The Generalized Dynamic Factor Model one-sided estimation and forecasting *

Mario Forni,
Università di Modena and CEPR

Marc Hallin
ISRO, ECARES, and Département de Mathématique
Université Libre de Bruxelles

Marco Lippi
Università di Roma La Sapienza

and

Lucrezia Reichlin
ECARES, Université Libre de Bruxelles and CEPR

Abstract

This paper proposes a new forecasting method which makes use of information from a large panel of time series. As in Forni, Hallin, Lippi and Reichlin (2000), and in Stock and Watson (2002a,b), the method is based on a dynamic factor model. We argue that our method improves upon a standard principal component predictor in that, first, it fully exploits all the dynamic covariance structure of the panel and, second, it 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.

JEL subject classification : C13, C33, C43. Key words and phrases : Dynamic factor models, principal components, time series, large cross-sections, panel data, forecasting.

1 Introduction

Economists and forecasters nowadays 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, hence 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

*Research supported by a P.A.I of the Belgian Federal Government, an A.R.C. 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).

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 to be defined more precisely below) of the idiosyncratic components ξ_{it}, which need not, as in traditional factor models, be mutually orthogonal across the panel.

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

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

In the present paper, still in a non-parametric spirit, we combine the advantages of FHLR and SW to propose a new predictor. Like in previous literature,

(a) 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}.

(b) Such prediction is obtained by, firstly, 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 paper lies both in the estimation of the factor space and in the way 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

(A) a new estimation of the factor space: we employ 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.

Both our two-step predictor and SW's 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, in simulations as well as on SW's own data set. Intuitively, our predictor indeed has a twofold advantage over SW's.

First, while SW's estimation of the h-step ahead projection matrix is based on the lag-h covariance matrix of the x's, our method employs the first-step estimate of the lag-h covariance matrix of the common components. Such 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 component method performs better than SW's standard principal component one in approaching the factor space, since 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 having larger idiosyncratic components, so that the idiosyncratic 'error' contained in the linear combination is minimized.

The paper 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 our two-step method is compared to SW's method by using simulated panels. The results of an in-depth comparison based on the empirical panel used in Stock and Watson (2002b) are briefly reported. Section 6 concludes. Some lemmas, which are used in Section 4, are stated and proved in the Appendix.

2 The model

The model used in the present paper is an *approximate* factor model, in that the idiosyncratic components are allowed to be weakly correlated, like in Chamberlain (1983) and Chamberlain and Rothschild (1983), and contrary to 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, like in FHLR (2000), Forni and Lippi (2001), SW (2002a,b), Bai and Ng (2002), and Bai (2003). However, unlike in 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 that

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 zero mean 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 require in addition 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 four, 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 $n, i, j \in \mathbb{N}$.

Assumptions A1 and A2 jointly will be referred to as assumption A.

To avoid heavy notation, 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 to be defined below, is not made explicit. In the same way, explicit reference to T will be avoided for estimated quantities. For example, an estimate of $\mathbf{\Gamma}_k$, which depends on n and T, will be denoted 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 as compared to n in empirical applications). In vector notation, defining $\boldsymbol{\chi}_t = (\chi_{1t}\ \cdots\ \chi_{nt})'$ and $\boldsymbol{\xi}_t = (\xi_{1t}\ \ldots\ \xi_{nt})'$, and denoting by $\mathbf{B}(L)$ as the $n \times q$ matrix whose (i,j) entry is $b_{ij}(L)$, our model is thus

$$\mathbf{x}_t = \boldsymbol{\chi}_t + \boldsymbol{\xi}_t = \mathbf{B}(L)\mathbf{f}_t + \boldsymbol{\xi}_t, \tag{2.1}$$

where the factors $\mathbf{f}_t$ follow a VAR scheme of the form $\mathbf{A}(L)\mathbf{f}_t = \mathbf{u}_t$. Our assumptions on (2.1) are the following.

B1.(a) For $i \in \mathbb{N}$ and $j = 1, \ldots, q$, 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 $\{\mathbf{u}_t = (u_{1t}\ \ldots\ u_{qt})',\ t \in \mathbb{Z}\}$ of *common shocks*, is a q-dimensional orthonormal white noise process orthogonal to $\{\xi_{it},\ i = 1, \ldots, n,\ t \in \mathbb{Z}\}$ (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 shall return to B1(b) in the next section.

Under Assumption A, finite second-order moments exist for all variables involved in the model. Let $\boldsymbol{\Sigma}^{\chi}(\theta)$, $\boldsymbol{\Sigma}^{\xi}(\theta)$, $\theta \in [-\pi, \pi]$, be the $n \times n$ spectral density matrices of $\boldsymbol{\chi}_t$ and $\boldsymbol{\xi}_t$, respectively, and denote by λ_k^{χ}, λ_k^{ξ} 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 $\lambda_k^{\xi}(\theta)$ stand for the k-th largest eigenvalues of $\boldsymbol{\Sigma}^{\chi}(\theta)$ and $\boldsymbol{\Sigma}^{\xi}(\theta)$, respectively. 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 $[-\pi\ \pi]$, $k = 1, \ldots, q$.

C2. There exists a 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). It makes proofs easier while not causing a serious loss of generality. Assumptions C1(a) and C2 are needed to guarantee identification of the common and the 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, though 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 the Introduction.

Letting $\mathbf{F}_t = (\mathbf{f}_t'\ \mathbf{f}_{t-1}'\ \ldots\ \mathbf{f}_{t-s}')'$ and $\mathbf{C} = (\mathbf{B}_0\ \mathbf{B}_1\ \cdots\ \mathbf{B}_s)$, model (2.1) also can be written as

$$\mathbf{x}_t = \mathbf{C}\mathbf{F}_t + \boldsymbol{\xi}_t, \tag{2.2}$$

in which $r = q(s+1)$ common factors are loaded only contemporaneously. Equation (2.2) looks like a static factor model. However, the dynamic nature of (2.1) 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 'static factors' the factors of the static representation (2.2), i.e. the r entries of $\mathbf{F}_t$, and 'dynamic factors' the q entries of $\mathbf{f}_t$.

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

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 $i \geq 2$, C1(a) clearly holds, with $q = 1$, but D does not, since $\mu_1^{\chi} \to \infty$ whereas μ_2^{χ} is bounded as $n \to \infty$. Further technical assumptions will be introduced in Section 4.

3 A two-step forecasting method

As already observed in the Introduction, the paper concentrates on forecasting the common components $\chi_{i,T+h}$. As a by-product 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 (i) the factors and (ii) the covariances between $\chi_{i,T+h}$, or χ_t, and the factors.

Formally, denote by $\mathcal{G}(\mathbf{F}, t)$ 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

$$\chi_{it} = \text{Proj}(\chi_{it} | \mathcal{G}(\mathbf{F}, t)). \tag{3.1}$$

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, based on F_{T-k}, $k \geq 0$, of $\chi_{i,T+h}$, denoted by $\chi_{i,T+h|T}$, is given by the linear projection

$$\chi_{i,T+h|T} = \text{Proj}(\chi_{i,T+h} | \mathcal{G}(\mathbf{F}, T)). \tag{3.2}$$

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, optimal prediction would require the projection on a space including past values of $\mathbf{F}_t$, this implying only minor changes in our statements (Proposition 4.1 in particular) and proofs.

Steps 1 and 2 of our procedure estimate (3.1) and (3.2) by estimating, in reverse order, the factor space and the relevant covariance matrices.

3.1 Step one: estimating $\mathbf{\Gamma}_k^{\chi}$ and $\mathbf{\Gamma}_k^{\xi}$

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

$$\underline{\hat{\mathbf{p}}}_j(L) = \frac{1}{2\pi} \sum_{k=-\infty}^{\infty} \left[\int_{-\pi}^{\pi} \hat{\mathbf{p}}_j(\theta) e^{ik\theta} d\theta \right] L^k, \tag{3.3}$$

(the inverse Fourier transform of $\hat{\mathbf{p}}_j(\theta)$), the j-th dynamic principal components of $\mathbf{x}_t$ is defined as $\underline{\hat{\mathbf{p}}}_j(L)\mathbf{x}_t$. The first q dynamic principal components are used to obtain a consistent (see FHLR 2000) estimation $\hat{\boldsymbol{\chi}}_t^D = \underline{\hat{\mathbf{P}}}^*(L)\left[\underline{\hat{\mathbf{P}}}(L)\mathbf{x}_t\right]$ of the χ's (with obvious notation $\underline{\hat{\mathbf{P}}}(L)$; star indicates complex conjugation and transposition; superscript ^{D} stands for 'dynamic method').

The trouble with this estimation method is that the filters (3.3) 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

$$\hat{\boldsymbol{\Sigma}}^{\chi}(\theta) = \hat{\lambda}_1(\theta)\hat{\mathbf{p}}_1^*(\theta)\hat{\mathbf{p}}_1(\theta) + \cdots + \hat{\lambda}_q(\theta)\hat{\mathbf{p}}_q^*(\theta)\hat{\mathbf{p}}_q(\theta) \tag{3.4}$$

$$\hat{\boldsymbol{\Sigma}}^{\xi}(\theta) = \hat{\lambda}_{q+1}(\theta)\hat{\mathbf{p}}_{q+1}^*(\theta)\hat{\mathbf{p}}_{q+1}(\theta) + \cdots + \hat{\lambda}_n(\theta)\hat{\mathbf{p}}_n^*(\theta)\hat{\mathbf{p}}_n(\theta). \tag{3.5}$$

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

$$\hat{\boldsymbol{\Gamma}}_k^{\chi} = \int_{-\pi}^{\pi} e^{ik\theta}\,\hat{\boldsymbol{\Sigma}}^{\chi}(\theta)d\theta \quad \text{and} \quad \hat{\boldsymbol{\Gamma}}_k^{\xi} = \int_{-\pi}^{\pi} e^{ik\theta}\,\hat{\boldsymbol{\Sigma}}^{\xi}(\theta)d\theta. \tag{3.6}$$

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 which are more efficient than standard principal components.

Consider all the linear combinations $\mathbf{a}\mathbf{x}_t = a_1x_{1t} + \cdots + a_nx_{nt}$ of the x's which 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 which 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

$$\hat{\mathbf{Z}}_j = \text{Arg}\max_{\mathbf{a}\in\mathbb{R}^n}\mathbf{a}\,\hat{\boldsymbol{\Gamma}}_0^{\chi}\,\mathbf{a}' \quad \text{s.t.} \quad \mathbf{a}\,\hat{\boldsymbol{\Gamma}}_0^{\xi}\,\mathbf{a}' = 1 \quad \text{and} \quad \mathbf{a}\,\hat{\boldsymbol{\Gamma}}_0^{\xi}\,\hat{\mathbf{Z}}_m' = 0, \quad 1 \le m \le j-1. \tag{3.7}$$

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* $\hat{\nu}_j$ of the couple of matrices $(\hat{\boldsymbol{\Gamma}}_0^{\chi},\,\hat{\boldsymbol{\Gamma}}_0^{\xi})$, i.e. the solutions of

$$\hat{\mathbf{Z}}_j\,\hat{\boldsymbol{\Gamma}}_0^{\chi} = \hat{\nu}_j\,\hat{\mathbf{Z}}_j\,\hat{\boldsymbol{\Gamma}}_0^{\xi} \qquad j = 1,2,\ldots,n, \tag{3.8}$$

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

$$\hat{\mathbf{Z}}\mathbf{x}_t = \left(\hat{\mathbf{Z}}_1\mathbf{x}_t \ \hat{\mathbf{Z}}_2\mathbf{x}_t \ \cdots \ \hat{\mathbf{Z}}_r\mathbf{x}_t\right)'. \tag{3.9}$$

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 (3.1) and (3.2), in vector form, are easily obtained as

$$\hat{\boldsymbol{\chi}}_t = \left[\hat{\boldsymbol{\Gamma}}_0^{\chi} \ \hat{\mathbf{Z}}' \ \left(\hat{\mathbf{Z}} \ \hat{\boldsymbol{\Gamma}}_0 \ \hat{\mathbf{Z}}'\right)^{-1}\right] \left[\hat{\mathbf{Z}} \ \mathbf{x}_t \ \right] \tag{3.10}$$

and

$$\hat{\boldsymbol{\chi}}_{T+h|T} = \left[\hat{\boldsymbol{\Gamma}}_h^{\chi} \ \hat{\mathbf{Z}}' \left(\hat{\mathbf{Z}} \ \hat{\boldsymbol{\Gamma}}_0 \ \hat{\mathbf{Z}}'\right)^{-1}\right] \left[\hat{\mathbf{Z}} \ \mathbf{x}_T\right], \tag{3.11}$$

respectively; these projections are the two-step estimators (predictors) we are proposing.

An alternative method to estimate the best linear predictor of the χ's might be a *factorisation* of $\hat{\boldsymbol{\Sigma}}^{\chi}(\theta)$ along the lines 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 non-singular submatrix. Moreover, the matrix to be inverted under our approach is only $r \times r$ (see (3.10) and (3.11)), with r independent of n, so that the motivation for factorization as opposed to the projection method (pp. 100-102 of Wiener and Masani 1958) loses much of its strength.

To get an intuition 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 $1/\sigma_i^{\xi}$, i.e. 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 ML estimators of F_t under the assumption of known diagonal idiosyncratic covariance matrix (see Lawley and Maxwell 1971, Chapter 4).

3.3 Comparison with Stock and Watson's 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 j-th 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 $\hat{\mathbf{S}} = \left(\hat{\mathbf{S}}_1' \ \cdots \ \hat{\mathbf{S}}_r'\right)'$ (a $r \times n$ matrix). Stock and Watson's estimation of the factor space and the projections are given by

$$\hat{\mathbf{S}}\mathbf{x}_t = \left(\hat{\mathbf{S}}_1\mathbf{x}_t \ \hat{\mathbf{S}}_2\mathbf{x}_t \ \cdots \ \hat{\mathbf{S}}_r\mathbf{x}_t\right)' \tag{3.12}$$

$$\hat{\boldsymbol{\chi}}_t^{SW} = \left[\hat{\boldsymbol{\Gamma}}_0 \ \hat{\mathbf{S}}' \ \left(\hat{\mathbf{S}} \ \hat{\boldsymbol{\Gamma}}_0 \ \hat{\mathbf{S}}'\right)^{-1}\right] \left[\hat{\mathbf{S}} \ \mathbf{x}_t\right] = \hat{\mathbf{S}}' \ [\hat{\mathbf{S}}\mathbf{x}_t] \tag{3.13}$$

$$\hat{\boldsymbol{\chi}}_{T+h|T}^{SW} = \left[\hat{\boldsymbol{\Gamma}}_h \ \hat{\mathbf{S}}' \ \left(\hat{\mathbf{S}} \ \hat{\boldsymbol{\Gamma}}_0 \ \hat{\mathbf{S}}'\right)^{-1}\right] \left[\hat{\mathbf{S}} \ \mathbf{x}_t\right] = \hat{\boldsymbol{\Gamma}}_h\hat{\mathbf{S}}'\hat{\mathbf{M}}^{-1} \ [\hat{\mathbf{S}}\mathbf{x}_T]. \tag{3.14}$$

Let us compare this with our two-step method.

(I) Since 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 therefore 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 $\hat{\mathbf{S}}\mathbf{x}_t$.

Though 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, a formal derivation of optimality properties or relative efficiency values is extremely difficult in such a general context, and will not be pursued in this paper. However, an important insight into the finite-sample relative performances of our method and SW's will be obtained 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, denote by $\hat{\mathbf{\Sigma}}(\theta) = (\hat{\sigma}_{ij}(\theta))$ 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 $\epsilon > 0$,

$$\lim_{T\to\infty} \mathrm{P}\left[\max_{1\le i,j\le n} \sup_{\theta\in[-\pi,\pi]} |\hat{\sigma}_{ij}(\theta) - \sigma_{ij}(\theta)| > \epsilon\right] = 0. \tag{4.1}$$

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 Theorem 11.7.2, p. 447, which extends Theorem 10.4.1 to the multidimensional case). Define

$$\begin{aligned}
\check{\mathbf{\Sigma}}^{\chi}(\theta) &= \lambda_1(\theta)\mathbf{p}_1^*(\theta)\mathbf{p}_1(\theta) + \cdots + \lambda_q(\theta)\mathbf{p}_q^*(\theta)\mathbf{p}_q(\theta) \\
\check{\mathbf{\Sigma}}^{\xi}(\theta) &= \lambda_{q+1}(\theta)\mathbf{p}_{q+1}^*(\theta)\mathbf{p}_{q+1}(\theta) + \cdots + \lambda_n(\theta)\mathbf{p}_n^*(\theta)\mathbf{p}_n(\theta),
\end{aligned}$$

where $\lambda_{nj}(\theta)$ and $\mathbf{p}_{nj}(\theta)$ 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 $\boldsymbol{\xi}_t$, respectively. They are, so to speak, estimates of such matrices for T infinite 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 (4.1) applies to the entries of $\hat{\mathbf{\Sigma}}^{\chi}$ and $\check{\mathbf{\Sigma}}^{\chi}$ respectively. More precisely, for any given n and any $\epsilon > 0$,

$$\lim_{T\to\infty} \mathrm{P}\left[\max_{1\le i,j\le n} \sup_{\theta\in[-\pi,\pi]} |\hat{\sigma}_{ij}^{\chi}(\theta) - \check{\sigma}_{ij}^{\chi}(\theta)| > \epsilon\right] = 0. \tag{4.2}$$

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 $\check{\mathbf{\Sigma}}^{\xi}$ respectively. Moreover, (4.2) implies that $\hat{\mathbf{\Gamma}}_k^{\chi}$ and $\hat{\mathbf{\Gamma}}_k^{\xi}$, as defined in (3.6), are consistent estimators, of

$$\check{\mathbf{\Gamma}}_k^{\chi} = \int_{-\pi}^{\pi} e^{ik\theta}\check{\mathbf{\Sigma}}^{\chi}(\theta)d\theta \quad \text{and} \quad \check{\mathbf{\Gamma}}_k^{\xi} = \int_{-\pi}^{\pi} e^{ik\theta}\check{\mathbf{\Sigma}}^{\xi}(\theta)d\theta, \tag{4.3}$$

respectively.

Denote by $\check{\mu}_k^{\chi}$ and $\check{\mu}_k^{\xi}$ the eigenvalues of $\check{\mathbf{\Gamma}}_0^{\chi}$ and $\check{\mathbf{\Gamma}}_0^{\xi}$ respectively. Lemma 7.2 (see the Appendix) proves that, under Assumptions A, B, C and D, $\lim_{n\to\infty} \check{\mu}_r^{\chi} = \infty$ while $\check{\mu}_1^{\xi}$ is bounded. We will need the following technical assumption

E1.(a) $\breve{\mu}_k^{\chi} > \breve{\mu}_{k+1}^{\chi}$, $k = 1, \ldots, r$.

(b) $\breve{\mu}_n^{\xi}$ is bounded away from zero as $n \to \infty$.

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

Lastly, let us introduce some new notation. (i) Set $\hat{\mathbf{K}}_h = \hat{\mathbf{\Gamma}}_{\boldsymbol{h}}^{\boldsymbol{\chi}}\ \hat{\mathbf{Z}}\ \left(\hat{\mathbf{Z}}\,\hat{\mathbf{\Gamma}}_0\,\hat{\mathbf{Z}}'\right)^{-1}\hat{\mathbf{Z}}$, so that $\hat{\boldsymbol{\chi}}_{T+h|T} = \hat{\mathbf{K}}_h \mathbf{x}_T$. (ii) Denote by $\hat{w}_{jt}$ the standardized version of $\hat{W}_{jt}$. Since

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

then $\hat{w}_{jt} = \hat{\mathbf{z}}_j \mathbf{x}_{tt}$, where $\hat{\mathbf{z}}_j = \hat{\mathbf{Z}}_j/\sqrt{1+\hat{\nu}_j}$. Note that since $\hat{\mathbf{Z}}_j\hat{\mathbf{\Gamma}}_0\hat{\mathbf{Z}}_k' = 0$ for $j \neq k$ (using the constraints of (3.7)), the linear combinations $\hat{w}_{jt}$, for $k = 1, 2, \ldots, r$, form an orthonormal system spanning a space of the same dimension as $\mathcal{G}(\mathbf{F}, t)$. (iii) Denote by $\check{\mathbf{Z}}_j$, $\check{\nu}_j$, $\check{\mathbf{K}}^h$, $\check{\chi}_{i,T+h|T}$, etc., the objects playing the same roles as $\hat{\mathbf{Z}}_j$, $\hat{\nu}_j$, $\hat{\mathbf{K}}_h$, $\hat{\chi}_{i,T+h|T}$, etc., but with respect to $\check{\mathbf{\Gamma}}_0^{\chi}$ and $\check{\mathbf{\Gamma}}_0^{\xi}$.

Proposition 4.1 *Suppose that Assumptions A, B, C, D and E hold for model (2.1). 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 $n \geq N_0$ and all $T \geq T_0$,*

$$\mathrm{P}\left[\left|\hat{\chi}_{i,T+h|T} - \chi_{i,T+h|T}\right| > \epsilon\right] \leq \eta. \tag{4.4}$$

The proof of Proposition 4.1 relies on the following two lemmas.

Lemma 4.1 *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 4.1,*

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

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 e.g. Lemma 3 of FHLR (2000).

Lemma 4.2 *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 zero 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 $\mathcal{K}$.*

For the proof, see the Appendix.

Proof of Proposition 4.1. Lemmas 7.1 and 7.2 (see 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 4.1. Indeed, $\check{\mathbf{Z}}_j$ is bounded in modulus, since $1 = \check{\mathbf{Z}}_j\,\check{\mathbf{\Gamma}}_0^{\xi}\,\check{\mathbf{Z}}_j' \geq \breve{\mu}_n^{\xi}\,\check{\mathbf{Z}}_j\,\check{\mathbf{Z}}_j'$, where, in view of Assumption E(b), $\breve{\mu}_n^{\xi}$ is bounded away from zero. Lemma 4.1 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 4.2, $\check{\chi}_{i,T+h|T}$ 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 $n > N_1$

$$\mathrm{P}\left[|\chi_{i,T+h|T} - \check{\chi}_{i,T+h|T}| > \epsilon\right] < \eta. \tag{4.5}$$

Convergence in probability of $\hat{\mathbf{\Gamma}}_0^{\chi}$ and $\hat{\mathbf{\Gamma}}_n^{\xi}$ to $\check{\mathbf{\Gamma}}_0^{\chi}$ and $\check{\mathbf{\Gamma}}_0^{\xi}$, respectively, 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 $\mathrm{P}\left[\max_{j=1,n} |x_{jt}| \geq M(n, \eta)\right] < \eta$. Then, given n, $\epsilon > 0$ and $\eta > 0$, there exists $T_2(n, \epsilon, \eta)$ such that, for $T > T_2$,

$$\mathrm{P}\left[\left|\check{\chi}_{i,T+h|T} - \hat{\chi}_{i,T+h|T}\right| > \epsilon\right] = \mathrm{P}\left[\left|\sum_{j=1}^{n} (\hat{K}_{h,ij} - \check{K}_{h,ij}) x_{jt}\right| > \epsilon\right] < \eta. \tag{4.6}$$

To see this, note that

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

so that (4.6) is obtained defining $T_2(n, \epsilon, \eta) = T_1(n, \epsilon/M(n, \eta/2), \eta/2)$. Lastly,

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

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 (4.5) and (4.6). QED

As the reader can easily check, the proof of Proposition 4.1 can be adapted with no difficulty to prove consistency of our two-step in-sample estimator (3.11). Moreover, Proposition 4.1 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, respectively, provided that $\hat{\mathbf{D}} - \check{\mathbf{D}}$ converges to zero in probability for any n, as $T \to \infty$, and that all eigenvalues of $\check{\mathbf{D}}$ are bounded and bounded away from zero as $n \to \infty$ (indeed Lemmas 7.1 and 7.2 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 which 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 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 which should favor the dynamic method. M2 has MA(3) loading filters, two serially uncorrelated factors and diagonal idiosyncratic variance-covariance matrix. M3 and M4 have one autoregressive factor and the common components of different groups are shifted

in time. The two models differ by the behaviours of their idiosyncratic components: in M3 they have different variances whereas in M4 variances are the same. The comparison between the last two models should help understanding the role of heterogeneity of the size of the idiosyncratic components. Finally, in M3 the idiosyncratic components are not mutually orthogonal (but condition C2 is still satisfied).

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

$$x_{it}^* = \lambda_i f_t + \alpha c_i \epsilon_{it}, \quad \text{with} \quad (1 - 0.5\ L) f_t = u_t, \tag{M1}$$

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, while the coefficients c_i are mutually independent, independent of the latter variables, and uniformly distributed on the interval $[0.1, 1.1]$ in order 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 one (the same holds for all models below). Here $q = 1$ and $s = 1$.

MODEL M2. Observations are generated by

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

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 while the c_i's are uniformly distributed over the interval [0.1 1.1], as for (M1). Here $q = 2$ and $s = 3$.

MODEL M3. The observations are generated by

$$x_{it}^* = \sum_{k=l_i}^{l_i+2} \lambda_{k-l_i,i} f_{t-k} + \xi_{it}^*, \quad \text{with} \quad (1 - aL) f_t = u_t \quad \text{and} \quad \xi_{it}^* = \alpha c_i (\epsilon_{it} + \epsilon_{i+1,t}), \tag{M3}$$

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 $2m + 1 \leq i \leq n$. In order for the three types ($l_i = 0, 1, 2$) to be equally present in the panel, we took $m = [n/3]$ (as usual, we denote by $[z]$ 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 $\text{var}(\lambda_i f_t)/\text{var}(x_{it}^*) = 0.5$; 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 $T = 20, 50, 100, 200$, meaning a total number of 64 experiments; each experiment was replicated 1000 times. Before estimation, all variables were taken in deviation from their sample means and divided by their standard deviations, i.e. spectral estimation was conducted 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 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 $M = [T^{1/2}]$. The spectra were evaluated 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 component decomposition, as explained in Section 3. In order to obtain $\hat{\mathbf{\Gamma}}_k^{\chi}$ and $\hat{\mathbf{\Gamma}}_k^{\xi}$ we used the inverse discrete Fourier transforms

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

with $\hat{\mathbf{\Sigma}}^{\xi}(\theta) = \hat{\mathbf{\Sigma}}(\theta) - \hat{\mathbf{\Sigma}}^{\chi}(\theta)$—except for $\hat{\mathbf{\Gamma}}_0^{\xi}$, as expained below. Throughout, we assumed both the number q of dynamic factors and the number $r = q(s+1)$ of static factors to be 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 to zero the off-diagonal entries of the estimated variance-covariance matrix $\hat{\mathbf{\Gamma}}_0^{\xi}$ of the idiosyncratic components significantly improves forecasting performances, even when the actual matrix is non diagonal. Our explanation for this somewhat counterintuitive result is the following. When computing $\hat{\mathbf{\Gamma}}_0^{\xi}$, we unavoidably get some spurious large covariances, even when the true covariance is zero. 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 $\hat{\mathbf{\Gamma}}_0^{\xi}$. On the other hand, by forcing to zero the off-diagonal entries of our estimated matrix, we ignore the true off-diagonal non-zero entries. Also in this case the error increases with n, but, owing to the boundedness of the eigenvalues, it increases only linearly in n. As mentioned at the end of Section 4, replacing $\hat{\mathbf{\Gamma}}_0^{\xi}$ with any symmetric positive semi-definite matrix with bounded eigenvalues does not affect consistency results. Therefore, we henceforth always set to zero the off-diagonal entries of $\hat{\mathbf{\Gamma}}_0^{\xi}$ before computing eigenvectors.

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

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

respectively. Results for models M1, M2, M3 and M4 are shown in Tables 5.1, 5.2, 5.3 and 5.4 respectively, with part (a) devoted to forecasts and part (b) devoted to within-sample estimation. We report the average value of the criterion, along with its empirical standard deviation (in brackets) across the 1000 replications, both for SW's and for the two-step method.

Table 5.1a: Model M1, forecasting results

	$n = 20$		$n = 50$		$n = 100$		$n = 200$	
	two-step	*SW*	*two-step*	*SW*	*two-step*	*SW*	*two-step*	*SW*
$T = 20$	0.9462 (1.2356)	0.9492 (1.1961)	0.9287 (1.2035)	0.9325 (1.1598)	0.9292 (1.2030)	0.9323 (1.1519)	0.9288 (1.1971)	0.9321 (1.1470)
$T = 50$	0.8642 (1.1493)	0.8606 (1.1430)	0.8584 (1.1472)	0.8529 (1.1288)	0.8555 (1.1451)	0.8486 (1.1190)	0.8553 (1.1439)	0.8488 (1.1174)
$T = 100$	0.7995 (1.0449)	0.8018 (1.0425)	0.7869 (1.034)	0.7881 (1.0294)	0.7864 (1.0321)	0.7864 (1.0288)	0.7851 (1.0326)	0.7818 (1.0256)
$T = 200$	0.7833 (1.0723)	0.7785 (1.0698)	0.7753 (1.0584)	0.7770 (1.0486)	0.7731 (1.0576)	0.7704 (1.0541)	0.7721 (1.0577)	0.7700 (1.0554)

Table 5.1b: 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*
$T = 20$	0.1463	0.1759	0.1070	0.1139	0.0969	0.0969	0.0924	0.0888
	(0.1030)	(0.1447)	(0.0551)	(0.0665)	(0.0447)	(0.0508)	(0.0403)	(0.0427)
$T = 50$	0.0631	0.0878	0.0408	0.0506	0.0354	0.0399	0.0328	0.0345
	(0.0351)	0.0471	(0.0156)	0.0190	(0.0111)	0.0126	(0.0091)	0.0099
$T = 100$	0.0413	0.0649	0.0225	0.0327	0.0182	0.0233	0.0161	0.0186
	(0.0232)	(0.0299)	(0.0075)	(0.0097)	(0.0046)	(0.0057)	(0.0034)	(0.0039)
$T = 200$	0.0313	0.0544	0.0143	0.0244	0.0103	0.0155	0.0085	0.0111
	(0.0181)	(0.0225)	(0.0044)	(0.0058)	(0.0022)	(0.0028)	(0.0014)	(0.0016)

Table 5.2a: Model M2, forecasting results

	$n = 20$		$n = 50$		$n = 100$		$n = 200$	
	two-step	*SW*	*two-step*	*SW*	*two-step*	*SW*	*two-step*	*SW*
$T = 20$	0.8901	0.9757	0.7773	0.8552	0.7227	0.7763	0.6911	0.7349
	(0.5266)	(0.5423)	(0.4790)	(0.4851)	(0.4276)	(0.4422)	(0.4132)	(0.4255)
$T = 50$	0.6514	0.7446	0.5025	0.5650	0.4613	0.4911	0.4412	0.4577
	(0.4360)	(0.4793)	(0.3446)	(0.3692)	(0.3165)	(0.3251)	(0.3078)	(0.3130)
$T = 100$	0.5385	0.6332	0.3944	0.4427	0.3552	0.3775	0.3402	0.3509
	(0.3844)	(0.4284)	(0.2895)	(0.3015)	(0.2692)	(0.2736)	(0.2645)	(0.2689)
$T = 200$	0.4949	0.5832	0.3660	0.4076	0.3278	0.3487	0.3127	0.3223
	(0.3367)	(0.3702)	(0.2852)	(0.2896)	(0.2694)	(0.2733)	(0.2673)	(0.2693)

Table 5.2b: 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*
$T = 20$	0.4587	0.7163	0.3683	0.5476	0.3290	0.4536	0.3057	0.3933
	(0.1523)	(0.2457)	(0.1049)	(0.1705)	(0.0836)	(0.1340)	(0.0708)	(0.1075)
$T = 50$	0.2838	0.5632	0.1827	0.3106	0.1496	0.2109	0.1340	0.1672
	(0.0751)	(0.1536)	(0.0347)	(0.0759)	(0.0235)	(0.0430)	(0.0180)	(0.0284)
$T = 100$	0.2154	0.4861	0.1238	0.2110	0.0931	0.1303	0.0788	0.0953
	(0.0525)	(0.1234)	(0.0207)	(0.0435)	(0.0120)	(0.0205)	(0.0081)	(0.0122)
$T = 200$	0.1842	0.4336	0.0921	0.1613	0.0613	0.0909	0.0471	0.0598
	(0.0448)	(0.0998)	(0.0157)	(0.0279)	(0.0076)	(0.0117)	(0.0042)	(0.0059)

Table 5.3a: Model M3, forecasting results

	$n = 20$		$n = 50$		$n = 100$		$n = 200$	
	two-step	*SW*	*two-step*	*SW*	*two-step*	*SW*	*two-step*	*SW*
$T = 20$	0.7644	0.8645	0.6680	0.7748	0.5713	0.6383	0.5292	0.5499
	(0.6472)	(0.6793)	(0.5975)	(0.6102)	(0.4619)	(0.4517)	(0.3919)	(0.3522)
$T = 50$	0.5273	0.5913	0.4636	0.4962	0.3565	0.3773	0.2792	0.2771
	(0.5886)	(0.5993)	(0.5620)	(0.5393)	(0.3770)	(0.3601)	(0.2355)	(0.2139)
$T = 100$	0.4482	0.4966	0.3935	0.4094	0.2873	0.2957	0.1958	0.1943
	(0.5312)	(0.5367)	(0.5136)	(0.4806)	(0.3470)	(0.3210)	(0.1758)	(0.1678)
$T = 200$	0.4131	0.4690	0.3493	0.3719	0.2488	0.2562	0.1521	0.1563
	(0.4779)	(0.5099)	(0.4213)	(0.4275)	(0.2826)	(0.2793)	(0.1440)	(0.1435)

Table 5.3b: 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*
$T = 20$	0.4143	0.9583	0.3104	0.6670	0.2970	0.5440	0.3025	0.4505
	(0.2145)	(0.4796)	(0.1326)	(0.2913)	(0.1096)	(0.2299)	(0.0985)	(0.1893)
$T = 50$	0.2285	0.7279	0.1444	0.4083	0.1335	0.2753	0.1326	0.1856
	(0.0900)	(0.2752)	(0.0433)	(0.1223)	(0.0317)	(0.0791)	(0.0268)	(0.0525)
$T = 100$	0.1682	0.6468	0.0921	0.3176	0.0831	0.1839	0.0809	0.1044
	(0.0613)	(0.2184)	(0.0234)	(0.0758)	(0.0157)	(0.0415)	(0.0119)	(0.0218)
$T = 200$	0.1386	0.5959	0.0662	0.2646	0.0554	0.1344	0.0501	0.0647
	(0.0469)	(0.1748)	(0.0151)	(0.0473)	(0.0092)	(0.0222)	(0.0059)	(0.0096)

Table 5.4a: Model M4, forecasting results

	$n = 20$		$n = 50$		$n = 100$		$n = 200$	
	two-step	*SW*	*two-step*	*SW*	*two-step*	*SW*	*two-step*	*SW*
$T = 20$	0.7649	0.8531	0.6885	0.7820	0.5855	0.6424	0.5373	0.5395
	(0.6897)	(0.7121)	(0.6447)	(0.6610)	(0.5072)	(0.4965)	(0.4042)	(0.3727)
$T = 50$	0.5288	0.5692	0.4577	0.4886	0.3632	0.3759	0.2933	0.2746
	(0.5340)	(0.5195)	(0.4869)	(0.4724)	(0.3427)	(0.3343)	(0.2286)	(0.1988)
$T = 100$	0.4745	0.4919	0.4192	0.4184	0.3061	0.3025	0.2027	0.1908
	(0.5263)	(0.5242)	(0.5171)	(0.4944)	(0.3459)	(0.3374)	(0.1810)	(0.1719)
$T = 200$	0.4207	0.4390	0.3575	0.3595	0.2534	0.2496	0.1567	0.1509
	(0.4575)	(0.4587)	(0.4391)	(0.4303)	(0.2941)	(0.2826)	(0.1463)	(0.1437)

Table 5.4b: 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*
$T = 20$	0.3883	0.8443	0.3011	0.6068	0.2881	0.4920	0.2937	0.4084
	(0.1775)	(0.3858)	(0.1297)	(0.2761)	(0.1100)	(0.2223)	(0.1038)	(0.1869)
$T = 50$	0.2284	0.6099	0.1487	0.3566	0.1359	0.2404	0.1324	0.1668
	(0.0641)	(0.1763)	(0.0393)	(0.0986)	(0.0317)	(0.0678)	(0.0283)	(0.0476)
$T = 100$	0.1799	0.5217	0.1024	0.2685	0.0896	0.1586	0.0839	0.0949
	(0.0371)	(0.1059)	(0.0189)	(0.0510)	(0.0148)	(0.0309)	(0.0122)	(0.0187)
$T = 200$	0.1594	0.4734	0.0811	0.2216	0.0656	0.1172	0.0561	0.0611
	(0.0262)	(0.0690)	(0.0111)	(0.0283)	(0.0080)	(0.0150)	(0.0059)	(0.0079)

The results can be summarized as follows.

(1) For model M1 the two competing methods yield similar performances 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 for models with heterogeneous dynamics, i.e. M2, M3 and M4; in-sample estimation performances are considerably better.

(3) Homogeneity in the common/idiosyncratic variance ratio (M4 versus M3) somewhat reduces the advantage of the two-step method over SW's, an advantage which 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) carry out an exhaustive comparison of the forecasting performances of our two-step method, SW's, and also some more standard ones. The exercise is based on the monthly series in Stock and Watson (2002b) dataset, which includes 150 US macroeconomic series, from 1959:1 to 1999:2 (so that $T = 482$). Such data are used to forecast two key macroeconomic variables: industrial production (IP, in log levels) and inflation (twelve-month log change of Consumer Price Index, CPI), by means of a simulated real-time forecasting exercise, i.e. 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 $T-h$; h is the forecast horizon.

In Table 5.5 we report some of their results. Like in simulations, we consider only prediction of the common components, as though the idiosyncratic components were white noise. Since the actual number r of static factors in real data obviously is unknown, we report results for r running from 6 to 15. The number of dynamic factors is set equal to 2. The forecast horizons are h =12 and 24 months, and the measure of performance used is (writing y_t for the variable to be predicted and $\hat{y}_{t+h|t}$ for the corresponding h-step ahead forecast)

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

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, two-step performs moderately better in all experiments. Both SW's and our method perform considerably better than AR based forecasts (the corresponding figures are not reported here, but can be found in D'Agostino and Giannone 2004).

The same result, two-step outperforming SW's method, is found in Gentile (2004) for Italian data. Prediction of aggregate inflation and industrial production there is obtained by applying the factor model to the panel of elementary price and production indexes respectively. The techniques introduced in Boivin and Ng (2004) to select the "good" variables in the panel are shown to produce considerable improvement.

Table 5.5: Mean square forecast errors for two-step and SW

	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

6 Summary and conclusions

This paper proposes 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 Forni, Hallin, Lippi, and Reichlin (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 Forni, Hallin, Lippi and Reichlin (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 show that the projection converges to the optimal forecast as n and T go to infinity. Being a linear combination of the x's which 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 Forni, Hallin, Lippi and Reichlin (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 component 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.

7 Appendix

We give here the proofs of Lemmas 7.1 and 7.2, used in the proof of Proposition 4.1, and of Lemma 4.2.

Lemma 7.1 *Given the integer $k > 0$, consider a sequence of real, symmetric, positive semi-definite $n \times n$ matrices $\mathbf{\Gamma}_n$ and a sequence of real, symmetric, positive definite $n \times n$ matrices $\mathbf{D}_n$, $n = k, k+1, \dots$ Assume that*

(i) $\mathbf{\Gamma}_n$'s k-th largest eigenvalue μ_{nk} diverges as $n \to \infty$, and

(ii) $\mathbf{D}_n$'s largest eigenvalue is bounded from above by δ.

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

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

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

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,

$$\mathbf{q}_n \mathbf{D}_n \mathbf{v}'_{nj} = 0, \quad j = 1, \dots, k-1, \quad \text{and} \quad \mathbf{q}_n \mathbf{D}_n \mathbf{q}'_n = 1 \tag{7.1}$$

(for $k = 1$, (7.1) 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 (ii) and (7.1) that $\alpha_{n1}^2 + \alpha_{n2}^2 + \cdots + \alpha_{nk}^2 \geq \frac{1}{\delta}$. 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 (7.1), $\nu_{nk} \geq \mathbf{q}_n \mathbf{\Gamma}_n \mathbf{q}'_n$. The conclusion follows. QED

Lemma 7.2 *Let $\check{\mathbf{\Gamma}}_0^{\chi}$ and $\check{\mathbf{\Gamma}}_0^{\xi}$ be as in (4.3), $\check{\mu}_k^{\chi}$ and $\check{\mu}_k^{\xi}$ being, respectively, their eigenvalues. Let $r = q(s+1)$. Then, under Assumptions A, B, C and D,*

(i) $\check{\mu}_r^{\chi} \to \infty$ as $n \to \infty$;

(ii) $\check{\mu}_1^{\xi}$ is bounded for $n \to \infty$.

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

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

We have (see Lancaster and Tismenetsky, 1985, p. 301, Theorem 1) $\lambda_{q+1}(\theta) \leq \lambda^{\chi}_{q+1}(\theta) + \lambda^{\xi}_1(\theta)$. Thus, since $\lambda^{\chi}_{q+1}(\theta) = 0$, Assumption C2 implies $\alpha \leq 2\pi\Lambda$. Part (ii) of the lemma follows. Obviously, C2 implies that $\mathbf{v}\mathbf{\Gamma}^{\xi}_0\mathbf{v}' \leq 2\pi\Lambda$. Setting $\mathbf{A} = \mathbf{\Gamma}^{\xi}_0 - \check{\mathbf{\Gamma}}^{\xi}_0$ and observing that $\mathbf{v}\mathbf{\Gamma}^{\xi}_0\mathbf{v}'$ and $\mathbf{v}\check{\mathbf{\Gamma}}^{\xi}_0\mathbf{v}'$ are non-negative, we obtain $|\mathbf{v}\mathbf{A}\mathbf{v}'| = |\mathbf{v}\mathbf{\Gamma}^{\xi}_0\mathbf{v}' - \mathbf{v}\check{\mathbf{\Gamma}}^{\xi}_0\mathbf{v}'| \leq 2\pi\Lambda$. Because

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

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

Proof of Lemma 4.2. Let $\mathbf{v} = (v_1 \ \cdots \ v_k)'$ 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

$$\mathbf{v}_n = \mathbf{a}_n\mathbf{v} + \mathbf{R}_n, \tag{7.2}$$

of $\mathbf{v}_n$ into its (componentwise) orthogonal projection $\mathbf{a}_n\mathbf{v}$ onto $\mathcal{K}$ and the orthogonal complement ($\mathbf{a}_n$ is a $k \times k$ matrix). By assumption, $\mathbf{R}_n \to 0$ in quadratic mean, while 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

$$\mathbf{v} = \mathbf{a}_n'\mathbf{v}_n + \mathbf{S}_n. \tag{7.3}$$

Taking covariances in (7.2) and (7.3), we obtain, in view of the orthonormality assumption on $\mathbf{v}$ and $\mathbf{v}_n$, that $\mathbf{I}_k = \mathbf{a}_n\mathbf{a}_n' + \mathbf{\Gamma}^R_n$, and $\mathbf{I}_k = \mathbf{a}_n'\mathbf{a}_n + \mathbf{\Gamma}^S_n$, so that $\mathbf{a}_n\mathbf{a}_n' + \mathbf{\Gamma}^R_n = \mathbf{a}_n'\mathbf{a}_n + \mathbf{\Gamma}^S_n$. Taking traces on both sides yields $\text{tr}\left(\mathbf{\Gamma}^R_n\right) = \text{tr}\left(\mathbf{\Gamma}^S_n\right)$. 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 $\mathcal{K}$, respectively ($\mathbf{b}_n$ 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

$$[(\mathbf{b}_n\mathbf{a}_n - \mathbf{b})\mathbf{v} - (s - s_n)] \longrightarrow 0 \tag{7.4}$$

in quadratic mean. Now, (7.3) and the fact that $\mathbf{S}_n \to 0$ imply that $\text{cov}(\mathbf{v}, s_n) \to 0$. Since $\mathbf{a}_n$ 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 (7.4), implies that $\lim_{n\to\infty}(\mathbf{b}_n\mathbf{a}_n - \mathbf{b})\mathbf{v} = \lim_{n\to\infty}(s - s_n) = 0$. QED

References

[1] Anderson, T. W. (1984). *An Introduction to Multivariate Statistical Analysis.* John Wiley & Sons: New York.

[2] Bai, J. (2003). Inferential theory for factor models of large dimension. *Econometrica* **71**, 135-71 .

[3] Bai, J. and S. Ng (2002). Determining the number of factors in approximate factor models. *Econometrica* **70**, 191-221.

[4] Boivin, J. and S. Ng (2004). Are more data always better for factor analysis? *Journal of Econometrics*, forthcoming.

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

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

[7] Chamberlain, G. (1983). Funds, factors, and diversification in arbitrage pricing models. *Econometrica* **51**, 1281-1304.

[8] Chamberlain, G. and M. Rothschild (1983). Arbitrage, factor structure and mean-variance analysis in large asset markets. *Econometrica* **51**, 1305-1324.

[9] D'Agostino, A. and D. Giannone (2004). Comparing alternative predictors based on large-panel dynamic factor models, ISRO Discussion Paper, Université libre de Bruxelles, Brussels.

[10] Doz, C., D. Giannone and L. Reichlin (2003). A maximum likelihood approach to dynamic factor analysis in large panels. Mimeo, ECARES, Université libre de Bruxelles, Brussels.

[11] Forni, M., M. Hallin, M. Lippi, and L. Reichlin (2000). The generalized factor model: identification and estimation. *The Review of Economics and Statistics* **82**, 540-554.

[12] Forni, M., M. Hallin, M. Lippi, and L. Reichlin (2004). The generalized dynamic factor model: consistency and rates. *Journal of Econometrics* **119**, 231-255.

[13] Forni, M. and M. Lippi (2001). The generalized dynamic factor model: representation theory. *Econometric Theory* **17**, 1113-1141.

[14] Forni, M. and L. Reichlin (1998). Let's get real: a factor analytical approach to disaggregated business cycle dynamics. *Review of Economic Studies* **65**, 453-473.

[15] Gentile, M. (2004). *Forecasting industrial production index, inflation and asset return volatility by generalized dynamic factor model*. Doctoral Dissertation, Sant'Anna School of Advanced Studies, Pisa.

[16] Geweke, J. (1977). The dynamic factor analysis of economic time series. In D.J. Aigner and A.S. Goldberger, Eds., *Latent Variables in Socio-Economic Models*, North Holland: Amsterdam.

[17] Lancaster, P. and M. Tismenetsky (1985). *The Theory of Matrices*, 2nd Edition, Academic Press: Orlando.

[18] Quah, D. and T.J. Sargent (1993). A dynamic index model for large cross sections. In J.H. Stock and M.W. Watson, Eds, *Business Cycles, Indicators, and Forecasting*, N.B.E.R. and University of Chicago Press: Chicago.

[19] Lawley, D.N. and A.E. Maxwell (1971). *Factor analysis as a statistical method.* Butterworths: London.

[20] Rozanov, Yu.A. (1967). *Stationary Random Processes.* Holden Day: San Francisco.

[21] Sargent, T.J. and C.A. Sims (1977). Business cycle modeling without pretending to have too much *a priori* economic theory. In C.A. Sims, Ed., *New Methods in Business Research*, Federal Reserve Bank of Minneapolis: Minneapolis.

[22] Stock, J.H. and M.W. Watson (2002a). Forecasting using principal components from a large number of predictors, *Journal of the American Statistical Association* **97**, 1167-79.

[23] Stock, J.H. and M.W. Watson (2002b). Macroeconomic forecasting using diffusion indexes. *Journal of Business and Economic Statistics* **20**, 147-162.

[24] Wiener, N. and P. Masani (1957, 1958). The prediction theory of multivariate stochastic processes I, *Acta Mathematica* **98**, 111-150 (1957) and II, *Acta Mathematica* **99**, 93-137 (1958).