

Notes on Dynamic Factor Models

Matteo Barigozzi *

April 2009

*ECARES, Université libre de Bruxelles.

This is just a summary of the works cited in the Bibliography, please do not quote it, but quote instead the original papers.

1 Introduction

The main idea underlying factor analysis is that a large set of variables can be explained by a small number of latent variables, the *factors*, which are responsible for all the relevant dynamics. Karl Pearson and Charles Spearman in the 1930s were the first to use this technique when studying the dimension of human intelligence, thus the first developments were obtained in the field of psychometry. In a factor model each observed variable is decomposed into two components: a *common* one that contains the information carried by the factors, and an *idiosyncratic* one which is simply the residual of the decomposition. Furthermore, the relation between the common part of the observable series and the factors is assumed to be linear. This decomposition is accomplished by analysing the covariance matrix of the observable variables. If these are all driven by a common set of factors, then such a structure should be recognizable in the observable relations among observable variables, which are best summarized by the covariance matrix. Usually the largest few eigenvalues of the matrix contain all the essential information. Factor analysis is therefore a technique of dimension reduction that takes the information contained in a large dataset and summarizes it by means of few unobservable common variables.

Geometrically we can think of T observations for n observable variables, where T for us is always a time dimension, but in general it can simply indicate different (cross) observations for each variable. At each point in time the observed variables are represented by a vector in an n -dimensional space. In factor analysis we look for the q directions, in the n -dimensional space, along which the observable variables have maximum variance, computed across the T observations. Such directions are mutually orthogonal and q is smaller than n . At each point in time, we project the n -dimensional vector of observable variables along such directions, thus obtaining the q common factors: one for each direction. The common component is the projection of the data on the q -dimensional subspace generated by the factors. We obtain it by simply taking a linear combination of the factors, with weights depending on the covariance between the observed series and the factors themselves. We move in this way from an n -dimensional space to a q -dimensional subspace, thus reducing the dimension of the problem. The similarity with principal components analysis is clear, and actually the difference is subtle and it often disappears.¹ While principal components are just a descriptive tool used for explaining the variance, factor models are used to explain the covariances among observed variables. There is a causal relation between the factor and the observable data, which is contained in the covariance matrix, and it implies some particular assumptions to be made on the orthogonality of factors and idiosyncratic parts. Concerning the latter ones, we assume here that some mild cross-correlation is present. This class of factor models is called *approximate* or *generalized*. In this dissertation we always deal with time series, therefore we can study the dynamics of the observed processes. We assume that the effect of the factors on observed series can be leading, lagging or contemporaneous. In this sense the models we consider are called *dynamic*.

We consider the so called Generalized Dynamic Factor Model (GDFM) presented in the papers by Forni and Lippi (2001), and Forni, Hallin, Lippi and Reichlin (2000, 2004 and 2005). Other factor models were already introduced in the literature. In particular, Chamberlain and Rothschild (1983) and Chamberlain (1983) consider an approximate factor model but static,

¹For an introduction on principal components and factor analysis see Peña (2002).

while Sargent and Sims (1977) and Geweke (1977) consider a dynamic factor model but with orthogonal idiosyncratic components (i.e. an exact factor model). The closest model to the GDFM is the approximate dynamic factor model by Stock and Watson (2002a and 2002b), and Bai (2003). They however estimate the model in a purely static way without explicitly considering the underlying dynamics of the common factors.

2 Definitions and preliminary assumptions

For every complex matrix or vector \mathbf{x} we denote with \mathbf{x}^\dagger the complex conjugate of its transposed and with \mathbf{x}' the transposed. With Θ we denote the real interval $[-\pi, \pi]$. Let then $\mathcal{P} = (\Omega, \mathcal{F}, P)$ be a probability space and let $L_2(\mathcal{P}, \mathbb{C})$ be the linear space of all complex-valued, zero-mean, square-integrable random variables defined on Ω . This space is a Hilbert space with the inner product defined as $\langle \mathbf{x}, \mathbf{y} \rangle = \mathbb{E}[\mathbf{x}\mathbf{y}^\dagger] = \text{cov}(\mathbf{x}, \mathbf{y})$, and the norm defined as $\|\mathbf{x}\|^2 = \mathbb{E}[\mathbf{x}\mathbf{x}^\dagger] = \text{var}(\mathbf{x})$. We always deal with an infinite double sequence

$$\mathbf{x} = \{x_{it}, i \in \mathbb{N}, t \in \mathbb{Z}\} ,$$

where $x_{it} \in L_2(\mathcal{P}, \mathbb{C})$. With \mathbf{x}_{nt} we denote the n -dimensional column vector $(x_{1t} \dots x_{nt})'$.

Assumption 1 For any $n \in \mathbb{N}$

1. the process \mathbf{x}_{nt} is covariance stationary, i.e. $\mathbb{E}[\mathbf{x}_{nt}\mathbf{x}_{nt-k}^\dagger] = \mathbf{\Gamma}_{nk}^x$;
2. \mathbf{x}_{nt} has a spectral density $\mathbf{\Sigma}_n^x$, which is Hermitian, nonnegative definite for any $\theta \in \Theta$, measurable and

$$\mathbf{\Sigma}_n^x(\theta) = \sum_{k=-\infty}^{+\infty} e^{-ik\theta} \mathbf{\Gamma}_{nk}^x \quad \text{or} \quad \mathbf{\Gamma}_{nk}^x = \frac{1}{2\pi} \int_{-\pi}^{\pi} e^{ik\theta} \mathbf{\Sigma}_n^x(\theta) d\theta .$$

With $\mathbf{\Sigma}^x$ we denote the spectral density of the infinite sequence \mathbf{x}_t . In what follows Assumption 1 always holds. The i -th largest eigenvalue of $\mathbf{\Sigma}_n^x(\theta)$, which we call dynamic eigenvalue, is denoted as λ_{ni}^x and it is a real valued function defined on Θ . It is a nondecreasing function of n for any θ . Moreover, for any i and θ we have that $\lim_{n \rightarrow \infty} \lambda_{ni}^x(\theta) = \lambda_i^x(\theta)$, which is the i -th largest eigenvalue of $\mathbf{\Sigma}^x(\theta)$.

We define the Hilbert space $L_2^\infty(\Theta, \mathbb{C}, \mathbf{\Sigma}^x)$ as the space of complex-valued infinite dimensional functions \mathbf{f} such that their components f_i are measurable on Θ and

$$\lim_{n \rightarrow \infty} \frac{1}{2\pi} \int_{-\pi}^{\pi} \mathbf{f}^{(n)} \mathbf{\Sigma}^x(\theta) \mathbf{f}^{(n)\dagger} d\theta = \frac{1}{2\pi} \int_{-\pi}^{\pi} \mathbf{f} \mathbf{\Sigma}^x(\theta) \mathbf{f}^\dagger d\theta = \|\mathbf{f}\|_{\mathbf{\Sigma}^x}^2 < \infty .$$

Analogously, we define $L_2^\infty(\Theta, \mathbb{C})$ where now we replace $\mathbf{\Sigma}^x$ with the infinite dimensional identity matrix, and $L_2^n(\Theta, \mathbb{C}, \mathbf{\Sigma}^x)$ and $L_2^n(\Theta, \mathbb{C})$ where we replace infinite dimensional vectors and matrices with n -dimensional ones.

If we introduce the isomorphism $\Xi : \overline{\text{span}}(\mathbf{x}) \rightarrow L_2^\infty(\Theta, \mathbb{C}, \mathbf{\Sigma}^x)$, this operator assures that every process belonging to $\overline{\text{span}}(\mathbf{x})$ and costationary with \mathbf{x} has a spectral density. For every function $\mathbf{f} \in L_2^\infty(\Theta, \mathbb{C}, \mathbf{\Sigma}^x)$ we then define the filtered process

$$\underline{\mathbf{f}}(L)\mathbf{x}_t = \Xi^{-1}(\mathbf{f}e^{it\cdot}) ,$$

which has as spectral density $\mathbf{f} \Sigma^x \mathbf{f}^\dagger$, and where L is the lag operator such that $L\mathbf{x}_t = \mathbf{x}_{t-1}$.

Finally, we add four definitions.

Definition 1 (Dynamic Averaging Sequence DAS) $\{\mathbf{a}_n, n \in \mathbb{N}\}$ such that $\mathbf{a}_n \in L_2^\infty(\Theta, \mathbb{C}) \cap L_2^\infty(\Theta, \mathbb{C}, \Sigma^x)$, is a DAS if

$$\lim_{n \rightarrow \infty} \frac{1}{2\pi} \int_{-\pi}^{\pi} \mathbf{a}_n(\theta) \mathbf{a}_n^\dagger(\theta) d\theta = 0 ,$$

i.e. the L_2 -norm of the sequence vanishes as n goes to infinity.

Definition 2 (Aggregate) Let $\mathbf{y}_t \in \overline{\text{span}}(\mathbf{x})$. We say that \mathbf{y}_t is an aggregate if there exists a DAS $\{\mathbf{a}_n, n \in \mathbb{N}\}$ such that

$$\lim_{n \rightarrow \infty} \underline{\mathbf{a}}_n(L)\mathbf{x}_t = \mathbf{y}_t .$$

The set of all aggregates is denoted as $\mathcal{G}(\mathbf{x})$ and it is a closed subspace of $\overline{\text{span}}(\mathbf{x})$.²

Definition 3 (Canonical Decomposition) The canonical decomposition of \mathbf{x} is, for any $i \in \mathbb{N}$ and $t \in \mathbb{Z}$,

$$x_{it} = \text{Proj}(x_{it} | \mathcal{G}(\mathbf{x})) + \delta_{it} .$$

Definition 4 (Idiosyncratic process) We say that \mathbf{x} is idiosyncratic if

$$\lim_{n \rightarrow \infty} \underline{\mathbf{a}}_n(L)\mathbf{x}_t = 0 ,$$

for any DAS $\{\mathbf{a}_n, n \in \mathbb{N}\}$. In this case we have that $\mathcal{G}(\mathbf{x}) = \{0\}$ and $x_{it} = \delta_{it}$.

3 On the existence of the Dynamic Factor representation

Given the previous definitions and Assumption 1, Forni and Lippi (2001) state the conditions in population under which the Generalized Dynamic Factor structure exists. We review briefly their results in order to summarize the theoretical foundations that lie beyond the study of dynamic factor models.

The first theorem concerns idiosyncratic processes as defined in definition 4.

Theorem 1 *The following two statements are equivalent:*

1. \mathbf{x} is idiosyncratic;
2. λ_1^x is essentially bounded, i.e. there exists a real constant c such that $|\lambda_1^x(\theta)| < c$ almost everywhere in Θ .

We then define the Generalized Dynamic Factor Model representation for a process $\mathbf{x} \in L_2(\mathcal{P}, \mathbb{C})$ for which Assumption 1 holds.

²As an example consider the series

$$\mathbf{d}_n = \frac{1}{n}(1 \dots 1 \ 0 \dots)$$

which is a DAS that produces arithmetic averages. Notice that dynamic averaging of \mathbf{x} is averaging simultaneously both in the cross section and in the time dimension.

Definition 5 (Generalized Dynamic Factor Model - Definition) Let q be a nonnegative integer. The process \mathbf{x} is a q -dynamic factor sequence (q -DFS) if $L_2(\mathcal{P}, \mathbb{C})$ contains an orthonormal q -dimensional white noise vector process

$$\mathbf{u} = \{u_{jt}, j = 1, \dots, q, t \in \mathbb{Z}\}$$

and a double sequence

$$\boldsymbol{\xi} = \{\xi_{it}, i \in \mathbb{N}, t \in \mathbb{Z}\},$$

fulfilling Assumption 1, such that:

1. we have a new double sequence

$$\boldsymbol{\chi} = \{\chi_{it}, i \in \mathbb{N}, t \in \mathbb{Z}\},$$

defined as

$$\mathbf{x}_t = \boldsymbol{\chi}_t + \boldsymbol{\xi}_t \tag{1}$$

$$\boldsymbol{\chi}_t = \underline{\mathbf{b}}_1(L)u_{1t} + \dots + \underline{\mathbf{b}}_q(L)u_{qt} = \underline{\mathbf{B}}(L)\mathbf{u}_t$$

where $\underline{\mathbf{B}}(L) = (\underline{\mathbf{b}}_1(L) \dots \underline{\mathbf{b}}_q(L))$. These filters are such that each entry $b_{ij} \in L_2(\Theta, \mathbb{C})$ for any $i \in \mathbb{N}$ and $j = 1, \dots, q$;

2. for any $i \in \mathbb{N}$, $j = 1, \dots, q$, $k \in \mathbb{Z}$, we have $\xi_{it} \perp u_{jt-k}$, i.e. $E[\xi_{it}u_{jt-k}^\dagger] = 0$;
3. λ_1^ξ is essentially bounded, i.e. ξ is idiosyncratic;
4. $\lambda_q^\chi(\theta) = \infty$ a.e. in Θ .

The processes $\boldsymbol{\chi}$ and $\boldsymbol{\xi}$ are called respectively common and idiosyncratic component, while \mathbf{u} is the vector of dynamic factors or common shocks.

The three theorems that follow state the conditions under which a q -DFS exists. The proofs of the theorems are not always trivial and are given in Forni and Lippi (2001). The first one is the fundamental theorem for the existence dynamic factor models.

Theorem 2 The process \mathbf{x} is a q -DFS if and only if

1. λ_{q+1}^x is essentially bounded;
2. $\lambda_q^x(\theta) = \infty$ a.e. in Θ .

Theorem 3 If \mathbf{x} is a q -DFS and representation (1) holds then

$$\overline{\text{span}}(\boldsymbol{\chi}) = \overline{\text{span}}(\mathbf{u}) = \mathcal{G}(\mathbf{x}) \quad \text{and} \quad \chi_{it} = \text{Proj}(x_{it} | \mathcal{G}(\mathbf{x})).$$

Thus, given definitions 5.2 and 5.3,

$$\chi_{it} = \text{Proj}(\chi_{it} | \mathcal{G}(\mathbf{x})).$$

Theorem 4 If \mathbf{x} is a q -DFS and representation (1) holds then q , $\boldsymbol{\chi}$, $\boldsymbol{\xi}$ are uniquely determined. However, \mathbf{u} and $\underline{\mathbf{B}}(L)$ are not unique, indeed it is always possible to write $\boldsymbol{\chi}_t = \underline{\mathbf{D}}(L)\mathbf{w}_t$ with $\mathbf{w}_t = \underline{\boldsymbol{\Omega}}(L)\mathbf{u}_t$ and $\mathbf{D} = \mathbf{B}\boldsymbol{\Omega}^\dagger$, provided that $\boldsymbol{\Omega}^\dagger\boldsymbol{\Omega} = \mathbf{I}_q$, and $\boldsymbol{\Omega} \in L_2^{q \times q}(\Theta, \mathbb{C})$.

4 Identification of the model

Given the previous definition of the Generalized Dynamic Factor Model, we can state the assumptions that are needed for (1) to hold. Following Forni et al. (2000) we require four assumptions. Hereafter we consider the Hilbert space $L_2(\Omega, \mathcal{F}, P, \mathbb{R})$ of all real-valued, zero-mean, square-integrable random variables defined on Ω . We still suppose that Assumption 1 holds.

Assumption 2 (Generalized Dynamic Factor Model - Assumptions) *Given the double sequence*

$$\mathbf{x} = \{x_{it} \in L_2(\Omega, \mathcal{F}, P, \mathbb{R}), i \in \mathbb{N}, t \in \mathbb{Z}\} ,$$

and representation (1)

$$\mathbf{x}_t = \boldsymbol{\chi}_t + \boldsymbol{\xi}_t = \underline{\mathbf{B}}(L)\mathbf{u}_t ,$$

we suppose that the following assumptions hold.

1. (a) The q -dimensional process $\mathbf{u} = \{u_{jt}, j = 1, \dots, q, t \in \mathbb{Z}\}$ is orthonormal white noise. That is $E[u_{jt}] = 0$, $E[u_{jt}u'_{jt}] = 1$ and $u_{jt} \perp u_{ht-k}$, for any $j, h = 1, \dots, q$ and $k \in \mathbb{Z}$.
- (b) The idiosyncratic component $\boldsymbol{\xi} = \{\xi_{it}, i \in \mathbb{N}, t \in \mathbb{Z}\}$ is a double sequence such that $\boldsymbol{\xi}_n$ is a zero-mean stationary process for any $n \in \mathbb{N}$, and $\xi_{it} \perp u_{jt-k}$ for any $i \in \mathbb{N}$, $j = 1, \dots, q$, $k \in \mathbb{Z}$, i.e. $E[\xi_{it}u'_{jt-k}] = 0$.
- (c) The entries of the matrix $\underline{\mathbf{B}}(L)$ are one-sided filters $\underline{b}_{ij}(L)$, with square-summable coefficients, for any $i \in \mathbb{N}$ and $j = 1, \dots, q$.
2. For any $i \in \mathbb{N}$, there exists a real $c_i > 0$ such that the entries of the spectral density of $\boldsymbol{\Sigma}_n^x$ are bounded, i.e. $\sigma_{ii}(\theta) \leq c_i$ for any $\theta \in \Theta$. This implies also that the entries $\sigma_{ij}(\theta)$ are bounded in modulus.
3. The first idiosyncratic dynamic eigenvalue λ_{n1}^ξ is uniformly bounded, i.e. there exists a real $c > 0$ such that $|\lambda_{n1}^\xi(\theta)| \leq c$ for any $\theta \in \Theta$, and any $n \in \mathbb{N}$.
4. The first q common dynamic eigenvalues diverge almost everywhere in Θ , i.e.

$$\lim_{n \rightarrow \infty} \lambda_{nj}^x(\theta) = \infty \quad \text{for } j = 1, \dots, q \quad \text{a.e. in } \Theta .$$

Assumption 2.3 is clearly satisfied for processes that are mutually orthogonal at any lead and lag and have uniformly bounded spectral densities, but it is more general since it allows for a limited amount of dynamic cross-correlation. Assumption 2.4 implies that each dynamic factor is present in infinitely many cross-sectional units, with nondecreasing importance, while Assumption 2.3 implies that idiosyncratic shocks, although possibly shared by many different units, have their effects concentrated on a finite number of them and tending to zero as the number of units tends to infinity. Theorem 2 is the fundamental result in the theory of the Generalized Dynamic Factor Model, it transforms statements on the common and idiosyncratic dynamic eigenvalues which are unobservable, into statements on the dynamic eigenvalues of \mathbf{x}_{nt} , which is observable. Under Assumption 2 the common and idiosyncratic components are identified, as stated in theorem 4, and can be consistently estimated. However, nothing can be said on the identification of the common shocks for which economic meaningful restrictions should be imposed. In this respect the assumption of one-sided filters is necessary only when we want to identify the factors, while it can be safely dropped in all other cases.

5 The two-sided estimator

We now move to the estimation of the common component. From theorem 3 we have that $\boldsymbol{\chi}_t \in \mathcal{G}(\mathbf{x})$, therefore the common component is an aggregate of \mathbf{x} . Moreover, this means that there exist a sequence of integers s_n and a sequence $\{\mathbf{a}_n, n \in \mathbb{N}\}$ such that $\mathbf{a}_n \in L_2^{s_n}(\Theta, \mathbb{C}) \cap L_2^{s_n}(\Theta, \mathbb{C}, \boldsymbol{\Sigma}^x)$ such that

$$\lim_{n \rightarrow \infty} \underline{\mathbf{a}}_n(L) \mathbf{x}_{s_n t} = \boldsymbol{\chi}_t .$$

Given the representation (1) for which $\boldsymbol{\xi}$ is an idiosyncratic process, we look for a DAS such that

$$\begin{aligned} \lim_{n \rightarrow \infty} \underline{\mathbf{a}}_n(L) \boldsymbol{\chi}_{s_n t} &= \boldsymbol{\chi}_t , \\ \lim_{n \rightarrow \infty} \underline{\mathbf{a}}_n(L) \boldsymbol{\xi}_{s_n t} &= 0 . \end{aligned} \tag{2}$$

Thus, assuming $s_n = n$, we are just looking for a dynamic average of \mathbf{x}_{nt} (or filters) such that, as n goes to infinity, the idiosyncratic part is wiped out, while the dynamic factors are present with nondecreasing importance through the common component. Forni et al. (2000) prove that the q largest dynamic principal components span the same space as the one spanned by the first q dynamic factors. Therefore, given theorem 3, the common component is estimated as the projection of \mathbf{x}_{nt} on all leads and lags of its q largest dynamic principal components.

We briefly review this estimation method. Forni et al. (2000) propose an estimator of $\boldsymbol{\chi}_t$ when starting from a finite process $\mathbf{x}_n^T = \{x_{it}, i = 1, \dots, n, t = 1, \dots, T\}$. Consistency for such an estimator is proved in two steps. Firstly, it is proved that the estimated common component $\boldsymbol{\chi}_{nt}$ converges to $\boldsymbol{\chi}_t$ in mean square as n goes to infinity. Secondly, it is proved that the finite sample counterpart, $\boldsymbol{\chi}_{nt}^T$, of the previously estimated common component, is a consistent estimator of $\boldsymbol{\chi}_{nt}$ for any n , as T goes to infinity. Combining the two results gives consistency of the proposed estimator. Here we are not interested in building finite sample estimators of the common component, therefore we assume that $t \in \mathbb{Z}$ and we consider later the case of samples with finite time dimension, for which we need consistent sample estimators of the covariance and spectral density matrices.

Given the spectral density $\boldsymbol{\Sigma}_n^x$ of \mathbf{x}_{nt} there exist n vectors of complex valued functions $\mathbf{p}_{ni}^x \in L_2^n(\Theta, \mathbb{C}) \cap L_2^n(\Theta, \mathbb{C}, \boldsymbol{\Sigma}^x)$, for $i = 1, \dots, n$, such that

1. they are eigenvectors of $\boldsymbol{\Sigma}_n^x(\theta)$, corresponding to the dynamic eigenvalues in decreasing order, i.e.

$$\mathbf{p}_{ni}^x(\theta) \boldsymbol{\Sigma}_n^x(\theta) = \lambda_{ni}^x(\theta) \mathbf{p}_{ni}^x(\theta) \quad \text{for any } \theta \in \Theta ,$$

and as usual $\lambda_{ni+1}^x(\theta) > \lambda_{ni}^x(\theta)$ for any $\theta \in \Theta$;

2. they have unitary length, i.e. $|\mathbf{p}_{ni}^x(\theta)|^2 = 1$ for any $\theta \in \Theta$;
3. they are orthogonal, i.e. $\mathbf{p}_{ni}^x(\theta) \mathbf{p}_{nj}^{x\dagger}(\theta) = 0$ for any $i \neq j$ and any $\theta \in \Theta$.

Such vectors are the dynamic eigenvectors associated with \mathbf{x}_{nt} . They can be expanded in

Fourier series and the corresponding linear filter can be built:

$$\begin{aligned}\mathbf{p}_{ni}^x(\theta) &= \frac{1}{2\pi} \sum_{k=-\infty}^{+\infty} \left[\int_{-\pi}^{\pi} \mathbf{p}_{ni}^x(\theta) e^{ik\theta} d\theta \right] e^{-ik\theta}, \\ \underline{\mathbf{p}}_{ni}^x(L) &= \frac{1}{2\pi} \sum_{k=-\infty}^{+\infty} \left[\int_{-\pi}^{\pi} \mathbf{p}_{ni}^x(\theta) e^{ik\theta} d\theta \right] L^k.\end{aligned}$$

For $i = 1, \dots, n$, the scalar process $\{\underline{\mathbf{p}}_{ni}^x(L)\mathbf{x}_{nt}, t \in \mathbb{Z}\}$ is the i -th dynamic principal component. Notice that, given property 3 of the dynamic eigenvectors, the dynamic principal components are orthogonal. Since $\mathbf{p}_{ni}^x \in L_2^n(\Theta, \mathbb{C})$, the dynamic eigenvectors are an orthonormal system of eigenvectors for the n -dimensional identity. Therefore

$$\mathbf{x}_{nt} = \sum_{i=1}^n \underline{\mathbf{p}}_{ni}^{x\dagger}(L) \underline{\mathbf{p}}_{ni}^x(L) \mathbf{x}_{nt}.$$

As said above, the common component χ_{nt} is simply the projection of \mathbf{x}_{nt} on the space spanned by the q -largest dynamic principal components, thus

$$\chi_{nt} = \sum_{i=1}^q \underline{\mathbf{p}}_{ni}^{x\dagger}(L) \underline{\mathbf{p}}_{ni}^x(L) \mathbf{x}_{nt} = \underline{\mathbf{K}}_n(L) \mathbf{x}_{nt}. \quad (3)$$

Forni et al. (2000) prove that this linear combination of the q largest dynamic principal components is the DAS that we are looking for, therefore

$$\lim_{n \rightarrow \infty} \underline{\mathbf{K}}_n(L) \mathbf{x}_{nt} = \chi_t.$$

The proof is not trivial and we limit ourselves to a simple example. Consider the one-factor process $\mathbf{x}_t = \mathbf{u}_t + \boldsymbol{\xi}_t$, with white noise orthonormal idiosyncratic parts. The first dynamic eigenvector of the finite process \mathbf{x}_{nt} is $\underline{\mathbf{p}}_{n1}^x(L) = (1/\sqrt{n} \dots 1/\sqrt{n})$, so that the i -th row of the common component is estimated as

$$\chi_{n,it} = \underline{\mathbf{K}}_{ni}(L) \mathbf{x}_{nt} = (1/n \dots 1/n) \mathbf{x}_{nt} = \mathbf{u}_t + \frac{1}{n} \sum_{j=1}^n \xi_{jt}.$$

The DAS in this case is simply the arithmetic mean of \mathbf{x}_{nt} . In this case it is clear that as n goes to infinity only the first term survives and it is the common component, while the idiosyncratic part disappears as required.

6 The static representation and the one-sided estimator

The previous estimator is built using leads and lags of the dynamic principal components. The trouble with such an estimator is that the filters used in (3) are two-sided. Although this leaves unaffected the estimate of the central part of the sample, the performance of the estimator deteriorates as we approach the end of the sample. Thus, the presence of two-sided filters makes this method not suitable for prediction. Stock and Watson (2002a) propose an estimation method based on principal components of contemporaneous values of \mathbf{x}_{nt} , which we call static principal components, that can also be used for forecasting. Forni et al. (2005)

(hereafter FHLR) combine the advantages of the methods by Forni et al. (2000) and Stock and Watson (2002a) and propose a two-steps one-sided estimator.

In the first the dynamic method by Forni et al. (2000) is used to obtain estimates of the covariance matrices of the common and idiosyncratic components. In the second step these covariances are used to produce:

1. a new estimation of the factor space as the space spanned by the generalized principal components, which are linear combinations of contemporaneous values of $\boldsymbol{\chi}_{nt}$ with minimum idiosyncratic to common variance ratio;
2. a new estimation of the common component as the projection of $\boldsymbol{\chi}_{nt}$ itself on the factor space, and a prediction of the common component based on its lagged covariance matrix.

Once again consistency is proved for n and t going to infinity. As the cross-size diverges, the idiosyncratic component cancels out, and the true common component is recovered. The same consistency result holds also for the estimator by Stock and Watson. However, their approach differs from the one by FHLR in that they do not have the first step, and just use static principal components in the second step. Stock and Watson therefore do not take into account the effective dynamics that lies beyond the data and do not minimize the error associated with the idiosyncratic part.

Static Principal Components

In order to appreciate the differences between the two approaches, we first consider the estimator by Stock and Watson and then we review the two-steps one-sided estimator by FHLR. Both methods require the static representation of the GDFM (1). We define the dynamic loadings as the matrix $\underline{\mathbf{B}}(L)$, whose generic entry is the linear filter $\underline{b}_{ij}(L)$. We assume that these filters have a maximum lag $s \geq 0$ for any $i \in \mathbb{N}$ and any $j = 1, \dots, q$. If we define the static factors as $\mathbf{F}_t = (\mathbf{u}'_t \dots \mathbf{u}'_{t-s})'$ and the loadings as $\mathbf{\Lambda} = (\mathbf{B}_0 \dots \mathbf{B}_s)$, we have the equivalent static representation

$$\mathbf{x}_t = \mathbf{\Lambda} \mathbf{F}_t + \boldsymbol{\xi}_t. \quad (4)$$

The common component is now a linear combination of the contemporaneous values of r static factors, where $r = q(s + 1)$. In the model by Stock and Watson equation (4) is given without any assumption on any underlying dynamic factor. Here instead we assume (1) and therefore we require that the spectral density of \mathbf{F}_t has reduced rank $q \leq r$, thus we keep explicitly into consideration the dynamic nature of our model. When dealing with the static representation of the GDFM we have a result on the eigenvalues that is the analogous of assumptions 2.3 and 2.4.

Lemma 1 *Denote with μ_{nj}^χ and μ_{nj}^ξ the j -th largest eigenvalue of $\mathbf{\Gamma}_{n0}^\chi$ and $\mathbf{\Gamma}_{n0}^\xi$ respectively. Under assumptions 1 and 2 the following hold:*

1. *the first r common eigenvalues diverge*

$$\lim_{n \rightarrow \infty} \mu_{nj}^\chi = \infty \quad \text{for } 1 \leq j \leq r$$

2. *the largest idiosyncratic eigenvalue μ_{n1}^ξ is bounded for n going to infinity.*

Stock and Watson estimate the static factors as the r largest static principal components of the finite process \mathbf{x}_{nt} . The first principal component is the linear combination of the observed variables that has maximum variance. It is defined as the vector $\mathbf{g}_{n1t} = \mathbf{s}_{n1}\mathbf{x}_{nt}$. The $1 \times n$ row vector \mathbf{s}_{n1} contains weights of the linear combination. Given that data are always zero-mean, the variance of the first principal component is

$$E[\mathbf{g}_{n1t}\mathbf{g}'_{n1t}] = E[\mathbf{s}_{n1}\mathbf{x}_{nt}\mathbf{x}'_{nt}\mathbf{s}'_{n1}] = \mathbf{s}_{n1}\mathbf{\Gamma}_{n0}^x\mathbf{s}'_{n1} .$$

We can in principle increase the variance without limit unless we impose the normalization constraint $\mathbf{s}_{n1}\mathbf{s}'_{n1} = 1$. The Lagrangian of the constrained maximization problem is

$$\mathcal{L} = \mathbf{s}_{n1}\mathbf{\Gamma}_{n0}^x\mathbf{s}'_{n1} - \mu_n(\mathbf{s}_{n1}\mathbf{s}'_{n1} - 1) .$$

The first order conditions are

$$\frac{\partial \mathcal{L}}{\partial \mathbf{s}_{n1}} = 2\mathbf{s}_{n1}\mathbf{\Gamma}_{n0}^x - 2\mu_n\mathbf{s}_{n1} = 0 ,$$

which have as a solution

$$\mathbf{s}_{n1}\mathbf{\Gamma}_{n0}^x = \mu_n\mathbf{s}_{n1} .$$

Thus \mathbf{s}_{n1} is an eigenvector of the covariance matrix of observed data, and μ_n the corresponding eigenvalue. If we multiply on the right by \mathbf{s}'_{n1}

$$\mathbf{s}_{n1}\mathbf{\Gamma}_{n0}^x\mathbf{s}'_{n1} = \mu_n\mathbf{s}_{n1}\mathbf{s}'_{n1} ,$$

and to maximize the variance of the first principal component we must take μ_n as the largest eigenvalue of $\mathbf{\Gamma}_{n0}^x$.

When we introduce the second principal component we require the variance of the sum of the two to be maximum. The Lagrangian is

$$\mathcal{L} = \mathbf{s}_{n1}\mathbf{\Gamma}_{n0}^x\mathbf{s}'_{n1} + \mathbf{s}_{n2}\mathbf{\Gamma}_{n0}^x\mathbf{s}'_{n2} - \mu_{n1}(\mathbf{s}_{n1}\mathbf{s}'_{n1} - 1) - \mu_{n2}(\mathbf{s}_{n2}\mathbf{s}'_{n2} - 1) .$$

From first order conditions and given the normalization constraint in its maximum the Lagrangian has value $\mu_{n1} + \mu_{n2}$. This value is maximized if we take the two largest eigenvalues of the covariance matrix of observed data. The second principal component is orthogonal to the first, indeed

$$\mathbf{g}_{n1t}\mathbf{g}'_{n2t} = \mathbf{s}_{n1}\mathbf{\Gamma}_{n0}^x\mathbf{s}'_{n2} = \mu_{n1}\mathbf{s}_{n1}\mathbf{s}'_{n2} = 0 ,$$

because eigenvectors are by definition orthogonal. By iterating this procedure r times we get all the required principal components. Computing the eigenvectors of the covariance of \mathbf{x}_{nt} is equivalent to solve a maximization problem. We look for r uncorrelated linear combinations $\mathbf{g}_{njt} = \mathbf{s}_{nj}\mathbf{x}_{nt}$, such that the weights are orthonormal row vectors.

These are the eigenvectors of the covariance matrix or, alternatively, they can be defined as

$$\begin{aligned} \mathbf{s}_{nj} &= \arg \max_{\mathbf{y} \in \mathbb{R}^n} \mathbf{y}\mathbf{\Gamma}_{n0}^x\mathbf{y}' \\ \text{s.t.} & \quad \mathbf{y}\mathbf{y}' = 1 \\ \text{and} & \quad \mathbf{y}\mathbf{s}'_{ni} = 0 \quad \text{for } 1 \leq i < j , \end{aligned}$$

for $j = 1, \dots, r$. As we proved above for $r = 2$, the solutions of this maximization problem are the eigenvectors corresponding to the r largest eigenvalues of $\mathbf{\Gamma}_{n0}^x$. We define the vector

of principal components as $\mathbf{G}_{nt} = (\mathbf{g}_{n1t} \dots \mathbf{g}_{nrt})'$. The space they span is identified and it is the same space as the one spanned by the static factors \mathbf{F}_t , although these latter are not identified. Indeed, $\mathbf{F}_t = \mathbf{\Omega} \mathbf{G}_{nt}$ for some orthogonal transformation $\mathbf{\Omega}$.

According to theorem 3 we should define the static principal components as $\mathbf{S}_n \boldsymbol{\chi}_{nt}$ so that they identify a basis for $\overline{\text{span}}(\boldsymbol{\chi})$. However, given that we do not have an estimate of the covariance matrix of the common component we cannot apply theorem 3. It can be proved that \mathbf{G}_{nt} , as defined above, is a consistent estimate for a basis of $\overline{\text{span}}(\boldsymbol{\chi})$ as n and T diverge (see Bai (2003)). Thus, in this case, the estimated common component is given by the projection of \mathbf{x}_{nt} on the space spanned by the static principal components. Since $\mathbf{x}_{nt} = \mathbf{\Lambda}_n \mathbf{G}_{nt} + \boldsymbol{\xi}_{nt}$ then, by multiplying on the right by \mathbf{G}'_{nt} , and remembering that the idiosyncratic part is orthogonal to the factors, the loadings are $\mathbf{\Lambda}_n = \text{E}[\mathbf{x}_{nt} \mathbf{G}'_{nt}] \text{E}[\mathbf{G}_{nt} \mathbf{G}'_{nt}]^{-1}$. They are the regression coefficients when we want to explain the observed variables with unobserved factors. Thus, since $\mathbf{G}_{nt} = \mathbf{S}_n \boldsymbol{\chi}_{nt}$

$$\boldsymbol{\chi}_{nt} = \mathbf{\Gamma}_{n0}^x \mathbf{S}'_n (\mathbf{S}_n \mathbf{\Gamma}_{n0}^x \mathbf{S}'_n)^{-1} \mathbf{S}_n \mathbf{x}_{nt}, \quad (5)$$

where we collect the weights into the $r \times n$ matrix $\mathbf{S}_n = (\mathbf{s}_{n1} \dots \mathbf{s}_{nr})'$. Consistency of the estimator (5) as n goes to infinity can be proved (see Bai (2003)), while in the next section we consider the result for samples of finite size T .

Finally, define $\mathbf{\Theta}_n$ as the diagonal matrix containing the r largest eigenvalues of the covariance of \mathbf{x}_{nt} . Then $\mathbf{S}_n \mathbf{\Gamma}_{n0}^x = \mathbf{\Theta}_n \mathbf{S}_n$, from (5) we have the normalization of the loadings

$$\mathbf{\Lambda}'_n \mathbf{\Lambda}_n = \mathbf{\Theta}_n^{-1} \mathbf{S}_n \mathbf{\Gamma}_{n0}^x \mathbf{\Gamma}_{n0}^x \mathbf{S}'_n \mathbf{\Theta}_n^{-1} = \mathbf{\Theta}_n^{-1} \mathbf{\Theta}_n \mathbf{S}_n \mathbf{S}'_n \mathbf{\Theta}_n^{-1} = \mathbf{I}_r,$$

where we use the normalization of the eigenvectors $\mathbf{S}_n \mathbf{S}'_n = \mathbf{I}_r$. From this relation $\mathbf{\Lambda}_n = \mathbf{S}'_n$, and, with reference to (2), we have the trivial result

$$\lim_{n \rightarrow \infty} [\mathbf{S}'_n \mathbf{S}_n] \mathbf{x}_{nt} = \lim_{n \rightarrow \infty} \boldsymbol{\chi}_{nt} = \boldsymbol{\chi}_t.$$

The one-sided two-step estimator

The method proposed by FHLR has two steps. In the first step the spectral density of \mathbf{x}_{nt} is estimated and then the q largest dynamic eigenvalues and corresponding eigenvectors are used to build the estimator for the spectral density of the common components. With the notation of the previous section

$$\boldsymbol{\Sigma}_n^x(\theta) = \sum_{j=1}^q \mathbf{p}_{nj}^{x\dagger}(\theta) \lambda_{nj}^x(\theta) \mathbf{p}_{nj}^x(\theta).$$

The contemporaneous and lagged covariance matrices of the common component are obtained by using the inverse Fourier transform

$$\mathbf{\Gamma}_{nk}^x = \frac{1}{2\pi} \int_{-\pi}^{\pi} e^{ik\theta} \boldsymbol{\Sigma}_n^x(\theta) d\theta,$$

and $\mathbf{\Gamma}_{nk}^\xi = \mathbf{\Gamma}_{nk}^x - \mathbf{\Gamma}_{nk}^x$.

In the second step these matrices are used to build the generalized principal components. Since now we have an estimate for the covariance of the common component we can apply

theorem 3. As a basis for $\overline{\text{span}}(\boldsymbol{\chi})$, we take r uncorrelated linear combinations such that the weights solve the maximization problem

$$\begin{aligned} \mathbf{z}_{nj} &= \arg \max_{\mathbf{y} \in \mathbb{R}^n} \mathbf{y} \boldsymbol{\Gamma}_{n0}^{\chi} \mathbf{y}' \\ \text{s.t.} & \quad \mathbf{y} \boldsymbol{\Gamma}_{n0}^{\xi} \mathbf{y}' = 1 \\ \text{and} & \quad \mathbf{y} \boldsymbol{\Gamma}_{n0}^{\xi} \mathbf{z}'_{ni} = 0 \quad \text{for } 1 \leq i < j, \end{aligned}$$

for $j = 1, \dots, r$. The idea is that the information contained in the covariance matrices estimated in the previous step can be used to determine linear combinations which are more efficient than standard principal components. It is possible to prove that the weights that solve the problem are the orthonormal eigenvectors corresponding to the r largest eigenvalues of $\boldsymbol{\Gamma}_{n0}^{\chi} (\boldsymbol{\Gamma}_{n0}^{\xi})^{-1}$. We define the generalized principal components as $\mathbf{Z}_n \boldsymbol{\chi}_{nt} = (\mathbf{z}_{n1} \boldsymbol{\chi}_{nt} \dots \mathbf{z}_{nr} \boldsymbol{\chi}_{nt})'$. As before they identify the factor space although the factors are not identified. We can apply theorem 3 and thus we define the common component as the projection of itself on the factor space. Hence $\boldsymbol{\chi}_{nt} = \boldsymbol{\Lambda}_n \mathbf{Z}_n \boldsymbol{\chi}_{nt}$ and the loadings are $\boldsymbol{\Lambda}_n = \text{E}[\boldsymbol{\chi}_{nt} \boldsymbol{\chi}'_{nt} \mathbf{Z}'_n] \text{E}[\mathbf{Z}_n \boldsymbol{\chi}_{nt} \boldsymbol{\chi}'_{nt} \mathbf{Z}'_n]^{-1}$. Therefore the common component is estimated as

$$\boldsymbol{\chi}_{nt} = \boldsymbol{\Gamma}_{n0}^{\chi} \mathbf{Z}'_n (\mathbf{Z}_n \boldsymbol{\Gamma}_{n0}^{\chi} \mathbf{Z}'_n)^{-1} \mathbf{Z}_n \mathbf{x}_{nt}. \quad (6)$$

Notice, however, that in this estimator we cannot use the generalized principal components, as 3 would require, because we would need a preliminary estimate of the common part. Therefore we substitute them with their consistent estimate $\mathbf{Z}_n \mathbf{x}_{nt}$. The consistency of (6) for n diverging is proved by Forni et al. (2005), while in the next section we consider the result for samples of finite size T .

By analogy with the case of static principal components, we can prove that $\boldsymbol{\Lambda}_n = (\boldsymbol{\Gamma}_{n0}^{\xi})^{-1/2} \mathbf{Z}'_n$. Therefore with reference to (2),

$$\lim_{n \rightarrow \infty} \left[(\boldsymbol{\Gamma}_{n0}^{\xi})^{-1/2} \mathbf{Z}'_n \mathbf{Z}_n \right] \boldsymbol{\chi}_{nt} = \boldsymbol{\chi}_t.$$

7 Consistent sample estimators and forecasting

In practice we always deal with finite samples $\mathbf{x}_n^T = \{x_{it}, i = 1, \dots, n, t = 1, \dots, T\}$, therefore, in order to apply the previous estimation methods, we need consistent sample estimators for the covariance and spectral density matrices. Once these estimators are given, we have consistency of the estimators of the common component also for T going to infinity. Moreover we can define out-of-sample estimators of the common component, which are needed for prediction.

First of all we need to replace assumption 2.2 with a stronger assumption.

Assumption 3 *The vector \mathbf{x}_{nt} admits a Wold representation*

$$\mathbf{x}_{nt} = \sum_{k=0}^{\infty} \mathbf{C}_k \mathbf{v}_{t-k} \quad \text{for any } n \in \mathbb{N},$$

where the full-rank n -dimensional innovations have finite moments of order 4 and the (i, j) -th entry of the matrices \mathbf{C}_k satisfies

$$\sum_{k=0}^{\infty} |C_{ij,k}| k^{1/2} < \infty \quad \text{for any } i, j \in \mathbb{N}.$$

The sample covariance matrix for the process \mathbf{x}_{nt}^T is the usual one:³

$$\hat{\mathbf{\Gamma}}_k^x = \frac{1}{T-k} \sum_{t=k+1}^T \mathbf{x}_{nt}^T \mathbf{x}_{nt-k}^{T'}.$$

We then replace the population spectral density Σ_n^x by its empirical counterpart by using the periodogram smoothing estimator

$$\hat{\Sigma}^x(\theta) = \sum_{k=-M_T}^{M_T} w_k \hat{\mathbf{\Gamma}}_k^x e^{-ik\theta}; \quad (7)$$

where the weights of the smoothing window are $w_k = 1 - \frac{|k|}{M_T+1}$ and the window size is usually $M_T \simeq \sqrt{T}$. We can use these estimators of the covariance and spectral density matrices to obtain a sample estimator of the common component, by applying the methods outlined in the previous section. Consistency of the estimator $\hat{\chi}_t$ for n and T going to infinity is proved by Forni et al. (2000), in the case of dynamic principal components, and by Forni et al. (2005), when using generalized principal components.

Theorem 5 (Consistency) *Assume that assumptions 1, 2, and 3 are satisfied. Then, for any $\epsilon > 0$ and $\eta > 0$, there exists $N_0 = N_0(\epsilon, \eta)$ with $N_0 > i$, and $T_0 = T_0(n, \epsilon, \eta)$ such that*

$$\text{Prob} [|\hat{\chi}_{it} - \chi_{it}| > \epsilon] \leq \eta,$$

for any $n \geq N_0$ and any $t \geq T_0$.⁴

Finally, given these results, we have that the estimate of the h -steps-ahead forecasted common component is

$$\hat{\chi}_{T+h|T} = \hat{\mathbf{\Gamma}}_h^x \hat{\mathbf{Z}}' (\hat{\mathbf{Z}} \hat{\mathbf{\Gamma}}_0^x \hat{\mathbf{Z}}')^{-1} \hat{\mathbf{Z}} \mathbf{x}_{nt}^T,$$

$$\hat{\chi}_{T+h|T} = \hat{\mathbf{\Gamma}}_h^x \hat{\mathbf{S}}' (\hat{\mathbf{S}} \hat{\mathbf{\Gamma}}_0^x \hat{\mathbf{S}}')^{-1} \hat{\mathbf{S}} \mathbf{x}_{nt}^T,$$

for the generalized or static principal components method respectively. Proof of the convergence in probability of $\hat{\chi}_{T+h|T}$ to the best linear forecast is given in Forni et al. (2005).

8 On the number of dynamic and static factors

Up to now we always made the hypothesis that the number of static and dynamic factors is smaller than the cross dimension of the dataset considered. However nothing is said on how

³Hereafter with $\hat{}$ we indicate sample estimators when dealing with the finite-time process \mathbf{x}_n^T .

⁴In the case of the two-sided estimator by Forni et al. (2000) this result holds only for the central part of the observed series.

to determine these numbers. Concerning the number of dynamic factors we review here the criterion by Hallin and Liška (2007), while for the number of static factors we present our refinement of the criterion by Bai and Ng (2002) as presented in Alessi et al. (2007b).

As usual, we first consider the population results for an n -dimensional vector process \mathbf{x}_{nt} , and then we consider the case for finite t . Given a GDFM with assumptions 2 holding, Hallin and Liška propose an information criterion aimed at minimizing the residual variance of the idiosyncratic components. In particular given k dynamic factors, the idiosyncratic component has variance

$$\frac{1}{n} \mathbb{E} \left[\sum_{i=1}^n (x_{it} - \chi_{n,it})^2 \right] = \frac{1}{n} \sum_{j=k+1}^n \int_{-\pi}^{\pi} \lambda_{nj}^x(\theta) d\theta .$$

where λ_{nj}^x is the j -th largest eigenvalue of the spectral density matrix Σ_n^x , which, for the moment, is supposed to be known. Therefore, denoting by q_{\max} some predefined upper bound and by $p(n)$ some adequate penalty, they propose to select the number of factors as

$$\hat{q}_n = \operatorname{argmin}_{0 \leq k \leq q_{\max}} IC_{0,n}(k) , \tag{8}$$

$$IC_{0,n}(k) = \frac{1}{n} \sum_{j=k+1}^n \int_{-\pi}^{\pi} \lambda_{nj}^x(\theta) d\theta + kp(n) .$$

This criterion is nothing else but testing the validity of assumption 2.3, indeed it requires the contribution of the eigenvalues to be small, when $k > \hat{q}_n$. The penalty function for n going to infinity should be not too large or we underestimate the true number of factors q , but it should be large enough to avoid overestimation. In order to impose consistency conditions on $p(n)$ we need an assumption on the rate of divergence of the smallest diverging eigenvalue. What we require is that the q -th eigenvalue of Σ_n^x diverges at least linearly in n . Given this assumption the following lemma holds.

Lemma 2 *Let \hat{q}_n be as defined in (8), and the penalty function be such that*

$$\lim_{n \rightarrow \infty} p(n) = 0 \quad \text{and} \quad \lim_{n \rightarrow \infty} np(n) = \infty .$$

Then, under the assumptions of GDFM,

$$\lim_{n \rightarrow \infty} \hat{q}_n = q .$$

Of course for any positive constant c a penalty function $cp(n)$ leads always to a consistent estimate of q . Although such a constant has no effect asymptotically, it can affect the actual result when n is finite.

When we deal with finite samples of size T , we replace the spectral density matrix with its estimator $\hat{\Sigma}^x$ (see (7)) with window size M_T . The penalty function depends also on T and the two associated information criteria proposed are

$$IC_{1,n}^T(k) = \frac{1}{n} \sum_{j=k+1}^n \frac{1}{2M_T + 1} \sum_{h=-M_T}^{M_T} \lambda_{nj}^{Tx}(\theta_h) + kp(n, T) ,$$

$$IC_{2,n}^T(k) = \log \left[\frac{1}{n} \sum_{j=k+1}^n \frac{1}{2M_T + 1} \sum_{h=-M_T}^{M_T} \lambda_{nj}^{Tx}(\theta_h) \right] + kp(n, T) ,$$

for $0 \leq k \leq q_{\max}$ and $\theta_h = \pi h / (M_T + 1/2)$. Depending on the criterion adopted the estimated number of factors is then

$$\hat{q}_n^T = \operatorname{argmin}_{0 \leq k \leq q_{\max}} IC_{a,n}^T(k) \quad \text{for } a = 1, 2 .$$

In practice, given the simulation results provided by Hallin and Liška, we always use the criterion in its logarithmic form $IC_{2,n}^T$. Consistency of this criterion is proved in Hallin and Liška (2007), provided that the penalty function satisfies

$$\lim_{\substack{n \rightarrow \infty \\ T \rightarrow \infty}} p(n, T) = 0 \quad \text{and} \quad \lim_{\substack{n \rightarrow \infty \\ T \rightarrow \infty}} \min(n, M_T^2, M_T^{-1/2} T^{1/2}) p(n, T) = \infty .$$

Once again, notice that if $p(n, T)$ is an appropriate penalty function, then $cp(n, T)$ is appropriate as well for any positive constant c . This degree of freedom can be exploited when implementing the criterion in practice. The only information we have about the asymptotic behaviour of $\hat{q}_{c,n}^T$ is given when considering subsamples of sizes $n_j \leq n$ and $T_j \leq T$ with $j = 1, \dots, J$. For each value of c we compute the number of factors $q_{c,n_j}^{T_j}$ for all possible subsamples, we obtain J possible numbers with a variability given by

$$S_c^2 = \frac{1}{J} \sum_{j=1}^J \left[q_{c,n_j}^{T_j} - \frac{1}{J} \sum_{j=1}^J q_{c,n_j}^{T_j} \right]^2 .$$

When $c = 0$ we always get $\hat{q}_{c,n}^T = q_{\max}$ and $S_c = 0$, when c increases we find stability intervals but also values of c with high variability. As c increases we increase penalization. In order to tune c , we look for intervals of c for which no dependence on sample size is present, i.e. $S_c = 0$ and a constant number of factors is obtained for $n_j = n$ and $T_j = T$, this number is the estimated number of dynamic factors $\hat{q}_{c,n}^T$. In figure 8, we look for the first plateau of the red line (representing the number of factors) corresponding to a stability interval (blue line equal to zero). Of course, the value q_{\max} is never considered as it is a boundary solution.

Bai and Ng (2002) propose an analogous information criterion to determine the number of static factors. They assume the static factor model (4) with r factors for an n -dimensional vector process of finite time length T . Common factors $\hat{\mathbf{F}}_t^{(k)}$ and their loadings $\hat{\mathbf{\Lambda}}^{(k)}$ are estimated using static principal components. We use the superscript k when we choose k factors. The variance of the idiosyncratic part, conveniently penalized, is computed as a function of k

$$PC_{1,n}^T(k) = \frac{1}{nT} \sum_{t=1}^T \sum_{i=1}^n (x_{it} - \hat{\mathbf{\Lambda}}_i^{(k)} \hat{\mathbf{F}}_t^{(k)})^2 + kp(n, T) ,$$

$$PC_{2,n}^T(k) = \log \left[\frac{1}{nT} \sum_{t=1}^T \sum_{i=1}^n (x_{it} - \hat{\mathbf{\Lambda}}_i^{(k)} \hat{\mathbf{F}}_t^{(k)})^2 \right] + kp(n, T) .$$

Depending on the chosen criterion, the estimated number of factors is

$$\hat{r}_n^T = \operatorname{argmin}_{0 \leq k \leq r_{\max}} PC_{a,n}^T(k) \quad \text{for } a = 1, 2 .$$

Consistency, is proven in Bai and Ng (2002), provided that

$$\lim_{\substack{n \rightarrow \infty \\ T \rightarrow \infty}} p(n, T) = 0 \quad \text{and} \quad \lim_{\substack{n \rightarrow \infty \\ T \rightarrow \infty}} p(n, T) [\min(\sqrt{n}, \sqrt{T})]^2 = \infty .$$

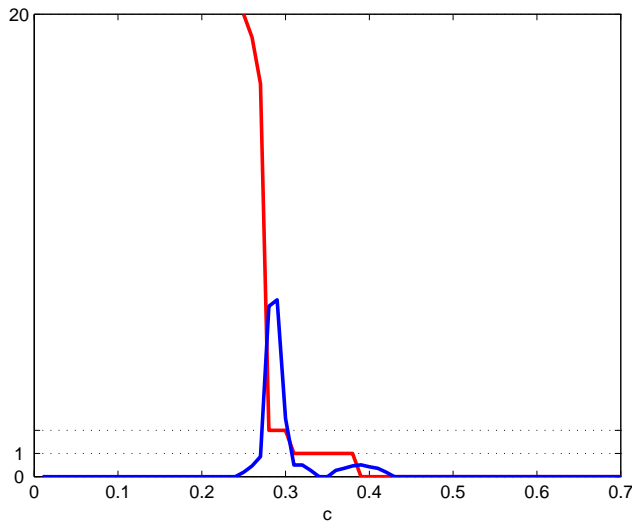


Figure 1: Estimated number of dynamic factors using the $IC_{2,n}^T$ criterion with penalty $p(n, T) = \min(n, M_T^2, M_T^{-1/2}T^{1/2})^{-1} \log(\min(n, M_T^2, M_T^{-1/2}T^{1/2}))$. Red line: $\hat{q}_{c,n}^T$. Blue line: S_c . We look for a plateau in the red line corresponding to a zero level blue line, therefore in this case $\hat{q}_{c,n}^T = 1$. Here $q_{\max} = 20$, $M_T = \sqrt{T}$, $n = 81$, and $T = 80$. Data are firm sales of the US chemicals sector (source: US Compustat).

However, in all empirical applications both criteria proposed by Bai and Ng turn out to heavily depend on the choice of r_{\max} (e.g. see Forni et al. (2007)). The PC_1 depends explicitly on it while the PC_2 depends on it only when implementing it in practice. For this reason we always consider the latter criterion. By analogy with the criterion by Hallin and Liška, we propose to multiply the penalty function by a positive constant c . This constant does not modify the asymptotic behaviour of the criterion but it can be used to tune the penalizing power of $p(n, T)$. Once we introduce c , the procedure to select the number of static factors is exactly the same as the one used for the number of dynamic factors. For each value of c , we consider J subsamples of the dataset and we compute $\hat{r}_{c,n_j}^{T_j}$ and the variability in the number of factors S_c for $j = 1, \dots, J$. The number of static factors is given by a plateau in the function $\hat{r}_{c,n}^T$ corresponding to the stability interval $S_c = 0$. In figure 8, taken from a simulated sample in Alessi et al. (2007b), we look for the first plateau of the red line (representing the number of factors) corresponding to a stability interval (blue line equal to zero). As before, the value r_{\max} is never considered as it is a boundary solution.

9 Comparing the static and dynamic representations

The equivalence between the dynamic and static representations of factor models is crucial in all what follows. Therefore, we explain in detail how the two are linked. There are two possible representations for a dynamic factor model

$$\mathbf{x}_{nt} = \underline{\mathbf{B}}_n(L)\mathbf{u}_t + \boldsymbol{\xi}_{nt} \quad \text{dynamic} \quad (9)$$

$$\mathbf{x}_{nt} = \boldsymbol{\Lambda}_n\mathbf{F}_t + \boldsymbol{\xi}_{nt} \quad \text{static} \quad (10)$$

We need an assumption that guarantees the possibility of switching from (9) to (10).

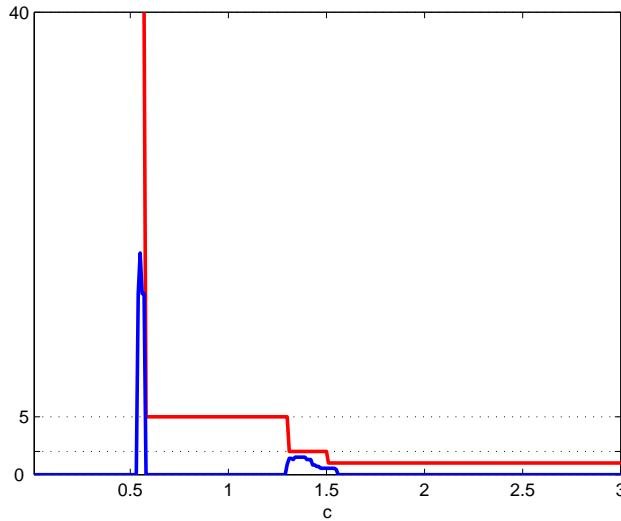


Figure 2: Estimated number of static factors using the $PC_{2,n}^T$ criterion with penalty $p(n, T) = (n + T/nT) \log[\min(\sqrt{n}, \sqrt{T})]^2$. Red line: $\hat{r}_{c,n}^T$. Blue line: S_c . We look for a plateau in the red line corresponding to a zero level blue line, therefore in this case $\hat{r}_{c,n}^T = 5$. Here $r_{\max} = 40$, $M_T = \sqrt{T}$, $n = 100$, and $T = 80$.

Assumption 4 *There exists an integer $r \geq q$, a nested sequence of $n \times r$ matrices Λ_n and a one-sided absolutely summable $r \times q$ matrix polynomial, in principle of infinite order, $\underline{\mathbf{N}}(L)$ such that*

$$\underline{\mathbf{B}}_n(L) = \Lambda_n \underline{\mathbf{N}}(L) \quad \text{and} \quad \mathbf{F}_t = \underline{\mathbf{N}}(L) \mathbf{u}_t$$

We know that lemma 1 holds as the analogous of assumptions 2.3 and 2.4. Moreover, the standard assumption on $\underline{\mathbf{N}}(L)$ is that it comes from the inversion of a VAR model

$$\underline{\mathbf{A}}(L) \mathbf{F}_t = \mathbf{H} \mathbf{u}_t . \quad (11)$$

This assumption implies that \mathbf{u}_t belongs to space spanned by past and present values of the common component. We say that the dynamic factors are fundamental for the χ 's. Actually it is enough that $\underline{\mathbf{A}}(L)$ is of order one to guarantee identification and fundamentalness of the dynamic factors. All details on the issues of fundamentalness and identification of the structural shocks are considered in chapter 6. Here we take for granted (11) and we prove the equivalence of the dynamic and static representations. By inverting equation (11), the static factors are an MA(∞) of the dynamic factors. Namely, if we consider a VAR(1) specification,

$$\mathbf{F}_t = (\mathbf{I}_r - \mathbf{A}L)^{-1} \mathbf{H} \mathbf{u}_t .$$

We are then back to the dynamic representation of the factor model

$$\mathbf{x}_{nt} = \Lambda_n (\mathbf{I}_r - \mathbf{A}L)^{-1} \mathbf{H} \mathbf{u}_t + \boldsymbol{\xi}_{nt} ,$$

where the loadings of (9) are

$$\underline{\mathbf{B}}_n(L) = \sum_{k=0}^{\infty} \mathbf{B}_{nk} L^k = \sum_{k=0}^{\infty} \Lambda_n \mathbf{A}^k \mathbf{H} L^k .$$

It is clear that a VAR(1) representation of the static factors is general enough for the common part to be an MA(∞) representation. This allows for both MA and AR loadings of the dynamic factors, which is an important generalization with respect to the static model by Stock and Watson (2002a). In practice, however, we assume that the maximum order of $\underline{\mathbf{B}}_n(L)$ is s and that the static factors are the dynamic ones counted with their lags

$$\mathbf{F}_t = (\mathbf{u}'_t \dots \mathbf{u}'_{t-s})' \quad \text{and} \quad \mathbf{\Lambda}_n = (\mathbf{B}_{n0} \dots \mathbf{B}_{ns}). \quad (12)$$

Therefore, the static rank of \mathbf{F}_t is $q(s+1)$ and for the static model to hold we need also $\mathbf{\Lambda}_n$ to be full rank. This is guaranteed if no restrictions hold among the loadings \mathbf{B}_{ni} , and, in this case, $r = q(s+1)$. Equations (11) and (12) are assumed in the next three chapters, while are rediscussed in chapter 6. e t .

10 Examples

Static factor model

One factor

$$x_{it} = a_i u_t + \xi_{it} \quad \text{var}(u_t) = 1 \quad \text{var}(\xi_{it}) = \sigma_i^2$$

The common component is given by a DAS of the variables, remember that it spans the same space as the factors, thus a basis for this space has dimension q equal to the number of factors. We need an aggregate of the variables for each factor. If $q = 1$ we could take the arithmetic average

$$\frac{1}{n} \sum_{i=1}^n x_{it} \equiv \bar{x}_t = u_t \frac{1}{n} \sum_{i=1}^n a_i + \frac{1}{n} \sum_{i=1}^n \xi_{it}$$

The variance of the DAS is made of two parts one diverging and one bounded

$$\text{var}(\bar{x}_t) = 1 \cdot \bar{a}_n^2 + \frac{1}{n^2} \sum_{i=1}^n \sigma_i^2 \xrightarrow{n \rightarrow \infty} \bar{a}^2$$

Remember that we always assume the eigenvalues of the idiosyncratic covariance to be bounded i.e. $\exists M$ s.t. $\sigma_i^2 \leq M \forall i$.

To be sure that $\bar{a}^2 \neq 0$ we can take a weighted average with weights $(a_1 \dots a_n)$ which has variance

$$\text{var} \left(\sum_{i=1}^n a_i x_{it} \right) = 1 \cdot \left(\sum_{i=1}^n a_i^2 \right)^2 + \left(\sum_{i=1}^n a_i^2 \sigma_i^2 \right) = O(n^2) + O(n)$$

However, we do not know the weights a in advance, we need weights $\mathbf{S} = (s_1 \dots s_n)$ s.t. maximize the variance of the weighted average

$$\begin{aligned} \mathbf{S}^{(1)} &= \underset{(s_1 \dots s_n)}{\text{argmax}} \left[\text{var} \left(\sum_{i=1}^n s_i x_{it} \right) \right] \\ \text{s.t.} \quad &\sum_{i=1}^n s_i^2 = 1 \end{aligned}$$

The vector $\mathbf{S}^{(1)}\mathbf{x}_t$ is the first static principal component. Remember that solving the previous maximization is equivalent to solve the eigenvalues problem

$$\mathbf{S}\mathbf{\Gamma}_0^x = \mu_1^x \mathbf{S}$$

In the one factor case recovering the factor space is the same as recovering the factor itself.

Two factors

$$x_{it} = a_i u_t + b_i v_t + \xi_{it} \quad \text{var}(u_t) = \text{var}(v_t) = 1 \quad \text{var}(\xi_{it}) = \sigma_i^2$$

If more than one factor is present we need another system of weights $\mathbf{S}^{(2)}$ orthogonal to the first one. The estimated factors are

$$\begin{aligned} \mathbf{S}^{(1)}\mathbf{x}_t &= \left(\sum_{i=1}^n a_i s_i^{(1)} \right) u_t + \left(\sum_{i=1}^n b_i s_i^{(1)} \right) v_t + \left(\sum_{i=1}^n s_i^{(1)} \xi_{it} \right) \xrightarrow{n \rightarrow \infty} \alpha_1 u_t + \beta_1 v_t \\ \mathbf{S}^{(2)}\mathbf{x}_t &= \left(\sum_{i=1}^n a_i s_i^{(2)} \right) u_t + \left(\sum_{i=1}^n b_i s_i^{(2)} \right) v_t + \left(\sum_{i=1}^n s_i^{(2)} \xi_{it} \right) \xrightarrow{n \rightarrow \infty} \alpha_2 u_t + \beta_2 v_t \end{aligned}$$

The space spanned by the factors is identified, but not the factors. We project \mathbf{x}_t on the estimated factors and since $\xi_{it} \perp u_t$ and $\xi_{it} \perp v_t$,

$$\text{Proj} \{ \xi_{it} | \overline{\text{span}}(\mathbf{S}^{(1)}\mathbf{x}_t, \mathbf{S}^{(2)}\mathbf{x}_t) \} = 0$$

$$\text{Proj} \{ u_t | \overline{\text{span}}(\mathbf{S}^{(1)}\mathbf{x}_t, \mathbf{S}^{(2)}\mathbf{x}_t) \} = u_t$$

$$\text{Proj} \{ v_t | \overline{\text{span}}(\mathbf{S}^{(1)}\mathbf{x}_t, \mathbf{S}^{(2)}\mathbf{x}_t) \} = v_t$$

The common component is given by

$$a_i u_t + b_i v_t = A_i \mathbf{S}^{(1)}\mathbf{x}_t + B_i \mathbf{S}^{(2)}\mathbf{x}_t$$

Dynamic factor model

One factor loaded with one lag - static estimation

$$x_{it} = a_i(L)u_t + \xi_{it} \quad \text{var}(u_t) = 1 \quad \text{var}(\xi_{it}) = \sigma_i^2$$

with $a_i(L) = \alpha_i + \beta_i L$. It is equivalent to a two static factor model by assuming $v_t = u_{t-1}$

$$x_{it} = \alpha_i u_t + \beta_i v_t + \xi_{it}$$

The “toy-model” for $k \geq 0$

$$\begin{cases} x_{(2k+1)t} = u_t + \xi_{(2k+1)t} \\ x_{(2k)t} = u_{t-1} + \xi_{(2k)t} \end{cases}$$

Assume $\text{var}(\xi_{it}) = 1 \forall i$ and n even, then

$$\mathbf{\Gamma}_0^x = \mathbf{E} \left[\begin{pmatrix} 1 & 0 \\ 0 & 1 \\ \vdots & \vdots \\ 0 & 1 \end{pmatrix} \underbrace{\begin{pmatrix} u_t \\ u_{t-1} \end{pmatrix} (u_t \ u_{t-1})}_{\mathbf{I}_2} \begin{pmatrix} 1 & 0 & \dots & 0 \\ 0 & 1 & \dots & 1 \end{pmatrix} \right] + \mathbf{I}_n = \begin{pmatrix} 1 & 0 & 1 & \dots & 0 \\ 0 & 1 & 0 & \dots & 1 \\ 1 & 0 & 1 & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 1 & 0 & \dots & 1 \end{pmatrix} + \mathbf{I}_n$$

The two eigenvectors and corresponding eigenvalues are

$$\mathbf{S}^{(1)} = (1 \ 0 \ 1 \ \dots \ 0) \rightarrow \mu_1 = \frac{n}{2} + 1$$

$$\mathbf{S}^{(2)} = (0 \ 1 \ 0 \ \dots \ 1) \rightarrow \mu_2 = \frac{n}{2} + 1$$

The maximum variance comes from either all odd variables or all even variables. The principal components are

$$\begin{aligned} \mathbf{S}^{(1)} \mathbf{x}_t &= \frac{n}{2} u_t + \sum_{i \text{ odd}} \xi_{it} \xrightarrow{n \rightarrow \infty} \frac{n}{2} u_t \\ \mathbf{S}^{(2)} \mathbf{x}_t &= \frac{n}{2} u_{t-1} + \sum_{i \text{ even}} \xi_{it} \xrightarrow{n \rightarrow \infty} \frac{n}{2} u_{t-1} \end{aligned}$$

Indeed, when n diverges the variance of the idiosyncratic component becomes negligible

$$\frac{\text{var}(\xi_{it})}{\text{var}(\chi_{it})} = \frac{n/2}{n^2/4} = \frac{2}{n} \xrightarrow{n \rightarrow \infty} 0$$

The common component is the projection of \mathbf{x}_t on the estimated factors, e.g. for the first variable

$$\chi_{1t} = u_t = A_1 \frac{n}{2} u_t + B_1 \frac{n}{2} u_{t-1}$$

moreover, $B_1 = 0$ since $u_t \perp u_{t-1}$ and thus $A_1 = \frac{2}{n}$ and the factor is identified.

One factor loaded with one lag - dynamic estimation

Instead of a simple weighted average, we can take a dynamic average of the variables, i.e. $\mathbf{p}(L)\mathbf{x}_t$, the order of the filter being the number of lags with which the factor is loaded. The dynamic averages that we are looking for have weights $\mathbf{p}(L) = (p_1(L) \dots p_n(L))$

$$\begin{aligned} \mathbf{p}^{(1)}(L) &= \underset{(p_1(L) \dots p_n(L))}{\text{argmax}} \left[\text{var} \left(\sum_{i=1}^n \sum_{k=-\infty}^{+\infty} p_{ik} x_{it-k} \right) = \text{var} \left(\sum_{i=1}^n p_i(L) x_{it} \right) = \text{var} \left(\sum_{k=-\infty}^{+\infty} \mathbf{p}_k \mathbf{x}_{t-k} \right) \right] \\ \text{s.t.} \quad &\sum_{i=1}^n \sum_{k=-\infty}^{+\infty} p_{ik}^2 = 1 \end{aligned}$$

$\mathbf{p}^{(1)}(L)\mathbf{x}_t$ is the first dynamic principal component. To solve the maximization problem we can move to the frequency domain

$$\text{var}(\mathbf{p}(L)\mathbf{x}_t) = \frac{1}{2\pi} \int_{-\pi}^{\pi} \mathbf{p}(\theta) \mathbf{\Sigma}^x(\theta) \mathbf{p}(\theta) d\theta$$

In order to maximize the integral it is enough to maximize the integrand $\forall \theta \in [-\pi, \pi]$. Once we have the first dynamic principal component $\mathbf{p}^{(1)}(\theta)$ we expand it in Fourier series and we use the coefficients \mathbf{p}_k to build the filter $\mathbf{p}^{(1)}(L)$

$$\begin{aligned}\mathbf{p}^{(1)}(\theta) &= \sum_{k=-\infty}^{+\infty} \mathbf{p}_k e^{-ik\theta} = \frac{1}{2\pi} \sum_{k=-\infty}^{+\infty} \left[\int_{-\pi}^{\pi} \mathbf{p}^{(1)}(\theta) e^{ik\theta} d\theta \right] e^{-ik\theta} \\ \mathbf{p}^{(1)}(L) &= \sum_{k=-\infty}^{+\infty} \mathbf{p}_k L^k = \frac{1}{2\pi} \sum_{k=-\infty}^{+\infty} \left[\int_{-\pi}^{\pi} \mathbf{p}^{(1)}(\theta) e^{ik\theta} d\theta \right] L^k\end{aligned}$$

Remember that solving this maximization problem is equivalent to solve the dynamic eigenvalues problem

$$\mathbf{p}(\theta) \Sigma^x(\theta) = \lambda_1^x(\theta) \mathbf{p}(\theta) \quad \forall \theta \in [-\pi, \pi]$$

The “toy-model” for $k \geq 0$

$$\begin{cases} x_{(2k+1)t} = u_t + \xi_{(2k+1)t} \\ x_{(2k)t} = u_{t-1} + \xi_{(2k)t} \end{cases}$$

Assume $\text{var}(\xi_{it}) = 1 \forall i$ and n even, then

$$\mathbf{\Gamma}_0^x = \begin{pmatrix} 2 & 0 & 1 & \dots & 0 \\ 0 & 2 & 0 & \dots & 1 \\ 1 & 0 & 2 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 1 & 0 & \dots & 2 \end{pmatrix} \quad \mathbf{\Gamma}_1^x = \begin{pmatrix} 0 & 0 & 0 & \dots & 0 \\ 1 & 0 & 1 & \dots & 0 \\ 0 & 0 & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & 0 & 1 & \dots & 0 \end{pmatrix} \quad \mathbf{\Gamma}_{-1}^x = \mathbf{\Gamma}_1^{x'}$$

Thus

$$\Sigma^x(\theta) = \begin{pmatrix} 1 & e^{i\theta} & 1 & \dots & e^{i\theta} \\ e^{-i\theta} & 1 & e^{-i\theta} & \dots & 1 \\ 1 & e^{i\theta} & 1 & \dots & e^{i\theta} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ e^{-i\theta} & 1 & e^{-i\theta} & \dots & 1 \end{pmatrix} + \mathbf{I}_n$$

The eigenvector we are looking for is $\mathbf{p}^{(1)}(\theta) = (1 \ e^{i\theta} \ 1 \ \dots \ e^{i\theta})$ which solves

$$\mathbf{p}^{(1)}(\theta) \Sigma^x(\theta) = (n+1) \mathbf{p}^{(1)}(\theta)$$

Take its Fourier expansion

$$\mathbf{p}^{(1)}(\theta) = \mathbf{p}_0 + \mathbf{p}_{-1} e^{i\theta}$$

$$\mathbf{p}_0 = (1 \ 0 \ 1 \ \dots \ 0)$$

$$\mathbf{p}_{-1} = (0 \ 1 \ 0 \ \dots \ 1)$$

$$\mathbf{p}^{(1)}(L) = (1 \ L^{-1} \ 1 \ \dots \ L^{-1})$$

Hence, the first dynamic principal component is

$$\mathbf{p}^{(1)}(L) \mathbf{x}_t = x_{1t} + x_{2t+1} + x_{3t} + \dots + x_{nt+1}$$

or

$$\mathbf{p}^{(1)}(L)\mathbf{x}_t = (\mathbf{p}_0 + \mathbf{p}_{-1}L^{-1})\mathbf{x}_t = (1 \ 0 \ 1 \ \dots \ 0)\mathbf{x}_t + (0 \ 1 \ 0 \ \dots \ 1)\mathbf{x}_{t+1} = nu_t + \sum_{i=1}^n \xi_{it}$$

In this way all variables lead the same factor u_t while before some were leading and some were lagging. Once again the variance of the idiosyncratic component becomes negligible

$$\frac{\text{var}(\xi_{it})}{\text{var}(\chi_{it})} = \frac{n}{n^2} = \frac{1}{n} \xrightarrow{n \rightarrow \infty} 0$$

The problem of the previous procedure is the presence of bilateral filters which makes forecasting impossible, thus we have to use the two-step estimator of section 6.

References

- Alessi, L., Barigozzi, M., and Capasso, M. (2006). Forecasting large datasets with conditionally heteroskedastic dynamic common factors. Working Paper 2009-005, ECARES.
- Alessi, L., Barigozzi, M., and Capasso, M. (2007a). A review of nonfundamentalness and identification in structural VAR models. Working Paper Series 922, European Central Bank.
- Alessi, L., Barigozzi, M., and Capasso, M. (2007b). A robust criterion for determining the number of static factors in approximate factor models. Working Paper Series 903, European Central Bank.
- Bai, J. (2003). Inferential theory for factor models of large dimensions. *Econometrica*, 71(1):135–171.
- Bai, J. and Ng, S. (2002). Determining the number of factors in approximate factor models. *Econometrica*, 70(1):191–221.
- Bai, J. and Ng, S. (2007). Determining the number of primitive shocks in factor models. *Journal of Business and Economic Statistics*, 25(1):52–60.
- Boivin, J. and Giannoni, M. (2006). DSGE models in a data-rich environment. NBER Working Papers 12772, National Bureau of Economic Research, Inc.
- Boivin, J. and Ng, S. (2005). Understanding and comparing factor-based forecasts. *International Journal of Central Banking*, 1(3):117–151.
- Boivin, J. and Ng, S. (2006). Are more data always better for factor analysis? *Journal of Econometrics*, 127(1):169–194.
- Brillinger, D. (1981). *Time Series: Data Analysis and Theory*. Holden-Day, Inc., San Francisco.
- Chamberlain, G. (1983). Funds, factors, and diversification in arbitrage pricing models. *Econometrica*, 51(5):1281–1304.
- Chamberlain, G. and Rothschild, M. (1983). Arbitrage, factor structure and mean-variance analysis in large asset markets. *Econometrica*, 51(5):1305–1324.
- D’Agostino, A. and Giannone, D. (2006). Comparing alternative predictors based on large-panel factor models. Working Paper Series 680, European Central Bank.
- Forni, M., Giannone, D., Lippi, M., and Reichlin, L. (2007). Opening the black box - structural factor models with large cross-sections. Working Paper Series 712, European Central Bank.
- Forni, M., Hallin, M., Lippi, M., and Reichlin, L. (2000). The generalized dynamic factor model: identification and estimation. *The Review of Economics and Statistics*, 82(4):540–554.
- Forni, M., Hallin, M., Lippi, M., and Reichlin, L. (2004). The generalized dynamic factor model: consistency and rates. *Journal of Econometrics*, 119(2):231–255.
- Forni, M., Hallin, M., Lippi, M., and Reichlin, L. (2005). The generalized dynamic factor model: one-sided estimation and forecasting. *Journal of the American Statistical Association*, 100(471):830–840.
- Forni, M. and Lippi, M. (2001). The generalized dynamic factor model: representation theory. *Econometric Theory*, 17(6):1113–1141.

- Geweke, J. (1977). The dynamic factor analysis of economic time series. In Aigner, D. J. and Goldberger, A. S., editors, *Latent Variables in Socio-Economic Models*. North-Holland.
- Giannone, D., Reichlin, L., and Sala, L. (2004). Monetary policy in real time. In Gertler, M. and K.Rogoff, editors, *NBER Macroeconomic Annual*. MIT Press.
- Hallin, M. and Liška, R. (2007). Determining the number of factors in the general dynamic factor model. *Journal of the American Statistical Association*, 102(478):603–617.
- Onatski, A. (2006). Determining the number of factors from empirical distribution of eigenvalues. Discussion papers 0405-19, Economics Department, Columbia University.
- Peña, D. (2002). *Análisis de datos multivariantes*. Mc Graw-Hill, Madrid.
- Sargent, T. J. and Sims, C. A. (1977). Business cycle modeling without pretending to have too much a priori economic theory. In Sims, C. A., editor, *New Methods in Business Research*. Federal Reserve Bank of Minneapolis.
- Stock, J. H. and Watson, M. W. (1998). Diffusion indexes. NBER Working Papers 6702, National Bureau of Economic Research, Inc.
- Stock, J. H. and Watson, M. W. (2002a). Forecasting using principal components from a large number of predictors. *Journal of the American Statistical Association*, 97(460):1167–1179.
- Stock, J. H. and Watson, M. W. (2002b). Macroeconomic forecasting using diffusion indexes. *Journal of Business and Economic Statistics*, 20(2):147–162.