DYNAMIC FACTOR MODELS

Matteo Barigozzi[†]

January 24, 2020

[†] London School of Economics and Political Science, Statistics Department, United Kingdom.

These notes cover part of the material taught in the courses on factor models held at
IHS in Vienna in March 2013 and CU Hong Kong in June 2016, jointly with Marc Hallin
University of Alicante in May 2013
LSE Methods Summer School in August 2014, 2015, 2016
University of London in November 2014 and 2015
Sant'Anna School of Advanced Studies, Pisa, April 2018

Section 8 benefited also from the help and comments of Matteo Luciani.

Discalimer. These notes might contain mistakes, they are meant as reference only. Please do not cite but refer to the original papers. Notice that there are many other works published on the subject and not all are cited here. Note also that Section 7 is incomplete and is part of my ongoing research and some further results are in Barigozzi and Luciani (2017, 2019a,b).

Contents

1 Introduction and taxonomy of factor models **3**

2 Principal component analysis and the exact factor model **9**

3 The approximate factor model and the blessing of dimensionality **12**

4 The generalised dynamic factor model - Time domain **23**

5 The generalised dynamic factor model - Frequency domain **26**

6 The approximate dynamic factor model in state-space form **35**

7 Maximum Likelihood Estimation, EM and Kalman Smoother **42**

7.1 Static estimation 43

7.2 Dynamic estimation 44

8 Forecasting **52**

9 Structural approximate dynamic factor models in macroeconomics **58**

10 Factor models in finance **66**

11 Number of factors **71**

1 Introduction and taxonomy of factor models

Factor models are a dimension reduction technique that exists in the statistical literature since almost one hundred years and were originally introduced in the filed of psychometrics. The recent availability of large datasets made them increasingly popular in the last twenty years. They are nowadays commonly used by public and private institutions as central banks and investment banks for analysing large panels of time series. Recently a class of models known as approximate (or generalised) dynamic factor models has been proposed in the literature can be considered as a pioneer dimension reduction technique in "big data econometrics" or also a pioneer technique of "unsupervised statistical learning".

In these notes we cover the following topics:

- large dimensional datasets of time series;
  - curse of dimensionality;
- factor models as dimension reduction techniques;
  - blessing of dimensionality;
- estimation;
- forecasting;
- macroeconomic and financial applications: policy analysis, risk management.

MOTIVATION. The large datasets available today cover a number of years which is finite thus the number of data points is limited and constrained by passage of time. However, more and more time series are collected and made available by statistical agencies. We denote by T the number of points in time and by n the number of series. Then, typically we are in a setting where $n \geq T$. For example, in macroeconomic datasets we have $n \simeq 100, 1000$ and $T \simeq 100$ (quarterly or monthly series), while in financial datasets we have $n \simeq 100, 1000$ and $T \simeq 1000$ (daily series). The case $n \geq T$ in statistics is known as high-dimensional setting and dealing with such datasets by means of standard techniques as linear regressions constitutes a hard problem to be solved due to the lack of degrees of freedom. Think for example of a VAR(1) model for n time series which would require estimation of about n^2 parameters with only T data points, as we know linear regression cannot be used in this setting.

In general, a factor model for a high-dimensional vector of time series is characterised by:

1. few latent factors, representing comovements;
2. idiosyncratic terms, representing measurement errors or individual/local features.

Examples are: GDP or inflation which are driven by few factors representing the business cycle plus some measurement errors; equity returns which are driven by few factors representing the market effect plus some factors specific of a given company or sector.

Throughout we consider an n dimensional vector of observed time series $\{\mathbf{x}_t = (x_{1t} \dots x_{nt})' \,|\, t = 1, \dots, T\}$ which we denote for simplicity as $\mathbf{x}_t$. Moreover, we always assume that the data are covariance stationary and purely non-deterministic, with zero mean and unit variance.[1]

STATIC VS. DYNAMIC. A first distinction is based on the effect of the factors on the data, $\mathbf{x}_t$. In the *static factor model* we have

$$\mathbf{x}_t = \mathbf{\Lambda} \mathbf{F}_t + \boldsymbol{\xi}_t, \tag{1}$$

where the factors $\mathbf{F}_t$ are an r-dimensional vector with $r < n$ and have only a contemporaneous effect on $\mathbf{x}_t$. They are called *static factors*.

[1] Non-stationarities, as unit roots and breaks, are considered in more recent works which are not considered in these lecture notes.

In the *dynamic factor model* we have[2]

$$\mathbf{x}_t = \mathbf{\Lambda}^*(L)\mathbf{f}_t + \boldsymbol{\xi}_t, \tag{2}$$

where the factors $\mathbf{f}_t$ are a q-dimensional vector with $q < n$ and have effect on $\mathbf{x}_t$ through their lags too. They are called *dynamic factors*.

The matrices $\mathbf{\Lambda}$ or $\mathbf{\Lambda}^*(L)$ are called *loadings* matrices (or filters in the dynamic case) and are of size $n \times r$ or $n \times q$ respectively. We denote $\boldsymbol{\chi}_t := \mathbf{\Lambda}\mathbf{F}_t$ or $\boldsymbol{\chi}_t := \mathbf{\Lambda}^*(L)\mathbf{f}_t$ the *common component*, while $\boldsymbol{\xi}_t$ is the *idiosyncratic component*. In both cases we always assume that common and idiosyncratic component are uncorrelated, that is

$$\mathsf{Cov}(\chi_{it}, \xi_{js}) = 0, \qquad t, s \in \mathbb{Z}, \quad i, j = 1, \dots, n.$$

or equivalently

$$\begin{aligned} \mathsf{Cov}(f_{it}, \xi_{js}) &= 0, \qquad t, s \in \mathbb{Z}, \quad i = 1, \dots, q, \ j = 1, \dots, n, \\ \mathsf{Cov}(F_{it}, \xi_{js}) &= 0, \qquad t, s \in \mathbb{Z}, \quad i = 1, \dots, r, \ j = 1, \dots, n. \end{aligned}$$

Now if the common and idiosyncratic component have to be the same both in (1) and in (2), then, typically (and intuitively), for the same time series $\mathbf{x}_t$ we need more factors in (1) than in (2). So if r is the dimension of $\mathbf{F}_t$ and q is the dimension of $\mathbf{f}_t$, then we often have $q \leq r$. Moreover, since we always assume $q \leq r \ll n$ we are saying that few dynamic or static factors capture most of the comovements in $\mathbf{x}_t$. The dynamic model is more realistic but harder to estimate, however its estimation can be accomplished by going through and equivalent static model, as explained below.

EXACT VS. APPROXIMATE. Another important distinction between classes of factor models is related to the idiosyncratic component. In the *exact factor model*, $\boldsymbol{\xi}_t$ has no cross-sectional dependence, thus it has a diagonal covariance matrix or even diagonal cross-autocovariances

$$\mathsf{Cov}(\xi_{it}, \xi_{js}) = 0, \qquad t, s \in \mathbb{Z}, \quad i, j = 1, \dots, n, \ i \neq j.$$

In the *approximate factor model*, $\boldsymbol{\xi}_t$ is allowed to have mild cross-sectional dependence, thus a covariance matrix which is not necessarily diagonal. If we even allow for non diagonal cross-autocovariances of $\boldsymbol{\xi}_t$, then we have a *generalised factor model*. These latter cases are more realistic but estimation of the model is possible only if a large cross-section is available, i.e. $n \to \infty$. The exact case imposes a more restrictive assumption but estimation is possible even for few time series, i.e. with a fixed n.

FACTOR REPRESENTATION FOR LARGE PANELS OF TIME SERIES. The approximate dynamic factor model is the most realistic but it is hard to deal with it and in particular we will require to work in a setting where $n \to \infty$. In particular, it is possible derive a representation for an infinite panel of time series which is the analogous of the Wold representation for finite dimensional panels, we call this result the *generalised dynamic factor representation* (or sometime model).

We know that a scalar causal and stationary purely non-deterministic univariate process can always be written as

$$x_t = b(L)\epsilon_t, \quad \epsilon_t \sim w.n.(0, \sigma_\epsilon^2), \tag{3}$$

where $b(L)$ is an infinite one-sided square-summable polynomial:

$$b(L) = b_0 + \sum_{k=1}^{\infty} b_k L^k, \qquad \sum_{k=0}^{\infty} b_k^2 < \infty,$$

with $b_0 = 1$. Representation (3) is known as the Wold representation.

[2] L is the lag operator such that $L^k \mathbf{y}_t = \mathbf{y}_{t-k}$ and a one-sided matrix polynomial of degree $s \in \mathbb{N}$ in L is

$$\mathbf{B}(L) = \sum_{k=0}^{s} \mathbf{B}_k L^k.$$

If we now assume to have an n-dimensional purely non-deterministic vector $\mathbf{x}_t$, under second order stationarity, we can generalize the Wold representation (3) as

$$\mathbf{x}_t = \mathbf{D}(L)\boldsymbol{\epsilon}_t, \quad \boldsymbol{\epsilon}_t \sim w.n.(\mathbf{0}, \boldsymbol{\Gamma}_\epsilon), \tag{4}$$

where $\mathbf{D}(L)$ is a $n \times n$ matrix of infinite one-sided square-summable matrix polynomials.

Typically in order to estimate (4) we consider the Vector Autoregressive (VAR) model of finite order

$$\mathbf{E}(L)\mathbf{x}_t = \boldsymbol{\epsilon}_t, \quad \boldsymbol{\epsilon}_t \sim w.n.(\mathbf{0}, \boldsymbol{\Gamma}_\epsilon), \tag{5}$$

where $\mathbf{E}(L)$ is a $n \times n$ matrix of one-sided matrix polynomials of finite order such that $\det(\mathbf{E}(z)) \neq 0$ for any $z \in \mathbb{C}$ such that $|z| \leq 1$ (that is (5) is a causal process and its inverse requires only positive powers of L, i.e. the past).

There are two areas where we typically use estimates of model (4) given by (5). In forecasting and in structural analysis where the object of interest are the impulse response functions (IRF) of data to unexpected shocks to the economy and which are obtained from an estimate of $\mathbf{D}(L)$ by means of economic restrictions (see later for details). In both cases we need to estimate an $n \times n$ matrix of coefficients and as n increases estimation becomes unfeasible. This is the *curse of dimensionality*, and factor models are a possible answer to it.

In particular, when $n \to \infty$ the Wold representation is replaced by the Generalized Dynamic Factor representation proved by Forni and Lippi (2001) and Hallin and Lippi (2013). In this case we have

$$\mathbf{x}_t = \mathbf{B}(L)\mathbf{u}_t + \boldsymbol{\xi}_t, \quad \mathbf{u}_t \sim w.n.(\mathbf{0}, \mathbf{I}), \tag{6}$$

where $\mathbf{B}(L)$ is a $n \times q$ matrix of infinite one-sided square-summable matrix polynomials and $\mathbf{u}_t$ is a q-dimensional vector of latent orthonormal white noise process with typically $q \ll n$ and $\boldsymbol{\xi}_t$ is allowed to be mildly cross-sectionally correlated and also autocorrelated. We will see that (6) can be derived from very few general assumptions and it is by all means the generalisation of the Wold representation theorem to the case of infinite dimensional panels of time series. In this sense is more primitive than the dynamic model in (2). Of course the two are related, in fact by using the Wold representation for the dynamic factors with orthonormal innovations[3]

$$\mathbf{f}_t = \mathbf{G}(L)\mathbf{u}_t, \quad \mathbf{u}_t \sim w.n.(\mathbf{0}, \mathbf{I}),$$

we see that (2) is equivalent to (6) just by setting $\mathbf{B}(L) = \boldsymbol{\Lambda}^*(L)\mathbf{G}(L)$. In this sense the approximate dynamic factor model (2), or equivalently (6), is the most general model we can consider when we have many series ($n \to \infty$). Finally, to avoid confusion, when, as in (6) the common factors are white noise, we call them *common shocks*.

The fact that we look for results when $n \to \infty$ is not just a mathematical artifact but it is required in order to deal with the approximate factor model case, indeed we will show that in this case common and idiosyncratic components can be disentangled only if $n \to \infty$, which in practice means having a large cross-section. In order words factors can be discovered more and more easily if we keep adding series to our dataset. If we can estimate (6) or equivalently (2), we have a divide and rule strategy to solve the problem of overparametrization: an infinite dimensional panel of time series is decomposed into a component driven by few factors and a component with small dependence. Since we can estimate the model only by letting $n \to \infty$, we can in this case speak of *blessing of dimensionality*. In Sections 2 and 3 we show the intuition of this concept in the i.i.d. case hence in a static factor model. Then in Sections 4 and 5 we describe representation theory and how to derive (6) using first a time-domain approach and then a spectral-domain approach.

[3]Notice that in general the innovations of the Wold representation are not necessarily orthonormal but for any generic innovation process $\boldsymbol{\eta}_t \sim w.n.(\mathbf{0}, \boldsymbol{\Gamma}_{\boldsymbol{\eta}})$ we can always build an orthonormal version of it by taking for example $\mathbf{u}_t = \boldsymbol{\Gamma}_{\boldsymbol{\eta}}^{-1/2}\boldsymbol{\eta}_t$, where $\boldsymbol{\Gamma}_{\boldsymbol{\eta}}^{-1/2} = \mathbf{D}_{\boldsymbol{\eta}}^{-1/2}\mathbf{W}_{\boldsymbol{\eta}}'$, such that $\mathbf{W}_{\boldsymbol{\eta}}$ is the matrix of normalised eigenvectors of $\boldsymbol{\Gamma}_{\boldsymbol{\eta}}$ and $\mathbf{D}_{\boldsymbol{\eta}}$ is the diagonal matrix of its eigenvalues.

STATE SPACE FORMULATION OF THE DYNAMIC MODEL. In order to estimate (6) or equivalently (2) the most common way is to consider the state space formulation by Giannone, Reichlin, and Sala (2005), Stock and Watson (2005), and Forni, Giannone, Lippi, and Reichlin (2009) which is given by

$$\begin{aligned} \mathbf{x}_t &= \mathbf{\Lambda}\mathbf{F}_t + \boldsymbol{\xi}_t \\ \mathbf{F}_t &= \mathbf{N}(L)\mathbf{u}_t, \quad \mathbf{u}_t \sim w.n.(\mathbf{0}, \mathbf{I}), \end{aligned} \tag{7}$$

where we have r static factors $\mathbf{F}_t$ which have a singular Wold representation, thus with $\mathbf{N}(L)$ being a $r \times q$ matrix of infinite one-sided square-summable matrix polynomials and $\mathbf{u}_t$ is q-dimensional orthonormal white noise. As noticed above, it is likely that $q \leq r$. Then, by merging the two equations in (7), we have

$$\mathbf{x}_t = \mathbf{\Lambda}\mathbf{N}(L)\mathbf{u}_t + \boldsymbol{\xi}_t \tag{8}$$

by setting $\mathbf{B}(L) = \mathbf{\Lambda}\mathbf{N}(L)$, we have again (6).

Moreover, while usually an inversion of an MA model implies an infinite VAR, in the singular case $r > q$ as above, it is shown by Anderson and Deistler (2008) that there exists an equivalent finite VAR representation for the factors, thus we can always write (7) as

$$\begin{aligned} \mathbf{x}_t &= \mathbf{\Lambda}\mathbf{F}_t + \boldsymbol{\xi}_t \\ \mathbf{A}(L)\mathbf{F}_t &= \mathbf{N}(0)\mathbf{u}_t, \quad \mathbf{u}_t \sim w.n.(\mathbf{0}, \mathbf{I}). \end{aligned} \tag{9}$$

As an example, assume that x_{it} is a macroeconomic variable driven by a cyclical indicator f_t such that $(1 - \alpha L)f_t = u_t$, where $|\alpha| < 1$ and $u_t \sim w.n.(0, 1)$ and assume the dynamic model with $q = 1$ (see also (2))

$$x_{it} = a_{0i}f_t + a_{1i}f_{t-1} + \xi_{it}, \qquad i = 1, \ldots, n. \tag{10}$$

Then if we define $F_{1t} = f_t$ and $F_{2t} = f_{t-1}$, (10) is also a static model with $r = 2$ (see also (1)). So we have one common shock and two static factors. Moreover, we have that the second equation in (7) reads as

$$\begin{pmatrix} F_{1t} \\ F_{2t} \end{pmatrix} = \begin{pmatrix} (1 - \alpha L)^{-1} \\ (1 - \alpha L)^{-1}L \end{pmatrix} u_t,$$

while by inverting the MA we can find that the second equation in (9) is a VAR(1) which reads as

$$\begin{pmatrix} (1 - \alpha L) & 0 \\ -L & 1 \end{pmatrix} \begin{pmatrix} F_{1t} \\ F_{2t} \end{pmatrix} = \begin{pmatrix} 1 \\ 0 \end{pmatrix} u_t,$$

or equivalently

$$\begin{pmatrix} F_{1t} \\ F_{2t} \end{pmatrix} = \begin{pmatrix} \alpha & 0 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} F_{1t-1} \\ F_{2t-1} \end{pmatrix} + \begin{pmatrix} 1 \\ 0 \end{pmatrix} u_t.$$

From the example is clear that if we just consider the first equation of (7) or (9), then this is equivalent to (6) if and only if $\mathbf{B}(L)$ has finite degree, which implies the existence of $\mathbf{F}_t$ with dimension r finite. This is a crucial issue which implies that when it comes to estimating (6) by means of the state space representation (9) and more assumptions are needed. In particular there is no representation result leading to (7), which is therefore called sometime *restricted approximate dynamic factor model* but we call it simply *approximate dynamic factor model*. In the applications, we consider mainly (7), which, once written as (9), is easy to estimate. This is the subject of Section 6, while estimation is discussed in Section 7.

ONE-SIDED REPRESENTATION. An alternative way of writing (6) using AR instead of MA polynomial is proposed by Forni, Hallin, Lippi, and Zaffaroni (2015a,b) and is based on the results of Anderson and Deistler (2008) for singular processes. In particular, we can write (6) as

$$\mathbf{A}(L)\mathbf{x}_t = \mathbf{G}\mathbf{u}_t + \boldsymbol{\xi}_t \quad \mathbf{u}_t \sim w.n.(\mathbf{0}, \mathbf{I}), \tag{11}$$

where $\mathbf{u}_t$ and $\boldsymbol{\xi}_t$ are the same as in (6), $\mathbf{G}$ is $n \times q$ and if we assume without loss of generality that

$n = m(q+1)$ for some integer $m < n$, then

$$\mathbf{A}(L) = \begin{pmatrix} \mathbf{A}^{(1)}(L) & \mathbf{0} & \dots & \mathbf{0} \\ \mathbf{0} & \mathbf{A}^{(2)}(L) & \dots & \mathbf{0} \\ \vdots & \ddots & \dots & \vdots \\ \mathbf{0} & \mathbf{0} & \dots & \mathbf{A}^{(m)}(L) \end{pmatrix}$$

where $\mathbf{A}^{(j)}(L)$ are polynomial matrices of finite degree and such that $\det(\mathbf{A}^{(j)}(z)) \neq 0$ for any $z \in \mathbb{C}$ such that $|z| \leq 1$, for $j = 1, \dots, m$. The result in (11) can be derived from the representation result (6) just by requiring the very mild additional assumption that an ARMA representation of the common components exists. For this reason (11) is called sometime *unrestricted generalised dynamic factor model*.

IDENTIFICATION. Notice that in general identification of $\mathbf{B}(L)$ in (6) is not possible. Indeed, for any invertible matrix $\mathbf{R}$ we can define an equivalent model

$$\mathbf{x}_t = \mathbf{E}(L)\mathbf{v}_t + \boldsymbol{\xi}_t = (\mathbf{B}(L)\mathbf{R}^{-1})(\mathbf{R}\mathbf{u}_t) + \boldsymbol{\xi}_t.$$

Analogously in (7) identification of $\mathbf{F}_t$ in (7) is not possible since for any invertible matrix $\mathbf{H}$ we can define an equivalent models

$$\mathbf{x}_t = \mathbf{M}\mathbf{G}_t + \boldsymbol{\xi}_t = (\boldsymbol{\Lambda}\mathbf{H}^{-1})(\mathbf{H}\mathbf{F}_t) + \boldsymbol{\xi}_t.$$

In terms of forecasting identification of the factors is not necessary and we just need to recover the space spanned by the factors, i.e. their linear combinations. Moreover, from the example (10) above, we see that in general we do not have to care about the economic interpretation of $\mathbf{F}_t$ and thus we are not interested in determining $\mathbf{H}$. But we are likely to be interested in interpreting (i.e. identifying) the effect of $\mathbf{u}_t$ on the elements of $\mathbf{x}_t$, thus we are often interested in determining $\mathbf{R}$ which must be such that $\mathbf{u}_t$ is orthonormal (see footnote 3 above). This is the subject of impulse response analysis (see below).

EXAMPLES OF APPLICATIONS. Let us consider forecasting and take an exact dynamic factor model. If we define the sets $\mathcal{X}_t = \overline{\text{span}}\{\mathbf{x}_s, s \leq t\}$ and similarly $\mathcal{U}_t$ and Ξ_t, the optimal one-step ahead forecast of series i is

$$\begin{aligned} \mathsf{E}[x_{it+1}|\mathcal{X}_t, \dots] &= \mathsf{E}[\boldsymbol{\lambda}_i^*(L)\mathbf{f}_{t+1} + \xi_{it+1}|\mathcal{X}_t \dots] \\ &= \mathsf{E}[\boldsymbol{\lambda}_i^*(L)\mathbf{f}_{t+1}|\mathcal{X}_t, \dots] + \mathsf{E}[\xi_{it+1}|\mathcal{X}_t, \dots] \\ &= \mathsf{E}[\boldsymbol{\lambda}_i^*(L)\mathbf{f}_{t+1}|\mathcal{U}_t \dots] + \mathsf{E}[\xi_{it+1}|\Xi_{it} \dots], \end{aligned}$$

thus in this case adding one series to the dataset does not increase the dimension of the problem. Forecasting is the subject of Section 8.

Another application is in structural economic analysis. Since Sims (1980) in structural VAR literature (SVAR) we have economic shocks driving observed variables and we are interested in their effect, i.e. in impulse response functions (IRF). These shocks are a linear combination of the Wold innovations, say $\mathbf{v}_t = \mathbf{R}\boldsymbol{\epsilon}_t$ where $\mathbf{R}$ is determined according to some economic restrictions that can be imposed. We know that a large n makes estimation of VAR impossible. On the other hand if n is small it is possible to show that the econometrician using standard estimation of SVAR models might not be able to recover space spanned the true shocks $\mathbf{v}_t$ driving the economy due to lack of information: this problem is called *nonfundamentalness*. A dynamic factor model has a common component $\boldsymbol{\chi}_t = \mathbf{B}(L)\mathbf{u}_t$ which is similar to the Wold representation (4) but has more variables than shocks, $n > q$. In this way we extend the information set of the econometrician, i.e. we use large datasets to solve nonfundamentalness and we use the factor model to solve the curse of dimensionality. As we have seen above $\mathbf{u}_t$ is not identified also in factor models, therefore we have to determine $\mathbf{R}$ also in this case. This can be done as in SVAR literature and the whole structural analysis in factor models is the subject of Section 9.

Finally, in portfolio optimisation the optimal weights to assign to each asset are usually determined by solving a risk minimisation problem which requires inversion of the covariance matrix of the data. However, in the large dimensional setting inverting the sample covariance is usually not feasible. With a factor model we are decomposing the second moment structure of the process into a reduced rank component plus

a sparse (weakly correlated) component and this approach can be used to obtain more regular estimators of large covariance matrices which have a well behaved inverse (Fan, Liao, and Mincheva, 2013). Other applications in finance show how to recover market volatility shocks from large panels of returns thus allowing to model volatility dynamics starting from observed data in the same spirit as GARCH models. This is the subject of Section 10.

HISTORY. The father of factor models is Spearman (1863-1945) who proposed an exact static model and its estimation by means of Maximum Likelihood was developed by Jöreskog (1969) and Lawley and Maxwell (1962, 1971) and by means of Principal Component Analysis by Tipping and Bishop (1999). This model is useful for i.i.d. data.

For stationary time series an exact dynamic model was proposed by Geweke (1977) and its estimation by means of the Kalman filter and the Expectation Maximisation algorithm was studied by Sargent and Sims (1977) in the spectral domain, and Watson and Engle (1983) and Quah and Sargent (1993) in the time domain. An approximate static model was proposed by Chamberlain and Rothschild (1983) and Connor and Korajczyk (1986) and its estimation by means of Principal Component Analysis was considered by Stock and Watson (2002a,b), Bai (2003), and Fan, Liao, and Mincheva (2013).

Combining the last two models Forni and Lippi (2001, 2011); Hallin and Lippi (2013) proposed the generalised dynamic factor model (aka GDFM) and its unrestricted representation. Estimation of the GDFM has been developed by Forni, Hallin, Lippi, and Reichlin (2000, 2005) and Forni, Hallin, Lippi, and Zaffaroni (2015a,b) using Dynamic Principal Component Analysis, that is in the spectral domain, and results on singular AR representations, respectively.

The restricted approximate dynamic factor model, i.e. in state-space form, has been proposed by Stock and Watson (2005). Two-step estimation by means of Principal Component Analysis and VAR was studied by Bai and Ng (2006, 2007) and Forni, Giannone, Lippi, and Reichlin (2009), while Doz, Giannone, and Reichlin (2011, 2012) proposed estimation by means of the Kalman filter and the Expectation Maximization algorithm. More details on estimation of this class of factor models follow.

OVERVIEW OF ESTIMATION TECHNIQUES. Depending on the use we want to make of the approximate dynamic factor model (7)-(9), we need estimators of

- the factors $\mathbf{F}_t$ or at least of the space they span;
- the parameters of the model, i.e. of $\mathbf{\Lambda}$, $\mathbf{G}(L)$, and $\mathbf{N}(0)$;
- the number of factors r and the number of shocks q.

There are two main classes of estimators of the approximate dynamic factor model in state-space form and here we list those proposed for the large n case

1. two-step estimators based on Principal Component Analysis (PCA) and a VAR on the estimated factors;
2. on-line techniques based on Kalman Filter (KF) and Expectation Maximisation (EM) algorithm which is equivalent to Quasi Maximum Likelihood (QML).

These methods are considered in Sections 6 and 7, respectively. In the same spirit of PCA is Forni, Hallin, Lippi, and Reichlin (2000) which estimates the space spanned by $\mathbf{u}_t$ using Dynamic PCA, it is not useful for forecasting but it is used as the first step of other estimators not considered here (Forni, Hallin, Lippi, and Reichlin, 2005; Forni, Hallin, Lippi, and Zaffaroni, 2015a,b).

For determining the number of factors and shocks there are many criteria available and most are based on the behaviour of the eigenvalues of the covariance and spectral density matrix of the data. This is the subject of Section 11.

2 Principal component analysis and the exact factor model

In this section we consider n fixed and finite. Moreover, in this and the next section we assume to have a panel of i.i.d. data, i.e. with no serial correlation, $\mathsf{Cov}(x_{it}, x_{jt-k}) = 0$ for any $k > 0$ and any $i, j = 1, \ldots, n$. Therefore, the index $t \in \mathbb{Z}$ represents different observations not necessarily in time. We consider an n-dimensional vector $\mathbf{x}_t = (x_{1t} \ldots x_{nt})'$ such that $\mathsf{E}[\mathbf{x}_t] = \mathbf{0}$, and we define $\mathbf{\Gamma}_\mathbf{x} = \mathsf{E}[\mathbf{x}_t \mathbf{x}_t']$. The aim is to explore the reaction between the common dimension reduction technique of principal component analysis and the simplest factor structure.

PRINCIPAL COMPONENT ANALYSIS. The covariance matrix $\mathbf{\Gamma}_\mathbf{x}$ is a symmetric and positive definite matrix and can always be factorised as

$$\mathbf{\Gamma}_\mathbf{x} = \mathbf{PDP}',$$

where $\mathbf{P}$ is an $n \times n$ orthogonal matrix of eigenvectors, i.e. $\mathbf{PP}' = \mathbf{P}'\mathbf{P} = \mathbf{I}_n$ and $\mathbf{P} = (\mathbf{p}_1 \ldots \mathbf{p}_n)$, and $\mathbf{D} = \mathrm{diag}(\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_n > 0)$ contains the ordered eigenvalues.

Then,

$$\mathbf{\Gamma}_\mathbf{x}\mathbf{p}_j = \lambda_j \mathbf{p}_j \text{ and } \mathbf{\Gamma}_\mathbf{x} = \sum_{j=1}^{n} \lambda_j \mathbf{p}_j \mathbf{p}_j'.$$

Recall that the eigenvalues solve

$$\det(\mathbf{\Gamma}_\mathbf{x} - \lambda_j \mathbf{I}_n) = 0,$$

then the once we have the eigenvalues, the eigenvectors can be found by solving

$$(\mathbf{\Gamma}_\mathbf{x} - \lambda_j \mathbf{I}_n)\mathbf{p}_j = \mathbf{0}.$$

Equivalently, eigenvalues and eigenvectors are the solutions of the problem

$$\begin{aligned} \mathbf{p}_1 &= \arg\max_{\mathbf{a}} \mathbf{a}'\mathbf{\Gamma}_\mathbf{x}\mathbf{a} \quad \text{s.t. } \mathbf{a}'\mathbf{a} = 1, \\ \mathbf{p}_j &= \arg\max_{\mathbf{a}} \mathbf{a}'\mathbf{\Gamma}_\mathbf{x}\mathbf{a} \quad \text{s.t. } \mathbf{a}'\mathbf{a} = 1 \text{ and } \mathbf{a}'\mathbf{p}_1 = \ldots = \mathbf{a}'\mathbf{p}_{j-1} = 0, \quad j > 1 \end{aligned}$$

the eigenvalues being the value of the functions at their maximum. In other words, if we write the first eigenvector as $\mathbf{p}_1 = (p_{11} \ldots p_{1n})'$ then its components are such that

$$(p_{11} \ldots p_{1n}) = \arg\max_{(a_1 \ldots a_n)} \mathsf{Var}\left(\sum_{i=1}^{n} a_i x_{it}\right) \text{ s.t. } \sum_{i=1}^{n} a_i^2 = 1.$$

The entries of the largest eigenvector are the normalised weights of the weighted average of the data with maximum variance which is given by the largest eigenvalue

$$\mathsf{Var}(\mathbf{p}_1'\mathbf{x}_t) = \mathsf{Var}\left(\sum_{i=1}^{n} p_{1i} x_{it}\right) = \sum_{i,j=1}^{n} p_{1i}p_{1j}\mathsf{E}[x_{it}x_{jt}] = \mathbf{p}_1'\mathbf{\Gamma}_\mathbf{x}\mathbf{p}_1 = \lambda_1.$$

If we consider all normalised linear combinations of the data $\mathbf{a}'\mathbf{x}_t$, then $\mathbf{p}_1'\mathbf{x}_t = \mathbf{x}_t'\mathbf{p}_1$ is the one with largest variance λ_1, and so on for the other eigenvectors. The linear combination $\mathbf{p}_j'\mathbf{x}_t$ is called j-th principal component of $\mathbf{x}_t$ and has variance λ_j. The projection of $\mathbf{x}_t$ onto the j-th principal component is the vector

$$\mathbf{y}_j = \left(\mathbf{p}_j'\mathbf{x}_t\right)\mathbf{p}_j,$$

which has length λ_j and direction $\mathbf{p}_j$.

Assume that we want to reduce the dimension of the data at hand from n to $q < n$, consider the de-

composition dictated by principal components

$$\mathbf{x}_t = \sum_{j=1}^{n} \left(\mathbf{p}_j' \mathbf{x}_t\right) \mathbf{p}_j = \underbrace{\sum_{j=1}^{q} \left(\mathbf{p}_j' \mathbf{x}_t\right) \mathbf{p}_j}_{\mathbf{x}_{t,[q]}} + \underbrace{\sum_{j=q+1}^{n} \left(\mathbf{p}_j' \mathbf{x}_t\right) \mathbf{p}_j}_{\mathbf{x}_t - \mathbf{x}_{t,[q]}}.$$

We define the projection of $\mathbf{x}_t$ orthogonal to $\mathbf{x}_{t,[q]}$ as $\boldsymbol{\xi}_t = \mathbf{x}_t - \mathbf{x}_{t,[q]}$ with covariance $\boldsymbol{\Gamma}_{\boldsymbol{\xi}}$. Then, consider any other projection say $\widetilde{\mathbf{x}}_{t,[q]}$ with residual $\widetilde{\boldsymbol{\xi}}_t = \mathbf{x}_t - \widetilde{\mathbf{x}}_{t,[q]}$ with covariance $\boldsymbol{\Gamma}_{\widetilde{\boldsymbol{\xi}}}$. It can be shown that $\mathbf{x}_{t,[q]}$ is the best linear q-dimensional representation of $\mathbf{x}_t$ in the sense that

$$\mathrm{tr}\Big(\boldsymbol{\Gamma}_{\boldsymbol{\xi}}\Big) \leq \mathrm{tr}\Big(\boldsymbol{\Gamma}_{\widetilde{\boldsymbol{\xi}}}\Big).$$

Thus, this decomposition minimises the sum of the residual variances which is equivalent to minimising the sum of all eigenvalues of the residual covariance $\boldsymbol{\Gamma}_{\boldsymbol{\xi}}$. Notice that the covariance matrix of $\mathbf{x}_{t,[q]}$ has just q eigenvalues different from zero and therefore has rank q, while the covariance matrix of $\boldsymbol{\xi}_t$ has rank $n - q$. This dimension reduction technique is called Principal Component Analysis (PCA).

FACTOR MODEL. Now assume the static exact factor model for $\mathbf{x}_t$

$$\mathbf{x}_t = \mathbf{B}\mathbf{f}_t + \boldsymbol{e}_t,$$

where $\mathbf{f}_t = (f_{1t} \dots f_{qt})'$ is a vector of q unobserved factors with $q < n$, $\mathbf{B}$ is a $n \times q$ matrix of loadings, and $\boldsymbol{e}_t$ is an idiosyncratic error component as opposed to the common component $\boldsymbol{\chi}_t = \mathbf{B}\mathbf{f}_t$.

We assume orthogonality between the common and idiosyncratic components, $\mathsf{E}[\mathbf{f}_t\, \boldsymbol{e}_t'] = \mathbf{0}$, and an exact factor structure, i.e. we require $\mathsf{E}[\boldsymbol{e}_t\, \boldsymbol{e}_t'] = \boldsymbol{\Gamma}_{\boldsymbol{e}}$ to be a positive definite and diagonal matrix. Finally, without loss of generality, we can impose $\mathsf{E}[\mathbf{f}_t\, \mathbf{f}_t'] = \mathbf{I}_q$. Therefore,

$$\boldsymbol{\Gamma}_{\mathbf{x}} = \mathbf{B}\mathbf{B}' + \boldsymbol{\Gamma}_{\boldsymbol{e}} = \boldsymbol{\Gamma}_{\boldsymbol{\chi}} + \boldsymbol{\Gamma}_{\boldsymbol{e}},$$

where $\mathbf{B}\mathbf{B}' = \boldsymbol{\Gamma}_{\boldsymbol{\chi}}$ has rank q and is accounting for all off-diagonal elements of $\boldsymbol{\Gamma}_{\mathbf{x}}$ while $\boldsymbol{\Gamma}_{\boldsymbol{e}}$ has rank n.

COMPARISON. The two models imply that any series can be written as

$$\begin{aligned} \text{PC} \quad & x_{it} = p_{1i}(\mathbf{p}_1'\mathbf{x}_t) + \ldots + p_{qi}(\mathbf{p}_q'\mathbf{x}_t) + \xi_{it}, \\ \text{FM} \quad & x_{it} = B_{i1} f_{1t} + \ldots + B_{iq} f_{qt} + e_{it}. \end{aligned}$$

Although these might seem different they both decompose of the covariance matrix into a rank q matrix plus another matrix:

$$\begin{aligned} \text{PC} \quad \boldsymbol{\Gamma}_{\mathbf{x}} &= \boldsymbol{\Gamma}_{\mathbf{x}_{[q]}} + \boldsymbol{\Gamma}_{\boldsymbol{\xi}} \\ \text{FM} \quad \boldsymbol{\Gamma}_{\mathbf{x}} &= \boldsymbol{\Gamma}_{\boldsymbol{\chi}} + \boldsymbol{\Gamma}_{\boldsymbol{e}} \end{aligned}$$

The difference between the two cases is clear by looking at the second term:

- in PCA the sum of variances of the idiosyncratic components $\boldsymbol{\xi}_t$ has to be minimised, this implies minimising the sum of all eigenvalues, hence the sum of the covariances is minimised, but $\boldsymbol{\Gamma}_{\boldsymbol{\xi}}$ is not necessarily diagonal, this is a *representation* result as it holds for any covariance matrix, it is just an algebraic result and it does not imply any statistical modelling;
- in an exact static factor model $\boldsymbol{\Gamma}_{\boldsymbol{e}}$ is assumed to be diagonal, therefore the common component $\boldsymbol{\chi}_t$ accounts for all covariances, now the covariance of the idiosyncratic component $\boldsymbol{e}_t$ is shrunk to zero while the idiosyncratic variances may still be large; this is a *model* as it is based on assumptions on the data generating process.

Ideally we would like to reconcile the two approaches since PCA is a very simple and intuitive dimension reduction technique. Can we have a factor model that is compatible with the PCA decomposition? In order

to do this we first need to allow for cross-sectional dependence among idiosyncratic components, i.e. to consider an approximate factor model. Such model can then be identified and then estimated only for an infinite dimensional panel, i.e. when $n \to \infty$.

3 The approximate factor model and the blessing of dimensionality

Now let $n \to \infty$ and for any $n \in \mathbb{N}$ consider the process $\mathbf{x}_t^{(n)}$, such that for a fixed q we have the factor model[4]

$$\mathbf{x}_t^{(n)} = \mathbf{B}^{(n)}\mathbf{f}_t + \boldsymbol{\xi}_t^{(n)}, \tag{12}$$

and define $\boldsymbol{\chi}_t^{(n)} = \mathbf{B}^{(n)}\mathbf{f}_t$. The covariance matrix is given by:

$$\boldsymbol{\Gamma}_{\mathbf{x}}^{(n)} = \mathbf{B}^{(n)}\mathbf{B}^{(n)'} + \boldsymbol{\Gamma}_{\boldsymbol{\xi}}^{(n)} = \boldsymbol{\Gamma}_{\boldsymbol{\chi}}^{(n)} + \boldsymbol{\Gamma}_{\boldsymbol{\xi}}^{(n)}.$$

CROSS-SECTIONAL AVERAGING - PART 1. The factors have an intuitive meaning in terms of aggregation. Consider as an aggregate the cross-sectional average of the data $\bar{x}_t = \frac{1}{n}\sum_{i=1}^n x_{it}$ and assume one factor, i.e. $q = 1$. In an exact factor model the variance of this aggregate is

$$\mathsf{Var}(\bar{x}_t) = \frac{1}{n^2}\left(\sum_{i=1}^n b_i\right)^2 \mathsf{Var}(f_t) + \frac{1}{n^2}\mathsf{Var}\left(\sum_{i=1}^n \xi_{it}\right) = \bar{b}^2 + \frac{1}{n^2}\sum_{i=1}^n \mathsf{Var}(\xi_{it}),$$

where b_i are the loadings and $\bar{b} = \frac{1}{n}\sum_{i=1}^n b_i$ and $\mathsf{Var}(f_t) = 1$. Then, if all b_i are non-zero, $\bar{b}^2$ is finite for any n, the above expression is such that

$$\mathsf{Var}(\bar{x}_t) \to \bar{b}^2 = \mathsf{Var}(\bar{\chi}_t), \qquad n \to \infty. \tag{13}$$

This means that all information contained in the factors as measured by second moments can be recovered by means of aggregation of observed data. We say that as $n \to \infty$ the model is identified, that is we can separate common and idiosyncratic component and therefore we can work only with the form thus effectively reducing the dimension of the problem. Notice that for (13) to hold we need all (or most) b_i to be non-zero, that is the factor has to be pervasive.

This example shows that factor modelling is about capturing covariances not variances. Consider an even simpler example of an exact factor model, where $b_i = 1$ for all i,

$$x_{it} = f_t + \xi_{it}$$

and with $\mathsf{Var}(f_t) = 1$ and $\mathsf{Var}(\xi_{it}) = C > 1$ for all i. Thus, for each series the factor accounts for a small portion of the total variance, indeed for any i we have

$$\frac{\mathsf{Var}(f_t)}{\mathsf{Var}(x_{it})} = \frac{1}{1+C} < \frac{1}{2}.$$

However, if we consider the cross-sectional average we have that

$$\frac{\mathsf{Var}(f_t)}{\mathsf{Var}(\bar{x}_t)} = \frac{\mathsf{Var}(f_t)}{\mathsf{Var}(f_t) + \frac{1}{n}\mathsf{Var}(\xi_{it})} = \frac{1}{1+\frac{C}{n}} \to 1, \qquad n \to \infty.$$

Asymptotically, the variance of the factor accounts for all the variance of the aggregate.

Now consider the case in which we allow for cross-sectional dependence among idiosyncratic component, then the variance of the aggregate is

$$\mathsf{Var}(\bar{x}_t) = \frac{1}{n^2}\left(\sum_{i=1}^n b_i\right)^2 \mathsf{Var}(f_t) + \frac{1}{n^2}\mathsf{Var}\left(\sum_{i=1}^n \xi_{it}\right) = \bar{b}^2 + \frac{1}{n^2}\sum_{i=1}^n\sum_{j=1}^n \mathsf{Cov}(\xi_{it}, \xi_{jt}). \tag{14}$$

Even if we assume pervasiveness of the factor we cannot obtain a result as in (13) unless we bound somehow the second term. Notice that since we are summing n^2 covariances and dividing by n^2 as $n \to \infty$ that term is in general $O(1)$ as the first one.

[4] In these sections we use the superscript n to emphasize the role of n.

Thus, as $n \to \infty$, we need both an assumption on the loadings and on the covariances of the idiosyncratic terms in order to disentangle common from idiosyncratic components.

IDENTIFYING ASSUMPTIONS. We assume that there exist $\alpha, \beta < \infty$ such that

(a) *bounded idiosyncratic variances*: $0 < \mathsf{Var}(\xi_{it}) < \alpha$ for any $i \in \mathbb{N}$;

(b) *pervasiveness of the factors*, i.e. each of the q factors has an impact on each component of $\mathbf{x}_t$ that is bounded away from zero:

$$\min_{j=1,\dots,q} \liminf_{i\to\infty} |b_{ij}| > \beta > 0,$$

and the factors have finite variance: $0 < \mathsf{Var}(f_{jt}) < \gamma$, for $j = 1, \dots, q$.

Assumptions *(a)* and *(b)* are very basic and not testable. Alternatively and equivalently, we can assume

(a') $\mathbf{\Gamma}_{\boldsymbol{\xi}}^{(n)}$ has all eigenvalues bounded for any n, that is there exists an M independent of n such that

$$\lambda_{1\boldsymbol{\xi}}^{(n)} \leq M < \infty, \quad n \in \mathbb{N};$$

(b') $\mathbf{\Gamma}_{\boldsymbol{\chi}}^{(n)}$ has all q eigenvalues diverging as $n \to \infty$, that is

$$\lim_{n\to\infty} \lambda_{q\boldsymbol{\chi}}^{(n)} = \infty.$$

Notice that if *(a)* holds then for any n

$$\frac{1}{n}\mathrm{Tr}(\mathbf{\Gamma}_{\boldsymbol{\xi}}^{(n)}) = O(1),$$

which is also implied by *(a')*. By assuming directly *(a')* we do not have to require $\mathbf{\Gamma}_{\boldsymbol{\xi}}^{(n)}$ to be diagonal and we are indirectly implying that only a limited amount of cross-sectional dependence is allowed for (see below for the formal implication).

Moreover, *(b)* implies *(b')*. For example if $q = 1$ and $\mathsf{Var}(f_t) = 1$, then $\mathbf{\Gamma}_{\boldsymbol{\chi}}^{(n)} = \mathbf{B}^{(n)}\mathbf{B}^{(n)'}$ has just one non-zero eigenvalue which is equal to $\mathbf{B}^{(n)'}\mathbf{B}^{(n)}$ (it is a scalar in this case) and such that because of *(b)*

$$\lim_{n\to\infty} \lambda_{1\boldsymbol{\chi}}^{(n)} = \lim_{n\to\infty} \mathbf{B}^{(n)'}\mathbf{B}^{(n)} = \lim_{n\to\infty} \sum_{i=1}^{n} b_i^2 > \lim_{n\to\infty} n\beta^2 = \infty.$$

Under assumptions *(a')* and *(b')* and using the PCA decomposition of the covariances $\mathbf{\Gamma}_{\boldsymbol{\chi}}^{(n)}$ and $\mathbf{\Gamma}_{\boldsymbol{\xi}}^{(n)}$, we have

$$\mathbf{\Gamma}_{\mathbf{x}}^{(n)} = \underbrace{\sum_{j=1}^{q} \lambda_{j\boldsymbol{\chi}}^{(n)} \mathbf{p}_{j\boldsymbol{\chi}}^{(n)} \mathbf{p}_{j\boldsymbol{\chi}}^{(n)'}}_{\mathbf{\Gamma}_{\boldsymbol{\chi}}^{(n)}} + \underbrace{\sum_{j=1}^{n} \lambda_{j\boldsymbol{\xi}}^{(n)} \mathbf{p}_{j\boldsymbol{\xi}}^{(n)} \mathbf{p}_{j\boldsymbol{\xi}}^{(n)'}}_{\mathbf{\Gamma}_{\boldsymbol{\xi}}^{(n)}}, \tag{15}$$

where, as $n \to \infty$, $\lambda_{j\boldsymbol{\chi}}^{(n)} \to \infty$, while $\lambda_{j\boldsymbol{\xi}}^{(n)}$ stay bounded.

Then from (15) and Weyl's inequality we have

$$\lambda_{j\boldsymbol{\chi}}^{(n)} + \lambda_{n\boldsymbol{\xi}}^{(n)} \leq \lambda_{j\mathbf{x}}^{(n)} \leq \lambda_{j\boldsymbol{\chi}}^{(n)} + \lambda_{1\boldsymbol{\xi}}^{(n)}, \quad j = 1, \dots n,$$

and therefore

1. for $j \leq q$, as $n \to \infty$, $\lambda_{j\boldsymbol{\chi}}^{(n)} \to \infty$, then $\lambda_{j\mathbf{x}}^{(n)} \to \infty$,
2. for $j > q$, $\lambda_{j\boldsymbol{\chi}}^{(n)} = 0$, then $\lambda_{j\mathbf{x}}^{(n)} \leq \lambda_{1\boldsymbol{\xi}}^{(n)} \leq M < \infty$.

This result implies that under a factor model with assumptions *(a')* and *(b')*, the q largest eigenvalues of $\mathbf{\Gamma}_{\mathbf{x}}^{(n)}$ diverge while the other $(n-q)$ stay bounded. However, the viceversa does not hold, therefore this is not a representation result, the behavior of the eigenvalues of $\mathbf{\Gamma}_{\mathbf{x}}^{(n)}$ does not in general imply assumptions *(a')* and *(b')*, i.e. a factor structure with pervasive shocks and mildly correlated idiosyncratic component.

ALTERNATIVE ASSUMPTIONS. Instead of assumptions *(a')* and *(b')* Bai and Ng (2002) assume that

(a") $\frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{n}|\mathsf{Cov}(\xi_{it},\xi_{jt})| \leq M < \infty$ for any n and for some positive constant M;

(b") $\frac{\mathbf{B}^{(n)\prime}\mathbf{B}^{(n)}}{n} \to \mathbf{H}$ as $n \to \infty$ where $\mathbf{H}$ is $q \times q$ and positive definite and $\mathbf{\Gamma_f}$ is positive definite.

Notice that since $\mathbf{H}$ does not depend on n all its eigenvalues are finite. Moreover, *(b")* implies *(b')*. Consider the eigenvalues decomposition $\mathbf{\Gamma_f} = \mathbf{P_f M_f P_f'}$, then

$$\frac{1}{n}\mathbf{P_f}\mathbf{M_f}^{1/2}\mathbf{B}^{(n)\prime}\mathbf{B}^{(n)}\mathbf{M_f}^{1/2}\mathbf{P_f'} \to \mathbf{\Gamma_f}^{1/2}\mathbf{H}\mathbf{\Gamma_f}^{1/2}, \quad \text{as } n \to \infty, \tag{16}$$

which is positive definite because of *(b")*. Then, consider

$$\frac{1}{n}\mathbf{\Gamma}_{\boldsymbol{\chi}}^{(n)} = \frac{1}{n}\mathbf{B}^{(n)}\mathbf{\Gamma_f}\mathbf{B}^{(n)\prime} = \frac{1}{n}\mathbf{B}^{(n)}\mathbf{P_f}\mathbf{M_f}^{1/2}\mathbf{M_f}^{1/2}\mathbf{P_f'}\mathbf{B}^{(n)\prime},$$

which as $n \to \infty$ has the same non-zero eigenvalues as (16) and these q and are all finite. Therefore, the eigenvalues of $\mathbf{\Gamma}_{\boldsymbol{\chi}}^{(n)}$ diverge linearly in n.

As for *(a')* and *(a")* we do not have a direct implication, but notice that if we assume that there exists an M independent of n such that

$$\max_{i=1,\dots,n}\sum_{j=1}^{n}|\mathsf{Cov}(\xi_{it},\xi_{jt})| \leq M < \infty, \tag{17}$$

then both *(a')* and *(a")* are implied, indeed[5]

$$\begin{aligned}
&\textit{(a')} \quad \lambda_{1\boldsymbol{\xi}} = \|\mathbf{\Gamma}_{\boldsymbol{\xi}}^{(n)}\| \leq \|\mathbf{\Gamma}_{\boldsymbol{\xi}}^{(n)}\|_1 = \max_{i=1,\dots,n}\sum_{j=1}^{n}|\mathsf{Cov}(\xi_{it},\xi_{jt})| \leq M < \infty,\\
&\textit{(a")} \quad \frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{n}|\mathsf{Cov}(\xi_{it},\xi_{jt})| \leq \max_{i=1,\dots,n}\sum_{j=1}^{n}|\mathsf{Cov}(\xi_{it},\xi_{jt})| \leq M < \infty.
\end{aligned}$$

In particular, Fan, Liao, and Mincheva (2013) show that assumption (17) is equivalent to asking for sparsity of the covariance matrix of the idiosyncratic component in the sense that its column norm is $o(n)$ which means that for large n many of its entries are likely to be vanishing.

Moreover, either by assuming *(a')* or *(a")* we have

$$\begin{aligned}
&\textit{(a')} \quad \frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{n}\mathsf{Cov}(\xi_{it},\xi_{jt}) = \frac{1}{n}\boldsymbol{\iota}'\mathbf{\Gamma}_{\boldsymbol{\xi}}^{(n)}\boldsymbol{\iota} \leq \lambda_{1\boldsymbol{\xi}} \leq M < \infty,\\
&\textit{(a")} \quad \frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{n}\mathsf{Cov}(\xi_{it},\xi_{jt}) \leq \frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{n}|\mathsf{Cov}(\xi_{it},\xi_{jt})| \leq M < \infty.
\end{aligned} \tag{18}$$

where $\boldsymbol{\iota}$ is an n-dimensional vector of ones. In other words, (18) shows that the total amount of cross-sectional dependence among idiosyncratic components is limited and we say that they are weakly or mildly

[5] In the first case we use Hölder inequality and the fact that column norm $\|\cdot\|_1$ and row norm $\|\cdot\|_\infty$ are identical for a symmetric matrix.

correlated. This is what characterises the approximate factor model.

Summing up, whichever set of assumptions we choose the main message is that we require

1. the factors to be pervasive i.e. to have non-negligible effects on most of the variables (assumptions *(b)* or *(b')* or *(b")*);
2. the idioscyncratic components to be weakly cross-correlated (assumptions *(a)* or *(a')* or *(a")*).

Under these conditions when $n \to \infty$ we can always disentangle the common from the idiosyncratic component by means of aggregation.

CROSS-SECTIONAL AVERAGING - PART 2. If we go back to the example above of a cross-sectional average of the data in an approximate factor model, we have

$$\mathsf{Var}(\bar{x}_t) = \bar{b}^2 + \frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n}\mathsf{Cov}(\xi_{it}, \xi_{jt}). \tag{19}$$

Now, under assumption *(b)* we have that $\bar{b}^2$ is a finite constant not depending on n, i.e. the first term in (19) is $O(1)$. Moreover, by assumption *(a')* or *(a")*, the second term in (19) is

$$\frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n}\mathsf{Cov}(\xi_{it}, \xi_{jt}) \leq \frac{M}{n}.$$

Therefore, as $n \to \infty$ we can disentangle the information in the common and idiosyncratic component

$$\mathsf{Var}(\bar{x}_t) = \bar{b}^2 + O\left(\frac{1}{n}\right) \to \bar{b}^2 = \mathsf{Var}(\bar{\chi}_t) \text{ as } n \to \infty, \tag{20}$$

and the factors represent the part of the data that under aggregation has a non-negligible contribution in terms of variance.

GENERALISATION TO WEIGHTED CROSS-SECTIONAL AVERAGES. We have seen that we can recover the factors' variance by means of cross-sectional averaging. The weights used in the previous example are all equal to $1/n$ and therefore they are not normalised. To keep the analogy with eigenvectors, which will be useful later, we hereafter consider weighted averages with normalised weights. The main message of (20) remains unchanged if we change scale of the weights.

In general, in a q factor model we look for q mutually orthogonal linear combinations

$$w_{jt}^{(n)} = (a_{1j} \dots a_{nj})\mathbf{x}_t^{(n)} = \mathbf{a}_j^{(n)'}\mathbf{x}_t^{(n)} = \mathbf{a}_j^{(n)'}\boldsymbol{\chi}_t^{(n)} + \mathbf{a}_j^{(n)'}\boldsymbol{\xi}_t^{(n)}, \quad j = 1, \dots, q,$$

such that $\mathbf{a}_j^{(n)'}\mathbf{a}_j^{(n)} = 1$ for any j and to avoid trivial combinations we assume $|a_{ij}| > 0$ for any $i \in \mathbb{N}$ and $j = 1, \dots q$. These two conditions in turn imply that we must have $a_{ij} = \frac{k_{ij}}{\sqrt{n}}$ for some finite constants $k_{ij} \neq 0$.

Then, because of assumption *(a')*

$$\mathsf{Var}(\mathbf{a}_j^{(n)'}\boldsymbol{\xi}_t^{(n)}) = \mathbf{a}_j^{(n)'}\boldsymbol{\Gamma}_{\boldsymbol{\xi}}^{(n)}\mathbf{a}_j^{(n)} \leq \lambda_{1\boldsymbol{\xi}}^{(n)} \leq M < \infty,$$

On the other hand because of assumption *(b")* we must have $\mathbf{b}_i'\mathbf{b}_k = \sum_{j=1}^{q} b_{ij}b_{kj} = O(n)$, and without loss of generality we can assume $\boldsymbol{\Gamma}_{\mathbf{f}} = \mathbf{I}_q$ and $\mathbf{B}^{(n)'}\mathbf{B}^{(n)}$ to be orthogonal, which implies

$$\mathsf{Var}(\mathbf{a}_j^{(n)'}\boldsymbol{\chi}_t^{(n)}) = \mathbf{a}_j^{(n)'}\mathbf{B}^{(n)}\mathbf{B}^{(n)'}\mathbf{a}_j^{(n)} = \left(\sum_{i=1}^{n} a_{ij}\mathbf{b}_i'\right)\left(\sum_{i=1}^{n} a_{ij}\mathbf{b}_i\right) = \frac{1}{n}\sum_{i}^{n} k_{ij}^2\mathbf{b}_i'\mathbf{b}_i = O(n).$$

If we instead rely only just on *(b)* or *(b')* we simply have that the last expression diverges as $n \to \infty$ but at unknown rate. In any case, as in the previous example, a cross-sectional weighted average of $\mathbf{x}_t$ is decomposed into a part with diverging variance corresponding to the common component, and a part with finite variance corresponding to the idiosyncratic component:

$$\mathsf{Var}(w_{jt}^{(n)}) = \mathsf{Var}(\mathbf{a}_j^{(n)'}\boldsymbol{\chi}_t^{(n)}) + \mathsf{Var}(\mathbf{a}_j^{(n)'}\boldsymbol{\xi}_t^{(n)}) = \mathsf{Var}(\mathbf{a}_j^{(n)'}\boldsymbol{\chi}_t^{(n)}) \left(1 + \frac{\mathsf{Var}(\mathbf{a}_j^{(n)'}\boldsymbol{\xi}_t^{(n)})}{\mathsf{Var}(\mathbf{a}_j^{(n)'}\boldsymbol{\chi}_t^{(n)})}\right), \tag{21}$$

the last term being negligible when $n \to \infty$.

In order to have finite limiting quantities we can standardise the aggregates. There are two possibilities

$$\dot{w}_{jt}^{(n)} := \frac{w_{jt}^{(n)}}{\sqrt{\mathsf{Var}(w_{jt}^{(n)})}} = \frac{\mathbf{a}_j^{(n)'}\boldsymbol{\chi}_t^{(n)}}{\sqrt{\mathsf{Var}(\mathbf{a}_j^{(n)'}\boldsymbol{\chi}_t^{(n)})}} + o_{q.m.}(1),$$
$$\ddot{w}_{jt}^{(n)} := \frac{w_{jt}^{(n)}}{\sqrt{n}} = \frac{\mathbf{a}_j^{(n)'}\boldsymbol{\chi}_t^{(n)}}{\sqrt{n}} + O_{q.m.}\left(\frac{1}{\sqrt{n}}\right),$$

where $q.m.$ stands for "quadratic mean".[6] Notice that in the first case is more general and holds as long as *(b')* holds, that is without specifying any rate of divergence of eigenvalues. While the second case holds if we make use of assumption *(b")* which implies that the divergence is linear in n. In both cases the first term has finite variance by construction while the second term tends to zero as $n \to \infty$. Notice that in the first case we just have convergence, while in the second case we also have information about the rate of convergence. The linear divergence is a natural choice since it means that we discover factors as the cross-sectional dimension increases linearly, i.e. each new series carries information about the factors. Hence, we consider only this case hereafter.[7]

RECONSTRUCTING THE FACTORS' SPACE. If we collect q aggregates in a vector $\mathbf{w}_t^{(n)} = (w_{1t}^{(n)} \dots w_{qt}^{(n)})'$, we can consider an $n \times q$ matrix of weights $\mathbf{A}^{(n)}$ with columns $\mathbf{a}_j^{(n)}$ of dimension $n \times 1$ for $j = 1, \dots, q$, and such that we impose the normalisation

$$\mathbf{A}^{(n)'}\mathbf{A}^{(n)} = \mathbf{I}_q. \tag{22}$$

Then, we will work with the standardised weighted average

$$\ddot{\mathbf{w}}_t^{(n)} = \frac{\mathbf{A}^{(n)'}\mathbf{x}_t^{(n)}}{\sqrt{n}}.$$

Notice that a matrix with as columns the eigenvectors of $\boldsymbol{\Gamma}_{\boldsymbol{\chi}}^{(n)}$ satisfies the normalisation requirement and delivers weighted averages $\ddot{\mathbf{w}}_t^{(n)}$ with maximum variance (see below for details).[8]

In order to recover the space spanned by the factors $\mathbf{f}$, we add one more assumption

$$\frac{\mathbf{A}^{(n)'}\mathbf{B}^{(n)}}{\sqrt{n}} \to \mathbf{H}, \text{ as } n \to \infty, \tag{23}$$

with $\mathbf{H}$ positive definite and again since it does not depend on n all its eigenvalues are finite. Notice that this is reasonable since by *(b")* $\|\mathbf{B}^{(n)}\| = O(\sqrt{n})$ and we are therefore just ruling out trivial combinations

[6] We have

$$\lim_{n\to\infty} \mathsf{E}[o_{q.m.}(1)^2] = 0,$$

and

$$\lim_{n\to\infty} n\mathsf{E}\left[O_{q.m.}\left(\frac{1}{\sqrt{n}}\right)^2\right] < \infty.$$

[7] Weak factors can be considered where divergence is more slowly than n, results for this case are in Onatski (2012).

[8] Sometime the weights are defined as $\mathbf{A}^{(n)*} = \sqrt{n}\mathbf{A}^{(n)}$.

in which $\mathbf{A}^{(n)}$ has too many zeros.

We now show that when $n \to \infty$, any weights matrix satisfying (22) and (23) generates aggregates that span the same space as $\mathbf{f}$. Indeed, from the definition and the factor model structure

$$\ddot{\mathbf{w}}_t^{(n)} = \frac{\mathbf{A}^{(n)'}\mathbf{x}_t^{(n)}}{\sqrt{n}} = \frac{\mathbf{A}^{(n)'}\mathbf{B}^{(n)}\mathbf{f}_t}{\sqrt{n}} + \frac{\mathbf{A}^{(n)'}\boldsymbol{\xi}_t^{(n)}}{\sqrt{n}}.$$

The first term converges to $\mathbf{H}\mathbf{f}_t$ by assumption (23). As for the second term notice that from the definition of eigenvalues, for any column $\mathbf{a}_j^{(n)}$ of $\mathbf{A}^{(n)}$

$$\mathsf{Var}\left(\frac{\mathbf{a}_j^{(n)'}\boldsymbol{\xi}_t^{(n)}}{\sqrt{n}}\right) = \frac{1}{n}\mathbf{a}_j^{(n)'}\boldsymbol{\Gamma}_{\boldsymbol{\xi}}^{(n)}\mathbf{a}_j^{(n)} \leq \frac{\lambda_{1\boldsymbol{\xi}}}{n} \to 0.$$

This result holds for one column of $\mathbf{A}^{(n)}$ and if we consider all columns jointly we have

$$\mathsf{Var}\left(\left\|\frac{\mathbf{A}^{(n)'}\boldsymbol{\xi}_t^{(n)}}{\sqrt{n}}\right\|\right) = \mathsf{E}\left[\left\|\frac{\mathbf{A}^{(n)'}\boldsymbol{\xi}_t^{(n)}}{\sqrt{n}}\right\|^2\right] \leq \frac{q\lambda_{1\boldsymbol{\xi}}}{n} \to 0. \tag{24}$$

Hence we have convergence in quadratic mean

$$\|\ddot{\mathbf{w}}_t^{(n)} - \mathbf{H}\mathbf{f}_t\| = O_{q.m.}\left(\frac{1}{\sqrt{n}}\right). \tag{25}$$

Moreover, by Chebychev's inequality we have also convergence in probability[9]

$$\|\mathbf{w}_t^{(n)} - \mathbf{H}\mathbf{f}_t\| = O_p\left(\frac{1}{\sqrt{n}}\right),$$

which means that for any $\epsilon > 0$ there exists an $M > 0$ such that

$$\mathrm{P}(\sqrt{n}\|\mathbf{w}_t^{(n)} - \mathbf{H}\mathbf{f}_t\| > M) < \epsilon.$$

Summing up, as $n \to \infty$ the q weighted cross-sectional averages $\ddot{\mathbf{w}}_t^{(n)}$ are spanning the same space of the space generated by the factors with probability one and also their variance converges to the variance of a linear combination $\mathbf{H}\mathbf{f}_t$ of the factors. In other words, in quadratic mean and in probability only the common component (driven by the factors) survives aggregation. It is important to notice that this limiting notion is related to variances of averages thus involving covariances and not to the variance of a single series. Hence, the factors do not necessarily represent the component with highest variance of a series.

Note that consistency of the factors estimates requires $n \to \infty$ and not T which is the sample size. This fact is the result of two main properties of the factor model:

1. for given value of the loadings we can always consider the factors as parameters to be estimated, then it is clear that we have a parameter space which increases as the sample T size increases, for this reason we need a large n in order to be able to still estimate consistently the factors;
2. the factors are recovered as weighted averages so the "sample size" is the number of units we use to build such averages, thus is n and not T.

The same comments apply to all other factor estimators discussed in Section 7): if n is fixed approximate factor models cannot be estimated consistently. In particular, we will see that although we can still estimate

[9]The Chebychev's inequality states that for any $\epsilon > 0$

$$\text{Prob}\left(|Y - \mathsf{E}[Y]|^2 \geq \epsilon\right) \leq \frac{\mathsf{Var}(Y)}{\epsilon}.$$

the loadings consistently (with Maximum Likelihood but not with PCA), but we cannot estimate consistently the factors.

CHOOSING THE OPTIMAL WEIGHTS. The weights $\mathbf{A}^{(n)}$ used to build $\ddot{\mathbf{w}}_t^{(n)}$ can be generic, as long as they satisfy (22) and (23). For example the cross-sectional average with all weights equal to $1/\sqrt{n}$ would satisfy this condition in the case $q = 1$. However, it is more desirable a data driven choice of $\mathbf{A}^{(n)}$.

The optimal weights must be such that when applied to the common component produce a weighted average with maximum variance. The right choice is given by PCA that is by choosing the weights matrix as the matrix of eigenvectors of $\mathbf{\Gamma}_{\boldsymbol{\chi}}^{(n)}$. Consider the spectral decomposition

$$\mathbf{\Gamma}_{\boldsymbol{\chi}}^{(n)} = \mathbf{P}^{(n)}\mathbf{D}^{(n)}\mathbf{P}^{(n)'}$$

where $\mathbf{P}^{(n)}$ is the $n \times q$ matrix of normalised eigenvectors and $\mathbf{D}^{(n)}$ is the $q \times q$ diagonal matrix of eigenvalues. We choose $\mathbf{A}^{(n)} = \mathbf{P}^{(n)}$. This is of course not feasible in practice since $\boldsymbol{\chi}^{(n)}$ is not observed but let's see what are the implications of this choice and we will then look for a feasible way to proceed. Without loss of generality we assume $\mathbf{\Gamma}_{\mathbf{f}} = \mathbf{I}_q$.

1. We trivially have the normalisation condition (22) and for (23) we have

$$\frac{\mathbf{A}^{(n)'}\mathbf{B}^{(n)}}{\sqrt{n}} = \frac{\mathbf{P}^{(n)'}\mathbf{B}^{(n)}}{\sqrt{n}} =: \mathbf{K}_n \tag{26}$$

   Now the matrix $\mathbf{K}_n$ is positive definite and has all eigenvalues which are finite. Indeed, since assumption *(b")* implies that the eigenvalues of $\mathbf{\Gamma}_{\boldsymbol{\chi}}^{(n)}$ diverge linearly in n, there exists a positive definite matrix $\mathbf{L}$ such that

$$\mathbf{K}_n\mathbf{K}_n' = \frac{\mathbf{P}^{(n)'}\mathbf{\Gamma}_{\boldsymbol{\chi}}^{(n)}\mathbf{P}^{(n)}}{n} = \frac{\mathbf{D}^{(n)}}{n} \to \mathbf{L}, \quad \text{as} \quad n \to \infty, \tag{27}$$

   and by setting $\mathbf{H} = \mathbf{L}^{1/2}$ we have shown that (23) holds for this choice of weights. Without loss of generality we can say that $\mathbf{K}_n \to \mathbf{H}$, as $n \to \infty$.

2. By definition of eigenvectors we have weighted averages of the common component with maximum variance given by the eigenvalues

$$\mathsf{Var}\left(\frac{\mathbf{P}^{(n)'}\boldsymbol{\chi}_t^{(n)}}{\sqrt{n}}\right) = \frac{1}{n}\mathbf{P}^{(n)'}\mathbf{\Gamma}_{\boldsymbol{\chi}}^{(n)}\mathbf{P}^{(n)'} = \frac{\mathbf{D}^{(n)}}{n}. \tag{28}$$

   By assumption *(b")* and linear divergence of eigenvalues, the last expression is always finite.

3. By construction we have

$$\frac{\mathbf{P}^{(n)'}\boldsymbol{\chi}_t^{(n)}}{\sqrt{n}} = \frac{\mathbf{P}^{(n)'}\mathbf{B}^{(n)}\mathbf{f}_t}{\sqrt{n}} = \mathbf{K}_n\mathbf{f}_t, \tag{29}$$

   the principal components of $\boldsymbol{\chi}_t^{(n)}$ span the same space as the factors. Notice that $\mathbf{K}_n$ is finite and this result holds for any n.

4. If we let $n \to \infty$ we have seen above that under (22) and (23) the variance of a weighted average of $\mathbf{x}_t^{(n)}$ converges to the variance of a weighted average of $\boldsymbol{\chi}_t^{(n)}$. By (24) and (28), we have

$$\mathsf{Var}\left(\frac{\mathbf{P}^{(n)'}\mathbf{x}_t^{(n)}}{\sqrt{n}}\right) = \mathsf{Var}\left(\frac{\mathbf{P}^{(n)'}\boldsymbol{\chi}_t^{(n)}}{\sqrt{n}}\right) + \mathsf{Var}\left(\frac{\mathbf{P}^{(n)'}\boldsymbol{\xi}_t^{(n)}}{\sqrt{n}}\right) = \frac{\mathbf{D}^{(n)}}{n} + O\left(\frac{1}{n}\right).$$

   As $n \to \infty$, the first term is always finite and positive definite while the second vanishes. Equivalently, we can write

$$\frac{\mathbf{P}^{(n)'}\mathbf{x}_t^{(n)}}{\sqrt{n}} = \frac{\mathbf{P}^{(n)'}\boldsymbol{\chi}_t^{(n)}}{\sqrt{n}} + O_{q.m.}\left(\frac{1}{\sqrt{n}}\right). \tag{30}$$

By combining (25) with (29) and (30) we have

$$\left\| \frac{\mathbf{P}^{(n)'}\mathbf{x}_t^{(n)}}{\sqrt{n}} - \mathbf{H}\mathbf{f}_t \right\| = O_{q.m.}\left(\frac{1}{\sqrt{n}}\right). \tag{31}$$

The last result shows that if we knew $\mathbf{P}^{(n)}$ we could recover the space spanned by the factors, but this is clearly impossible. We consider instead the $n \times q$ matrix of eigenvectors $\mathbf{P}_{\mathbf{x}}^{(n)}$ corresponding to the q largest eigenvalues of $\mathbf{\Gamma}_{\mathbf{x}}^{(n)}$. This matrix can be computed from data. The Davis-Kahan theorem and assumptions *(a')* and *(b')* imply

$$\|\mathbf{P}_{\mathbf{x}}^{(n)} - \mathbf{P}^{(n)}\mathbf{O}\| \leq \frac{c\|\mathbf{\Gamma}_{\boldsymbol{\xi}}^{(n)}\|}{\lambda_{q\boldsymbol{\chi}}} = \frac{c\lambda_{1\boldsymbol{\xi}}}{\lambda_{q\boldsymbol{\chi}}} \leq \frac{cM}{\lambda_{q\boldsymbol{\chi}}} \to 0, \text{ as } n \to \infty, \tag{32}$$

for some orthogonal matrix $\mathbf{O}$ and finite constant $c > 0$. Moreover under *(b")* convergence is with rate n^{-1}. Therefore, because of (32), as $n \to \infty$ all previous conclusions still hold when we use the weights $\mathbf{A}^{(n)} = \mathbf{P}_{\mathbf{x}}^{(n)}$. In particular, (31) becomes

$$\left\| \frac{\mathbf{P}_{\mathbf{x}}^{(n)'}\mathbf{x}_t^{(n)}}{\sqrt{n}} - \mathbf{O}'\mathbf{H}\mathbf{f}_t \right\| = O_{q.m.}\left(\frac{1}{\sqrt{n}}\right). \tag{33}$$

Thus, (29) and (33) show that PCA of $\boldsymbol{\chi}_t^{(n)}$ and of $\mathbf{x}_t^{(n)}$ are asymptotically equivalent. As $n \to \infty$ the two approaches of PCA and factor analysis are reconciled: this is the blessing of dimensionality.

RECONSTRUCTING THE LOADINGS' SPACE. Since principal components are orthogonal they are also a linear basis of the space of the factors. Moreover, projecting on any linear basis of factors' space allows us to recover the loadings or a linear transformation of them. Since projecting on any basis of the same space is equivalent up to a linear transformation, this means that up to terms $o_{q.m.}(1)$ we can recover the loadings by any of the following equivalent projections:

1. of $\mathbf{x}_t^{(n)}$ onto the space spanned by its q first normalised principal components, $\frac{\mathbf{P}_{\mathbf{x}}^{(n)'}\mathbf{x}_t^{(n)}}{\sqrt{n}}$;
2. of $\mathbf{x}_t^{(n)}$ onto the space spanned by $\boldsymbol{\chi}^{(n)}$'s non-zero principal components, $\frac{\mathbf{P}^{(n)'}\boldsymbol{\chi}_t^{(n)}}{\sqrt{n}}$;
3. of $\mathbf{x}_t^{(n)}$ onto the space spanned by the q-factors, $\mathbf{f}_t$.

In particular, consider projection 1, which is the only feasible one, we can show that

$$\begin{aligned}
&\left(\mathsf{E}\left[\frac{\mathbf{P}_{\mathbf{x}}^{(n)'}\mathbf{x}_t^{(n)}\mathbf{x}_t^{(n)'}\mathbf{P}_{\mathbf{x}}^{(n)}}{n}\right]\right)^{-1}\left(\mathsf{E}\left[\frac{\mathbf{P}_{\mathbf{x}}^{(n)'}\mathbf{x}_t^{(n)}\mathbf{x}_t^{(n)'}}{\sqrt{n}}\right]\right) \\
&= \left(\frac{\mathbf{D}_{\mathbf{x}}^{(n)}}{n}\right)^{-1}\left(\frac{\mathbf{P}_{\mathbf{x}}^{(n)'}\mathbf{\Gamma}_{\boldsymbol{\chi}}^{(n)}}{\sqrt{n}} + \frac{\mathbf{P}_{\mathbf{x}}^{(n)'}\mathbf{\Gamma}_{\boldsymbol{\xi}}^{(n)}}{\sqrt{n}}\right) \\
&= \left(\frac{\mathbf{D}^{(n)}}{n}\right)^{-1}\left(\frac{\mathbf{P}^{(n)'}\mathbf{\Gamma}_{\boldsymbol{\chi}}^{(n)}}{\sqrt{n}} + \frac{\mathbf{P}^{(n)'}\mathbf{\Gamma}_{\boldsymbol{\xi}}^{(n)}}{\sqrt{n}}\right) + O\left(\frac{1}{n}\right) \\
&= \left(\frac{\mathbf{D}^{(n)}}{n}\right)^{-1}\left(\frac{\mathbf{P}^{(n)'}\mathbf{\Gamma}_{\boldsymbol{\chi}}^{(n)}}{\sqrt{n}}\right) + O\left(\frac{1}{\sqrt{n}}\right) \\
&= \left(\frac{\mathbf{D}^{(n)}}{n}\right)^{-1}\left(\frac{\mathbf{D}^{(n)}\mathbf{P}^{(n)}}{\sqrt{n}}\right) + O\left(\frac{1}{\sqrt{n}}\right) \\
&= \sqrt{n}\mathbf{P}^{(n)} + O\left(\frac{1}{\sqrt{n}}\right),
\end{aligned} \tag{34}$$

where $\mathbf{D}_{\mathbf{x}}^{(n)}$ is the $q \times q$ diagonal matrix of eigenvalues of $\mathbf{\Gamma}_{\mathbf{x}}^{(n)}$ and we used assumption *(b")*, (24), (32), and

$$\frac{1}{n}\|\mathbf{D}_{\mathbf{x}}^{(n)} - \mathbf{D}^{(n)}\| \leq \frac{1}{n}\|\mathbf{\Gamma}_{\mathbf{x}}^{(n)} - \mathbf{\Gamma}_{\boldsymbol{\chi}}\| = \frac{1}{n}\|\mathbf{\Gamma}_{\boldsymbol{\xi}}^{(n)}\| = \frac{\lambda_{1\boldsymbol{\xi}}^{(n)}}{n} \leq \frac{M}{n} \tag{35}$$

which follows from Weyl's inequality and assumption *(a')*. Therefore, from (34) we see that we can take the loadings to be proportional to the eigenvectors.

ESTIMATION. Now consider a time series setting and we write the approximate static factor model as

$$\mathbf{x}_t^{(n)} = \mathbf{B}^{(n)}\mathbf{f}_t + \boldsymbol{\xi}_t^{(n)}, \qquad t \in \mathbb{Z}.$$

We can use PCA to estimate the factors, the loadings and the common component. Let us first assume that the number of factors q is given. When we observe only T realisations of $\mathbf{x}^{(n)}$, i.e. we observe $(\mathbf{x}_1^{(n)}, \ldots, \mathbf{x}_T^{(n)})$, the covariance matrix has to be estimated. We use a suffix T to indicate estimated quantities. We start with the sample covariance

$$\mathbf{\Gamma}_{\mathbf{x}}^{(nT)} = \frac{1}{T}\sum_{t=1}^{T}\mathbf{x}_t^{(n)}\mathbf{x}_t^{(n)'}.$$

When $T \to \infty$, it is known that

$$\frac{1}{n}\|\mathbf{\Gamma}_{\mathbf{x}}^{(nT)} - \mathbf{\Gamma}_{\mathbf{x}}^{(n)}\| = O_{q.m.}\left(\frac{1}{\sqrt{T}}\right) \tag{36}$$

as long as fourth order cumulants of $\mathbf{x}_t^{(n)}$ exist. Then, following (34), we collect the normalised eigenvectors corresponding to the q largest eigenvalues of $\mathbf{\Gamma}_{\mathbf{x}}^{(nT)}$ in the $n \times q$ matrix $\mathbf{P}^{(nT)}$ and we define the loadings estimator as

$$\mathbf{B}^{(nT)} = \sqrt{n}\mathbf{P}^{(nT)},$$

then by definition $\frac{\mathbf{B}^{(nT)'}\mathbf{B}^{(nT)}}{n} = \mathbf{I}_q$, for any n, which satisfies assumption *(b")*. Clearly, an estimator of the factors is estimated by projecting the data onto the space spanned by the estimated loadings, which, given the chosen loadings, in this case is equivalent to projecting onto the space of the q largest principal components of $\mathbf{x}_t^{(n)}$,

$$\mathbf{f}_t^{(nT)} = \left(\mathbf{B}^{(nT)'}\mathbf{B}^{(nT)}\right)^{-1}\mathbf{B}^{(nT)'}\mathbf{x}_t^{(n)} = \frac{\mathbf{B}^{(nT)'}\mathbf{x}_t^{(n)}}{n} = \frac{\mathbf{P}^{(nT)'}\mathbf{x}_t^{(n)}}{\sqrt{n}}. \tag{37}$$

Notice that the last expression is the sample analogue of (29) and (30).

The estimator depends on n and T and if we assume that the common component has asymptotically distinct eigenvalues then we can show that it is consistent. In particular, there are two possible equivalent identifying constraints that we can choose:

1. Approach by Stock and Watson (2002a), common in econometrics:
   (a) $\mathbf{\Gamma}_{\mathbf{f}}$ is positive definite with distinct diagonal entries;
   (b) $\frac{\mathbf{B}^{(n)}\mathbf{B}^{(n)'}}{n} \to \mathbf{I}_q$ as $n \to \infty$;
   (c) $\mathbf{f}_t = \frac{\mathbf{P}^{(n)'}\boldsymbol{\chi}_t^{(n)}}{\sqrt{n}}$ (non-normalised principal components of $\boldsymbol{\chi}_t^{(n)}$).

   Then $\mathbf{\Gamma}_{\mathbf{f}} = \frac{\mathbf{D}^{(n)}}{n}$ which must be positive definite for any n as requested by (a), moreover this matrix is diagonal and (a) implies that $\mathbf{\Gamma}_{\mathbf{f}}$ has distinct eigenvalues. Moreover, we have

   $$\mathbf{f}_t = \frac{\mathbf{P}^{(n)'}\boldsymbol{\chi}_t^{(n)}}{\sqrt{n}} = \frac{\mathbf{P}^{(n)'}\mathbf{P}^{(n)}\mathbf{P}^{(n)'}\boldsymbol{\chi}_t^{(n)}}{\sqrt{n}} = \frac{\mathbf{P}^{(n)'}\mathbf{P}^{(n)}\mathbf{P}^{(n)'}\mathbf{B}^{(n)}\mathbf{f}_t}{\sqrt{n}},$$

   which implies $\mathbf{B}^{(n)} = \sqrt{n}\mathbf{P}^{(n)}$ and thus (b) holds for any n (this is because the definition of $\mathbf{f}_t$ depends on n so as n changes the factors change). This is the approach followed here.

2. Approach by Fan, Liao, and Mincheva (2013), common in statistics:

   (a) $\mathbf{\Gamma_f} = \mathbf{I}_q$;

   (b) $\frac{\mathbf{B}^{(n)}\mathbf{B}^{(n)'}}{n} \to \mathbf{H}$ as $n \to \infty$, with $\mathbf{H}$ positive definite with distinct eigenvalues;

   (c) $\mathbf{f}_t = (\mathbf{D}^{(n)})^{-1/2}\mathbf{P}^{(n)'}\boldsymbol{\chi}_t^{(n)}$ (normalised principal components of $\boldsymbol{\chi}_t^{(n)}$).

   Then $\mathbf{\Gamma_f} = \mathbf{I}_q$ as requested by (a). Moreover, we have

$$\mathbf{f}_t = (\mathbf{D}^{(n)})^{-1/2}\mathbf{P}^{(n)'}\boldsymbol{\chi}_t^{(n)} = (\mathbf{D}^{(n)})^{-1/2}\mathbf{P}^{(n)'}\mathbf{P}^{(n)}\mathbf{P}^{(n)'}\boldsymbol{\chi}_t^{(n)} = (\mathbf{D}^{(n)})^{-1/2}\mathbf{P}^{(n)'}\mathbf{P}^{(n)}\mathbf{P}^{(n)'}\mathbf{B}^{(n)}\mathbf{f}_t,$$

   which implies $\mathbf{B}^{(n)} = \mathbf{P}^{(n)}(\mathbf{D}^{(n)})^{1/2}$ and thus (b) holds only asymptotically as $n \to \infty$.

Note that both 1(a)-1(b) or 2(a)-2(b) imply that the eigenvalues of $\mathbf{\Gamma}_{\boldsymbol{\chi}}$ diverge linearly in n (this is already shown above since these are special versions of *(b")*) and moreover the eigenvalues are asymptotically distinct. Regardless of which set of identification constraints is chosen, for large n, T, using (36) and a result similar to (32) for estimated quantities, we can show that the eigenvectors $\mathbf{P}_{\mathbf{x}}^{(n)}$ estimate $\mathbf{P}^{(n)}$ up to a sign. This is true regardless of the rate of divergence of n and T as long as we consider just the eigenvectors corresponding to the largest (diverging) eigenvalues. Then, for large n, the result in (33) applies. Therefore, as $n, T \to \infty$, there exists a diagonal matrix $\mathbf{J}$ with entries ± 1 such that

$$\|\mathbf{f}_t^{(nT)} - \mathbf{J}\mathbf{f}_t\| = O_{q.m.}\left(\max\left(\frac{1}{\sqrt{n}}, \frac{1}{\sqrt{T}}\right)\right), \qquad t = 1, \ldots, T,$$

and it is also possible to prove

$$\|\mathbf{b}_i^{(nT)'} - \mathbf{b}_i^{(n)'}\mathbf{J}\| = O_{q.m.}\left(\max\left(\frac{1}{\sqrt{n}}, \frac{1}{\sqrt{T}}\right)\right), \qquad i = 1, \ldots, n,$$

where $\mathbf{b}_i^{(n)'}$ is the i-th row of $\mathbf{B}^{(n)}$ thus of dimension $1 \times q$. These results are proved for example in Forni, Giannone, Lippi, and Reichlin (2009).

Some care has to be taken when $n > T$ as in this case consistency of eigenvectors in general does not hold. However, it can be proved that when the largest eigenvalues are ultra-spiked, in the sense that they diverge as $n \to \infty$, as in this case, then the corresponding eigenvectors are still consistently estimated using the eigenvectors of the sample covariance (Fan, Liao, and Mincheva, 2013). Therefore, PCA delivers always consistent estimates of the loadings, thus the previous results still hold. Alternatively one can consider the $T \times T$ covariance as in Bai and Ng (2002) but this requires more restrictions on the serial dependence of the idiosyncratic components.

The estimated common component is defined as

$$\chi_{it}^{(nT)} = \mathbf{b}_i^{(nT)}\mathbf{f}_t^{(nT)} = \sqrt{n}\mathbf{p}_i^{(nT)'}\frac{\mathbf{P}^{(nT)'}\mathbf{x}_t^{(n)}}{\sqrt{n}} = \mathbf{p}_i^{(nT)'}\mathbf{P}^{(nT)'}\mathbf{x}_t^{(n)}, \tag{38}$$

where $\mathbf{p}_i^{(nT)'}$ is a $1 \times r$ row vector corresponding to the i-th row of $\mathbf{P}^{(nT)}$. It follows that the common component is identified and estimated consistently, as $n, T \to \infty$:

$$|\chi_{it}^{(nT)} - \chi_{it}| = O_{q.m.}\left(\max\left(\frac{1}{\sqrt{n}}, \frac{1}{\sqrt{T}}\right)\right).$$

Convergence in probability follows straightforwardly.

FACTORS AS LINEAR PROJECTIONS. To conclude, notice that PCA estimators of loadings and factors can be obtained as a solution of the problem

$$\min \frac{1}{nT}\sum_{i=1}^{n}\sum_{t=1}^{T}(x_{it} - \mathbf{b}_i'\mathbf{f}_t)^2 = \min \frac{1}{nT}\|\mathbf{x}_t^{(n)} - \mathbf{B}^{(n)}\mathbf{f}_t\|_F^2 = \min \frac{1}{nT}\|\boldsymbol{\xi}_t^{(n)}\|_F^2, \tag{39}$$

subject to one of the sets of identifying constraints 1 or 2 stated above. Indeed, given the true loadings $\mathbf{B}^{(n)}$, we obtain the factors estimator by linear projection onto the loadings (see also (37))

$$\mathbf{f}_t^{(n)*} = \left(\frac{\mathbf{B}^{(n)'}\mathbf{B}^{(n)}}{n}\right)^{-1} \frac{\mathbf{B}^{(n)'}\mathbf{x}_t^{(n)}}{n}, \tag{40}$$

while, given the true factors $\mathbf{f}_t$, we obtain the loadings estimator by linear projection onto the factors (recall that $\mathsf{E}[\boldsymbol{\xi}_t^{(n)}\mathbf{f}_t'] = \mathbf{0}$ by assumption)

$$\mathbf{B}^{(n)*} = \left(\mathsf{E}[\mathbf{x}_t^{(n)}\mathbf{f}_t']\right)\left(\mathsf{E}[\mathbf{f}_t\mathbf{f}_t']\right)^{-1} = \left(\mathsf{E}[\boldsymbol{\chi}_t^{(n)}\mathbf{f}_t']\right)\left(\mathsf{E}[\mathbf{f}_t\mathbf{f}_t']\right)^{-1}. \tag{41}$$

Now the minimisation in (39) means finding both loadings and factors such that the trace of the covariance of the idiosyncratic component is minimum, i.e. the one of the common component is maximum, which is solved by PCA, which implies that the optimal factors can be built by taking a weighted average of $\mathbf{x}^{(n)}$. We see from (40) that a factors estimator is a weighted average of the data, then under the set of identification constraints 1, the optimal estimated weights are given by the normalised eigenvectors of the sample covariance, i.e. we have to choose estimated weights $\mathbf{B}^{(nT)}$ such that $\mathbf{B}^{(nT)} = \sqrt{n}\mathbf{P}^{(nT)}$ (see above). If we replace this choice in (40) we get the factors estimator $\mathbf{f}_t^{(nT)} = \frac{\mathbf{P}^{(nT)'}\mathbf{x}_t^{(n)}}{\sqrt{n}}$. Now by replacing in (41) expectations with sums (data is stationary and assuming is also ergodic), we obtain the estimated common component as above:

$$\begin{aligned}
\chi_{it}^{(nT)} &= \mathbf{b}_i^{(nT)'}\mathbf{f}_t^{(nT)} = \left(\sum_{t=1}^{T} x_{it}^{(n)}\mathbf{f}_t^{(nT)'}\right)\left(\sum_{t=1}^{T}\mathbf{f}_t^{(nT)}\mathbf{f}_t^{(nT)'}\right)^{-1}\frac{\mathbf{P}^{(nT)'}\mathbf{x}_t^{(n)}}{\sqrt{n}} \\
&= \mathbf{b}_i^{(nT)'}\left(\sum_{t=1}^{T}\mathbf{f}_t^{(nT)}\mathbf{f}_t^{(nT)'}\right)\left(\sum_{t=1}^{T}\mathbf{f}_t^{(nT)}\mathbf{f}_t^{(nT)'}\right)^{-1}\frac{\mathbf{P}^{(nT)'}\mathbf{x}_t^{(n)}}{\sqrt{n}} \\
&\quad + \left(\frac{1}{T}\sum_{t=1}^{T}\xi_{it}^{(nT)}\mathbf{f}_t^{(nT)'}\right)\left(\frac{1}{T}\sum_{t=1}^{T}\mathbf{f}_t^{(nT)}\mathbf{f}_t^{(nT)'}\right)^{-1}\frac{\mathbf{P}^{(nT)'}\mathbf{x}_t^{(n)}}{\sqrt{n}} \\
&= \sqrt{n}\mathbf{p}_i^{(nT)}\frac{\mathbf{P}^{(nT)'}\mathbf{x}_t^{(n)}}{\sqrt{n}} + O_p\left(\frac{1}{\sqrt{T}}\right)
\end{aligned}$$

which is coincident with the expression found before up to an estimation error, which however could be removed by using the second expression in (41) which depends only on the common component. This is similar to the approach adopted by Bai (2003).

4 The generalised dynamic factor model - Time domain

For the reasons above it is then natural to turn to infinite dimensional panels of times series and to study their aggregation not only cross-sectionally but, given the time series context, also serially. This approach will lead to a more general result, which is the representation theorem showing that under the simple assumption of stationarity every infinite dimensional panel of time series admits a dynamic factor representation.

DYNAMIC AGGREGATION. Since we are in a time series context we consider now aggregation across series and time, therefore, we are interested in studying the dynamic aggregates

$$w_t^{(n)} = \sum_{i=1}^{n} \sum_{k=-\infty}^{\infty} a_{ik} x_{it-k}, \tag{42}$$

such that $\sum_{i=1}^{n} \sum_{k=-\infty}^{\infty} a_{ik}^2 = 1$.

Let us start from a static one-factor model with unit loadings

$$\mathbf{x}_t^{(n)} = f_t + \boldsymbol{\xi}_t^{(n)}, \tag{43}$$

and we assume $\mathsf{E}[f_t\, \xi_{is}] = 0$ for any i and $s, t \in \mathbb{Z}$.

First, consider an exact factor structure with no serial dependence in ξ_{it} and f_t, that is $\mathsf{E}[f_t f_s] = 0$ and $\mathsf{E}[\xi_{it}\xi_{is}] = 0$ for any $s \neq t$. Then, (42) becomes

$$w_t^{(n)} = \sum_{i=1}^{n} \sum_{k=-\infty}^{\infty} a_{ik} f_{t-k} + \sum_{i=1}^{n} \sum_{k=-\infty}^{\infty} a_{ik} \xi_{it-k}, \tag{44}$$

and we have

$$\mathsf{Var}\left(w_t^{(n)}\right) = \sum_{k=-\infty}^{\infty} \left(\sum_{i=1}^{n} a_{ik}\right)^2 \mathsf{Var}(f_t) + \sum_{i=1}^{n} \sum_{k=-\infty}^{\infty} a_{ik}^2 \mathsf{Var}(\xi_{it}).$$

If, for simplicity we also assume $a_{ik} = a_k$ for any i, then due to normalisation we must have $\sum_{k=-\infty}^{\infty} a_k^2 = 1/n$, and

$$\mathsf{Var}\left(w_t^{(n)}\right) = n\mathsf{Var}(f_t) + \frac{1}{n} \sum_{i=1}^{n} \mathsf{Var}(\xi_{it}).$$

Thus, also under dynamic aggregation, when $n \to \infty$ we separate common and idiosyncratic component and only the former gives a contribute to the total variance of the aggregate.

Now let us consider an approximate factor structure, as assumed in *(a')*, and we allow for autocorrelation in f_t and $\boldsymbol{\xi}_t^{(n)}$, provided they are both covariance stationary. Let us assume a dynamic aggregation as (44) with $a_{ik} = a_k$ for any i, so that $\sum_{k=-\infty}^{\infty} a_k^2 = 1/n$. We have

$$\mathsf{Var}\left(w_t^{(n)}\right) \;=\; n^2 \sum_{k,h=-\infty}^{\infty} a_k a_h \mathsf{Cov}(f_{t-k}, f_{t-h}) + \sum_{k,h=-\infty}^{\infty} a_k a_h \sum_{i,j=1}^{n} \mathsf{Cov}(\xi_{it-k}, \xi_{jt-k}).$$

Because of stationarity we can write $\mathsf{Cov}(f_{t-k}, f_{t-h}) = \gamma_f(|k-h|)$ and $\mathsf{Cov}(\xi_{it-k}, \xi_{jt-k}) = \gamma_\xi(ij, |k-h|)$ and all these autocovariances are square summable but let us assume that they are also absolutely summable.

Under the chosen normalisation it is always possible to have $\sum_{k=-\infty}^{\infty} a_k = K/\sqrt{n}$ for some constant

K.[10] Then, setting $K = 1$ for simplicity, (45) becomes

$$\mathsf{Var}\left(w_t^{(n)}\right) = n\mathsf{Var}(f_t) + n \sum_{\substack{k,h=-\infty \\ k\neq h}}^{\infty} \gamma_f(|k-h|) + \frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{n}\mathsf{Cov}(\xi_{it},\xi_{jt}) + \frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{\substack{k,h=-\infty \\ k\neq h}}^{\infty}\gamma_\xi(ij,|k-h|).$$

The first two terms diverge as $n \to \infty$, the third is finite under assumption *(a')* of approximate factor structure while for the last it is reasonable to assume that is also bounded (the necessary assumption is made in Section 5 when considering spectral densities). Thus, also in the most general case under dynamic aggregation when $n \to \infty$ we separate common and idiosyncratic component and only the former gives a contribute to the total variance of the aggregate.

More precisely, if we define $w_{tf}^{(n)} = \sum_{i=1}^{n}\sum_{k=-\infty}^{\infty} a_{ik} f_{t-k}$ we have shown that for the standardised dynamic aggregate we have

$$\dot{w}_t^{(n)} := \frac{w_t^{(n)}}{\sqrt{\mathsf{Var}(w_t^{(n)})}} = \frac{w_{tf}^{(n)}}{\sqrt{\mathsf{Var}(w_{tf}^{(n)})}} + o_{q.m.}(1).$$

REPRESENTATION THEOREMS. We now change approach and show that actually a dynamic factor structure always exists under very general conditions. Hereafter, we consider a double-indexed stochastic process $\mathbf{x} = \{x_{it} | i \in \mathbb{N},\ t \in \mathbb{Z}\}$, of which $\mathbf{x}^{(n)}$ is a subprocess of dimension n and $\mathbf{x}_t^{(n)}$ is a realization. We make two assumptions.

ASSUMPTION (A1). *The process* $\mathbf{x}$ *is second-order stationary, i.e. for any* $i, j \in \mathbb{N}$, $t \in \mathbb{Z}$, *and* $k \in \mathbb{Z}$, $\mathsf{Cov}(x_{it}, x_{jt-k})$ *exist, are finite and do not depend on* t.

ASSUMPTION (A2). *For any* $i \in \mathbb{N}$ *and* $t \in \mathbb{Z}$, $\mathsf{E}[x_{it}] = 0$ *and* $\mathsf{Var}(x_{it}) > 0$.

We denote the Hilbert space spanned by $\mathbf{x}$ as $\mathcal{H}^{\mathbf{x}} := L^2(\Omega, \mathcal{F}, P)$ for some probability space $(\Omega, \mathcal{F}, P)$. Then $\mathcal{H}^{\mathbf{x}}$ is equipped with the covariance scalar product

$$(\mathbf{x}, \mathbf{y}) = \int_\Omega \mathbf{x}\,\mathbf{y}'\,\mathrm{d}F_{\mathbf{x},\mathbf{y}}.$$

We also denote the space $\mathcal{H}_t^{\mathbf{x}} \subset \mathcal{H}^{\mathbf{x}}$ as the space spanned by $\{x_{is} | i \in \mathbb{N},\ s \leq t\}$. Notice that $\mathcal{H}^{\mathbf{x}}$ contains any convergent (in quadratic mean) linear combination of x_{it} thus it contains scalar processes $w = \{w_t | t \in \mathbb{Z}\}$ such that

$$w_t = \sum_{i=1}^{\infty}\sum_{s=-\infty}^{\infty} a_{is} x_{it-s}, \quad t \in \mathbb{Z}.$$

We have the definition of commonness and idiosyncracy.

DEFINITION. *A random variable* $\chi \in \mathcal{H}^{\mathbf{x}}$ *satisfying Assumptions A1 and A2 is common if*

$$\frac{w_t^{(n)}}{\sqrt{\mathsf{Var}(w_t^{(n)})}} = \frac{\chi_t}{\sqrt{\mathsf{Var}(\zeta_t)}} + o_{q.m.}(1),$$

[10] If we choose $a_k^2 = 1/(\alpha n^{(|k|+1)})$ and we define

$$\alpha = \sum_{k=-\infty}^{\infty}\frac{1}{n^{|k|}} = \left(2\sum_{k=0}^{\infty}\frac{1}{n^k} - 1\right) = \left(\frac{2n}{n-1} - 1\right) = n\left(\frac{2n-n+1}{n-1}\right) = \frac{(n+1)}{(n-1)},$$

then $\sum_{k=-\infty}^{\infty} a_k^2 = 1/n$.

for some dynamic aggregation scheme, as for example the one defined in (44)*, and such that* $\lim_{n\to\infty} \mathsf{Var}(w_t^{(n)}) = \infty$*. If we then define the common Hilbert space* $\mathcal{H}_{com}^{\mathbf{x}}$ *as the space spanned by all common variables in* $\mathcal{H}^{\mathbf{x}}$ *then its orthogonal complement defined as* $\mathcal{H}_{idio}^{\mathbf{x}} \equiv \mathcal{H}_{com}^{\mathbf{x}\perp}$ *is the idiosyncratic space.*

Given these definitions we have:

THEOREM 1 (Hallin and Lippi (2013)). *Under Assumptions A1 and A2 there exist two uniquely defined mutually orthogonal processes* $\boldsymbol{\chi} = \{\chi_{it}\} \in \mathcal{H}_{com}^{\mathbf{x}}$ *and* $\boldsymbol{\xi} = \{\xi_{it}\} \in \mathcal{H}_{idio}^{\mathbf{x}}$ *such that*

$$x_{it} = \chi_{it} + \xi_{it}, \quad i \in \mathbb{N},\ t \in \mathbb{Z}.$$

*and we call this representation Generalized Dynamic Factor Model (*GDFM*) for* $\mathbf{x}$*.*

The reason is that since the space $\mathcal{H}^{\mathbf{x}}$ can always be decomposed into $\mathcal{H}_{com}^{\mathbf{x}}$ and $\mathcal{H}_{idio}^{\mathbf{x}}$ (which are orthogonal) then χ_{it} and ξ_{it} are uniquely obtained by projecting x_{it} onto these subspaces. This is a very general statement and it is of difficult application. We therefore add another assumption:

ASSUMPTION (A3). *Assume that there exists a* q*-dimensional process* $\mathbf{y} = \{y_{jt} | j = 1, , \ldots, q,\ t \in \mathbb{Z}\} \in \mathcal{H}^{\mathbf{x}}$ *with innovation process* $\mathbf{u}$ *having full-rank* q *and such that* $\mathcal{H}_t^{\mathbf{u}} = \mathcal{H}_t^{\mathbf{y}} = \mathcal{H}_t^{\boldsymbol{\chi}}$ *for any* $t \in \mathbb{Z}$*.*

We say that $\boldsymbol{\chi}$ has dynamic rank q. We then have:

THEOREM 2 (Hallin and Lippi (2013)). *Under Assumptions A1, A2, and A3, there exist*

1. *a* q*-dimensional process* $\mathbf{u} = \{u_{jt} | j = 1, , \ldots, q,\ t \in \mathbb{Z}\}$ *of mutually orthogonal white noises, i.e. such that* $\mathsf{Cov}(u_{jt}, u_{ks}) = 0$ *for any* $j, k = 1, \ldots, q$*,* $t, s \in \mathbb{Z}$ *and* $j \neq k$ *and* $t \neq s$*;*
2. *a collection of one-sided square-summable filters* $b_{ij}(L)$*,* $i \in \mathbb{N}$*,* $j = 1, \ldots, q$*;*

such that $\mathcal{H}_t^{\mathbf{u}} = \mathcal{H}_t^{\boldsymbol{\chi}}$ *and*

$$x_{it} = \chi_{it} + \xi_{it}, \quad i \in \mathbb{N},\ t \in \mathbb{Z} \tag{45}$$

$$\chi_{it} = \sum_{j=1}^{q} b_{ij}(L) u_{jt}. \tag{46}$$

In matrix notation we can write

$$\mathbf{x}_t^{(n)} = \mathbf{B}^{(n)}(L)\mathbf{u}_t + \boldsymbol{\xi}_t^{(n)}, \quad t \in \mathbb{Z}. \tag{47}$$

5 The generalised dynamic factor model - Frequency domain

We have seen in example (43) that the common component can be recovered when considering dynamic linear combinations of $\mathbf{x}$ as $n \to \infty$. Intuitively, as in traditional PCA, among all possible dynamic aggregations we are interested in those having maximum variance. With respect to the i.i.d. case, we need to take into consideration the autocorrelation present in $\mathbf{x}$ thus we have to generalize PCA to dynamic PCA. This will give us other representation results and a tool for estimation of (47). First, we quickly review the basics of spectral analysis.

SPECTRAL DENSITY. We make an additional assumption.

ASSUMPTION (A4). *For all $n \in \mathbb{N}$, the spectral measure of $\mathbf{x}^{(n)}$ is absolutely continuous with respect to the Lebesgue measure on $[-\pi, \pi]$, that is, $\mathbf{x}^{(n)}$ has a spectral density matrix $\mathbf{\Sigma}_{\mathbf{x}}^{(n)}(\theta)$ with entries $\sigma_{ij}(\theta)$, for $i, j = 1, \dots, n$.*

Notice that, since the spectral density matrices for different values of n are nested, then their entries do not depend on n. The spectral density matrix of $\mathbf{x}$ is defined as

$$\mathbf{\Sigma}_{\mathbf{x}}^{(n)}(\theta) = \frac{1}{2\pi} \sum_{k=-\infty}^{\infty} e^{-ik\theta} \mathbf{\Gamma}_{\mathbf{x}}^{(n)}(k), \quad \theta \in [-\pi, \pi], \tag{48}$$

where $\mathbf{\Gamma}_{\mathbf{x}}^{(n)}(k) = \mathsf{E}[\mathbf{x}_t \mathbf{x}_{t-k}']$. It is an Hermitian matrix, i.e. $\mathbf{\Sigma}_{\mathbf{x}}^{(n)}(\theta) = \mathbf{\Sigma}_{\mathbf{x}}^{(n)\dagger}(\theta)$, where we indicate $\mathbf{\Sigma}_{\mathbf{x}}^{(n)\dagger}(\theta) = \bar{\mathbf{\Sigma}}_{\mathbf{x}}^{(n)'}(\theta)$, i.e. the transposed and complex conjugate matrix. The sum in (48) exists because of Assumption A4, which means that for any i, j, $\sum_{k=-\infty}^{\infty} \left(\mathbf{\Gamma}_{\mathbf{x}}^{(n)}(k)\right)_{ij}^2 < \infty$, a reasonable assumption for second order stationary processes.

By using the inverse Fourier transform of (48) we have

$$\mathbf{\Gamma}_{\mathbf{x}}^{(n)}(k) = \int_{-\pi}^{\pi} \mathbf{\Sigma}_{\mathbf{x}}^{(n)}(\theta) e^{ik\theta} \mathrm{d}\theta, \quad k \in \mathbb{Z}.$$

Any mapping $\theta \mapsto \mathbf{M}(\theta)$ defined over $[-\pi, \pi]$, which is measurable and has square-integrable elements, i.e. $\mathbf{M} \in L^2([-\pi, \pi])$, admits a Fourier expansion

$$\mathbf{M}(\theta) = \frac{1}{2\pi} \sum_{k=-\infty}^{\infty} \mathbf{M}(k) e^{-ik\theta}, \tag{49}$$

where the series converges in quadratic-mean and the coefficients are defined as

$$\mathbf{M}(k) = \int_{-\pi}^{\pi} \mathbf{M}(\theta) e^{ik\theta} \mathrm{d}\theta. \tag{50}$$

This relation defines a filter with the same coefficients

$$\underline{\mathbf{M}}(L) = \frac{1}{2\pi} \sum_{k=-\infty}^{\infty} \mathbf{M}(k) L^k, \tag{51}$$

i.e. such that $\underline{\mathbf{M}}(e^{-i\theta}) = \mathbf{M}(\theta)$. Therefore, if $\mathbf{x}^{(n)}$ has spectral density $\mathbf{\Sigma}_{\mathbf{x}}^{(n)}(\theta)$ then $\mathbf{y}^{(n)} = \{\underline{\mathbf{M}}(L)\mathbf{x}_t^{(n)}\}$ has spectral density

$$\mathbf{\Sigma}_{\mathbf{y}}^{(n)}(\theta) = \mathbf{M}(\theta) \mathbf{\Sigma}_{\mathbf{x}}^{(n)}(\theta) \mathbf{M}^{\dagger}(\theta) = \underline{\mathbf{M}}(e^{-i\theta}) \mathbf{\Sigma}_{\mathbf{x}}^{(n)}(\theta) \underline{\mathbf{M}}'(e^{i\theta}). \tag{52}$$

Consider the example

$$\mathbf{y}_t = \mathbf{v}_t + \mathbf{M}\mathbf{v}_{t-1}, \quad \mathbf{v}_t \sim w.n.(\mathbf{0}, \mathbf{\Gamma}_{\mathbf{v}}(0)).$$

If we want to compute the spectral density of $\mathbf{y}$ we can proceed in different ways.

1. We can compute

$$\mathbf{\Gamma}_{\mathbf{y}}(0) = \mathbf{\Gamma}_{\mathbf{v}}(0) + \mathbf{M}\mathbf{\Gamma}_{\mathbf{v}}(0)\mathbf{M}', \quad \mathbf{\Gamma}_{\mathbf{y}}(1) = \mathbf{M}\mathbf{\Gamma}_{\mathbf{v}}(0), \text{ and } \mathbf{\Gamma}_{\mathbf{y}}(-1) = \mathbf{\Gamma}_{\mathbf{y}}'(1)$$

   and the spectral density of $\mathbf{y}_t$ is

$$\mathbf{\Sigma}_{\mathbf{y}}(\theta) = \frac{1}{2\pi}\left(\mathbf{\Gamma}_{\mathbf{v}}(0) + \mathbf{M}\mathbf{\Gamma}_{\mathbf{v}}(0)\mathbf{M}' + \mathbf{M}\mathbf{\Gamma}_{\mathbf{v}}(0)e^{-i\theta} + \mathbf{\Gamma}_{\mathbf{v}}(0)\mathbf{M}'e^{i\theta}\right).$$

   By inverting the last relation and by noticing that $\int_{-\pi}^{\pi} e^{ik\theta}\mathrm{d}\theta = 0$ for any $k \in \mathbb{Z} \setminus \{0\}$ we can go back to $\mathbf{\Gamma}_{\mathbf{y}}(0)$ and $\mathbf{\Gamma}_{\mathbf{y}}(1)$.

2. Alternatively, if we define the filter $\underline{\mathbf{M}}(L) = \mathbf{I} + \mathbf{M}L$ then $\mathbf{y}_t = \underline{\mathbf{M}}(L)\mathbf{v}_t$ and, by comparing with (51), the Fourier coefficients are $\mathbf{M}(0) = 2\pi\mathbf{I}$ and $\mathbf{M}(1) = 2\pi\mathbf{M}$. Therefore, from (49)

$$\mathbf{M}(\theta) = \frac{1}{2\pi}\left(\mathbf{M}(0) + \mathbf{M}(1)e^{-i\theta}\right) = \left(\mathbf{I} + \mathbf{M}e^{-i\theta}\right),$$

   and from (52) and since $\mathbf{\Sigma}_{\mathbf{v}}(\theta) = \frac{1}{2\pi}\mathbf{\Gamma}_{\mathbf{v}}(0)$

$$\mathbf{\Sigma}_{\mathbf{y}}(\theta) = \mathbf{M}(\theta)\mathbf{\Sigma}_{\mathbf{v}}(\theta)\mathbf{M}^{\dagger}(\theta) = \frac{1}{2\pi}\left(\mathbf{I} + \mathbf{M}e^{-i\theta}\right)\mathbf{\Gamma}_{\mathbf{v}}(0)\left(\mathbf{I} + \mathbf{M}'e^{i\theta}\right),$$

   which gives the same formula as before.

3. In general, when things are more complex we can use directly the filter $\underline{\mathbf{M}}(L) = \mathbf{I} + \mathbf{M}L$ and from (52) and we have

$$\mathbf{\Sigma}_{\mathbf{y}}(\theta) = \underline{\mathbf{M}}(e^{-i\theta})\mathbf{\Sigma}_{\mathbf{v}}(\theta)\underline{\mathbf{M}}'(e^{i\theta}) = \frac{1}{2\pi}\left(\mathbf{I} + \mathbf{M}e^{-i\theta}\right)\mathbf{\Gamma}_{\mathbf{v}}(0)\left(\mathbf{I} + \mathbf{M}'e^{i\theta}\right).$$

DYNAMIC PRINCIPAL COMPONENTS. We now show how to recover the common component in the generalised dynamic factor models defined in Theorem 2. The spectral density matrix can be factorised as

$$\mathbf{\Sigma}_{\mathbf{x}}^{(n)}(\theta) = \bar{\mathbf{P}}_{\mathbf{x}}^{(n)}(\theta)\mathbf{D}_{\mathbf{x}}^{(n)}(\theta)\mathbf{P}_{\mathbf{x}}^{(n)'}(\theta),$$

where $\mathbf{P}_{\mathbf{x}}^{(n)}(\theta)$ is an $n \times n$ unitary matrix of complex normalized eigenvectors, i.e. $\mathbf{P}_{\mathbf{x}}^{(n)\dagger}(\theta)\mathbf{P}_{\mathbf{x}}^{(n)}(\theta) = \mathbf{P}_{\mathbf{x}}^{(n)}(\theta)\mathbf{P}_{\mathbf{x}}^{(n)\dagger}(\theta) = \mathbf{I}$ and $\mathbf{P}_{\mathbf{x}}^{(n)}(\theta) = (\mathbf{p}_{1\mathbf{x}}^{(n)}(\theta)\dots\mathbf{p}_{n\mathbf{x}}^{(n)}(\theta))$, and $\mathbf{D}_{\mathbf{x}}^{(n)} = \mathrm{diag}(\lambda_{1\mathbf{x}}^{(n)}(\theta) \geq \lambda_{2\mathbf{x}}^{(n)}(\theta) \geq \dots \geq \lambda_{n\mathbf{x}}^{(n)}(\theta) > 0)$ contains the ordered eigenvalues. Since the spectral density is Hermitian, its eigenvalues are real.

Then,

$$\mathbf{\Sigma}_{\mathbf{x}}^{(n)}(\theta)\bar{\mathbf{p}}_{j\mathbf{x}}^{(n)}(\theta) = \lambda_{j\mathbf{x}}^{(n)}(\theta)\bar{\mathbf{p}}_{j\mathbf{x}}^{(n)}(\theta) \text{ and } \mathbf{\Sigma}_{\mathbf{x}}^{(n)}(\theta) = \sum_{j=1}^{n}\lambda_{j\mathbf{x}}^{(n)}(\theta)\bar{\mathbf{p}}_{j\mathbf{x}}^{(n)}(\theta)\mathbf{p}_{j\mathbf{x}}^{(n)'}(\theta). \tag{53}$$

As in traditional PCA, for any $\theta \in [-\pi, \pi]$ eigenvalues and eigenvectors are the solutions of the problem

$$\begin{aligned}
\mathbf{p}_{1\mathbf{x}}^{(n)}(\theta) &= \arg\max_{\mathbf{u}(\theta)} \mathbf{u}^{\dagger}(\theta)\mathbf{\Sigma}_{\mathbf{x}}^{(n)}(\theta)\mathbf{u}(\theta) \text{ s.t. } \mathbf{u}^{\dagger}(\theta)\mathbf{u}(\theta) = 1, \\
\mathbf{p}_{j\mathbf{x}}^{(n)}(\theta) &= \arg\max_{\mathbf{u}(\theta)} \mathbf{u}^{\dagger}(\theta)\mathbf{\Sigma}_{\mathbf{x}}^{(n)}(\theta)\mathbf{u}(\theta) \text{ s.t. } \mathbf{u}^{\dagger}(\theta)\mathbf{u}(\theta) = 1 \text{ and } \mathbf{u}^{\dagger}(\theta)\mathbf{p}_{1\mathbf{x}}^{(n)}(\theta) = \dots = \mathbf{u}^{\dagger}(\theta)\mathbf{p}_{j-1\mathbf{x}}^{(n)}(\theta) = 0,
\end{aligned}$$

the eigenvalues being the value of the functions at their maximum.

The mapping $\theta \mapsto \mathbf{p}_{j\mathbf{x}}^{(n)}(\theta)$ is square-integrable, therefore it admits a Fourier expansion

$$\mathbf{p}_{j\mathbf{x}}^{(n)}(\theta) = \frac{1}{2\pi}\sum_{k=-\infty}^{\infty}\left[\int_{-\pi}^{\pi}\mathbf{p}_{j\mathbf{x}}^{(n)}(\theta)e^{ik\theta}\mathrm{d}\theta\right]e^{-ik\theta}$$

and we have the associated filters

$$\underline{\mathbf{p}}_{j\mathbf{x}}^{(n)}(L) = \frac{1}{2\pi}\sum_{k=-\infty}^{\infty}\left[\int_{-\pi}^{\pi}\mathbf{p}_{j\mathbf{x}}^{(n)}(\theta)e^{ik\theta}\mathrm{d}\theta\right]L^k.$$

Therefore, given the definition of dynamic eigenvectors as those directions that maximize the spectral density, the aggregated univariate process $w_j^{(n)} = \{w_{jt}^{(n)} = \underline{\mathbf{p}}_{j\mathbf{x}}^{(n)'}(L)\mathbf{x}_t^{(n)}\}$ is called j-th dynamic principal component is a dynamic linear combination of $\mathbf{x}^{(n)}$ with maximum spectral density given by (see (53))

$$\underline{\mathbf{p}}_{j\mathbf{x}}^{(n)'}(e^{-i\theta})\mathbf{\Sigma}_{\mathbf{x}}^{(n)}(\theta)\underline{\mathbf{p}}_{j\mathbf{x}}^{(n)}(e^{i\theta}) = \mathbf{p}_{j\mathbf{x}}^{(n)'}(\theta)\mathbf{\Sigma}_{\mathbf{x}}^{(n)}(\theta)\bar{\mathbf{p}}_{j\mathbf{x}}^{(n)}(\theta) = \lambda_{j\mathbf{x}}^{(n)}(\theta).$$

Moreover, the filters are $n\times 1$ column vectors such that $\underline{\mathbf{p}}_{j\mathbf{x}}^{(n)}(L) = (\underline{p}_{j1\mathbf{x}}(L)\dots\underline{p}_{jn\mathbf{x}}(L))'$ with components defined as

$$\underline{p}_{ji\mathbf{x}}(L) = \sum_{k=-\infty}^{\infty}\underline{p}_{jik\mathbf{x}}L^k, \quad i = 1,\dots,n.$$

Now define a generic filter $\mathbf{a}^{(n)}(L) = (a_1(L)\dots a_n(L))'$ such that

$$a_i(L) = \sum_{k=-\infty}^{\infty}a_{ik}L^k, \quad i = 1,\dots,n.$$

It is possible to prove that

$$\underline{\mathbf{p}}_{1\mathbf{x}}^{(n)}(L) = \underset{\mathbf{a}^{(n)}(L)}{\operatorname{argmax}}\ \mathsf{Var}\left(\sum_{i=1}^{n}\sum_{k=-\infty}^{\infty}a_{ik}x_{it-k}\right),\ \text{s.t.}\ \sum_{i=1}^{n}\sum_{k=-\infty}^{\infty}a_{ik}^2 = 1$$

i.e. the first dynamic eigenvector defines a filter which has elements that deliver the dynamic linear combination of x_{it} with maximum variance. Such aggregated process, which we defined as $w_{1t}^{(n)} = \underline{\mathbf{p}}_{1\mathbf{x}}^{(n)'}(L)\mathbf{x}_t^{(n)}$, has variance

$$\mathsf{Var}\left(w_{1t}^{(n)}\right) = \mathsf{Var}\left(\sum_{i=1}^{n}\sum_{k=-\infty}^{\infty}\underline{p}_{1ik\mathbf{x}}x_{it-k}\right) = \int_{-\pi}^{\pi}\lambda_{1\mathbf{x}}^{(n)}(\theta)\mathrm{d}\theta.$$

Similarly the j-th dynamic principal component has maximum variance conditional on being orthogonal to the first $j-1$ dynamic principal components. If we consider the vector process $\mathbf{w}^{(n)} = \{(w_{1t}^{(n)}\dots w_{nt}^{(n)})'\}$ it can be seen that has a diagonal spectral density matrix having as elements the dynamic eigenvalues. Therefore, $\mathbf{w}^{(n)}$ has elements which are orthogonal at every lead and lag thus it constitutes an orthogonal basis of $\mathcal{H}^{\mathbf{x}^{(n)}}$. Such basis defines those dynamic linear combinations of x_{it} with maximum variance and are called dynamic principal components.

As in the static case, we can use dynamic principal components for dimension reduction. We know that $\mathbf{P}_{\mathbf{x}}^{(n)}(\theta)\mathbf{P}_{\mathbf{x}}^{(n)\dagger}(\theta) = \bar{\mathbf{P}}_{\mathbf{x}}^{(n)}(\theta)\mathbf{P}_{\mathbf{x}}^{(n)'}(\theta) = \mathbf{I}_n$, therefore

$$\mathbf{I}_n = \bar{\mathbf{p}}_{1\mathbf{x}}^{(n)}(\theta)\mathbf{p}_{1\mathbf{x}}^{(n)'}(\theta) + \ldots + \bar{\mathbf{p}}_{n\mathbf{x}}^{(n)}(\theta)\mathbf{p}_{n\mathbf{x}}^{(n)'}(\theta),$$

and for the corresponding filters we have

$$\mathbf{I}_n = \bar{\underline{\mathbf{p}}}_{1\mathbf{x}}^{(n)}(L)\underline{\mathbf{p}}_{1\mathbf{x}}^{(n)'}(L) + \ldots + \bar{\underline{\mathbf{p}}}_{n\mathbf{x}}^{(n)}(L)\underline{\mathbf{p}}_{n\mathbf{x}}^{(n)'}(L).$$

So that,

$$\begin{aligned} x_{it} &= \bar{\underline{p}}_{1i\mathbf{x}}^{(n)}(L)\underline{\mathbf{p}}_{1\mathbf{x}}^{(n)'}(L)\mathbf{x}_t^{(n)} + \ldots + \bar{\underline{p}}_{ni\mathbf{x}}^{(n)}(L)\underline{\mathbf{p}}_{n\mathbf{x}}^{(n)'}(L)\mathbf{x}_t^{(n)} = \\ &= \bar{\underline{p}}_{1i\mathbf{x}}^{(n)}(L)w_{1t}^{(n)} + \ldots + \bar{\underline{p}}_{ni\mathbf{x}}^{(n)}(L)w_{nt}^{(n)}, \quad i = 1,\dots,n. \end{aligned}$$

Since $\{(w_{1t}^{(n)}\dots w_{nt}^{(n)})'\}$ is an orthogonal basis of $\mathcal{H}^{\mathbf{x}^{(n)}}$, then we have decomposed x_{it} according to this

basis. If we project x_{it} onto the first q dynamic principal components, we recover the common component defined in Theorem 2 (compare with (38) for static PCA)

$$\chi_{it}^{(n)} = \underline{\bar{p}}_{1i\mathbf{x}}^{(n)}(L)\underline{\mathbf{p}}_{1\mathbf{x}}^{(n)'}(L)\mathbf{x}_t^{(n)} + \ldots + \underline{\bar{p}}_{qi\mathbf{x}}^{(n)}(L)\underline{\mathbf{p}}_{q\mathbf{x}}^{(n)'}(L)\mathbf{x}_t^{(n)} \quad i = 1, \ldots, n. \tag{54}$$

This representation in terms of a q-dimensional process is optimal in the sense that, for any other q-tuple in $\mathcal{H}^{\mathbf{x}^{(n)}}$, the projection $\widetilde{\chi}_{it}^{(n)}$ has larger residual variance:

$$\mathsf{E}[(x_{it} - \chi_{it}^{(n)})^2] \leq \mathsf{E}[(x_{it} - \widetilde{\chi}_{it}^{(n)})^2].$$

It is then possible to show that under Assumptions A1, A2, and A4 (Forni, Hallin, Lippi, and Reichlin, 2000)

$$|\chi_{it}^{(n)} - \chi_{it}| = o_p(1), \text{ as } n \to \infty, \quad i \in \mathbb{N}, \quad t \in \mathbb{Z}.$$

Therefore, projecting the data on the first q dynamic principal components allows to recover consistently the common component of each series.

This is again the blessing of dimensionality: the factor model which is unidentified under finite n gets asymptotically identified as $n \to \infty$, and boils down, in the limit to a dynamic PCA dimension reduction technique.

REPRESENTATION THEOREMS. By considering the analogy between static and dynamic PCA as a way to recover the common component in an approximate static or dynamic factor model, it is then clear that there exists a relation between idiosyncratic behaviour and dynamic eigenvalues. Indeed, we have seen that an idiosyncratic process is such that it has a negligible contribution to the total variance of an aggregated process when $n \to \infty$. If among all possible dynamic aggregations we consider those having maximum variance, then we know that their variances are given by the integrals of the dynamic eigenvalues, and we have the following result:

THEOREM 3 (Hallin and Lippi (2013)). *Under Assumptions A1, A2, and A4, a vector process $\boldsymbol{\xi} \in \mathcal{H}^{\mathbf{x}}$ is purely idiosyncratic if and only if, for any $n \in \mathbb{N}$, its largest dynamic eigenvalue $\lambda_{1\boldsymbol{\xi}}^{(n)}(\theta)$ is bounded a.e. in $[-\pi, \pi]$.*

Since by definition a purely idiosyncratic process is killed in the limit $n \to \infty$ by dynamic aggregation (see before the definition of purely common and purely idiosyncratic processes), then its dynamic eigenvalues which are the largest variance of all possible dynamic aggregations must be bounded. Notice that, since the spectral density matrices for different values of n are nested, then the eigenvalues increase monotonically with n and since they are non-negative, then either they are bounded or diverge to infinity.

The representation Theorem 2, jointly with Assumption A3 of finite dynamic dimension q for the common component has other implications for the dynamic eigenvalues.

THEOREM 4 (Hallin and Lippi (2013)). *Under Assumptions A1, A2, A3, and A4,*

(i) the dynamic eigenvalues of $\boldsymbol{\chi}^{(n)}$ are such that

$$\lim_{n\to\infty} \lambda_{q\boldsymbol{\chi}}^{(n)}(\theta) = \infty,$$

while $\lambda_{q+1\boldsymbol{\chi}}^{(n)}(\theta) = 0$, a.e. in $[-\pi, \pi]$ as $n \to \infty$;

(ii) the dynamic eigenvalues of $\mathbf{x}^{(n)}$ are such that

$$\lim_{n\to\infty} \lambda_{q\mathbf{x}}^{(n)}(\theta) = \infty,$$

while $\lambda_{q+1\mathbf{x}}^{(n)}(\theta)$ is bounded, a.e. in $[-\pi, \pi]$ as $n \to \infty$.

Theorem 4 is equivalent to the following original representation theorem used in all applied works on dynamic factor models.

THEOREM 5 (Forni and Lippi (2001)). *Under Assumptions A1, A2, and A4, the following statements are equivalent*

(a) the components of the panel admit a unique decomposition

$$x_{it} = \chi_{it} + \xi_{it}, \quad i \in \mathbb{N}, \quad t \in \mathbb{Z},$$

such that

(i) $\mathsf{E}[\chi_{it}\xi_{js}] = 0$ *for any* $i, j \in \mathbb{N}$ *and any* $t, s \in \mathbb{Z}$*;*

(ii) $\boldsymbol{\xi} = \{\xi_{it} | i \in \mathbb{N},\ t \in \mathbb{Z}\}$ *is purely idiosyncratic, i.e. for any* $n \in \mathbb{N}$ *its largest dynamic eigenvalue,* $\lambda^{(n)}_{1\boldsymbol{\xi}}(\theta)$*, is bounded a.e. in* $[-\pi, \pi]$*;*

(iii) there exists a q*-dimensional process* $\mathbf{u} = \{u_{jt} | j = 1, , \ldots, q,\ t \in \mathbb{Z}\}$ *of mutually orthogonal white noises, and a collection of square-summable filters* $b_{ij}(L)$*,* $i \in \mathbb{N}$*,* $j = 1, \ldots, q$ *such that*

$$\chi_{it} = \sum_{j=1}^{q} b_{ij}(L) u_{jt}, \quad i \in \mathbb{N}, \quad t \in \mathbb{Z};$$

Moreover, the dynamic eigenvalues of $\boldsymbol{\chi}^{(n)}$ *are such that* $\lim_{n\to\infty} \lambda^{(n)}_{q\boldsymbol{\chi}}(\theta) = \infty$*;*

(b) the dynamic eigenvalues of $\mathbf{x}^{(n)}$ *are such that* $\lim_{n\to\infty} \lambda^{(n)}_{q\mathbf{x}}(\theta) = \infty$*, while* $\lambda^{(n)}_{q+1\mathbf{x}}(\theta)$ *is bounded, a.e. in* $[-\pi, \pi]$ *as* $n \to \infty$*.*

Here we do not need Assumption A3 as the q-dynamic factor representation is given by the behavior of the dynamic eigenvalues. But the part *(b)* is equivalent to part *(ii)* of Theorem 4 which is a consequence of Assumption A3. Therefore, either we make Assumptions A1, A2, A3 and we have Theorem 2 or we make Assumptions A1, A2, A4 and we have Theorem 5. Let us compare the two theorems

- Theorem 2:
  - advantages: filters are one-sided;
  - disadvantages: Assumption A3 of q-dimensional innovations of the common component is strong;
- Theorem 5:
  - advantages: Assumption A4 of existence of spectral density is quite reasonable and the behaviour of idiosyncratic eigenvalues is implied by Theorem 3, it also defines q through behaviour of common eigenvalues;
  - disadvantages: filters are two-sided.

ESTIMATION VIA DYNAMIC PCA. Let us first assume that the number of factors q is given. When we observe only T realisations of $\mathbf{x}^{(n)}$ the spectral density has to be estimated. We use a suffix T to indicate estimated quantities. We start with the sample autocovariances

$$\boldsymbol{\Gamma}^{(nT)}_{\mathbf{x}}(k) = \frac{1}{T} \sum_{t=1}^{T-|k|} \mathbf{x}^{(n)}_t \mathbf{x}^{(n)'}_{t+k}, \quad k = 0 \ldots M_T, \quad \boldsymbol{\Gamma}^{(nT)}_{\mathbf{x}}(-k) = \boldsymbol{\Gamma}^{(nT)'}_{\mathbf{x}}(k). \tag{55}$$

One possible estimator of the spectral density is then

$$\boldsymbol{\Sigma}^{(nT)}_{\mathbf{x}}(\theta) = \sum_{k=-M_T}^{M_T} \boldsymbol{\Gamma}^{(nT)}_{\mathbf{x}}(k) \omega_k e^{-ik\theta}, \tag{56}$$

where ω_k are some weights, e.g. for the Bartlett window are such that $\omega_k = 1 - \frac{|k|}{M_T+1}$. Such estimator has a bias $O(M_T^{-1})$, while it has a variance $O(M_T/T)$. The resolution of the estimator in frequency domain is

$O(1/M_T)$ which motivates us to take M_T frequencies on a grid $\theta_j = \frac{2\pi j}{M_T}$, with $j = 0, \ldots, (M_T - 1)$. For consistency, we then must have, as $T \to \infty$, that $M_T \to \infty$ but $\frac{M_T}{T} \to 0$. It can then be shown that under suitable assumptions on the serial dependence of $\mathbf{x}_t^{(n)}$, the entries of the spectral density are estimated consistently with rate $O(\sqrt{(M_T \log M_T)/T})$ uniform over all frequencies (Wu and Zaffaroni, 2018, for details).

Once we have a consistent estimator of the spectral density matrix, then the dynamic eigenvectors are also consistently estimated by the eigenvectors of $\mathbf{\Sigma}_{\mathbf{x}}^{(nT)}(\theta)$ which we denote as $\mathbf{p}_{j\mathbf{x}}^{(nT)}(\theta)$. Now define the estimated common component as:

$$\chi_{it}^{(nT)} = \underline{\bar{p}}_{1i\mathbf{x}}^{(nT)}(L)\underline{\mathbf{p}}_{1\mathbf{x}}^{(nT)'}(L)\mathbf{x}_t^{(n)} + \ldots + \underline{\bar{p}}_{qi\mathbf{x}}^{(nT)}(L)\underline{\mathbf{p}}_{q\mathbf{x}}^{(nT)'}(L)\mathbf{x}_t^{(n)}. \tag{57}$$

In practice we have to define a grid of values of θ say we choose $2M_T + 1$ equally spaced values. The filters in (57) are then obtained as

$$\underline{p}_{ji\mathbf{x}}^{(nT)}(L) = \sum_{k=-M_T}^{M_T} \underline{p}_{jik\mathbf{x}}^{(nT)} L^k, \tag{58}$$

$$\underline{p}_{jik\mathbf{x}}^{(nT)} = \frac{1}{2M_T + 1} \sum_{h=-M_T}^{M_T} p_{ji\mathbf{x}}^{(nT)}(\theta_h) e^{ik\theta_h}, \tag{59}$$

where $p_{jik\mathbf{x}}^{(nT)}(\theta_h)$ is the i-th row of the dynamic eigenvector $\mathbf{p}_{j\mathbf{x}}^{(nT)}(\theta_h)$. Summarizing estimation is based on the following steps

1. compute the lagged covariances $\mathbf{\Gamma}_{\mathbf{x}}^{(nT)}(k)$ as in (55);
2. compute the spectral density $\mathbf{\Sigma}_{\mathbf{x}}^{(nT)}(\theta)$ as in (56);
3. compute the first q dynamic eigenvectors of the spectral density, $\mathbf{p}_{j\mathbf{x}}^{(nT)}(\theta)$, $j = 1, \ldots, q$;
4. compute the coefficients of the filters $\underline{p}_{jik\mathbf{x}}^{(nT)}$ as in (59);
5. compute the filters $\underline{p}_{ji\mathbf{x}}^{(nT)}(L)$ as in (58);
6. compute the common component $\chi_{it}^{(nT)}$ as in (57).

We have the following consistency result (Forni, Hallin, Lippi, and Reichlin, 2000). Under Assumptions A1-A4, as $n, T \to \infty$

$$|\chi_{it}^{(nT)} - \chi_{it}| = o_P(1), \quad i = 1, \ldots, n.$$

The rate of convergence depends on n, T and also on the chosen estimator of the spectral density, so for our choice it will also depend on M_T. Some details on rates are given in Forni, Hallin, Lippi, and Reichlin (2004). Therefore, as for the PCA estimator of the static model, we have consistency, i.e. convergence in probability, provided that both n and T are large. While the role of $T \to \infty$ is necessary to have consistent estimates of the spectral densities, the role for $n \to \infty$ is needed to disentangle and identify the common and idiosyncratic components (see Theorem 6).

LIMITATIONS OF DYNAMIC PRINCIPAL COMPONENTS. Although dynamic PCA is the way to proceed consistently with the representation results, it presents two main practical problems: (i) the common component is in general estimated by means of two-sided filters, and (ii) we cannot recover the space spanned by the dynamic factors.

First, consider the following "toy" example:

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

and assume $\mathsf{Var}(\xi_{it}) = \sigma_\xi^2 < \infty$ and $\mathsf{E}[\xi_{it}\xi_{jt}] = 0$, for any $i, j = 1, \ldots, n$ with $i \neq j$, and n even. Then,

$$\begin{aligned}
\boldsymbol{\Sigma}_{\mathbf{x}}^{(n)}(\theta) &= \frac{1}{2\pi}\begin{pmatrix} e^{-i\theta} \\ 1 \\ \vdots \\ e^{-i\theta} \\ 1 \end{pmatrix}\left(e^{i\theta}\ 1 \ldots e^{i\theta}\ 1\right) + \frac{\sigma_\xi^2}{2\pi}\mathbf{I} = \\
&= \frac{1}{2\pi}\begin{pmatrix} 1 & e^{-i\theta} & \ldots & 1 & e^{-i\theta} \\ e^{i\theta} & 1 & \ldots & e^{i\theta} & 1 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 1 & e^{-i\theta} & \ldots & 1 & e^{-i\theta} \\ e^{i\theta} & 1 & \ldots & e^{i\theta} & 1 \end{pmatrix} + \frac{\sigma_\xi^2}{2\pi}\mathbf{I},
\end{aligned}$$

which shows that the common component has spectral density of rank 1 since we have just one dynamic factor.

If we define $\mathbf{p}_{1\mathbf{x}}^{(n)}(\theta) = \frac{1}{\sqrt{n}}\left(e^{-i\theta}\ 1 \ldots e^{-i\theta}\ 1\right)'$, this is such that $\mathbf{p}_{1\mathbf{x}}^{(n)\dagger}(\theta)\mathbf{p}_{1\mathbf{x}}^{(n)}(\theta) = 1$ and it is an eigenvector with dynamic eigenvalue

$$\lambda_{1\mathbf{x}}^{(n)}(\theta) = \mathbf{p}_{1\mathbf{x}}^{(n)\dagger}(\theta)\boldsymbol{\Sigma}_{\mathbf{x}}^{(n)}(\theta)\mathbf{p}_{1\mathbf{x}}^{(n)}(\theta) = \frac{1}{2\pi}(n + \sigma_\xi^2).$$

The corresponding filter is such that $\underline{\mathbf{p}}_{1\mathbf{x}}^{(n)}(e^{-i\theta}) = \mathbf{p}_{1\mathbf{x}}^{(n)}(\theta)$ therefore

$$\underline{\mathbf{p}}_{1\mathbf{x}}^{(n)}(L) = \frac{1}{\sqrt{n}}(L\ 1 \ldots L\ 1)'.$$

Since $q = 1$, the first dynamic principal component is enough to estimate the common component. By projecting the data on the first dynamic principal component we have (see (54)):

$$\begin{aligned}
\chi_{(2i+1)t}^{(n)} &= \underline{\bar{p}}_{1(2i+1)\mathbf{x}}^{(n)}(L)\underline{\mathbf{p}}_{1\mathbf{x}}^{(n)'}(L)\mathbf{x}_t^{(n)} \\
&= \frac{L^{-1}}{n}(x_{1t-1} + x_{2t} + \ldots + x_{n-1t-1} + x_{nt}) \\
&= \frac{1}{n}(x_{1t} + x_{2t+1} + \ldots + x_{n-1t} + x_{nt+1}).
\end{aligned}$$

Analogously,

$$\chi_{(2i)t}^{(n)} = \underline{\bar{p}}_{1(2i)\mathbf{x}}^{(n)}(L)\underline{\mathbf{p}}_{1\mathbf{x}}^{(n)'}(L)\mathbf{x}_t^{(n)} = \frac{1}{n}(x_{1t-1} + x_{2t} + \ldots + x_{n-1t-1} + x_{nt}).$$

If we compute the variance of the previous expressions, we have for any i

$$\mathrm{Var}\left(\chi_{it}^{(n)}\right) = \mathsf{Var}(u_t) + \frac{1}{n}\sigma_\xi^2 \to \mathsf{Var}(u_t) = 1, \ \text{ as } \ n \to \infty.$$

which proves Theorem 6 for this simple example.

The problem of the previous procedure is the presence of two-sided (bilateral) filters, that is we need to use $\mathbf{x}_{t+1}$ to estimate $\boldsymbol{\chi}_t$. This makes impossible the use of the data up to time T. We need a way to estimate the model using only one-sided filters and we know that such representation exists from Theorem 2.

Second, notice that contrary to the static case we cannot recover the factor space by means of dynamic PCA. Indeed by analogy with the static case one would be tempted to write the estimated dynamic loadings and factors as as

$$\mathbf{B}^{(nT)}(L) = \sqrt{n}\underline{\mathbf{P}}_{\mathbf{x}}^{(nT)}(L), \qquad \mathbf{u}_t^{(nT)} = \frac{\underline{\mathbf{P}}_{\mathbf{x}}^{(nT)'}(L)\mathbf{x}_t^{(n)}}{n} \tag{60}$$

where $\underline{\mathbf{P}}_{\mathbf{x}}^{(nT)}(L) = (\underline{\mathbf{p}}_{1\mathbf{x}}^{(nT)}(L) \dots \underline{\mathbf{p}}_{q\mathbf{x}}^{(nT)}(L))'$. However, while the true common shocks $\mathbf{u}_t$ are a white noise process by definition, in general $\mathbf{u}_t^{(nT)}$ will be autocorrelated. Indeed in such a setting factors and loadings are identified only up to an invertible linear filter, thus also autocorrelation of the estimated factors can be affected by identification. Thus, the above definitions cannot give consistent estimators of the dynamic loadings and factors separately. Only the common component is recovered consistent by means of dynamic PCA.

ONE-SIDED REPRESENTATION AND ESTIMATION. A common component estimated in the way describe above is not useful for empirical purposes. In particular, while for forecasting it is only important to get an estimate using one-sided filters, for impulse-response analysis we also need an estimator of the common shocks and for portfolio optimisation it might be useful to estimate consistently the conditional second order structure of the factors themselves. Forni and Lippi (2011) and Forni, Hallin, Lippi, and Zaffaroni (2015a,b) solve this problem by exploiting results on singular processes by Anderson and Deistler (2008). From Theorems 2 and 5, we have

$$\boldsymbol{\chi}_t^{(n)} = \mathbf{B}^{(n)}(L)\mathbf{u}_t, \tag{61}$$

and, if the common component has rational spectral density, we can always write (61) as

$$\mathbf{A}^{(n)}(L)\boldsymbol{\chi}_t^{(n)} = \mathbf{H}^{(n)}\mathbf{u}_t, \tag{62}$$

where $\mathbf{H}^{(n)}$ is $n \times q$, assuming without loss of generality that $n = (q+1)m$ for some integer m,

$$\mathbf{A}^{(n)}(L) = \begin{pmatrix} \mathbf{A}_1^{(n)}(L) & \mathbf{0}_{q+1} & \dots & \mathbf{0}_{q+1} \\ \mathbf{0}_{q+1} & \mathbf{A}_2^{(n)}(L) & \dots & \mathbf{0}_{q+1} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0}_{q+1} & \mathbf{0}_{q+1} & \dots & \mathbf{A}_m^{(n)}(L) \end{pmatrix},$$

with $\mathbf{A}_j^{(n)}(L)$ are $(q+1) \times 1(q+1)$ finite filters such that $\det(\mathbf{A}_j^{(n)}(z)) \neq 0$ for $|z| \leq 1$. Then, the GDFM can be written as

$$\mathbf{A}^{(n)}(L)\mathbf{x}_t^{(n)} = \mathbf{H}^{(n)}\mathbf{u}_t + \mathbf{A}^{(n)}(L)\boldsymbol{\xi}_t^{(n)} \tag{63}$$

and it can be shown the the last term on the right hand side of (63) is still idiosyncratic in the sense of Theorem 5 (Forni and Lippi, 2011).

Estimation of (63) is based on the following steps (Forni, Hallin, Lippi, and Zaffaroni, 2015a,b).

1. Estimate the spectral density matrix of $\mathbf{x}_t^{(n)}$ using (56) and by means of dynamic PCA compute $\boldsymbol{\Sigma}_{\boldsymbol{\chi}}^{(nT)}(\theta_j)$, for $\theta_j = \frac{2\pi j}{M_T}$, an estimator of the spectral density matrix of $\boldsymbol{\chi}_t^{(n)}$.
2. By means of inverse Fourier transform compute

$$\boldsymbol{\Gamma}_{\boldsymbol{\chi}}^{(nT)}(k) = \frac{1}{2\pi}\sum_{j=0}^{M_T-1} \boldsymbol{\Sigma}_{\boldsymbol{\chi}}^{(nT)}(\theta_j)e^{ij\theta},$$

   the estimated autocovariances of $\boldsymbol{\chi}_t^{(n)}$.
3. By means of Yule-Walker equations compute $\mathbf{A}_j^{(nT)}(L)$ for $j = 1, \dots, m$ and the $(q+1)$-dimensional vectors $\boldsymbol{\epsilon}_{t,j}^{(nT)}$ of residuals of each of the m VAR models.
4. Consider the stacked n-dimensional vector $\boldsymbol{\epsilon}_t^{(nT)} = (\boldsymbol{\epsilon}_{t,1}^{(nT)'} \dots \boldsymbol{\epsilon}_{t,m}^{(nT)'})'$ and by PCA compute $\mathbf{H}^{(nT)}$ and $\mathbf{u}_t^{(nT)}$ as the first q principal components of $\boldsymbol{\epsilon}_t^{(nT)}$.

Define the estimated filters as $\mathbf{B}^{(nT)}(L) = [\mathbf{A}^{(nT)}(L)]^{-1}\mathbf{H}^{(nT)}$, having as rows $\mathbf{b}_i^{(nT)}(L)$, then we have

the following consistency results (Forni, Hallin, Lippi, and Zaffaroni, 2015b):

$$\|\mathbf{b}_i^{(nT)}(L) - \mathbf{b}_i(L)\mathbf{R}\| = O_p\left(\max\left(\frac{1}{\sqrt{n}}, \sqrt{\frac{M_T \log M_T}{T}}\right)\right),$$
$$\|\mathbf{u}_t^{(nT)} - \mathbf{R}^{-1}\mathbf{u}_t\| = O_p\left(\max\left(\frac{1}{\sqrt{n}}, \sqrt{\frac{M_T \log M_T}{T}}\right)\right),$$

for an invertible $q \times q$ matrix $\mathbf{R}$. Consistency of the estimated common component, $\chi_{it}^{(nT)} = \mathbf{b}_i^{(nT)'}(L)\mathbf{u}_t^{(nT)}$, is then immediate.

6 The approximate dynamic factor model in state-space form

A second approach to solve the two-sidedness problem is proposed by Forni, Giannone, Lippi, and Reichlin (2009) and is based on a state-space representation of the dynamic factor model, which in turn requires some additional assumption on the model and for this reason is sometimes called *restricted approximated dynamic factor model*. Here and in the next chapter we focus on this approach which is strictly related to the works by Stock and Watson (2005) and Bai and Ng (2006, 2007), which are in turn the theoretical basis also of those models usually called Factor Augmented VAR (FAVAR) and firstly proposed by Bernanke, Boivin, and Eliasz (2005). In fact, we show in this section that FAVAR are just a restricted version of the model considered in this section, where some constraints are imposed on the parameters.

SUMMARY OF REPRESENTATION RESULTS. We here summarise the representation results of Theorems 2-5 presented in Sections 4 and 5, which were proved by Forni, Hallin, Lippi, and Reichlin (2000); Forni and Lippi (2001); Hallin and Lippi (2013). Under Assumptions A1-A3 we have the approximate dynamic factor representation of Theorem 2

$$x_{it} = \chi_{it} + \xi_{it} = \mathbf{b}_i^{'}(L)\mathbf{u}_t + \xi_{it}, \quad i \in \mathbb{N}, \quad t \in \mathbb{Z}, \tag{64}$$

where

0. $\mathsf{E}[\chi_{it}\xi_{js}] = 0$ for any $i, j \in \mathbb{N}$ and any $t, s \in \mathbb{Z}$;
1. $\mathbf{u} = \{\mathbf{u}_t = (u_{1t} \dots u_{qt})' | t \in \mathbb{Z}\}$ is a q-dimensional process of orthonormal white noises and q does not depend on n;
2. $\mathbf{b}_i^{'}(L)$ is a $1 \times q$ one-sided infinite vector polynomial with square-summable coefficients;
3. $\boldsymbol{\chi}^{(n)} = \{\boldsymbol{\chi}_t^{(n)} = (\chi_{1t} \dots \chi_{nt})' | t \in \mathbb{Z}\}$ is common, that is, if A4 holds, $\lim_{n\to\infty} \lambda_{q\boldsymbol{\chi}}^{(n)}(\theta) = \infty$ (see also Theorem 4);
4. $\boldsymbol{\xi}^{(n)} = \{\boldsymbol{\xi}_t^{(n)} = (\xi_{1t} \dots \xi_{nt})' | t \in \mathbb{Z}\}$ is idiosyncratic, that is, if A4 holds, its largest dynamic eigenvalue $\lambda_{1\boldsymbol{\xi}}^{(n)}(\theta)$ is bounded a.e. in $[-\pi, \pi]$ and for any $n \in \mathbb{N}$ (see also Theorem 3).

Moreover, under Assumption A4 we know that (64) is equivalent to the following conditions on dynamic eigenvalues (see Theorem 5)

5. $\lim_{n\to\infty} \lambda_{q\mathbf{x}}^{(n)}(\theta) = \infty$;
6. $\lambda_{q+1\mathbf{x}}^{(n)}(\theta)$ is bounded a.e. in $[-\pi, \pi]$ and for any $n \in \mathbb{N}$.

EXAMPLE OF EQUIVALENCE OF DYNAMIC AND STATIC REPRESENTATIONS. In order to have an equivalent factor structure which we can then estimate by means of PCA, we would like to be able to write (64) as a static approximate factor model. Consider the example by Stock and Watson (2005) of a dynamic factor model where q factors are loaded with $(p-1)$ lags

$$x_{it} = \sum_{k=0}^{p-1} \boldsymbol{\lambda}_{ik}^{*'} \mathbf{f}_{t-k} + \xi_{it} = \boldsymbol{\lambda}_i^{*'}(L)\mathbf{f}_t + \xi_{it}, \quad i \in \mathbb{N}, \quad t \in \mathbb{Z}, \tag{65}$$

and we allow the factors to follow a VAR(p) model

$$\mathbf{f}_t = \sum_{k=1}^{p} \mathbf{C}_k \mathbf{f}_{t-k} + \boldsymbol{\eta}_t = \mathbf{C}(L)\mathbf{f}_{t-1} + \boldsymbol{\eta}_t, \quad \boldsymbol{\eta}_t \sim w.n.(\mathbf{0}, \boldsymbol{\Gamma}_{\boldsymbol{\eta}}), \quad t \in \mathbb{Z}. \tag{66}$$

This is now a model since more assumptions on the data generating process are made, still we have that (65)-(66) can be written as (64) just by setting

$$\mathbf{u}_t = \mathbf{R}^{-1}\boldsymbol{\eta}_t, \qquad \mathbf{b}_i^{'}(L) = \boldsymbol{\lambda}_i^{*'}(L)(\mathbf{I} - \mathbf{C}(L))^{-1}\mathbf{R},$$

where $\mathbf{R}$ is any $q \times q$ invertible matrix that makes $\mathbf{u}_t$ orthonormal (for example we can take $\mathbf{R} = \mathbf{\Gamma}_{\boldsymbol{\eta}}^{1/2}$).

Now by defining

$$\begin{aligned} \boldsymbol{\lambda}_i^{'} &= (\boldsymbol{\lambda}_{i0}^{*'} \dots \boldsymbol{\lambda}_{ip-1}^{*'}), \qquad i \in \mathbb{N}, \\ \mathbf{F}_t &= (\mathbf{f}_t \dots \mathbf{f}_{t-p+1})', \qquad t \in \mathbb{Z}, \end{aligned}$$

where $\boldsymbol{\lambda}_i^{'}$ is $1 \times qp$ and $\mathbf{F}_t$ is $qp \times 1$, we can write (65) as

$$x_{it} = \boldsymbol{\lambda}_i^{'} \mathbf{F}_t + \xi_{it}, \quad i \in \mathbb{N}, \quad t \in \mathbb{Z}, \tag{67}$$

which is a static factor model with $r = qp$ factors. Moreover, (66) can now be written as a singular VAR(1) model for the static factors:

$$\mathbf{F}_t = \mathbf{A}\mathbf{F}_{t-1} + \mathbf{G}\boldsymbol{\eta}_t, \qquad \boldsymbol{\eta}_t \sim w.n.(\mathbf{0}, \mathbf{\Gamma}_{\boldsymbol{\eta}}) \qquad t \in \mathbb{Z}, \tag{68}$$

where $\mathbf{A}$ is $r \times r$ and $\mathbf{G}$ is $r \times q$, and such that

$$\mathbf{A} = \begin{pmatrix} \mathbf{C}_1 & \mathbf{C}_2 & \dots & \mathbf{C}_{p-1} & \mathbf{C}_p \\ \mathbf{I}_q & \mathbf{0} & \dots & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_q & \dots & \mathbf{0} & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \mathbf{0} & \mathbf{0} & \dots & \mathbf{I}_q & \mathbf{0} \end{pmatrix}, \qquad \mathbf{G} = \begin{pmatrix} \mathbf{I}_q \\ \mathbf{0} \\ \mathbf{0} \\ \vdots \\ \mathbf{0} \end{pmatrix}. \tag{69}$$

Therefore, also (67)-(68) is equivalent to (64) just by setting

$$\mathbf{u}_t = \mathbf{R}^{-1}\boldsymbol{\eta}_t, \qquad \mathbf{b}_i^{'}(L) = \boldsymbol{\lambda}_i^{'}(\mathbf{I} - \mathbf{A}L)^{-1}\mathbf{G}\mathbf{R},$$

where, as before, $\mathbf{R}$ is any $q \times q$ invertible matrix that makes $\mathbf{u}_t$ orthonormal (for example we can take $\mathbf{R} = \mathbf{\Gamma}_{\boldsymbol{\eta}}^{1/2}$).

RESTRICTIONS IMPLIED BY THE STATE-SPACE MODEL. We say that a static factor model with a singular VAR representation for the factors as in (67)-(68) is written in state-space. It has to be noticed though that the example above and the derived equivalent representations hold if and only if we impose additional assumptions. We now look for the assumptions needed in order to allow us to always write (64) in state-space form.

First, let us consider the eigenvalues of the covariance matrices. It is clear that we must have a common component with covariance of rank r. In particular we assume

ASSUMPTION (A5). *The eigenvalues of the covariance matrix $\mathbf{\Gamma}_{\boldsymbol{\chi}}^{(n)}$ are such that there exist finite constants $\underline{c}_j, \overline{c}_j > 0$, $j = 1, \dots, r$, such that $\underline{c}_j > \overline{c}_{j+1}$, $j = 1, \dots, r-1$, and*

$$0 < \underline{c}_j < \liminf_{n\to\infty} \frac{1}{n}\lambda_{j\boldsymbol{\chi}}^{(n)} \leq \limsup_{n\to\infty} \frac{1}{n}\lambda_{j\boldsymbol{\chi}}^{(n)} \leq \overline{c}_j, \quad j = 1, \dots r.$$

We known that Assumption A5, is equivalent to conditions *(b')* and *(b")* (we assume divergence linear in n) given in Section 3 when considering the static model. So equivalently we can assume (see also the discussion at the end of Section 2):

ASSUMPTION (A5'). *There exists a full-rank $r \times r$ matrix $\mathbf{H}$ such that $\frac{\mathbf{\Lambda}'\mathbf{\Lambda}}{n} \to \mathbf{H}$, as $n \to \infty$, the covariance matrix $\mathbf{\Gamma}_{\mathbf{F}}$ is positive definite.*

Moreover, condition 4 above on the dynamic eigenvalues of the idiosyncratic component implies

7. $\lambda_{1\boldsymbol{\xi}}^{(n)}$ is bounded for any $n \in \mathbb{N}$.

That is, also the eigenvalues of the idiosyncratic covariance matrix, $\mathbf{\Gamma}_{\boldsymbol{\xi}}^{(n)}$, are bounded for any n. Therefore, condition *(a')* is also satisfied. We know that under *(a')* and *(b')*, PCA allows us to recover consistently the space spanned by the factors which in this case are the elements of the r-dimensional vector $\mathbf{F} = \{F_{jt} | j = 1, \dots, r,\ t \in \mathbb{Z}\}$. Moreover, we know that under those conditions the eigenvalues of the covariance of the data satisfy:

8. $\lim_{n\to\infty} \lambda_{r\mathbf{x}}^{(n)} = \infty$;
9. $\lambda_{r+1\mathbf{x}}^{(n)}$ is bounded for any $n \in \mathbb{N}$.

Since a VAR for the factors implies an ARMA for the common component we must require that:

ASSUMPTION (A6). *The process $\boldsymbol{\chi}^{(n)}$ has rational spectral density.*

Finally, in order for the static model to be equivalent to the dynamic one we must require the following condition.

ASSUMPTION (A7). *Denoting by $\mathcal{X}_t$ the space spanned by χ_{it}, $i \in \mathbb{N}$, $t \in \mathbb{Z}$, we assume that, for any t, $\dim \mathcal{X}_t = r < \infty$, with $q < r < n$.*

Notice that while $\mathcal{X}_t = \overline{\text{span}}\{\chi_{it},\ i \in \mathbb{N}\}$ contains only contemporaneous linear combinations of χ_{it}, the Hilbert space $\mathcal{H}_t^{\boldsymbol{\chi}} = \overline{\text{span}}\{\chi_{it-k},\ i \in \mathbb{N}, k \geq 0\}$, considered in Assumption A3, contains also dynamic linear combinations using observations up to time t. In particular, Assumption A7 implies that $p < \infty$ in the previous example and therefore rules out AR loadings of the common factors.[11] In Assumption A7 we assume that the static rank of $\boldsymbol{\chi}$ is r, which is the rank of its covariance, while in Assumption A3 we assumed that the dynamic rank of $\boldsymbol{\chi}$ is q, which is the rank of its spectral density.

We have the following result that states the existence of the state-space representation of the approximate dynamic factor model.

THEOREM 6 (Forni, Giannone, Lippi, and Reichlin (2009)). *Under Assumptions A1-A7 (with A5 or A5'),* (64) *can always be written as*

$$x_{it} = \chi_{it} + \xi_{it} = \boldsymbol{\lambda}_i' \mathbf{F}_t + \xi_{it}, \quad i \in \mathbb{N},\ t \in \mathbb{Z} \tag{70}$$

$$\mathbf{F}_t = \mathbf{N}(L)\mathbf{u}_t. \tag{71}$$

such that conditions 0-9 hold, $\boldsymbol{\lambda}_i'$ is $1 \times r$, $\mathbf{F}_t$ is $r \times 1$, $\mathbf{N}(L)$ is a $r \times q$ one-sided infinite matrix polynomial with square-summable coefficients, and, in general, $q < r$. Moreover, $\mathcal{H}_t^{\boldsymbol{\chi}} = \mathcal{H}_t^{\mathbf{u}}$.

As in the example above, by combining (70)-(71), we get an expression equivalent to (64)

$$x_{it} = \boldsymbol{\lambda}_i' \mathbf{N}(L)\mathbf{u}_t + \xi_{it}, \quad i \in \mathbb{N}, \quad t \in \mathbb{Z}. \tag{72}$$

We use the following terminology:

- $\mathbf{F}$ are called static factors: they are loaded only contemporaneously, however they are dynamic in the sense that their autocorrelation is not zero;
- $\mathbf{u}$ are called common shocks: they are loaded dynamically, i.e. with leads and lags, however they are static in the sense that their autocorrelation zero, being white noises.

[11] For example, under A7 we cannot have

$$\chi_{it} = \frac{1}{1 - \alpha_i L} u_t,$$

with $|\alpha_i| < 1$, which would require $r = \infty$ to be written in state-space form. This case is considered by Forni and Lippi (2011); Forni, Hallin, Lippi, and Zaffaroni (2015a,b).

IDENTIFICATION. Notice that in Theorem 6 no autoregressive representation is assumed for the factors. The factors by assumption follow a singular MA process and in general its inversion might lead to an infinite VAR or an ARMA. However, Anderson and Deistler (2008) show that, under Assumption A6 of rational spectral density, (71) admits a left-inverse of finite order such that[12]

$$\mathbf{G}(L)\mathbf{F}_t = \mathbf{N}(0)\mathbf{u}_t. \tag{73}$$

The last relation is still not operative since nothing is said about the VAR order (we just know it is finite) and $\mathbf{N}(0)$ cannot be identified. However, by recalling the example above, it is clear that without loss of generality a VAR(1) is sufficient, and hereafter we replace (71) with the following result.

COROLLARY 1. *Under the same assumptions of Theorem 6,* (71) *can be written as*

$$\mathbf{F}_t = \mathbf{A}\mathbf{F}_{t-1} + \mathbf{G}\boldsymbol{\eta}_t, \quad \boldsymbol{\eta}_t \sim w.n.(\mathbf{0}, \boldsymbol{\Gamma}_{\boldsymbol{\eta}}), \tag{74}$$

such that $\det(\mathbf{I}_r - \mathbf{A}z) \neq 0$ *for* $|z| \leq 1$*,* $\mathbf{G}$ *is* $r \times q$*,* $\boldsymbol{\eta}_t \sim w.n.(\mathbf{0}, \boldsymbol{\Gamma}_{\boldsymbol{\eta}})$ *and* $\mathbf{G}\boldsymbol{\eta}_t = \mathbf{N}(0)\mathbf{u}_t$*. Therefore, there exists a* $q \times q$ *invertible matrix* $\mathbf{R}$*, such that* $\mathbf{u}_t = \mathbf{R}^{-1}\boldsymbol{\eta}_t$*, which implies* $\mathbf{GR} = \mathbf{N}(0)$*.*

By combining (70) and (74), we get an expression equivalent to (64)

$$x_{it} = \boldsymbol{\lambda}_i^{'}(\mathbf{I} - \mathbf{A}L)^{-1}\mathbf{G}\mathbf{R}\mathbf{u}_t + \xi_{it}, \quad i \in \mathbb{N}, \quad t \in \mathbb{Z}.$$

The choice of $\mathbf{R}$ is not unique and is necessary only if we are interested in identifying the shocks (as for example in impulse response analysis). In particular, estimation of (74) might be needed only if we want to compute recursive forecasts or impulse responses, while if we are interested in just recovering the common component we can just focus on (70). Moreover, for forecasting we usually do not need to estimate the matrix $\mathbf{G}$ as it does not enter the h-step-ahead forecast of the VAR.

EXAMPLE OF EQUIVALENCE OF DYNAMIC AND STATIC PCA. Let us consider again example (60) where we have $q = 1$ and we have seen that the common component can be recovered by means of dynamic PCA. We now show that a static approach with $r = 2$ static factors can allow us to recover the common component exactly as in the dynamic PCA case but without requiring the use of two-sided filters. The model is equivalent to a model with two factors u_t and u_{t-1} loaded statically, and we can write

$$\begin{aligned}
\boldsymbol{\Gamma}_{\mathbf{x}}^{(n)} &= \begin{pmatrix} 0 & 1 \\ 1 & 0 \\ \vdots & \vdots \\ 0 & 1 \\ 1 & 0 \end{pmatrix} \underbrace{\mathsf{E}\left[\begin{pmatrix} u_t \\ u_{t-1} \end{pmatrix}(u_t \;\; u_{t-1})\right]}_{\mathbf{I}} \begin{pmatrix} 0 & 1 & \dots & 0 & 1 \\ 1 & 0 & \dots & 1 & 0 \end{pmatrix} + \sigma_\xi^2\mathbf{I} = \\
&= \begin{pmatrix} 1 & 0 & \dots & 1 & 0 \\ 0 & 1 & \dots & 0 & 1 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 1 & 0 & \dots & 1 & 0 \\ 0 & 1 & \dots & 0 & 1 \end{pmatrix} + \sigma_\xi^2\mathbf{I},
\end{aligned}$$

which has rank 2 since $r = 2$. We then need the first two eigenvectors and corresponding eigenvalues which are

$$\begin{aligned}
\mathbf{p}_{1\mathbf{x}}^{(n)} &= \frac{1}{\sqrt{n/2}}(0\;1\;0\dots0\;1)', \quad \lambda_{1\mathbf{x}}^{(n)} = \frac{n}{2} + \sigma_\xi^2, \\
\mathbf{p}_{2\mathbf{x}}^{(n)} &= \frac{1}{\sqrt{n/2}}(1\;0\;1\dots1\;0)', \quad \lambda_{2\mathbf{x}}^{(n)} = \frac{n}{2} + \sigma_\xi^2.
\end{aligned}$$

As explained in Section 3, the common component of series i is obtained by taking weights as the eigen-

[12] This result is a consequence of Theorem 7 in Section 9.

vectors (see (38)):

$$\begin{aligned}\chi_{(2i+1)t} &= \left(p^{(n)}_{1(2i+1)\mathbf{x}}\mathbf{p}^{(n)'}_{1\mathbf{x}} + p^{(n)'}_{2(2i+1)\mathbf{x}}\mathbf{p}^{(n)}_{2\mathbf{x}}\right)\mathbf{x}^{(n)}_t = \frac{1}{n/2}(x_{1t} + x_{3t} + \ldots + x_{n-3t} + x_{n-1t}) \\ &= u_{t-1} + \frac{1}{n/2}(\xi_{1t} + \xi_{3t} + \ldots + \xi_{n-3t} + \xi_{n-1t}).\end{aligned}$$

Analogously,

$$\begin{aligned}\chi_{(2i)t} &= \left(p^{(n)}_{1(2i)\mathbf{x}}\mathbf{p}^{(n)'}_{1\mathbf{x}} + p^{(n)'}_{2(2i)\mathbf{x}}\mathbf{p}^{(n)}_{2\mathbf{x}}\right)\mathbf{x}^{(n)}_t = \frac{1}{n/2}(x_{2t} + x_{4t} + \ldots + x_{n-2t} + x_{nt}) \\ &= u_t + \frac{1}{n/2}(\xi_{2t} + \xi_{4t} + \ldots + \xi_{n-2t} + \xi_{nt}).\end{aligned}$$

If we compute the variance of the previous expressions, we have for any i

$$\mathrm{Var}\left(\chi^{(n)}_{it}\right) = \mathsf{Var}(u_t) + \frac{n/2}{n^2/4}\sigma^2_\xi \to \mathsf{Var}(u_t) = 1, \text{ as } n \to \infty.$$

In this case since the space spanned by the common component at time t has dimension 2, which is finite as requested by Assumption A7, therefore we can use PCA to recover the common component and we do not need two-sided filters.

RELATION WITH FAVAR. An alternative approach used for dimension reduction is based on a VAR model made of few observed variables, say $\mathbf{Y}$ of dimension m, and of r unobserved factors, $\mathbf{F}$, containing all relevant information of a larger panel $\mathbf{x}^{(n)}$.

$$\begin{pmatrix} \mathbf{Y}_t \\ \mathbf{F}_t \end{pmatrix} = \mathbf{D}(L)\begin{pmatrix} \mathbf{Y}_{t-1} \\ \mathbf{F}_{t-1} \end{pmatrix} + \begin{pmatrix} \boldsymbol{\epsilon}^Y_t \\ \boldsymbol{\epsilon}^F_t \end{pmatrix}. \tag{75}$$

This is a mr dimensional VAR which is called Factor Augmented Vector Autoregression (FAVAR) and once the factors are recovered it can be estimated in the usual way (Bernanke, Boivin, and Eliasz, 2005; Bai and Ng, 2006). From the previous result we see that if the factors are recovered via PCA and $n > T$, then standard $\sqrt{T}$-asymptotics applies, and therefore we can use estimated factors as observed: we say there is no generated regressors problem. Otherwise, if $n < T$ we have to take into account the estimation error of the factors which will be of order $\sqrt{n}$.

If we choose $\mathbf{Y}$ to be a subset of $\mathbf{x}^{(n)}$ then also $\mathbf{Y}$ has a factor structure. However, in order to have a VAR representation like (75) for $(\mathbf{Y}_t, \mathbf{F}_t)$, the innovations $(\boldsymbol{\epsilon}^Y_t, \boldsymbol{\epsilon}^F_t)$ must be white noise. Therefore, if the idiosyncratic component is autocorrelated, we must write the dynamic factor model for the components of $\mathbf{Y}$ as (see also (65))

$$\begin{aligned} y_{it} &= \boldsymbol{\lambda}^{*'}_i(L)\mathbf{f}_t + \xi_{it} \\ &= \boldsymbol{\lambda}^{*'}_i(L)\mathbf{f}_t + \delta_i(L)\xi_{it-1} + \nu_{it} \\ &= \boldsymbol{\lambda}^{\dagger'}_i(L)\mathbf{f}_t + \delta_i(L)y_{it-1} + \nu_{it}, \qquad i = 1, \ldots, m, \quad t \in \mathbb{Z}, \end{aligned} \tag{76}$$

where $\boldsymbol{\lambda}^{\dagger'}_i(L) = (1 - \delta_i(L)L)\boldsymbol{\lambda}^{*'}_i(L)$ and $\nu_{it} \sim w.n.(0, \sigma^2_i)$. Using the static representation (70)-(74) and (76), the VAR (75) becomes (Stock and Watson, 2005)

$$\begin{aligned}\begin{pmatrix} \mathbf{Y}_t \\ \mathbf{F}_t \end{pmatrix} &= \begin{pmatrix} \mathbf{D}_{11}(L) & \boldsymbol{\lambda}'_Y\mathbf{A} \\ \mathbf{0} & \mathbf{A} \end{pmatrix}\begin{pmatrix} \mathbf{Y}_{t-1} \\ \mathbf{F}_{t-1} \end{pmatrix} + \begin{pmatrix} \boldsymbol{\lambda}'_Y\mathbf{G}\boldsymbol{\eta}_t \\ \mathbf{G}\boldsymbol{\eta}_t \end{pmatrix} + \begin{pmatrix} \boldsymbol{\nu}_t \\ \mathbf{0} \end{pmatrix} \\ &= \begin{pmatrix} \mathbf{D}_{11}(L) & \boldsymbol{\lambda}'_Y\mathbf{A} \\ \mathbf{0} & \mathbf{A} \end{pmatrix}\begin{pmatrix} \mathbf{Y}_{t-1} \\ \mathbf{F}_{t-1} \end{pmatrix} + \begin{pmatrix} \boldsymbol{\epsilon}^Y_t \\ \boldsymbol{\epsilon}^F_t \end{pmatrix},\end{aligned} \tag{77}$$

where $\mathbf{D}_{11}(L)$ is a diagonal $m \times m$ matrix polynomial with entries $d_i(L)$ and $\boldsymbol{\lambda}'_Y$ is the $m \times r$ loadings matrix for $\mathbf{Y}_t$. It has to be noticed that while traditional FAVAR models do not impose any constraint on (75), the following restrictions are implied by (77):

1. the covariance of $\mathbf{Y}_t$ has an r-factor structure;
2. $\boldsymbol{\epsilon}_t^Y = \boldsymbol{\lambda}_Y' \mathbf{G}\boldsymbol{\eta}_t + \boldsymbol{\nu}_t$, the innovations of $\mathbf{Y}_t$, has a q-factor structure;
3. $\boldsymbol{\epsilon}_t^F$, the innovations of $\mathbf{F}_t$, have a reduced rank covariance.

TWO-STEP ESTIMATION. The simplest way to estimate the model is in two steps as proposed by Giannone, Reichlin, and Sala (2005) and Forni, Giannone, Lippi, and Reichlin (2009)

1. PCA to estimate factors and loadings;
2. VAR on the estimated factors to estimate the other parameters.

For estimation we need two additional identifying assumptions.

ASSUMPTION (A8). *Consider $\mathbf{H}$ defined in Assumption (A5'), then $\boldsymbol{\Gamma}_{\mathbf{F}}^{1/2}\mathbf{H}\boldsymbol{\Gamma}_{\mathbf{F}}^{1/2}$ has distinct eigenvalues. In particular, we can choose one of the two equivalent set of identifying conditions:*

(i) $\mathbf{H} = \mathbf{I}_r$, $\boldsymbol{\Gamma}_{\mathbf{F}}$ *has distinct diagonal entries, and* $\mathbf{F}_t$ *are the principal components of* $\boldsymbol{\chi}_t^{(n)}$ *rescaled by* $\sqrt{n}$*;*

(ii) $\boldsymbol{\Gamma}_{\mathbf{F}} = \mathbf{I}_r$, $\mathbf{H}$ *has distinct eigenvalues, and* $\mathbf{F}_t$ *are the the principal components of* $\boldsymbol{\chi}_t^{(n)}$ *rescaled by* $\sqrt{\lambda_{j\boldsymbol{\chi}}^{(n)}}$.

ASSUMPTION (A9). *Consider $\mathbf{G}$ defined in Corollary 1, then $\mathbf{G}'\mathbf{G}$ has q non-zero distinct eigenvalues.*

Note that $\mathbf{G}$ must have full-rank for condition 5 to hold. We assume that r is known, then, given Assumption A8 with identification (i) and conditions 8-9, we can use the normalised eigenvectors corresponding to the r largest eigenvalues of the sample covariance matrix $\boldsymbol{\Gamma}_{\mathbf{x}}^{(nT)}$ to estimate the loadings and by means of the corresponding principal components we recover the space spanned by the factors. Denote by $\mathbf{P}^{(nT)}$ the $n \times r$ matrix of eigenvectors, then

$$\begin{aligned} \boldsymbol{\lambda}_i^{(nT)'} &= \sqrt{n}\mathbf{p}_i^{(nT)'}, \qquad i = 1, \dots, n, \\ \mathbf{F}_t^{(nT)} &= \frac{\mathbf{P}^{(nT)'}\mathbf{x}_t^{(n)}}{\sqrt{n}}, \quad t = 1, \dots, T, \\ \chi_{it} &= \boldsymbol{\lambda}_i^{(nT)'}\mathbf{F}_t^{(nT)} = \mathbf{p}_i^{(nT)'}\mathbf{P}^{(nT)'}\mathbf{x}_t^{(n)}, \end{aligned} \tag{78}$$

where $\mathbf{p}_i^{(nT)'}$ is the i-th row of $\mathbf{P}^{(nT)}$. We showed how to derive these estimators and we discussed their consistency in Section 3.

In the second step we estimate the VAR parameters by ordinary least squares

$$\mathbf{A}^{(nT)} = \left(\frac{1}{T}\sum_{t=2}^{T}\mathbf{F}_{t-1}^{(nT)}\mathbf{F}_{t-1}^{(nT)'}\right)^{-1}\left(\frac{1}{T}\sum_{t=2}^{T}\mathbf{F}_{t-1}^{(nT)}\mathbf{F}_{t}^{(nT)'}\right). \tag{79}$$

Finally, denote by $\boldsymbol{\epsilon}^{(nT)}$ the VAR residuals, with sample covariance

$$\boldsymbol{\Gamma}_{\boldsymbol{\epsilon}^{(nT)}}^{(nT)} = \frac{1}{T}\sum_{t=2}^{T}\boldsymbol{\epsilon}^{(nT)}\boldsymbol{\epsilon}^{(nT)'} = \frac{1}{T}\sum_{t=2}^{T}\left(\mathbf{F}_t^{(nT)} - \mathbf{A}^{(nT)}\mathbf{F}_{t-1}^{(nT)}\right)\left(\mathbf{F}_t^{(nT)} - \mathbf{A}^{(nT)}\mathbf{F}_{t-1}^{(nT)}\right)'.$$

An estimator of $\mathbf{G}$ can be obtained as the normalised eigenvectors corresponding to the q-largest eigenvalues of $\boldsymbol{\Gamma}_{\boldsymbol{\epsilon}^{(nT)}}^{(nT)}$. We have the estimator of the non-identified IRF

$$\mathbf{b}_i^{(nT)}(L) = \boldsymbol{\lambda}_i^{(nT)'}(\mathbf{I} - \mathbf{A}^{(nT)}L)^{-1}\mathbf{G}^{(nT)}, \qquad i = 1, \dots, n.$$

If we assume that the identification matrix $\mathbf{R}$ is known, then we have consistency of the identified IRF

$$|\mathbf{b}_i^{(nT)}(L)\mathbf{R} - \mathbf{b}_i(L)| = O_p\left(\max\left(\frac{1}{\sqrt{n}}, \frac{1}{\sqrt{T}}\right)\right), \qquad i = 1, \ldots, n.$$

7 Maximum Likelihood Estimation, EM and Kalman Smoother

THIS CHAPTER IS INCOMPLETE

We now consider more in detail estimation of the approximate dynamic factor model in state-space form given in (70) and (74):

$$\mathbf{x}_t^{(n)} = \mathbf{\Lambda}^{(n)}\mathbf{F}_t + \boldsymbol{\xi}_t^{(n)}, \tag{80}$$

$$\mathbf{F}_t = \mathbf{A}\mathbf{F}_{t-1} + \mathbf{G}\boldsymbol{\eta}_t, \tag{81}$$

$$\boldsymbol{\eta}_t = \mathbf{R}\mathbf{u}_t.$$

Throughout this section we assume that Assumptions A1-A4 in Sections 4 and 5 and A5'-A9 in Section 6 hold. These assumptions in turn imply that conditions 0-9 in Section 6, hold too.[13] Throughout we assume $\mathbf{R}$ to be known since this is usually determined ex-post and only if impulse response functions to macroeconomic shocks are of interest. Note that some of the asymptotic results stated in this chapter are proved in Barigozzi and Luciani (2017, 2019b) in a more general non-stationary setting (not covered here).

The two-step approach of Section 6 is very easy to implement however it has some limitations. First, PCA estimation is based only on the sample covariance matrix and therefore it does not take into account cross-sectional heteroschedasticity of idiosyncratic components (in other words heterogeneity in the signal to noise ratio). Second, since the focus is on the common component and therefore on equation (80), the fact that we do not use the second equation (81) for estimation of the common component might seems to be harmless. However, there is a feedback from the estimation of the factors' equation (81) (also called state equation) to the observables equation (80) (also called measurement equation). In other words when using PCA to estimate the common component we do not take into account the dynamics of the factors themselves. The dynamic equation (81) has also the effect that the variables are cross-correlated at different lags, hence in order to capture co-movements we need to take into account also these dynamic relations (in as similar way as done with spectral analysis in Section 5). Moreover, the idiosyncratic components might also be autocorrelated and this should also be taken into account. Although PCA estimators are consistent, all these mis-specifications when using PCA have a possibly non-negligible impact in terms of efficiency of the estimators, i.e. their variance is affected. We know that in a parametric model as (80)-(81), the most efficient estimator is given by Maximum Likelihood (ML). So in this chapter we consider different estimation techniques and their relation with the ML estimator.

Throughout, we also impose the following.

ASSUMPTION (A10). *The idiosyncratic components and the common shocks have a multivariate Gaussian distribution:*

(i) $\boldsymbol{\xi}_t^{(n)} \sim N(\mathbf{0}_n, \mathbf{\Gamma}_{\boldsymbol{\xi}})$;

(ii) $\boldsymbol{\eta}_t \sim wnN(\mathbf{0}_q, \mathbf{\Gamma}_{\boldsymbol{\eta}})$.

Hence, with respect to the previous chapters we introduce Gaussianity of the whole model for $\mathbf{x}_t$. This is reasonable as long as we deal with macroeconomic data. If Gaussianity is not the case but data still have a distribution of the Exponential family then the estimators discussed below are Quasi Maximum Likelihood estimators. Note that we already assumed in previous chapter that $\mathbf{u}_t$ is an orthonormal white noise, hence also $\boldsymbol{\eta}_t$ is a white noise, but we have not assumed Gaussianity before.

LIKELIHOOD. The model (80)-(81) depends on a set of parameters

$$\boldsymbol{\varphi}^{(n)} = (\text{vec}(\mathbf{\Lambda}^{(n)}), \text{vech}(\mathbf{\Gamma}_{\mathbf{F}}), \text{vech}(\mathbf{\Gamma}_{\boldsymbol{\xi}}^{(n)}), \text{vec}(\mathbf{A}), \text{vec}(\mathbf{G}), \text{vech}(\mathbf{\Gamma}_{\boldsymbol{\eta}})),$$

and on the unknown factors which can be collected into an rT-dimensional vector, which we denote as $\boldsymbol{F}_T = (\mathbf{F}_1' \dots \mathbf{F}_T')'$. We also define, the nT-dimensional vector of all data as $\boldsymbol{X}_T^{(n)} = (\mathbf{x}_1^{(n)'} \dots \mathbf{x}_T^{(n)'})'$

[13] In other words we are saying that Theorem 6 and Corollary 1 hold.

and the the nT-dimensional vector of all idiosyncratic components as $\boldsymbol{\Xi}_T^{(n)} = (\boldsymbol{\xi}_1^{(n)'} \dots \boldsymbol{\xi}_T^{(n)'})'$. Hereafter, when computing expectations we highlight the fact that these are taken with respect to the distribution of $\boldsymbol{X}^{(n)}$ computed using the a given value of the parameters $\boldsymbol{\varphi}^{(n)}$.

The covariance matrix of $\boldsymbol{X}_T^{(n)}$ is given by the $nT \times nT$ matrix (recall that $\mathsf{E}_{\boldsymbol{\varphi}^{(n)}}[\boldsymbol{X}_T^{(n)}] = \mathbf{0}$ and that $\boldsymbol{\Gamma}_{\mathbf{x}}^{(n)}(k) = \boldsymbol{\Gamma}_{\mathbf{x}}^{(n)'}(-k)$):

$$\boldsymbol{\Omega}_{T,\mathbf{x}}^{(n)} = \mathsf{E}_{\boldsymbol{\varphi}^{(n)}}[\boldsymbol{X}_T^{(n)} \boldsymbol{X}_T^{(n)'}] = \begin{pmatrix} \boldsymbol{\Gamma}_{\mathbf{x}}^{(n)} & \boldsymbol{\Gamma}_{\mathbf{x}}^{(n)}(1) & \boldsymbol{\Gamma}_{\mathbf{x}}^{(n)}(2) & \cdots & \boldsymbol{\Gamma}_{\mathbf{x}}^{(n)}(T-1) \\ \boldsymbol{\Gamma}_{\mathbf{x}}^{(n)'}(1) & \boldsymbol{\Gamma}_{\mathbf{x}}^{(n)} & \boldsymbol{\Gamma}_{\mathbf{x}}^{(n)}(1) & \cdots & \boldsymbol{\Gamma}_{\mathbf{x}}^{(n)}(T-2) \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \boldsymbol{\Gamma}_{\mathbf{x}}^{(n)'}(T-2) & \boldsymbol{\Gamma}_{\mathbf{x}}^{(n)'}(T-1) & \cdots & \boldsymbol{\Gamma}_{\mathbf{x}}^{(n)} & \boldsymbol{\Gamma}_{\mathbf{x}}^{(n)}(1) \\ \boldsymbol{\Gamma}_{\mathbf{x}}^{(n)'}(T-1) & \boldsymbol{\Gamma}_{\mathbf{x}}^{(n)'}(T-2) & \cdots & \boldsymbol{\Gamma}_{\mathbf{x}}^{(n)}(1) & \boldsymbol{\Gamma}_{\mathbf{x}}^{(n)} \end{pmatrix}$$

Note that this matrix contains all auto-covariances and thus contains all the information which is also contained in the spectral density matrix. Let also $\boldsymbol{L}_T^{(n)} = \mathbf{I}_T \otimes \boldsymbol{\Lambda}^{(n)}$ be the $nT \times rT$ matrix containing all factor loadings and define $\boldsymbol{\Omega}_{T,\mathbf{F}}^{(n)} = \mathsf{E}_{\boldsymbol{\varphi}^{(n)}}[\boldsymbol{F}_T \boldsymbol{F}_T']$ and $\boldsymbol{\Omega}_{T,\boldsymbol{\xi}}^{(n)} = \mathsf{E}_{\boldsymbol{\varphi}^{(n)}}[\boldsymbol{\Xi}_T^{(n)} \boldsymbol{\Xi}_T^{(n)'}]$ the $rT \times rT$ and $nT \times nT$ covariance matrices of $\boldsymbol{F}_T$ and of $\boldsymbol{\Xi}_T^{(n)}$ defined analogously to $\boldsymbol{\Omega}_{T,\mathbf{x}}^{(n)}$. Then, we have

$$\boldsymbol{\Omega}_{T,\mathbf{x}}^{(n)} = \boldsymbol{L}_T^{(n)} \boldsymbol{\Omega}_{T,\mathbf{F}}^{(n)} \boldsymbol{L}_T^{(n)'} + \boldsymbol{\Omega}_{T,\boldsymbol{\xi}}^{(n)},$$

which shows the dependence on all parameters. Then, under Assumption (A10) the log-likelihood of (80) is given by (omitting constant terms):

$$\begin{aligned} \ell(\boldsymbol{X}_T^{(n)}; \boldsymbol{\varphi}^{(n)}) &\simeq -\frac{1}{2} \log\det(\boldsymbol{\Omega}_T^{(n)}) - \frac{1}{2}\big[\boldsymbol{X}_T^{(n)'} \big(\boldsymbol{\Omega}_T^{(n)}\big)^{-1} \boldsymbol{X}_T^{(n)}\big]. \qquad (82) \\ &= -\frac{1}{2} \log\det(\boldsymbol{L}_T^{(n)} \boldsymbol{\Omega}_{T,\mathbf{F}}^{(n)} \boldsymbol{L}_T^{(n)'} + \boldsymbol{\Omega}_{T,\boldsymbol{\xi}}^{(n)}) - \frac{1}{2}\big[\boldsymbol{X}_T^{(n)'} \big(\boldsymbol{L}_T^{(n)} \boldsymbol{\Omega}_{T,\mathbf{F}}^{(n)} \boldsymbol{L}_T^{(n)'} + \boldsymbol{\Omega}_{T,\boldsymbol{\xi}}^{(n)}\big)^{-1} \boldsymbol{X}_T^{(n)}\big]. \end{aligned}$$

The optimal estimator of the parameters is then such that it maximises the above log-likelihood function (82). However, as it stands this problem is very hard (almost impossible) to be solved for large n. Note that the parameters $\mathbf{A}, \mathbf{G}, \boldsymbol{\Gamma}_{\boldsymbol{\eta}}$ are hidden into the autocovariances of the factors, thus they are in $\boldsymbol{\Omega}_{T,\mathbf{F}}^{(n)}$. Thus, we need to know the second moments of the factors in order to estimate the other parameters. For example, we have the Yule-Walker relations

$$\boldsymbol{\Gamma}_{\mathbf{F}} = \mathbf{A}\boldsymbol{\Gamma}_{\mathbf{F}}(-1) + \boldsymbol{\Gamma}_{\boldsymbol{\eta}}, \qquad \boldsymbol{\Gamma}_{\mathbf{F}}(1) = \mathbf{A}\boldsymbol{\Gamma}_{\mathbf{F}}$$

which imply[14]

$$\mathrm{vec}(\boldsymbol{\Gamma}_{\mathbf{F}}) = (\mathbf{I}_{r^2} - \mathbf{A} \otimes \mathbf{A})^{-1} \mathrm{vec}(\boldsymbol{\Gamma}_{\boldsymbol{\eta}}), \qquad \mathrm{vec}(\boldsymbol{\Gamma}_{\mathbf{F}}(1)) = (\mathbf{I}_r \otimes \mathbf{A})(\mathbf{I}_{r^2} - \mathbf{A} \otimes \mathbf{A})^{-1} \mathrm{vec}(\boldsymbol{\Gamma}_{\boldsymbol{\eta}}).$$

Similar relations would hold if we were to take into account also the serial correlation in the idiosyncratic components. In the next sections we will make various assumptions on $\boldsymbol{\Omega}_{T,\boldsymbol{\xi}}^{(n)}$ and $\boldsymbol{\Omega}_{T,\mathbf{F}}^{(n)}$ in order to simplify the problem.

OPTIMAL FACTOR ESTIMATOR. INCOMPLETE SECTION

GENERALISED LEAST SQUARES FACTOR ESTIMATOR. INCOMPLETE SECTION

MAXIMUM LIKELIHOOD ESTIMATOR OF THE FACTORS. INCOMPLETE SECTION

7.1 Static estimation

INCOMPLETE SECTION

[14] From Yule-Walker relations we have $\mathrm{vec}(\boldsymbol{\Gamma}_{\mathbf{F}}) = \mathrm{vec}(\mathbf{A}\boldsymbol{\Gamma}_{\mathbf{F}}\mathbf{A}') + \mathrm{vec}(\boldsymbol{\Gamma}_{\boldsymbol{\eta}})$. Then since $\mathrm{vec}(ABC) = (C' \otimes A)\mathrm{vec}(B)$ we have the results.

7.2 Dynamic estimation

In this section we consider the dynamic factor model where we now allow for autocorrelated factors, thus we consider the dynamic factor model

$$\mathbf{x}_t^{(n)} = \mathbf{\Lambda}^{(n)}\mathbf{F}_t + \boldsymbol{\xi}_t^{(n)}, \tag{83}$$
$$\mathbf{F}_t = \mathbf{A}\mathbf{F}_{t-1} + \mathbf{G}\boldsymbol{\eta}_t, \tag{84}$$
$$\boldsymbol{\eta}_t = \mathbf{R}\mathbf{u}_t, \tag{85}$$

where now we assume all Assumptions A1-A4 and A5'-A9 to hold together with conditions 0-9 and we assume $\mathbf{R}$ to be given. As before, we consider different restricted versions of A10. The parameters of the model are collected in the vector

$$\boldsymbol{\varphi}^{(n)} = (\text{vec}(\mathbf{\Lambda}^{(n)}), \text{vech}(\mathbf{\Gamma}_{\mathbf{F}}), \text{vech}(\mathbf{\Gamma}_{\boldsymbol{\xi}}^{(n)}), \text{vec}(\mathbf{A}), \text{vec}(\mathbf{G}), \text{vech}(\mathbf{\Gamma}_{\boldsymbol{\eta}})).$$

KALMAN FILTER. Let us first consider the case of known parameters. Then the factors can be estimated via the Kalman Filter (KF), which is a recursive procedure for computing the optimal estimator of the latent states, $\mathbf{F}_t$, at time t given all information up to time t. That is, for given values of the parameters, the KF gives $\mathsf{E}_{\boldsymbol{\varphi}^{(n)}}[\mathbf{F}_t|\boldsymbol{X}_t^{(n)}] = \text{Proj}_{\boldsymbol{\varphi}^{(n)}}\{\mathbf{F}_t|\boldsymbol{X}_t^{(n)}\}$. It was originally proposed by Kalman (1960) and Kalman and Bucy (1961). As noticed above this is not the ML estimator of the factors even if we use the true value of the parameters. Still this procedure is quite appealing and often used and, moreover, as seen below, is necessary as the first step for computing the ML estimator of the factors (see Moore and Anderson, 1979, Harvey, 1990 and Durbin and Koopman, 2012, for the finite n case, and Doz, Giannone, and Reichlin, 2011 for large n case).

Let us now describe the KF in detail. We use the traditional notation and denote the conditional mean and variance of the factors given all information up to time τ, with $1 \leq \tau \leq T$, as

$$\mathbf{F}_{t|\tau}^{(n)} = \mathsf{E}_{\boldsymbol{\varphi}^{(n)}}[\mathbf{F}_t|\boldsymbol{X}_\tau^{(n)}], \qquad \mathbf{P}_{t|s}^{(n)} = \mathsf{Var}_{\boldsymbol{\varphi}^{(n)}}[\mathbf{F}_t|\boldsymbol{X}_\tau^{(n)}] = \mathsf{E}_{\boldsymbol{\varphi}^{(n)}}[(\mathbf{F}_t - \mathbf{F}_{t|\tau}^{(n)})(\mathbf{F}_t - \mathbf{F}_{t|\tau}^{(n)})'|\boldsymbol{X}_\tau^{(n)}].$$

Then, the KF gives the filtered process $\mathbf{F}_{t|t}^{(n)}$ and its covariance $\mathbf{P}_{t|t}^{(n)}$, as well as the one-step-ahead prediction $\mathbf{F}_{t|t-1}^{(n)}$ and its covariance $\mathbf{P}_{t|t-1}^{(n)}$ (the case $\tau = T$ is considered below). Assume the initial conditions, $\mathbf{F}_{0|0}$ and $\mathbf{P}_{0|0}$, and the parameters of the model, to be known. Assuming to have the conditioning set $\boldsymbol{X}_{t-1}^{(n)}$ and that $\mathbf{F}_{t-1|t-1}^{(n)}$ and $\mathbf{P}_{t-1|t-1}^{(n)}$ are given, then the one-step-ahead predictions are obtained from the *prediction equations*

$$\begin{aligned} \mathbf{F}_{t|t-1}^{(n)} &= \mathbf{A}\mathbf{F}_{t-1|t-1}^{(n)}, \\ \mathbf{P}_{t|t-1}^{(n)} &= \mathbf{A}\mathbf{P}_{t-1|t-1}^{(n)}\mathbf{A}' + \mathbf{G}\mathbf{\Gamma}_{\boldsymbol{\eta}}\mathbf{G}'. \end{aligned} \tag{86}$$

Once a new observation $\mathbf{x}_t^{(n)}$ becomes available, so that the conditioning set becomes $\boldsymbol{X}_t^{(n)}$, we have the *update equations*

$$\begin{aligned} \mathbf{F}_{t|t}^{(n)} &= \mathbf{F}_{t|t-1}^{(n)} + \mathbf{P}_{t|t-1}^{(n)}\mathbf{\Lambda}^{(n)\prime}(\mathbf{\Lambda}^{(n)}\mathbf{P}_{t|t-1}^{(n)}\mathbf{\Lambda}^{(n)\prime} + \mathbf{\Gamma}_{\boldsymbol{\xi}}^{(n)})^{-1}(\mathbf{x}_t^{(n)} - \mathbf{\Lambda}^{(n)}\mathbf{F}_{t|t-1}^{(n)}), \\ \mathbf{P}_{t|t}^{(n)} &= \mathbf{P}_{t|t-1}^{(n)} - \mathbf{P}_{t|t-1}^{(n)}\mathbf{\Lambda}^{(n)\prime}(\mathbf{\Lambda}^{(n)}\mathbf{P}_{t|t-1}^{(n)}\mathbf{\Lambda}^{(n)\prime} + \mathbf{\Gamma}_{\boldsymbol{\xi}}^{(n)})^{-1}\mathbf{\Lambda}^{(n)}\mathbf{P}_{t|t-1}^{(n)}. \end{aligned} \tag{87}$$

The KF estimator of the factors is given by the first equation in (87): $\mathbf{F}_{t,\text{KF}}^{(n)} \equiv \mathbf{F}_{t|t}^{(n)}$. We see that this estimator is obtained as a weighted average of the one-step-ahead prediction of the factors and the prediction error of $\mathbf{x}_t^{(n)}$.[15]

[15] Notice that, by substituting (87) into (86), the KF recursions can be written as

$$\begin{aligned} \mathbf{F}_{t+1|t}^{(n)} &= \mathbf{A}\mathbf{F}_{t|t-1}^{(n)} + \mathbf{A}\mathbf{P}_{t|t-1}^{(n)}\mathbf{\Lambda}^{(n)\prime}(\mathbf{\Lambda}^{(n)}\mathbf{P}_{t|t-1}^{(n)}\mathbf{\Lambda}^{(n)\prime} + \mathbf{\Gamma}_{\boldsymbol{\xi}}^{(n)})^{-1}(\mathbf{x}_t^{(n)} - \mathbf{\Lambda}^{(n)}\mathbf{F}_{t|t-1}^{(n)}), \\ \mathbf{P}_{t+1|t}^{(n)} &= \mathbf{A}\mathbf{P}_{t|t-1}^{(n)}\mathbf{A}' + \mathbf{G}\mathbf{\Gamma}_{\boldsymbol{\eta}}\mathbf{G}' - \mathbf{A}\mathbf{P}_{t|t-1}^{(n)}\mathbf{\Lambda}^{(n)\prime}(\mathbf{\Lambda}^{(n)}\mathbf{P}_{t|t-1}^{(n)}\mathbf{\Lambda}^{(n)\prime} + \mathbf{\Gamma}_{\boldsymbol{\xi}}^{(n)})^{-1}\mathbf{\Lambda}^{(n)}\mathbf{P}_{t|t-1}^{(n)}\mathbf{A}'. \end{aligned}$$

The above recursion starts from $t = 0$ with some given initial conditions and ends at $t = T$. The KF is usually initialised using the first observation of the factors, $\mathbf{F}_{0|0}$, estimated via PCA and by setting $\mathbf{P}_{0|0} = \infty$ (or in practice a large value). The effect of such initialisation is limited in time. Indeed, it is well known that, if $\mathbf{\Lambda}^{(n)}$ and $\mathbf{G}$ have full column rank as in our Assumptions A5' and A9, then after few iteration the one-step-ahead covariance matrix $\mathbf{P}^{(n)}_{t|t-1}$ converges to a steady-state $\mathbf{P}^{(n)}$, in particular this happens exponentially fast (see Harvey, 1990). Moreover, we can show that, for large n, the steady-state is given by $\mathbf{P}^{(n)} = \mathbf{G}\mathbf{\Gamma}_{\boldsymbol{\eta}}\mathbf{G}' + O(n^{-1})$, and we define $\mathbf{P} = \mathbf{G}\mathbf{\Gamma}_{\boldsymbol{\eta}}\mathbf{G}'$ (see Barigozzi and Luciani, 2019a).

Then, using the steady-state and (87), the KF estimator of the factors is given by

$$\mathbf{F}^{(n)}_{t,\text{KF}} = \mathbf{F}^{(n)}_{t|t-1} + \mathbf{P}\mathbf{\Lambda}^{(n)\prime}\left(\mathbf{\Lambda}^{(n)}\mathbf{P}\mathbf{\Lambda}^{(n)\prime} + \mathbf{\Gamma}^{(n)}_{\boldsymbol{\xi}}\right)^{-1}\left(\mathbf{x}^{(n)}_t - \mathbf{\Lambda}^{(n)}\mathbf{F}^{(n)}_{t|t-1}\right) + O\left(\frac{1}{n}\right). \tag{88}$$

where it is intended that this relation holds after the system reaches the steady-state, hence for a given $\bar{t}(T) > 0$ such that there exists a finite positive M for which

$$\lim_{T\to\infty}\sup Te^{-\bar{t}(T)} \leq M < \infty. \tag{89}$$

Consistency of $\mathbf{F}^{(n)}_{t,\text{KF}}$ in the general singular case $q \leq r$ is given in Barigozzi and Luciani (2019a). Here, for simplicity let us assume $q = r$ so that $\mathbf{P}$ is invertible, then by using the Woodbury formula (88) becomes

$$\begin{aligned}
\mathbf{F}^{(n)}_{t,\text{KF}} &= \mathbf{F}^{(n)}_{t|t-1} + \left(\mathbf{\Lambda}^{(n)\prime}(\mathbf{\Gamma}^{(n)}_{\boldsymbol{\xi}})^{-1}\mathbf{\Lambda}^{(n)} + \mathbf{P}^{-1}\right)^{-1}\mathbf{\Lambda}^{(n)\prime}(\mathbf{\Gamma}^{(n)}_{\boldsymbol{\xi}})^{-1}\left(\mathbf{x}^{(n)}_t - \mathbf{\Lambda}^{(n)}\mathbf{F}^{(n)}_{t|t-1}\right) + O\left(\frac{1}{n}\right) \\
&= \mathbf{F}^{(n)}_{t|t-1} + \left(\mathbf{\Lambda}^{(n)\prime}(\mathbf{\Gamma}^{(n)}_{\boldsymbol{\xi}})^{-1}\mathbf{\Lambda}^{(n)}\right)^{-1}\mathbf{\Lambda}^{(n)\prime}(\mathbf{\Gamma}^{(n)}_{\boldsymbol{\xi}})^{-1}\left(\mathbf{x}^{(n)}_t - \mathbf{\Lambda}^{(n)}\mathbf{F}^{(n)}_{t|t-1}\right) + O\left(\frac{1}{n}\right) + O\left(\frac{1}{n}\right) \\
&= \left(\mathbf{\Lambda}^{(n)\prime}(\mathbf{\Gamma}^{(n)}_{\boldsymbol{\xi}})^{-1}\mathbf{\Lambda}^{(n)}\right)^{-1}\mathbf{\Lambda}^{(n)\prime}(\mathbf{\Gamma}^{(n)}_{\boldsymbol{\xi}})^{-1}\mathbf{x}^{(n)}_t + O\left(\frac{1}{n}\right) \\
&= \mathbf{F}_t + \left(\mathbf{\Lambda}^{(n)\prime}(\mathbf{\Gamma}^{(n)}_{\boldsymbol{\xi}})^{-1}\mathbf{\Lambda}^{(n)}\right)^{-1}\mathbf{\Lambda}^{(n)\prime}(\mathbf{\Gamma}^{(n)}_{\boldsymbol{\xi}})^{-1}\boldsymbol{\xi}^{(n)}_t + O\left(\frac{1}{n}\right) \\
&= \mathbf{F}_t + O_p\left(\frac{1}{\sqrt{n}}\right) + O\left(\frac{1}{n}\right),
\end{aligned} \tag{90}$$

where we used condition 9 of weak idiosyncratic dependence together with Assumption A8 (i), which allow us to simplify the denominator. For given values of the parameters the KF estimator of the factors is consistent.

KALMAN SMOOTHER. As discussed above the KF estimator of the factors is not the ML estimator of the factors even if we use the true value of the parameters. We known that the ML estimator of the factors is given by the conditional expectation of the factors given all available data $\boldsymbol{X}^{(n)}_T$. Consistently with the notation above the ML estimator of the factors is then given by

$$\mathbf{F}^{(n)}_{t,\text{ML}} \equiv \mathbf{F}^{(n)}_{t|T} = \mathsf{E}_{\boldsymbol{\varphi}^{(n)}}[\mathbf{F}_t|\boldsymbol{X}^{(n)}_T] = \text{Proj}_{\boldsymbol{\varphi}^{(n)}}\{\mathbf{F}_t|\boldsymbol{X}^{(n)}_T\},$$

since we always assume Gaussianity. An expression for this estimator is given in (**??**), where we also proved its consistency given the true value of the parameters. However, note that computing that expression directly requires inverting $\mathbf{\Omega}^{(n)}_{T,\boldsymbol{\xi}}$ which is $nT \times nT$ and even when this is assumed to be diagonal, we still have to invert $\mathbf{\Omega}_{T,\mathbf{F}}$ which is $rT \times rT$ which is smaller but still large when T is large.

There are two other alternative ways of computing $\mathbf{F}^{(n)}_{t,\text{ML}}$ a first one is in time domain and is considered here, while a second one requires spectral analysis and is considered at the end of this section. The first

The second equation is the so-called Riccati difference equation, and the quantity

$$\mathbf{K}^{(n)}_t = \mathbf{A}\mathbf{P}^{(n)}_{t|t-1}\mathbf{\Lambda}^{(n)\prime}\mathbf{\Lambda}^{(n)\prime}(\mathbf{\Lambda}^{(n)}\mathbf{P}^{(n)}_{t|t-1}\mathbf{\Lambda}^{(n)\prime} + \mathbf{\Gamma}^{(n)}_{\boldsymbol{\xi}})^{-1}$$

is called *Kalman gain* and is used to update predictions.

approach is by means of the Kalman Smoother (KS), which is a recursion following the KF and starting from $t = T$ and going backwards to $t = 0$ and it delivers the smoothed process $\mathbf{F}^{(n)}_{t|T}$ and its covariance $\mathbf{P}^{(n)}_{t|T}$. The KS is initialised by $\mathbf{F}^{(n)}_{T|T}$, which is obtained from the last iteration of the KF, and is then given by the recursion

$$\mathbf{F}^{(n)}_{t|T} = \mathbf{F}^{(n)}_{t|t} + \mathbf{P}^{(n)}_{t|t}\mathbf{A}'(\mathbf{P}^{(n)}_{t+1|t})^{-1}(\mathbf{F}^{(n)}_{t+1|T} - \mathbf{F}^{(n)}_{t+1|t}), \tag{91}$$

where $\mathbf{P}^{(n)}_{t+1|t}$ and $\mathbf{F}^{(n)}_{t|t} \equiv \mathbf{F}^{(n)}_{t,\text{KF}}$ are given by the KF (see (86) and (87)). The ML estimator of the factors is then given by $\mathbf{F}^{(n)}_{t,\text{ML}} \equiv \mathbf{F}^{(n)}_{t|T}$ and we also denote it as $\mathbf{F}^{(n)}_{t,\text{KS}}$ when computed using (91).

Using the steady-state and (91), the KS, i.e. the ML, estimator of the factors can be written as the recursion

$$\mathbf{F}^{(n)}_{t,\text{KS}} = \mathbf{F}^{(n)}_{t,\text{KF}} + \mathbf{P}^{(n)}_{t|t}\mathbf{A}'(\mathbf{P})^{-1}(\mathbf{F}^{(n)}_{t+1,\text{ML}} - \mathbf{F}^{(n)}_{t+1,\text{KF}}) + O\left(\frac{1}{n}\right). \tag{92}$$

where it is intended that this relation holds after the system reaches the steady-state, hence for a given $\bar{t}(T) > 0$ such that (89) holds. In order to show consistency of (91), for simplicity let us assume $q = r$ so that $\mathbf{P}$ is invertible. Then, by consistency of the KF estimator in (90) we have also that $\mathbf{P}^{(n)}_{t|t} = O_p(n^{-1})$, therefore

$$\mathbf{F}^{(n)}_{t,\text{KS}} = \mathbf{F}^{(n)}_{t,\text{KF}} + O_p\left(\frac{1}{n}\right) = \mathbf{F}_t + O_p\left(\frac{1}{\sqrt{n}}\right) + O_p\left(\frac{1}{n}\right). \tag{93}$$

This is an alternative proof of consistency of the ML estimator of the factors when writing it using the KS.

Note that in the singular case $q < r$, we cannot prove consistency using (**??**) since that expression would depend on $\mathbf{\Omega}_{T,\mathbf{F}}$ which is singular when $q < r$. We cannot use the steady-state in (91) either, because $\mathbf{P}$ is not be invertible. But we could instead use the steady-state approach in the recursion in Durbin and Koopman (2012) based on a different expression, which does not require inversion of $\mathbf{P}$. Consistency of $\mathbf{F}^{(n)}_{t,\text{ML}}$ in the general singular case $q < r$ is given in Barigozzi and Luciani (2019a) and is based on such an approach.

KALMAN SMOOTHER WITH PRE-ESTIMATED PARAMETERS. In order to implement the KS we need an estimator of the parameters. Let us first modify Assumption (A10) to the following.

ASSUMPTION (A10-D1). *$\boldsymbol{\eta}_t \sim wnN(\mathbf{0}_q, \mathbf{\Gamma}_{\boldsymbol{\eta}})$ and $\boldsymbol{\xi}^{(n)}_t \sim wnN(\mathbf{0}_n, \mathbf{\Gamma}^{(n)}_{\boldsymbol{\xi}})$, with $\mathbf{\Gamma}^{(n)}_{\boldsymbol{\xi}}$ diagonal.*

Hence, we now assume that idiosyncratic components are uncorrelated both serially and cross-sectionally. The parameter vector is now given by

$$\boldsymbol{\varphi}^{(n)} = (\text{vec}(\mathbf{\Lambda}^{(n)}), \text{vech}(\mathbf{\Gamma}_{\mathbf{F}}), \text{diag}(\mathbf{\Gamma}^{(n)}_{\boldsymbol{\xi}}), \text{vec}(\mathbf{A}), \text{vec}(\mathbf{G}), \text{vech}(\mathbf{\Gamma}_{\boldsymbol{\eta}}))..$$

A first possibility to estimate the model is to replace the unknown parameters $\boldsymbol{\varphi}^{(n)}$, with those obtained as described in Section 6, that is based on PCA and VAR on estimated factors. We know that these estimators denoted as $\boldsymbol{\varphi}^{(nT)}_{\text{PCA-VAR}}$ are consistent with a rate is $O_p(\max(1/\sqrt{T}, 1/\sqrt{n}))$ (see Section 6 and Forni, Giannone, Lippi, and Reichlin, 2009). Note that thanks to A10-D1, we only retain the diagonal of the PCA estimator $\mathbf{\Gamma}^{(n)}_{\boldsymbol{\xi},\text{PCA}}$ of the idiosyncratic covariance, which is always positive definite. Then, when using $\boldsymbol{\varphi}^{(nT)}_{\text{PCA-VAR}}$ in (90), we can still prove consistency of the KF estimator, by including also that additional estimation error. As a consequence and because of (92), (93), and consistency of the estimated parameters $\boldsymbol{\varphi}^{(nT)}_{\text{PCA-VAR}}$, we still have consistency also of the KS estimator.

MAXIMUM LIKELIHOOD ESTIMATION OF EXACT DYNAMIC FACTOR MODEL. The method above although consistent is not the most efficient. Indeed, the estimated parameters are not obtained as ML estimators. Let us assume that A10-D1 still holds. Then, instead of estimating once the parameters and then running the KS with those estimates, we can think of iterating between the two estimators. Indeed, given the ML estimator of the parameters the KS gives the ML estimator of the factors, and viceversa, given the ML estimator of the factors the parameters can be estimated by maximising the log-likelihood. Formally, such an iterative procedure can be implemented through the Expectation Maximisation EM algorithm, which was originally proposed by Dempster, Laird, and Rubin (1977) in a general setting and then applied to factor

models by Shumway and Stoffer (1982) (see also Watson and Engle, 1983 and Quah and Sargent, 1993 for the fixed n case and Doz, Giannone, and Reichlin, 2012 and Barigozzi and Luciani, 2019a, for the large n case).

In order to have an intuition of the EM algorithm, let us write the log-likelihood (82) as

$$\ell(\boldsymbol{X}_T^{(n)};\boldsymbol{\varphi}^{(n)}) = \ell(\boldsymbol{X}_T^{(n)},\boldsymbol{F}_T;\boldsymbol{\varphi}^{(n)}) - \ell(\boldsymbol{F}_T|\boldsymbol{X}_T^{(n)};\boldsymbol{\varphi}^{(n)}). \tag{94}$$

Then, denote as $\mathrm{F}(\boldsymbol{F}_T|\boldsymbol{X}_T^{(n)};\boldsymbol{\varphi}^{(n)})$ the distribution of the factors given the data and the parameters. Define the expectations of the rhs of (94) with respect to this distribution computed using an estimator of the parameters $\boldsymbol{\varphi}^{(nT)}$ as

$$\mathcal{Q}\left(\boldsymbol{\varphi}^{(n)};\boldsymbol{\varphi}^{(nT)}\right) = \int_{\mathbb{R}^{rT}} \ell(\boldsymbol{X}_T^{(n)},\boldsymbol{F}_T;\boldsymbol{\varphi}^{(n)})\mathrm{dF}(\boldsymbol{F}_T|\boldsymbol{X}_T^{(n)};\boldsymbol{\varphi}^{(nT)}) = \mathsf{E}_{\boldsymbol{\varphi}^{(nT)}}[\ell(\boldsymbol{X}_T^{(n)},\boldsymbol{F}_T;\boldsymbol{\varphi}^{(n)})|\boldsymbol{X}_T^{(n)}],$$
$$\mathcal{H}\left(\boldsymbol{\varphi}^{(n)};\boldsymbol{\varphi}^{(nT)}\right) = \int_{\mathbb{R}^{rT}} \ell(\boldsymbol{F}_T|\boldsymbol{X}_T^{(n)};\boldsymbol{\varphi}^{(n)})\mathrm{dF}(\boldsymbol{F}_T|\boldsymbol{X}_T^{(n)};\boldsymbol{\varphi}^{(nT)}) = \mathsf{E}_{\boldsymbol{\varphi}^{(nT)}}[\ell(\boldsymbol{F}_T|\boldsymbol{X}_T^{(n)};\boldsymbol{\varphi}^{(n)})|\boldsymbol{X}_T^{(n)}].$$

Notice that, since the lhs of (94) does not depend on $\boldsymbol{F}_T$, we have

$$\ell(\boldsymbol{X}_T^{(n)};\boldsymbol{\varphi}^{(n)}) = \mathcal{Q}\left(\boldsymbol{\varphi}^{(n)};\boldsymbol{\varphi}^{(nT)}\right) - \mathcal{H}\left(\boldsymbol{\varphi}^{(n)};\boldsymbol{\varphi}^{(nT)}\right). \tag{95}$$

Thus, if we have the conditional distribution of the factors given the data we can maximise the rhs (95), thus maximising the log-likelihood. This is the intuition of the EM algorithm.

More precisely, the EM algorithm is based on two steps, which at iteration j are given by:

1. Expectation step (E-step): for a given estimator of the parameters $\boldsymbol{\varphi}_j^{(nT)}$ compute $\mathcal{Q}(\boldsymbol{\varphi}^{(n)};\boldsymbol{\varphi}_j^{(nT)})$, for which we need the conditional moments of the factors $\mathsf{E}_{\boldsymbol{\varphi}_j^{(nT)}}[\mathbf{F}_t|\boldsymbol{X}_T^{(n)}]$ and $\mathsf{Var}_{\boldsymbol{\varphi}_j^{(nT)}}[\mathbf{F}_t|\boldsymbol{X}_T^{(n)}]$, which can be computed via the KS using $\boldsymbol{\varphi}_j^{(nT)}$;
2. Maximisation step (M-step): define a new estimator of the parameters as

$$\boldsymbol{\varphi}_{j+1}^{(nT)} = \arg\max_{\boldsymbol{\varphi}^{(n)}} \mathcal{Q}\left(\boldsymbol{\varphi}^{(n)};\boldsymbol{\varphi}_j^{(nT)}\right)$$

We iterate until the change in the values of the parameters does not exceed a given fixed threshold. At each iteration, the log-likelihood increases, indeed by definition of Kullback-Leibler divergence, we have

$$\mathcal{H}(\boldsymbol{\varphi}_{j+1}^{(nT)};\boldsymbol{\varphi}_j^{(nT)}) \leq \mathcal{H}(\boldsymbol{\varphi}_j^{(nT)};\boldsymbol{\varphi}_j^{(nT)}).$$

If $\mathcal{Q}(\boldsymbol{\varphi}^{(n)};\boldsymbol{\varphi}^{(n)^*})$ and its gradient are continuous in $\boldsymbol{\varphi}^{(n)}$ and $\boldsymbol{\varphi}^{(n)*}$, then the parameters obtained at the last iteration and denoted as $\boldsymbol{\varphi}_{\text{EM}}^{(nT)}$ converge to the ML estimator $\boldsymbol{\varphi}_{\text{ML}}^{(nT)}$ (see Wu, 1983 for details). The convergence is not exact since we stop the algorithm after a finite number of iterations, hence there is a numerical approximation error between $\boldsymbol{\varphi}_{\text{EM}}^{(nT)}$ and $\boldsymbol{\varphi}_{\text{ML}}^{(nT)}$ (see Meng and Rubin, 1994 for details). If the algorithm is initialised with a consistent estimator of the parameters, e.g. with $\boldsymbol{\varphi}_{\text{PCA-VAR}}^{(nT)}$ thus such approximation becomes smaller provided we iterate few times, or in other words the sensibility to the threshold imposed is minimised .

We now derive the estimators of the parameters explicitly. First, we compute the expected log-likelihood. It is convenient to write

$$\ell(\boldsymbol{X}_T^{(n)},\boldsymbol{F}_T;\boldsymbol{\varphi}^{(n)}) = \ell(\boldsymbol{X}_T^{(n)}|\boldsymbol{F}_T;\boldsymbol{\varphi}^{(n)}) + \ell(\boldsymbol{F}_T;\boldsymbol{\varphi}^{(n)}).$$

General expressions for these log-likelihoods are given in (**??**)-(**??**), but under A10-D1 these can be simpli-

fied further. First, if $\boldsymbol{\xi}_t^{(n)}$ is a white noise, then $\boldsymbol{\Omega}_{T,\boldsymbol{\xi}}^{(n)} = \mathbf{I}_T \otimes \boldsymbol{\Gamma}_{\boldsymbol{\xi}}^{(n)}$ and (**??**) becomes

$$\ell(\boldsymbol{X}_T^{(n)}, \boldsymbol{F}_T; \boldsymbol{\varphi}^{(n)}) = -\frac{T}{2}\log\det\boldsymbol{\Gamma}_{\boldsymbol{\xi}}^{(n)} - \frac{1}{2}\mathrm{tr}\Big\{(\boldsymbol{\Gamma}_{\boldsymbol{\xi}}^{(n)})^{-1}\sum_{t=1}^{T}(\mathbf{x}_t^{(n)} - \boldsymbol{\Lambda}^{(n)}\mathbf{F}_t)(\mathbf{x}_t^{(n)} - \boldsymbol{\Lambda}^{(n)}\mathbf{F}_t)'\Big\}. \quad (96)$$

Turning, to (**??**), we have

$$\ell(\boldsymbol{F}_T; \boldsymbol{\varphi}^{(n)}) = \ell(\mathbf{F}_0; \boldsymbol{\varphi}^{(n)}) + \sum_{t=1}^{T}\ell(\mathbf{F}_t|\mathbf{F}_{t-1}; \boldsymbol{\varphi}^{(n)}), \quad (97)$$

and since $\boldsymbol{\eta}_t$ is a white noise and Gaussian, then $\mathbf{F}_t|\mathbf{F}_{t-1} \sim N(\mathbf{A}\mathbf{F}_{t-1}, \mathbf{G}\boldsymbol{\Gamma}_{\boldsymbol{\eta}}\mathbf{G}')$. Let us first consider the case $r = q$, so that without loss of generality we can set $\mathbf{G} = \mathbf{I}_r$, then up to an initial condition (which we can assume to be irrelevant for large T), we have

$$\ell(\boldsymbol{F}_T; \boldsymbol{\varphi}^{(n)}) = -\frac{T}{2}\log\det\boldsymbol{\Gamma}_{\boldsymbol{\eta}} - \frac{1}{2}\mathrm{tr}\Big\{\boldsymbol{\Gamma}_{\boldsymbol{\eta}}^{-1}\sum_{t=1}^{T}(\mathbf{F}_t - \mathbf{A}\mathbf{F}_{t-1})(\mathbf{F}_t - \mathbf{A}\mathbf{F}_{t-1})'\Big\}. \quad (98)$$

Therefore, from (96) and (98) the E-step gives

$$\begin{aligned}\mathcal{Q}\left(\boldsymbol{\varphi}^{(n)}; \boldsymbol{\varphi}_j^{(nT)}\right) = &-\frac{T}{2}\log\det\boldsymbol{\Gamma}_{\boldsymbol{\xi}}^{(n)} - \frac{1}{2}\mathrm{tr}\Big\{(\boldsymbol{\Gamma}_{\boldsymbol{\xi}}^{(n)})^{-1}\sum_{t=1}^{T}\mathsf{E}_{\boldsymbol{\varphi}^{(n)}}\big[(\mathbf{x}_t^{(n)} - \boldsymbol{\Lambda}^{(n)}\mathbf{F}_t)(\mathbf{x}_t^{(n)} - \boldsymbol{\Lambda}^{(n)}\mathbf{F}_t)'|\boldsymbol{X}_T^{(n)}\big]\Big\} \\ &-\frac{T}{2}\log\det\boldsymbol{\Gamma}_{\boldsymbol{\eta}} - \frac{1}{2}\mathrm{tr}\Big\{\boldsymbol{\Gamma}_{\boldsymbol{\eta}}^{-1}\sum_{t=1}^{T}\mathsf{E}_{\boldsymbol{\varphi}^{(n)}}\big[(\mathbf{F}_t - \mathbf{A}\mathbf{F}_{t-1})(\mathbf{F}_t - \mathbf{A}\mathbf{F}_{t-1})'|\boldsymbol{X}_T^{(n)}\big]\Big\}.\end{aligned} \quad (99)$$

In the M-step we maximise (99), and in order to do this we need to compute the sufficient statistics:

$$\begin{aligned}&\mathsf{E}_{\boldsymbol{\varphi}_j^{(nT)}}[\mathbf{x}_t^{(n)}\mathbf{F}_t'|\boldsymbol{X}_T] = \mathbf{x}_t^{(n)}\mathbf{F}_{t|T}^{(nT)'}, \\ &\mathsf{E}_{\boldsymbol{\varphi}_j^{(nT)}}[\mathbf{F}_t\mathbf{F}_t'|\boldsymbol{X}_T] = \mathbf{F}_{t|T}^{(nT)}\mathbf{F}_{t|T}^{(nT)'} + \mathbf{P}_{t|T}^{(nT)}, \\ &\mathsf{E}_{\boldsymbol{\varphi}_j^{(nT)}}[\mathbf{F}_t\mathbf{F}_{t-1}'|\boldsymbol{X}_T] = \mathbf{F}_{t|T}^{(nT)}\mathbf{F}_{t-1|T}^{(nT)'} + \mathsf{Cov}[\mathbf{F}_{t|T}^{(nT)}, \mathbf{F}_{t-1|T}^{(nT)}|\boldsymbol{X}_T^{(n)}],\end{aligned} \quad (100)$$

where $\mathbf{F}_{t|T}^{(nT)} \equiv \mathsf{E}_{\boldsymbol{\varphi}_j^{(nT)}}[\mathbf{F}_t|\boldsymbol{X}_T^{(n)}]$ and $\mathbf{P}_{t|T}^{(nT)} \equiv \mathsf{Var}_{\boldsymbol{\varphi}_j^{(nT)}}[\mathbf{F}_t|\boldsymbol{X}_T^{(n)}]$ are computed from the KS. Then, we

have the estimators[16]

$$
\begin{aligned}
\boldsymbol{\Lambda}_{j+1}^{(nT)} &= \left(\sum_{t=1}^{T} \mathsf{E}_{\boldsymbol{\varphi}_j^{(nT)}}[\mathbf{x}_t^{(n)}\mathbf{F}_t'|\boldsymbol{X}_T^{(n)}]\right)\left(\sum_{t=1}^{T} \mathsf{E}_{\boldsymbol{\varphi}_j^{(nT)}}[\mathbf{F}_t\mathbf{F}_t'|\boldsymbol{X}_T^{(n)}]\right)^{-1}, \\
\mathbf{A}_{j+1}^{(nT)} &= \left(\sum_{t=1}^{T} \mathsf{E}_{\boldsymbol{\varphi}_j^{(nT)}}[\mathbf{F}_t\mathbf{F}_{t-1}'|\boldsymbol{X}_T^{(n)}]\right)\left(\sum_{t=1}^{T} \mathsf{E}_{\boldsymbol{\varphi}_j^{(nT)}}[\mathbf{F}_{t-1}\mathbf{F}_{t-1}'|\boldsymbol{X}_T^{(n)}]\right)^{-1}, \\
\boldsymbol{\Gamma}_{\boldsymbol{\xi},j+1}^{(nT)} &= \operatorname{diag}\left\{\frac{1}{T}\sum_{t=1}^{T} \mathsf{E}_{\boldsymbol{\varphi}_j^{(nT)}}\left[(\mathbf{x}_t^{(n)} - \boldsymbol{\Lambda}_{j+1}^{(nT)}\mathbf{F}_t)(\mathbf{x}_t^{(n)} - \boldsymbol{\Lambda}_{j+1}^{(nT)}\mathbf{F}_t)'|\boldsymbol{X}_T^{(n)}\right]\right\} \\
&= \operatorname{diag}\left\{\frac{1}{T}\left(\sum_{t=1}^{T} \mathsf{E}_{\boldsymbol{\varphi}_j^{(nT)}}\left[\mathbf{x}_t^{(n)}\mathbf{x}_t^{(n)'}|\boldsymbol{X}_T^{(n)}\right] - \boldsymbol{\Lambda}_{j+1}^{(nT)}\sum_{t=1}^{T} \mathsf{E}_{\boldsymbol{\varphi}_j^{(nT)}}\left[\mathbf{F}_t\mathbf{x}_t^{(n)'}|\boldsymbol{X}_T^{(n)}\right]\right)\right\}, \\
\boldsymbol{\Gamma}_{\boldsymbol{\eta},j+1}^{(nT)} &= \frac{1}{T}\sum_{t=1}^{T} \mathsf{E}_{\boldsymbol{\varphi}_j^{(nT)}}\left[(\mathbf{F}_t - \mathbf{A}_{j+1}^{(nT)}\mathbf{F}_{t-1})(\mathbf{F}_t - \mathbf{A}_{j+1}^{(nT)}\mathbf{F}_{t-1})'|\boldsymbol{X}_T^{(n)}\right] \\
&= \frac{1}{T}\left(\sum_{t=1}^{T} \mathsf{E}_{\boldsymbol{\varphi}_j^{(nT)}}\left[\mathbf{F}_t\mathbf{F}_t'|\boldsymbol{X}_T^{(n)}\right] - \mathbf{A}_{j+1}^{(nT)}\sum_{t=1}^{T} \mathsf{E}_{\boldsymbol{\varphi}_j^{(nT)}}\left[\mathbf{F}_{t-1}\mathbf{F}_t'|\boldsymbol{X}_T^{(n)}\right]\right). \qquad (101)
\end{aligned}
$$

Notice that these expressions resemble the ordinary least squares solution to the ML estimation for (auto)-regressions with complete data with the difference that the sufficient statistics are replaced by their expectations. In particular, by using (100) in (101) we can compute the estimators $\boldsymbol{\varphi}_{j+1}^{(nT)}$ using the output of the KS run with the parameters $\boldsymbol{\varphi}_j^{(nT)}$. For non-linear problems like factors with conditionally heteroskedasticity, where closed forms of the estimators, like those above, are not available, it is possible to use the Expectation Conditional Maximisation (ECM) algorithm by Meng and Rubin (1993), thus running an additional maximisation with respect to those other parameters after (101) has been computed.

Once, we have $\boldsymbol{\varphi}_{\text{EM}}^{(nT)}$ by running a last time the KF and then the KS and using those parameters, we obtain the KF estimator $\mathbf{F}_{t,\text{KF}}^{(nT)}$ and the ML estimator of the factors, $\mathbf{F}_{t,\text{KS}}^{(nT)}$. Moreover, because we use ML estimators of the parameters, these are $\sqrt{T}$-consistent and by combining this result with consistency of the KF and of the KS for given parameters (see (90) and (92)) we can see that for the case $q = r$ we have:

$$
\begin{aligned}
\mathbf{F}_{t,\text{KF}}^{(nT)} &= \mathbf{F}_t + O_p\left(\frac{1}{\sqrt{T}}\right) + O_p\left(\frac{1}{\sqrt{n}}\right), \\
\mathbf{F}_{t,\text{KS}}^{(nT)} &= \mathbf{F}_{t,\text{KF}}^{(nT)} + \mathbf{P}_{t|t}^{(nT)}\mathbf{A}^{(nT)'}(\mathbf{P}_{t|t-1}^{(nT)})^{-1}(\mathbf{F}_{t+1,\text{KS}}^{(nT)} - \mathbf{F}_{t+1,\text{KF}}^{(nT)}) \\
&= \mathbf{F}_t + O_p\left(\frac{1}{\sqrt{T}}\right) + O_p\left(\frac{1}{\sqrt{n}}\right),
\end{aligned}
$$

Finally, note that because of consistency of the KS, it can be shown that in (100) we have $\mathbf{P}_{t|T}^{(nT)} = O_p(\max(1/\sqrt{T}, 1/n))$ and therefore, as $n, T \to \infty$ the sufficient statistics converge to those we would have if we used the true factors. As a consequence the estimated parameters converge to those we would have if we maximised the true log-likelihood or in other words converge to the ML estimators of the parameters.

When $q < r$ then we cannot invert the conditional covariance matrix of $\mathbf{F}_t$ in (98). This implies that the log-likelihood does not have a maximum and we cannot estimate $\mathbf{G}$ directly in the M-step, however is important to note that such inverse is not needed to estimate all other parameters, thus the first three definitions in (101) remain valid. If we then define $\boldsymbol{\Gamma}_{\boldsymbol{\zeta}} = \mathbf{G}\boldsymbol{\Gamma}_{\boldsymbol{\eta}}\mathbf{G}'$, then the last definition in (101) provides the estimator of $\boldsymbol{\Gamma}_{\boldsymbol{\zeta}}$ in the singular case, defined as the sample covariance of the VAR residuals, denoted

[16] For the covariances recall that in an AR(1), $y_t = ay_{t-1} + e_t$, with $\mathsf{E}[y_{t-1}e_t] = 0$, we have

$$\mathsf{E}[(y_t - ay_{t-1})^2] = \mathsf{E}[y_t^2 + a^2y_{t-1}^2 - 2ay_ty_{t-1}] = \mathsf{E}[y_t^2] - a\mathsf{E}[y_ty_{t-1}],$$

because $\mathsf{E}[a^2y_{t-1}^2] = \mathsf{E}[ay_ty_{t-1}]$, since $\mathsf{E}[y_{t-1}e_t] = 0$.

as $\mathbf{\Gamma}_{\boldsymbol{\zeta},j+1}^{(nT)}$. An estimator $\mathbf{G}_{j+1}^{(nT)}$ can then be obtained as the normalised eigenvectors corresponding to the q-largest eigenvalues of $\mathbf{\Gamma}_{\boldsymbol{\zeta},j+1}^{(nT)}$. However, at convergence of the EM algorithm the estimator obtained in this way, although consistent is not converging to the ML estimator of $\mathbf{G}$, thus the estimated factors obtained from the KS will not be ML estimators.

MAXIMUM LIKELIHOOD ESTIMATION OF APPROXIMATE DYNAMIC FACTOR MODEL. Let us now consider explicitly the and cross and serial dependence in idiosyncratic components. In order to address serial dependence, we treat each serially correlated idiosyncratic component as an additional latent state. In particular let $\mathcal{I} = \{i \in \mathbb{N} : \mathsf{Cov}(\xi_{it}, \xi_{it-k}) \neq 0, \text{ for at least one } k \in \mathbb{Z}\}$. Hence, we allow only for few idiosyncratic components to autocorrelated. Thus, we modify (83)-(84) into

$$\mathbf{x}_t^{(n)} = \mathbf{\Lambda}^{(n)}\mathbf{F}_t + \boldsymbol{\xi}_t^{(n)} + \boldsymbol{\nu}_t^{(n)}, \tag{102}$$

$$\mathbf{F}_t = \mathbf{A}\mathbf{F}_{t-1} + \mathbf{G}\boldsymbol{\eta}_t, \tag{103}$$

$$\xi_{it} = \rho_i \mathbb{I}_{(i\in\mathcal{I})}\xi_{it-1} + e_{it}, \qquad i = 1, \dots, n. \tag{104}$$

Cross-sectional dependence is then introduced across the innovations of (104). Specifically, let $\mathbf{e}_t^{(n)} = (e_{1t} \cdots e_{nt})'$, then we replace A10-D1 with

ASSUMPTION (A10-D2).

a. $\mathbf{e}_t^{(n)} \sim wnN(\mathbf{0}_n, \mathbf{\Gamma}_{\mathbf{e}}^{(n)})$*, such that* $\lambda_{1\mathbf{e}}^{(n)}$ *is bounded for any* $n \in \mathbb{N}$*;*

b. $\boldsymbol{\eta}_t \sim wnN(\mathbf{0}_q, \mathbf{\Gamma}_{\boldsymbol{\eta}})$*;*

c. $\boldsymbol{\nu}_t \sim wnN(\mathbf{0}_n, \nu\mathbf{I}_n)$*, with* $\nu = o\left(\frac{1}{n}\right)$*;*

d. $|\mathcal{I}| = m < \infty$*, for any* $n \in \mathbb{N}$*.*

Note that with condition a. we are replacing condition 7. of weak cross-sectional dependence of idiosyncratic components. If we define $\mathbf{S} = \mathrm{diag}(\rho_1\mathbb{I}_{(1\in\mathcal{I})} \cdots \rho_n\mathbb{I}_{(n\in\mathcal{I})})$, the new parameter vector is given by

$$\boldsymbol{\varphi}^{(n)} = (\mathrm{vec}(\mathbf{\Lambda}^{(n)}), \mathrm{vech}(\mathbf{\Gamma}_{\mathbf{F}}), \mathrm{vech}(\mathbf{\Gamma}_{\mathbf{e}}^{(n)}), \mathrm{vec}(\mathbf{A}), \mathrm{vec}(\mathbf{G}), \mathrm{vech}(\mathbf{\Gamma}_{\boldsymbol{\eta}}), \mathbf{S}).$$

Note that since most of the entries of $\mathbf{S}$ are zero we do not highlight its dependence on n. Letting $\mathbf{H}_t = (\mathbf{F}_t' \; \boldsymbol{\xi}_t^{(n)'})'$ and $\mathbf{B}^{(n)} = (\mathbf{\Lambda}^{(n)} \; \mathbf{I}_n)$, we can also write (102) as

$$\mathbf{x}_t^{(n)} = \mathbf{B}^{(n)}\mathbf{H}_t + \boldsymbol{\nu}_t^{(n)}. \tag{105}$$

Note that although the new state vector has dimension $(r + n)$, most of its components do not evolve in time hence there is no additional complexity in estimating the above model and we do not highlight the dependence of the state vector on n.

Let us now for simplicity assume that $|\mathcal{I}| = 1$ and that $\rho_1 \neq 0$, so that $\xi_{it} = e_{it}$ for $i > 1$. At iteration j of the algorithm, from the KS we also obtain estimates of the additional latent idiosyncratic state $\xi_{1t}^{(nT)} = \mathsf{E}_{\boldsymbol{\varphi}_j^{(nT)}}[\xi_{1t}|\boldsymbol{X}_T^{(n)}]$. In the E-step we now have the expected log-likelihood (see also (99))

$$\begin{aligned}
\mathcal{Q}\left(\boldsymbol{\varphi}^{(n)}; \boldsymbol{\varphi}_j^{(nT)}\right) =& \mathsf{E}_{\boldsymbol{\varphi}_j^{(nT)}}[\ell(\boldsymbol{X}_T^{(n)}, \boldsymbol{H}_T; \boldsymbol{\varphi}^{(n)})|\boldsymbol{X}_T^{(n)}] \qquad (106)\\
=& -\frac{T}{2}\log\det(\nu\mathbf{I}_n) - \frac{1}{2}\mathrm{tr}\left\{\frac{1}{\nu}\sum_{t=1}^{T}\mathsf{E}_{\boldsymbol{\varphi}_j^{(nT)}}\left[(\mathbf{x}_t^{(n)} - \mathbf{B}^{(n)}\mathbf{H}_t^{(n)})(\mathbf{x}_t^{(n)} - \mathbf{B}^{(n)}\mathbf{H}_t^{(n)})'|\boldsymbol{X}_T^{(n)}\right]\right\}\\
&- \frac{T}{2}\log\det\mathbf{\Gamma}_{\boldsymbol{\zeta}} - \frac{1}{2}\mathrm{tr}\left\{\mathbf{\Gamma}_{\boldsymbol{\zeta}}^{-1}\sum_{t=1}^{T}\mathsf{E}_{\boldsymbol{\varphi}_j^{(nT)}}\left[(\mathbf{F}_t - \mathbf{A}\mathbf{F}_{t-1})(\mathbf{F}_t - \mathbf{A}\mathbf{F}_{t-1})'|\boldsymbol{X}_T^{(n)}\right]\right\},\\
&- \frac{T}{2}\log\det\mathbf{\Gamma}_{\mathbf{e}} - \frac{1}{2}\mathrm{tr}\left\{\mathbf{\Gamma}_{\mathbf{e}}^{-1}\sum_{t=1}^{T}\mathsf{E}_{\boldsymbol{\varphi}_j^{(nT)}}\left[(\boldsymbol{\xi}_t^{(n)} - \mathbf{S}\xi_{t-1}^{(n)})(\boldsymbol{\xi}_t^{(n)} - \mathbf{S}\xi_{t-1}^{(n)})'|\boldsymbol{X}_T^{(n)}\right]\right\}.
\end{aligned}$$

In the M-step the estimators of $\mathbf{\Lambda}^{(n)}$, $\mathbf{A}$, $\mathbf{\Gamma}_{\boldsymbol{\zeta}}$ are the same as in (101). Then, by noticing that $\mathbf{S}$ is diagonal and has just one non zero entry which is ρ_1 we have

$$\rho_{1,j+1}^{(nT)} = \left(\sum_{t=1}^{T} \mathsf{E}_{\boldsymbol{\varphi}_j^{(nT)}} [\xi_{1t}\xi_{1t-1} | \boldsymbol{X}_T^{(n)}] \right) \left(\sum_{t=1}^{T} \mathsf{E}_{\boldsymbol{\varphi}_j^{(nT)}} [\xi_{1t-1}^2 | \boldsymbol{X}_T^{(n)}] \right)^{-1}.$$

and consequently we have $\mathbf{S}_{j+1}^{(nT)}$. Finally, an estimator of $\mathbf{\Gamma}_{\mathbf{e}}$ can be found as the sample covariance of the residuals of the state equation

$$\mathbf{\Gamma}_{\mathbf{e},j+1}^{(nT)} = \frac{1}{T} \sum_{t=1}^{T} \mathsf{E}_{\boldsymbol{\varphi}_j^{(nT)}} \left[(\boldsymbol{\xi}_t^{(n)} - \mathbf{S}_{j+1}^{(nT)} \xi_{t-1}^{(n)}) (\boldsymbol{\xi}_t^{(n)} - \mathbf{S} \xi_{t-1}^{(n)})' | \boldsymbol{X}_T^{(n)} \right]. \tag{107}$$

Convergence of these estimators to ML estimators and their $\sqrt{T}$-consistency are straightforward to prove.[17]

[17] Note that while we have element-wise consistency of (107) the only meaningful consistency for the whole matrix is when $\mathbf{\Gamma}_{\mathbf{e},j+1}^{(nT)}$ is rescaled by n.

8 Forecasting

There are two main ways to make forecasts with dynamic factor models

1. direct forecasts, where data at $t+h$ is projected onto the factors at time t;
2. recursive forecasts, where data at $t+h$ is computed using data at time $t+h-1$ and so on.

For the first case it is enough to have an estimate of the factors, so usually PCA is used, while for the second the dynamics of the model has to be specified, so usually the state-space model is used.

DIRECT FORECASTS. Given the static factor model for a sample of length T

$$x_{it} = \boldsymbol{\lambda}_i' \mathbf{F}_t + \xi_{it} = \chi_{it} + \xi_{it}, \qquad i = 1, \dots, n, \quad t = 1, \dots, T,$$

we have that the h-step-ahead forecast of x_{it} is given by

$$x_{iT+h|T} = \chi_{iT+h|T} + \xi_{iT+h|T}, \qquad h > 0. \tag{108}$$

In particular, start by considering the case in which the observed data is predicted only by the factors

$$x_{it+h} = \boldsymbol{\beta}' \mathbf{F}_t + \epsilon_{it+h}, \quad t = 1, \dots, T, \quad h > 0, \tag{109}$$

where we assume $\mathsf{E}[\epsilon_{it} \mathbf{F}_t] = 0$. Therefore, the best linear factor based h-step-ahead prediction of the observed data is given by the linear projection

$$x_{iT+h|T} = \text{Proj}\{x_{iT+h} | \mathbf{F}_T\} = \boldsymbol{\beta}' \mathbf{F}_T, \qquad h > 0. \tag{110}$$

Notice that, since we do not exploit the dynamics of the factors, forecasts depend only on their last observation. Now, since the common and idiosyncratic component are orthogonal at every lead and lag, then $\text{Proj}\{\xi_{iT+h} | \mathbf{F}_T\} = 0$, thus (110) is also the h-step-ahead prediction of the common component

$$\chi_{iT+h|T} = \text{Proj}\{x_{iT+h} | \mathbf{F}_T\} = \text{Proj}\{\chi_{iT+h} | \mathbf{F}_T\} = \boldsymbol{\beta}' \mathbf{F}_T, \qquad h > 0.$$

Now consider again the model (109). The assumption $\mathsf{E}[\epsilon_{it} \mathbf{F}_t] = 0$ is crucial and might not hold, there might of course be omitted variables contained in ϵ_{it}. We have three cases.

1. There are no omitted variables, then we just have to estimate the factors and this is typically done by means of PCA or some variation of this technique.
2. The omitted variables are orthogonal to the factors, then they are by definition idiosyncratic (when $h = 0$ in (109) then $\boldsymbol{\beta} = \boldsymbol{\lambda}_i'$). By modelling the idiosyncratic component of each series separately, we can take these into account. A complete forecast of x_{it} is then given by (108), where a forecast for the idiosyncratic component, $\xi^{(nT)}_{iT+h}$, can be also computed for example by using univariate or sparse autoregressive models (Boivin and Ng, 2005; Bai and Ng, 2008).
3. The omitted variables are correlated with the factors, for example these are lagged values of x_{it} or of $\mathbf{F}_t$. Then, assumption $\mathsf{E}[\epsilon_{it} \mathbf{F}_t] = 0$ in (109) does not hold and we have to consider the more general model
$$x_{it+h} = \boldsymbol{\beta}' \mathbf{F}_t + \boldsymbol{\gamma}' \mathbf{w}_t + \epsilon_{it+h}, \tag{111}$$
where $\mathbf{w}_t$ are the omitted variables. Then,
$$x_{iT+h|T} = \boldsymbol{\beta}' \mathbf{F}_T + \boldsymbol{\gamma}' \mathbf{w}_T = \chi_{iT+h|T} + \boldsymbol{\gamma}' \mathbf{w}_T.$$
The predictor is then computed in two steps, first the factors are extracted, then a linear regression using the estimated factors and the variables $\mathbf{w}_t$ is estimated. This is the general model considered originally by Stock and Watson (2002a). Now since factors are pervasive and idiosyncratic forecasts are negligible, the only omitted variables can indeed be lagged values of $\mathbf{F}_t$, but then by considering

a fully dynamic factor model it is enough to consider a forecast of the common component to obtain a forecast of x_{it}.

Regarding case 2, Luciani (2014) shows that at least in macroeconomic applications, the role of $\xi_{iT+h}^{(nT)}$ forecasts is negligible, i.e. adding them to (108) does not improve the forecasts of x_{it} based only on the factors, $\chi_{iT+h}^{(nT)}$. So we do not consider this case further. Cases 1 and 3 are considered next.

PCA REGRESSION APPROACH. Given an estimator of the factors, $\mathbf{F}_t^{(nT)}$, obtained via PCA, the regression coefficient in (109) is estimated by OLS as

$$\boldsymbol{\beta}^{(nT)} = \left(\sum_{t=1}^{T} \mathbf{F}_t^{(nT)} \mathbf{F}_t^{(nT)'} \right)^{-1} \sum_{t=1}^{T} \mathbf{F}_t^{(nT)} x_{it+h}.$$

Then, since, by definition, the PCA factor estimator is given by $\mathbf{F}_T^{(nT)} = \frac{\boldsymbol{\Lambda}^{(nT)'} \mathbf{x}_T}{n}$, we have

$$\begin{aligned} \chi_{iT+h|T}^{(nT)} &= \boldsymbol{\beta}^{(nT)'} \mathbf{F}_T^{(nT)} = \\ &= \left(\sum_{t=1}^{T} x_{it+h} \mathbf{F}_t^{(nT)'} \right) \left(\sum_{t=1}^{T} \mathbf{F}_t^{(nT)} \mathbf{F}_t^{(nT)'} \right)^{-1} \mathbf{F}_T^{(nT)} \\ &= \left(\sum_{t=1}^{T} x_{it+h} \mathbf{x}_t' \right) \boldsymbol{\Lambda}^{(nT)} \left(\boldsymbol{\Lambda}^{(nT)'} \sum_{t=1}^{T} \mathbf{x}_t \mathbf{x}_t' \boldsymbol{\Lambda}^{(nT)} \right)^{-1} \boldsymbol{\Lambda}^{(nT)'} \mathbf{x}_T \\ &= \left(\sum_{t=1}^{T} x_{it+h} \mathbf{x}_t' \right) \mathbf{P}^{(nT)} \left(\mathbf{P}^{(nT)'} \sum_{t=1}^{T} \mathbf{x}_t \mathbf{x}_t' \mathbf{P}^{(nT)} \right)^{-1} \mathbf{P}^{(nT)'} \mathbf{x}_T, \end{aligned} \tag{112}$$

where in the last step we used the fact that the loadings estimator can be written as

$$\boldsymbol{\Lambda}^{(nT)'} = \sqrt{n} \mathbf{P}^{(nT)'},$$

with $\mathbf{P}^{(nT)}$ being the $n \times r$ matrix of normalised eigenvectors of the sample covariance matrix $\boldsymbol{\Gamma}_{\mathbf{x}}^{(nT)}$ and corresponding to the r largest eigenvalues. From (112), we then have the h-step-ahead predictor of the common component in vector notation

$$\boldsymbol{\chi}_{T+h|T}^{(nT)} = \left[\boldsymbol{\Gamma}_{\mathbf{x}}^{(nT)}(h) \mathbf{P}^{(nT)} \left(\mathbf{P}^{(nT)'} \boldsymbol{\Gamma}_{\mathbf{x}}^{(nT)} \mathbf{P}^{(nT)} \right)^{-1} \right] \left[\mathbf{P}^{(nT)'} \mathbf{x}_T^{(n)} \right]. \tag{113}$$

where $\boldsymbol{\Gamma}_{\mathbf{x}}^{(nT)}(h)$ is the sample h-lags autocovariance matrix. Notice that this predictor contains lagged values of the common component and therefore of the factors therefore it is dynamic but the way in which the factors are estimated, PCA, is not accounting for the factors' dynamics.

Notice that, as $n \to \infty$, we have, by assumption on eigenvalues like *(a')*[18]

$$\left\| \frac{\boldsymbol{\Gamma}_{\mathbf{x}}}{n} - \frac{\boldsymbol{\Gamma}_{\boldsymbol{\chi}}}{n} \right\| = \left\| \frac{\boldsymbol{\Gamma}_{\boldsymbol{\xi}}}{n} \right\| = \frac{\lambda_{1\boldsymbol{\xi}}}{n} \to 0.$$

Using this last result in (113), we can replace the covariance of $\mathbf{x}_t$ with the covariance of $\boldsymbol{\chi}_t$. Therefore, we have a new predictor of the common component which holds for large n:

$$\boldsymbol{\chi}_{T+h|T}^{(nT)} = \left[\boldsymbol{\Gamma}_{\boldsymbol{\chi}}^{(nT)}(h) \mathbf{P}^{(nT)} \left(\mathbf{P}^{(nT)'} \boldsymbol{\Gamma}_{\mathbf{x}}^{(nT)} \mathbf{P}^{(nT)} \right)^{-1} \right] \left[\mathbf{P}^{(nT)'} \mathbf{x}_T^{(n)} \right], \tag{114}$$

[18] An implication of this result is that

$$\left\| \frac{1}{\sqrt{n}} \mathsf{E}[\mathbf{x}_t x_{it}] - \frac{1}{\sqrt{n}} \mathsf{E}[\boldsymbol{\chi}_t \chi_{it}] \right\| \to 0.$$

or

$$\boldsymbol{\chi}_{T+h|T}^{(nT)} = \left[\boldsymbol{\Gamma}_{\boldsymbol{\chi}}^{(nT)}(h)\mathbf{P}^{(nT)}\left(\mathbf{P}^{(nT)'}\boldsymbol{\Gamma}_{\boldsymbol{\chi}}^{(nT)}\mathbf{P}^{(nT)}\right)^{-1}\right]\left[\mathbf{P}^{(nT)'}\mathbf{x}_T^{(n)}\right]. \tag{115}$$

In order to compute (114) and (115), the estimated covariance, $\boldsymbol{\Gamma}_{\boldsymbol{\chi}}^{(nT)}$, and lagged covariance matrices, $\boldsymbol{\Gamma}_{\boldsymbol{\chi}}^{(nT)}(h)$, of the common component must be computed. This can be done using the estimated common component obtained from PCA.

GENERALIZED AND DYNAMIC PCA REGRESSION APPROACH. Forni, Hallin, Lippi, and Reichlin (2005) assume an underlying dynamic factor model for $\mathbf{x}_t^{(n)}$ driven by q dynamic factors. Then, $\boldsymbol{\Gamma}_{\boldsymbol{\chi}}^{(nT)}(h)$ can be estimated using the first q dynamic principal components of the spectral density matrix:

$$\boldsymbol{\Sigma}_{\boldsymbol{\chi}}^{(nT)}(\theta) = \sum_{j=1}^{q}\lambda_{j\boldsymbol{\chi}}^{(nT)}(\theta)\bar{\mathbf{p}}_{j\mathbf{x}}^{(nT)}(\theta)\mathbf{p}_{j\mathbf{x}}^{(nT)'}(\theta), \qquad \boldsymbol{\Gamma}_{\boldsymbol{\chi}}^{(nT)}(h) = \int_{-\pi}^{\pi} e^{ih\theta}\,\boldsymbol{\Sigma}_{\boldsymbol{\chi}}^{(nT)}(\theta)\mathrm{d}\theta, \tag{116}$$

where in practice the integral has to be computed on a grid of values of θ as explained in the previous sections.

Consistently with the assumptions behind the state-space representation of the dynamic factor model, the common component has a spectral density of rank q but a covariance of rank $r > q$. If we now assume that $r = q(s+1)$ for some $s > 1$, then a modification of (114) is also possible where instead of using the r eigenvectors of $\boldsymbol{\Gamma}_{\mathbf{x}}^{(nT)}$ we compute the r eigenvectors of $\boldsymbol{\Gamma}_{\boldsymbol{\chi}}^{(nT)}$ or a generalised version of these eigenvectors which accounts also for cross-sectional heteroschedasticty of the idiosyncratic components. In particular, we define generalised eigenvectors as the solutions of the problem

$$\begin{aligned}
\mathbf{q}_1 &= \arg\max_{\mathbf{u}} \mathbf{u}'\boldsymbol{\Gamma}_{\boldsymbol{\chi}}^{(n)}\mathbf{u} \text{ s.t. } \mathbf{u}'\boldsymbol{\Gamma}_{\boldsymbol{\xi}}^{(n)}\mathbf{u} = 1,\\
\mathbf{q}_j &= \arg\max_{\mathbf{u}} \mathbf{u}'\boldsymbol{\Gamma}_{\boldsymbol{\chi}}^{(n)}\mathbf{u} \text{ s.t. } \mathbf{u}'\boldsymbol{\Gamma}_{\boldsymbol{\xi}}^{(n)}\mathbf{u} = 1 \text{ and } \mathbf{u}'\mathbf{q}_1 = \ldots = \mathbf{u}'\mathbf{q}_{j-1} = 0,
\end{aligned}$$

the eigenvalues being the value of the functions at their maximum. This problem is equivalent to find eigenvector and eigenvalues of the matrix $(\boldsymbol{\Gamma}_{\boldsymbol{\xi}}^{(n)})^{-1}\boldsymbol{\Gamma}_{\boldsymbol{\chi}}^{(n)}$. Given the estimators $\boldsymbol{\Gamma}_{\boldsymbol{\chi}}^{(nT)}$ (see (116)) and $\boldsymbol{\Gamma}_{\boldsymbol{\xi}}^{(nT)} = (\boldsymbol{\Gamma}_{\mathbf{x}}^{(nT)} - \boldsymbol{\Gamma}_{\boldsymbol{\chi}}^{(nT)})$, we collect the normalised generalised eigenvectors corresponding to the r largest generalised eigenvalues in a $n \times r$ matrix $\mathbf{Q}^{(nT)}$, then we have a new loadings estimator defined as $\mathbf{A}^{(nT)} = \sqrt{n}\mathbf{Q}^{(nT)}$ and the space spanned by the factors is then recovered in the same way as in PCA, i.e. as

$$\mathbf{F}_t^{(nT)} = \frac{\mathbf{Q}^{(nT)'}\mathbf{x}_t}{\sqrt{n}} = \frac{\mathbf{A}^{(nT)'}\mathbf{x}_t}{n}, \quad t = 1, \ldots, T.$$

Since we have also in this case a linear combination of the data such that the normalization conditions are satisfied it is straightforward to prove consistency also of this estimator. Notice that this is not the PCA estimator but a different one.

By using $\mathbf{Q}^{(nT)}$ instead of $\mathbf{P}^{(nT)}$ in (114) and (115), we have the predictor of the common component proposed by Forni, Hallin, Lippi, and Reichlin (2005), is given by

$$\boldsymbol{\chi}_{T+h|T}^{(nT)} = \left[\boldsymbol{\Gamma}_{\boldsymbol{\chi}}^{(nT)}(h)\mathbf{Q}^{(nT)}\left(\mathbf{Q}^{(nT)'}\boldsymbol{\Gamma}_{\mathbf{x}}^{(nT)}\mathbf{Q}^{(nT)}\right)^{-1}\right]\left[\mathbf{Q}^{(nT)'}\mathbf{x}_T^{(n)}\right], \tag{117}$$

or

$$\boldsymbol{\chi}_{T+h|T}^{(nT)} = \left[\boldsymbol{\Gamma}_{\boldsymbol{\chi}}^{(nT)}(h)\mathbf{Q}^{(nT)}\left(\mathbf{Q}^{(nT)'}\boldsymbol{\Gamma}_{\boldsymbol{\chi}}^{(nT)}\mathbf{Q}^{(nT)}\right)^{-1}\right]\left[\mathbf{Q}^{(nT)'}\mathbf{x}_T^{(n)}\right]. \tag{118}$$

This latter estimator in practice poses problems when computing the generalised eigenvectors $\mathbf{Q}^{(nT)}$. Indeed, when n is large inversion of the idiosyncratic covariance matrix, $\boldsymbol{\Gamma}_{\boldsymbol{\xi}}^{(nT)}$, is in general not possible (it has by construction rank $n - q$). Regularised estimators for this matrix or of its inverse are necessary (see e.g. Fan, Liao, and Mincheva, 2013 for $\boldsymbol{\Gamma}_{\boldsymbol{\xi}}^{(nT)}$ and Barigozzi and Brownlees, 2018 for its inverse). The simplest case is when $\boldsymbol{\Gamma}_{\boldsymbol{\xi}}^{(nT)}$ is diagonal, i.e. the most sparse estimator.

Forni, Hallin, Lippi, and Reichlin (2005) prove consistency for (117), as $n, T \to \infty$,

$$\chi_{iT+h|T}^{(nT)} \xrightarrow{p} \chi_{iT+h}, \quad i = 1, \dots, n, \quad h > 0.$$

Similar consistency results can be proved for (113), (114), (115), and (118).

STATE SPACE APPROACH. We know that under some additional assumptions we can write the approximate dynamic factor model in state space form, that is a static factor model where we also specify the dynamics of the factor as in (70) and (74). The forecast of the common component is given by the recursive equation[19]

$$\boldsymbol{\chi}_{T+h|T}^{(nT)} = \boldsymbol{\Lambda}^{(nT)} \big(\mathbf{A}^{(nT)}\big)^h \mathbf{F}_T^{(nT)}, \qquad h > 0,$$

where we can use estimates of the state space form of the model obtained either by PCA and VAR or by means of the EM algorithm. From this equation it is clear that for forecasting the fact that in the original model the r factors $\mathbf{F}_t$ are driven by $q < r$ shocks is irrelevant. This is true in everything that follows except when we consider spectral densities. Consistency of this predictor is then an immediate consequence of consistency of the estimated parameters.

MISSING VALUES AND MIXED FREQUENCIES. Model (70) and (74) when estimated by means of the EM algorithm can be used for real-time applications. The use of real-time data presents two main problems:

1. how to manage missing values, in particular at the end of the sample due to the delay in macroeconomic data releases which come from different sources;
2. how to handle mixed frequencies, which most of the times reduces to how to bridge monthly and quarterly indicators, eq. GDP is released only quarterly while other indicators as Industrial Production are released monthly.

In the case of missing values the E-step of the EM algorithm can be modified and there are two possibilities. At iteration j a new panel $\tilde{\mathbf{x}}_{t,j}^{(n)}$, is built as follows:

a. Stock and Watson (2002b) modify the observed data $\tilde{x}_{it,j} = x_{it}$ if observed or $\tilde{x}_{it,j} = \boldsymbol{\lambda}_i'(j-1)\mathbf{F}_{t|T,j-1}$ otherwise, where $\boldsymbol{\lambda}_i(j-1)$ and $\mathbf{F}_{t|T,j-1}$ are the estimated loadings and factors at iteration $j-1$;

b. Bańbura and Modugno (2014) impose the restrictions on the variance of the idiosyncratic component: $(\tilde{\boldsymbol{\Gamma}}_{\boldsymbol{\xi}})_{ii} = (\boldsymbol{\Gamma}_{\boldsymbol{\xi}})_{ii}$ if x_{it} observed or $(\tilde{\boldsymbol{\Gamma}}_{\boldsymbol{\xi}})_{ii} = \infty$ otherwise.

Missing frequencies are instead included in the model by imposing restrictions on the loadings. In particular, consider the case in which we have monthly and quarterly series. If Y_t denotes the log-level of a variable at month t then $y_t^M = Y_t - Y_{t-1}$ is the month on month growth rate. However, assume that the variable is recorded only at quarterly frequency then what we observe is the variable Y_t^Q and if t indicates months then this variable is available only at $t, t+3, t+6 \dots$. Its growth rate is $y_t^Q = Y_t - Y_{t-3}$. Mariano and Murasawa (2003) propose the approximation

$$Y_t^Q \simeq Y_t + Y_{t-1} + Y_{t-2},$$

which for the growth rates implies

$$y_t^Q = y_t^M + 2y_{t-1}^M + 3y_{t-2}^M + 2y_{t-3}^M + y_{t-4}^M.$$

The quarter on quarter growth rate that we want to use in the database is a weighted average of a month on month growth rate. If in a given dataset we have both monthly and quarterly growth rates denoted as $\mathbf{x}_t^M$

[19] Compare this with the AR(1) forecast where

$$y_{T+h|T} = a^h y_T, \qquad h \geq 0.$$

and $\mathbf{x}_t^Q$ respectively, then we have either two separate factor models

$$\begin{aligned}\mathbf{x}_t^M &= \mathbf{\Lambda}^M \mathbf{F}_t^M + \boldsymbol{\xi}_t^M,\\ \mathbf{x}_{3t+1}^Q &= \mathbf{\Lambda}^Q \mathbf{F}_{3t+1}^Q + \boldsymbol{\xi}_{3t+1}^Q,\end{aligned}$$

or a unique model

$$\begin{pmatrix} \mathbf{x}_t^M \\ \mathbf{x}_t^Q \end{pmatrix} = \begin{pmatrix} \mathbf{\Lambda}^M & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{\Lambda}^Q & 2\mathbf{\Lambda}^Q & 3\mathbf{\Lambda}^Q & 2\mathbf{\Lambda}^Q & \mathbf{\Lambda}^Q \end{pmatrix} \begin{pmatrix} \mathbf{F}_t^M \\ \mathbf{F}_{t-1}^M \\ \mathbf{F}_{t-2}^M \\ \mathbf{F}_{t-3}^M \\ \mathbf{F}_{t-4}^M \end{pmatrix} + \begin{pmatrix} \boldsymbol{\xi}_t^M \\ \boldsymbol{\xi}_t^Q \end{pmatrix}.$$

In a first iteration, the estimated factor $\mathbf{F}_{t|T,1}^M$ can be extracted from the dataset of monthly data $\mathbf{x}_t^M$ only using KF and KS and then $\mathbf{\Lambda}^M$ can be easily estimated. Then, in successive iterations (for $j > 1$), we can build a new dataset $\tilde{\mathbf{x}}_{t,j}^{(n)}$ containing both quarterly and monthly data where the quarterly data at iteration j are computed as

$$\begin{aligned}\tilde{x}_{it,j} &= \boldsymbol{\lambda}_i'(j-1)\left(\mathbf{F}_{t|T,j-1}^M + 2\mathbf{F}_{t-1|T,j-1}^M + 3\mathbf{F}_{t-2|T,j-1}^M + 2\mathbf{F}_{t-3|T,j-1}^M + \mathbf{F}_{t-4|T,j-1}^M\right) =\\ &= \boldsymbol{\lambda}_i'(j-1)(1+L+L^2)^2\mathbf{F}_{t|T,j-1}^M.\end{aligned}$$

This model is used by Stock and Watson (2002b) to predict Industrial Production and CPI inflation and by Bernanke and Boivin (2003) to forecast also the unemployment rate. A similar approach not described here is proposed by Bańbura and Modugno (2014). Two other applications of these techniques follow.

EUROCOIN. Altissimo, Cristadoro, Forni, Lippi, and Veronese (2010) developed a business cycle indicators for the Euro Area known as Eurocoin which is computed each month by the staff of the Banca d'Italia, and it is published also by the Centre for Economic Policy Research. This indicator was constructed in order to have an assessment of economic activity that is free from short-run fluctuations. With the term "economic activity" it is meant that the goal is to capture only those fluctuations due to the structural/macroeconomic shocks $\mathbf{u}_t$, i.e. the business cycle. With the expression "free from short-run fluctuations" it is meant that the goal is not to estimate all the fluctuations induced by the common shocks, but only those producing medium-run and long-run fluctuations (i.e. larger then one year).

Let $\mathbf{x}_t^{(n)}$ be a large vector of month on month growth rates, and let y_t^Q be quarter on quarter GDP growth rate, it is assumed that $(\mathbf{x}_t^{(n)}, y_t^Q)$ have a joint approximate factor representation:

$$\begin{aligned}\mathbf{x}_t^{(n)} &= \mathbf{\Lambda}^{(n)}\mathbf{F}_t + \boldsymbol{\xi}_t^{(n)},\\ y_t^Q &= \boldsymbol{\lambda}_y'\mathbf{F}_t^Q + \xi_{yt}.\end{aligned}$$

where $\mathbf{F}_t^Q = (1+L+L^2)^2\mathbf{F}_t$. Let $w(L)$ be a band-pass filter such that $c_t^Q = w(L)y_t^Q$ is the medium-run to long-run GDP growth, i.e. related to fluctuations large than 1 year thus keeping frequencies $\theta \in [-\pi/6, \pi/6]$. Then Eurocoin, denoted as the estimator $e_t^{(nT)}$, is the linear projection of c_t^Q onto the space spanned by the quarterly smoothed factors $\mathbf{F}_t^Q$. Formally

$$e_t^{(nT)} = \text{Proj}\left\{c_t^Q | \mathcal{F}_t^Q\right\}$$

where $\mathcal{F}_t^Q = \overline{\text{span}}\{\mathbf{F}_{t-h}^Q, h \geq 0\}$. Then, following the same reasoning as for the predictor of the common component we have

$$e_t^{(nT)} = \bar{y}_t^Q + \mathbf{\Gamma}_{c\mathbf{F}^Q}^{(nT)}\left(\mathbf{\Gamma}_{\mathbf{F}^Q}^{(nT)}\right)^{-1}\mathbf{F}_t^Q$$

where $\mathbf{\Gamma}_{c\mathbf{F}^Q}^{(nT)}$ is an estimator of the covariance between c_t^Q and the factors $\mathbf{F}_t^Q$, and $\bar{y}_t^Q$ is the sample average of y_t^Q. In order to estimate the quarterly factors from monthly data, we can use the KF approach described above thus computing also the common component of y_t^Q which once filtered would give an estimator of

Eurocoin.

Alternatively, Altissimo, Bassanetti, Cristadoro, Forni, Hallin, Lippi, Reichlin, and Veronese (2001) propose the use of generalised PCA. Denote by $\boldsymbol{\phi}_t^{(n)}$ the filtered common component keeping only medium-run and long-run frequencies. Then, an estimator of the factors is obtained using the eigenvectors corresponding to the r-largest eigenvalues of the matrix $(\boldsymbol{\Gamma}_{\boldsymbol{\chi}}^{(n)} + \boldsymbol{\Gamma}_{\boldsymbol{\xi}}^{(n)})^{-1}\boldsymbol{\Gamma}_{\boldsymbol{\phi}}^{(n)}$. All matrices can be computed from spectral density of the common component $\boldsymbol{\Sigma}_{\boldsymbol{\chi}}^{(n)}(\theta)$ as

$$\boldsymbol{\Gamma}_{\boldsymbol{\chi}}^{(n)} = \int_{-\pi}^{\pi} \boldsymbol{\Sigma}_{\boldsymbol{\chi}}^{(n)}(\theta)\mathrm{d}\theta, \qquad \boldsymbol{\Gamma}_{\boldsymbol{\phi}}^{(n)} = \int_{-\pi/6}^{\pi/6} \boldsymbol{\Sigma}_{\boldsymbol{\chi}}^{(n)}(\theta)\mathrm{d}\theta, \qquad \boldsymbol{\Gamma}_{\boldsymbol{\xi}}^{(n)} = \boldsymbol{\Gamma}_{\mathbf{x}}^{(n)} - \boldsymbol{\Gamma}_{\boldsymbol{\chi}}^{(n)}.$$

By replacing the spectral density with its estimates obtained by means of the q largest dynamic PCA, we can compute the generalized eigenvectors and we have the factor estimates denoted as $\mathbf{F}_t^{Q(nT)}$. Differently from the previous approach, in this case, the difference between r and q is crucial and must be considered. Then $\boldsymbol{\Gamma}_{\mathbf{F}^Q}^{(nT)}$ is estimated using the sample covariance of the estimated factors. Finally, the covariance $\boldsymbol{\Gamma}_{c\mathbf{F}^Q}^{(n)}$ is given by

$$\boldsymbol{\Gamma}_{c\mathbf{F}^Q}^{(n)} = \int_{-\pi/6}^{\pi/6} \boldsymbol{\Sigma}_{y\mathbf{f}^Q}^{(n)}(\theta)\mathrm{d}\theta,$$

where $\boldsymbol{\Sigma}_{y\mathbf{f}^Q}^{(n)}(\theta)$ is the cross-spectrum between y_t^Q and the factors $\mathbf{F}_t^Q$ and is given by

$$\boldsymbol{\Sigma}_{y\mathbf{F}^Q}^{(n)}(\theta) = \frac{1}{2\pi}\sum_{k=-\infty}^{\infty} e^{-ik\theta}\boldsymbol{\Gamma}_{y\mathbf{F}^Q}^{(n)}(k), \quad \theta \in [-\pi, \pi].$$

with $\mathsf{E}[y_t^Q \mathbf{F}_{t-k}^Q]$. By replacing, the factors with their estimates obtained above with generalised PCA and estimating cross-covariances and cross-spectra as usual we have an estimator of the covariance between the factors and c_t^Q.

In an analogous way, Cristadoro, Forni, Reichlin, and Veronese (2005) build a core euro inflation indicator as the projection of HICP on the quarterly factors.

NOWCASTING. Eurocoin is computed without using the state space representation, that is with a purely dynamic approach based on spectral analysis. On the other hand, Giannone, Reichlin, and Small (2008) use the state-space representation for computing the quarter on quarter growth rate of GDP. Namely, (notice that as mentioned above here we do not make any distinction between r and q)

$$\begin{aligned} \mathbf{x}_t^{(n)} &= \boldsymbol{\Lambda}^{(n)}\mathbf{F}_t + \boldsymbol{\xi}_t^{(n)}, \\ \mathbf{F}_t &= \mathbf{A}\mathbf{F}_{t-1} + \boldsymbol{\epsilon}_t, \\ y_t^Q &= c + \boldsymbol{\beta}\mathbf{F}_t + \eta_t. \end{aligned} \tag{119}$$

The factors and loadings can be estimated with the EM algorithm as explained above, while the last equation is estimated in a second step using OLS.

While GDP is released at the end of a quarter (and then revised in later quarters), Central Banks need to set policies in real-time, thus often without knowing the present status of the economy. Moreover, also other variables needed for policy making are released with different timing and their values are often revised later thus also the recent past might be unknown. Predicting the current and past state of the economy (as current quarter GDP) is called nowcasting. The model above given an estimates of the factors up to any given time t delivers, via (119), the nowcast of quarter on quarter growth rate of GDP. Direct forecasts of the same variables can also be computed as in Stock and Watson (2002a).

9 Structural approximate dynamic factor models in macroeconomics

FACTOR MODELS AND STRUCTURAL VAR. The state-space representation of the approximate dynamic factor model is particularly useful also for studying impulse response functions (IRF) of observed variables to unexpected shocks to the economy. The classical tool for studying IRF are Structural VAR (SVAR) models. These are employed as tools for validating more complex economic models known as Dynamic Stochastic General Equilibrium (DSGE) models. In particular, the theoretical IRF implied by DSGE can be compared with those estimated on observed data, thus validating or not the underlying economic theory.

However, SVAR suffer of two main problems:

1. a VAR with n variables has a number of parameters that grows with n^2, so if n is large and T is not large enough we encounter the curse of dimensionality problem;
2. there is the possibility that the space spanned by VAR innovations might not coincide with the space spanned by structural shocks, this problem is called nonfundamentalness and is discussed in Lippi and Reichlin (1993, 1994) and Alessi, Barigozzi, and Capasso (2011) and also below.

The availability of large datasets, combined with factor models, helps in solving both problems. In particular, the curse of dimensionality has already been discussed. Here we focus on nonfundamentalness and we show that factor models do not suffer of this problem. We then discuss identification of IRF using the the approximate dynamic factor model (see Section 7 for estimation).

INNOVATIONS AND STRUCTURAL SHOCKS. DSGE models are often used in academic research and Central Banks for modelling and analysing national and worldwide economies. These are analytical models based on finite difference equations with stochastic disturbances. They are used for aggregating the unobserved behaviour of single agents (e.g. consumers or firms) into observed macroeconomic time series. Let us call $\mathbf{y}$ a large, possibly infinite, dimensional vector process containing all available economic information. If we assume that this process is purely non-deterministic, then although unobservable, it has the Wold representation

$$\mathbf{y}_t = \boldsymbol{\epsilon}_t^y + \sum_{k=1}^{\infty} \boldsymbol{\Psi}_k \boldsymbol{\epsilon}_{t-k}^y, \quad \boldsymbol{\epsilon}_t^y \sim \text{w.n.}(\mathbf{0}, \boldsymbol{\Gamma}_{\boldsymbol{\epsilon}^y}), \quad t \in \mathbb{Z}, \tag{120}$$

and by definition $\mathcal{H}_t^{\boldsymbol{\epsilon}} = \mathcal{H}_t^{\mathbf{y}}$.

Moreover, DSGE models can be reduced to VARMA models by log-linearization of the equations around their steady-state. If $\mathbf{y}$ has rational spectral density, we say that it is a VARMA(p_1, p_2) process if it is a stationary solution of

$$\boldsymbol{\Phi}(L)\mathbf{y}_t = \boldsymbol{\Theta}(L)\mathbf{u}_t, \quad \mathbf{u}_t \sim \text{w.n.}(\mathbf{0}, \mathbf{I}), \quad t \in \mathbb{Z}. \tag{121}$$

The AR and MA filters are respectively defined as $\boldsymbol{\Phi}(L) = \mathbf{I} - \sum_{k=1}^{p_1} \boldsymbol{\Phi}_k L^k$ and $\boldsymbol{\Theta}(L) = \mathbf{I} + \sum_{k=1}^{p_2} \boldsymbol{\Theta}_k L^k$, where the roots of $\det(\boldsymbol{\Phi}(z))$ must be outside the unit disc. Therefore, there exists two white noise process $\boldsymbol{\epsilon}^y$ and $\mathbf{u}$ associated with $\mathbf{y}$.

When $\mathbf{u}$ is orthonormal (it can always be transformed to be like that), econometricians interpret it as a vector of unexpected, uncorrelated, shocks affecting the economy as a whole and since $\mathbf{y}$ contains all available information, then it is reasonable to assume $\mathcal{H}_t^{\mathbf{u}} = \mathcal{H}_t^{\boldsymbol{\epsilon}^y} = \mathcal{H}_t^{\mathbf{y}}$, for any $t \in \mathbb{Z}$. In other words, an economy-wide VARMA representation of the DSGE model is driven by shocks that belong only to the past of $\mathbf{y}_t$. In order to emphasize their economic meaning, we call $\mathbf{u}$ *structural shocks*. Typically, these shocks are very few and their nature can be, for example, technological (e.g. a new technical innovation), or monetary (e.g. a change in the interest rate), or fiscal (e.g. an increase in taxes or in government spending), or financial (e.g. a crash in the stock market). Notice that under the additional assumption of Gaussianity of $\mathbf{u}$ the shocks are also independent. This assumption is made throughout the rest of the section.

NONFUNDAMENTALNESS. However, econometricians observe only a finite and often small amount of information, i.e. they observe only a subset $\mathbf{x}^{(n)}$ of the economy wide process $\mathbf{y}$. Clearly, the structural shocks $\mathbf{u}$, affecting the whole economy $\mathbf{y}$, affect also a subset $\mathbf{x}^{(n)}$ of it. However, there is evidence of

many economic examples for which the space generated by $\mathbf{u}$, which coincides with the space generated by the innovations of $\mathbf{y}$, is larger than the space generated by the innovations of $\mathbf{x}^{(n)}$, that is:

$$\begin{array}{ccccc} \mathcal{H}_t^{\mathbf{x}^{(n)}} & \subset & \mathcal{H}_t^{\mathbf{u}} & = & \mathcal{H}_t^{\mathbf{y}} \\ \| & & & & \| \\ \mathcal{H}_t^{\boldsymbol{\epsilon}^{x(n)}} & \subset & \mathcal{H}_t^{\mathbf{u}} & = & \mathcal{H}_t^{\boldsymbol{\epsilon}^{y}} \end{array}, \qquad t \in \mathbb{Z}. \tag{122}$$

In particular, we have the definition.

DEFINITION. *A white noise process* $\mathbf{u}$ *is* $\mathbf{x}^{(n)}$*-fundamental if* $\mathcal{H}_t^{\mathbf{x}^{(n)}} = \mathcal{H}_t^{\mathbf{u}}$*, for any* $t \in \mathbb{Z}$*.*

Then, if we have a situation as in (122), we say that $\mathbf{u}$ is $\mathbf{x}^{(n)}$-nonfundamental, or that $\mathbf{x}^{(n)}$ is a nonfundamental VARMA process which implies that at any $t \in \mathbb{Z}$, $\mathbf{u}_t$ belongs to the space spanned by also future values of $\mathbf{x}_t^{(n)}$. How can we then retrieve the space spanned by $\mathbf{u}_t$ using only present and past values of the observables? What is the meaning of nonfundamental shocks? From (122) we see that $\mathbf{u}$ is $\mathbf{x}^{(n)}$-nonfundamental because the econometricians' information space $\mathcal{H}_t^{\mathbf{x}^{(n)}}$ is smaller than the information space $\mathcal{H}_t^{\mathbf{y}}$ of an agent able to observe the whole economy $\mathbf{y}$. In this sense nonfundamentalness is a problem related to information.

The traditional way of retrieving the space generated by the structural shocks $\mathcal{H}_t^{\mathbf{u}}$ consists in assuming finite lag, causal, VAR approximations of a VARMA for a process $\mathbf{x}^{(n)}$. By definition of causal VAR, the associated white noise, say $\mathbf{v}$, is always $\mathbf{x}^{(n)}$-fundamental. Thus, a VAR approximation is valid only as long as the structural shocks $\mathbf{u}$ that we want to recover are $\mathbf{x}^{(n)}$-fundamental, and in this case $\mathcal{H}_t^{\mathbf{v}} = \mathcal{H}_t^{\mathbf{u}}$, for any $t \in \mathbb{Z}$. If instead economic theory tells us that $\mathbf{u}$ is $\mathbf{x}^{(n)}$-nonfundamental, then VAR are not the appropriate model to use since in this case $\mathcal{H}_t^{\mathbf{v}} \subset \mathcal{H}_t^{\mathbf{u}}$, for any $t \in \mathbb{Z}$.

Two solutions are possible to solve the problem of nonfundamentalness. Either we apply to $\mathbf{v}_t$ *ad hoc* filters called Blaschke matrices in order to obtain a new white noise process $\widetilde{\mathbf{v}}$ such that $\mathcal{H}_t^{\mathbf{v}} \subset \mathcal{H}_t^{\tilde{\mathbf{v}}} = \mathcal{H}_t^{\mathbf{u}}$, for any $t \in \mathbb{Z}$. Or we add new information to $\mathbf{x}_t^{(n)}$ in order to have a larger panel $\mathbf{x}_t^{(m)}$ with $m > n$ (i.e. we enlarge the econometricians' information space) in such a way that $\mathcal{H}_t^{\mathbf{x}^{(n)}} \subset \mathcal{H}_t^{\mathbf{x}^{(m)}} = \mathcal{H}_t^{\mathbf{y}}$, for any $t \in \mathbb{Z}$. An example of the first way is in Forni, Gambetti, Lippi, and Sala (2014) and we do not consider it here. The second way is mainly based on factor models, which are apt to deal with the case $n \to \infty$, thus mimicking the economy-wide process $\mathbf{y}$. In particular, we show that if $\mathbf{x}^{(n)}$ has a factor structure, then for the common component $\boldsymbol{\chi}^{(n)}$ we have $\mathcal{H}_t^{\boldsymbol{\chi}} = \mathcal{H}_t^{\mathbf{u}}$, where $\mathbf{u}$ are the q common shocks driving $\boldsymbol{\chi}^{(n)}$.

CONDITIONS FOR FUNDAMENTALNESS. Let us consider an n-dimensional MA covariance stationary process $\mathbf{x}_t^{(n)}$ obtained from a given DSGE

$$\mathbf{x}_t^{(n)} = \mathbf{C}^{(n)}(L)\mathbf{u}_t, \quad \mathbf{u}_t \sim \text{w.n.}(\mathbf{0}, \mathbf{I}), \quad t \in \mathbb{Z}, \tag{123}$$

where $\mathbf{u}_t$ is a q-dimensional vector white noise process of structural shocks with q not necessarily equal to n; $\mathbf{C}^{(n)}(L)$ is an $n \times q$ matrix of rational, one-sided, and square-summable filters. In principle $\mathbf{C}^{(n)}(L)$ can be of infinite order thus including also an AR component, which we always assume to be causal.

When $n = q$, $\mathbf{u}$ is $\mathbf{x}^{(n)}$-fundamental if and only if:

a) $\mathbf{u}_t$ is a weak white noise vector;

b) $\mathbf{C}^{(n)}(z)$ has no poles inside the unit circle;

c) $\det \mathbf{C}^{(n)}(z)$ has no roots lying inside the unit circle:

$$\det \mathbf{C}^{(n)}(z) \neq 0 \quad \forall\, z \in \mathbb{C} \quad \text{s.t. } |z| < 1\,.$$

If, on the other hand, $n > q$, $\mathbf{u}$ is $\mathbf{x}^{(n)}$-fundamental if and only if:

a) $\mathbf{u}_t$ is a weak white noise vector;

b) $\mathbf{C}(z)^{(n)}$ has no poles inside the unit circle;

c) $\mathbf{C}(z)^{(n)}$ has full rank inside the unit circle

$$\text{rank } \mathbf{C}^{(n)}(z) = q \quad \forall\, z \in \mathbb{C} \quad \text{s.t. } |z| < 1\,.$$

Intuitively, singularity makes nonfundamentalness a non-generic problem. Consider the case $q = 1$. If $n > 1$ we have n submatrices $\mathbf{C}_j^{(n)}(z)$ and the representation is nonfundamental if their determinants have a common root smaller than one in modulus. Nonfundamentalness is non generic because to have a common root we must satisfy $\binom{n}{q} - 1$ equality constraints. Intuitively, when n is large, it is highly improbable to have a common root for all submatrices $\mathbf{C}_j^{(n)}(L)$, as the number of restrictions we should satisfy increases with n.

We say that $\mathbf{C}^{(n)}(z)$ is left-invertible if it exists a $n \times q$ full-rank matrix $\mathbf{D}^{(n)}(z)$ of rational complex functions such that $\mathbf{D}^{(n)}(z)'\mathbf{C}^{(n)}(z) = \mathbf{I}_q$. We have three possibilities:

a) $\mathbf{D}^{(n)}(z)$ exists and its Laurent series expansion is composed only of positive powers of z, thus present and past values of $\mathbf{x}_t$ are enough to recover $\mathbf{u}_t$: we have fundamentalness and invertibility or, better, invertibility in the past;

b) $\mathbf{D}^{(n)}(z)$ exists and its Laurent series expansion contains also negative powers of z, thus we need also future values of $\mathbf{x}_t$ to recover $\mathbf{u}_t$: we have nonfundamentalness and noninvertibility or, better, invertibility but in the future;

c) $\mathbf{D}^{(n)}(z)$ does not exist and conditions for fundamentalness still hold, we do not have invertibility either in the past or in the future. For example, in squared systems this happens when $\det \mathbf{C}^{(n)}(z)$ has at least one root on the unit circle and all other roots outside.

For simplicity, being the latter a borderline case, we do not deal with it here and in what follows we speak about roots "outside the unit circle", meaning "outside and, possibly, on the unit circle".

As an example consider the permanent income Friedman-Muth model. Income y_t is decomposed in a permanent part y_{1t} and a transitory part y_{0t} which are independently affected by uncorrelated shocks

$$(1 - L)y_{1t} = u_{1t}, \qquad y_{0t} = u_{0t}.$$

If consumption c_t follows the permanent income hypothesis, we have: $(1 - L)c_t = u_{1t} + (1 - \beta)u_{0t}$ where $\beta \in (0, 1)$ is the agent discount factor. Therefore, we have

$$\mathbf{x}_t = \begin{pmatrix} (1 - L)y_t \\ (1 - L)c_t \end{pmatrix} = \begin{pmatrix} 1 & 1 - L \\ 1 & 1 - \beta \end{pmatrix} \begin{pmatrix} u_{1t} \\ u_{0t} \end{pmatrix} = \mathbf{C}(L)\mathbf{u}_t.$$

In this case $\det \mathbf{C}(z) = (z - \beta)$ and hence it has the only root in $z = \beta$, which by definition lies inside the unit circle. The above econometric model representing the permanent income model is nonfundamental. The space spanned by permanent and transitory shocks to income is not recoverable by considering only present and past values of income and consumption.

Then let us consider the "toy" example with $n = q = 2$

$$x_{1t} = (1 - aL)u_{1t}, \quad x_{2t} = (1 - bL)u_{2t},$$

thus we need both $|a| < 1$ and $|b| < 1$ in order to have fundamentalness.[20] Consider instead the case $n = 2$ and $q = 1$:

$$x_{1t} = (1 - aL)u_t, \quad x_{2t} = (1 - bL)u_t.$$

[20] Indeed, we have

$$u_{1t} = x_{1t} + \sum_{k=1}^{\infty} (-1)^{k-1} a^k x_{1t-k} \tag{124}$$

and the series converges only if $|a| < 1$.

Then u_t is (x_1, x_2)-fundamental for any real numbers a and b provided that $a \neq b$, indeed we can write[21]

$$u_t = \frac{bx_{1t} - ax_{2t}}{b - a}.$$

In this sense in the case $n > q$ it is easier to have fundamentalness as a less restrictive condition on the parameters has to be satisfied.

FUNDAMENTALNESS IN FACTOR MODELS - INTUITION. From Theorem 2, we know that as $n \to \infty$ the correct representation of $\mathbf{x}^{(n)}$ is not (123) but (see also (47))

$$\mathbf{x}_t^{(n)} = \mathbf{B}^{(n)}(L)\mathbf{u}_t + \boldsymbol{\xi}_t^{(n)}, \quad t \in \mathbb{Z}, \tag{125}$$

where $\mathbf{u}$ is an orthonormal q-dimensional vector white noise process with $q < n$ and $\mathbf{B}^{(n)}(L)$ is an $n \times q$ matrix of rational, one-sided, and square-summable filters. Moreover, from Theorem 6, we have the state-space representation (see also (70)-(71))

$$\mathbf{x}_t^{(n)} = \boldsymbol{\Lambda}^{(n)}\mathbf{F}_t + \boldsymbol{\xi}_t^{(n)}, \quad t \in \mathbb{Z}, \tag{126}$$

$$\mathbf{F}_t = \mathbf{N}(L)\mathbf{u}_t. \tag{127}$$

According to DSGE we can interpret:

1. $\mathbf{x}^{(n)}$ as the observed variables of a typical macroeconomic dataset;
2. $\mathbf{F}$ as some proxies for the state variables, which are unobserved;
3. $\mathbf{u}$ as the structural shocks;
4. $\boldsymbol{\xi}^{(n)}$ as measurement errors.

Under these conditions we are interested in the effect of the structural shocks on the common component $\boldsymbol{\chi}^{(n)}$, i.e. we are neglecting the idiosyncratic component. It is then possible to show that actually nonfundamentalness is a non-generic problem in factor models, and, under reasonable assumptions, we can always guarantee that the structural shocks $\mathbf{u}$ are $\boldsymbol{\chi}^{(n)}$-fundamental. In factor models we always have $n > q$, therefore, we refer to the condition for fundamentalness in singular systems. It is the singularity of dynamic factor models that makes the property of nonfundamentalness non generic.

To have fundamentalness we ask for $\mathbf{B}^{(n)}(L)$ to have maximum rank q inside the unit circle, i.e. its columns must be linearly independent. This is equivalent to asking for heterogeneity of the IRF (which are the elements of $\mathbf{B}^{(n)}(L)$) of the n variables to the q structural shocks. In very large datasets this is very likely to happen and it is therefore reasonable to assume fundamentalness. Roughly speaking, although in principle the econometrician has a smaller information space than the agents' one (i.e. there is room for potential nonfundamentalness), she can include additional series in the system, and, if dynamic heterogeneity is guaranteed, then these series contain truly new and useful information.

To clarify this point further, let us consider again the permanent income model. We have seen that the model is nonfundamental. However, if the econometrician observes also some additional variables $\mathbf{x}_t^*$ such that $(1 - L)x_{it}^* = \mathbf{b}_i(L)\mathbf{u}_t$, then $\mathbf{u}_t$ is fundamental for the whole system unless $\mathbf{b}_i(z)$ has the same root as the original system, i.e. unless $\mathbf{b}_i(\beta) = \mathbf{0}$ for every variable x_{it}^* added, which is extremely unlikely. Notice

[21] We have

$$x_{1t} = a\left(\frac{1}{a} - L\right)u_t, \quad x_{2t} = b\left(\frac{1}{b} - L\right)u_t$$

and then

$$\frac{x_{1t}}{a} - \frac{x_{2t}}{b} = \frac{1}{a}u_t - \frac{1}{b}u_t = \frac{b-a}{ab}u_t$$

that adding savings $s_t = y_t - c_t$ as an additional variable, for example, would not work, indeed

$$\mathbf{x}_t = \begin{pmatrix} (1-L)y_t \\ (1-L)c_t \\ (1-L)s_t \end{pmatrix} = \begin{pmatrix} 1 & 1-L \\ 1 & 1-\beta \\ 0 & (\beta - L) \end{pmatrix} \begin{pmatrix} u_{1t} \\ u_{0t} \end{pmatrix} = \begin{pmatrix} \mathbf{C}(L) \\ \mathbf{b}(L) \end{pmatrix} \mathbf{u}_t.$$

The determinants of the three $q \times q$ submatrices will all have their only root in β, so it will still be impossible to recover $\mathbf{u}_t$ using only past and present observations of income, consumption and savings. This makes sense, as, being savings just a linear combination of income and consumption, they do not contain any useful additional information, i.e. we do not have any dynamic heterogeneity in the system.

FUNDAMENTALNESS IN FACTOR MODELS - THEORY. Fundamentalness in singular models is proved in the following Theorem.

THEOREM 7 (Anderson and Deistler (2008)). *Consider an n-dimensional process $\boldsymbol{\chi}^{(n)}$, with rational spectral density $\boldsymbol{\Sigma}_{\boldsymbol{\chi}}^{(n)}(\theta)$ with the first q dynamic eigenvalues diverging as $n \to \infty$, and $n > q$. Then,*

$$\boldsymbol{\Sigma}_{\boldsymbol{\chi}}^{(n)}(\theta) = \frac{1}{2\pi}\mathbf{B}^{(n)}(e^{-i\theta})\mathbf{B}^{(n)'}(e^{i\theta}),$$

where $\mathbf{B}^{(n)}(z)$ is $n \times q$, rational, has no poles and no zeros for $|z| \leq 1$ and is associated to the representation

$$\boldsymbol{\chi}_t^{(n)} = \mathbf{B}^{(n)}(L)\mathbf{u}_t, \quad t \in \mathbb{Z},$$

where $\mathbf{u}_t$ is a q-dimensional orthonormal weak white noise. Moreover, $\mathbf{B}^{(n)}(z)$ admits a finite order left-inverse $\mathbf{A}^{(n)}(z)$ such that we can always write

$$\mathbf{A}^{(n)}(L)\boldsymbol{\chi}_t^{(n)} = \mathbf{B}^{(n)}(0)\mathbf{u}_t, \quad t \in \mathbb{Z}.$$

Given the properties of $\mathbf{B}^{(n)}(z)$ this Theorem guarantees fundamentalness of the shocks $\mathbf{u}$ for the common component $\boldsymbol{\chi}_t$. The estimation technique proposed in Forni, Hallin, Lippi, and Zaffaroni (2015b,a) exploits directly the autoregressive representation of the common component as given in Theorem 7, while here we consider only the implications of this theorem for the state-space version of the approximate dynamic factor model.

The idiosyncratic component can be neglected and considered as a vanishing measurement error and indeed we can show that as $n \to \infty$ the shocks $\mathbf{u}$ are also $\mathbf{x}^{(n)}$-fundamental. As we have already seen in Section 6, Theorem 7, when applied to the factors process, (127), guarantees the existence of a VAR model, which can be exploited when estimating the model. Therefore, we can consider again the state-space representation (70)-(73) of Section 6, where (127) is replaced by

$$\mathbf{G}(L)\mathbf{F}_t = \mathbf{N}(0)\mathbf{u}_t,$$

and $\mathbf{G}(L)$ is one-sided and such that $\mathbf{G}(L)\mathbf{N}(L) = \mathbf{N}(0)$. Under Assumption A5', condition *(b")* given in Section 3 is satisfied and $\frac{\boldsymbol{\Lambda}^{(n)'}\boldsymbol{\Lambda}^{(n)}}{n} \to \mathbf{H}$, with $\mathbf{H}$ being $r \times r$ and full-rank. Moreover, from Theorem 3, condition *(a')* in Section 3 is also satisfied. Define the $r \times n$ one-sided filter $\mathbf{S}^{(n)}(L) = \frac{\mathbf{G}(L)\mathbf{H}^{-1}\boldsymbol{\Lambda}^{(n)'}}{n}$[22], then, in a way similar to what we did in the PCA case in Section 3 we can show that, as $n \to \infty$,

$$\begin{aligned} \mathbf{S}^{(n)}(L)\mathbf{x}_t^{(n)} &= \mathbf{G}(L)\mathbf{H}^{-1}\frac{\boldsymbol{\Lambda}^{(n)'}\boldsymbol{\Lambda}^{(n)}}{n}\mathbf{F}_t + \mathbf{G}(L)\mathbf{H}^{-1}\frac{\boldsymbol{\Lambda}^{(n)'}\boldsymbol{\xi}_t^{(n)}}{n} \\ &= \mathbf{G}(L)\mathbf{H}^{-1}\frac{\boldsymbol{\Lambda}^{(n)'}\boldsymbol{\Lambda}^{(n)}}{n}\mathbf{N}(L)\mathbf{u}_t + \mathbf{G}(L)\mathbf{H}^{-1}\frac{\mathbf{A}^{(n)'}\boldsymbol{\xi}_t^{(n)}}{n} \xrightarrow{q.m.} \mathbf{N}(0)\mathbf{u}_t, \end{aligned} \tag{128}$$

where convergence is in quadratic mean (and therefore also in probability). As $n \to \infty$, $\mathbf{u}_t$ lies in the space spanned by the present and past values of $\mathbf{x}_t^{(n)}$, i.e. the structural shocks are $\mathbf{x}^{(n)}$-fundamental. In terms of Hilbert spaces, by recognizing that $\mathcal{H}_t^{\mathbf{x}^{(n)}} = \mathcal{H}_t^{\boldsymbol{\chi}^{(n)}} \oplus \mathcal{H}_t^{\boldsymbol{\xi}^{(n)}}$, then, as $n \to \infty$, we have

[22] Notice that the autoregressive polynomial $\mathbf{S}^{(n)}(L)$ is one-sided since $\mathbf{G}(L)$ is one-sided.

$\mathcal{H}_t^{\mathbf{x}^{(n)}} \to \mathcal{H}_t^{\chi^{(n)}} = \mathcal{H}_t^{\mathbf{u}}$, for any $t \in \mathbb{Z}$.

Summing up, fundamentalness is equivalent to the assumptions of reduced rank (i.e. comovements) and dynamic heterogeneity, which are perfectly reasonable assumptions when dealing with large datasets. As already pointed out, in empirical applications with large cross sections we often find evidence of reduced static and dynamic ranks, i.e. $r < n$ and $q < n$, which suggest the presence of few pervasive factors. Dynamic heterogeneity is then a reasonable property in a factor model with large cross sectional dimension n as economic variables react differently to the few structural shocks. Even if the structural shocks are fundamental for the whole system, they may not be fundamental for some subsamples of series. However, this is not a major problem in this context. Indeed, thanks to dynamic heterogeneity, the missing information due to local nonfundamentalness is completed with additional cross-sectional information contained in other series. In this sense, we can consider nonfundamentalness a non-generic problem for dynamic factor models.

IDENTIFICATION OF IMPULSE RESPONSES. We know from Section 7 how to estimate IRF. Without loss of generality, we know that a VAR(1) for the factors is sufficient, and we replace (127) with

$$\mathbf{F}_t = \mathbf{A}\mathbf{F}_{t-1} + \mathbf{G}\boldsymbol{\eta}_t, \quad \boldsymbol{\eta}_t \sim (\mathbf{0}, \boldsymbol{\Gamma}_{\boldsymbol{\eta}}),$$

with $\mathbf{G}$ being $r \times q$ and such that $\mathbf{G}\boldsymbol{\eta}_t = \mathbf{N}(0)\mathbf{u}_t$. Therefore, we know that it is always possible to find an invertible $q \times q$ matrix $\mathbf{R}$ such that $\boldsymbol{\eta}_t = \mathbf{R}\mathbf{u}_t$ and the identified IRF are given by

$$\begin{aligned} x_{it} &= \boldsymbol{\lambda}_i^{'}(\mathbf{I} - \mathbf{A}L)^{-1}\mathbf{G}\mathbf{R}\mathbf{u}_t + \xi_{it} \\ &= \mathbf{c}_i'(L)\mathbf{R}\mathbf{u}_t + \xi_{it} \\ &= \mathbf{b}_i'(L)\mathbf{u}_t + \xi_{it}, \quad i \in \mathbb{N}, \quad t \in \mathbb{Z}. \end{aligned} \tag{129}$$

and we refer to $\mathbf{c}_i'(L)$ as to the non-identified IRF and to $\mathbf{b}_i'(L)$ as to the identified IRF, these are collected into the $n \times q$ matrix polynomials $\mathbf{C}^{(n)}(L)$ and $\mathbf{B}^{(n)}(L)$, respectively. Since we are interested in identifying the shocks we must deal with the choice of $\mathbf{R}$. Since we typically we assume normality of the shocks $\mathbf{u}$, then in order to impose independence of the identified shocks $\mathbf{u}$ it is enough to impose orthogonality, thus we can restrict $\mathbf{R}$ to be an orthogonal matrix, i.e. such that $\mathbf{R}'\mathbf{R} = \mathbf{I}_q$. The main advantage with respect to SVAR is that in this case identification is always reduced to the choice of an orthogonal matrix $\mathbf{R}$ with only $q(q-1)/2$ parameters, independently of the number of series n considered, which is typically much larger than q. In contrast with the SVAR case, we do not have to impose any limitation on the size of the panel in order to estimate and identify the structural shocks. Notice that if one were to use FAVAR for structural analysis then also $\mathbf{F}_t$ should be identified, thus increasing the number of restrictions that we need to impose. On restrictions see also Bai and Wang (2015) and Han (2018).

As we can assume that the idiosyncratic component contains negligible information (e.g. it contains measurement errors), after estimating a dynamic factor model on a large panel $\mathbf{x}^{(n)}$, we can compare the estimated impulse responses of a q-dimensional subset of series $\mathbf{x}_t^{(q)}$, with those obtained by estimating a q-dimensional SVAR using only $\mathbf{x}_t^{(q)}$. This is the approach used in Forni and Gambetti (2010); Barigozzi, Conti, and Luciani (2014); Juvenal and Petrella (2015); Luciani (2015); Dahlhaus (2017).

There are many possibility for choosing the identification strategy, i.e. the choice and estimation of $\mathbf{R}$. Three main strategies are possible:

1. recursive or Cholesky identification used by Forni and Gambetti (2010) in factor models;
2. identification based on long-run restrictions as proposed by Blanchard and Quah (1989) and used by Forni, Giannone, Lippi, and Reichlin (2009) in factor models;
3. identification based on sign restrictions as proposed by Uhlig (2005) and then applied in factor models by Luciani (2015) and Barigozzi, Conti, and Luciani (2014).

In the recursive scheme Forni and Gambetti (2010) identify a monetary policy shock by assuming that it does not affect at impact neither Industrial Production nor the CPI. Consider the $q \times q$ subset of IRF

corresponding to the rows of Industrial Production, CPI, Federal Funds rate, and CHF/USD exchange rate (the order is important here), and define this sub-matrix as $\mathbf{C}_q(L)$. The Cholesky triangular factorization is such that

$$\mathbf{C}_q(0)\mathbf{C}_q'(0) = \mathbf{K}\mathbf{K}'$$

with $\mathbf{K}$ being lower triangular. The matrix $\mathbf{K}$ has 6 zeros, corresponding to the 6 restrictions we need when $q = 4$. The rotation matrix is then defined as

$$\mathbf{R} = [\mathbf{C}_q(0)]^{-1}\mathbf{K}.$$

In this way, from (129), the impact of the 4 shocks on the 4 selected variables are

$$\begin{pmatrix} \chi_{ipi,t} \\ \chi_{cpi,t} \\ \chi_{ffr,t} \\ \chi_{exr,t} \end{pmatrix} = \mathbf{B}_q(0)\mathbf{u}_t = \mathbf{C}_q(0)[\mathbf{C}_q(0)]^{-1} \begin{pmatrix} k_1 & 0 & 0 & 0 \\ k_2 & k_3 & 0 & 0 \\ k_4 & k_5 & k_6 & 0 \\ k_7 & k_8 & k_9 & k_{10} \end{pmatrix} \begin{pmatrix} u_{1t} \\ u_{2t} \\ u_{3t} \\ u_{4t} \end{pmatrix},$$

and we call u_{3t} the monetary policy shock and the related IRF for all variables are in the third column of $\mathbf{C}(L)\mathbf{R}$.

For identification based on long-run restrictions consider the simple case $q = 2$. A generic rotation matrix in 2 dimensions is always written as

$$\mathbf{R} = \begin{pmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{pmatrix};$$

with $\theta \in [0, 2\pi]$ being the rotation angle. We have to impose one restriction, i.e. fix the rotation angle θ. Assume that one of the two shocks say u_{2t} is a transitory shock (e.g. a demand shock), while the other say u_{1t} has a permanent effect on all variables (e.g. technology or productivity shock). Following the example in Blanchard and Quah (1989) we then assume that u_{2t} has no long-run effect on real variables, e.g. on Industrial Production and unemployment rate (we need $q = 2$ variables to impose restrictions on as we have two shocks). Then we have

$$\begin{aligned} \begin{pmatrix} \chi_{ipi,t} \\ \chi_{ur,t} \end{pmatrix} &= \begin{pmatrix} b_{11}(L) & b_{12}(L) \\ b_{21}(L) & b_{22}(L) \end{pmatrix} \begin{pmatrix} u_{1t} \\ u_{2t} \end{pmatrix} = \\ &= \begin{pmatrix} c_{11}(L) & c_{12}(L) \\ c_{21}(L) & c_{22}(L) \end{pmatrix} \begin{pmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{pmatrix} \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix} \begin{pmatrix} v_{1t} \\ v_{2t} \end{pmatrix} \end{aligned}$$

The restriction we impose is

$$b_{12}(1) = c_{11}(1)\sin\theta + c_{12}(1)\cos\theta = 0$$

and solving for θ we have the estimated angle

$$\theta = \tan^{-1}\left(\frac{-c_{12}(1)}{c_{11}(1)}\right).$$

Things get more complex with $q > 2$, e.g. with $q = 3$ we need 3 angles and with $q = 6$ we need 15 angles.

When we have many shocks we can use sign restrictions. For example, as in Barigozzi, Conti, and Luciani (2014), we can identify a monetary policy shock by imposing that at impact it has positive effect on the Federal Funds rate, and negative on CPI, Industrial Production and on other variables. Assume that the

monetary policy shock is u_{1t} then we look for a rotation matrix such that the identified system is

$$\begin{pmatrix} \chi_{ipi,t} \\ \\ \chi_{ffr,t} \\ \\ \chi_{cpi,t} \\ \vdots \end{pmatrix} = \begin{pmatrix} \underbrace{b_{11}(0)}_{<0} & \dots & b_{1q}(0) \\ \\ \underbrace{b_{21}(0)}_{<0} & \dots & b_{2q}(0) \\ \\ \underbrace{b_{31}(0)}_{>0} & \dots & b_{3q}(0) \\ \vdots \end{pmatrix} \begin{pmatrix} u_{1t} \\ \vdots \\ u_{qt} \end{pmatrix}$$

u_{1t} is the monetary policy shock and the others are left unidentified. In order to determine the rotation we extract a $q(q-1)/2$ dimensional uniform distribution of angles defined on $[0, 2\pi]$. Thus, we build a distribution of $q \times q$ rotation matrices and among these we keep only those satisfying the restrictions. In factor models, we can impose up to n sign restrictions for any shock, however if we impose too many restrictions we are likely to find no rotation satisfying them. This strategy gives only partial identification as there might be more than one rotation satisfying the restrictions. On the other hand, both the recursive and the long-run identification schemes have exact solutions, hence we have just identified models.

10 Factor models in finance

Factor models for financial data, are aimed at analysing either stock returns or their volatility (conditional variance). We consider here the following topics

1. the Arbitrage Pricing Theory (APT) and optimal investment portfolios;
2. network analysis of interdependencies in idiosyncratic components to measure risk;
3. parametric models for large conditional variance matrices.

APT AND THE STATIC APPROXIMATE FACTOR MODEL. Stock returns are often seen to have very low autocorrelation and therefore are often modelled as white noises, and, in this case, a static factor model seems to be an appropriate choice. Therefore, Chamberlain and Rothschild (1983) consider a static approximate r factor model for returns

$$\mathbf{x}_t^{(n)} = \boldsymbol{\mu}^{(n)} + \boldsymbol{\Lambda}^{(n)}\mathbf{F}_t + \boldsymbol{\xi}_t^{(n)}, \tag{130}$$

where $\boldsymbol{\mu}^{(n)} = (\mu_1 \dots \mu_n)'$. Notice that differently from the usual case the variables have mean μ_i which is not zero, and $\boldsymbol{\mu}^{(n)} = \mathsf{E}[\mathbf{x}_t^{(n)}]$ is the vector of means. The usual assumptions of the approximate factor models are made. In particular, with reference to the properties in Section 7, we make Assumption A5', implies diverging eigenvalues of the common component (linearly in n) and we assume bounded eigenvalues of the idiosyncratic covariance matrix, which is consequence of condition 4. The factors and the idiosyncratic components are orthogonal because of condition 0.

Now, consider the orthogonal projection of the expected return vector $\boldsymbol{\mu}_n$ onto the space spanned by the $(r+1)$ columns of $(\boldsymbol{\iota}_n, \boldsymbol{\Lambda}^{(n)})$, where $\boldsymbol{\iota}_n$ is an n-dimensional column of ones:

$$\boldsymbol{\mu}^{(n)} = \boldsymbol{\iota}_n\rho + \sum_{j=1}^{r} \boldsymbol{\lambda}_j^{(n)}\tau_j + \mathbf{c}^{(n)}, \tag{131}$$

where ρ and τ_j are the projection coefficients and $\boldsymbol{\lambda}_j^{(n)}$ is the jth column of $\boldsymbol{\Lambda}^{(n)}$ and is n-dimensional. Since the above is an orthogonal projection, the residuals $\mathbf{c}^{(n)}$ must satisfy the orthogonality conditions[23]

$$\mathbf{c}^{(n)'}\boldsymbol{\lambda}_j^{(n)} = \sum_{i=1}^{n} c_i\lambda_{ij} = 0, \quad j = 1, \dots, r, \tag{132}$$

$$\mathbf{c}^{(n)'}\boldsymbol{\iota}_n = \sum_{i=1}^{n} c_i = 0. \tag{133}$$

A portfolio is a weighted average of the stock returns where the weights represent the fraction of total wealth allocated to each stock. So if we consider a portfolio with weights given by $\mathbf{c}^{(n)}$, and denoted as $p_t^{(n)} = \mathbf{c}^{(n)'}\mathbf{x}_t^{(n)}$ then by (133), the portfolio is costless. Thus, by (130), using (131), (132), (133), we have

$$\begin{aligned} p_t^{(n)} = \mathbf{c}^{(n)'}\mathbf{x}_t^{(n)} &= \mathbf{c}^{(n)'}\boldsymbol{\mu}^{(n)} + \mathbf{c}^{(n)'}\boldsymbol{\Lambda}^{(n)}\mathbf{F}_t + \mathbf{c}^{(n)'}\boldsymbol{\xi}_t^{(n)} = \mathbf{c}^{(n)'}\boldsymbol{\mu}^{(n)} + \mathbf{c}^{(n)'}\boldsymbol{\xi}_t^{(n)} \\ &= \mathbf{c}^{(n)'}\boldsymbol{\iota}_n\rho + \mathbf{c}^{(n)'}\sum_{j=1}^{r}\boldsymbol{\lambda}_j^{(n)}\tau_j + \mathbf{c}^{(n)'}\mathbf{c}^{(n)} + \mathbf{c}^{(n)'}\boldsymbol{\xi}_t^{(n)} \\ &= \mathbf{c}^{(n)'}\mathbf{c}^{(n)} + \mathbf{c}^{(n)'}\boldsymbol{\xi}_t^{(n)}. \end{aligned} \tag{134}$$

The expected return on this portfolio is then

$$\mathsf{E}[p_t^{(n)}] = \mathsf{E}[\mathbf{c}^{(n)'}\mathbf{x}_t^{(n)}] = \mathbf{c}^{(n)'}\mathbf{c}^{(n)}. \tag{135}$$

[23] It is customary to consider nested sequences thus when moving from n to $n+1$ the first n entries of all vectors remain unchanged and therefore c_i and λ_{ij} do not depend on n.

Now let us consider the variance of this portfolio, which is a measure of risk associated to the portfolio, using (132) or directly (134), we have

$$\begin{aligned}\mathsf{Var}(p_t^{(n)}) = \mathsf{Var}(\mathbf{c}^{(n)'}\mathbf{x}_t^{(n)}) &= \mathbf{c}^{(n)'}\boldsymbol{\Gamma}_{\mathbf{x}}^{(n)}\mathbf{c}^{(n)} \\ &= \mathbf{c}^{(n)'}\boldsymbol{\Lambda}^{(n)}\boldsymbol{\Gamma}_{\mathbf{F}}\boldsymbol{\Lambda}^{(n)'}\mathbf{c}^{(n)} + \mathbf{c}^{(n)'}\boldsymbol{\Gamma}_{\boldsymbol{\xi}}^{(n)}\mathbf{c}^{(n)} \\ &= \mathbf{c}^{(n)'}\boldsymbol{\Gamma}_{\boldsymbol{\xi}}^{(n)}\mathbf{c}^{(n)}. \end{aligned} \tag{136}$$

Thus, from (136) and the definition of eigenvalues, we have the bounds on the portfolio variance (recall that $\mathbf{c}^{(n)'}\mathbf{c}^{(n)} = \|\mathbf{c}^{(n)}\|^2$)

$$\lambda_{n\boldsymbol{\xi}}^{(n)}\|\mathbf{c}^{(n)}\|^2 \leq \mathsf{Var}(p_t^{(n)}) \leq \lambda_{1\boldsymbol{\xi}}^{(n)}\|\mathbf{c}^{(n)}\|^2 \tag{137}$$

Define the Sharpe ratio as the ratio between the expected return and the risk of a given portfolio

$$\delta^{(n)} = \frac{|\mathsf{E}[p_t^{(n)}]|}{\sqrt{\mathsf{Var}(p_t^{(n)})}} = \frac{\|\mathbf{c}^{(n)}\|^2}{\sqrt{\mathsf{Var}(p_t^{(n)})}}. \tag{138}$$

This gives the trade-off mean-variance and the larger the Sharpe ratio the better it is. Then, from (135), (136), and (137), we have the bounds

$$\frac{\|\mathbf{c}^{(n)}\|^2}{\lambda_{1\boldsymbol{\xi}}^{(n)}} \leq (\delta^{(n)})^2 \leq \frac{\|\mathbf{c}^{(n)}\|^2}{\lambda_{n\boldsymbol{\xi}}^{(n)}}. \tag{139}$$

In the case $\delta^{(n)} \to \infty$ we say that we have arbitrage, indeed, if we assume $\lambda_{n\boldsymbol{\xi}}^{(n)} > 0$, the only possibility for this to happen is that $\|\mathbf{c}^{(n)}\|^2 \to \infty$, which in turn implies $\mathsf{E}[p_t^{(n)}] \to \infty$ and $\mathsf{Var}(p_t^{(n)}) \to 0$, and this is a portfolio with infinite return and zero risk.

So to impose no arbitrage, $\delta^{(n)} < \infty$, we must require $\|\mathbf{c}^{(n)}\|^2 < \infty$. In particular, from (139), we also find that

$$(\delta^{(n)})^2\lambda_{n\boldsymbol{\xi}}^{(n)} \leq \|\mathbf{c}^{(n)}\|^2 \leq (\delta^{(n)})^2\lambda_{1\boldsymbol{\xi}}^{(n)} < \infty, \tag{140}$$

since by assumption the eigenvalues of the idiosyncratic covariance are bounded. As a consequence of the upper bound in (140), for any i we must have $c_i \to 0$ as $n \to \infty$, therefore from (131) we have that the return on the ith stock tends to

$$\mu_i \to \rho + \sum_{j=1}^{r} \lambda_{ij}\tau_j, \quad n \to \infty \tag{141}$$

which is the same result as in the Capital Asset Pricing Model (CAPM). If we consider the factors as coming from the market we are saying that we can price each stock based on its relation to the market.[24] In other words, we have the well known result that, under a factor structure, after aggregation (i.e. building a portfolio) only the factor component survives when $n \to \infty$.

PORTFOLIO OPTIMISATION AND LARGE COVARIANCE MATRICES. We have seen how to build a no arbitrage portfolio. However, in portfolio optimisation we want to build portfolios with maximum expected return or minimum risk. We here focus on the second case, thus,[25]

$$\min_{\mathbf{u}} \mathsf{Var}(\mathbf{u}'\mathbf{x}_t^{(n)}) = \min_{\mathbf{u}} \mathbf{u}'\boldsymbol{\Gamma}_{\mathbf{x}}^{(n)}\mathbf{u}, \qquad \text{s.t. } \mathbf{u}'\boldsymbol{\iota}_n = 1. \tag{142}$$

[24] The loadings are sometimes denoted as β_{ij} and in finance are called betas.

[25] If instead we also set a given a pre-specified portfolio return μ^*, we want to solve

$$\min_{\mathbf{u}} \mathsf{Var}(\mathbf{u}'\mathbf{x}_t^{(n)}) = \min_{\mathbf{u}} \mathbf{u}'\boldsymbol{\Gamma}_{\mathbf{x}}^{(n)}\mathbf{u}, \qquad \text{s.t. } \mathbf{u}'\mathbf{u} = 1, \quad \mathbf{u}'\boldsymbol{\mu}^{(n)} = \mathsf{E}[\mathbf{u}'\mathbf{x}_t^{(n)}] = \mu^*.$$

The solution of (142) is given by the optimal weights

$$\mathbf{w}^{(n)} = \beta\big(\boldsymbol{\Gamma}_{\mathbf{x}}^{(n)}\big)^{-1}\boldsymbol{\iota}_n + \gamma\big(\boldsymbol{\Gamma}_{\mathbf{x}}^{(n)}\big)^{-1}\boldsymbol{\mu}^{(n)}$$

The solution of (142) is given by the optimal weights

$$\mathbf{w}^{(n)} = \frac{\left(\mathbf{\Gamma}_{\mathbf{x}}^{(n)}\right)^{-1}\boldsymbol{\iota}_n}{\boldsymbol{\iota}_n'\left(\mathbf{\Gamma}_{\mathbf{x}}^{(n)}\right)^{-1}\boldsymbol{\iota}_n}, \tag{143}$$

which has Sharpe ratio

$$\delta^{(n)} = \frac{\left|\boldsymbol{\iota}_n'\left(\mathbf{\Gamma}_{\mathbf{x}}^{(n)}\right)^{-1}\boldsymbol{\mu}^{(n)}\right|}{\left[\boldsymbol{\iota}_n'\left(\mathbf{\Gamma}_{\mathbf{x}}^{(n)}\right)^{-1}\boldsymbol{\iota}_n\right]^{1/2}}.$$

It is then clear that to solve the problem we need an estimator of the inverse covariance matrix of $\mathbf{x}^{(n)}$, which for large n might be very hard to obtain. In particular, if we let $n \to \infty$ then the sample covariance matrix is not a consistent estimator of the true covariance unless $n^2/T \to 0$, which is hardly the case (see (36)). Many alternative estimators of the covariance matrix which deliver an invertible matrix even in the large n case have been proposed, however most of them are for independent data (hence not time series). Moreover, a whole class of estimators assumes that the covariance matrix is a sparse matrix, and the usual definition of sparsity is that

$$\|\mathbf{\Gamma}_{\mathbf{x}}^{(n)}\|_1 = \max_{j=1,\dots,n}\sum_{i=1}^{n}\left|\left[\mathbf{\Gamma}_{\mathbf{x}}^{(n)}\right]_{ij}\right| = o(n). \tag{144}$$

Since $\|\mathbf{\Gamma}_{\mathbf{x}}^{(n)}\|_1 \geq \|\mathbf{\Gamma}_{\mathbf{x}}^{(n)}\|_2 = \lambda_{1\mathbf{x}}^{(n)}$, condition (144) clearly violates the assumption of a factor structure where $\lambda_{1\mathbf{x}}^{(n)} = O(n)$.

Fan, Liao, and Mincheva (2013) provide a regularised estimator of the covariance matrix which holds for time dependent data and is compatible with factor structure. The covariance matrix is written as

$$\mathbf{\Gamma}_{\mathbf{x}}^{(n)} = \mathbf{\Lambda}^{(n)'}\mathbf{\Gamma}_{\mathbf{F}}\mathbf{\Lambda}^{(n)} + \mathbf{\Gamma}_{\boldsymbol{\xi}}^{(n)} = \mathbf{\Gamma}_{\boldsymbol{\chi}}^{(n)} + \mathbf{\Gamma}_{\boldsymbol{\xi}}^{(n)},$$

which shows that it is made of a reduced rank matrix plus a matrix which we know has small cross-sectional dependence, that is $\boldsymbol{\lambda}_{1\boldsymbol{\xi}}^{(n)}$ is bounded for any n. Hence, we can assume sparsity of the idiosyncratic covariance, indeed it is always true that

$$\boldsymbol{\lambda}_{1\boldsymbol{\xi}}^{(n)} = \|\mathbf{\Gamma}_{\boldsymbol{\xi}}^{(n)}\|_2 \leq \|\mathbf{\Gamma}_{\boldsymbol{\xi}}^{(n)}\|_1 = o(n).$$

The covariance matrix is then estimated in two steps:

1. PCA to estimate the covariance of the common component $\mathbf{\Gamma}_{\boldsymbol{\chi}}^{(nT)}$ as in Sections 3 and 7, and define $\mathbf{\Gamma}_{\boldsymbol{\xi}}^{(nT)} = \mathbf{\Gamma}_{\mathbf{x}}^{(nT)} - \mathbf{\Gamma}_{\boldsymbol{\chi}}^{(nT)}$, where $\mathbf{\Gamma}_{\mathbf{x}}^{(nT)}$ is the sample covariance of $\mathbf{x}^{(n)}$;

2. penalise $\mathbf{\Gamma}_{\boldsymbol{\xi}}^{(nT)}$ by forcing to zero all entries that satisfy a certain condition, for example define $\mathbf{\Gamma}_{\boldsymbol{\xi},\tau}^{(nT)}$ such that its entries satisfy

$$\left[\mathbf{\Gamma}_{\boldsymbol{\xi},\tau}^{(nT)}\right]_{ij} = \left[\mathbf{\Gamma}_{\boldsymbol{\xi}}^{(nT)}\right]_{ij}\mathbb{I}_{\left|\left[\mathbf{\Gamma}_{\boldsymbol{\xi}}^{(nT)}\right]_{ij}\right|>\tau},$$

   which is called hard-thresholding, while other penalisations are also possible.

where

$$\beta = \frac{(\boldsymbol{\mu}^{(n)'} - \mu^*\boldsymbol{\iota}_n')(\mathbf{\Gamma}_{\mathbf{x}}^{(n)})^{-1}\boldsymbol{\mu}^{(n)}}{\boldsymbol{\iota}_n'(\mathbf{\Gamma}_{\mathbf{x}}^{(n)})^{-1}\boldsymbol{\iota}_n\boldsymbol{\mu}^{(n)'}(\mathbf{\Gamma}_{\mathbf{x}}^{(n)})^{-1}\boldsymbol{\mu}^{(n)} - \left[\boldsymbol{\iota}_n'(\mathbf{\Gamma}_{\mathbf{x}}^{(n)})^{-1}\boldsymbol{\mu}^{(n)}\right]^2},$$
$$\gamma = \frac{\boldsymbol{\iota}_n'(\mathbf{\Gamma}_{\mathbf{x}}^{(n)})^{-1}(\mu^*\boldsymbol{\iota}_n - \boldsymbol{\mu}^{(n)})}{\boldsymbol{\iota}_n'(\mathbf{\Gamma}_{\mathbf{x}}^{(n)})^{-1}\boldsymbol{\iota}_n\boldsymbol{\mu}^{(n)'}(\mathbf{\Gamma}_{\mathbf{x}}^{(n)})^{-1}\boldsymbol{\mu}^{(n)} - \left[\boldsymbol{\iota}_n'(\mathbf{\Gamma}_{\mathbf{x}}^{(n)})^{-1}\boldsymbol{\mu}^{(n)}\right]^2}.$$

The resulting estimated covariance is defined as

$$\mathbf{\Gamma}_{\mathbf{x},\tau}^{(nT)} = \mathbf{\Gamma}_{\boldsymbol{\chi}}^{(nT)} + \mathbf{\Gamma}_{\boldsymbol{\xi},\tau}^{(nT)}, \tag{145}$$

and it can proved that is a consistent estimator of the true covariance and that $(\mathbf{\Gamma}_{\mathbf{x},\tau}^{(nT)})^{-1}$ exists. Therefore, (145) can be used into (143) to estimate the optimal weights of the minimum risk portfolio.

NETWORKS AND PARTIAL CORRELATION MATRICES. Another object of interest in financial application is the partial correlation matrix, that is the matrix with as generic (i,j)th entry, ρ^{ij}, the correlation between stock returns x_i and x_j given all other returns. The partial correlation is directly related to the inverse covariance matrix by the relation

$$\rho^{ij} = -\frac{\left[(\mathbf{\Gamma}_{\mathbf{x}}^{(n)})^{-1}\right]_{ij}}{\sqrt{\left[(\mathbf{\Gamma}_{\mathbf{x}}^{(n)})^{-1}\right]_{ii}\left[(\mathbf{\Gamma}_{\mathbf{x}}^{(n)})^{-1}\right]_{jj}}}. \tag{146}$$

Therefore, the method explained above is useful to compute also partial correlation.

On the other hand, graphical models literature has shown that partial correlations have a useful representation as networks, which in turn can be used to study the propagation of shocks through the market in what can be seen as an assessment of systemic risk by looking at most connected nodes in the network (Diebold and Yılmaz, 2014). To this end it is customary to consider sparse networks rather than full ones. In this sense the estimator proposed above is not ideal since having defined $\mathbf{\Gamma}_{\boldsymbol{\xi},\tau}^{(nT)}$ as sparse in general does not imply that also its inverse is sparse. Moreover, once again the factor structure is in contradiction with the idea that $(\mathbf{\Gamma}_{\mathbf{x}}^{(n)})^{-1}$ is sparse. To see this, consider, as an example, the static approximate factor model (130) with one factor ($r = 1$). The inverse covariance matrix of $\mathbf{x}^{(n)}$ is given by

$$(\mathbf{\Gamma}_{\mathbf{x}}^{(n)})^{-1} = (\mathbf{\Gamma}_{\boldsymbol{\xi}}^{(n)})^{-1} - \frac{(\mathbf{\Gamma}_{\boldsymbol{\xi}}^{(n)})^{-1}\boldsymbol{\lambda}^{(n)}\boldsymbol{\lambda}^{(n)'}(\mathbf{\Gamma}_{\boldsymbol{\xi}}^{(n)})^{-1}}{1 + \boldsymbol{\lambda}^{(n)'}(\mathbf{\Gamma}_{\boldsymbol{\xi}}^{(n)})^{-1}\boldsymbol{\lambda}^{(n)}},$$

where $\boldsymbol{\lambda}^{(n)}$ is the $n \times 1$ vector of factor loadings. Under the assumption of diverging eigenvalues like Assumption A5' in Section 6, we have $\|\boldsymbol{\lambda}^{(n)}\|^2 = \boldsymbol{\lambda}^{(n)'}\boldsymbol{\lambda}^{(n)} = \text{Tr}(\boldsymbol{\lambda}^{(n)}\boldsymbol{\lambda}^{(n)'}) = O(n)$, and the matrix $(\mathbf{\Gamma}_{\mathbf{x}}^{(n)})^{-1}$ cannot be sparse in the sense of (144) even if $(\mathbf{\Gamma}_{\boldsymbol{\xi}}^{(n)})^{-1}$ is sparse. Moreover, if we think of the factor as a weighted average of all other series, say $f_t = \mathbf{a}'\mathbf{x}_t = \sum_{i=1}^{n} a_i x_{it}$, then though a factor model x_i is function of all other x's, that is $\rho^{ij} \neq 0$ for all j.

Therefore, in large dimensional settings networks should be considered for the idiosyncratic component only. A way to obtain sparse estimators of the inverse idiosyncratic covariance is by considering the regression for the estimated idiosyncratic components (for example obtained from PCA)

$$\xi_{it}^{(nT)} = \sum_{\substack{i=1 \\ i \neq j}}^{n} \beta_{ij}\xi_{jt}^{(nT)} + v_{it}, \qquad i = 1, \ldots, n. \tag{147}$$

Since $\beta_{ij} \neq 0$ if and only if $\left[(\mathbf{\Gamma}_{\mathbf{x}}^{(nT)})^{-1}\right]_{ij} \neq 0$, then, by estimating (147) for example with LASSO, we obtain through (146) an estimated sparse partial correlation network.

This approach can be applied to returns, but it is more interesting on volatility time series as these can be seen as risk measures. In the latter case conditional lagged dependencies should also be considered and hence (147) should be replaced by sparse VAR models (Barigozzi and Hallin, 2017).

FACTOR GARCH MODELS. In the previous application the dynamics of the factors has not be considered, indeed as we said returns are usually modelled as white noise. However, we also know that squared returns are likely to be correlated, that is returns are not independent. A way to model this feature is by assuming that returns are conditionally heteroschedastic that is their variance and covariance conditional on the past evolve with time. In univariate setting this is captured with GARCH models, but in a multivariate setting

these models are hardly generalised due to the too many parameters that would be needed. A parsimonious solution is provided by the factor GARCH model (Diebold and Nerlove, 1989)

$$\begin{aligned}\mathbf{x}^{(n)} &= \mathbf{\Lambda}^{(n)}\mathbf{F}_t + \boldsymbol{\xi}_t^{(n)},\\ \mathbf{F}_t &= \mathbf{H}_t^{1/2}\boldsymbol{\nu}_t, \qquad \boldsymbol{\nu}_t \sim i.i.d.N(\mathbf{0}, \mathbf{I}_r),\\ \big[\mathbf{H}_t\big]_{ii} &= \omega_i + \alpha_i F_{it-1}^2 + \beta_i \big[\mathbf{H}_{t-1}\big]_{ii}, \qquad i = 1, \dots, r,\end{aligned} \tag{148}$$

with $\omega_i > 0$, $\alpha_i, \beta_i \geq 0$ to guarantee positivity, and $\alpha_i + \beta_i < 1$ to ensure weak stationarity. Then it is straightforward to see that

$$\mathsf{E}[\mathbf{F}_t|\mathcal{X}_{t-1}] = \mathbf{0}, \qquad \mathsf{Var}(\mathbf{F}_t|\mathcal{X}_{t-1}) = \mathbf{H}_t.$$

Moreover, we can also assume $\big[\mathbf{H}_t\big]_{ij}$ to be either constant (which would result in a CCC-GARCH) or time dependent (which would result in a DCC-GARCH or a BEKK-GARCH).

From (148) we immediately have that the conditional covariance of $\mathbf{x}^{(n)}$ is given by

$$\mathsf{Var}(\mathbf{x}_t^{(n)}|\mathcal{X}_{t-1}) = \mathbf{\Lambda}^{(n)}\mathbf{H}_t\mathbf{\Lambda}^{(n)'} + \mathsf{Var}(\boldsymbol{\xi}_t^{(n)}|\mathcal{X}_{t-1})$$

In the simplest case we assume $\boldsymbol{\xi}_t^{(n)}$ to be white noise and also homoschedastic so that $\mathsf{Var}(\boldsymbol{\xi}_t^{(n)}|\mathcal{X}_{t-1}) = \mathbf{\Gamma}_{\boldsymbol{\xi}}^{(n)}$. Generalisations to univariate idiosyncratic GARCH models are also possible, thus making $\mathsf{Var}(\boldsymbol{\xi}_t^{(n)}|\mathcal{X}_{t-1})$ diagonal and time dependent.

Estimation of (148) in two steps is possible by first extracting the factors via PCA and then estimating the appropriate GARCH model (Alessi, Barigozzi, and Capasso, 2009; Aramonte, del Giudice Rodriguez, and Wu, 2013). This approach is however not efficient since it does not exploit conditional heteroscedasticity in estimating the factors. Alternatively, the KF approach can be used, possibly embedded in an EM algorithm, to deliver more efficient estimates (Harvey, Ruiz, and Sentana, 1992; Sentana, Calzolari, and Fiorentini, 2008, although in the finite n case).

11 Number of factors

EXPLAINED VARIANCES. Given the equivalence between a static model with r factors and a dynamic model with q factors, where typically $q \leq r$, it is natural to assume that the variance explained by the q largest dynamic eigenvalues:

$$EVD = \frac{\sum_{j=1}^{q} \int_{-\pi}^{\pi} \lambda_{j\mathbf{x}}^{(n)}(\theta) \mathrm{d}\theta}{\sum_{j=1}^{n} \int_{-\pi}^{\pi} \lambda_{j\mathbf{x}}^{(n)}(\theta) \mathrm{d}\theta}$$

is the same as the variance explained by the r largest eigenvalues of the covariance matrix:

$$EVS = \frac{\sum_{j=1}^{r} \lambda_{j\mathbf{x}}^{(n)}}{\sum_{j=1}^{n} \lambda_{j\mathbf{x}}^{(n)}}$$

By comparison of the two, we often find that in order to have $EVD = EVS$ we need $r > q$. This intuition is also confirmed by the results of the criteria to determine q and r presented below.

DYNAMIC FACTORS. We consider here the criterion proposed by Hallin and Liška (2007) which is based on the representation Theorem 2, thus on the GDFM.[26] The criterion is based on the properties of dynamic eigenvalues of the data and looks for the value q that minimizes the contribution of the idiosyncratic component. Clearly, in order to avoid overestimation of q a penalty $p(n,T)$ is required and due to the large n,T setting such function must depend on both dimensions. Moreover, the penalty is tuned by means of a constant c. Thus for a grid of values of $q = 0, \ldots, q_{\max}$ and $c \in [0, c_{\max}]$, we need to compute the criterion

$$IC^{(nT)}(c,q) = \frac{1}{n} \sum_{i=q+1}^{n} \frac{1}{2M_T+1} \sum_{h=-M_T}^{M_T} \lambda_{i\mathbf{x}}^{(nT)}(\theta_h) + c\, q\, p(n,T),$$

where we considered a grid of frequencies θ_h on which we estimated the spectral density and its dynamic eigenvalues. Then, for any c we have to solve

$$q^{(nT)}(c) = \underset{q=0,\ldots,q_{\max}}{\arg\min} \; IC^{(nT)}(c,q). \tag{149}$$

In order to determine c and the estimate the number of dynamic factors, we have to compute the criterion for $J < n$ sub-blocks of series of size $n_j = n - J + j$ with $j = 1, \ldots, J$. For each sub-block we can compute the number of dynamic factors by minimising the analogous of (149) which we denote by $q^{(n_j T)}(c)$. Thus the variation in the number of dynamic factors as the size of the panel varies is

$$S(c) = \frac{1}{J} \sum_{j=1}^{J} \left(q^{(n_j T)}(c) - \frac{1}{J} \sum_{j=1}^{J} q^{(n_j T)}(c) \right)^2.$$

This procedure mimics the idea that the number of dynamic factors is constant for a panel of size n_j when n_j is large enough, indeed q in Theorem 2 is defined for an infinite dimensional panel. Thus, we look for values of c such that:

1. $S(c) = 0$;
2. $q^{(nT)}(c)$ is constant and less than $q_{\max}$.

Starting from the smallest value of c denote by c^* the first value of c such that the two conditions are met. Then the estimator of the number of dynamic factors is $q^{(nT)}(c^*)$. We have the consistency result, as $n, T \to \infty$ (Hallin and Liška, 2007)

$$q^{(nT)}(c^*) \xrightarrow{p} q,$$

[26] An an alternative method is a test proposed by Onatski (2009) and based on the properties of dynamic eigenvalues. While the methods by Amengual and Watson (2007) and Bai and Ng (2007) require first the estimation of the number of static factors r.

provided that $p(n,T) \to 0$ and $\min(n, M_T^2, M_T^{-1/2}T^{1/2})p(n,T) \to \infty$, where M_T is the truncation parameter used to estimate the spectral density and is such that $M_T \to \infty$ and $M_T T^{-1} \to 0$ as $T \to \infty$.

STATIC FACTORS. We consider here the criterion by Bai and Ng (2002) with the penalty modification proposed in Alessi, Barigozzi, and Capasso (2010).[27] The criterion works exactly as the previous criterion for dynamic factors but considers the eigenvalues of the covariance matrix. Thus for a grid of values of $r = 0, \dots, r_{\max}$ and $c \in [0, c_{\max}]$, we need to compute the criterion[28]

$$IC^{(nT)}(c,r) = \frac{1}{n} \sum_{i=r+1}^{n} \lambda_{i\mathbf{x}}^{(nT)} + c\, r\, p(n,T),$$

Then, for any c we have to solve

$$r^{(nT)}(c) = \arg\min_{r=0,\dots,r_{\max}} IC^{(nT)}(c,r). \tag{150}$$

The optimal value of the constant denoted as c^* is determined exactly as in the case of dynamic factors and the estimator of the number of static factors is obtained as $r^{(nT)}(c^*)$. We then have consistency, as $n, T \to \infty$ (Bai and Ng, 2002)

$$r^{(nT)}(c^*) \xrightarrow{p} r,$$

provided that $p(n,T) \to 0$ and $\min(n,T)p(n,T) \to \infty$.

[27] Other methods based on the behaviour of eigenvalues are proposed by Onatski (2010), Ahn and Horenstein (2013), and Trapani (2017).

[28] Equivalently, the criterion is often written as

$$IC(c,r) = \frac{1}{T} \sum_{t=1}^{T} \left(\mathbf{x}_t^{(n)} - \mathbf{\Lambda}_r^{(nT)} \mathbf{F}_{t,r}^{(nT)}\right)' \left(\mathbf{x}_t^{(n)} - \mathbf{\Lambda}_r^{(nT)} \mathbf{F}_{t,r}^{(nT)}\right) + c\, r\, p(n,T) = \text{tr}(\mathbf{\Gamma}_\xi^{(nT)}) + c\, r\, p(n,T).$$

References

AHN, S. C., AND A. R. HORENSTEIN (2013): "Eigenvalue Ratio Test for the Number of Factors," *Econometrica*, 81, 1203–1227.

ALESSI, L., M. BARIGOZZI, AND M. CAPASSO (2009): "Estimation and forecasting in large datasets with conditionally heteroskedastic dynamic common factors," Working Paper 1115, ECB.

——— (2010): "Improved Penalization for Determining the Number of Factors in Approximate Static Factor Models," *Statistics and Probability Letters*, 80, 1806–1813.

——— (2011): "Nonfundamentalness in Structural Econometric Models: A Review," *International Statistical Review*, 79, 16–47.

ALTISSIMO, F., R. BASSANETTI, R. CRISTADORO, M. FORNI, M. HALLIN, M. LIPPI, L. REICHLIN, AND G. VERONESE (2001): "A real time coincident indicator for the euro area business cycle," Discussion Paper 3108, CEPR.

ALTISSIMO, F., R. CRISTADORO, M. FORNI, M. LIPPI, AND G. VERONESE (2010): "New Eurocoin: Tracking Economic Growth in Real Time," *The Review of Economics and Statistics*, 92, 1024–1034.

AMENGUAL, D., AND M. W. WATSON (2007): "Consistent Estimation of the Number of Dynamic Factors in a Large N and T Panel," *Journal of Business and Economic Statistics*, 25, 91–96.

ANDERSON, B. D., AND M. DEISTLER (2008): "Generalized linear dynamic factor models. A Structure Theory," *IEE Conference on Decision and Control*.

ARAMONTE, S., M. DEL GIUDICE RODRIGUEZ, AND J. WU (2013): "Dynamic factor Value-at-Risk for large heteroskedastic portfolios," *Journal of Banking & Finance*, 37, 4299–4309.

BAI, J. (2003): "Inferential theory for factor models of large dimensions," *Econometrica*, 71, 135–171.

BAI, J., AND S. NG (2002): "Determining the number of factors in approximate factor models," *Econometrica*, 70, 191–221.

——— (2006): "Confidence Intervals for Diffusion Index Forecasts and Inference for Factor-Augmented Regressions," *Econometrica*, 74, 1133–1150.

——— (2007): "Determining the Number of Primitive Shocks in Factor Models," *Journal of Business and Economic Statistics*, 25, 52–60.

——— (2008): "Forecasting economic time series using targeted predictors," *Journal of Econometrics*, 146, 304–317.

BAI, J., AND P. WANG (2015): "Identification and bayesian estimation of dynamic factor models," *Journal of Business & Economic Statistics*, 33, 221–240.

BAŃBURA, M., AND M. MODUGNO (2014): "Maximum likelihood estimation of factor models on data sets with arbitrary patterns of missing data," *Journal of Applied Econometrics*, 29, 133–160.

BARIGOZZI, M., AND C. T. BROWNLEES (2018): "NETS: Network Estimation for Time Series," *Journal of Applied Econometrics*, forthcoming.

BARIGOZZI, M., A. M. CONTI, AND M. LUCIANI (2014): "Do Euro Area Countries Respond Asymmetrically to the Common Monetary Policy?," *Oxford Bulletin of Economics and Statistics*, 76, 693–714.

BARIGOZZI, M., AND M. HALLIN (2017): "Networks, dynamic factors, and the volatility analysis of high-dimensional financial series," *Journal of the Royal Statistical Society, Series C*, 66, 581-605.

BARIGOZZI, M., AND M. LUCIANI (2017): "Common factors, trends, and cycles in large datasets," arxiv.1709.01445.

——— (2019a): "Quasi maximum likelihood estimation and inference of large approximate dynamic factor models via the EM algorithm," arxiv.1910.03821.

——— (2019b): "Quasi maximum likelihood estimation of non-stationary large approximate dynamic factor models," arxiv.1910.09841.

BERNANKE, B. S., AND J. BOIVIN (2003): "Monetary Policy in a Data–Rich Environment," *Journal of Monetary Economics*, 50, 525–546.

BERNANKE, B. S., J. BOIVIN, AND P. S. ELIASZ (2005): "Measuring the Effects of Monetary Policy: A Factor-Augmented Vector Autoregressive (FAVAR) Approach," *The Quarterly Journal of Economics*, 120, 387–422.

BLANCHARD, O., AND D. QUAH (1989): "The Dynamic Effects of Aggregate Demand and Aggregate Supply Disturbances," *American Economic Review*, 79, 665–673.

BOIVIN, J., AND S. NG (2005): "Understanding and Comparing Factor-Based Forecasts," *International Journal of Central Banking*, 1, 117–151.

CHAMBERLAIN, G., AND M. ROTHSCHILD (1983): "Arbitrage, Factor Structure, and Mean–Variance Analysis on Large Asset Markets," *Econometrica*, 51, 1281–304.

CONNOR, G., AND R. A. KORAJCZYK (1986): "Performance measurement with the arbitrage pricing theory. A new framework for analysis," *Journal of Financial Economics*, 15, 373–394.

CRISTADORO, R., M. FORNI, L. REICHLIN, AND G. VERONESE (2005): "A core inflation indicator for the euro area," *Journal of Money Credit and Banking*, 37, 539–560.

DAHLHAUS, T. (2017): "Conventional monetary policy transmission during financial crises: an empirical analysis," *Journal of Applied Econometrics*, 32, 401–421.

DEMPSTER, A. P., N. M. LAIRD, AND D. B. RUBIN (1977): "Maximum likelihood from incomplete data via the EM algorithm," *Journal of the Royal Statistical Society. Series B (methodological)*, pp. 1–38.

DIEBOLD, F. X., AND M. NERLOVE (1989): "The dynamics of exchange rate volatility: a multivariate latent factor ARCH model," *Journal of Applied Econometrics*, 4, 1–21.

DIEBOLD, F. X., AND K. YILMAZ (2014): "On the network topology of variance decompositions: Measuring the connectedness of financial firms," *Journal of Econometrics*, forthcoming.

DOZ, C., D. GIANNONE, AND L. REICHLIN (2011): "A two–step estimator for large approximate dynamic factor models based on Kalman filtering," *Journal of Econometrics*, 164, 188–205.

——— (2012): "A quasi maximum likelihood approach for large approximate factor models," *The Review of Economics and Statistics*, 94, 1014–1024.

DURBIN, J., AND S. J. KOOPMAN (2012): *Time series analysis by state space methods*. Oxford University Press.

FAN, J., Y. LIAO, AND M. MINCHEVA (2013): "Large covariance estimation by thresholding principal orthogonal complements," *Journal of the Royal Statistical Society, Series B*, 75, 603–680.

FORNI, M., AND L. GAMBETTI (2010): "The Dynamic Effects of Monetary Policy: A Structural Factor Model Approach," *Journal of Monetary Economics*, 57, 203–216.

FORNI, M., L. GAMBETTI, M. LIPPI, AND L. SALA (2014): "Noisy News in Business Cycles," mimeo.

FORNI, M., D. GIANNONE, M. LIPPI, AND L. REICHLIN (2009): "Opening the Black Box: Structural Factor Models versus Structural VARs," *Econometric Theory*, 25, 1319–1347.

FORNI, M., M. HALLIN, M. LIPPI, AND L. REICHLIN (2000): "The Generalized Dynamic Factor Model: Identification and Estimation," *The Review of Economics and Statistics*, 82, 540–554.

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

——— (2005): "The Generalized Dynamic Factor Model: One-Sided Estimation and Forecasting," *Journal of the American Statistical Association*, 100, 830–840.

FORNI, M., M. HALLIN, M. LIPPI, AND P. ZAFFARONI (2015a): "Dynamic factor models with infinite–dimensional factor spaces: Asymptotic results," .

——— (2015b): "Dynamic factor models with infinite–dimensional factor spaces: One–sided representations," *Journal of Econometrics*, forthcoming.

FORNI, M., AND M. LIPPI (2001): "The Generalized Dynamic Factor Model: Representation Theory," *Econometric Theory*, 17, 1113–1141.

——— (2011): "The General Dynamic Factor Model: One–sided representation results," *Journal of Econometrics*, 163, 23–28.

GEWEKE, J. (1977): "The Dynamic Factor Analysis of Economic Time Series," in *Latent Variables in Socio-Economic Models*, ed. by D. J. Aigner, and A. S. Goldberger. North Holland, Amsterdam.

GIANNONE, D., L. REICHLIN, AND L. SALA (2005): "Monetary Policy in Real Time," in *NBER Macroeconomics Annual 2004*, ed. by M. Gertler, and K. Rogoff. MIT Press.

GIANNONE, D., L. REICHLIN, AND D. SMALL (2008): "Nowcasting: The real–time informational content of macroeconomic data," *Journal of Monetary Economics*, 55, 665–676.

HALLIN, M., AND M. LIPPI (2013): "Factor models in high–dimensional time series. A time-domain approach," *Stochastic Processes and their Applications*, 123, 2678–2695.

HALLIN, M., AND R. LIŠKA (2007): "Determining the number of factors in the general dynamic factor model," *Journal of the American Statistical Association*, 102, 603–617.

HAN, X. (2018): "Estimation and inference of dynamic structural factor models with over-identifying restrictions," *Journal of Econometrics*, 202, 125–147.

HARVEY, A., E. RUIZ, AND E. SENTANA (1992): "Unobserved component time series models with ARCH disturbances," *Journal of Econometrics*, 52, 129–157.

HARVEY, A. C. (1990): *Forecasting, structural time series models and the Kalman filter*. Cambridge university press.

JÖRESKOG, K. G. (1969): "A general approach to confirmatory maximum likelihood factor analysis," *Psychometrika*, 34, 183–202.

JUVENAL, L., AND I. PETRELLA (2015): "Speculation in the oil market," *Journal of Applied Econometrics*, 30, 621–649.

KALMAN, R. E. (1960): "A new approach to linear filtering and prediction problems," *Journal of Basic Engineering*, 82, 35–45.

KALMAN, R. E., AND R. S. BUCY (1961): "New results in linear filtering and prediction theory," *Journal of Basic Engineering*, 83, 95–108.

LAWLEY, D., AND A. MAXWELL (1962): "Factor analysis as a statistical method," *Journal of the Royal Statistical Society. Series D (The Statistician)*, 12, 209–229.

LAWLEY, D. N., AND A. E. MAXWELL (1971): *Factor Analysis as a Statistical Method*. Butterworths, London.

LIPPI, M., AND L. REICHLIN (1993): "The Dynamic Effects of Aggregate Demand and Supply Disturbances: Comment," *American Economic Review*, 83, 644–652.

——— (1994): "VAR analysis, nonfundamental representations, Blaschke matrices," *Journal of Econometrics*, 63, 307–325.

LUCIANI, M. (2014): "Forecasting with Approximate Dynamic Factor Models: the role of non-pervasive shocks," *International Journal of Forecasting*, 30, 20–29.

——— (2015): "Monetary Policy and the Housing Market: A Structural Factor Analysis," *Journal of Applied Econometrics*, 30, 199–218.

MARIANO, R. S., AND Y. MURASAWA (2003): "A new coincident index of business cycles based on monthly and quarterly series," *Journal of Applied Econometrics*, 18, 427–443.

MENG, X.-L., AND D. B. RUBIN (1993): "Maximum likelihood estimation via the ECM algorithm: A general framework," *Biometrika*, 80, 267–278.

——— (1994): "On the global and componentwise rates of convergence of the EM algorithm," *Linear Algebra and its Applications*, 199, 413–425.

MOORE, J. B., AND B. ANDERSON (1979): *Optimal filtering*. Prentice-Hall.

ONATSKI, A. (2009): "Testing Hypotheses about the Number of Factors in Large Factor Models," *Econometrica*, 77, 1447–1479.

ONATSKI, A. (2010): "Determining the Number of Factors from Empirical Distribution of Eigenvalues," *The Review of Economics and Statistics*, 92, 1004–1016.

ONATSKI, A. (2012): "Asymptotics of the principal components estimator of large factor models with weakly influential factors," *Journal of Econometrics*, 168, 244–258.

QUAH, D., AND T. J. SARGENT (1993): "A dynamic index model for large cross sections," in *Business cycles, indicators and forecasting*, pp. 285–310. University of Chicago Press.

SARGENT, T. J., AND C. A. SIMS (1977): "Business cycle modeling without pretending to have too much a-priori economic theory," in *New Methods in Business Cycle Research*, ed. by C. A. Sims. Federal Reserve Bank of Minneapolis, Minneapolis.

SENTANA, E., G. CALZOLARI, AND G. FIORENTINI (2008): "Indirect estimation of large conditionally heteroskedastic factor models, with an application to the Dow 30 stocks," *Journal of Econometrics*, 146, 10–25.

SHUMWAY, R. H., AND D. S. STOFFER (1982): "An approach to time series smoothing and forecasting using the EM algorithm," *Journal of Time Series Analysis*, 3, 253–264.

SIMS, C. A. (1980): "Macroeconomics and Reality," *Econometrica*, 48, 1–48.

STOCK, J. H., AND M. W. WATSON (2002a): "Forecasting using principal components from a large number of predictors," *Journal of the American Statistical Association*, 97, 1167–1179.

——— (2002b): "Macroeconomic forecasting using diffusion indexes," *Journal of Business and Economic Statistics*, 20, 147–162.

——— (2005): "Implications of Dynamic Factor Models for VAR Analysis," Working Paper 11467, NBER.

TIPPING, M. E., AND C. M. BISHOP (1999): "Probabilistic principal component analysis," *Journal of the Royal Statistical Society: Series B*, 61, 611–622.

TRAPANI, L. (2017): "A randomised sequential procedure to determine the number of factors," *Journal of the American Statistical Association*, (forthcoming).

UHLIG, H. (2005): "What are the effects of monetary policy on output? Results from an agnostic identification procedure," *Journal of Monetary Economics*, 52, 381–419.

WATSON, M. W., AND R. F. ENGLE (1983): "Alternative algorithms for the estimation of dynamic factor, mimic and varying coefficient regression models," *Journal of Econometrics*, 23, 385–400.

WU, C. J. (1983): "On the convergence properties of the EM algorithm," *The Annals of statistics*, 11, 95–103.

WU, W. B., AND P. ZAFFARONI (2018): "Asymptotic theory for spectral density estimates of general multivariate time series," *Econometric Theory*, 34, 1–22.