The Generalized Dynamic Factor Model: One-Sided Estimation and Forecasting

Mario FORNI, Marc HALLIN, Marco LIPPI, and Lucrezia REICHLIN

This article proposes a new forecasting method that makes use of information from a large panel of time series. Like earlier methods, our method is based on a dynamic factor model. We argue that our method improves on a standard principal component predictor in that it fully exploits all the dynamic covariance structure of the panel and also weights the variables according to their estimated signal-to-noise ratio. We provide asymptotic results for our optimal forecast estimator and show that in finite samples, our forecast outperforms the standard principal components predictor.

KEY WORDS: Dynamic factor model; Forecasting; Large cross-section; Panel data; Principal components; Time series.

1. INTRODUCTION

Economists and forecasters currently typically have access to information scattered through a huge number of observed aggregated and disaggregated economic time series. Intuition suggests that concentrating on a few series, and thus disregarding potentially relevant information; or performing "naive" aggregation, always produces suboptimal forecasts; the more scattered the information, the more severe this loss of forecasting efficiency. Yet most multivariate forecasting methods in the literature are restricted to vector time series of low dimension and allow for incorporating only a limited number of key variables. Such methods are thus of little help in large panels of time series, where the cross-sectional dimension is often of the same order as, or even larger than, the series lengths.

As a solution to this large-size problem, recent literature has given much attention to *dynamic factor models*, whose main features are (a) an infinite number of cross-sectional units; (b) a decomposition of each observed variable x_{it} in the panel into two mutually orthogonal unobservable components, the *common* component χ_{it} and the *idiosyncratic* component ξ_{it}; (c) a small dynamic dimension of the common components χ_{it}, which are determined by dynamic loading of a finite number, q, of *common factors*; and (d) a "weak correlation structure" (a notion defined more precisely later) of the idiosyncratic components ξ_{it}, which need not be (in contrast with traditional factor models) mutually orthogonal across the panel.

Adopting a parametric approach, Quah and Sargent (1993) estimated by maximum likelihood such a large cross-sectional model, under the restriction of orthogonal idiosyncratic components. Doz, Giannone, and Reichlin (2003) implemented a maximum likelihood estimator (MLE) by forcing orthogonality among the idiosyncratic components, and showed that the impact of the resulting misspecification is negligible as the cross-section size tends to infinity.

Weakly correlated idiosyncratic components are dealt with directly in the nonparametric approach adopted by Forni and Reichlin (1998), Forni, Hallin, Lippi, and Reichlin (2000, 2004), Forni and Lippi (2001), and Stock and Watson (2002a,b). The approach of Stock and Watson (SW hereinafter), based on principal components of *contemporaneous* values of the x's, can also be used for forecasting. The approach of Forni, Hallin, Lippi, and Reichlin (FHLR hereinafter), based on frequency-domain principal components and thus on two-sided filtering of the x's, although more efficient than SW's for estimation of the common components (see FHLR 2000) is not directly suitable for prediction.

In the present article, still in a nonparametric spirit, we combine the advantages of the FHLR and SW methods to propose a new predictor. Following those previous methods, we start with the observation that the forecast of any of the x's can be obtained as the sum of the forecasts of the common and the idiosyncratic components, each based on its own past values. The idiosyncratic component, being mildly cross-correlated, can be predicted by means of traditional univariate or low-dimensional forecasting methods. Thus we concentrate on prediction of the common components χ_{it}. Such prediction is obtained by first estimating the factor space by linear combinations of the x's. As the cross-section size tends to infinity, the idiosyncratic components (being poorly correlated) cancel out, and the factor space is approached. The predictor is then obtained by projecting future values of the χ's on the estimated factor space.

The novelty of this article lies both in the estimation of the factor space and in the way in which projections onto this space are performed. We proceed in two steps. The first step uses the dynamic techniques of FHLR (2000) to obtain estimates of the covariance matrices of common and idiosyncratic components. In the second step, these covariances are used to produce two estimations:

(A) A new estimation of the factor space. We use generalized eigenvectors associated with the estimated covariance matrices of common and idiosyncratic components to obtain (unlike in FHLR 2000) linear combinations—referred to as *generalized principal components*—of *contemporaneous* x's with minimum idiosyncratic–common variance ratio.

(B) A new estimation of the projection of future values of the χ's on the factor space, based on the estimated lagged covariance matrices of the χ's.

Mario Forni is Professor of Political Economy, Department of Political Economy, University of Modena and Reggio Emilia, Modena, Italy (E-mail: *forni@unimo.it*). Marc Hallin is Professor, ISRO, ECARES, and Department of Mathematics, Free University of Brussels, Brussels, Belgium (E-mail: *mhallin@ulb.ac.be*). Marco Lippi is Professor, Department of Economic Science, University of Rome La Sapienza, Rome, Italy (E-mail: *ml@lippi.ws*). Lucrezia Reichlin is Professor, ECARES, Free University of Brussels, Brussels, Belgium (E-mail: *lreichli@ulb.ac.be*). This research was supported by a PAI of the Belgian Federal Government, an ARC contract of the Communauté française de Belgique, the Training and Mobility of Researchers Programme of the European Commission (Contract ERBFMRX-CT98-0213), and a COFIN 2002 grant by Ministero dell'Istruzione, dell'Università e della Ricerca (Italy).

Journal of the American Statistical Association
September 2005, Vol. 100, No. 471, Theory and Methods
DOI 10.1198/016214504000002050

Both our two-step predictor and SW's predictor are consistent, in the sense that, as the cross-section size n and the number of time observations T tend to infinity, both predictors tend in probability to the population-optimal predictor. However, we show that our predictor outperforms SW's predictor in simulations as well as on SW's own dataset. Intuitively, our predictor indeed has a twofold advantage over SW's.

First, whereas SW's estimation of the h-step-ahead projection matrix is based on the lag-h covariance matrix of the x's, our method uses the first-step estimate of the lag-h covariance matrix of the common components. Such an estimate is based on the frequency domain principal components (as in FHLR 2000), which allow efficient aggregation of variables that may be out of phase, so that the information contained in all cross-covariances of the x's, both lagged and contemporaneous, is fully exploited to obtain the h-step-ahead projection matrix.

Second, our generalized principal components method performs better than SW's standard principal components method in approaching the factor space, because it exploits preliminary estimation of the contemporaneous covariance matrices of common and idiosyncratic components. Roughly speaking, our first step enables us to place smaller weights on variables with larger idiosyncratic components, so that the idiosyncratic "error" contained in the linear combination is minimized.

The article is organized as follows. In Section 2 we set up the model and the main assumptions. In Section 3 we provide a detailed presentation of the two-step method and our predictor. In Section 4 we prove consistency. In Section 5 we compare our two-step method with SW's method using simulated panels, and briefly report the results of an in-depth comparison based on the empirical panel used by SW (2002b). We conclude in Section 6. Some mathematical results, needed in Section 4, are given in the Appendix.

2. THE MODEL

The model used in this article is an *approximate* factor model, in that the idiosyncratic component are allowed to be weakly correlated, as was done by Chamberlain (1983) and Chamberlain and Rothschild (1983) and in contrast to the approaches of Sargent and Sims (1977), Geweke (1977), and Quah and Sargent (1993). It is a *dynamic* factor model in that the common factors are loaded through a lag structure, as was done by FHLR (2000), Forni and Lippi (2001), SW (2002a,b), Bai and Ng (2002), and Bai (2003). However, unlike the approaches of FHLR (2000) and Forni and Lippi (2001), the lag structure is assumed to be finite.

Denote by $\mathbf{X} = (x_{it})_{i=1,\dots,n;\, t=1,\dots,T}$ an $n \times T$ rectangular array of observations. Throughout, we assume the following:

A1. $\mathbf{X}$ is a finite realization of a real-valued stochastic process $\{x_{it} \in L_2(\Omega, \mathcal{F}, \mathrm{P}),\ i \in \mathbb{N},\ t \in \mathbb{Z}\}$, where all n-dimensional vector processes $\{\mathbf{x}_t = (x_{1t}\ \cdots\ x_{nt})',\ t \in \mathbb{Z}\}$, $n \in \mathbb{N}$, are stationary, with mean 0 and finite second-order moments, $\mathbf{\Gamma}_k = \mathrm{E}[\mathbf{x}_t \mathbf{x}'_{t-k}]$, $k \in \mathbb{N}$.

The spectral techniques to be used in the sequel also require the following technical assumption:

A2. For all $n \in \mathbb{N}$, the process $\{\mathbf{x}_t, t \in \mathbb{Z}\}$ admits a Wold representation $\mathbf{x}_t = \sum_{k=0}^{\infty} \mathbf{C}_k \mathbf{w}_{t-k}$, where the full-rank n-dimensional innovations $\mathbf{w}_t$ have finite moments of order 4 and the $n \times n$ matrices $\mathbf{C}_k = (C_{ij,k})$ satisfy $\sum_{k=0}^{\infty} |C_{ij,k}| k^{1/2} < \infty$ for all $n, i, j \in \mathbb{N}$.

We refer to assumptions A1 and A2 jointly as assumption A.

To avoid heavy notation, we do not make explicit the dependence on n of the vectors $\mathbf{x}_t$ and $\mathbf{w}_t$, of the matrices $\mathbf{\Gamma}_k$ and $\mathbf{C}_k$, and of many other scalar, vector, and matrix quantities defined later. In the same way, we avoid explicit reference to T for estimated quantities. For example, we denote an estimate of $\mathbf{\Gamma}_k$, which depends on n and T, by $\hat{\mathbf{\Gamma}}_k$.

The basic idea in dynamic factor analysis is that each process x_{it}, $i \in \mathbb{N}$, is the sum of a *common component*, χ_{it}, and an *idiosyncratic component*, ξ_{it}. The common component is driven by a q-dimensional vector of *common factors*, $\mathbf{f}_t = (f_{1t}\ f_{2t}\ \cdots\ f_{qt})'$, which are loaded with possibly different coefficients and lags,

$$\chi_{it} = b_{i1}(L) f_{1t} + b_{i2}(L) f_{2t} + \cdots + b_{iq}(L) f_{qt}.$$

Note that q is independent of n (and small compared with n in empirical applications). In vector notation, defining $\boldsymbol{\chi}_t = (\chi_{1t}\ \cdots\ \chi_{nt})'$ and $\boldsymbol{\xi}_t = (\xi_{1t}\ \cdots\ \xi_{nt})'$, and letting $\mathbf{B}(L)$ denote the $n \times q$ matrix whose (i,j) entry is $b_{ij}(L)$, our model is thus

$$\mathbf{x}_t = \boldsymbol{\chi}_t + \boldsymbol{\xi}_t = \mathbf{B}(L)\mathbf{f}_t + \boldsymbol{\xi}_t, \tag{1}$$

where the factors $\mathbf{f}_t$ follow a vector autoregressive scheme of the form $\mathbf{A}(L)\mathbf{f}_t = \mathbf{u}_t$. Our assumptions on (1) are as follows:

B1. (a) For $i \in \mathbb{N}$ and $j = 1, \dots, q$, the orders of the $b_{ij}(L)$'s have a finite maximum $s \geq 0$; thus $\mathbf{B}(L) = \mathbf{B}_0 + \mathbf{B}_1 L + \cdots + \mathbf{B}_s L^s$ is an $n \times q$ matrix polynomial in the lag operator L, where $\mathbf{B}_s \neq \mathbf{0}$ for n greater than or equal to some $m \geq 1$.
(b) $\mathbf{A}(L) = \mathbf{I} - \mathbf{A}_1 L - \cdots - \mathbf{A}_S L^S$ is a $q \times q$ matrix polynomial, with $\mathbf{A}_S \neq \mathbf{0}$ and $S < s + 1$.
(c) All solutions of $\det[\mathbf{A}(z)] = 0$, $z \in \mathbb{C}$, lie outside the unit ball.

B2. The vector $\{\mathbf{u}_t = (u_{1t}\ \cdots\ u_{qt})',\ t \in \mathbb{Z}\}$ of *common shocks* is a q-dimensional orthonormal white noise process orthogonal to $\{\xi_{it},\ i = 1, \dots, n,\ t \in \mathbb{Z}\}$. (This implies that χ_{it} and ξ_{jt} are orthogonal at any lead and lag for all $i, j \in \mathbb{N}$.)

Of course, the matrices $\mathbf{B}_j$ are nested as n increases. Assumption B1(c) on the characteristic roots of $\mathbf{A}(L)$ guarantees the existence of the inverse operator $[\mathbf{A}(L)]^{-1}$. We return to B1(b) in the next section.

Under assumption A, finite second-order moments exist for all variables involved in the model. Let $\mathbf{\Sigma}^{\chi}(\theta)$ and $\mathbf{\Sigma}^{\xi}(\theta)$, $\theta \in [-\pi, \pi]$, be the $n \times n$ spectral density matrices of $\boldsymbol{\chi}_t$ and $\boldsymbol{\xi}_t$, and let λ_k^{χ} and λ_k^{ξ} denote the corresponding eigenvalues, namely the mappings $\theta \mapsto \lambda_k^{\chi}(\theta)$ and $\theta \mapsto \lambda_k^{\xi}(\theta)$, where $\lambda_k^{\chi}(\theta)$ and $\lambda_k^{\xi}(\theta)$ represent the kth largest eigenvalues of $\mathbf{\Sigma}^{\chi}(\theta)$ and $\mathbf{\Sigma}^{\xi}(\theta)$. On these spectral densities, we make the following assumptions:

C1. (a) $\lambda_q^{\chi}(\theta) \to \infty$ as $n \to \infty$, θ-a.e. in $[-\pi, \pi]$.
(b) $\lambda_k^{\chi}(\theta) > \lambda_{k+1}^{\chi}(\theta)$ θ-a.e. in $[-\pi, \pi]$, $k = 1, \dots, q$.

C2. There exists a positive real Λ such that $\lambda_1^{\xi}(\theta) \leq \Lambda$ for any $\theta \in [-\pi, \pi]$ and any $n \in \mathbb{N}$.

Assumption C1(b) requires that the first $q+1$ eigenvalues be distinct for almost all θ [note that $\lambda_j^{\chi}(\theta)=0$ for $j>q$ and all θ]. This makes the proofs easier without causing a serious loss of generality. Assumptions C1(a) and C2 are needed to guarantee identification of the common and idiosyncratic components (see Forni and Lippi 2001). Note that condition C2, on the asymptotic behavior of $\lambda_k^{\xi}(\theta)$, includes the case in which the idiosyncratic components are mutually orthogonal, with uniformly bounded variances. Mutual orthogonality is a standard, although highly unrealistic assumption in finite-n factor models; condition C2 relaxes this assumption, while giving a precise meaning to the expression "weak correlation" used in Section 1.

Letting $\mathbf{F}_t=(\mathbf{f}_t' \ \mathbf{f}_{t-1}' \ \cdots \ \mathbf{f}_{t-s}')'$ and $\mathbf{C}=(\mathbf{B}_0 \ \mathbf{B}_1 \ \cdots \ \mathbf{B}_s)$, model (1) can also be written as

$$\mathbf{x}_t=\mathbf{C}\mathbf{F}_t+\boldsymbol{\xi}_t, \tag{2}$$

in which $r=q(s+1)$ common factors are loaded only contemporaneously. Equation (2) looks like a static factor model. However, the dynamic nature of (1) implies that $\mathbf{F}_t$ has a special structure; indeed, the spectral density matrix of $\mathbf{F}_t$ has rank q, which is smaller than r if $s>0$. In the sequel we call the factors of the static representation (2) (i.e., the r entries of $\mathbf{F}_t$) "static factors" and the q entries of $\mathbf{f}_t$ "dynamic factors."

Finally, let $\boldsymbol{\Gamma}_k^{\chi}$ and $\boldsymbol{\Gamma}_k^{\xi}$ be the lag-k covariance matrices of the vectors $\boldsymbol{\chi}_t$ and $\boldsymbol{\xi}_t$, and let μ_j^{χ} and μ_j^{ξ} be the jth eigenvalues of $\boldsymbol{\Gamma}_0^{\chi}$ and $\boldsymbol{\Gamma}_0^{\xi}$. We assume that the following holds:

D. $\mu_r^{\chi}\to\infty$ as $n\to\infty$.

Assumption D rules out the case in which some of the elements in $\mathbf{F}_t$ are loaded only by a finite number of the x's. Note that C1(a) does not imply D; for example, if $\chi_{1t}=u_{t-1}$ and $\chi_{it}=u_t$ for $i\geq 2$, then C1(a) clearly holds with $q=1$, but D does not, because $\mu_1^{\chi}\to\infty$, whereas μ_2^{χ} is bounded as $n\to\infty$. Further technical assumptions are introduced in Section 4.

3. A TWO–STEP FORECASTING METHOD

As observed in Section 1, in this article we concentrate on forecasting the common components $\chi_{i,T+h}$. As a byproduct, we also provide an estimator for in-sample values of the χ's. Both problems—forecasting and in-sample estimation—can be reduced to estimating the factors and the covariances between $\chi_{i,T+h}$, or $\boldsymbol{\chi}_t$, and the factors.

Formally, let $\mathcal{G}(\mathbf{F},t)$ denote the linear space spanned by F_{jt}, for $j=1,\ldots,r$. Quite obviously, the common component of χ_{it} coincides with the linear projection

$$\chi_{it}=\operatorname{Proj}(\chi_{it}|\mathcal{G}(\mathbf{F},t)) \tag{3}$$

of χ_{it} on $\mathcal{G}(\mathbf{F},t)$. Moreover, using the inequality $S<s+1$ [see assumption B1(b)], it is easily seen that the best linear predictor of $\chi_{i,T+h}$ based on F_{T-k}, $k\geq 0$, denoted by $\chi_{i,T+h|T}$, is given by the linear projection

$$\chi_{i,T+h|T}=\operatorname{Proj}(\chi_{i,T+h}|\mathcal{G}(\mathbf{F},T)). \tag{4}$$

Note that Assumption B1(b) implies that enlarging the projection space with past values of $\mathbf{F}_t$ does not improve prediction. Note, however, that B1(b) is just a convenience. If S were larger than s, then optimal prediction would require the projection on a space including past values of $\mathbf{F}_t$, implying only minor changes in our statements (Prop. 1 in particular) and proofs.

Steps 1 and 2 of our procedure estimate (3) and (4) by estimating, in reverse order, the factor space and the relevant covariance matrices.

3.1 Step One: Estimating $\boldsymbol{\Gamma}_k^{\chi}$ and $\boldsymbol{\Gamma}_k^{\xi}$

In the method of FHLR (2000), estimation of the common components is based on the *dynamic principal components* method (see Brillinger 1981, chap. 9). Let $\hat{\boldsymbol{\Sigma}}(\theta)=(\hat{\sigma}_{ij}(\theta))$, $\theta\in[-\pi,\pi]$, denote a consistent periodogram-smoothing or lag-window estimator of the $n\times n$ spectral density matrix, $\boldsymbol{\Sigma}(\theta)=(\sigma_{ij}(\theta))$, of $\mathbf{x}_t$. Let $\hat{\lambda}_j(\theta)$ be $\hat{\boldsymbol{\Sigma}}(\theta)$'s jth largest eigenvalue, let $\hat{\mathbf{p}}_j(\theta)=(\hat{p}_{j1}(\theta) \ \cdots \ \hat{p}_{jn}(\theta))$ be the corresponding row eigenvector, and let $\hat{\mathbf{P}}(\theta)=(\hat{\mathbf{p}}_1'(\theta) \ \hat{\mathbf{p}}_2'(\theta) \ \cdots \ \hat{\mathbf{p}}_q'(\theta))'$ be a $q\times n$ matrix. Defining

$$\underline{\hat{\mathbf{p}}}_j(L)=\frac{1}{2\pi}\sum_{k=-\infty}^{\infty}\left[\int_{-\pi}^{\pi}\hat{\mathbf{p}}_j(\theta)e^{ik\theta}\,d\theta\right]L^k \tag{5}$$

[the inverse Fourier transform of $\hat{\mathbf{p}}_j(\theta)$], the jth dynamic principal components of $\mathbf{x}_t$ is defined as $\underline{\hat{\mathbf{p}}}_j(L)\mathbf{x}_t$. The first q dynamic principal component is used to obtain a consistent (see FHLR 2000) estimation $\hat{\boldsymbol{\chi}}_t^D=\underline{\hat{\mathbf{P}}}^*(L)[\underline{\hat{\mathbf{P}}}(L)\mathbf{x}_t]$ of the χ's (with the star indicating complex conjugation and transposition and superscript "D" indicating dynamic method).

The trouble with this estimation method is that the filters (5) used in the definition of $\hat{\boldsymbol{\chi}}_t^D$ are two-sided in general. This creates no problem in the central part of the sample, but the performance of $\hat{\boldsymbol{\chi}}_t^D$ as an estimator of $\boldsymbol{\chi}_t$ deteriorates as t approaches T or 1. For the same reason, dynamic principal components cannot be used for prediction.

However, the spectral density of $\hat{\boldsymbol{\chi}}_t^D$ provides estimates of the spectral density matrices of $\boldsymbol{\chi}_t$ and $\boldsymbol{\xi}_t$,

$$\hat{\boldsymbol{\Sigma}}^{\chi}(\theta)=\hat{\lambda}_1(\theta)\hat{\mathbf{p}}_1^*(\theta)\hat{\mathbf{p}}_1(\theta)+\cdots+\hat{\lambda}_q(\theta)\hat{\mathbf{p}}_q^*(\theta)\hat{\mathbf{p}}_q(\theta) \tag{6}$$

and

$$\hat{\boldsymbol{\Sigma}}^{\xi}(\theta)=\hat{\lambda}_{q+1}(\theta)\hat{\mathbf{p}}_{q+1}^*(\theta)\hat{\mathbf{p}}_{q+1}(\theta)+\cdots+\hat{\lambda}_n(\theta)\hat{\mathbf{p}}_n^*(\theta)\hat{\mathbf{p}}_n(\theta) \tag{7}$$

(see FHLR 2000 for details). Therefore, the covariance matrices of $\boldsymbol{\chi}_t$ and $\boldsymbol{\xi}_t$ can be estimated as

$$\begin{aligned}\hat{\boldsymbol{\Gamma}}_k^{\chi}&=\int_{-\pi}^{\pi}e^{ik\theta}\hat{\boldsymbol{\Sigma}}^{\chi}(\theta)\,d\theta \qquad \text{and}\\ \hat{\boldsymbol{\Gamma}}_k^{\xi}&=\int_{-\pi}^{\pi}e^{ik\theta}\hat{\boldsymbol{\Sigma}}^{\xi}(\theta)\,d\theta.\end{aligned} \tag{8}$$

3.2 Step Two: Estimating the Factor Space and the Best Linear Forecast

The general idea underlying our method is that the factor space can be consistently estimated by linear combinations of the x's, as n tends to infinity. Different linear combinations produce different estimators. We argue that the information contained in the covariance matrices estimated in Section 3.1 can be used to determine linear combinations of the x's that are more efficient than standard principal components.

Consider all of the linear combinations $\mathbf{a}\mathbf{x}_t = a_1 x_{1t} + \cdots + a_n x_{nt}$ of the x's that fulfill the constraint $\text{var}(\mathbf{a}\mathbf{x}_t) = 1$. Observe that

$$\mathbf{a}\mathbf{x}_t = \mathbf{a}\boldsymbol{\chi}_t + \mathbf{a}\boldsymbol{\xi}_t = \text{Proj}(\mathbf{a}\mathbf{x}_t | \mathcal{G}(\mathbf{F}, t)) + \mathbf{a}\boldsymbol{\xi}_t,$$

so that under the constraint, the linear combination of the x's that is closest to the factor space is obtained by solving $\min_{\mathbf{a}\in\mathbb{R}^n} \text{var}(\mathbf{a}\boldsymbol{\xi}_t)$, s.t. $\text{var}(\mathbf{a}\mathbf{x}_t) = 1$, which is obviously equivalent to $\max_{\mathbf{a}\in\mathbb{R}^n} \text{var}(\mathbf{a}\boldsymbol{\chi}_t)$, s.t. $\text{var}(\mathbf{a}\boldsymbol{\xi}_t) = 1$, or, using the variances and covariances estimated in Section 3.1,

$$\max_{\mathbf{a}\in\mathbb{R}^n} \mathbf{a}\hat{\boldsymbol{\Gamma}}_0^{\chi}\mathbf{a}' \qquad \text{s.t.} \qquad \mathbf{a}\hat{\boldsymbol{\Gamma}}_0^{\xi}\mathbf{a}' = 1.$$

Extending this simple argument, we want to find r [the dimension of $\mathcal{G}(\mathbf{F}, t)$] independent linear combinations $\hat{W}_{jt} = \hat{\mathbf{Z}}_j\mathbf{x}_t$, where the weights $\hat{\mathbf{Z}}_j$ are defined recursively as

$$\hat{\mathbf{Z}}_j = \text{Arg}\max_{\mathbf{a}\in\mathbb{R}^n} \mathbf{a}\hat{\boldsymbol{\Gamma}}_0^{\chi}\mathbf{a}' \qquad \text{s.t.} \qquad \mathbf{a}\hat{\boldsymbol{\Gamma}}_0^{\xi}\mathbf{a}' = 1 \qquad \text{and}$$

$$\mathbf{a}\hat{\boldsymbol{\Gamma}}_0^{\xi}\hat{\mathbf{Z}}_m' = 0, \qquad 1 \le m \le j-1, \quad (9)$$

for $j = 1, \ldots, r$. (For $j = 1$, only the first constraint applies.) The solutions $\hat{\mathbf{Z}}_j$ of this problem are the *generalized eigenvectors* associated with the *generalized eigenvalues*, $\hat{v}_j$, of the couple of matrices ($\hat{\boldsymbol{\Gamma}}_0^{\chi}$ and $\hat{\boldsymbol{\Gamma}}_0^{\xi}$), that is, the solutions of

$$\hat{\mathbf{Z}}_j\hat{\boldsymbol{\Gamma}}_0^{\chi} = \hat{v}_j\hat{\mathbf{Z}}_j\hat{\boldsymbol{\Gamma}}_0^{\xi}, \qquad j = 1, 2, \ldots, n, \quad (10)$$

with the normalization constraints $\hat{\mathbf{Z}}_j\hat{\boldsymbol{\Gamma}}_0^{\xi}\hat{\mathbf{Z}}_j' = 1$ and $\hat{\mathbf{Z}}_i\hat{\boldsymbol{\Gamma}}_0^{\xi}\hat{\mathbf{Z}}_j' = 0$ for $i \neq j$ (see thm. A.2.4 of Anderson 1984, p. 590). The linear combinations $\hat{W}_{jt}$ are the *generalized principal components* of $\mathbf{x}_t$ relative to the couple $(\hat{\boldsymbol{\Gamma}}_0^{\chi}, \hat{\boldsymbol{\Gamma}}_0^{\xi})$.

Defining $\hat{\mathbf{Z}} = (\hat{\mathbf{Z}}_1' \ \cdots \ \hat{\mathbf{Z}}_r')'$, the space $\mathcal{G}(\mathbf{F}, t)$ is estimated by the first r generalized principal components of the x's, that is, by the r components of

$$\hat{\mathbf{Z}}\mathbf{x}_t = (\hat{\mathbf{Z}}_1\mathbf{x}_t \ \hat{\mathbf{Z}}_2\mathbf{x}_t \ \cdots \ \hat{\mathbf{Z}}_r\mathbf{x}_t)'. \quad (11)$$

Observing that the covariance between $\boldsymbol{\chi}_{T+h}$ and $\hat{\mathbf{Z}}\mathbf{x}_T$ (or $\boldsymbol{\chi}_t$ and $\hat{\mathbf{Z}}\mathbf{x}_t$) equals the covariance between $\boldsymbol{\chi}_{T+h}$ and $\hat{\mathbf{Z}}\boldsymbol{\chi}_T$ (or $\boldsymbol{\chi}_t$ and $\hat{\mathbf{Z}}\boldsymbol{\chi}_t$), the estimators of the projections (3) and (4), in vector form, are easily obtained as

$$\hat{\boldsymbol{\chi}}_t = \left[\hat{\boldsymbol{\Gamma}}_0^{\chi}\hat{\mathbf{Z}}'(\hat{\mathbf{Z}}\hat{\boldsymbol{\Gamma}}_0\hat{\mathbf{Z}}')^{-1}\right][\hat{\mathbf{Z}}\mathbf{x}_t] \quad (12)$$

and

$$\hat{\boldsymbol{\chi}}_{T+h|T} = \left[\hat{\boldsymbol{\Gamma}}_h^{\chi}\hat{\mathbf{Z}}'(\hat{\mathbf{Z}}\hat{\boldsymbol{\Gamma}}_0\hat{\mathbf{Z}}')^{-1}\right][\hat{\mathbf{Z}}\mathbf{x}_T]. \quad (13)$$

These projections are the two-step estimators (predictors) that we are proposing.

An alternative method for estimating the best linear predictor of the χ's might be a *factorization* of $\hat{\boldsymbol{\Sigma}}^{\chi}(\theta)$ along the lines of the approach of Wiener and Masani (1957, 1958) or Rozanov (1967). However, adapting this factorization approach to the present context is far from trivial. Indeed, our spectral density matrix is singular (its rank is q, irrespective of n), which raises a serious problem with the choice of a nonsingular submatrix. Moreover, the matrix to be inverted under our approach is only $r \times r$ [see (12) and (13)], with r independent of n, so that the motivation for factorization as opposed to the projection method (Wiener and Masani 1958, pp. 100–102) loses much of its strength.

To get a sense of how generalized principal components are constructed, consider the very simple example under which $\chi_{it} = u_t$ for any i, with orthogonal idiosyncratic terms. In this case, assuming that the relevant covariance matrices are estimated without error, the entries of $\hat{\mathbf{Z}}_1$ are proportional to $1/\sigma_i^{\xi}$, that is, inversely proportional to the "size" of the idiosyncratic components. More generally, it is easily seen that the generalized principal components of $\mathbf{x}_t$ are equal to the standard principal components of the transformed vector $\tilde{\mathbf{x}}_t = (\hat{\boldsymbol{\Gamma}}_0^{\xi})^{-1/2}\mathbf{x}_t$. Such principal components are the MLEs of $\mathbf{F}_t$ under the assumption of known diagonal idiosyncratic covariance matrix (see Lawley and Maxwell 1971, chap. 4).

3.3 Comparison With the Stock and Watson Method

Consider the k-lag *sample* cross-covariance matrix $\hat{\boldsymbol{\Gamma}}_k = (T-k)^{-1}\sum_{t=k+1}^{T}\mathbf{x}_t\mathbf{x}_{t-k}'$ of $\mathbf{x}_t$. Let $\hat{m}_j$ be the jth largest eigenvalue of $\hat{\boldsymbol{\Gamma}}_0$, with row eigenvector $\hat{\mathbf{S}}_j$. Moreover, let $\hat{\mathbf{M}}$ be the $r \times r$ diagonal matrix with diagonal elements $\hat{m}_1, \ \hat{m}_2, \ldots, \hat{m}_r$, and put $\hat{\mathbf{S}} = (\hat{\mathbf{S}}_1' \ \cdots \ \hat{\mathbf{S}}_r')'$, an $r \times n$ matrix. The SW estimation of the factor space and the projections are given by

$$\hat{\mathbf{S}}\mathbf{x}_t = (\hat{\mathbf{S}}_1\mathbf{x}_t \ \hat{\mathbf{S}}_2\mathbf{x}_t \ \cdots \ \hat{\mathbf{S}}_r\mathbf{x}_t)', \quad (14)$$

$$\hat{\boldsymbol{\chi}}_t^{SW} = \left[\hat{\boldsymbol{\Gamma}}_0\hat{\mathbf{S}}'(\hat{\mathbf{S}}\hat{\boldsymbol{\Gamma}}_0\hat{\mathbf{S}}')^{-1}\right][\hat{\mathbf{S}}\mathbf{x}_t] = \hat{\mathbf{S}}'[\hat{\mathbf{S}}\mathbf{x}_t], \quad (15)$$

and

$$\hat{\boldsymbol{\chi}}_{T+h|T}^{SW} = \left[\hat{\boldsymbol{\Gamma}}_h\hat{\mathbf{S}}'(\hat{\mathbf{S}}\hat{\boldsymbol{\Gamma}}_0\hat{\mathbf{S}}')^{-1}\right][\hat{\mathbf{S}}\mathbf{x}_T] = \hat{\boldsymbol{\Gamma}}_h\hat{\mathbf{S}}'\hat{\mathbf{M}}^{-1}[\hat{\mathbf{S}}\mathbf{x}_T]. \quad (16)$$

We compare this with our two-step method:

(I) Because no preliminary estimation of the matrices $\boldsymbol{\Gamma}_h^{\chi}$ is available, estimated values of $\boldsymbol{\chi}_t$ and $\boldsymbol{\chi}_{T+h}$ in SW are obtained by projecting $\mathbf{x}_t$ and $\mathbf{x}_{T+h}$ on the estimated factors. Hence, SW use $\hat{\boldsymbol{\Gamma}}_h$ in the projection matrix, whereas we use $\hat{\boldsymbol{\Gamma}}_h^{\chi}$, which is obtained via the dynamic method of FHLR (2000) and thus conveys information contained in the whole covariance sequence $\{\hat{\boldsymbol{\Gamma}}_s, \ s \in \mathbb{Z}\}$.

(II) Moreover, with the two-step method, as illustrated in Section 3.2, the covariance matrices estimated in step 1 are used to obtain efficient weights in step 2, so that the factor space is approximated by $\hat{\mathbf{Z}}\mathbf{x}_t$ instead of SW's standard principal components $\hat{\mathbf{S}}\mathbf{x}_t$.

Although the arguments in (I) and (II) provide a strong heuristic support for the claim that forecasts based on the two-step method outperform those based on SW's method, a formal derivation of optimality properties or relative efficiency values is extremely difficult in such a general context and is not pursued in this article. However, we obtain an important insight into the finite-sample relative performances of our method and SW's method in Section 5 by comparing forecast results on simulated and empirical panels.

4. CONSISTENCY

In this section we prove convergence in probability of $\hat{\chi}_{it}$ to χ_{it} and of $\hat{\chi}_{i,T+h|T}$ to the best linear forecast of $\chi_{i,T+h}$, for each i, as T and n tend to infinity. As in Section 3, let $\hat{\mathbf{\Sigma}}(\theta) = (\hat{\sigma}_{ij}(\theta))$ denote any consistent estimator of the $n \times n$ spectral density matrix $\mathbf{\Sigma}(\theta) = (\sigma_{ij}(\theta))$. Under assumption A2, for a given n and any $\epsilon > 0$,

$$\lim_{T\to\infty} \mathrm{P}\Big[\max_{1\le i,j\le n} \sup_{\theta\in[-\pi,\pi]} |\hat{\sigma}_{ij}(\theta) - \sigma_{ij}(\theta)| > \epsilon\Big] = 0. \tag{17}$$

This is an easy consequence of remark 1 to theorem 10.4.1 of Brockwell and Davis (1991, p. 353). (Note that remark 1 applies, mutatis mutandis, to their thm. 11.7.2, p. 447, which extends thm. 10.4.1 to the multidimensional case.) Define

$$\check{\mathbf{\Sigma}}^{\chi}(\theta) = \lambda_1(\theta)\mathbf{p}_1^*(\theta)\mathbf{p}_1(\theta) + \cdots + \lambda_q(\theta)\mathbf{p}_q^*(\theta)\mathbf{p}_q(\theta)$$

and

$$\check{\mathbf{\Sigma}}^{\xi}(\theta) = \lambda_{q+1}(\theta)\mathbf{p}_{q+1}^*(\theta)\mathbf{p}_{q+1}(\theta) + \cdots + \lambda_n(\theta)\mathbf{p}_n^*(\theta)\mathbf{p}_n(\theta),$$

$\lambda_j(\theta)$ and $\mathbf{p}_j(\theta)$ are, as in Section 2, eigenvalues and eigenvectors of $\mathbf{\Sigma}(\theta)$. Note that $\check{\mathbf{\Sigma}}^{\chi}(\theta)$ and $\check{\mathbf{\Sigma}}^{\xi}(\theta)$ are *not* the population spectral density matrices of $\boldsymbol{\chi}_t$ and $\boldsymbol{\xi}_t$. They are, so to speak, estimates of such matrices for infinite T but finite n.

Under Assumption C1(b), continuity of the eigenvalues and first q eigenvectors as functions of the entries of $\hat{\mathbf{\Sigma}}(\theta)$ [the somewhat inaccurate expression "continuity of the eigenvectors" stands for continuity, for all $j = 1, \ldots, q$, of $\hat{\mathbf{p}}_j^*(\theta)\hat{\mathbf{p}}_j(\theta)$] implies that (17) applies to the entries of $\hat{\mathbf{\Sigma}}^{\chi}$ and $\check{\mathbf{\Sigma}}^{\chi}$. More precisely, for any given n and any $\epsilon > 0$,

$$\lim_{T\to\infty} \mathrm{P}\Big[\max_{1\le i,j\le n} \sup_{\theta\in[-\pi,\pi]} |\hat{\sigma}_{ij}^{\chi}(\theta) - \check{\sigma}_{ij}^{\chi}(\theta)| > \epsilon\Big] = 0. \tag{18}$$

The same property holds for $\hat{\mathbf{\Sigma}}^{\xi}$ and $\check{\mathbf{\Sigma}}^{\xi}$, so that $\hat{\mathbf{\Sigma}}^{\chi}$ and $\hat{\mathbf{\Sigma}}^{\xi}$, for fixed n, are consistent estimators of $\check{\mathbf{\Sigma}}^{\chi}$ and $\check{\mathbf{\Sigma}}^{\xi}$. Moreover, (18) implies that $\hat{\mathbf{\Gamma}}_k^{\chi}$ and $\hat{\mathbf{\Gamma}}_k^{\xi}$, as defined in (8), are consistent estimators of

$$\begin{aligned} \check{\mathbf{\Gamma}}_k^{\chi} &= \int_{-\pi}^{\pi} e^{ik\theta}\check{\mathbf{\Sigma}}^{\chi}(\theta)\,d\theta \qquad \text{and} \\ \check{\mathbf{\Gamma}}_k^{\xi} &= \int_{-\pi}^{\pi} e^{ik\theta}\check{\mathbf{\Sigma}}^{\xi}(\theta)\,d\theta. \end{aligned} \tag{19}$$

Let $\check{\mu}_k^{\chi}$ and $\check{\mu}_k^{\xi}$ denote the eigenvalues of $\check{\mathbf{\Gamma}}_0^{\chi}$ and $\check{\mathbf{\Gamma}}_0^{\xi}$. Lemma A.2 (see the App.) proves that, under assumptions A, B, C, and D, $\lim_{n\to\infty} \check{\mu}_r^{\chi} = \infty$, whereas $\check{\mu}_1^{\xi}$ is bounded. We need the following technical assumption:

E1. (a) $\check{\mu}_k^{\chi} > \check{\mu}_{k+1}^{\chi}, \qquad k = 1, \ldots, r.$
(b) $\check{\mu}_n^{\xi}$ is bounded away from 0 as $n \to \infty$.

Assumption E(a), like assumption C1(b), does not imply any significant loss of generality.

Finally, we introduce some further notation:

a. Set $\hat{\mathbf{K}}_h = \hat{\mathbf{\Gamma}}_h^{\chi}\hat{\mathbf{Z}}(\hat{\mathbf{Z}}\hat{\mathbf{\Gamma}}_0\hat{\mathbf{Z}}')^{-1}\hat{\mathbf{Z}}$, so that $\hat{\boldsymbol{\chi}}_{T+h|T} = \hat{\mathbf{K}}_h\mathbf{x}_T$.

b. Let $\hat{w}_{jt}$ denote the standardized version of $\hat{W}_{jt}$. Because

$$\hat{\mathbf{Z}}_j\hat{\mathbf{\Gamma}}_0\hat{\mathbf{Z}}_j' = \hat{\mathbf{Z}}_j\hat{\mathbf{\Gamma}}_0^{\chi}\hat{\mathbf{Z}}_j' + \hat{\mathbf{Z}}_j\hat{\mathbf{\Gamma}}_0^{\xi}\hat{\mathbf{Z}}_j' = 1 + \hat{\nu}_j,$$

we have that $\hat{w}_{jt} = \hat{\mathbf{z}}_j\mathbf{x}_t$, where $\hat{\mathbf{z}}_j = \hat{\mathbf{Z}}_j/\sqrt{1+\hat{\nu}_j}$. Note that because $\hat{\mathbf{Z}}_j\hat{\mathbf{\Gamma}}_0\hat{\mathbf{Z}}_k' = 0$ for $j \neq k$ [using the constraints of (9)], the linear combinations $\hat{w}_{jt}$, for $k = 1, 2, \ldots, r$, form an orthonormal system spanning a space of the same dimension as $\mathcal{G}(\mathbf{F}, t)$.

c. Let $\check{\mathbf{Z}}_j$, $\check{\nu}_j$, $\check{\mathbf{K}}^h$, $\check{\chi}_{i,T+h|T}$, and so on denote the objects playing the same roles as $\hat{\mathbf{Z}}_j$, $\hat{\nu}_j$, $\hat{\mathbf{K}}_h$, $\hat{\chi}_{i,T+h|T}$, and so on, but with respect to $\check{\mathbf{\Gamma}}_0^{\chi}$ and $\check{\mathbf{\Gamma}}_0^{\xi}$.

We now can state the main consistency result of this article.

Proposition 1. Suppose that assumptions A, B, C, D, and E hold for model (1). Then, for any given $i, \epsilon > 0$, and $\eta > 0$, there exist $N_0 = N_0(\epsilon, \eta)$, with $N_0 > i$, and $T_0 = T_0(n, \epsilon, \eta)$ such that, for all $n \ge N_0$ and all $T \ge T_0$,

$$\mathrm{P}\big[\big|\hat{\chi}_{i,T+h|T} - \chi_{i,T+h|T}\big| > \epsilon\big] \le \eta. \tag{20}$$

The proof of Proposition 1 relies on the following two lemmas.

Lemma 1. Let $\mathbf{a}_n = (a_{n1}\ a_{n2}\ \cdots\ a_{nn})$, $n \in \mathbb{N}$, denote a triangular array of real numbers such that $\lim_{n\to\infty}\sum_{i=1}^n a_{ni}^2 = 0$. Then, under the assumptions of Proposition 1,

$$\mathbf{a}_n\boldsymbol{\xi}_t = (a_{n1}\ a_{n2}\ \cdots\ a_{nn})(\xi_{1t}\ \xi_{2t}\ \cdots\ \xi_{nt})' \longrightarrow 0$$

in quadratic mean as $n \to \infty$. It follows that $\mathbf{a}_n\mathbf{x}_t$ converges to $\mathcal{G}(\mathbf{F}, t)$ in quadratic mean.

For a proof, see, for example, lemma 3 of FHLR (2000).

Lemma 2. Let $\mathcal{K}$ denote a subspace of a Hilbert space $\mathcal{H}$ of centered, square-integrable random variables, with covariance scalar product. Assume that $\mathcal{K}$ is generated by the independent k-tuple $(v_1, v_2, \ldots, v_k)$, $v_j \in \mathcal{H}$. Let $\{(v_{n1}, v_{n2}, \ldots, v_{nk}), n \in \mathbb{N}\}$ be a sequence of orthonormal k-tuples of $\mathcal{H}$ such that $v_{nj} - \mathrm{Proj}(v_{nj}|\mathcal{K})$ converges to 0 in quadratic mean as $n \to \infty$. Then the projection of $v \in \mathcal{H}$ onto the space $\mathcal{K}_n$ spanned by $(v_{n1}, \ldots, v_{nk})$ converges in quadratic mean, as $n \to \infty$, to the projection of v onto $\mathcal{K}$.

The proof of this lemma is given in the Appendix.

Proof of Proposition 1. Lemmas A.1 and A.2 in the Appendix imply that $\check{\nu}_r$ tends to infinity as $n \to \infty$. As a consequence, each of the r sequences $\{\check{\mathbf{Z}}_j/\sqrt{1+\check{\nu}_j}, n \in \mathbb{N}\}$, $j = 1, \ldots, r$, is a triangular array fulfilling the assumption of Lemma 1. Indeed, $\check{\mathbf{Z}}_j$ is bounded in modulus, because $1 = \check{\mathbf{Z}}_j\check{\mathbf{\Gamma}}_0^{\xi}\check{\mathbf{Z}}_j' \ge \check{\mu}_n^{\xi}\check{\mathbf{Z}}_j\check{\mathbf{Z}}_j'$, where, in view of assumption E(b), $\check{\mu}_n^{\xi}$ is bounded away from 0. Lemma 1 implies that $\check{w}_{jt} = \check{\mathbf{Z}}_j\mathbf{x}_t/\sqrt{1+\check{\nu}_j}$ converges in quadratic mean to the space $\mathcal{G}(\mathbf{F}, t)$ as $n \to \infty$, for $j = 1, \ldots, r$. Then, by Lemma 2, $\check{\chi}_{i,T+h|T}$ converges to $\chi_{i,T+h|T}$ in quadratic mean and therefore in probability. Thus, given $\epsilon > 0$ and $\eta > 0$, there exists $N_1(\epsilon, \eta)$, such that for $n > N_1$,

$$\mathrm{P}\big[\big|\chi_{i,T+h|T} - \check{\chi}_{i,T+h|T}\big| > \epsilon\big] < \eta. \tag{21}$$

Convergence in probability of $\hat{\mathbf{\Gamma}}_0^{\chi}$ and $\hat{\mathbf{\Gamma}}_n^{\xi}$ to $\check{\mathbf{\Gamma}}_0^{\chi}$ and $\check{\mathbf{\Gamma}}_0^{\xi}$, as $T \to \infty$, and assumption E(a), imply that $\hat{\mathbf{K}}_h$ converges in probability to $\check{\mathbf{K}}_h$ for $T \to \infty$. This implies that, given n, $\epsilon > 0$ and $\eta > 0$, there exists $T_1(n, \epsilon, \eta)$ such that, for $T > T_1$,

$$\mathrm{P}\left[\sum_{j=1}^{n} |\hat{K}_{h,ij} - \check{K}_{h,ij}| > \epsilon\right] < \eta.$$

Moreover, given n and $\eta > 0$, let $M(n, \eta)$ be a positive real such that $\mathrm{P}[\max_{j=1,n} |x_{jt}| \geq M(n, \eta)] < \eta$. Then, given n, $\epsilon > 0$ and $\eta > 0$, there exists $T_2(n, \epsilon, \eta)$ such that, for $T > T_2$,

$$\mathrm{P}\left[\left|\check{\chi}_{i,T+h|T} - \hat{\chi}_{i,T+h|T}\right| > \epsilon\right] = \mathrm{P}\left[\left|\sum_{j=1}^{n} (\hat{K}_{h,ij} - \check{K}_{h,ij}) x_{jt}\right| > \epsilon\right] < \eta. \quad (22)$$

To see this, note that

$$\begin{aligned}
&\mathrm{P}\left[\left|\sum_{j=1}^{n} (\hat{K}_{h,ij} - \check{K}_{h,ij}) x_{jt}\right| > \epsilon\right] \\
&\quad \leq \mathrm{P}\left[\sum_{j=1}^{n} |(\hat{K}_{h,ij} - \check{K}_{h,ij})| M\left(n, \frac{\eta}{2}\right) > \epsilon \text{ and } \max_{j=1,n} |x_{jt}| < M(n, \eta/2)\right] \\
&\quad + \mathrm{P}\left[\max_{j=1,n} |x_{jt}| \geq M\left(n, \frac{\eta}{2}\right)\right],
\end{aligned}$$

so that (22) is obtained defining $T_2(n, \epsilon, \eta) = T_1(n, \epsilon/M(n, \eta/2), \eta/2)$. Finally,

$$\begin{aligned}
&\mathrm{P}\left[\left|\chi_{i,T+h|T} - \hat{\chi}_{i,T+h|T}\right| > \epsilon\right] \\
&\quad \leq \mathrm{P}\left[\left|\chi_{i,T+h|T} - \check{\chi}_{i,T+h|T}\right| > \epsilon/2\right] \\
&\quad + \mathrm{P}\left[\left|\check{\chi}_{i,T+h|T} - \hat{\chi}_{i,T+h|T}\right| > \epsilon/2\right].
\end{aligned}$$

Defining $N_0(\epsilon, \eta) = N_1(\epsilon/2, \eta/2)$ and $T_0(n, \epsilon, \eta) = T_2(n, \epsilon/2, \eta/2)$, the conclusion follows from (21) and (22).

As can be easily checked, the proof of Proposition 1 can be adapted with no difficulty to prove consistency of our two-step in-sample estimator (13). Moreover, Proposition 1 holds if the matrices $\hat{\mathbf{\Gamma}}_0^{\xi}$ and $\check{\mathbf{\Gamma}}_0^{\xi}$ are replaced by any other sequence of couples of positive-definite $n \times n$ symmetric matrices $\hat{\mathbf{D}}$, depending on n and T, and $\check{\mathbf{D}}$, depending on n, provided that $\hat{\mathbf{D}} - \check{\mathbf{D}}$ converges to 0 in probability for any n, as $T \to \infty$, and that all eigenvalues of $\check{\mathbf{D}}$ are bounded and bounded away from 0 as $n \to \infty$ (indeed, Lemmas A.1 and A.2 hold).

5. FINITE–SAMPLE PERFORMANCES: SIMULATED AND EMPIRICAL PANELS

In this section we apply both our two-step and SW's estimator to simulated panels that differ by the degree of heterogeneity of the idiosyncratic variances and the dynamic structure of the common components. We also briefly report some of the results of two empirical exercises comparing the predictive performances of the two-step and SW predictors (the first being based on SW's own dataset).

5.1 Simulation Results

The first model, M1, has one autoregressive factor, loaded only contemporaneously, and spherical idiosyncratic components. This is a case where in principle, SW's method should perform comparatively well. Models M2, M3, and M4 have a richer and more heterogeneous dynamic structure—a feature that should favor the dynamic method. Model M2 has MA(3) loading filters, two serially uncorrelated factors, and a diagonal idiosyncratic variance-covariance matrix. Models M3 and M4 have one autoregressive factor, and the common components of different groups are shifted in time. The two models differ by the behaviors of their idiosyncratic components; in Model M3, they have different variances, whereas in Model M4, the variances are the same. The comparison between the last two models should improve understanding the role of heterogeneity of the size of the idiosyncratic components. Finally, in Model M3, the idiosyncratic components are not mutually orthogonal, but condition C2 is still satisfied.

Model M1. Under this model, the observations are generated by

$$x_{it}^* = \lambda_i f_t + \alpha c_i \epsilon_{it} \quad \text{with } (1 - .5L) f_t = u_t, \qquad \text{(M1)}$$

where the shocks u_t and ϵ_{it}, $t = 1, \ldots, T$, $i = 1, \ldots, n$, and the coefficients λ_i, $i = 1, \ldots, n$, are mutually independent standard normal variables, whereas the coefficients c_i are mutually independent, independent of the latter variables, and uniformly distributed on the interval [.1, 1.1] to avoid nearly zero idiosyncratic components. The constant α is set so as to guarantee that the average idiosyncratic–common variance ratio is equal to 1. (The same holds for all of the other models.) Here $q = 1$ and $s = 1$.

Model M2. Observations are generated by

$$x_{it}^* = \sum_{k=0}^{3} a_{ik} u_{1,t-k} + \sum_{k=0}^{3} b_{ik} u_{2,t-k} + \alpha c_i \epsilon_{it}. \qquad \text{(M2)}$$

Again, a_{ik} and b_{ik}, $k = 0, 1, 2, 3$, $i = 1, \ldots, n$, and the shocks u_{1t}, u_{2t} and ϵ_{it}, $t = 1, \ldots, T$, $i = 1, \ldots, n$, are standard normal variables, whereas the c_i's are uniformly distributed over the interval [.1 1.1], as for (M1). Here $q = 2$ and $s = 3$.

Model M3. The observations are generated by

$$x_{it}^* = \sum_{k=l_i}^{l_i+2} \lambda_{k-l_i,i} f_{t-k} + \xi_{it}^* \quad \text{with}$$

$$(1 - aL) f_t = u_t \text{ and } \xi_{it}^* = \alpha c_i (\epsilon_{it} + \epsilon_{i+1,t}), \qquad \text{(M3)}$$

where $l_i = 0$ for $1 \leq i \leq m$, $l_i = 1$ for $m + 1 \leq i \leq 2m$, and $l_i = 2$ for $2m + 1 \leq i \leq n$. For the three types ($l_i = 0, 1, 2$) to be equally present in the panel, we took $m = [n/3]$. (As usual, we let $[z]$ denote the largest integer less than or equal to z.) Here $q = 1$ and $s = 5$. Note that ξ_{it}^* is positively correlated with $\xi_{i+1,t}^*$, but is orthogonal to $\xi_{i+k,t}^*$ at any lead and lag for $k > 1$.

Model M4. The observations are generated as in M3, but idiosyncratic components are no longer cross-sectionally correlated ($\xi_{it}^* = \alpha c_i \epsilon_{it}$), and the coefficients c_i are such that $\mathrm{var}(\lambda_i f_t)/\mathrm{var}(x_{it}^*) = 0.5$; the percentage of idiosyncratic variance then is the same for all i.

We generated data from each model, with $n = 20, 50, 100, 200$ and $T = 20, 50, 100, 200$, for a total number of 64 experiments; we replicated each experiment 1,000 times. Before estimation, we took all variables in deviation from their sample means and divided them by their standard deviations; that is, we conducted spectral estimation on the standardized observations $x_{it} = (x^*_{it} - \bar{x}^*_i)/s_i$, where $\bar{x}^*_i = \sum_{t=1}^T x^*_{it}/T$ and $s_i^2 = \sum_{t=1}^T (x^*_{it} - \bar{x}^*_i)^2/(T-1)$. For each replication, we computed the in-sample estimates and the one-step-ahead forecasts using both SW's method and the two-step method. We estimated the spectral density matrix of the x's as

$$\hat{\mathbf{\Sigma}}(\theta) = \frac{1}{2\pi} \sum_{k=-M}^{M} w_k \hat{\mathbf{\Gamma}}_k e^{-i\theta k},$$

where $w_k = 1 - \frac{|k|}{M+1}$ with window size $M = [T^{1/2}]$. We evaluated the spectra at 101 equally spaced frequencies in the interval $[-\pi, \pi]$, namely at a grid of frequencies $\theta_h = \frac{2\pi h}{100}$, $h = -50, \ldots, 50$. We then computed the dynamic principal components decomposition, as explained in Section 3. To obtain $\hat{\mathbf{\Gamma}}^{\chi}_k$ and $\hat{\mathbf{\Gamma}}^{\xi}_k$, we used the inverse discrete Fourier transforms

$$\hat{\mathbf{\Gamma}}^{\chi}_k = \frac{2\pi}{101} \sum_{h=-50}^{50} \hat{\mathbf{\Sigma}}^{\chi}(\theta_h) e^{i\theta_h k} \qquad \text{and}$$

$$\hat{\mathbf{\Gamma}}^{\xi}_k = \frac{2\pi}{101} \sum_{h=-50}^{50} \hat{\mathbf{\Sigma}}^{\xi}(\theta_h) e^{i\theta_h k},$$

with $\hat{\mathbf{\Sigma}}^{\xi}(\theta) = \hat{\mathbf{\Sigma}}(\theta) - \hat{\mathbf{\Sigma}}^{\chi}(\theta)$—except for $\hat{\mathbf{\Gamma}}^{\xi}_0$, as expained later. Throughout, we assumed that both the number q of dynamic factors and the number $r = q(s+1)$ of static factors were known.

An important empirical finding of our simulations is that when the cross-sectional dimension n is large with respect to the period of observation T, forcing the off-diagonal entries of the estimated variance–covariance matrix $\hat{\mathbf{\Gamma}}^{\xi}_0$ of the idiosyncratic components to 0 significantly improves forecasting performance, even when the actual matrix is nondiagonal. Our explanation for this somewhat counterintuitive result is as follows. When computing $\hat{\mathbf{\Gamma}}^{\xi}_0$, we unavoidably get some spurious large covariances, even when the true covariance is 0. When n increases and T is held fixed, the number of such errors increases as n^2, the order of the number of elements in the $n \times n$ matrices $\hat{\mathbf{\Gamma}}^{\xi}_0$. In contrast, by forcing the off-diagonal entries of our estimated matrix to 0, we ignore the true off-diagonal non-0 entries. Also, in this case the error increases with n, but, due to the boundedness of the eigenvalues, it increases only linearly in n. As mentioned at the end of Section 4, replacing $\hat{\mathbf{\Gamma}}^{\xi}_0$ with any symmetric positive-semidefinite matrix with bounded eigenvalues does not affect consistency results. Therefore, we henceforth always set the off-diagonal entries of $\hat{\mathbf{\Gamma}}^{\xi}_0$ to 0 before computing eigenvectors.

We measured the performance of one-step-ahead forecasts and within-sample estimates by means of the criteria

$$\frac{\sum_{i=1}^n (\hat{\chi}_{i,T+1|T} - \chi_{i,T+1})^2}{\sum_{i=1}^n \sum_{t=1}^T \chi^2_{it}/T} \qquad \text{and}$$

$$\frac{\sum_{i=1}^n \sum_{t=1}^T (\hat{\chi}_{it} - \chi_{it})^2}{\sum_{i=1}^n \sum_{t=1}^T \chi^2_{it}}.$$

Results for Models M1, M2, M3, and M4 are given in Tables 1–8. We report the average value of the criterion, along with its empirical standard deviation (in brackets) across the 1,000 replications, for both SW's method and the two-step method.

The results of this simulation study can be summarized as follows.

1. For Model M1, the two competing methods yield similar performance for all n and T. (Recall that this model, under which the unique factor is loaded only contemporaneously, in principle is favorable to the SW method.)
2. The two-step method performs better than SW's method for models with heterogeneous dynamics, namely for Models M2, M3, and M4; in-sample estimation performance is considerably better.
3. Homogeneity in the common–idiosyncratic variance ratio (Model M4 vs. Model M3) somewhat reduces the advantage of the two-step method over SW's method, an advantage that nevertheless remains quite significant. This is an indication that in these two models a substantial gain is obtained simply from the estimation of the matrix used to project the χ's on the common factor space, whereas the advantage stemming from a better estimation of the space itself is more modest.

5.2 Some Empirical Results

D'Agostino and Giannone (2004) carried out an exhaustive comparison of the forecasting performance of our two-step method, SW's method, and some more standard methods. The exercise was based on the monthly series in the SW (2002b) dataset, which includes 150 U.S. macroeconomic series from

Table 1. Model M1, Forecasting Results

	$n = 20$		$n = 50$		$n = 100$		$n = 200$	
	Two-step	*SW*	*Two-step*	*SW*	*Two-step*	*SW*	*Two-step*	*SW*
$T = 20$	.9462	.9492	.9287	.9325	.9292	.9323	.9288	.9321
	(1.2356)	(1.1961)	(1.2035)	(1.1598)	(1.2030)	(1.1519)	(1.1971)	(1.1470)
$T = 50$	.8642	.8606	.8584	.8529	.8555	.8486	.8553	.8488
	(1.1493)	(1.1430)	(1.1472)	(1.1288)	(1.1451)	(1.1190)	(1.1439)	(1.1174)
$T = 100$	.7995	.8018	.7869	.7881	.7864	.7864	.7851	.7818
	(1.0449)	(1.0425)	(1.034)	(1.0294)	(1.0321)	(1.0288)	(1.0326)	(1.0256)
$T = 200$	.7833	.7785	.7753	.7770	.7731	.7704	.7721	.7700
	(1.0723)	(1.0698)	(1.0584)	(1.0486)	(1.0576)	(1.0541)	(1.0577)	(1.0554)

Table 2. Model M1, Within-Sample Results

	$n = 20$		$n = 50$		$n = 100$		$n = 200$	
	Two-step	*SW*	*Two-step*	*SW*	*Two-step*	*SW*	*Two-step*	*SW*
$T = 20$	.1463	.1759	.1070	.1139	.0969	.0969	.0924	.0888
	(.1030)	(.1447)	(.0551)	(.0665)	(.0447)	(.0508)	(.0403)	(.0427)
$T = 50$	.0631	.0878	.0408	.0506	.0354	.0399	.0328	.0345
	(.0351)	.0471	(.0156)	.0190	(.0111)	.0126	(.0091)	.0099
$T = 100$	.0413	.0649	.0225	.0327	.0182	.0233	.0161	.0186
	(.0232)	(.0299)	(.0075)	(.0097)	(.0046)	(.0057)	(.0034)	(.0039)
$T = 200$	.0313	.0544	.0143	.0244	.0103	.0155	.0085	.0111
	(.0181)	(.0225)	(.0044)	(.0058)	(.0022)	(.0028)	(.0014)	(.0016)

Table 3. Model M2, Forecasting Results

	$n = 20$		$n = 50$		$n = 100$		$n = 200$	
	Two-step	*SW*	*Two-step*	*SW*	*Two-step*	*SW*	*Two-step*	*SW*
$T = 20$	.8901	.9757	.7773	.8552	.7227	.7763	.6911	.7349
	(.5266)	(.5423)	(.4790)	(.4851)	(.4276)	(.4422)	(.4132)	(.4255)
$T = 50$	.6514	.7446	.5025	.5650	.4613	.4911	.4412	.4577
	(.4360)	(.4793)	(.3446)	(.3692)	(.3165)	(.3251)	(.3078)	(.3130)
$T = 100$	.5385	.6332	.3944	.4427	.3552	.3775	.3402	.3509
	(.3844)	(.4284)	(.2895)	(.3015)	(.2692)	(.2736)	(.2645)	(.2689)
$T = 200$	.4949	.5832	.3660	.4076	.3278	.3487	.3127	.3223
	(.3367)	(.3702)	(.2852)	(.2896)	(.2694)	(.2733)	(.2673)	(.2693)

Table 4. Model M2, Within-Sample Results

	$n = 20$		$n = 50$		$n = 100$		$n = 200$	
	Two-step	*SW*	*Two-step*	*SW*	*Two-step*	*SW*	*Two-step*	*SW*
$T = 20$	.4587	.7163	.3683	.5476	.3290	.4536	.3057	.3933
	(.1523)	(.2457)	(.1049)	(.1705)	(.0836)	(.1340)	(.0708)	(.1075)
$T = 50$	.2838	.5632	.1827	.3106	.1496	.2109	.1340	.1672
	(.0751)	(.1536)	(.0347)	(.0759)	(.0235)	(.0430)	(.0180)	(.0284)
$T = 100$	.2154	.4861	.1238	.2110	.0931	.1303	.0788	.0953
	(.0525)	(.1234)	(.0207)	(.0435)	(.0120)	(.0205)	(.0081)	(.0122)
$T = 200$	.1842	.4336	.0921	.1613	.0613	.0909	.0471	.0598
	(.0448)	(.0998)	(.0157)	(.0279)	(.0076)	(.0117)	(.0042)	(.0059)

Table 5. Model M3, Forecasting Results

	$n = 20$		$n = 50$		$n = 100$		$n = 200$	
	Two-step	*SW*	*Two-step*	*SW*	*Two-step*	*SW*	*Two-step*	*SW*
$T = 20$	.7644	.8645	.6680	.7748	.5713	.6383	.5292	.5499
	(.6472)	(.6793)	(.5975)	(.6102)	(.4619)	(.4517)	(.3919)	(.3522)
$T = 50$	.5273	.5913	.4636	.4962	.3565	.3773	.2792	.2771
	(.5886)	(.5993)	(.5620)	(.5393)	(.3770)	(.3601)	(.2355)	(.2139)
$T = 100$	.4482	.4966	.3935	.4094	.2873	.2957	.1958	.1943
	(.5312)	(.5367)	(.5136)	(.4806)	(.3470)	(.3210)	(.1758)	(.1678)
$T = 200$	.4131	.4690	.3493	.3719	.2488	.2562	.1521	.1563
	(.4779)	(.5099)	(.4213)	(.4275)	(.2826)	(.2793)	(.1440)	(.1435)

Table 6. Model M3, Within-Sample Results

	$n = 20$		$n = 50$		$n = 100$		$n = 200$	
	Two-step	*SW*	*Two-step*	*SW*	*Two-step*	*SW*	*Two-step*	*SW*
$T = 20$	.4143	.9583	.3104	.6670	.2970	.5440	.3025	.4505
	(.2145)	(.4796)	(.1326)	(.2913)	(.1096)	(.2299)	(.0985)	(.1893)
$T = 50$	.2285	.7279	.1444	.4083	.1335	.2753	.1326	.1856
	(.0900)	(.2752)	(.0433)	(.1223)	(.0317)	(.0791)	(.0268)	(.0525)
$T = 100$	.1682	.6468	.0921	.3176	.0831	.1839	.0809	.1044
	(.0613)	(.2184)	(.0234)	(.0758)	(.0157)	(.0415)	(.0119)	(.0218)
$T = 200$	.1386	.5959	.0662	.2646	.0554	.1344	.0501	.0647
	(.0469)	(.1748)	(.0151)	(.0473)	(.0092)	(.0222)	(.0059)	(.0096)

Table 7. Model M4, Forecasting Results

	$n = 20$		$n = 50$		$n = 100$		$n = 200$	
	Two-step	*SW*	*Two-step*	*SW*	*Two-step*	*SW*	*Two-step*	*SW*
$T = 20$	.7649	.8531	.6885	.7820	.5855	.6424	.5373	.5395
	(.6897)	(.7121)	(.6447)	(.6610)	(.5072)	(.4965)	(.4042)	(.3727)
$T = 50$	.5288	.5692	.4577	.4886	.3632	.3759	.2933	.2746
	(.5340)	(.5195)	(.4869)	(.4724)	(.3427)	(.3343)	(.2286)	(.1988)
$T = 100$	.4745	.4919	.4192	.4184	.3061	.3025	.2027	.1908
	(.5263)	(.5242)	(.5171)	(.4944)	(.3459)	(.3374)	(.1810)	(.1719)
$T = 200$	.4207	.4390	.3575	.3595	.2534	.2496	.1567	.1509
	(.4575)	(.4587)	(.4391)	(.4303)	(.2941)	(.2826)	(.1463)	(.1437)

1959:1 to 1999:2 (so that $T = 482$). These data were used to forecast two key macroeconomic variables: industrial production (IP, in log levels) and inflation (12-month log change of the Consumer Price Index (CPI)], by means of a simulated real-time forecasting exercise; that is, by comparing actual and predicted figures as the models are estimated on the time span $[1,\ \tau]$, with τ running from τ_0 (corresponding to 1969:1) to $T - h$, where h is the forecast horizon.

Table 9 reports some of their results. As in the simulations, we consider only prediction of the common components, as though the idiosyncratic components were white noise. Because the actual number r of static factors in real data obviously is unknown, we report results for r running from 6 to 15. We set the number of dynamic factors equal to 2. The forecast horizons were h =12 and 24 months, and the measure of performance used (writing y_t for the variable to be predicted and $\hat{y}_{t+h|t}$ for the corresponding h-step-ahead forecast) was

$$\sum_{\tau=\tau_0}^{T-h} (\hat{y}_{\tau+h|\tau} - y_{\tau+h})^2/(T - h - \tau_0 + 1).$$

The best result for each variable and method is in bold type. The two-step method outperforms SW, for any given r, except for IP with $h = 12$ and $r = 6, 7, 8, 11$. Comparing the minima, the two-step method performs moderately better in all experiments. Both SW's method and our method perform considerably better than autoregression-based forecasts. (The corresponding figures are not reported here, but can be found in D'Agostino and Giannone 2004.)

The same result, with the two-step method outperforming SW's method, was given by Gentile (2004) for Italian data. In that study, prediction of aggregate inflation and industrial production was obtained by applying the factor model to the panel of elementary price and production indexes respectively. The techniques introduced by Boivin and Ng (2005) to select the "good" variables in the panel were shown to produce considerable improvement.

6. SUMMARY AND CONCLUSIONS

In this article we have proposed a new forecasting method that exploits information from a large panel of time series. The method is based on the dynamic factor model proposed by FHLR (2000) and proceeds in two steps. In the first step, we estimate the lagged covariances of the common and idiosyncratic components using the frequency domain approach proposed by FHLR (2000). In the second step, we use information about the "degree of commonality" of each variable to estimate the common factors and project the variables to be predicted onto the linear space spanned by these factors. We have shown that the projection converges to the optimal forecast as n and T go to infinity. Being a linear combination of the x's that does not involve future observations, the two-step predictor solves the end-of-sample problems caused by two-sided filtering in the estimation method of FHLR (2000), while exploiting the advantages of dynamic information. Both theoretical arguments and simulation results suggest that our predictor can provide a substantial improvement over SW's principal components predictor when the various cross-sectional items differ significantly in the lag structure of the factor loadings, particularly if in addition there is heterogeneity in the fraction of total variance explained by the idiosyncratic components. Empirical exercises conducted so far are consistent with such arguments and results.

APPENDIX: PROOFS

We first present Lemmas A.1 and A.2, used in the proofs of Proposition 1 and Lemma 2.

Table 8. Model M4, Within-Sample Results

	$n = 20$		$n = 50$		$n = 100$		$n = 200$	
	Two-step	*SW*	*Two-step*	*SW*	*Two-step*	*SW*	*Two-step*	*SW*
$T = 20$	.3883	.8443	.3011	.6068	.2881	.4920	.2937	.4084
	(.1775)	(.3858)	(.1297)	(.2761)	(.1100)	(.2223)	(.1038)	(.1869)
$T = 50$	.2284	.6099	.1487	.3566	.1359	.2404	.1324	.1668
	(.0641)	(.1763)	(.0393)	(.0986)	(.0317)	(.0678)	(.0283)	(.0476)
$T = 100$	.1799	.5217	.1024	.2685	.0896	.1586	.0839	.0949
	(.0371)	(.1059)	(.0189)	(.0510)	(.0148)	(.0309)	(.0122)	(.0187)
$T = 200$	.1594	.4734	.0811	.2216	.0656	.1172	.0561	.0611
	(.0262)	(.0690)	(.0111)	(.0283)	(.0080)	(.0150)	(.0059)	(.0079)

Table 9. Mean Squared Forecast Errors for the Two-Step and SW Methods

	log IP				$100 \times (1 - L^{12})$ log CPI			
	$h = 12$		$h = 24$		$h = 12$		$h = 24$	
r	Two-step	SW	Two-step	SW	Two-step	SW	Two-step	SW
6	12.01	11.77	29.53	**30.05**	2.83	2.90	6.75	6.88
7	11.85	11.43	29.54	30.30	**2.79**	2.88	6.76	**6.84**
8	11.51	**11.29**	29.44	31.57	2.81	**2.86**	6.68	6.88
9	11.32	11.53	29.30	31.74	2.83	2.86	6.57	6.95
10	11.24	11.35	29.32	32.55	2.84	2.93	6.65	6.97
11	11.31	11.30	29.29	32.79	2.87	2.94	6.61	7.09
12	11.29	11.39	29.38	33.20	2.90	2.97	6.48	6.95
13	11.19	11.66	29.38	34.28	2.97	3.09	6.42	7.04
14	**10.96**	11.50	29.24	33.63	2.96	3.16	6.29	7.18
15	11.09	11.59	**29.07**	35.17	2.95	3.24	**6.27**	7.07

Lemma A.1. Given the integer $k > 0$, consider a sequence of real, symmetric, positive semidefinite $n \times n$ matrices $\mathbf{\Gamma}_n$ and a sequence of real, symmetric, positive definite $n \times n$ matrices $\mathbf{D}_n$, $n = k, k+1, \ldots$. Assume that the following hold:

(a) $\mathbf{\Gamma}_n$'s kth largest eigenvalue μ_{nk} diverges as $n \to \infty$.
(b) $\mathbf{D}_n$'s largest eigenvalue is bounded from above by δ.

Then the kth largest generalized eigenvalue of $(\mathbf{\Gamma}_n, \mathbf{D}_n)$, ν_{nk}, diverges as $n \to \infty$.

Proof. Let $\mathbf{v}_{nj}$, for $j = 1, \ldots, k-1$, be the generalized eigenvectors corresponding to the $(k-1)$ generalized eigenvalues of the couple $(\mathbf{\Gamma}_n, \mathbf{D}_n)$, and let $\mathbf{w}_{nj}, j = 1, \ldots, k$, denote the standard (unit-modulus) eigenvectors corresponding to the first k eigenvalues of $\mathbf{\Gamma}_n$. Let $\alpha_{n1}, \alpha_{n2}, \ldots, \alpha_{nk}$ be any nontrivial solution of the linear system of $(k-1)$ equations in the k unknowns y_j,

$$(y_1\mathbf{w}_{n1} + y_2\mathbf{w}_{n2} + \cdots + y_k\mathbf{w}_{nk})\mathbf{D}_n\mathbf{v}_{nj}' = 0, \qquad j = 1, \ldots, k-1.$$

Define $\mathbf{q}_n = \alpha_{n1}\mathbf{w}_{n1} + \alpha_{n2}\mathbf{w}_{n2} + \cdots + \alpha_{nk}\mathbf{w}_{nk}$. The vectors $\mathbf{w}_{nj}$ are orthonormal, so that $\mathbf{q}_n \neq 0$. Therefore, because $\mathbf{D}_n$ is positive definite, $\mathbf{q}_n\mathbf{D}_n\mathbf{q}_n' > 0$. Thus, rescaling the α's,

$$\mathbf{q}_n\mathbf{D}_n\mathbf{v}_{nj}' = 0, \qquad j = 1, \ldots, k-1, \qquad \text{and} \qquad \mathbf{q}_n\mathbf{D}_n\mathbf{q}_n' = 1. \tag{A.1}$$

[For $k = 1$, (A.1) does not apply, and we are just setting $\alpha_{n1} = 1/\sqrt{\mathbf{w}_{n1}\mathbf{D}_n\mathbf{w}_{n1}{}'}$.] It follows from assumption (b) and (A.1) that $\alpha_{n1}^2 + \alpha_{n2}^2 + \cdots + \alpha_{nk}^2 \geq 1/\delta$. This and the definition of $\mathbf{w}_{nj}$ imply that

$$\mathbf{q}_n\mathbf{\Gamma}_n\mathbf{q}_n{}' = \alpha_1^2\mu_{n1} + \alpha_{n2}^2\mu_{n2} + \cdots + \alpha_{nk}^2\mu_{nk} \geq \frac{1}{\delta}\mu_{nk}.$$

But, in view of (A.1), $\nu_{nk} \geq \mathbf{q}_n\mathbf{\Gamma}_n\mathbf{q}_n{}'$. The conclusion follows.

Lemma A.2. Let $\check{\mathbf{\Gamma}}_0^{\chi}$ and $\check{\mathbf{\Gamma}}_0^{\xi}$ be as in (19), $\check{\mu}_k^{\chi}$ and $\check{\mu}_k^{\xi}$ being, respectively, their eigenvalues. Let $r = q(s+1)$. Then, under assumptions A, B, C, and D, the following results holds:

(a) $\check{\mu}_r^{\chi} \to \infty$ as $n \to \infty$.
(b) $\check{\mu}_1^{\xi}$ is bounded for $n \to \infty$.

Proof. For any n-dimensional unit-modulus row vector $\mathbf{v}$, we have

$$\mathbf{v}\check{\mathbf{\Gamma}}_0^{\xi}\mathbf{v}' = \mathbf{v}\left[\int_{-\pi}^{\pi} \check{\mathbf{\Sigma}}^{\xi}(\theta)\, d\theta\right]\mathbf{v}'$$
$$= \int_{-\pi}^{\pi} \mathbf{v}\check{\mathbf{\Sigma}}^{\xi}(\theta)\mathbf{v}'\, d\theta \leq \int_{-\pi}^{\pi} \lambda_{q+1}(\theta)\, d\theta = \alpha, \qquad \text{say.}$$

We have (see Lancaster and Tismenetsky 1985, p. 301, thm. 1) $\lambda_{q+1}(\theta) \leq \lambda_{q+1}^{\chi}(\theta) + \lambda_1^{\xi}(\theta)$. Thus, because $\lambda_{q+1}^{\chi}(\theta) = 0$, assumption C2 implies that $\alpha \leq 2\pi\Lambda$. Part (b) of the lemma follows. Obviously, C2 implies that $\mathbf{v}\mathbf{\Gamma}_0^{\xi}\mathbf{v}' \leq 2\pi\Lambda$. Setting $\mathbf{A} = \mathbf{\Gamma}_0^{\xi} - \check{\mathbf{\Gamma}}_0{}^{\xi}$ and observing that $\mathbf{v}\mathbf{\Gamma}_0^{\xi}\mathbf{v}'$ and $\mathbf{v}\check{\mathbf{\Gamma}}_0^{\xi}\mathbf{v}'$ are nonnegative, we obtain $|\mathbf{v}\mathbf{A}\mathbf{v}'| = |\mathbf{v}\mathbf{\Gamma}_0^{\xi}\mathbf{v}' - \mathbf{v}\check{\mathbf{\Gamma}}_0^{\xi}\mathbf{v}'| \leq 2\pi\Lambda$. Because

$$\mathbf{A} = \int_{-\pi}^{\pi} (\mathbf{\Sigma}^{\xi}(\theta) - \check{\mathbf{\Sigma}}^{\xi}(\theta))\, d\theta$$
$$= \int_{-\pi}^{\pi} (\check{\mathbf{\Sigma}}^{\chi}(\theta) - \mathbf{\Sigma}^{\chi}(\theta))\, d\theta = \check{\mathbf{\Gamma}}_0^{\chi} - \mathbf{\Gamma}_0^{\chi},$$

it follows that $\check{\mathbf{\Gamma}}_0^{\chi} + 2\pi\Lambda\mathbf{I} = \mathbf{\Gamma}_0^{\chi} + [2\pi\Lambda\mathbf{I} + \mathbf{A}]$. Because the matrix in square brackets is positive semidefinite, the result of Lancaster and Tismenetsky mentioned earlier implies that the eigenvalues of the sum on the left side are larger than or equal to the corresponding eigenvalues of $\mathbf{\Gamma}_0^{\chi}$. This entails that $\check{\mu}_r^{\chi} + 2\pi\Lambda \geq \mu_r^{\chi}$; part (a) of the lemma thus follows from assumption D.

Proof of Lemma 2. Let $\mathbf{v} = (\mathbf{v}_1 \ \cdots \ \mathbf{v}_k)'$ and $\mathbf{v}_n = (v_{n1} \ \cdots \ v_{nk})'$. With no loss of generality, we can suppose that $\mathbf{v}$ is orthonormal. Consider the decomposition

$$\mathbf{v}_n = \mathbf{a}_n\mathbf{v} + \mathbf{R}_n \tag{A.2}$$

of $\mathbf{v}_n$ into its (componentwise) orthogonal projection $\mathbf{a}_n\mathbf{v}$ onto $\mathcal{K}$ and the orthogonal complement (where $\mathbf{a}_n$ is a $k \times k$ matrix). By assumption, $\mathbf{R}_n \to 0$ in quadratic mean, whereas the assumption $\text{var}(v_{nj}) = 1$ implies that $\mathbf{a}_n$ is bounded. Then consider the projection of $\mathbf{v}$ on $\mathcal{K}_n$. It is easily seen that

$$\mathbf{v} = \mathbf{a}_n'\mathbf{v}_n + \mathbf{S}_n. \tag{A.3}$$

Taking covariances in (A.2) and (A.3), we obtain, in view of the orthonormality assumption on $\mathbf{v}$ and $\mathbf{v}_n$, that $\mathbf{I}_k = \mathbf{a}_n\mathbf{a}_n' + \mathbf{\Gamma}_n^R$ and $\mathbf{I}_k = \mathbf{a}_n'\mathbf{a}_n + \mathbf{\Gamma}_n^S$, so that $\mathbf{a}_n\mathbf{a}_n' + \mathbf{\Gamma}_n^R = \mathbf{a}_n'\mathbf{a}_n + \mathbf{\Gamma}_n^S$. Taking traces on both sides yields $\text{tr}(\mathbf{\Gamma}_n^R) = \text{tr}(\mathbf{\Gamma}_n^S)$. Thus $\mathbf{S}_n \to 0$ in quadratic mean as $n \to \infty$.

Decomposing similarly v into $v = \mathbf{b}_n\mathbf{v}_n + s_n = \mathbf{b}_n\mathbf{a}_n\mathbf{v} + \mathbf{b}_n\mathbf{R}_n + s_n$ and $v = \mathbf{b}\mathbf{v} + s$, where $\mathbf{b}_n\mathbf{v}_n$ and $\mathbf{b}\mathbf{v}$ denote the orthogonal projections of v onto $\mathcal{K}_n$ and $\mathcal{K}$ (and $\mathbf{b}_n$ and $\mathbf{b}$ are $1 \times k$), we obtain

$$\text{Proj}(v|\mathcal{K}_n) - \text{Proj}(v|\mathcal{K}) = \mathbf{b}_n\mathbf{v}_n - \mathbf{b}\mathbf{v}$$
$$= (\mathbf{b}_n\mathbf{a}_n - \mathbf{b})\mathbf{v} + \mathbf{b}_n\mathbf{R}_n$$
$$= s - s_n.$$

The assumption that $\text{var}(v_{nj}) = 1$ implies that $\mathbf{b}_n$ is bounded. Hence $\mathbf{b}_n\mathbf{R}_n \to 0$ in quadratic mean, and

$$[(\mathbf{b}_n\mathbf{a}_n - \mathbf{b})\mathbf{v} - (s - s_n)] \longrightarrow 0 \tag{A.4}$$

in quadratic mean. Now, (A.3) and the fact that $\mathbf{S}_n \to 0$ imply that $\text{cov}(\mathbf{v}, s_n) \to 0$. Because $\mathbf{a}_n$ and $\mathbf{b}_n$ are bounded, $|\text{cov}((\mathbf{b}_n\mathbf{a}_n - \mathbf{b})\mathbf{v}, s - s_n)| = |\text{cov}((\mathbf{b}_n\mathbf{a}_n - \mathbf{b})\mathbf{v}, s_n)| \to 0$. This, combined with (A.4), implies that $\lim_{n\to\infty}(\mathbf{b}_n\mathbf{a}_n - \mathbf{b})\mathbf{v} = \lim_{n\to\infty}(s - s_n) = 0$.

[Received June 2003. Revised October 2004.]

REFERENCES

Anderson, T. W. (1984), *An Introduction to Multivariate Statistical Analysis*, New York: Wiley.

Bai, J. (2003), "Inferential Theory for Factor Models of Large Dimension," *Econometrica*, 71, 135–171.

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

Boivin, J., and Ng, S. (2005), "Are More Data Always Better for Factor Analysis?" *Journal of Econometrics*, to appear.

Brillinger, D. R. (1981), *Time Series: Data Analysis and Theory*, San Francisco: Holden-Day.

Brockwell, P. J., and Davis, R. A. (1991), *Time Series: Theory and Methods* (2nd ed.), New York: Springer-Verlag.

Chamberlain, G. (1983), "Funds, Factors, and Diversification in Arbitrage Pricing Models," *Econometrica*, 51, 1281–1304.

Chamberlain, G., and Rothschild, M. (1983), "Arbitrage, Factor Structure and Mean-Variance Analysis in Large Asset Markets," *Econometrica*, 51, 1305–1324.

D'Agostino, A., and Giannone, D. (2004), "Comparing Alternative Predictors Based on Large-Panel Dynamic Factor Models," ISRO discussion paper, Free University of Brussels.

Doz, C., Giannone, D., and Reichlin, L. (2003), "A Maximum Likelihood Approach to Dynamic Factor Analysis in Large Panels," mimeo, ECARES, Free University of Brussels.

Forni, M., Hallin, M., Lippi, M., and Reichlin, L. (2000), "The Generalized Factor Model: Identification and Estimation," *The Review of Economics and Statistics*, 82, 540–554.

——— (2004), "The Generalized Dynamic Factor Model: Consistency and Rates," *Journal of Econometrics*, 119, 231–255.

Forni, M., and Lippi, M. (2001), "The Generalized Dynamic Factor Model: Representation Theory," *Econometric Theory*, 17, 1113–1141.

Forni, M., and Reichlin, L. (1998), "Let's Get Real: A Factor Analytical Approach to Disaggregated Business Cycle Dynamics," *Review of Economic Studies*, 65, 453–473.

Gentile, M. (2004), "Forecasting Industrial Production Index, Inflation and Asset Return Volatility by Generalized Dynamic Factor Model," unpublished doctoral dissertation, Sant'Anna School of Advanced Studies, Pisa.

Geweke, J. (1977), "The Dynamic Factor Analysis of Economic Time Series," in *Latent Variables in Socio-Economic Models*, eds. D. J. Aigner and A. S. Goldberger, Amsterdam: North-Holland, pp. 365–383.

Lancaster, P., and Tismenetsky, M. (1985), *The Theory of Matrices* (2nd ed.), Orlando, FL: Academic Press.

Lawley, D. N., and Maxwell, A. E. (1971), *Factor Analysis as a Statistical Method*, London: Butterworths.

Quah, D., and Sargent, T. J. (1993), "A Dynamic Index Model for Large Cross-Sections," in *Business Cycles, Indicators, and Forecasting*, eds. J. H. Stock and M. W. Watson, Chicago: National Bureau of Economic Statistics and University of Chicago Press, pp. 285–309.

Rozanov, Y. A. (1967), *Stationary Random Processes*, San Francisco: Holden-Day.

Sargent, T. J., and Sims, C. A. (1977), "Business Cycle Modeling Without Pretending to Have too Much a priori Economic Theory," in *New Methods in Business Research*, ed. C. A. Sims, Minneapolis: Federal Reserve Bank of Minneapolis, pp. 45–109.

Stock, J. H., and Watson, M. W. (2002a), "Forecasting Using Principal Components From a Large Number of Predictors," *Journal of the American Statistical Association*, 97, 1167–1179.

——— (2002b), "Macroeconomic Forecasting Using Diffusion Indexes," *Journal of Business & Economic Statistics*, 20, 147–162.

Wiener, N., and Masani, P. (1957), "The Prediction Theory of Multivariate Stochastic Processes I," *Acta Mathematica*, 98, 111–150.

——— (1958), "The Prediction Theory of Multivariate Stochastic Processes II," *Acta Mathematica*, 99, 93–137.