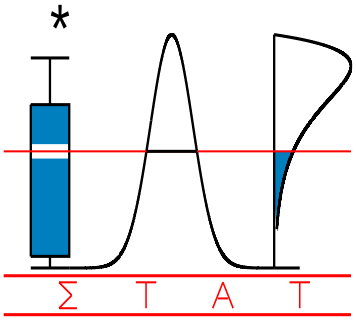
T E C H N I C A L

R E P O R T

0205

THE GENERALIZED DYNAMIC FACTOR MODEL: ONE-SIDED ESTIMATION AND FORECASTING

M. FORNI, M. HALLIN, M. LIPPI and L. REICHLIN



I A P S T A T I S T I C S

N E T W O R K

INTERUNIVERSITY ATTRACTION POLE

http://www.stat.ucl.ac.be/IAP

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

Mario Forni,
Universitá di Modena and CEPR

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

Marco Lippi
Universitá di Roma La Sapienza

and

Lucrezia Reichlin
ECARES, Université Libre de Bruxelles and CEPR

June 6, 2002

Abstract

This paper proposes a new forecasting method that exploits information from a large panel of time series. The method is based on the generalized dynamic factor model proposed in Forni, Hallin, Lippi, and Reichlin (2000), and takes advantage of the information on the dynamic covariance structure of the whole panel. We first use our previous method to obtain an estimation for the covariance matrices of common and idiosyncratic components. The generalized eigenvectors of this couple of matrices are then used to derive a consistent estimate of the optimal forecast, which is constructed as a linear combination of present and past observations only (one-sided filter). This two-step approach solves the end-of-sample problems caused by two-sided filtering (as in our previous work), while retaining the advantages of an estimator based on dynamic information. Both simulation results and an empirical illustration on the forecast of the Euro area industrial production and inflation, based on a panel of 447 monthly time series show very encouraging results.

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 huge numbers of observed time series – aggregated and disaggregated, real and nominal variables. In-

*Research supported by a P.A.I of the Belgian Federal Government, an A.R.C. contract of the Communauté française de Belgique, and the Training and Mobility of Researchers Programme of the European Commission (Contract ERBFMRX-CT98-0213).

tuition suggests that disregarding potentially useful information 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 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. The challenge for econometricians is to develop alternative techniques that are sufficiently powerful as to overcome this dimensionality problem, yet flexible enough to provide an adequate picture of economic reality.

Recently, Forni and Reichlin (1998), Stock and Watson (1999), and Forni, Hallin, Lippi, and Reichlin (2000, 2001) have developed factor model techniques that are tailored to exploit a large cross-sectional dimension. Under such models, each time series in the panel is represented as the sum of two mutually orthogonal components : the *common* component, which is "strongly correlated" with the rest of the panel and has reduced stochastic dimension, and the *idiosyncratic* component. These idiosyncratic components are either mutually orthogonal or "mildly cross-correlated" across the panel. The common component is (non-parametrically) consistently estimated as both the size n of the cross-section and the series length T go to infinity. These results are obtained under conditions that look reasonable in empirical situations, whenever there are comovements between time series, as it is generally the case for macroeconomic data (for a documentation of this point, see Altissimo et al. 2001).

In a factor model, multivariate information can help forecasting the common component, while the idiosyncratic, being mildly cross-correlated, can be reasonably well predicted by means of traditional univariate methods (or methods based on low dimension models such as VARs). Therefore, the forecast of the future of any given series in the panel can be obtained as the sum of the forecast of the common component, where we exploit multivariate information, and the forecast of the idiosyncratic component, where multivariate information safely can be disregarded. The common component being of reduced stochastic dimension, its forecast can be expressed as a projection on the span of a small number of appropriately constructed aggregates.

The two methods proposed in the literature, Stock and Watson (1999) and Forni et al. (2001) are both based on this general idea, but each of them presents a serious shortcoming. Forni et al. (2000) base their estimation of the common and idiosyncratic components on the dynamic covariance structure of the data. This is a highly desirable feature, since economic time series in general are non-synchronized, and the leading variables should play a crucial role in the forecast of the lagging ones. Typically, provided that leading series are included in the panel, such methods should allow for forecasting even those lagging series that are unforecastable at univariate level. The Forni et al. (2000) estimator, however, is derived from the spectral density of the data and, as a consequence, is based on a two-sided filtering of the observations (linear combination of present, past and future observations). This two-sidedness feature is not a problem when within-sample estimation of the common component is the objective; but it is most unpleasant in the forecasting context, since at the end of the sample future observations are not available.

Stock and Watson (1999), on the other hand, propose to estimate the common component by projecting onto the static principal components of the data. Their estimator relies on a one-sided filtering of the observations, which does not cause any problems at the end of the sample. However, being based on contemporaneous covariances only, it fails to exploit the potentially crucial information contained in the leading-lagging relations between the elements of the panel.

The method we propose in this paper aims at combining the advantages of both approaches. Firstly, using the dynamic techniques developed in Forni et al. (2000), we obtain estimates

of common and idiosyncratic cross-covariance matrices at all leads and lags as inverse Fourier transforms of the corresponding estimated spectral density matrices. Secondly, we use these estimates in the construction of the contemporaneous linear combinations of the observations having smallest idiosyncratic-common variance ratio. The resulting aggregates can be obtained as the solution of a *generalized principal component* problem (see Section 6). Our h-step ahead forecast is obtained as the projection of the h step ahead observation onto these estimated generalized principal components.

We prove, under finite-order VARMA structure, that this forecast is a consistent estimator (for n and T going to infinity) of the optimal h-step ahead forecast. The same method can be used to reestimate the within sample common component, thus improving the accuracy of the estimator based on the first step. These projections do not involve future observations and hence do not suffer the end-of-sample problems of the Forni et al. (2000) method.

Both Stock and Watson's estimators and ours are linear combinations of present and past observations, but the weighting schemes used are completely different. Theoretically, we both provide a consistent forecast. Empirical relative performance, however, are difficult to establish *a priori*. Indeed, though our weighting scheme, being tailored with the purpose of minimizing the impact of the idiosyncratic in the aggregate, should perform better in approximating the common factor space, relative performance depends in a very complicated way on the underlying model, on T, and on n. Our simulation study confirms the conjecture that our method, at least for the models studied, performs better than Stock and Watson's. The empirical illustration from a simulated out-of-sample forecasting experiment on a panel of monthly time series for the Euro area confirms this result. The same illustration also shows that our method outperforms the univariate autoregressive model.

The paper is organized as follows. Section 2 defines the framework, establishes notation, and provides a brief summary of previous work. Section 3 illustrates the idea of our two-step procedure, and Section 4 provides consistency results under the assumption of a finite-order VARMA structure. Section 5 reports the results of a simulation study, and describes an empirical illustration. Technical proofs are concentrated in Section 6.

2 Factor models in the analysis of (large) panels of time series data : a brief review.

2.1 Panels of time series data.

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

(F0a) $\mathbf{X}_n^T$ is a finite realization of a real-valued stochastic process $\mathbf{X} := \{x_{it}, i \in \mathbb{N}, t \in \mathbb{Z}\}$ indexed by $\mathbb{N} \times \mathbb{Z}$, where the n-dimensional vector processes $\{\mathbf{x}_{nt} := (x_{1t} \ \cdots \ x_{nt})', \ t \in \mathbb{Z}\}$, $n \in \mathbb{N}$ are stationary, with mean $\mathbf{0}_n$ and finite second-order moments $\mathbf{\Gamma}_{nk} := \mathrm{E}[\mathbf{x}_{nt}\mathbf{x}'_{n,t-k}]$, $k \in \mathbb{N}$.

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

(F0b) For all $n \in \mathbb{N}$, the process $\{\mathbf{x}_{nt}, t \in \mathbb{Z}\}$ admits a Wold representation $\mathbf{x}_{nt} = \sum_{k=0}^{\infty} \mathbf{C}_k^n \mathbf{w}_{n,t-k}$, where the full-rank innovations $\mathbf{w}_{nt}$ have finite moments of order four, and the matrices $\mathbf{C}_k^n = (C_{ij,k}^n)$ satisfy $\sum_{k=0}^{\infty} |C_{ij,k}^n| k^{1/2} < \infty$ for all $n, i, j \in \mathbb{N}$.

Assumptions (F0a) and (F0b) jointly will be referred to as (F0).

It is important to point out that the roles played by the cross-sectional and the time dimensions in panel data are strongly asymmetric. A strong order in the *time index* $t \in \mathbb{Z}$ is implicitly imposed, allowing for meaningful measures of serial dependence at various lags, for the stationarity assumption (F0), for the definition of spectral densities, etc. By contrast, the *cross-sectional index* i does not possess much structure : no past, no future, no meaningful ordering. As a rule, all assumptions involving the finite-dimensional processes $\{\mathbf{x}_{nt}, t \in \mathbb{Z}\}$ should remain unaffected by a permutation of the cross-sectional indexes $i = 1, \ldots, n$. This precludes, in particular, any concept of *cross-sectional stationarity* or *cross-sectional mixing* : for example, there is no reason for the correlation between x_{1t} and x_{2t} being the same as the correlation between x_{3t} and x_{4t}, or to be less than the correlation between x_{1t} and $x_{987,t}$. Nevertheless, some regularity in the cross-sectional dimension is needed, providing the red thread connecting all these series together and ensuring that the panel is not just an artificial collection of mutually unrelated observations. This regularity feature takes the form of an assumption on the eigenvalues of the spectral density matrices of the vectors $\mathbf{x}_t^n$ as $n \to \infty$ (see Section 2.2 below), and remains invariant, as the reader can easily check, under any permutation of the n cross-sectional items in the panel. As a rule, in the various factor models we are now describing, both the assumptions and the results are independent of the particular cross-sectional indexing.

2.2 Static, dynamic, and generalized dynamic factor models.

Traditional factor models, as well as traditional vector time-series models, deal with situations in which T is large as compared to n; the classical asymptotic theory, which considers n as fixed, with T tending to infinity, can be regarded as relevant. The situation is quite different in the context of panel data, where n typically is large or even very large compared to T. The challenge then is twofold. First, an alternative is to be found to traditional time series methods, which cannot handle the number of parameters implied by such values of n. Second, a relevant asymptotic theory should address the case where both n and T go to infinity. The dynamic factor models considered in Forni et al. (2000) and in the present paper allow for handling large values of n, and quite naturally enter this nonstandard asymptotic framework under which both n and T tend to infinity.

Standard factor models decompose a given finite number n of observed processes $\{x_{it}\}$ into two non-observable components, a *common* component $\{\chi_{it}\}$ driven by q common *factors* $\{F_{jt}\}$, $j = 1, \ldots, q$, where q is very small as compared to n, and an *idiosyncratic* component $\{\xi_{it}\}$. In the simplest case the factors are *loaded* only contemporaneously, yielding the *static factor model*, of the form

$$x_{it} = \chi_{it} + \xi_{it} = a_{i1}F_{1t} + a_{i2}F_{2t} + \cdots + a_{iq}F_{qt} + \xi_{it} = \mathbf{a}_i\mathbf{F}_t + \xi_{it}, \tag{2.1}$$

where $\mathbf{a}_i := (a_{i1}, \ldots, a_{iq})$ and $\mathbf{F}_t := (F_{1t}, \ldots, F_{qt})'$; the usual assumptions are

(F1a) the processes $\{F_{jt},\ j = 1, \ldots, q,\ t \in \mathbb{Z}\}$ and $\{\xi_{it},\ i = 1, \ldots, n,\ t \in \mathbb{Z}\}$ are mutually orthogonal (hence also the processes $\{\chi_{it},\ i = 1, \ldots, n,\ t \in \mathbb{Z}\}$ and $\{\xi_{it},\ i = 1, \ldots, n,\ t \in \mathbb{Z}\}$);

(F1b) the idiosyncratic processes $\{\xi_{i't},\ t \in \mathbb{Z}\}$ and $\{\xi_{i''t},\ t \in \mathbb{Z}\}$ are mutually orthogonal for all $i' \neq i''$.

Under Assumptions (F1a) and (F1b), the static model (2.1) can be identified (up to rotations of the factors) and estimated. This model has been extended by Sargent and Sims (1977) and

Geweke (1977) to allow for a *dynamic* loading of the factors. More precisely, the *dynamic factor model* is characterized by a decomposition of $\{x_{it}\}$ of the form

$$x_{it} = \chi_{it} + \xi_{it} = b_{i1}(L)u_{1t} + b_{i2}(L)u_{2t} + \cdots + b_{iq}(L)u_{qt} + \xi_{it} = \mathbf{b}_i(L)\mathbf{u}_t + \xi_{it}, \tag{2.2}$$

where $\{\mathbf{u}_t := (u_{1t} \ \dots \ u_{qt})', \ t \in \mathbb{Z}\}$, the vector of *common shocks*, is a q-dimensional (q very small with respect to n) non-observed orthonormal white noise process, and $b_{ij}(L) := \sum_{k=0}^{\infty} b_{ij,k}L^k$, $i = 1, \dots, n$, $j = 1, \dots, q$ are nq square-summable filters. The assumptions on $\{\mathbf{u}_t, t \in \mathbb{Z}\}$ and the idiosyncratic components $\{\xi_{it}\}$, $i = 1, \dots, n$ are quite similar to (F1a) and (F1b) :

(F2a) the processes $\{\mathbf{u}_t, \ t \in \mathbb{Z}\}$ and $\{\xi_{it}, \ i = 1, \dots, n, \ t \in \mathbb{Z}\}$ are mutually orthogonal (hence, also $\{\chi_{it}, \ i = 1, \dots, n, \ t \in \mathbb{Z}\}$ and $\{\xi_{it}, \ i = 1, \dots, n, \ t \in \mathbb{Z}\}$);

(F2b) same as (F1b).

In most applications, the assumption (F1b)=(F2b) that the idiosyncratic processes associated with distinct cross-sectional items are mutually orthogonal is much too restrictive. On the other hand, dropping this orthogonality assumptions destroys the identifiability of models (2.1) and (2.2) for any given finite n. Chamberlain and Rothschild (1983) have introduced a more general definition of idiosyncratic components under the assumption of an infinite cross-sectional dimension for a static model of the form (2.1), and Forni et al. (2000) have provided the following generalization to the dynamic case. Considering the spectral density matrices $\mathbf{\Sigma}_n^{\chi}(\theta)$ and $\mathbf{\Sigma}_n^{\xi}(\theta)$, $\theta \in [-\pi, \pi]$, associated with $\{\boldsymbol{\chi}_{nt} := (\chi_{1t} \ \dots \ \chi_{nt})', t \in \mathbb{Z}\}$ and $\{\boldsymbol{\xi}_{nt} := (\xi_{1t} \ \dots \ \xi_{nt})', t \in \mathbb{Z}\}$, respectively, denote by λ_{nk}^{χ} and λ_{nk}^{ξ} the corresponding *dynamic eigenvalues*, namely, the mappings $\theta \mapsto \lambda_{nk}^{\chi}(\theta)$ and $\theta \mapsto \lambda_{nk}^{\xi}(\theta)$, where $\lambda_{nk}^{\chi}(\theta)$ and $\lambda_{nk}^{\xi}(\theta)$ stand for the kth largest eigenvalues of $\mathbf{\Sigma}_n^{\chi}(\theta)$ and $\mathbf{\Sigma}_n^{\xi}(\theta)$, respectively. The *generalized dynamic factor model* analyzed in Forni et al. (2000) and Forni and Lippi (2001) still relies on the decomposition (2.2), but with the following assumptions :

(F3a) same as (F2a);

(F3b) $\lambda_{nq}^{\chi}(\theta) \to \infty$ as $n \to \infty$, θ-a.e. in $[-\pi \ \pi]$, and

(F3c) there exists a real Λ such that $\lambda_{n1}^{\xi}(\theta) \leq \Lambda$ for any $\theta \in [-\pi \ \pi]$ and any $n \in \mathbb{N}$.

Assumptions (F3a), (F3b) and (F3c) jointly will be referred to as Assumption (F3), and have been discussed in detail in Forni et al. (2000). We only recall here that the condition on the asymptotic behavior of $\lambda_{nk}^{\xi}(\theta)$ includes the case in which the idiosyncratic components are mutually orthogonal with an upper bound for the variances (this condition is trivially satisfied when n is finite and (F1b)=(F2b) holds). Assumption (F3) relaxes the orthogonality condition (F1b)=F2b) by allowing for a limited amount of cross-correlation among the various idiosyncratic components. Moreover, as Forni and Lippi (2001) show, the existence of a q-dimensional white noise $\{\mathbf{u}_t\}$ satisfying Assumptions (F3) is equivalent to the following condition on the eigenvalues $\lambda_{nk}(\theta)$ of the "observable" spectral density matrix $\mathbf{\Sigma}_n(\theta) = \mathbf{\Sigma}_n^{\chi}(\theta) + \mathbf{\Sigma}_n^{\xi}(\theta)$:

(F4) $\lambda_{nq}(\theta) \to \infty$ θ-a.e. in $[-\pi \ \pi]$ as $n \to \infty$, and there exists a real Λ such that $\lambda_{nq+1}(\theta) \leq \Lambda$ for any $\theta \in [-\pi \ \pi]$ and any $n \in \mathbb{N}$.

Forni et al. (2000) prove that under Assumptions (F0) and (F3), or equivalently (F0) and (F4), model (2.2) is identified, that is, the common and the idiosyncratic components are

uniquely characterized (not the vector of common shocks $\mathbf{u}_t$, though). Moreover, Forni et al. (2000) construct a consistent finite-sample estimator of the components χ_{it} and ξ_{it}, the definition of which we briefly recall here.

Denote by $\mathbf{\Sigma}_n^T(\theta) = \left(\sigma_{ij}^T(\theta)\right)$, $\theta \in [-\pi, \pi]$, an arbitrary periodogram-smoothing or lag-window estimator of the spectral density $\mathbf{\Sigma}_n(\theta) = (\sigma_{ij}(\theta))$ of $\mathbf{x}_{nt}$, based on the empirical cross-covariance matrices $\mathbf{\Gamma}_k^{nT} := (n-k)^{-1} \sum_{t=k+1}^{T} \mathbf{x}_{nt}\mathbf{x}_{n,t-k}$. Let $\lambda_{nj}^T(\theta)$ be $\mathbf{\Sigma}_n^T(\theta)$'s j-th largest eigenvalue, and let $\mathbf{p}_{nj}^T(\theta) = (p_{nj,1}^T(\theta), \dots, p_{nj,n}^T(\theta))$ be the corresponding row-eigenvector. Defining

$$\underline{\mathbf{p}}_{nj}^T(L) := \frac{1}{2\pi} \sum_{k=-\infty}^{\infty} \left[\int_{-\pi}^{\pi} \mathbf{p}_{nj}^T(\theta) e^{ik\theta} d\theta \right] L^k, \tag{2.3}$$

Forni et al. (2000) propose to estimate $\boldsymbol{\chi}_{nt} := (\chi_{1t} \ \cdots \ \chi_{nt})'$ by means of

$$\boldsymbol{\chi}_{nt}^T := (\chi_{1t,n}^T \ \cdots \ \chi_{nt,n}^T)' := \tilde{\underline{\mathbf{p}}}_{n1}^T(L)\underline{\mathbf{p}}_{n1}^T(L)\mathbf{x}_{nt} + \cdots + \tilde{\underline{\mathbf{p}}}_{nq}^T(L)\underline{\mathbf{p}}_{nq}^T(L)\mathbf{x}_{nt}, \tag{2.4}$$

where tilde denotes complex conjugation and transposition, and show that $\chi_{it,n}^T$ tends in probability to χ_{it}, as T and n tend to infinity, for t in the "central part" of the sample.

The problem with the filters (2.3) used in the definition of the estimated principal components $\chi_{it,n}^T$ is that they are in general two-sided. Therefore, the performance of (2.4) as an estimator of χ_{it} deteriorates at the ends of the sample, so that $\chi_{iT,n}^T$ is of little help when forecasting $\chi_{i,T+h}$ or $x_{i,T+h}$ is the main objective. This does not imply, however, that the dynamic factor approach just described is helpless in the forecasting context. The next section proposes a two-step strategy, in which the dynamic principal components of $\mathbf{\Sigma}_n^T(\theta)$ play the major role, for constructing one-sided filters that are suitable for end-of-sample estimation and forecasting.

2.3 Further assumptions.

The techniques developed in Forni et al. (2000) are valid under extremely general dynamic loadings $b_{ij}(L)$. The forecasting methods we are proposing here are slightly more restrictive in this respect, and require finite-order VARMA loadings. More precisely, we make the following assumption.

(F2c) The dynamic factor model (2.2) holds with VARMA(S, s) transfer function, that is,

$$\mathbf{x}_{nt} = \boldsymbol{\chi}_{nt} + \boldsymbol{\xi}_{nt} = \mathbf{B}_n(L)[\mathbf{A}(L)]^{-1}\mathbf{u}_t + \boldsymbol{\xi}_{nt}, \tag{2.5}$$

where $\mathbf{B}_n(L) := \mathbf{B}_0^n + \mathbf{B}_1^n L + \ldots + \mathbf{B}_s^n L^s$ is a $n \times q$ polynomial of order s, and $\mathbf{A}(L) := \mathbf{I} - \mathbf{A}_1 L - \ldots - \mathbf{A}_S L^S$ a $q \times q$ polynomial of order S; moreover, all solutions of $\det[\mathbf{A}(z)] = 0$ and $\det[\mathbf{A}(z)] = 0$, $z \in \mathbb{C}$, lie outside the unit ball, and there exists an m such that $\mathbf{B}_s^n \neq \mathbf{0}$ for $n \geq m$, and $S \leq s + 1$.

Note that the matrices $\mathbf{B}_k^n$ for fixed k and $n \in \mathbb{N}$ are nested, whereas the polynomial $\mathbf{A}(L)$ does not depend on n; the assumption on its characteristic roots guarantees the existence of the inverse operator $[\mathbf{A}(L)]^{-1}$. Finally, without the assumption that $\mathbf{B}_s^n \neq \mathbf{0}$ for $n \geq m$, the inequality $S \leq s+1$ would be meaningless. Writing $\mathbf{f}_t := (f_{1,t} \dots f_{q,t})'$ for $[\mathbf{A}(L)]^{-1}\mathbf{u}_t$ and $\mathbf{F}_t$ for $(\mathbf{f}_t' \ \mathbf{f}_{t-1}' \ \dots \ \mathbf{f}_{t-s}')'$, also note that the dynamic model (2.5) actually reduces to the static factor model

$$\mathbf{x}_{nt} = \mathbf{B}_n(L)\mathbf{f}_t + \boldsymbol{\xi}_{nt} = \mathbf{C}_n\mathbf{F}_t + \boldsymbol{\xi}_{nt} \tag{2.6}$$

with $r = q(s+1)$ factors and $\mathbf{C}_n := (\mathbf{B}_0^n\ \mathbf{B}_1^n\ \cdots\ \mathbf{B}_s^n)$.

Denote by $\mathbf{\Gamma}_{nk}^{\chi}$ and $\mathbf{\Gamma}_{nk}^{\xi}$ the k-lag covariance matrices of the vectors $\boldsymbol{\chi}_{nt}$ and $\boldsymbol{\xi}_{nt}$ respectively. Letting μ_{nj}^{χ} be the j-th eigenvalue of $\mathbf{\Gamma}_{n0}^{\chi}$, we also need strengthening (F3b) into

(F3b′) $\lim_{n\to\infty} \mu_{nr}^{\chi} = \infty$, where $r := q(s+1)$.

Assumption (F3b) indeed does not imply (F3b′). For example, if $\chi_{1t} = u_{t-1}$ and $\chi_{it} = u_t$ for $i \geq 2$, (F3b) clearly holds, with $q = 1$, and (F2c) also holds with $s = 1$ (and $S = 0$). However, (F3b′) does not hold since $\mu_{n1}^{\chi} \to \infty$ whereas μ_{n2}^{χ} is bounded as $n \to \infty$. In general, Assumption (F3b′) rules out the case in which some of the $f_{j,t-s}$ is loaded only by a finite number of the x's, so that the assumption that $\mathbf{B}_s^n \neq \mathbf{0}$ for $n \geq m$ is also strengthened.

Finally, for the sake of convenience, and without much loss of generality, we also will make the following assumption on the idiosyncratic processes :

(F3d) denoting by μ_{nj}^{ξ}, $j = 1, \ldots, n$, the eigenvalues of $\mathbf{\Gamma}_{n0}^{\xi}$, the smallest of them, μ_{nn}^{ξ}, is bounded away from zero as $n \to \infty$.

This last assumption avoids trivial but unpleasant problems of asymptotic degeneracy of idiosyncratics.

3 A two-step forecasting method based on dynamic factors

3.1 Outline of the method.

In view of the orthogonality, at any lead and lag, of χ_{it} and ξ_{it} in the decomposition of x_{it} into $\chi_{it} + \xi_{it}$, the problem of forecasting x_{it} can be split into forecasting the common component χ_{it} and forecasting the idiosyncratic ξ_{it} separately. Since we have assumed orthogonality (F2) or weak cross-correlation (F3) between the idiosyncratic components of distinct cross-sectional items, the forecast of ξ_{it} safely can be based on traditional univariate or low dimensional vector time series methods. Hence, we concentrate on forecasting χ_{it}.

Suppose that $\chi_{i,T+h}$ is to be predicted, at time T, for some given i and h. Denote by $\mathcal{H}(\chi, \tau)$ the Hilbert space spanned by the variables $\{\chi_{jt},\ t \leq \tau,\ j \in \mathbb{N}\}$. If $\mathcal{H}(\chi, T)$ were known, then the optimal linear forecast of $\chi_{i,T+h}$ in the mean square sense would be the projection

$$\phi_{i,T+h|T} := \text{proj}\left(\chi_{i,T+h}\middle|\mathcal{H}(\chi,T)\right) = \sum_{j=1}^{q}\sum_{k=0}^{\infty} b_{ij,k}\text{proj}\left(f_{j,T+h}\middle|\mathcal{H}(\chi,T)\right) \tag{3.1}$$

of $\chi_{i,T+h}$ onto $\mathcal{H}(\chi, T)$, where $b_{ij,k}$ is the coefficient of L^k in the filter $b_{ij}(L)$ (see (2.2)). Defining $\mathcal{G}(\mathbf{F}, \tau)$ as the r-dimensional space spanned by the r omponents of $\mathbf{F}_\tau$, for $k = 1, 2, \ldots, r$ (note that only values of $\mathbf{F}_t$ at time $t = \tau$ are taken into account), and using Assumption (F2c), (3.1) can be rewritten as

$$\phi_{i,T+h|T} = \text{proj}\left(\chi_{i,T+h}\middle|\mathcal{G}(\mathbf{F},T)\right) := \mathbf{M}_{ih}\mathbf{F}_T \tag{3.2}$$

where $\mathbf{M}_{ih}$ is an r-dimensional vector whose components can be easily obtained from the coefficients of $\mathbf{A}(L)$ and the coefficients $b_{ij,k}$.

Formula (3.2) of course is of little help in practice, since all quantities involved are unobservable. Our objective is to provide a sequence $\phi_{i,T+h|T}^{nT}$ of predictors computable from the observed $\mathbf{X}_n^T$, and converging, as n and T tend to infinity, to the optimal forecast of $\chi_{i,T+h}$.

Our construction is in two steps.

Step 1. We construct, so as to say, a *consistent estimate of* $\mathcal{G}(\mathbf{F},T)$. More precisely, we define an r-tuple

$$W_{nt}^{kT} := \mathbf{Z}_{nk}^{T}\mathbf{x}_{nt}, \quad k=1,\dots,r \tag{3.3}$$

of linear aggregates of $x_{1t},\dots,x_{nt}$, and prove that r-dimensional space $\mathcal{G}_n^T(\mathbf{F},T)$ spanned by $W_{nT}^{1T},\dots,W_{nT}^{rT}$ converges to $\mathcal{G}(\mathbf{F},T)$ as n and T go to infinity (see Proposition 3.1 for a precise statement). This space $\mathcal{G}_n^T(\mathbf{F},T)$ thus can be seen as an empirical substitute for the unknown $\mathcal{G}(\mathbf{F},T)$.

Step 2. We project $\chi_{i,T+h}$ onto $\mathcal{G}_n^T(\mathbf{F},T)$. This projection is our forecast. Proposition 3.2 provides a proof of the intuitive fact that when n and T tend to infinity, so that $\mathcal{G}_n^T(\mathbf{F},T)$ tends to $\mathcal{G}(\mathbf{F},T)$, this forecast converges to the optimal forecast $\phi_{i,T+h|T}$.

Estimators of the covariance matrices $\mathbf{\Gamma}_{n0}^{\chi}$ and $\mathbf{\Gamma}_{n0}^{\xi}$ are obtained as a simple application of the dynamic analysis developed in Forni et al. (2000). First we construct estimates $\mathbf{\Sigma}_n^{\chi T}(\theta)$ and $\mathbf{\Sigma}_n^{\xi T}(\theta)$ of the spectral density matrices of $\boldsymbol{\chi}_{nt}$ and $\boldsymbol{\xi}_{nt}$, respectively. These estimated spectral matrices are used to compute estimators $\mathbf{\Gamma}_{nk}^{\chi T}$ and $\mathbf{\Gamma}_{nk}^{\xi T}$ of the common and idiosyncratic autocovariance matrices for all lags k. The weights $\mathbf{Z}_{nk}^{T}$ are then obtained as those vectors that maximize the variance of the aggregate common component, given the variance of the aggregate idiosyncratic component. As we shall see, these vectors $\mathbf{Z}_{nk}^{T}$ are solutions of a generalized eigenvalue problem.

3.2 Consistent estimation of the autocovariance matrices of $\boldsymbol{\chi}_{nt}$ and $\boldsymbol{\xi}_{nt}$.

As explained in Section 3.1, the first step of our method consists in defining an r-tuple of aggregates (3.3), the span of which eventually coincides with that of the unobserved common shocks. Letting $r=q(s+1)$, our aggregates are constructed in a way that takes advantage of the common/idiosyncratic decomposition of the observations through an estimation of the covariance matrices $\mathbf{\Gamma}_{nk}^{\chi}$ and $\mathbf{\Gamma}_{nk}^{\xi}$.

We first compute the estimators $\mathbf{\Sigma}_n^T(\theta)=\left(\sigma_{ij}^T(\theta)\right)_{1\le i,j\le n}$ of the spectral density matrices $\mathbf{\Sigma}_n(\theta)=(\sigma_{ij}(\theta))_{1\le i,j\le n}$ as explained in Section 6.3. Under (F0), these estimators are such that, for all $n\in\mathbb{N}$ and $\epsilon>0$,

$$\lim_{T\to\infty} \mathrm{P}\left[\max_{1\le i,j\le n}\sup_{\theta\in[-\pi,\pi]}|\sigma^T{}_{ij}(\theta)-\sigma_{ij}(\theta)|>\epsilon\right]=0 \tag{3.4}$$

(see for instance Brockwell and Davis 1987, p. 433). Denoting by $\lambda_{nj}^T(\theta)$ and $\mathbf{p}_{nj}^T(\theta)$, $j=1,\dots,n$, the eigenvalues and eigenvectors of $\mathbf{\Sigma}_n^T(\theta)$, put

$$\mathbf{P}_n^T(\theta):=(\mathbf{p}_{n1}^{T\prime}(\theta)\ \dots\ \mathbf{p}_{nq}^{T\prime}(\theta))' \quad\text{and}\quad \mathbf{Q}_n^T(\theta):=(\mathbf{p}_{n,q+1}^{T\prime}(\theta)\ \dots\ \mathbf{p}_{nn}^{T\prime}(\theta))',$$

and define

$$\mathbf{\Sigma}_n^{\chi T}(\theta):=\tilde{\mathbf{P}}_n^T(\theta)\mathrm{Diag}\left(\lambda_{n1}^T(\theta),\ \dots,\ \lambda_{nq}^T(\theta)\right)\mathbf{P}_n^T(\theta)$$

and

$$\mathbf{\Sigma}_n^{\xi T}(\theta):=\tilde{\mathbf{Q}}_n^T(\theta)\mathrm{Diag}\left(\lambda_{n,q+1}^T(\theta),\ \dots,\ \lambda_{nn}^T(\theta)\right)\mathbf{Q}_n^T(\theta).$$

Moreover, let

$$\check{\mathbf{\Sigma}}_n^{\chi}(\theta) := \tilde{\mathbf{P}}_n(\theta)\text{Diag}\left(\lambda_{n1}(\theta),\ \ldots,\ \lambda_{nq}(\theta)\right)\mathbf{P}_n(\theta)$$

and

$$\check{\mathbf{\Sigma}}_n^{\xi}(\theta) := \tilde{\mathbf{Q}}_n(\theta)\text{Diag}\left(\lambda_{n,q+1}(\theta),\ \ldots,\ \lambda_{nn}(\theta)\right)\mathbf{Q}_n(\theta),$$

where $\lambda_{n,j}$, $\mathbf{P}_n(\theta)$ and $\mathbf{Q}_n(\theta)$ are the population counterparts of $\lambda_{n,j}^T$, $\mathbf{P}_n^T(\theta)$, and $\mathbf{Q}_n^T(\theta)$, respectively.

Now, (3.4) and the continuity of eigenvalues and eigenvectors as functions of the corresponding matrix entries imply that, for a given n, the matrices $\mathbf{\Sigma}_{nt}^{\chi T}$ and $\mathbf{\Sigma}_{nt}^{\xi T}$ are consistent estimators of $\check{\mathbf{\Sigma}}_{nt}^{\chi}$ and $\check{\mathbf{\Sigma}}_{nt}^{\xi}$ respectively, as $T \to \infty$, whereas

$$\mathbf{\Gamma}_{nk}^{\chi T} = \frac{1}{2\pi}\int_{-\pi}^{\pi} e^{ik\theta}\mathbf{\Sigma}_n^{\chi T}(\theta)d\theta \quad \text{and} \quad \mathbf{\Gamma}_{nk}^{\xi T} = \frac{1}{2\pi}\int_{-\pi}^{\pi} e^{ik\theta}\mathbf{\Sigma}_n^{\xi T}(\theta)d\theta \tag{3.5}$$

are consistent estimators, as $T \to \infty$, of

$$\check{\mathbf{\Gamma}}_{nk}^{\chi} = \frac{1}{2\pi}\int_{-\pi}^{\pi} e^{ik\theta}\check{\mathbf{\Sigma}}_n^{\chi}(\theta)d\theta \quad \text{and} \quad \check{\mathbf{\Gamma}}_{nk}^{\xi} = \frac{1}{2\pi}\int_{-\pi}^{\pi} e^{ik\theta}\check{\mathbf{\Sigma}}_n^{\xi}(\theta)d\theta, \tag{3.6}$$

respectively.

3.3 Step 1: consistent estimation of the space of common factors.

A crucial feature of our method is the choice of the r aggregates providing an estimation of $\mathcal{H}(\chi, T)$. Our r aggregates are of the form $W_{nt}^{kT} := \mathbf{Z}_{nk}^T\mathbf{x}_{nt}$, where the weights $\mathbf{Z}_{nk}^T$ are defined as the solutions, for $1 \le \ell \le r$, of the optimization problems

$$\begin{array}{rcl} \mathbf{Z}_{n\ell}^T & := & \text{Arg}\max\limits_{\boldsymbol{a}\in\mathbb{R}^n} \text{var}\left(\mathbf{a}\boldsymbol{\chi}_{nt}^T\right) \\ \text{subject to} & & \text{var}\left(\mathbf{a}\boldsymbol{\xi}_{nt}^T\right) = 1 \\ & & \mathbf{a}\boldsymbol{\xi}_{nt}^T \perp \mathbf{Z}_{nm}^T\boldsymbol{\xi}_{nt}^T \ \text{ for } \ 1 \le m < \ell, \ \ 1 \le \ell \le n. \end{array} \tag{3.7}$$

The rationale behind this choice is that, given the estimated variance-covariance matrices $\mathbf{\Gamma}_{n0}^{\chi T}$ and $\mathbf{\Gamma}_{n0}^{\xi T}$, the weights $\mathbf{Z}_{n\ell}^T$ obtained in (3.7) are maximizing the common-to-idiosyncratic variance ratio in the resulting aggregates.

Now, an equivalent form of (3.7) is

$$\begin{array}{rcl} \mathbf{Z}_{n\ell}^T & := & \text{Arg}\max\limits_{\boldsymbol{a}\in\mathbb{R}^n} \mathbf{a}\mathbf{\Gamma}_{n0}^{\chi T}\tilde{\mathbf{a}} \\ \text{subject to} & & \mathbf{a}\mathbf{\Gamma}_{n0}^{\xi T}\tilde{\mathbf{a}} = 1 \\ & & \mathbf{a}\mathbf{\Gamma}_{n0}^{\xi T}\tilde{\mathbf{Z}}_{nm}^T = 0 \ \text{ for } \ 1 \le m < \ell\,, 1 \le \ell \le n. \end{array} \tag{3.8}$$

Under this form, (3.7) reduces (see Lemma 6.1) to a *generalized eigenvalue problem*. Precisely, the vectors $\mathbf{Z}_{nj}^T$, $j = 1, \ldots, n$, are the *generalized eigenvectors*, associated to the *generalized eigenvalues* ν_{nj}^T, of the couple of matrices $\left(\mathbf{\Gamma}_{n0}^{\chi T}, \mathbf{\Gamma}_{n0}^{\xi T}\right)$, that is

$$\mathbf{Z}_{nj}^T\mathbf{\Gamma}_{n0}^{\chi T} = \nu_{nj}^T\mathbf{Z}_{nj}^T\mathbf{\Gamma}_{n0}^{\xi T} \quad j = 1, 2, \ldots, n \tag{3.9}$$

(the reader is referred to Section 6.1 for a brief description of generalized eigenvalue problems).

The following proposition provides a formal statement of the intuitive idea that the space spanned by the aggregates thus obtained approximates the unobserved space $\mathcal{G}(\mathbf{F},t)$. Let us first introduce some new notation. By w_{nt}^{kT} we denote the standardized version of W_{nt}^{kT}. Since

$$\mathbf{Z}_{nj}^T\mathbf{\Gamma}_{n0}^T\tilde{\mathbf{Z}}_{nj}^T = \mathbf{Z}_{nj}^T\mathbf{\Gamma}_{n0}^{\chi T}\tilde{\mathbf{Z}}_{nj}^T + \mathbf{Z}_{nj}^T\mathbf{\Gamma}_{n0}^{\xi T}\tilde{\mathbf{Z}}_{nj}^T = 1+\nu_{nj}^T,$$

then $w_{nt}^{kT} = \mathbf{z}_{nj}^T\mathbf{x}_{nt}$, where $\mathbf{z}_{nj}^T := \mathbf{Z}_{nj}^T/\sqrt{1+\nu_{nj}^T}$. Note that since $\mathbf{Z}_{nj}^T\mathbf{\Gamma}_{n0}^T\tilde{\mathbf{Z}}_{nk}^T = 0$ for $j \neq k$ (using the constraints of (3.8)), the vectors w_{nt}^{kT}, for $k = 1,2,\ldots,r$, form an orthonormal system spanning a space of the same dimension as $\mathcal{G}(\mathbf{F},t)$. By $\mathbf{Z}_{nj}$, ν_{nj} etc., we denote the objects playing the same role as $\mathbf{Z}_{nj}^T$, ν_{nj}^T etc., but with respect to $\check{\Gamma}_{n0}^{\chi}$ and $\check{\Gamma}_{n0}^{\xi}$.

Proposition 3.1 *Assume that (2.2), Assumptions (F0), (F2a)-(F2c), (F3a), (F3b′), (F3c) and (F3d) hold. Then, for all $\epsilon > 0$ and $\eta > 0$, there exist $N_0 = N_0(\epsilon,\eta)$ and $T_0 = T_0(n,\epsilon,\eta)$ such that*

$$\mathrm{P}\left[\left|w_{nt}^{kT} - \mathrm{proj}\left(w_{nt}^{kT}\middle|\mathcal{G}(\mathbf{F},t)\right)\right| > \epsilon\right] \leq \eta \tag{3.10}$$

for all $1 \leq t \leq T$, all $n \geq N_0$, all $T \geq T_0$, and all $1 \leq k \leq r$.

The proof of this Proposition relies on the following lemma, the proof of which is given in the Appendix (Section 6).

Lemma 3.1 *Let $\mathbf{a}_n := (a_{n1},\ldots,a_{nn})$ 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 3.1, $\mathbf{a}_n\boldsymbol{\xi}_{nt} \to 0$ in quadratic mean as $n \to \infty$.*

Proof of Proposition 3.1. Lemmas 6.2 and 6.3 imply that ν_{nr} tends to infinity as $n \to \infty$. As a consequence, each of the r sequences $\{\mathbf{Z}_{nj}/\sqrt{1+\nu_{nj}}, n \in \mathbb{N}\}$, $j = 1,\ldots,r$, is a triangular array fulfilling the assumption of Lemma 3.1. Indeed, $\mathbf{Z}_{nj}$ is bounded in modulus, since $1 = \mathbf{Z}_{nj}\mathbf{\Sigma}_{n0}^{\xi}\tilde{\mathbf{Z}}_{nj} \geq \mu_{nn}^{\xi}\mathbf{Z}_{nj}\tilde{\mathbf{Z}}_{nj}$ where, in view of Assumption (F3d), μ_{nn}^{ξ} is bounded away from zero. It follows that the sequences $\mathbf{Z}_{nj}\mathbf{x}_{nt}/\sqrt{1+\nu_{nj}}$ converge in probability to the space $\mathcal{G}(\mathbf{F},t)$ as $n \to \infty$ for any $j = 1,\ldots,r$, i.e. for all $\epsilon > 0$ and $\eta > 0$, there exist $N_1 := N_1(\epsilon,\eta)$ such that for all $1 \leq t \leq T$, all $n \geq N_1$, and all $1 \leq k \leq r$,

$$\mathrm{P}\left[\left|w_{nt}^{k} - \mathrm{proj}\left(w_{nt}^{k}\middle|\mathcal{G}(\mathbf{F},t)\right)\right| > \epsilon\right] \leq \eta. \tag{3.11}$$

Turning to w_{nt}^{kT}, we have

$$\begin{aligned}\left|w_{nt}^{kT} - \mathrm{proj}\left(w_{nt}^{kT}\middle|\mathcal{G}(\mathbf{F},t)\right)\right| &\leq \left|w_{nt}^{kT} - \mathrm{proj}\left(w_{nt}^{k}\middle|\mathcal{G}(\mathbf{F},t)\right)\right| \\ &\leq \left|w_{nt}^{kT} - w_{nt}^{k}\right| + \left|w_{nt}^{k} - \mathrm{proj}\left(w_{nt}^{k}\middle|\mathcal{G}(\mathbf{F},t)\right)\right|.\end{aligned}$$

Hence,

$$\begin{aligned}\mathrm{P}\left[\left|w_{nt}^{kT} - \mathrm{proj}\left(w_{nt}^{kT}\middle|\mathcal{G}(\mathbf{F},t)\right)\right| > \epsilon\right] &\leq \mathrm{P}\left[\left|w_{nt}^{k} - \mathrm{proj}\left(w_{nt}^{k}\middle|\mathcal{G}(\mathbf{F},t)\right)\right| > \frac{\epsilon}{2}\right] + \mathrm{P}\left[\left|w_{nt}^{kT} - w_{nt}^{k}\right| > \frac{\epsilon}{2}\right] \\ &:= P_1^n + P_2^{nT}, \qquad \text{say.}\end{aligned}$$

From (3.11), $P_1^n \leq \frac{\eta}{2}$ for all $n \geq N_1 = N_1(\frac{\epsilon}{2}, \frac{\eta}{2})$. As for P_2^{nT}, the convergence in probability, as $T \to \infty$, of $\mathbf{\Gamma}_{n0}^{\chi T}$ to $\check{\mathbf{\Gamma}}_{n0}^{\chi}$ and $\mathbf{\Gamma}_{n0}^{\xi T}$ to $\check{\mathbf{\Gamma}}_{n0}^{\xi}$, and the continuous mapping theorem, imply the existence, for all n, of $T_2 := T_2(n, \frac{\epsilon}{2}, \frac{\eta}{2})$ such that $P_2^{nT} \leq \frac{\eta}{2}$ for T larger that T_2 and all $1 \leq k \leq r$. The proposition follows, with $N_0(\epsilon, \eta) = N_1(\frac{\epsilon}{2}, \frac{\eta}{2})$ and $T_0(n, \epsilon, \eta) = T_2(n, \frac{\epsilon}{2}, \frac{\eta}{2})$. QED

3.4 Step 2: reconstructing the optimal forecast.

The forecast we propose for $\chi_{i,T+h}$ is the estimated projection of $x_{i,T+h}$ onto the space spanned by the r aggregates W_{nT}^{kT}, $k = 1, \ldots, r$, i.e.

$$\phi_{i,T+h|T}^{nT} := \left(\mathbf{\Gamma}_{nh}^{\chi T} \mathbf{Z}_n^T (\tilde{\mathbf{Z}}_n^T \mathbf{\Gamma}_{n0}^T \mathbf{Z}_n^T)^{-1} \tilde{\mathbf{Z}}_n^T \mathbf{x}_{nT} \right)_i \tag{3.12}$$

where $\mathbf{Z}_n^T := (\mathbf{Z}_{n1}^{T\prime} \ \cdots \ \mathbf{Z}_{nr}^{T\prime})'$.

The theoretical motivation of this forecast lies the following proposition, which shows that $\phi_{i,T+h|T}^{nT}$ is asymptotically equivalent, as n and T go to infinity, to the best possible forecast $\phi_{i,T+h|T}$.

Proposition 3.2 *Assume that (2.2), Assumptions (F0), (F2a)-(F2c), (F3a), (F3b′), (F3c) and (F3d) hold. Then, for all $\epsilon > 0$ and $\eta > 0$, there exist $N_3 = N_3(\epsilon, \eta)$ and $T_3 = T_3(n, \epsilon, \eta)$ such that, for all $n \geq N_3$, all $T \geq T_3$, and all $1 \leq i \leq N_3$,*

$$\mathrm{P}\left[\left| \phi_{i,T+h|T}^{nT} - \phi_{i,T+h|T} \right| > \epsilon \right] \leq \eta. \tag{3.13}$$

The proof of Proposition 3.2 relies on Proposition 3.1 and the following lemma.

Lemma 3.2 *Let $\mathcal{K}$ denote a subspace of the Hilbert space $\mathcal{H}$ of centered, square-integrable random variables, with covariance scalar product. Assume that $\mathcal{K}$ is generated by a subset $(v_1, \ldots, v_k)$ of k linearly independent elements of $\mathcal{H}$. Let $\{v_{n1}, \ldots, v_{nk}, n \in \mathbb{N}\}$ be a sequence of k-tuples of $\mathcal{H}$ such that, denoting by $\mathrm{proj}(.|\mathcal{K})$ the projection onto $\mathcal{K}$,*

(i) $v_{nj} - \mathrm{proj}(v_{nj}|\mathcal{K})$ converges to zero in probability as $n \to \infty$,

(ii) the determinant of the covariance matrix $(\mathrm{Cov}(v_{ni}, v_{nj}))_{i,j=1,\ldots,k}$ is bounded away from zero 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 probability, as $n \to \infty$, to the projection of v onto $\mathcal{K}$.

Proof. Let $\mathbf{v} := (v_1, \ldots, v_k)'$ and $\mathbf{v}_n := (v_{n1}, \ldots, v_{nk})'$. Consider the decomposition $\mathbf{v}_n = \mathbf{a}_n \mathbf{v} + \mathbf{R}_n$, of $\mathbf{v}_n$ into its (componentwise) orthogonal projection $\mathbf{a}_n \mathbf{v}$ onto $\mathcal{K}$ and the orthogonal complement. Assumption (i) implies that $\|\mathbf{R}_n\| := \mathbf{R}_n \mathbf{R}_n' \xrightarrow{\mathrm{P}} 0$. Decomposing similarly v into

$$v = \mathbf{b}_n \mathbf{v}_n + \mathbf{s}_n = \mathbf{b}_n \mathbf{a}_n \mathbf{v} + \mathbf{b}_n \mathbf{R}_n + \mathbf{s}_n \quad \text{and} \quad v = \mathbf{b}\mathbf{v} + \mathbf{s},$$

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, we obtain

$$\mathrm{proj}(v|\mathcal{K}_n) - \mathrm{proj}(v|\mathcal{K}) = \mathbf{b}_n \mathbf{v}_n - \mathbf{b}\mathbf{v} = \mathbf{b}_n \mathbf{a}_n \mathbf{v} + \mathbf{b}_n \mathbf{R}_n - \mathbf{b}\mathbf{v} = \mathbf{s} - \mathbf{s}_n. \tag{3.14}$$

Assumption (ii) implies that $\mathbf{b}_n$ is bounded. As a consequence $\mathbf{b}_n\mathbf{R}_n \xrightarrow{\text{P}} 0$, hence $\mathbf{b}_n\mathbf{v}_n - \mathbf{b}_n\mathbf{a}_n\mathbf{v} \xrightarrow{\text{P}} 0$. This and (3.14) in turn implies that $\mathbf{s} - \mathbf{s}_n \xrightarrow{\text{P}} 0$. The result follows. QED

Proof of Proposition 3.2. Choose two (otherwise arbitrary) strictly decreasing sequences $(\epsilon_n)_{n\in\mathbb{N}}$ and $(\eta_n)_{n\in\mathbb{N}}$ of strictly positive constants such that $\epsilon_n \downarrow 0$ and $\eta_n \downarrow 0$ as $n \to \infty$. Defining $T(n) := T_0(n, \epsilon_n, \eta_n) + 1$, the path $(n, T(n))_{n\in\mathbb{N}}$ has, in view of Theorem 3.1, the property that, for any $\epsilon > 0$ and $\eta > 0$, there exists a $N^* := N^*(\epsilon, \eta)$ such that

$$\text{P}\left[\left|w_{nt}^{k,T(n)} - \text{proj}\left(w_{nt}^{k,T(n)}\middle|\mathcal{G}(\mathbf{F}, t)\right)\right| > \epsilon\right] \leq \eta \tag{3.15}$$

for all $1 \leq t \leq T(n)$, all $n \geq N^*$, and all $1 \leq k \leq r$. Assumptions (i) of Lemma 3.2 thus holds for $(v_{n1}, \ldots, v_{nk}) = (w_{n,T(n)}^{1,T(n)}, \ldots, w_{n,T(n)}^{r,T(n)})$; since the vectors $w_{n,T(n)}^{k,T(n)}$ are an orthonormal system, assumption (ii) holds. Projecting onto the space spanned by the aggregates $W_{nt}^{k,T(n)}$ or by their standardized versions $w_{nt}^{k,T(n)}$ of course does not make any difference. Applying Lemma 3.2 to $v = x_{i,T+h}$, we thus obtain that (3.13) holds, with $N_3 = N^*$ and $T_3(n, \epsilon, \eta) = T(n)$, and with $\text{proj}\left(x_{i,T+h}\middle|\mathcal{G}(\mathbf{F}, T)\right)$ instead of $\phi_{i,T+h|T} := \text{proj}\left(\chi_{i,T+h}\middle|\mathcal{G}(\mathbf{F}, T)\right)$. However, since the common component $\chi_{i,T+h}$ is orthogonal, at all leads and lags, to the idiosyncratic process, $\text{proj}\left(x_{i,T+h}\middle|\mathcal{G}(\mathbf{F}, T)\right)$ and $\text{proj}\left(\chi_{i,T+h}\middle|\mathcal{G}(\mathbf{F}, T)\right)$ coincide, which completes the proof. QED

Throughout this section, the values of q and s, hence the value of r, have been assumed to be correctly specified. In Section 4.2, we show how to deal with this specification issue in practical situations. Note however that letting $r > q(s+1)$ clearly does not affect consistency results.

3.5 One-sided estimation of common/idiosyncratic components.

Obviously, within the sample, the method just described yields the projection

$$\phi_{i,t}^{nT} := \left(\mathbf{\Gamma}_{n0}^{\chi T}\mathbf{Z}_n^T(\tilde{\mathbf{Z}}_n^T\mathbf{\Gamma}_{n0}^T\mathbf{Z}_n^T)^{-1}\tilde{\mathbf{Z}}_n^T\mathbf{x}_{nt}\right)_i \tag{3.16}$$

of x_{it}, $t \leq T$, onto the space spanned by W_{nt}^{kT}, $k = 1, \ldots, r$. Under the same conditions as above, $\phi_{i,t}^{nT}$ converges in probability to χ_{it}, yielding a one-sided consistent estimation of the common component χ_{it} for $t \leq T$ which, for fixed t, avoids the end-of-sample inconsistency problems mentioned in Section 2.2.

3.6 Stock and Watson's one-step forecasting method.

Stock and Watson (1999) also propose a forecast based on a one-sided filtering of past and present observations. Their weights $\mathbf{S}_{nk}^T$ (playing the role of our weights $\mathbf{Z}_{nk}^T$) are defined (in our notation) as the static principal components of the observations, defined by

$$\begin{aligned} \mathbf{S}_{n\ell}^T &:= \text{Arg}\max_{\mathbf{a}\in\mathbb{R}^n} \mathbf{a}\mathbf{\Gamma}_{n0}^T \\ \text{subject to} \quad & \mathbf{a}\tilde{\mathbf{a}} = 1 \\ & \mathbf{a}\tilde{\mathbf{S}}_{nm}^T = 0 \ \text{ for } \ 1 \leq m < \ell\,,\, 1 \leq \ell \leq n. \end{aligned} \tag{3.17}$$

The resulting forecasts are obtained by projecting $x_{i,T+h}$ onto the space spanned by the r scalar aggregates

$$\mathbf{S}_{nj}^T\mathbf{x}_{nt}, \qquad j = 1, \ldots, r, \tag{3.18}$$

i.e. by substituting $\mathbf{S}_{nj}^T$ for $\mathbf{Z}_{nj}^T$ and $\mathbf{\Gamma}_{nh}^T$ for $\mathbf{\Gamma}_{nh}^{\chi T}$ in equations (3.12) and (3.16).

A comparison between the **Z**-weights of our two-step method ans the **S**-weights of Stock and Watson's gives a useful insight into the respective merits of the two methods (see Section 4 for a numerical comparison). Both methods indeed provide a consistent estimation of $\mathcal{G}(\mathbf{F}, T)$ and $\phi_{i,T+h|T}$ (under the assumptions of Section 2.3). If the matrices $\mathbf{\Gamma}_{nk}^{\chi T}$ and $\mathbf{\Gamma}_{nk}^{\xi T}$ are accurate estimates of $\mathbf{\Gamma}_{nk}^{\chi}$ and $\mathbf{\Gamma}_{nk}^{\xi}$ respectively, then our **Z**-weights, which are tailored with the purpose of minimizing the impact of the idiosyncratic in the aggregate, should perform better than the **S**-weights in approximating the space $\mathcal{G}(\mathbf{F}, T)$. However, the performance of such estimates depends in a very complicated way on the underlying model, on T, and on n, so that an *a priori* comparison between the two methods seems very difficult. A numerical assessment of their respective merits is made in Section 4.

4 Simulations and empirical results

4.1 Simulations

In order to evaluate the performance of our within-sample estimation and forecasting procedures for finite values of n and T, we carried out Monte Carlo experiments on three s-dependent models. In the first one (M1), the common components are generated by moving averages of order three, with two common factors ($q = 2$, $s = 3$), while the idiosyncratic components are mutually orthogonal. More precisely, the observations x_{it}^* were generated from the model

$$x_{it}^* = \sum_{k=0}^{3} a_{ki} u_{1,t-k} + \sum_{k=0}^{3} b_{ki} u_{2,t-k} + 3(c_i + 1)\epsilon_{it}, \tag{M1}$$

where the shocks u_{1t}, u_{2t}, ϵ_{it}, $t = 1, \dots, T$, $i = 1, \dots, n$ and the coefficients c_i, a_{ki}, and b_{ki}, $k = 0, 1, 2, 3$, $i = 1, \dots, n$ are mutually independent standard normal random variables.

The second model (M2) is identical to the first one ($q = 2$, $s = 3$) but for the idiosyncratic components, which now have a non-diagonal variance-covariance matrix; the model is

$$x_{it}^* = \sum_{k=0}^{3} a_{ki} u_{1,t-k} + \sum_{k=0}^{3} b_{ki} u_{2,t-k} + \xi_{it}, \tag{M2}$$

with

$$\xi_{it} = (1.5c_i + 1)\epsilon_{it} + (1.5c_{i+1} + 1)\epsilon_{i+1,t}$$

and $\epsilon_{n+1,t} = 0$ for all t.

In the third model (M3), the common components are generated by moving averages of order two, which are shifted in time in order to create leading, coincident, and lagging variables. Also in this case $q = 2$, and all shocks and coefficients are drawn from mutually independent standard normal distributions. More precisely, the observations are generated from the model

$$x_{it}^* = \sum_{k=l_i}^{l_i+2} a_{k-l_i,i} u_{1,t-k} + \sum_{k=l_i}^{l_i+2} b_{k-l_i,i} u_{2,t-k} + 2(c_i + 1)\epsilon_{it}, \tag{M3}$$

where $l_i = 0$ for $1 \leq i \leq m$ (leading variables), $l_i = 1$ for $m + 1 \leq i \leq 2m$ (coincident variables), and $l_i = 2$ for $2m + 1 \leq i \leq n$ (lagging variables); hence, $s = 4$. In order for the three types

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).

Before estimation of the spectral matrices, all variables were taken in deviation from their sample means and divided by their standard deviations, i.e., spectral estimation was conducted from the standardized observations

$$x_{it} = (x^*_{it} - \bar{x}^*_i)/s_i, \tag{4.1}$$

where $\bar{x}^*_i := \sum_{t=1}^T x^*_{it}/T$ and $s_i^2 := \sum_{t=1}^T (x^*_{it} - \bar{x}^*_i)^2/(T-1)$.

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 $\mathbf{\Gamma}^{\xi T}_{n0}$ of the idiosyncratic components improves forecasting performance, even when the actual matrix is non diagonal (see results from model M2 below). Our explanation for this somewhat counterintuitive result is the following. When computing $\mathbf{\Gamma}^{\xi T}_{n0}$, 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 $\mathbf{\Gamma}^{\xi T}_{n0}$. On the other hand, by diagonalizing our estimated matrix, we ignore the true off-diagonal non-zero entries. Also in this case the error increases with n, but, so to speak, owing to the boundedness of the eigenvalues, it increases only linearly in n.

Therefore, we henceforth diagonalize the matrices $\mathbf{\Gamma}^{\xi T}_{n0}$ before computing eigenvectors. Consistency is still ensured : indeed, it is easily seen from Lemmas 6.2 and 6.3 in the Appendix that replacing $\mathbf{\Gamma}^{\xi T}_{n0}$ with any symmetric semi-positive definite matrix with bounded eigenvalues does not affect consistency results.

We generated data from each model with $n = 20, 50, 100, 200$ and $T = 20, 50, 100, 200$. Then we applied the procedures described in Section 3 (see Section 6.3 of the Appendix for details) in order to compute the within-sample estimates ϕ^{nT}_{it}, $t = 1, \dots, T$ and the one-step-ahead forecasts $\phi^{nT}_{i,T+1|T}$. For the sake of simplicity, we assumed both the number q of dynamic factors and the number $r = q(s+1)$ of static factors to be known. Each experiment was replicated 400 times.

An important term of comparison is given by the corresponding estimates and forecasts obtained by using the static factor model method proposed by Stock and Watson (1999), i.e. by projecting the x's on the first r static principal components (see Section 3.6). This actually consists in using the same algorithm as for $\phi^{nT}_{i,T+1|T}$, but with weights $\mathbf{S}^T_n$ instead of $\mathbf{Z}^T_n$ (see (3.7) and (3.17)), and $\mathbf{\Gamma}^T_{nh}$ for $\mathbf{\Gamma}^{\chi T}_{nh}$.

We measured the performance of one-step-ahead forecasts by means of the criterion

$$\frac{\sum_{i=1}^n (\phi^{nT}_{i,T+1|T} - \phi_{i,T+1|T})^2}{\sum_{i=1}^n \mathrm{var}(\phi_{i,t+1|t})}$$

and performance of within-sample estimates by means of

$$\frac{\sum_{i=1}^n \sum_{t=1}^T (\phi^{nT}_{i,t} - \phi_{it})^2}{T \sum_{i=1}^n \mathrm{var}(\phi_{it})}.$$

Note that, in these two criteria, we normalize by dividing by the theoretical variances.

Results for models M1, M2, and M3 are shown in Tables 5.1, 5.2, and 5.3, respectively, with part (a) devoted to forecasts and part (b) devoted to within-sample estimation. In each cell

we report (i) the average value of the criterion, across the 400 replications, for our two-step method, (ii) (in brackets) the empirical standard deviation, across the 400 replications, of the same criterion, still for our two-step method, and (iii) (in square brackets) the average value of the criterion, across the 400 replications, for the one-step static principal component method.

Table 5.1a: Model M1, forecasting results

	$n = 20$	$n = 50$	$n = 100$	$n = 200$
	0.8769	0.6369	0.4894	0.3767
$T = 20$	(0.5461)	(0.4219)	(0.3601)	(0.2590)
	[0.9679]	[0.7600]	[0.6329]	[0.5610]
	0.7520	0.3790	0.2323	0.1534
$T = 50$	(0.4313)	(0.1954)	(0.1206)	(0.0839)
	[0.8693]	[0.5067]	[0.3383]	[0.2346]
	0.6336	0.3097	0.1657	0.1035
$T = 100$	(0.3309)	(0.1588)	(0.0838)	(0.0498)
	[0.7249]	[0.3878]	[0.2182]	[0.1465]
	0.6138	0.2712	0.1442	0.0806
$T = 200$	(0.3058)	(0.1521)	(0.0715)	(0.0420)
	[0.6883]	[0.3188]	[0.1772]	[0.1020]

Table 5.1b: Model M1, within-sample results

	$n = 20$	$n = 50$	$n = 100$	$n = 200$
	0.9259	0.5196	0.3679	0.2920
$T = 20$	(0.2974)	(0.1649)	(0.1155)	(0.0973)
	[1.1457]	[1.0111]	[0.9072]	[0.8213]
	0.7339	0.3375	0.2127	0.1475
$T = 50$	(0.1563)	(0.0719)	(0.0446)	(0.0334)
	[0.8879]	[0.5783]	[0.4077]	[0.2987]
	0.6514	0.2753	0.1613	0.1014
$T = 100$	(0.1116)	(0.0385)	(0.0226)	(0.0151)
	[0.7436]	[0.3783]	[0.2358]	[0.1609]
	0.6075	0.2505	0.1360	0.0800
$T = 200$	(0.0854)	(0.0260)	(0.0138)	(0.0082)
	[0.6450]	[0.2943]	[0.1666]	[0.1041]

Table 5.2a: Model M2, forecasting results

	$n = 20$	$n = 50$	$n = 100$	$n = 200$
	0.9540	0.6098	0.4724	0.3606
$T = 20$	(0.5666)	(0.4028)	(0.3352)	(0.2371)
	[0.9913]	[0.6781]	[0.5938]	[0.4992]
	0.7339	0.3645	0.2330	0.1638
$T = 50$	(0.4130)	(0.2047)	(0.1272)	(0.0938)
	[0.8140]	[0.4807]	[0.3372]	[0.2471]
	0.6132	0.2811	0.1653	0.1060
$T = 100$	(0.3445)	(0.1416)	(0.0864)	(0.0528)
	[0.6878]	[0.3459]	[0.2074]	[0.1387]
	0.5641	0.2561	0.1325	0.0828
$T = 200$	(0.3197)	(0.1501)	(0.0723)	(0.0369)
	[0.6287]	[0.2882]	[0.1553]	[0.0993]

Table 5.2b: Model M2, within-sample results

	$n = 20$	$n = 50$	$n = 100$	$n = 200$
	0.8639	0.5444	0.4058	0.3445
$T = 20$	(0.2802)	(0.1739)	(0.1239)	(0.1138)
	[1.0205]	[0.9942]	[0.9504]	[0.9525]
	0.6608	0.3253	0.2177	0.1648
$T = 50$	(0.1475)	(0.0709)	(0.0460)	(0.0333)
	[0.7907]	[0.5713]	[0.4360]	[0.3481]
	0.5688	0.2557	0.1559	0.1064
$T = 100$	(0.0885)	(0.0372)	(0.0212)	(0.0160)
	[0.6511]	[0.3564]	[0.2257]	[0.1609]
	0.5299	0.2211	0.1272	0.0778
$T = 200$	(0.0613)	(0.0228)	(0.0131)	(0.0072)
	[0.5614]	[0.2578]	[0.1521]	[0.0963]

Table 5.3a: Model M 3, forecasting results

	$n = 20$	$n = 50$	$n = 100$	$n = 200$
$T = 20$	0.7990	0.4962	0.4109	0.3595
	(0.5107)	(0.3011)	(0.2771)	(0.2467)
	[0.9046]	[0.6136]	[0.5406]	[0.5116]
$T = 50$	0.6172	0.3346	0.1957	0.1479
	(0.3740)	(0.1876)	(0.1094)	(0.0961)
	[0.7151]	[0.4544]	[0.3071]	[0.2552]
$T = 100$	0.5799	0.2587	0.1552	0.0896
	(0.3521)	(0.1340)	(0.0894)	(0.0447)
	[0.6683]	[0.3264]	[0.2067]	[0.1292]
$T = 200$	0.5490	0.2231	0.1251	0.0679
	(0.3057)	(0.1308)	(0.0730)	(0.0318)
	[0.6145]	[0.2593]	[0.1446]	[0.0852]

Table 5.3b: Model M3, within-sample results

	$n = 20$	$n = 50$	$n = 100$	$n = 200$
$T = 20$	0.6048	0.3533	0.2595	0.2162
	(0.1908)	(0.1053)	(0.0704)	(0.0561)
	[0.7847]	[0.7312]	[0.6750]	[0.6407]
$T = 50$	0.4921	0.2413	0.1608	0.1175
	(0.1057)	(0.0468)	(0.0282)	(0.0203)
	[0.6411]	[0.4470]	[0.3425]	[0.2668]
$T = 100$	0.4523	0.2058	0.1245	0.0828
	(0.0774)	(0.0278)	(0.0159)	(0.0108)
	[0.5577]	[0.3117]	[0.1922]	[0.1330]
$T = 200$	0.4308	0.1865	0.1081	0.0670
	(0.0567)	(0.0195)	(0.0099)	(0.0059)
	[0.4999]	[0.2355]	[0.1337]	[0.0835]

The tables show that for all models and all choices of n and T the two-step method outperforms the one-step method. This is true both for forecasting and within-sample reestimation. Our relative advantage is particularly evident when n is large relative to T.

4.2 Empirics.

In this Section, we evaluate the performance of our methodology empirically. The database used has been constructed by the Research Department of the Banca d'Italia within a Bank of Italy-CEPR project. Here we used the $n = 447$ variable subset fully documented in Cristadoro et al. (2001). The data consist of monthly time series on key aggregates and sectorial variables for the six main economies in the Euro area—Germany, France, Italy, Spain, the Netherlands, Belgium—and, when available, for the Euro area as a whole. The aim is to forecast two target

variables : the Euro aggregate industrial production and the Euro area harmonized consumer price index for the period 1987:1-2001:3. Data are logged and first-differenced whenever needed to achieve stationarity. For details on data treatment, see Cristadoro et al. (2001).

The forecasting exercise is a traditional simulated out-of-sample experiment. Three different models (AR, FHLR, and SW; see below for a precise description) are considered successively, with the corresponding forecasting method. For each of these models/methods, we proceed as follows. Assume that the target variable x_{it} is to be predicted. First, we estimate the model using observations from 1987:2 to 1997:1 and, based on this estimation, we compute $h = 1, 3, 6, 12$-step ahead forecasts; these forecasts are h-step ahead predictions $x_{i,T+h|T}$, made at time $T = 1997 : 1$, of $x_{i,T+h}$. The corresponding forecasting error is $x_{i,T+h|T} - x_{i,T+h}$. We then reestimate the model using one additional observation, and compute again h-step ahead forecasts $x_{i,T+h|T}$, and the corresponding forecasting error, for $T = 1997 : 2$. Adding one observation at a time, we repeat the same exercise until all observations, from 1987:2 to 2000:3, are used in the estimation. Finally, for each method, and for each value of the forecasting horizon h, we average the squared prediction errors thus obtained over 50 successive values of T. After dividing by the corresponding empirical variance, we obtain

$$MSE_h := \sum_T (x_{i,T+h|T} - x_{i,T+h})^2 / \sum_T (x_{i,T+h} - \bar{x}_i^*)^2, \quad \text{where} \quad \bar{x}_i^* := \frac{1}{50} \sum_T x_{i,T+h}. \tag{4.2}$$

This MSE_h is used as an assesment of the forecasting performance of the model/method used, at lag h.

Without loss of generality, assume again that x_{it} is the target variable to be predicted (in our case, either inflation or the industrial production index, both in difference of logs). The three models/forecasting methods we are considering are

(AR) (univariate AR models). The model used for predicting x_{it} is a simple univariate AR model, of the form

$$x_{i,t} = \sum_{j=1}^{l} \alpha_j x_{i,t-j} + \epsilon_{i,t}. \tag{AR}$$

First assume that l is fixed. After traditional OLS estimation, a mean squared prediction error MSE_h^l (4.2) is obtained for each horizon h. This exercise is performed for $l = 1, \ldots, 36$, and the value l_h^* of l minimizing MSE_h^l is adopted : $MSE_h^* := MSE_h^{l^*}$ is thus the best forecasting performance obtainable, at horizon h, via a pure univariate AR approach.

(FHLR) (our method). Here, x_{it} decomposes into $\chi_{it} + \xi_{it}$. The common component χ_{it} is predicted as explained in the previous sections, whereas the idiosyncratic component ξ_{it} is predicted from its own past only; the autoregressive coefficients are obtained from the lagged covariances computed by using the estimated covariances $\mathbf{\Gamma}_{nh}^{\xi T}$ in the OLS formula. The h-step ahead forecast $x_{i,T+h|T}$ of $x_{i,T+h}$ then is obtained by adding the respective forecasts of $\chi_{i,T+h}$ and $\xi_{i,T+h}$. The model used for predicting x_{it} thus has the form

$$\begin{cases} x_{i,t+h} &= \chi_{i,t+h} + \xi_{i,t+h} \\ \chi_{i,t+h} &= \sum_{k=1}^{r} \beta_k^h w_{nt}^k + \epsilon_t^{\chi,h} \\ \xi_{i,t+h} &= \sum_{j=1}^{m} \gamma_j \xi_{i,t-j} + \epsilon_t^{\xi,h}, \end{cases} \tag{FHLR}$$

where the two OLS estimations are to be performed independently of each other, leading for each h to a MSE_h (4.2). The MSE_h was computed for each combination of $h = 1, 3, 6, 12$, $r = 1, \ldots, 50$, and $m = 1, \ldots, 30$. The values r_h^*, and m_h^* minimizing this MSE_h were selected, yielding a minimal value MSE_h^*. This MSE_h^* thus is the best forecasting performance obtainable, at horizon h, via the two-step procedure based on the dynamic factor approach.

(SW) (one-step method based on the static factor approach). The prediction model now is

$$x_{i,t+h} = \delta^h(L)x_{it} + \sum_{k=1}^{r} \hat{\beta}_k^h \hat{w}_{nt}^k + \eta_t^h. \qquad \text{(SW)}$$

where $\delta^h(L)$ is polynomial of order p, and $\hat{w}_{nt}^k$, $k = 1, \ldots, r$ denote the aggregates obtained by solving (3.17) rather than (3.8). Note that the observable autoregressive terms $\delta^h(L)x_{1t}$ here help forecasting the idiosyncratic component, while the aggregates $\hat{w}_{nt}^k$ help forecasting the common component. Mean square errors are obtained, along the same lines as in (FHLR), then minimized in order to identify, for each horizon h, optimal values of p and r, yielding again, for each h, the best forecasting performance MSE_h^* that can be achieved by means of the one-step procedure based on static factors.

For the FHLR method, two more parameters need to be set for estimation : the number of common factors q and the window size M used in spectral estimation. Both are needed for the computation of the covariances in the first step and were set as $M = 24$, $q = 3$ (see Section 6.3 in the appendix).

The results are presented in Table 5.4 for industrial production, in Table 5.5 for the harmonized inflation index. The numbers in squared brackets in the first column are the optimal lags l^* for (AR); the numbers in round brackets in the other columns are the optimal static ranks r^* and optimal idiosyncratic lags (m^* for FHLR, p^* for SW), respectively.

Table 5.4 Industrial Production

horizon	AR		FHLR			SW		
	MSE	lag l^*	MSE	static rank r^*	idiosyncratic lag m^*	MSE	static rank r^*	idiosyncratic lag p^*
h=1	0.72	11	0.68	6	3	0.73	4	11
h=3	0.81	18	0.77	3	3	0.81	2	1
h=6	0.94	18	0.93	3	3	0.96	1	3
h=12	0.91	34	0.96	6	6	0.97	17	7

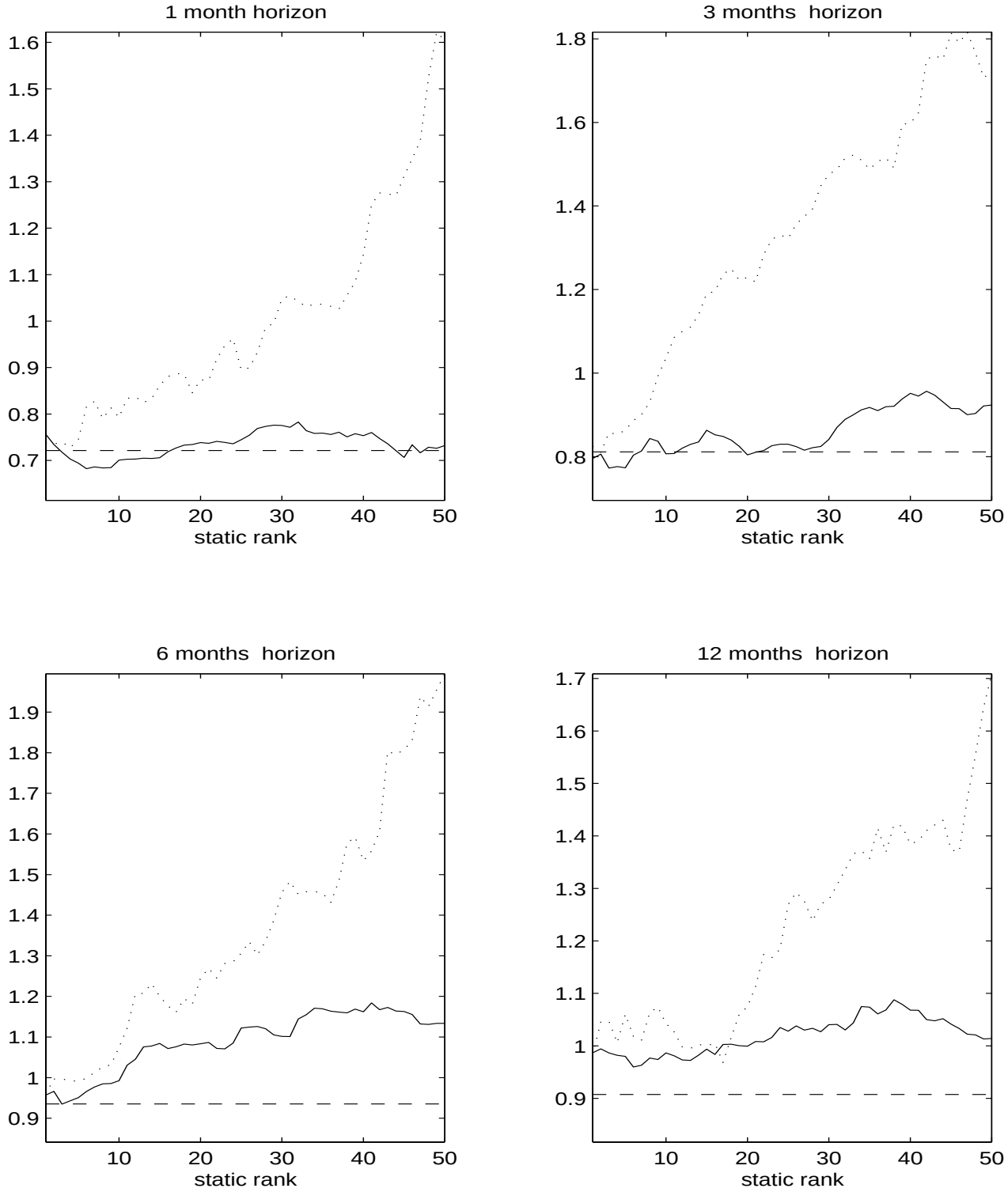
Table 5.5 Harmonized Inflation Index

horizon	AR		FHLR			SW		
	MSE	lag l^*	MSE	static rank r^*	idiosyncratic lag m^*	MSE	static rank r^*	LAG lag p^*
h=1	0.66	9	0.53	10	1	0.67	10	9
h=3	0.66	10	0.66	5	2	0.62	6	8
h=6	0.68	10	0.67	2	3	0.67	1	6
h=12	0.83	13	0.89	5	1	0.89	4	8

Results indicate that the FHLR method outperforms both the AR and the SW methods at horizons $h = 1, 3$, and 6. For inflation, the superiority of our method is clear only at one month horizon, while for larger values of h, all methods yield roughly the same predictive performances.

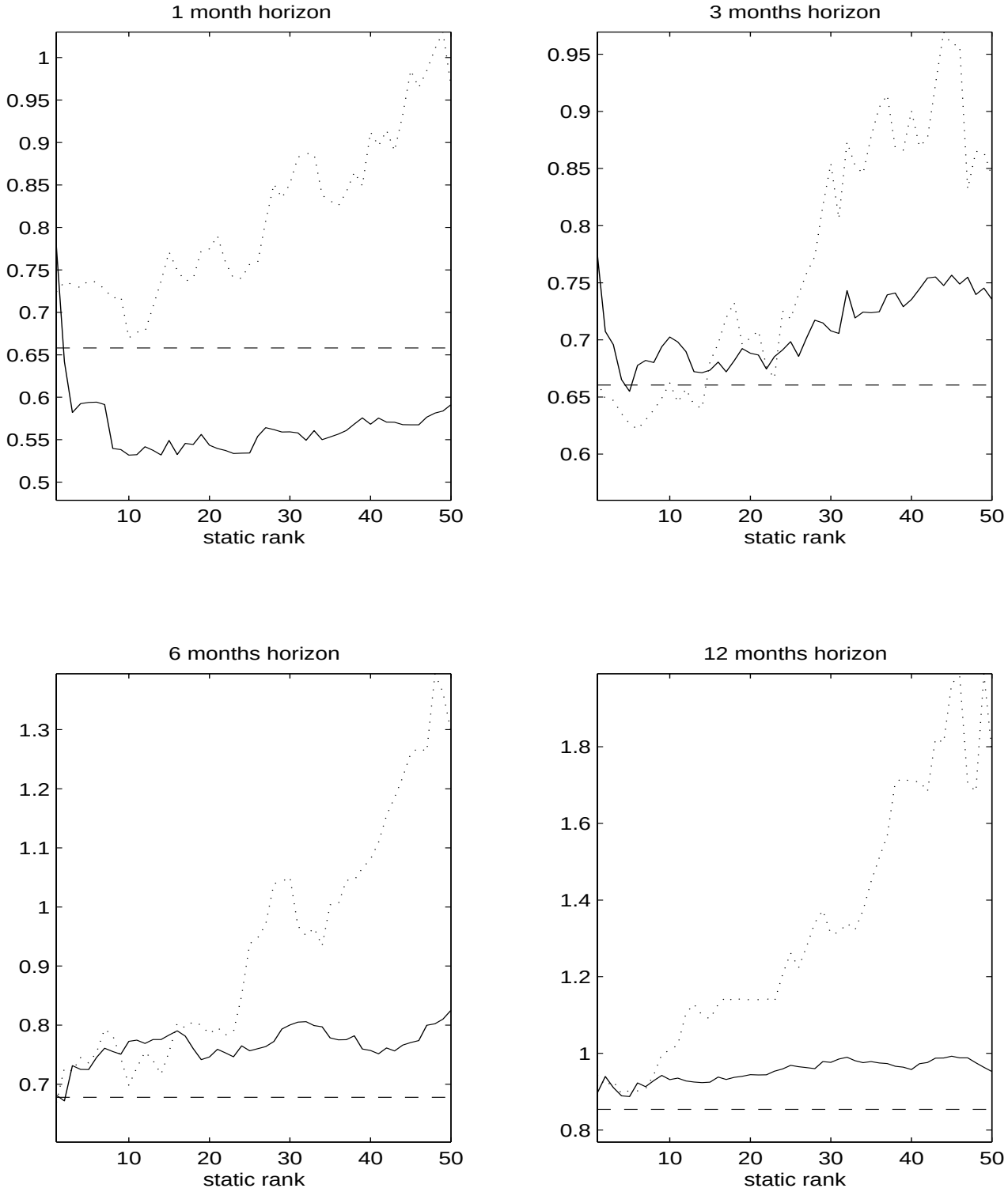
It is interesting also to analyze the stability of the ranks r^* that minimize the MSE in the out-of-sample forecast exercise. This analysis is illustrated in Figures 5.1 and 5.2, where MSE is plotted against different values of the static rank r. The results based on the FHLR method appear to be more stable, in this respect, than those based on the SW method.

Figure 5.1. Industrial Production: mean squared errors
FHLR, SW and AR over static rank



FHLR (solid line), SW (dotted line), AR (dashed line)

Figure 5.2. Consumer Price Index: mean squared errors
FHLR, SW and AR over static rank



FHLR (solid line), SW (dotted line), AR (dashed line)

5 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 et al. (2000) for computing linear aggregates of present and past observations in the panel that minimize the ratio of the variance of the idiosyncratic component with respect to that of the common component. We show that these aggregates allow for a consistent estimation of the space spanned by the unobserved common components up to time T. The second step consists in projecting the observation to be predicted onto the linear spacespanned by these aggregates. We show that this projection converges to the optimal forecast as n and T go to infinity. Being a linear combination of present and past observations only, the proposed predictor solves the end-of-sample problems caused by two-sided filtering in the estimation method of Forni et al. (2000), while exploiting the advantages of dynamic information. Simulation results and empiric results show very encouraging results.

6 Appendix

6.1 Generalized eigenvalues.

Denoting by $\mathbf{\Gamma}$ and $\mathbf{D}$ two real, symmetric, positive semidefinite $n \times n$ matrices, consider all complex numbers ν_j, and all $1 \times n$ complex vectors $\mathbf{v}_j$ such that

$$\begin{aligned} \mathbf{\Gamma}\tilde{\mathbf{v}}_j &= \nu_j \mathbf{D}\tilde{\mathbf{v}}_j, \\ \mathbf{v}_j \mathbf{D}\tilde{\mathbf{v}}_j &= 1, \\ \mathbf{v}_j \mathbf{D}\tilde{\mathbf{v}}_m &= 0, \quad j \neq m. \end{aligned} \tag{6.1}$$

The solutions ν_j and $\mathbf{v}_j$ of (6.1) are called the *generalized eigenvalues* and *generalized eigenvectors* of the couple $(\mathbf{\Gamma}, \mathbf{D})$, respectively (see Wilkinson, 1965). The problem (6.1) of obtaining all generalized eigenvalues and eigenvectors of a given couple $(\mathbf{\Gamma}, \mathbf{D})$ is known as a *generalized eigenvalue problem*.

Since $\mathbf{\Gamma}$ and $\mathbf{D}$ in (6.1) have been assumed positive semidefinite, the eigenvalues ν_j are nonnegative real numbers; throughout, we tacitly assume that they are ranked in decreasing order of magnitude. The relation between generalized eigenvalues and maximization problems of the form (3.8) is established in the following Lemma.

Lemma 6.1 *Let $\mathbf{\Gamma}$ and $\mathbf{D}$ denote two real, symmetric, positive semidefinite $n \times n$ matrices, and consider the n-tuple of maximization problems*

$$\begin{aligned} \mathbf{a}_j \quad &:= \quad \operatorname{Arg}\max_{\mathbf{a}\in\mathbb{R}^n} \mathbf{a}\mathbf{\Gamma}\tilde{\mathbf{a}} \\ \text{subject to} \quad & \quad \mathbf{a}\mathbf{D}\tilde{\mathbf{a}} = 1, \\ & \quad \mathbf{a}\mathbf{D}\tilde{\mathbf{a}}_m = 0 \quad \text{for} \quad 1 \leq m \leq j-1, \end{aligned} \tag{6.2}$$

$j = 1, \ldots, n$. The solutions of (6.2) are those of the generalized eigenvalue problem (6.1). More precisely, denoting by ν_j and $\mathbf{v}_j$ the generalized eigenvalues and eigenvectors of the couple $(\mathbf{\Gamma}, \mathbf{D})$, the solutions of (6.2) are $\mathbf{a}_j = \mathbf{v}_j$, and $\mathbf{a}_j \mathbf{\Gamma}\tilde{\mathbf{a}}_j = \nu_j$.

Proof. The Lagrangian for $j = 1$ in (6.2) is $\mathbf{a}\mathbf{\Gamma}\tilde{\mathbf{a}} - \lambda(\mathbf{a}\mathbf{D}\mathbf{a}' - 1)$, so that the first order conditions are

$$\mathbf{\Gamma}\tilde{\mathbf{a}} = \lambda \mathbf{D}\tilde{\mathbf{a}} \quad \text{and} \qquad \mathbf{a}\mathbf{D}\tilde{\mathbf{a}} = 1.$$

These conditions are satisfied only by the generalized eigenvectors $\mathbf{v}_j$, $j = 1, \ldots, n$. Since $\nu_j = \mathbf{v}_j \boldsymbol{\Gamma} \tilde{\mathbf{v}}_j$ is the value of the objective function, $\mathbf{a}_1 = \mathbf{v}_1$ solves the problem for $j = 1$.

The Lagrangian for $j = 2$ in (6.2) is $\mathbf{a}\boldsymbol{\Gamma}\tilde{\mathbf{a}} - \lambda(\mathbf{a}\mathbf{D}\tilde{\mathbf{a}} - 1) - \mu(\mathbf{a}\mathbf{D}\tilde{\mathbf{v}}_1)$, so that the first order conditions take the form

$$2\boldsymbol{\Gamma}\tilde{\mathbf{a}} - 2\lambda\mathbf{D}\tilde{\mathbf{a}} - \mu\mathbf{D}\tilde{\mathbf{v}}_1 = 0, \quad \mathbf{a}\mathbf{D}\tilde{\mathbf{a}} = 1, \quad \text{and} \quad \mathbf{a}\mathbf{D}\tilde{\mathbf{v}}_m = 0. \tag{6.3}$$

Premultiplying by $\mathbf{v}_1$ and taking (6.1) into account, we get $2\mathbf{v}_1\boldsymbol{\Gamma}\tilde{\mathbf{a}} - \mu = 2\lambda\mathbf{v}_1\mathbf{D}\tilde{\mathbf{a}} - \mu = -\mu = 0$, so that (6.3) reduces to

$$\boldsymbol{\Gamma}\tilde{\mathbf{a}} = \lambda\mathbf{D}\tilde{\mathbf{a}}, \quad \mathbf{a}\mathbf{D}\tilde{\mathbf{a}} = 1, \quad \text{and} \quad \mathbf{a}\mathbf{D}\tilde{\mathbf{v}}_1 = 0.$$

These conditions are satisfied only by the generalized eigenvectors $\mathbf{v}_j$, $j = 2, \ldots, n$, so that (6.2) for $j = 2$ is solved by $\mathbf{a}_2 = \mathbf{v}_2$. The lemma follows recursively. QED

Lemma 6.2 *Consider a sequences of real, symmetric, semi-positive definite $n \times n$ matrices $(\boldsymbol{\Gamma}_n, \mathbf{D}_n)$, $n = k, k+1, \ldots$, and assume that*

(i) $\boldsymbol{\Gamma}_n$'s kth largest eigenvalue λ_{nk} diverges as $n \to \infty$, and

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

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

Proof. Denote by $\mathbf{W}_n$ the $k \times n$ matrix whose rows are the k first eigenvectors of $\boldsymbol{\Gamma}_n$, and consider the representation

$$\mathbf{W}_n\mathbf{D}_n\mathbf{W}_n' = \mathbf{U}_n\mathbf{R}_n\tilde{\mathbf{U}}_n,$$

where $\mathbf{U}_n$ is a unitary $k \times k$ matrix and $\mathbf{R}_n = \text{Diag}(r_{n1}, \ldots, r_{nj})$ is the diagonal matrix of the k largest eigenvalues of $\mathbf{D}_n$. Clearly, $r_{nj} \leq \mu$ for all j, whereas the rows of $\mathbf{R}^{-1/2}\tilde{\mathbf{U}}_n\mathbf{W}_n$ satisfy the normalization and the orthogonality constraints of problem (6.2), since

$$\mathbf{R}^{-1/2}\tilde{\mathbf{U}}_n\mathbf{W}_n\mathbf{D}_n\tilde{\mathbf{W}}_n\mathbf{U}_n\mathbf{R}^{-1/2} = \mathbf{I}_k.$$

Hence, Lemma 6.1 implies that $\mathbf{R}^{-1/2}\tilde{\mathbf{U}}_n\mathbf{W}_n$'s last row, namely, $\boldsymbol{\rho}_n\tilde{\mathbf{U}}_n\mathbf{W}_n$, where $\boldsymbol{\rho}_n := (0 \ \cdots \ 0 \ r_{nk}^{-1/2})$, is such that $\boldsymbol{\rho}_n\tilde{\mathbf{U}}_n\mathbf{W}_n\boldsymbol{\Gamma}_n\tilde{\mathbf{W}}_n\mathbf{U}_n\tilde{\boldsymbol{\rho}}_n \leq \nu_{nk}$. It follows that, denoting by $\boldsymbol{\Lambda}_n := \text{Diag}(\lambda_{n1}, \ldots, \lambda_{nk})$ the diagonal matrix of $\boldsymbol{\Gamma}_n$'s first k eigenvalues,

$$\nu_{nk} \geq \boldsymbol{\rho}_n\tilde{\mathbf{U}}_n\boldsymbol{\Lambda}_n\mathbf{U}_n\tilde{\boldsymbol{\rho}}_n \geq \frac{\lambda_{nk}}{r_k} \geq \frac{\lambda_{nk}}{\mu}.$$

The lemma follows. QED

Lemma 6.3 *Let $\check{\boldsymbol{\Gamma}}_{n0}^{\chi}$ and $\check{\boldsymbol{\Gamma}}_{n0}^{\xi}$ denote the common and idiosyncratic covariance matrices, as defined in (3.6). Then, under the assumptions of Proposition 4.1,*

(i) the $r = q(s+1)$ largest eigenvalues of $\check{\boldsymbol{\Gamma}}_{n0}^{\chi}$ go to infinity as $n \to \infty$, and

(ii) the eigenvalues of $\check{\boldsymbol{\Gamma}}_{n0}^{\xi}$ are bounded as $n \to \infty$.

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

$$\begin{aligned} \mathbf{v}\check{\boldsymbol{\Gamma}}_{n0}^{\xi}\tilde{\mathbf{v}} &= \mathbf{v}\int_{-\pi}^{\pi}\check{\boldsymbol{\Sigma}}_n^{\xi}(\theta)d\theta \ \tilde{\mathbf{v}} = \int_{-\pi}^{\pi}\mathbf{v}\check{\boldsymbol{\Sigma}}_n^{\xi}(\theta)\tilde{\mathbf{v}}d\theta \\ &\leq \int_{-\pi}^{\pi}\boldsymbol{\lambda}_{n,q+1}(\theta)d\theta \leq 2\pi\Lambda \end{aligned}$$

by (F3c). Part (ii) of the lemma follows. By the same argument, it follows from (F3a) that $\mathbf{v}\boldsymbol{\Gamma}_{n0}^{\xi}\tilde{\mathbf{v}} \leq 2\pi\Lambda$. Setting $\mathbf{A}_n := \boldsymbol{\Gamma}_{n0}^{\xi} - \check{\boldsymbol{\Gamma}}_{n0}^{\xi}$ and observing that $\mathbf{v}\boldsymbol{\Gamma}_{n0}^{\xi}\tilde{\mathbf{v}}$ cannot be negative, we obtain

$$|\mathbf{v}\mathbf{A}_n\tilde{\mathbf{v}}| = |\mathbf{v}\boldsymbol{\Gamma}_{n0}^{\xi}\tilde{\mathbf{v}} - \mathbf{v}\check{\boldsymbol{\Gamma}}_{n0}^{\xi}\tilde{\mathbf{v}}| \leq 2\pi\Lambda.$$

Since

$$\mathbf{A}_n = \int_{-\pi}^{\pi} \left(\mathbf{\Sigma}_n^{\xi}(\theta) - \check{\mathbf{\Sigma}}_n^{\xi}(\theta) \right) d\theta = \int_{-\pi}^{\pi} \left(\check{\mathbf{\Sigma}}_n^{\chi}(\theta) - \mathbf{\Sigma}_n^{\chi}(\theta) \right) d\theta = \check{\mathbf{\Gamma}}_{n0}^{\chi} - \mathbf{\Gamma}_{n0}^{\chi},$$

it follows that

$$\check{\mathbf{\Gamma}}_{n0}^{\xi} + 2\pi\Lambda\mathbf{I}_n = \mathbf{\Gamma}_{n0}^{\xi} + [2\pi\Lambda\mathbf{I}_n - \mathbf{A}_n].$$

Since the matrix in square brackets is positive semi-definite, Theorem 1 in Lancaster and Tismenetsky (1985, p. 301) implies that the eigenvalues of the sum on the left-hand side are larger than or equal to the corresponding eigenvalues of $\mathbf{\Gamma}_{n0}^{\chi}$. Denoting by $\check{\mu}_{nj}^{\chi}$, $j = 1, \ldots, n$ the eigenvalues of $\check{\mathbf{\Gamma}}_{n0}^{\chi}$, this entails that $\check{\mu}_{nr}^{\chi} + 2\pi\Lambda \geq \mu_{nr}^{\chi}$; part (i) of the lemma thus follows from Assumption (F3b′).

6.2 Proof of Lemma 3.1.

Chamberlain (1983, p. 1312) shows that, for any sequence $a_n := (a_{n1}, \ldots, a_{nn})$ such that $\sum_{i=1}^{n} a_{ni}^2 \to 0$, $\mathbf{a}_n \boldsymbol{\xi}_{nt} \longrightarrow 0$ in quadratic mean as $n \to \infty$ provided that the first eigenvalue of the covariance matrix $\mathbf{\Gamma}_{n0}^{\xi}$ of $\boldsymbol{\xi}_{nt}$ is bounded as $n \to \infty$—a condition that holds in view of Assumption (F3a). It follows that $\mathbf{a}_n \mathbf{x}_{nt}$ converges to $\mathcal{G}(\mathbf{F}, \tau)$ for any $\tau \geq t$. QED

6.3 Spectral estimation.

The first step of our forecasting method relies on the estimation

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

of the spectral density matrix $\mathbf{\Sigma}_n(\theta)$. In Section 4.1 we used a Bartlett window of size $M = [T^{1/3}] + 1$, while in the empirical application of Section 4.2 we used a Bartlett window of $M = 24$, i.e. with weights $w_k = 1 - \frac{|k|}{M+1} = 1 - \frac{|k|}{25}$. 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 3.2, in order to obtain $\mathbf{\Gamma}_{nk}^{\chi T}$ and $\mathbf{\Gamma}_{nk}^{\xi T}$ via the inverse discrete Fourier transforms (see (3.5))

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

with $\mathbf{\Sigma}_n^{\xi T}(\theta) = \mathbf{\Sigma}_n(\theta) - \mathbf{\Sigma}_n^{\chi T}(\theta)$.

In the simulation exercise of Section 4.1 we simply assumed q as known. In Section 4.2 we identified $q = 3$ (as well as $M = 24$) using the information criterion suggested in Liška (2002).

6.4 Stacking.

While the first step of our forecasting method, involving dynamic eigenvalues and eigenvectors, is dynamic, the second step, where projections are made onto a space of static principal components, has a more static flavour. Intuitively, an improvement of predictive performance could be expected from introducing some dynamic features into this second step. This could be achieved, as in Stock and Watson (1999), by stacking the observations made at time $t, t-1, \ldots, t-g$, for some $g \geq 1$. The theory developed (for $g = 0$) in Section 3 goes through with little changes. The consistency results of Section 3.4 hold without stacking, and the empirical study of Section 4 all indicates that stacking does not bring any significant improvement;

We briefly indicate how the proofs of Section 3.4 are to be modified in order to handle such stacking. For some $g \in \mathbb{N}$, define the $n(g+1) \times 1$ stacks $\mathbf{x}_{nt,g} := (\mathbf{x}_{nt}' \; \mathbf{x}_{n,t-1}' \; \ldots \; \mathbf{x}_{n,t-g}')'$. In the same way, stacks $\boldsymbol{\chi}_{nt,g}^T$ and $\boldsymbol{\xi}_{nt,g}^T$ are obtained from the estimates $\boldsymbol{\chi}_{nt}^T$ and $\boldsymbol{\xi}_{nt}^T$, with covariance matrices $\mathbf{\Upsilon}_{n,g}^T$, $\mathbf{\Upsilon}_{n,g}^{\chi T}$, and

$\mathbf{\Upsilon}_{n,g}^{\xi T}$ respectively; these matrices result in an obvious way from rearranging the elements of the covariance matrices $\mathbf{\Gamma}_{nk}^{T}$, $\mathbf{\Gamma}_{nk}^{\chi T}$ and $\mathbf{\Gamma}_{nk}^{\xi T}$. The generalized eigenvalue problem (3.8) then takes the form

$$\begin{aligned} \mathbf{Z}_{n\ell,g}^{T} \quad &:= \quad \text{Arg} \max_{\boldsymbol{a}\in\mathbb{R}^{n(g+1)}} \mathbf{a}\mathbf{\Upsilon}_{n,g}^{\chi T}\tilde{\mathbf{a}} \\ \text{subject to} \quad & \mathbf{a}\mathbf{\Upsilon}_{n,g}^{\xi T}\tilde{\mathbf{a}} = 1 \\ & \mathbf{a}\mathbf{\Upsilon}_{n,g}^{\xi T}\tilde{\mathbf{Z}}_{nm,g}^{T} = 0 \quad \text{for} \quad 1 \le m < \ell\,,\, 1 \le \ell \le n(g+1), \end{aligned} \tag{6.4}$$

leading to the r aggregates $W_{nt,g}^{kT} := \mathbf{Z}_{nk,g}^{T}\mathbf{x}_{nt,g},\ k = 1, \ldots, r\ (1 \le r \le n(g+1))$. The resulting forecasts are $\phi_{i,T+h|T}^{nT} := \left(\mathbf{\Upsilon}_{n,gh}^{\chi T}\mathbf{Z}_{n}^{T}(\tilde{\mathbf{Z}}_{n}^{T}\mathbf{\Upsilon}_{n,g}^{T}\mathbf{Z}_{n}^{T})^{-1}\tilde{\mathbf{Z}}_{n}^{T}\mathbf{x}_{nT,g}\right)_{i}$, where $\mathbf{Z}_{n}^{T} := (\mathbf{Z}_{n1,g}^{T\prime} \ \cdots \ \mathbf{Z}_{nr,g}^{T\prime})'$. Letting $h = 0$ and $T = t$, the same method yields $\phi_{i,t}^{nT} := \left(\mathbf{\Upsilon}_{n0}^{\chi T}\mathbf{Z}_{n}^{T}(\tilde{\mathbf{Z}}_{n}^{T}\mathbf{\Upsilon}_{n,g}^{T}\mathbf{Z}_{n}^{T})^{-1}\tilde{\mathbf{Z}}_{n}^{T}\mathbf{x}_{nt,g}\right)_{i}$.

References

[1] Altissimo, F., A. Bassanetti, R. Cristadoro, M. Forni, M. Lippi, L. Reichlin, and G. Veronese (2001). A real time coincident indicator of the Euro area business cycle, Mimeo (August), Banca d'Italia, Rome.

[2] Bai, J. (2001). Inference on factor models of large dimension. *Econometrica*, to appear.

[3] Bai, J. and S. Ng (2000). Determining the number of factors in approximate factor models. Unpublished manuscript.

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

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

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

[7] Engle, R.F. and M.F. Watson (1981). A one-factor multivariate time series model of metropolitan wage rates. *Journal of the American Statistical Association* **76**, 774-781.

[8] 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.

[9] Forni, M., M. Hallin, M. Lippi, and L. Reichlin (2001). Coincident and leading indicators for the Euro area. *The Economic Journal* **111**, 62-85.

[10] Forni, M., M. Hallin, M. Lippi, and L. Reichlin (2002). The generalized dynamic factor model : consistency and rates. *Journal of Econometrics*, to appear.

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

[12] 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.

[13] 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.

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

[15] 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.

[16] Sargent, T.J. and C.A. Sims (1977). Business cycle modelling 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.

[17] Stock, J.H. and M.W. Watson (1999). Diffusion indexes. Unpublished manuscript.

[18] Wilkinson, J.H. (1965). *The Algebraic Eigenvalue Problem*. Clarendon Press, Oxford.