.19	1.06	1.00	0.99	1.19
8	100	1.14	1.04	0.99	1.00	1.14
12	100	1.24	1.09	1.00	0.96	1.23
DGP6: Dynamic, Factor 1: iid; Factor 2: AR(1)						
1	100	0.51	0.50	0.50	0.51	0.51
2	100	0.54	0.52	0.51	0.54	0.52
4	100	0.61	0.59	0.56	0.60	0.60
6	100	0.77	0.73	0.69	0.73	0.76
8	100	0.85	0.81	0.74	0.77	0.84
12	100	0.94	0.87	0.81	0.84	0.95
DGP7: Dynamic iid Factors (FHLR DGP)						
1	100	0.70	0.67	0.67	0.68	0.69
2	100	0.86	0.82	0.84	0.82	0.85
4	100	0.99	0.96	0.94	0.95	0.99
6	100	0.96	0.94	0.92	0.91	0.95
8	100	1.04	1.02	0.95	0.98	1.03
12	100	1.04	0.99	0.94	0.94	1.01

5.2 A Calibrated Monte Carlo

The simulations in the previous subsection assume that the idiosyncratic errors are serially and cross-sectionally uncorrelated to focus on the factor processes. We now consider a Monte Carlo that replicates the error structure of the macroeconomic data set with 147 series from 1959:1 to 1998:12. The same data have been used

Table 1b. RMSE (MSE Relative AR(4)) from 1,000 Simulations of DGP 1–7

h	T	SU/AR	SD/AR	SS/AR	DN/AR	DU/AR
DGP1: Static AR(1) Factors						
1	300	0.87	0.87	0.87	0.86	0.87
2	300	0.82	0.81	0.81	0.82	0.82
4	300	0.84	0.82	0.81	0.83	0.84
6	300	0.92	0.92	0.90	0.92	0.92
8	300	0.97	0.97	0.95	0.97	0.96
12	300	0.97	0.97	0.94	0.96	0.97
DGP2: Dynamic AR(1) Factors						
1	300	0.52	0.52	0.52	0.52	0.51
2	300	0.46	0.45	0.45	0.45	0.44
4	300	0.61	0.60	0.59	0.59	0.59
6	300	0.70	0.69	0.68	0.67	0.68
8	300	0.72	0.71	0.70	0.71	0.72
12	300	0.81	0.81	0.78	0.79	0.80
DGP3: Factor 1: Static AR(1); Factor 2: Dynamic AR(1)						
1	300	0.57	0.56	0.56	0.57	0.56
2	300	0.54	0.53	0.53	0.52	0.52
4	300	0.55	0.54	0.54	0.56	0.55
6	300	0.68	0.67	0.68	0.70	0.70
8	300	0.75	0.74	0.74	0.75	0.75
12	300	0.84	0.84	0.82	0.85	0.86
DGP4: Static, Factor 1: ARMA(0.8,0.5); Factor 2: AR(0.5)						
1	300	0.83	0.78	0.78	0.78	0.82
2	300	0.79	0.77	0.77	0.80	0.78
4	300	0.85	0.82	0.82	0.86	0.84
6	300	0.92	0.89	0.86	0.88	0.91
8	300	0.96	0.93	0.91	0.94	0.96
12	300	0.96	0.94	0.92	0.96	0.96

Table 1b (continued). RMSE (MSE Relative AR(4)) from 1,000 Simulations of DGP 1–7

h	T	SU/AR	SD/AR	SS/AR	DN/AR	DU/AR
DGP5: Static, Factor 1: ARMA(0.8,–0.5); Factor 2: AR(0.5)						
1	300	0.95	0.96	0.96	0.95	0.95
2	300	0.97	0.97	0.97	0.95	0.97
4	300	0.96	0.94	0.94	0.96	0.96
6	300	0.99	0.98	0.96	0.97	0.99
8	300	1.02	0.99	0.97	0.99	1.03
12	300	1.04	0.97	0.94	0.97	1.04
DGP6: Dynamic, Factor 1: iid; Factor 2: AR(1)						
1	300	0.47	0.47	0.47	0.46	0.45
2	300	0.49	0.49	0.49	0.48	0.47
4	300	0.56	0.56	0.56	0.56	0.55
6	300	0.64	0.63	0.63	0.64	0.63
8	300	0.75	0.75	0.74	0.75	0.74
12	300	0.76	0.75	0.72	0.77	0.76
DGP7: Dynamic iid Factors (FHLR DGP)						
1	300	0.72	0.70	0.70	0.73	0.75
2	300	0.75	0.75	0.75	0.78	0.79
4	300	0.91	0.90	0.92	0.93	0.89
6	300	0.93	0.92	0.91	0.93	0.92
8	300	0.98	0.96	0.95	0.97	0.92
12	300	0.96	0.95	0.93	0.96	0.95

by Stock and Watson in a number of studies.[6] We consider the forecasts for eight series, as in the empirical analysis to follow. These are industrial production (IP), real personal income less transfers (GMYXSPQ), real manufacturing trade and sales (MSMTQ), number of employees on nonagricultural payrolls (LPNAG), the consumer price index (PUNEW), the personal consumption expenditure

[6]See, for example, Stock and Watson (1999) and Stock and Watson (2002b).

deflator (GMDC), the CPI less food and energy (PUXX), and the producer price index for finished goods (PWFSA).

The calibration exercise aims to preserve the relative importance of the common and the idiosyncratic errors, the serial and cross-section correlation in the idiosyncratic errors, as well as potential parameter instability in the data. In the empirical analysis to follow, the first forecast is based on estimation over a sample with 133 observations (corresponding to 1959:3 to 1970:1), while the last estimation sample has 445 observations (corresponding to 1997:1). Recalibrating the model parameters and performing a Monte Carlo at each T would be extremely time consuming. Instead, starting with $T = 133$, T is extended every twelve months, a new model is recalibrated, and new $T + h$ forecasts are obtained. This exercise still takes two weeks to execute.

The calibration consists of estimating r static factors by the method of principal components at every T, where r is determined as discussed below. The estimated common component, $\hat{\chi}_{it} = \hat{\lambda}_i' \hat{F}_t, t = 1, \ldots T$ is then treated as fixed. Data are simulated by adding to $\hat{\chi}_{it}$ new draws of the idiosyncratic errors. More precisely, least-squares regression of the AR(1) model $\hat{e}_{it} = \rho_i \hat{e}_{it-1} + v_{it}$ yields $\hat{\rho}_{i,T}$ and $\hat{v}_{it}, i = 1, \ldots N$, noting that if the t-statistic for $\hat{\rho}_{i,T}$ is less than 2 in absolute value, we set $\hat{\rho}_{i,T}$ to zero. Resampling $\hat{v}_{\bullet,t}$ with replacement yields a new set of residuals—say, $\tilde{v}_{\bullet,t}$—which, along with $\hat{\rho}_{i,T}$, yields a T by N matrix of idiosyncratic errors that preserve the cross-correlation structure.[7] That is to say, if $\tau_{ij,T} = \frac{1}{T}\sum_{t=1}^{T} \hat{v}_{it}\hat{v}_{jt}$ is the cross-section covariance over a sample of size T, the simulated errors $\tilde{v}_{it}$ and $\tilde{v}_{jt}$ will have the same covariance on average. Allowing the serial and cross-section covariance structure to change over time permits us to evaluate the forecasts when the data are not covariance stationary. This might be of empirical relevance as macroeconomic data are known to exhibit substantial parameter instability.

The appendix provides summary statistics on the common and idiosyncratic components, computed at twenty-six values of T at which the model is calibrated. The mean of the nonzero $\hat{\rho}_{i,T}$ indicates

[7] An alternative procedure is to first estimate $\hat{\Omega}_v = \frac{1}{T}\sum_{t=1}^{T} \hat{v}_t \hat{v}_t'$, and then multiple random draws of N(0,1) errors into the choleski decomposition of $\hat{\Omega}$. By construction, the rank of $\hat{\Omega}_v$ is N minus the number of factors. The procedure would require dropping six series from the simulations, since we estimated six factors.

that many of the $\hat{e}_{it}$ are serially correlated, while the mean of $\hat{\tau}_{ij,T}$ indicates that some $\hat{v}_{it}$ pairs are quite strongly cross-correlated. While $\hat{\rho}_{i,T}$ appears quite stable over time, $\hat{\tau}_{ij,T}$ seems to have fallen over time on average. The number of series that are serial and cross-correlated has also fallen over the sample. It would be difficult to use simple, parametric models to capture such heterogeneity and parameter instability. The present calibrated Monte Carlo aims to shed light on the factor forecasts in a realistic setting.

The number of factors, the dynamic structure of the factors, and the dynamic structure of the idiosyncratic errors are now treated as unknown parameters. The BIC is used to jointly determine p_x, p_F, and r in SU, based on the static principal components. The r used in SU is also used for SD and SS. The BIC is then used to determine q_x, q_F, and r in DU, which is based on the generalized principal components. The r used for DU is also used for DN with q fixed to 2. The BIC is also used to determine p_e^D and p_e^S, the lag order of the idiosyncratic errors in SS and SD, as well as p_F^S and p_F^D, the lag order of the VAR in the factors in SS and SD. Note that these parameters are repeatedly reoptimized because the model is recalibrated each time the sample is extended. We fix throughout q to 2, $M = H = 20$ when constructing $\hat{\Gamma}_\chi(0)$ and $\hat{\Gamma}_e(0)$. The benchmark forecast is based on an AR(4).

The results in tables 2a and 2b are averaged over forecasts made at the twenty-six values of T. On average, the estimation consists of 289 observations. Our eight series can be classified into two groups—four nominal and four real. For the four series reported in table 2a, we see that the SU and DU now have *significantly* smaller RMSEs than the other methods at long horizons. In fact, the DN loses the edge in long-horizon forecasts that it enjoyed in the simple Monte Carlo experiments. In this calibrated setting, the two unconstrained forecasts of the real variables tend to outperform the three forecasts that are based on models with more structure.

For the four series considered in table 2b, a notable result is that the factor forecasts yield more modest improvements over the AR forecasts. Only in long-horizon SU and DU forecasts for PUNEW and GMDC did we witness an RMSE below .80. At long horizons, the SD can produce forecasts that are much inferior to the AR. Of all the methods considered, the SU and DU appear to make larger gains over the AR forecasts at every horizon.

Table 2a. RMSE for Calibrated DGP from 300 Simulations, Real Variables

Horizon	Variable	Forecasting Methods				
		SU/AR	SD/AR	SS/AR	DN/AR	DU/AR
1	IP	0.87	0.80	0.80	0.88	0.84
	GMYXSPQ	0.75	0.79	0.79	0.84	0.74
	MSMTQ	0.83	0.87	0.87	0.87	0.83
	LPNAG	0.79	0.71	0.71	0.84	0.79
2	IP	0.73	0.70	0.69	0.81	0.72
	GMYXSPQ	0.72	0.76	0.76	0.83	0.70
	MSMTQ	0.77	0.80	0.80	0.83	0.75
	LPNAG	0.83	0.70	0.69	0.86	0.83
4	IP	0.64	0.58	0.64	0.86	0.67
	GMYXSPQ	0.69	0.75	0.75	0.90	0.69
	MSMTQ	0.69	0.74	0.75	0.85	0.70
	LPNAG	0.67	0.63	0.65	0.74	0.69
6	IP	0.56	0.58	0.59	0.80	0.62
	GMYXSPQ	0.63	0.71	0.73	0.83	0.60
	MSMTQ	0.64	0.70	0.73	0.81	0.64
	LPNAG	0.57	0.69	0.65	0.69	0.59
8	IP	0.63	0.68	0.69	0.81	0.68
	GMYXSPQ	0.67	0.76	0.80	0.82	0.60
	MSMTQ	0.64	0.70	0.77	0.81	0.64
	LPNAG	0.64	0.81	0.71	0.69	0.62
12	IP	0.65	0.77	0.75	0.80	0.68
	GMYXSPQ	0.64	0.77	0.85	0.81	0.61
	MSMTQ	0.63	0.71	0.79	0.81	0.66
	LPNAG	0.61	0.88	0.76	0.73	0.61

Table 2b. RMSE for Calibrated DGP from 300 Simulations, Nominal Variables

Horizon	Variable	Forecasting Methods				
		SU/AR	SD/AR	SS/AR	DN/AR	DU/AR
1	PUNEW	0.91	0.93	0.93	0.93	0.91
	GMDC	0.88	0.90	0.90	0.90	0.89
	PUXX	0.97	0.98	0.98	0.97	0.97
	PWFSA	0.94	0.95	0.95	0.95	0.95
2	PUNEW	0.91	0.92	0.92	0.95	0.92
	GMDC	0.87	0.93	0.93	0.92	0.89
	PUXX	0.96	0.97	0.97	0.97	0.97
	PWFSA	0.94	0.95	0.95	0.96	0.94
4	PUNEW	0.84	0.88	0.88	0.92	0.86
	GMDC	0.78	0.94	0.94	0.88	0.80
	PUXX	0.95	0.97	0.97	0.96	0.96
	PWFSA	0.91	0.93	0.93	0.94	0.92
6	PUNEW	0.83	0.89	0.88	0.91	0.85
	GMDC	0.76	0.97	0.96	0.85	0.78
	PUXX	0.94	0.99	0.98	0.96	0.96
	PWFSA	0.91	0.94	0.94	0.94	0.92
8	PUNEW	0.86	0.93	0.91	0.92	0.88
	GMDC	0.77	1.03	0.99	0.85	0.80
	PUXX	0.95	1.03	1.01	0.95	0.96
	PWFSA	0.94	0.97	0.96	0.95	0.94
12	PUNEW	0.87	1.04	0.96	0.92	0.90
	GMDC	0.76	1.13	1.04	0.83	0.83
	PUXX	0.94	1.10	1.04	0.95	0.97
	PWFSA	0.94	1.02	0.97	0.95	0.95

Comparing DN with DU in tables 2a and 2b, the DU generally has significantly smaller errors than the DN. Comparing SU with SD and SS, the SU tends to be the better of the three. These differences underscore the point that step F can generate important differences in forecast errors. As noted earlier, even when the factors and the idiosyncratic errors are observed, how much forecast improvement the factors can provide depends on λ_i and the difference between the dynamics of the factors and the errors. While no one method systematically outperforms the others for the real series at short horizons (i.e., $h = 1, 2, 4$), the SU always has the smallest error for forecasting the nominal series, and the SU and DU are best for forecasting the real series at longer horizons.

An overview of the results in tables 1 and 2 is as follows. For simple data-generating processes such as those in table 1, the SU or DU cannot be supported as the best method. However, once we consider more complex error structures such as the ones encountered in the data, the two unrestricted forecasts become noticeably better than all other methods. Unlike the SS and SD, the SU and DU do not specify the dynamics of $\hat{F}_t$. As well, many auxiliary parameters need to be chosen in order to estimate the dynamic factors. Simple implementation and leaving the forecasting equation with the flexibility to adapt to the complex properties data may explain why the unconstrained forecasts perform much better when the errors are heterogeneous in many dimensions. It remains to consider whether the various forecasting methods perform in the data as they do in simulations.

6. Application to Eight Series

Results for various diffusion index forecasts are available in the literature. As is clear from the discussion above, there are many ways to construct forecasts using the factor estimates. FHLR reported results for static forecasts, but they implemented what would have been SN in our notation, not the SU that Stock and Watson used. On the other hand, Stock and Watson evaluated the DU forecasts,[8] but these are not the same as the DN that FHLR proposed. As well, the results are reported for different forecast horizons, and using

[8]See for example, Stock and Watson (2004a) and Forni et al. (2005).

different criteria. Here, we provide an objective comparison of the various methods, making clear the role of steps E and F.

We apply the five methods to eight series: IP, GMYXSPQ, MSMTQ, LPNAG, PUNEW, GMDC, PUXX, and PWFSA. The goal is to forecast the growth rates of the real variables and inflation rates h periods hence. For x_{it}, we consider a balanced panel of N=147 monthly series available from 1959:1 to 1998:12. Following Stock and Watson (2002b), the data are standardized and transformed to achieve stationarity where necessary before the factors are estimated. The logarithms of the four real variables are assumed to be $I(1)$, while the logarithms of the four prices are assumed to be $I(2)$.

The forecasting exercise begins with data from 1959:3–1970:1. An h-period-ahead forecast is formed by using values of the regressors at 1970:1 to give $y^h_{1970:1+h}$. The sample is updated by one month, the factors and the forecasting equation are both reestimated, and an h-month forecast for 1971:2 is formed. The final forecast is made for 1998:12 in 1998:12–h. Recursive AR(p) forecasts are likewise constructed. Several auxiliary parameters must again be chosen. As in the calibrated Monte Carlo, we determine q_x, r, and q_F jointly using the BIC. Given this r, we then use the BIC to determine p^D_F, p^S_F, p^D_e, p^S_e. For the dynamic factors, we again set M to 20 and fix q to 2.

Table 3 reports the MSE relative to the optimal AR(p) model, where p is also chosen by the BIC. Overall, the improvements of the factor forecasts over the autoregressive forecasts for the inflation series (table 3b) are modest.[9] However, the SU and DU are noticeably better than the three methods that maintain a factor structure in the forecasting equation. Interestingly, the SU did not do as well in forecasting GMYXSPQ and MSMTQ in the calibrated simulations, and they also do not perform favorably in the empirical exercise.

For the real series (table 3a), all methods are quite similar at $h = 1$. However, there are nontrivial differences at other horizons. The RMSEs are all below unity at all forecast horizons, and they generally fall with h. Observe that (i) SS and SU tend to outperform SD, (ii) DU outperforms DN in twenty-three of twenty-four cases,

[9]Results in Inoue and Kilian (2003) suggest that for data with a weak factor structure, factor forecasts might be less effective.

Table 3a. RMSE for Real Variables, 1970:1998

Horizon	Variable	Forecasting Methods				
		SU/AR	SD/AR	SS/AR	DN/AR	DU/AR
1	IP	0.83	0.86	0.86	0.85	0.82
	GMYXSPQ	0.81	0.84	0.84	0.90	0.87
	MSMTQ	0.87	0.88	0.88	0.90	0.88
	LPNAG	0.83	0.97	0.97	0.89	0.85
2	IP	0.73	0.79	0.77	0.80	0.73
	GMYXSPQ	0.77	0.78	0.76	0.82	0.75
	MSMTQ	0.83	0.84	0.84	0.87	0.82
	LPNAG	0.74	0.88	0.86	0.79	0.78
4	IP	0.66	0.70	0.67	0.73	0.78
	GMYXSPQ	0.76	0.71	0.69	0.77	0.67
	MSMTQ	0.76	0.75	0.74	0.83	0.75
	LPNAG	0.64	0.87	0.82	0.71	0.66
6	IP	0.55	0.67	0.63	0.74	0.64
	GMYXSPQ	0.69	0.72	0.69	0.75	0.65
	MSMTQ	0.76	0.74	0.73	0.83	0.72
	LPNAG	0.60	0.74	0.71	0.69	0.59
8	IP	0.56	0.71	0.64	0.75	0.58
	GMYXSPQ	0.74	0.73	0.70	0.75	0.66
	MSMTQ	0.80	0.78	0.76	0.86	0.75
	LPNAG	0.59	0.73	0.68	0.70	0.60
12	IP	0.49	0.60	0.53	0.71	0.55
	GMYXSPQ	0.70	0.74	0.72	0.76	0.63
	MSMTQ	0.80	0.75	0.74	0.88	0.78
	LPNAG	0.49	0.60	0.57	0.64	0.54

and (iii) all of the twenty-four best forecasts are generated by SU or DU, which do not impose a factor or a dynamic structure on the forecasts. Looking across series and forecast horizon, SU has the

Table 3b. RMSE for Nominal Variables, 1970:1998

Horizon	Variable	Forecasting Methods				
		SU/AR	SD/AR	SS/AR	DN/AR	DU/AR
1	PUNEW	0.96	0.99	0.99	1.00	0.97
	GMDC	0.95	0.99	0.99	0.99	0.95
	PUXX	0.93	1.00	1.00	1.00	0.93
	PWFSA	0.94	0.98	0.98	0.98	0.94
2	PUNEW	0.86	0.89	0.89	0.91	0.85
	GMDC	0.93	0.96	0.96	0.95	0.91
	PUXX	0.85	0.98	0.98	0.93	0.84
	PWFSA	0.91	0.97	0.97	0.98	0.91
4	PUNEW	0.68	0.77	0.77	0.84	0.66
	GMDC	0.84	0.92	0.92	0.91	0.85
	PUXX	0.80	0.98	0.98	0.92	0.80
	PWFSA	0.80	0.87	0.87	0.86	0.84
6	PUNEW	0.65	1.42	1.42	0.95	0.65
	GMDC	0.83	0.96	0.95	0.96	0.87
	PUXX	0.79	0.98	0.98	0.93	0.81
	PWFSA	0.75	0.90	0.90	0.85	0.77
8	PUNEW	0.65	1.71	1.67	1.01	0.67
	GMDC	0.82	1.31	1.30	0.98	0.90
	PUXX	0.81	0.97	0.97	0.99	0.81
	PWFSA	0.75	0.96	0.95	0.87	0.76
12	PUNEW	0.55	1.53	1.51	0.89	0.62
	GMDC	0.73	1.29	1.27	0.94	0.82
	PUXX	0.77	0.99	0.98	0.94	0.78
	PWFSA	0.71	0.92	0.90	0.83	0.71

smallest error in thirteen cases. The DU is best in eleven cases. These results are broadly similar to our calibrated Monte Carlo.

Result (i) is consistent with Marcellino, Stock, and Watson (2004) that sequential forecasts tend to outperform direct forecasts. Results (i) and (ii) reinforce the point that step F can yield important

differences in the forecasting errors. Result (iii) points to the robustness of the unconstrained forecasts. The choice of static versus dynamic factor estimates is less important. This is at odds with the perception that a fully specified dynamic factor model (the DN) should yield better forecasts. We offer several explanations. The first is simply that the static factor model is a better characterization of the data. The dynamic estimates would then unnecessarily smooth the spectrum over different frequencies and suffer efficiency loss. Second, the DN has only been shown to be more efficient in counterfactually simple examples. Whether the dynamic estimator remains efficient when, for example, $s > 0$ for only a subset of the factors is unclear. Furthermore, it should be kept in mind the object of interest is forecast of a series, not precise estimation of the factors or the common component per se. That the Stock-Watson unconstrained method produces better forecasts does not mean it will produce a more precise estimate of the common component.

Third, our results might reflect the fact that we have not tuned the parameters to maximize the efficiency of the dynamic estimator. To get a sense of this issue, table 4 reports additional results for the DN. Instead of selecting s as described above, we first compute the out-of-sample forecasts for every configuration of the parameters for the full sample of 456 observations. The optimal s is the one that minimizes the time-averaged MSE. This method, which we refer to as method B in table 4, thus fixes s over the entire sample period.[10] The results in table 4 suggest that method B produces better forecasts for the real series, but not the nominal series. We next consider the sensitivity of the DN to the choice of M and H. In addition to the base case of $M = H = 20$, table 4 also provides results for $M = H =$ 12, 50 and $\sqrt{T}$, as well as alternative values of M holding H fixed at 50. The RMSEs are generally smaller with $H = M = 50$. Overall, table 4 suggests that the dynamic factor forecasts can indeed be improved with better choices of the auxiliary parameters. However, it is far less clear that the improvements will be large enough to beat the SU or the DU, at least for the series considered.

Taking the empirical and simulation results together, we find that the SU and DU perform better in practice and in simulations

[10]This method was considered in the working paper version of Forni et al. (2005). Note that in contrast to other methods, this is not, strictly speaking, an out-of-sample exercise since the auxiliary parameters are selected on the basis of the out-of-sample performance.

Table 4a. Robustness Check for DN: RMSE for Real Variables, 1970:1998

Horizon	Variable	Forecasting Methods							
		M=20 H=20	Method B	M=12 H=12	M=50 H=50	M=$\sqrt{T}$ H=$\sqrt{T}$	M=12 H=50	M=20 H=50	M=$\sqrt{T}$ H=50
1	IP	0.85	0.84	0.83	0.87	0.83	0.98	0.98	0.98
	GMYXSPQ	0.90	0.80	0.88	0.83	0.88	1.01	1.01	1.01
	MSMTQ	0.90	0.90	0.90	0.90	0.90	1.01	1.01	1.01
	LPNAG	0.89	0.86	0.88	0.88	0.89	1.00	1.00	1.00
2	IP	0.80	0.76	0.75	0.79	0.76	0.96	0.96	0.96
	GMYXSPQ	0.82	0.72	0.78	0.72	0.78	1.00	1.00	1.00
	MSMTQ	0.87	0.84	0.87	0.83	0.88	0.99	0.99	0.99
	LPNAG	0.79	0.78	0.76	0.79	0.77	1.01	1.01	1.01
4	IP	0.73	0.69	0.72	0.73	0.72	0.96	0.96	0.96
	GMYXSPQ	0.77	0.66	0.75	0.65	0.76	1.00	1.00	1.00
	MSMTQ	0.83	0.76	0.82	0.77	0.84	1.01	1.01	1.01
	LPNAG	0.71	0.70	0.68	0.70	0.69	0.98	0.98	0.98
6	IP	0.74	0.64	0.70	0.67	0.72	0.98	0.98	0.98
	GMYXSPQ	0.75	0.64	0.75	0.64	0.75	0.99	0.99	0.99
	MSMTQ	0.83	0.76	0.84	0.74	0.85	1.03	1.03	1.03
	LPNAG	0.69	0.65	0.65	0.64	0.66	0.99	0.99	0.99
8	IP	0.75	0.61	0.72	0.62	0.74	0.98	0.98	0.98
	GMYXSPQ	0.75	0.63	0.77	0.62	0.76	0.99	0.99	0.99
	MSMTQ	0.86	0.75	0.87	0.74	0.89	1.04	1.04	1.04
	LPNAG	0.70	0.63	0.65	0.62	0.66	0.99	0.99	0.99
12	IP	0.71	0.60	0.69	0.56	0.71	0.97	0.97	0.97
	GMYXSPQ	0.76	0.64	0.82	0.63	0.80	1.00	1.00	1.00
	MSMTQ	0.88	0.78	0.92	0.78	0.92	1.02	1.02	1.02
	LPNAG	0.64	0.64	0.65	0.61	0.63	0.99	0.99	0.99

in which the parameters of the model are unknown. As seen from (5), what matters for forecasting is whether $\hat{F}_t$ has predictive power beyond information in lags of x_{it}. The forecasting equation need not be a full-blown factor model. The stronger the adherence to a factor structure, the more likely that the sampling variability of the

Table 4b. Robustness Check for DN: RMSE for Nominal Variables, 1970:1998

Horizon	Variable	Forecasting Methods							
		M=20 H=20	Method B	M=12 H=12	M=50 H=50	M=$\sqrt{T}$ H=$\sqrt{T}$	M=12 H=50	M=20 H=50	M=$\sqrt{T}$ H=50
1	PUNEW	1.00	0.96	1.00	1.01	1.00	1.00	1.00	0.97
	GMDC	0.99	1.00	0.99	0.99	0.99	0.99	0.99	0.95
	PUXX	1.00	0.96	1.00	1.00	1.00	1.00	1.00	0.93
	PWFSA	0.98	0.98	0.98	0.98	0.98	0.98	0.98	0.94
2	PUNEW	0.91	0.93	0.91	0.91	0.90	0.98	0.98	0.85
	GMDC	0.95	0.99	0.96	0.96	0.96	0.98	0.98	0.91
	PUXX	0.93	0.95	0.92	0.94	0.91	1.02	1.02	0.82
	PWFSA	0.98	0.97	0.97	0.97	0.97	0.98	0.98	0.93
4	PUNEW	0.84	0.91	0.82	0.83	0.83	0.98	0.98	0.69
	GMDC	0.91	0.97	0.92	0.90	0.92	0.99	0.99	0.83
	PUXX	0.92	0.96	0.90	0.93	0.89	1.11	1.11	0.77
	PWFSA	0.86	0.94	0.85	0.87	0.86	0.92	0.92	0.84
6	PUNEW	0.95	0.88	0.95	0.87	0.94	0.99	0.99	0.68
	GMDC	0.96	0.96	0.94	0.90	0.95	0.98	0.98	0.85
	PUXX	0.93	0.94	0.97	0.95	0.95	1.16	1.16	0.81
	PWFSA	0.85	0.95	0.84	0.85	0.84	0.92	0.92	0.79
8	PUNEW	1.01	0.85	0.99	0.97	0.97	0.99	0.99	0.72
	GMDC	0.98	0.96	0.91	0.92	0.94	1.00	1.00	0.88
	PUXX	0.99	0.92	0.95	0.96	0.97	1.15	1.15	0.81
	PWFSA	0.87	0.98	0.85	0.88	0.85	0.95	0.95	0.80
12	PUNEW	0.89	0.79	0.82	0.90	0.83	0.97	0.97	0.63
	GMDC	0.94	0.92	0.87	0.89	0.86	1.00	1.00	0.80
	PUXX	0.94	0.93	0.94	0.94	0.94	1.11	1.11	0.76
	PWFSA	0.83	0.91	0.80	0.83	0.79	0.95	0.95	0.79

factor estimates will enter the forecasts. When the parameters of the data-generating process are also unknown and/or are unstable, misspecification of the dynamics can magnify the sampling variability of the factor estimates. The favorable properties of the unconstrained

forecasts are likely due to the minimal factor structure imposed on the forecasting equation and the simplicity in its implementation.

Finally, given that the SU and DU have rather similar finite sample properties, a comment on SU versus DU is in order. Of the two, the SU is much easier to implement. The SU only requires the user to determine r, p_x, and p_F by applying the BIC to the forecasting equation. Estimation of the dynamic factors necessitates choosing various parameters for which we have no guide. It should also be mentioned that in practice and following FHLR, the dynamic factors are constructed as the generalized principal components of $\hat{\Sigma}_x - \text{diag}(\tilde{\Gamma}_e(0))$, not $\tilde{\Gamma}_\chi(0) = \hat{\Sigma}_x - \tilde{\Gamma}_e(0)$ as theory suggests. A symmetric treatment would have the static factors estimated as the generalized principal components of $\hat{\Sigma}_x - \text{diag}(\hat{\Omega})$. Boivin and Ng (2004) suggest that this could further improve the SU. Such a method is not implemented because Ω is not a diagonal matrix in an approximate factor model. While the methods work in practice, more work is required at the theoretical level to justify its use.

7. Conclusion

In this paper, we seek to better understand how factors estimated from a large panel of data can be used in forecasting exercises. In principle, how the factors are estimated and how the forecasts are formed can both affect the mean-squared forecast error. We find that for simple error structure, differences across methods exist, especially when T is small, but the differences are not so strong as to immediately favor a particular method. When more complicated but realistic error structures are considered, the unconstrained method proposed by Stock and Watson works systematically better. The method also behaves noticeably better in the empirical analysis. This method simply augments estimates of the factors to an autoregressive forecasting equation. We attribute its performance to not imposing a tight factor structure and having to choose a small number of auxiliary parameters. This leaves the forecasting equation with more flexibility to adapt to the data. The fully specified dynamic factor forecasts have the potential for improvements, but more needs to be learned about how to adapt the auxiliary parameters to the data on hand. Finally, the static factors are easier to construct than dynamic factors, and are favored on practical grounds.

Appendix. Descriptive Statistics of the Idiosyncratic Term in the Calibrated DGP

Sample Size	$\lvert\rho_{i,T}\rvert$			$max_j \lvert\tau_{ij}\rvert$		$\frac{1}{N}\sum_j \lvert\tau_{ij}\rvert$		$\%\lvert\tau_{ij}\rvert > 0.2$	R_i^2	
	$1 - N^*/N$	Mean	Std	Mean	Std	Mean	Std		Mean	Std
133	0.21	0.43	0.19	0.71	0.22	0.12	0.03	0.15	0.36	0.28
145	0.23	0.42	0.18	0.71	0.22	0.12	0.04	0.16	0.36	0.27
157	0.18	0.41	0.19	0.71	0.22	0.11	0.03	0.14	0.41	0.27
169	0.16	0.39	0.18	0.69	0.23	0.11	0.03	0.15	0.46	0.28
181	0.21	0.39	0.19	0.66	0.23	0.11	0.03	0.13	0.55	0.26
193	0.15	0.39	0.20	0.67	0.23	0.11	0.03	0.12	0.52	0.27
205	0.15	0.39	0.20	0.67	0.23	0.11	0.03	0.12	0.52	0.27
217	0.13	0.39	0.21	0.67	0.23	0.10	0.03	0.11	0.52	0.27
229	0.17	0.40	0.21	0.67	0.23	0.10	0.03	0.11	0.52	0.27
241	0.17	0.39	0.21	0.67	0.23	0.11	0.03	0.13	0.53	0.26
253	0.18	0.38	0.19	0.67	0.23	0.11	0.03	0.12	0.53	0.26
265	0.20	0.38	0.20	0.64	0.23	0.10	0.03	0.11	0.60	0.26
277	0.17	0.37	0.20	0.63	0.23	0.10	0.03	0.10	0.60	0.26
289	0.17	0.38	0.20	0.63	0.23	0.10	0.02	0.10	0.60	0.26
301	0.14	0.38	0.20	0.63	0.23	0.10	0.02	0.10	0.60	0.26
313	0.12	0.37	0.20	0.63	0.22	0.10	0.02	0.10	0.60	0.26
325	0.14	0.38	0.20	0.63	0.22	0.10	0.02	0.10	0.59	0.26
337	0.12	0.38	0.21	0.63	0.22	0.09	0.02	0.10	0.59	0.26
349	0.16	0.38	0.21	0.60	0.22	0.09	0.02	0.10	0.64	0.24
361	0.15	0.39	0.21	0.61	0.22	0.09	0.02	0.10	0.64	0.24
373	0.12	0.38	0.21	0.63	0.22	0.09	0.02	0.10	0.59	0.26
385	0.16	0.38	0.21	0.61	0.22	0.09	0.03	0.10	0.64	0.24
397	0.15	0.38	0.20	0.59	0.22	0.09	0.02	0.09	0.66	0.24
409	0.14	0.38	0.20	0.59	0.22	0.09	0.02	0.09	0.66	0.24
421	0.15	0.39	0.20	0.59	0.22	0.09	0.02	0.09	0.65	0.24
433	0.14	0.39	0.20	0.59	0.22	0.09	0.02	0.08	0.65	0.24
445	0.14	0.39	0.21	0.59	0.22	0.09	0.02	0.08	0.65	0.24
Mean	0.16	0.39	0.20	0.64	0.23	0.10	0.03	0.11	0.56	0.26
Std	0.02	0.01	0.01	0.04	0.00	0.01	0.00	0.02	0.09	0.01

Note: The mean and standard deviation reported in the columns are overtaken over i. N^* is the number of $\hat{\rho}_i$ that are statistically different from zero at the two-tailed 5 percent level.

References

Anderson, Theodore W., and H. Rubin. 1956. "Statistical Inference in Factor Analysis." In *Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability,* Vol. V, ed. J. Neyman, 114–50. Berkeley: University of California Press.

Bai, Jushan. 2003. "Inferential Theory for Factor Models of Large Dimensions." *Econometrica* 71 (1): 135–72.

Bai, Jushan, and Serena Ng. 2002. "Determining the Number of Factors in Approximate Factor Models." *Econometrica* 70 (1): 191–221.

Boivin, Jean, and Serena Ng. 2004. "Are More Data Always Better for Factor Analysis?" Forthcoming in *Journal of Econometrics.*

Brillinger, David R. 1981. *Time Series: Data Analysis and Theory.* San Francisco: Wiley.

Chamberlain, Gary, and Michael Rothschild. 1983. "Arbitrage, Factor Structure and Mean-Variance Analysis on Large Asset Markets." *Econometrica* 51:1305–24.

Connor, Gregory, and Robert A. Korajzcyk. 1986. "Performance Measurement with the Arbitrage Pricing Theory: A New Framework for Analysis." *Journal of Financial Economics* 15 (3): 373–94.

Forni, Mario, Marc Hallin, Marco Lippi, and Lucrezia Reichlin. 2000. "The Generalized Dynamic Factor Model: Identification and Estimation." *Review of Economics and Statistics* 82 (4): 540–54.

———. 2001. "Coincident and Leading Indicators for the Euro Area." *Economic Journal* 111 (471): C82–5.

———. 2004. "The Generalized Factor Model: Consistency and Rates." *Journal of Econometrics* 119 (2): 231–55.

———. 2005. "The Generalized Dynamic Factor Model: One-Sided Estimation and Forecasting." *Journal of the American Statistical Association* 100 (471): 830–40.

Forni, Mario, and Lucrezia Reichlin. 2001. "Federal Policies and Local Economies: Europe and the US." *European Economic Review* 45 (1): 109–34.

Grenouilleau, Daniel. 2004. "A Sorted Leading Indicators Dynamic (SLID) Factor Model for Short-Run Euro-Area DGP Forecasting." European Commission Economic Papers 219.

Inoue, Atsushi, and Lutz Kilian. 2003. "Bagging Time Series Models." Unpublished Manuscript, North Carolina University.

Kapetanios, George, and Massimiliano Marcellino. 2002. "A Comparison of Estimation Methods for Dynamic Factor Models of Large Dimensions." Draft, Bocconi University.

Kitchen, John, and Ralph M. Monaco. 2003. "Real-Time Forecasting in Practice: The U.S. Treasury Staff's Real-Time GDP Forecast System." *Business Economics* (October): 10–19.

Marcellino, Massimiliano, James Stock, and Mark Watson. 2004. "A Comparison of Direct and Iterated AR Methods for Forecasting Macroeconomic Time Series h-steps Ahead." Mimeo, Princeton University.

Stock, James H., and Mark W. Watson. 1999. "Forecasting Inflation." *Journal of Monetary Economics* 44 (2): 293–335.

———. 2002a. "Forecasting Using Principal Components from a Large Number of Predictors." *Journal of the American Statistical Association* 97 (460): 1167–79.

———. 2002b. "Macroeconomic Forecasting Using Diffusion Indexes." *Journal of Business and Economic Statistics* 20 (2): 147–62.

———. 2004a. "An Empirical Comparison of Methods for Forecasting Using Many Predictors." Mimeo, Princeton University.

———. 2004b. "Forecasting with Many Predictors." In *The Handbook of Economic Forecasting*, ed. G. Elliott. Elsevier Science.

Stock, James H., Mark W. Watson, and Massimiliano Marcellino. 2003. "Macroeconomic Forecasting in the Euro Area: Country Specific versus Area-Wide Information." *European Economic Review* 47 (1): 1–18.