Dynamic Factor Models

Matteo Barigozzi[†]

[†]Università di Bologna

CREST
March 2024

Outline

- Introduction
- Taxonomy of Factor Models
- Approximate Factor Model - Identification
- Principal Components Analysis
- The Likelihood
- Approximate Static Factor Model - Quasi Maximum Likelihood
- Approximate Dynamic Factor Model - Expectation Maximization
- Generalized Dynamic Factor Model
- Applications and Extensions

- **Introduction**

- Factor analysis is one of the earliest proposed multivariate statistical techniques.
- It dates back to the studies of Spearman (1904) in experimental psychology.
- Main idea:

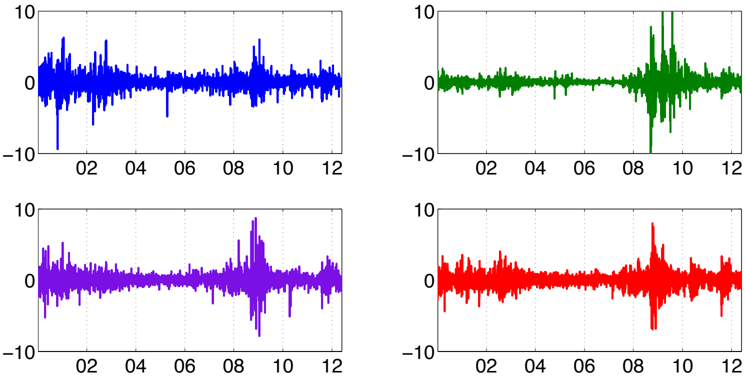
  a vector of n observed random variables/time series decomposed into the sum of

  1. few, less than n, latent factors
     - capturing co-movements;
  2. many idiosyncratic factors
     - capturing item specific or local features or measurement errors.
- We can retrospectively consider factor analysis as a pioneering technique in the filed of unsupervised statistical learning.

Examples:

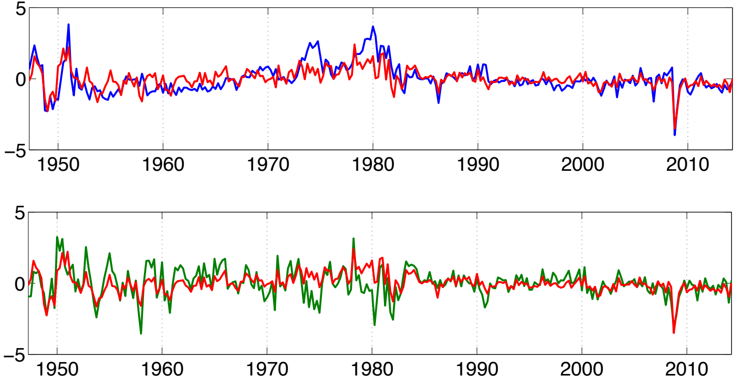
- equity returns are driven by few factors representing the “market” plus some factors specific of a given company or sector;
- GDP or inflation are driven by few factors representing the “business cycle” plus some measurement errors.

Finance example stock returns:



Blue: IBM;
Green: AIG;
Purple: Goldman Sachs;
Red: S&P500 (weighted average) capturing the co-movements.

Macro example:



Blue: CPI quarterly inflation;
Green: GDP quarterly growth rate;
Red: Average of GDP and CPI capturing the co-movements.

Main intuition:

CO-MOVEMENTS ARE CAPTURED BY
AGGREGATING THE DATA (DYNAMICALLY)
i.e. BY CROSS-SECTIONAL (WEIGHTED*) AVERAGES!

(* the weights are selected starting from the data, not a priori.)

IN LARGE SYSTEMS BY FOCUSSING ON CO-MOVEMENTS
WE ACHIEVE DIMENSION REDUCTION!

Features of large datasets of time series available today:

- number of periods for which we have data is limited and constrained by passage of time;
- more and more time series are collected and made available by statistical agencies;
- we denote by
    - T the the sample size, points in time;
    - n the number of series;
- we are in a setting where $n \simeq T$ or even $n > T$:
    - hard problem in statistics: high-dimensional setting;
- in macro $n \simeq 100, 1000$ and $T \simeq 100, 1000$ (quarterly or monthly series);
- in finance $n \simeq 100, 1000$ and $T \simeq 1000, 10000$ (daily series).
- (moderately) big data!

Two main fields of applications:

1. psychometrics in a low-dimensional setting (Spearman, 1904);
2. econometrics in a low- and high-dimensional setting with applications to
   - the analysis of financial markets (Connor, Korajczyk & Linton, 2006; Aït-Sahalia & Xiu, 2017; Barigozzi & Hallin, 2020);
   - the measurement and prediction of macroeconomic aggregates (De Mol, Giannone & Reichlin, 2008; Giannone, Reichlin & Small, 2008; Barigozzi & Luciani, 2021);
   - the study of the dynamic effects of unexpected shocks to the economy (Bernanke, Boivin & Eliasz, 2005; Forni & Gambetti, 2010; Barigozzi, Lippi & Luciani, 2021);
   - the analysis of demand systems (Stone, 1945; Barigozzi & Moneta, 2014).

A Google search on "Dynamic Factor Model" brings no less than 435 million entries–as many "as the stars of the heaven and as the sand which is upon the seashore!"

- **Taxonomy of Factor Models**

- We model a panel of n time series $\{\mathbf{x}_t = (x_{1t} \cdots x_{nt})',\ t \in \mathbb{Z}\}$ as

$$x_{it} = \chi_{it} + \xi_{it},$$

  where

  - χ_{it} **common** component, i.e. driven by factors common to all x_i's;
  - ξ_{it} **idiosyncratic** component;
  - $\mathsf{Cov}(\chi_{it}, \xi_{js}) = 0$ for any i, j, t, s (orthogonal at all leads and lags).

- Throughout, for simplicity we work with centered data so $\mathsf{E}[\chi_{it}] = \mathsf{E}[\xi_{it}] = 0$.
- We assume weak stationarity of $\{\mathbf{x}_t,\ t \in \mathbb{Z}\}$.

- There are different kind of factor models:
  - Exact vs. **Approximate**, this refers to idiosyncratic components;
  - Static vs. **Dynamic**, this refers to common components.

Exact vs. Approximate.

Let $\boldsymbol{\xi}_t = (\xi_{1t} \cdots \xi_{nt})'$.

- Exact: the elements of $\boldsymbol{\xi}_t$ are not correlated:
  - $\boldsymbol{\Gamma}^{\xi} = \mathsf{E}[\boldsymbol{\xi}_t \boldsymbol{\xi}_t']$ is diagonal;
- Approximate: mild cross-sectional correlations are allowed:
  - $\boldsymbol{\Gamma}^{\xi} = \mathsf{E}[\boldsymbol{\xi}_t \boldsymbol{\xi}_t']$ is not diagonal;

The distinction is about contemporaneous correlations only.

About autocorrelations:

- exact model: natural to assume also $\boldsymbol{\Gamma}_k^{\xi} = \mathsf{E}[\boldsymbol{\xi}_t \boldsymbol{\xi}_{t-k}'] = \mathbf{0}_{n\times n}$ for all $k \neq 0$.
- approximate model: we can allow for $\boldsymbol{\Gamma}_k^{\xi} = \mathsf{E}[\boldsymbol{\xi}_t \boldsymbol{\xi}_{t-k}'] \neq \mathbf{0}_{n\times n}$ for some $k \neq 0$, or even for all $k \in \mathbb{Z}$ provided we control for serial dependence.

The term generalized is used only for the dynamic case and only under certain additional conditions.

- Classical factor analysis considers an exact model, n is small and fixed;
- In an exact model we can estimate the loadings even if n fixed, but the factors are not estimated consistently, unless $n \to \infty$;
- In a high-dimensional setting, $n \to \infty$, an exact model is not realistic;
- Modern factor analysis considers the approximate model $\Rightarrow$ curse of dimensionality;
- An approximate model can be identified and estimated only if $n \to \infty \Rightarrow$ blessing of dimensionality;
- The condition on mild idiosyncratic cross-sectional correlations must depend on n. The most common are:
  - $\sup_{n\in\mathbb{N}} \mu_1^\xi < M$, with μ_1^ξ the max eigenvalue of $\mathbf{\Gamma}^\xi$;
  - $\sup_{n\in\mathbb{N}} n^{-1} \sum_{i,j=1}^n |\mathsf{E}[\xi_{it}\xi_{jt}]| < M$;
  - $\sup_{n\in\mathbb{N}} \max_{i=1,\dots,n} \sum_{j=1}^n |\mathsf{E}[\xi_{it}\xi_{jt}]| < M$;
  - $|\mathsf{E}[\xi_{it}\xi_{jt}]| \le M_{ij}$ s.t. $\sup_{n\in\mathbb{N}} \sum_{i=1}^n M_{ij} < M$ and $\sup_{n\in\mathbb{N}} \sum_{j=1}^n M_{ij} < M$.

Ex: static 1-factor model:

$$x_{it} = F_t + \xi_{it},$$

Consider an exact homoskedastic static factor model, then as $n \to \infty$,

$$\mathsf{E}\left[\left(\frac{1}{n}\sum_{i=1}^{n} x_{it} - F_t\right)^2\right] = \mathsf{E}\left[\left(\frac{1}{n}\sum_{i=1}^{n} \xi_{it}\right)^2\right] = \frac{1}{n^2}\sum_{i=1}^{n} \mathsf{E}[\xi_{it}^2] = \frac{\mathsf{E}[\xi_{it}^2]}{n} \to 0.$$

Under heteroskedasticity

$$\mathsf{E}\left[\left(\frac{1}{n}\sum_{i=1}^{n} \xi_{it}\right)^2\right] = \frac{1}{n^2}\sum_{i=1}^{n} \mathsf{E}[\xi_{it}^2] \leq \frac{\max_{i=1,\dots,n} \mathsf{E}[\xi_{it}^2]}{n} \to 0.$$

We need $n \to \infty$ to consistently estimate the factors. Classically n fixed and factors are incidental parameters.

Ex: static 1-factor model (cont.):

$$x_{it} = F_t + \xi_{it},$$

The same argument would hold also for an approximate model as long as

$$\mathsf{E}\left[\left(\frac{1}{n}\sum_{i=1}^{n}\xi_{it}\right)^2\right] = \frac{1}{n^2}\sum_{i,j=1}^{n}\mathsf{E}[\xi_{it}\xi_{jt}] = \frac{\boldsymbol{\iota}'\boldsymbol{\Gamma}^{\xi}\boldsymbol{\iota}}{n^2} \leq \frac{\max_{\boldsymbol{v}:\boldsymbol{v}'\boldsymbol{v}=1}\boldsymbol{v}'\boldsymbol{\Gamma}^{\xi}\boldsymbol{v}}{n} = \frac{\mu_1^{\xi}}{n} \to 0,$$

where $\boldsymbol{\iota} = (1\cdots 1)'$.

The max eigenvalue of $\boldsymbol{\Gamma}^{\chi} = \boldsymbol{\iota}\mathsf{E}[F_t^2]\boldsymbol{\iota}'$ is $\mu_1^{\chi} = n\mathsf{E}[F_t^2]$.

As $n \to \infty$ eigengap increases: we can identify the common component, and we can recover the factors. $\Rightarrow$ **blessing of dimensionality!**

Static vs. Dynamic.

- Static:

$$x_{it} = \underbrace{\boldsymbol{\lambda}_i' \mathbf{F}_t}_{\chi_{it}} + \xi_{it}, \tag{1}$$

the factors $\mathbf{F}_t$ and the loadings $\boldsymbol{\lambda}_i$ are r-dimensional vectors with $r < n$. $\mathbf{F}_t$ have only a contemporaneous effect on x_{it} and are called static factors.

- Dynamic:

$$x_{it} = \underbrace{\sum_{k=0}^{s} \boldsymbol{\lambda}_{ki}^{*\prime} \mathbf{f}_{t-k}}_{\boldsymbol{\lambda}_i^{*\prime}(L)\mathbf{f}_t = \chi_{it}} + \xi_{it}, \tag{2}$$

the factors $\mathbf{f}_t$ and the loadings $\boldsymbol{\lambda}_{ki}^*$ are q-dimensional vectors with $q < n$. $\mathbf{f}_t$ have effect on x_{it} through their lags too and are called dynamic factors.

- If $s < \infty$ and ξ_{it} is the same in (1) and (2) then $q \leq r$.
- If $s = \infty$ then (2) is the most general dynamic factor model.

- Approximate static factor model

$$x_{it} = \boldsymbol{\lambda}_i' \mathbf{F}_t + \xi_{it}$$

Estimation:
Principal Components (Chamberlain & Rothschild, 1983; Stock & Watson, 2002; Bai, 2003).
Quasi Maximum Likelihood (Bai & Li, 2016).

- Exact static factor model
Estimation:
Principal Components (Hotelling, 1933).
Maximum Likelihood (Thomson, 1936; Bartlett, 1937; Lawley, 1940; Anderson & Rubin, 1956; Jöreskog, 1969; Lawley & Maxwell, 1971; Amemiya, Fuller & Pantula, 1987; Tipping & Bishop, 1999; Bai & Li, 2012).

- Approximate dynamic factor model (DFM)

$$x_{it} = \boldsymbol{\lambda}_i' \mathbf{F}_t + \xi_{it},$$
$$\mathbf{F}_t = \mathbf{N}(L)\mathbf{u}_t.$$

Estimation:
Principal Components plus VAR (Forni, Giannone, Lippi & Reichlin, 2009).
Principal Components plus Kalman smoother (Doz, Giannone & Reichlin, 2011).
Expectation Maximization algorithm (Watson & Engle, 1983; Quah & Sargent, 1993; Doz, Giannone & Reichlin, 2012; Barigozzi & Luciani, 20xx).

- Approximate dynamic factor model (DFM)

$$x_{it} = \boldsymbol{\lambda}_i' \mathbf{F}_t + \xi_{it},$$
$$\mathbf{F}_t = \mathbf{A}\mathbf{F}_{t-1} + \mathbf{H}\mathbf{u}_t.$$

Estimation:
Principal Components plus VAR (Forni, Giannone, Lippi & Reichlin, 2009).
Principal Components plus Kalman smoother (Doz, Giannone & Reichlin, 2011).
Expectation Maximization algorithm (Watson & Engle, 1983; Quah & Sargent, 1993; Doz, Giannone & Reichlin, 2012; Barigozzi & Luciani, 20xx).

- Restricted generalized dynamic factor model (GDFM)

$$x_{it} = \sum_{k=0}^{s} \boldsymbol{\lambda}_{ki}^{*'} \boldsymbol{f}_{t-k} + \xi_{it},$$
$$\boldsymbol{f}_t = \mathbf{G}(L)\mathbf{u}_t$$

  Estimation:
  Spectral Principal Components plus Principal Components (Forni, Hallin, Lippi & Reichlin, 2005).

- Exact dynamic factor model
  Estimation:
  Spectral Expectation Maximization algorithm (Sargent & Sims, 1977).

- Restricted generalized dynamic factor model (GDFM)

$$x_{it} = \boldsymbol{\lambda}_i^{*'}(L)\boldsymbol{f}_t + \xi_{it},$$
$$\boldsymbol{f}_t = \boldsymbol{\Phi}\boldsymbol{f}_{t-1} + \mathbf{u}_t.$$

Estimation:
Spectral Principal Components plus Principal Components (Forni, Hallin, Lippi & Reichlin, 2005).

- Exact dynamic factor model
Estimation:
Spectral Expectation Maximization algorithm (Sargent & Sims, 1977).

- Unrestricted generalized dynamic factor model (GDFM)

$$x_{it} = \sum_{k=0}^{\infty} \boldsymbol{\lambda}_{ki}^{*'} \boldsymbol{f}_{t-k} + \xi_{it},$$
$$\boldsymbol{f}_t = \mathbf{G}(L)\mathbf{u}_t$$

Estimation:
Spectral Principal Components (Forni, Hallin, Lippi & Reichlin, 2000).
Spectral Principal Components plus VAR (Forni, Hallin, Lippi & Zaffaroni, 2017; Barigozzi, Hallin, Luciani & Zaffaroni, 2023).

- Unrestricted generalized dynamic factor model (GDFM)

$$x_{it} = b_i^{'}(L)\mathbf{u}_t + \xi_{it},$$

Estimation:
Spectral Principal Components (Forni, Hallin, Lippi & Reichlin, 2000).
Spectral Principal Components plus VAR (Forni, Hallin, Lippi & Zaffaroni, 2017; Barigozzi, Hallin, Luciani & Zaffaroni, 2023).

- Compare the approximate DFM with the unrestricted GDFM

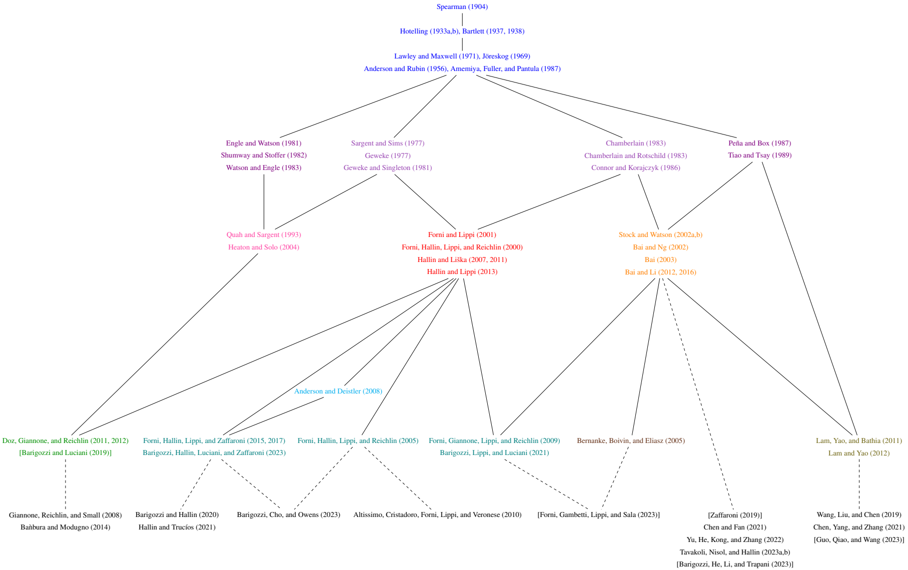
$$\text{(A) } x_{it} = \boldsymbol{\lambda}_i'\mathbf{F}_t + \xi_{it}, \qquad \text{(B) } x_{it} = \boldsymbol{\lambda}_i^{*'}(L)\boldsymbol{f}_t + \xi_{it},$$
$$\mathbf{F}_t = \mathbf{A}\mathbf{F}_{t-1} + \mathbf{H}\mathbf{u}_t, \qquad \boldsymbol{f}_t = \boldsymbol{\Phi}\boldsymbol{f}_{t-1} + \mathbf{u}_t.$$

- Let $\mathbf{F}_t = (\boldsymbol{f}_t' \cdots \boldsymbol{f}_{t-s}')'$ s.t. $r = q(s+1) \geq q$, then (B) reads (say $s = 1$)

$$x_{it} = [\boldsymbol{\lambda}_{0i}^{*'}\ \boldsymbol{\lambda}_{1i}^{*'}]\ \mathbf{F}_t + \xi_{it},$$
$$\mathbf{F}_t = \begin{pmatrix} \boldsymbol{\Phi} & \mathbf{0}_{q\times q} \\ \mathbf{I}_q & \mathbf{0}_{q\times q} \end{pmatrix} \mathbf{F}_{t-1} + \begin{pmatrix} \mathbf{I}_q \\ \mathbf{0}_{q\times q} \end{pmatrix} \mathbf{u}_t$$

- Estimating (A) is not equivalent to estimating (B). We can find the loadings and factors in (A) if we know (B), but not the viceversa!

Taxonomy of Factor Models



Source: Barigozzi and Hallin, 2024.

- Scalar notation ($i = 1, \ldots, n$ and $t = 1, \ldots, T$):

$$x_{it} = \underbrace{\underbrace{\boldsymbol{\lambda}_i'}_{1\times r}\underbrace{\mathbf{F}_t}_{r\times 1}}_{\chi_{it}} + \xi_{it}.$$

- Vector notation ($i = 1, \ldots, n$ or $t = 1, \ldots, T$):

$$\underbrace{\mathbf{x}_t}_{n\times 1} = \underbrace{\underbrace{\boldsymbol{\Lambda}}_{n\times r}\underbrace{\mathbf{F}_t}_{r\times 1}}_{\boldsymbol{\chi}_t} + \underbrace{\boldsymbol{\xi}_t}_{n\times 1}, \qquad \underbrace{\boldsymbol{x}_i}_{T\times 1} = \underbrace{\underbrace{\boldsymbol{F}}_{T\times r}\underbrace{\boldsymbol{\lambda}_i}_{r\times 1}}_{\boldsymbol{\chi}_i} + \underbrace{\boldsymbol{\zeta}_i}_{T\times 1}.$$

- Matrix notation:

$$\underbrace{\boldsymbol{X}}_{T\times n} = \underbrace{\underbrace{\boldsymbol{F}}_{T\times r}\underbrace{\boldsymbol{\Lambda}'}_{r\times n}}_{\boldsymbol{C}} + \underbrace{\boldsymbol{\Xi}}_{T\times n}.$$

- Stacked notation:

$$\underbrace{\boldsymbol{\mathcal{X}}}_{nT\times 1} = \underbrace{\underbrace{\boldsymbol{\mathcal{L}}}_{(\boldsymbol{\Lambda}\otimes\mathbf{I}_T)}}_{nT\times rT}\underbrace{\boldsymbol{\mathcal{F}}}_{rT\times 1} + \underbrace{\boldsymbol{\mathcal{E}}}_{nT\times 1}.$$

- **Approximate Factor Model - Identification**

Weighted averages. Large n to recover factors.

- Take any $n \times r$ weight matrix $\boldsymbol{W}_F = (\boldsymbol{w}_{F,1} \cdots \boldsymbol{w}_{F,n})'$ and such that

$$n^{-1}\boldsymbol{W}_F'\boldsymbol{\Lambda} = \mathbf{K} \succ 0, \qquad n^{-1}\boldsymbol{W}_F'\boldsymbol{W}_F = \mathbf{I}_r$$

  and $\|\boldsymbol{w}_{F,i}\| \le c$ for some $c > 0$ independent of i.

- For any given t an estimator of a linear combination of the factors is

$$\check{\mathbf{F}}_t = \frac{\boldsymbol{W}_F'\mathbf{x}_t}{n} = \frac{\boldsymbol{W}_F'\boldsymbol{\Lambda}\mathbf{F}_t}{n} + \frac{\boldsymbol{W}_F'\boldsymbol{\xi}_t}{n} = \mathbf{K}\mathbf{F}_t + \frac{1}{n}\sum_{i=1}^n \boldsymbol{w}_{F,i}'\xi_{it}.$$

- Then we have $\sqrt{n}$-consistency if as $n \to \infty$ (assume $r = 1$ for simplicity):

$$\mathsf{E}\left[\left|\frac{1}{n}\sum_{i=1}^n w_{F,i}\xi_{it}\right|^2\right] \le \begin{cases} \frac{c^2}{n}\frac{\boldsymbol{\iota}'\boldsymbol{\Gamma}^\xi\boldsymbol{\iota}}{n} \le \frac{c^2}{n}\mu_1^\xi = O\left(\frac{1}{n}\right), \\ \text{or} \\ \frac{c^2}{n}\left(\frac{1}{n}\sum_{i=1}^n\sum_{j=1}^n |\mathsf{E}[\xi_{it}\xi_{jt}]|\right) = O\left(\frac{1}{n}\right), \end{cases}$$

  which are standard assumptions in approximate factor model.

- It is enough to have $n^{-1}\boldsymbol{W}_F'\boldsymbol{\Lambda} \to \mathbf{K}$ and $n^{-1}\boldsymbol{W}_F'\boldsymbol{W}_F \to \mathbf{I}_r$ as $n \to \infty$.

Weighted averages. Large n to recover factors. Example.

- For known $\mathbf{\Lambda}$, the OLS estimator of the factors is, for any given t,

$$\mathbf{F}_t^{\mathsf{OLS}} = (\mathbf{\Lambda}'\mathbf{\Lambda})^{-1}\mathbf{\Lambda}'\mathbf{x}_t = (\mathbf{\Lambda}'\mathbf{\Lambda})^{-1}\mathbf{\Lambda}'(\mathbf{\Lambda}\mathbf{F}_t + \boldsymbol{\xi}_t)$$
$$= \mathbf{F}_t + \left(\frac{1}{n}\sum_{i=1}^n \boldsymbol{\lambda}_i\boldsymbol{\lambda}_i'\right)^{-1}\left(\frac{1}{n}\sum_{i=1}^n \boldsymbol{\lambda}_i\xi_{it}\right).$$

- For consistency it is enough that, as $n \to \infty$:
  1. $\frac{1}{n}\sum_{i=1}^n \boldsymbol{\lambda}_i\xi_{it} \to_p \mathbf{0}_r$;
  2. $\frac{1}{n}\sum_{i=1}^n \boldsymbol{\lambda}_i\boldsymbol{\lambda}_i' = \frac{\mathbf{\Lambda}'\mathbf{\Lambda}}{n} \to \mathbf{\Sigma}_\Lambda \succ 0$;

  and 1 is ensured by $\|\boldsymbol{\lambda}_i\| \le M_\lambda$ plus weak cross-sectional dependence of idiosyncratic components:

$$\sup_{n\in\mathbb{N}} \frac{1}{n}\sum_{i=1}^n\sum_{j=1}^n |\mathsf{E}[\xi_{it}\xi_{jt}]| \le M_\xi,$$

- This is equivalent to choose the optimal unfeasible weights $\boldsymbol{W}_F = n\mathbf{\Lambda}(\mathbf{\Lambda}'\mathbf{\Lambda})^{-1}$, then $\mathbf{K} = n^{-1}\boldsymbol{W}_F'\mathbf{\Lambda} = \mathbf{I}_r$.

Weighted averages. Large T to recover loadings.

- Take any $T \times r$ weight matrix $\boldsymbol{W}_\Lambda = (\boldsymbol{w}_{\Lambda,1} \cdots \boldsymbol{w}_{\Lambda,T})'$ and such that

$$T^{-1}\boldsymbol{W}_\Lambda'\boldsymbol{F} = \mathbf{K} \succ 0, \qquad T^{-1}\boldsymbol{W}_\Lambda'\boldsymbol{W}_\Lambda = \mathbf{I}_r$$

  and $\|\boldsymbol{w}_{\Lambda,t}\| \le c$ for some $c > 0$ independent of t.

- For any given i an estimator of a linear combination of the loadings is

$$\check{\boldsymbol{\lambda}}_i = \frac{\boldsymbol{W}_\Lambda'\boldsymbol{x}_i}{T} = \frac{\boldsymbol{W}_\Lambda'\boldsymbol{F}\boldsymbol{\lambda}_i}{T} + \frac{\boldsymbol{W}_\Lambda'\boldsymbol{\zeta}_i}{T} = \mathbf{K}\boldsymbol{\lambda}_i + \frac{1}{T}\sum_{t=1}^{T}\boldsymbol{w}_{\Lambda,t}'\xi_{it}.$$

- Then we have $\sqrt{T}$-consistency if as $T \to \infty$ (assume $r = 1$ for simplicity):

$$\mathsf{E}\left[\left|\frac{1}{T}\sum_{t=1}^{T} w_{\Lambda,t}\xi_{it}\right|^2\right] \le \frac{c^2}{T}\left(\frac{1}{T}\sum_{t=1}^{T}\sum_{s=1}^{T}|\mathsf{E}[\xi_{it}\xi_{is}]|\right) = O\left(\frac{1}{T}\right),$$

  which is a standard assumption for stationary time series.

- It is enough to have $T^{-1}\boldsymbol{W}_\Lambda'\boldsymbol{F} \to \mathbf{K}$ and $T^{-1}\boldsymbol{W}_\Lambda'\boldsymbol{W}_\Lambda \to \mathbf{I}_r$ as $T \to \infty$.

Weighted averages. Large T to recover factors. Example.

- For known $\boldsymbol{F}$, the OLS estimator of the loadings is, for any given i,

$$\boldsymbol{\lambda}_i^{\mathsf{OLS}} = (\boldsymbol{F}'\boldsymbol{F})^{-1}\boldsymbol{F}'\boldsymbol{x}_i = (\boldsymbol{F}'\boldsymbol{F})^{-1}\boldsymbol{F}'(\boldsymbol{F}\boldsymbol{\lambda}_i + \boldsymbol{\zeta}_i)$$
$$= \boldsymbol{\lambda}_i + \left(\frac{1}{T}\sum_{t=1}^{T}\mathbf{F}_t\mathbf{F}_t'\right)^{-1}\left(\frac{1}{T}\sum_{t=1}^{T}\mathbf{F}_t\xi_{it}\right).$$

- For consistency it is enough that, as $T \to \infty$:
  1. $\frac{1}{T}\sum_{t=1}^{T}\mathbf{F}_t\xi_{it} \to_p \mathbf{0}_r$;
  2. $\frac{1}{T}\sum_{t=1}^{T}\mathbf{F}_t\mathbf{F}_t' = \frac{\boldsymbol{F}'\boldsymbol{F}}{T} \to_p \boldsymbol{\Gamma}^F \succ 0$;

  and 1 and 2 are ensured by standard time series assumptions: finite fourth order cumulants, strong mixing, ergodicity....plus

$$\sup_{T\in\mathbb{N}} \frac{1}{T}\sum_{t=1}^{T}\sum_{s=1}^{T}|\mathsf{E}[\xi_{it}\xi_{is}]| \le M_\xi'.$$

- This is equivalent to choose the optimal unfeasible weights $\boldsymbol{W}_\Lambda = T\boldsymbol{F}(\boldsymbol{F}'\boldsymbol{F})^{-1}$, then $\mathbf{K} = T^{-1}\boldsymbol{W}_\Lambda'\boldsymbol{F} = \mathbf{I}_r$.

Identification problem.

- We can always rewrite the model as:

$$\mathbf{x}_t = \underbrace{\mathbf{\Lambda H}}_{\boldsymbol{P}} \underbrace{\mathbf{H}^{-1}\mathbf{F}_t}_{\mathbf{G}_t} + \boldsymbol{\xi}_t,$$

  for some invertible $r \times r$ matrix $\mathbf{H}$.
- To pin down $\mathbf{H}$ we need r^2 constraints.
- The common component $\boldsymbol{\chi}_t = \mathbf{\Lambda F}_t = \boldsymbol{P}\mathbf{G}_t$ is always identified.

Main assumptions.

0 $\mathsf{E}[\mathbf{F}_t] = \mathbf{0}_r$, $\mathsf{E}[\boldsymbol{\xi}_t] = \mathbf{0}_n$;

1 $\frac{\boldsymbol{F}'\boldsymbol{F}}{T} \to_p \boldsymbol{\Gamma}^F \succ 0$ as $T \to \infty$;

2 $\frac{\boldsymbol{\Lambda}'\boldsymbol{\Lambda}}{n} \to \boldsymbol{\Sigma}_\Lambda \succ 0$ as $n \to \infty$;

3 $\boldsymbol{\Gamma}^\xi \succ 0$ and $\sup_{n,T\in\mathbb{N}} \frac{1}{nT} \sum_{i=1}^n \sum_{j=1}^n \sum_{t=1}^T \sum_{s=1}^T |\mathsf{E}[\xi_{it}\xi_{js}]| \leq M$;

4 finite fourth order moments of $\{\xi_{it}\}$ summable over t and i;

5 $\{\mathbf{F}_t\}$ and $\{\boldsymbol{\xi}_t\}$ are mutually independent;

6 the r eigenvalues of $\frac{\boldsymbol{\Gamma}^\chi}{n} = \frac{\boldsymbol{\Lambda}\boldsymbol{\Gamma}^F\boldsymbol{\Lambda}'}{n}$ are distinct (coincide with those of $\boldsymbol{\Sigma}_\Lambda\boldsymbol{\Gamma}^F$);

7 CLTs, as $n, T \to \infty$,

$$\frac{1}{\sqrt{n}} \sum_{i=1}^n \boldsymbol{\lambda}_i \xi_{it} \to_d \mathcal{N}(\mathbf{0}_r, \boldsymbol{\Gamma}_t), \qquad \frac{1}{\sqrt{T}} \sum_{t=1}^T \mathbf{F}_t \xi_{it} \to_d \mathcal{N}(\mathbf{0}_r, \boldsymbol{\Phi}_i).$$

Alternatively to A.1 we can make assumptions on the process $\{\mathbf{F}_t\}$ such that

$$\mathsf{E}\left[\left\|\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\left\{\mathbf{F}_t\mathbf{F}_t' - \mathbf{\Gamma}^F\right\}\right\|^2\right] \leq M$$

e.g. assume finite fourth order moments of $\{\mathbf{F}_t\}$ summable over t.

Alternatively to A.2 and part of A.3 we can assume

2' largest r eigenvalues of $\mathbf{\Gamma}^\chi$ diverge (linearly) as $n \to \infty$

$$\underline{c}_j \leq \liminf_{n\to\infty} \frac{\mu_j^\chi}{n} \leq \limsup_{n\to\infty} \frac{\mu_j^\chi}{n} \leq \bar{c}_j, \quad j = 1, \dots, r$$

3' largest eigenvalue of $\mathbf{\Gamma}^\xi$ is bounded for all n

$$\sup_{n\in\mathbb{N}} \mu_1^\xi \leq M$$

By Weyl's inequality, since $\boldsymbol{\Gamma}^x = \boldsymbol{\Gamma}^\chi + \boldsymbol{\Gamma}^\xi$,

$$\lim_{n\to\infty} \frac{\mu_j^x}{n} \geq \lim_{n\to\infty} \frac{\mu_j^\chi}{n} + \lim_{n\to\infty} \frac{\mu_n^\xi}{n} \geq \underline{c}_j, \quad j = 1, \dots, r,$$

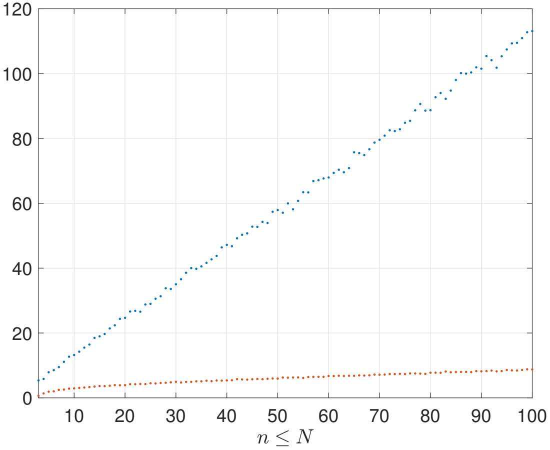
$$\lim_{n\to\infty} \frac{\mu_j^x}{n} \leq \lim_{n\to\infty} \frac{\mu_j^\chi}{n} + \lim_{n\to\infty} \frac{\mu_1^\xi}{n} \leq \bar{c}_j, \quad j = 1, \dots, r,$$

and

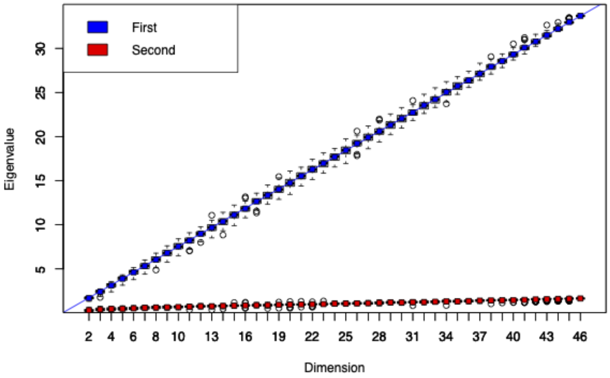
$$\sup_{n\in\mathbb{N}} \mu_j^x \leq \sup_{n\in\mathbb{N}} \mu_{r+1}^\chi + \sup_{n\in\mathbb{N}} \mu_1^\xi \leq M, \quad j = r+1, \dots, n,$$

- Eigen-gap in eigenvalues μ_j^x of $\boldsymbol{\Gamma}^x$
- As $n \to \infty$ we identify the number of factors!
- In general an observed eigen-gap is just a necessary condition to have a factor structure but it is not a sufficient condition.

Plot of μ_j^x when $r = 1$, simulated data



Plot of μ_j^x when $r = 1$, real data



We consider the classical identification conditions used in exploratory factor analysis:

1. $\frac{\mathbf{\Lambda}'\mathbf{\Lambda}}{n}$ is diagonal for all n;
2. $\frac{\boldsymbol{F}'\boldsymbol{F}}{T} = \mathbf{I}_r$ for all T;

To achieve global identification we need also to fix the sign, e.g. of one row of $\mathbf{\Lambda}$ or $\boldsymbol{F}$.

Identification of loadings.

- By SVD $\boldsymbol{\Lambda} = \boldsymbol{V}\boldsymbol{D}\boldsymbol{U}$.
- From $\boldsymbol{\Lambda}'\boldsymbol{\Lambda} = \boldsymbol{U}'\boldsymbol{D}\boldsymbol{V}'\boldsymbol{V}\boldsymbol{D}\boldsymbol{U}' = \boldsymbol{U}'\boldsymbol{D}^2\boldsymbol{U}$, and to make it diagonal we can set $\boldsymbol{U} = \mathbf{I}_r$.
- Since $\boldsymbol{\Gamma}^\chi = \mathbf{V}^\chi \mathbf{M}^\chi \mathbf{V}^{\chi'} = \boldsymbol{\Lambda}\boldsymbol{\Lambda}' = \boldsymbol{V}\boldsymbol{D}^2\boldsymbol{V}'$
  1. the columns of $\boldsymbol{V}$ span the same space as the columns of $\mathbf{V}^\chi$.
  2. $\boldsymbol{D}^2 = \mathbf{M}^\chi$.
- Therefore:
  - $\boldsymbol{\Lambda} = \mathbf{V}^\chi(\mathbf{M}^\chi)^{1/2}$ and $\frac{\boldsymbol{\Lambda}'\boldsymbol{\Lambda}}{n} = \frac{\mathbf{M}^\chi}{n}$;
  - $\boldsymbol{F} = \boldsymbol{C}\mathbf{V}^\chi(\mathbf{M}^\chi)^{-1/2}$ by linear projection of $\boldsymbol{C}$ onto $\boldsymbol{\Lambda}$;
  - $\boldsymbol{\Sigma}_\Lambda = \lim_{n\to\infty} \frac{\mathbf{M}^\chi}{n}$;
  - $\boldsymbol{\Gamma}^F = \mathbf{I}_r$.

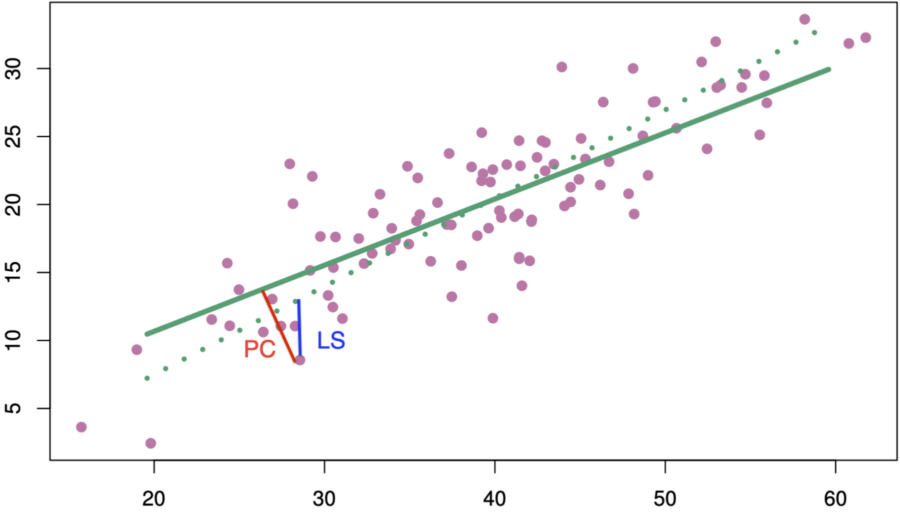
- **Principal Components Analysis**

PC for dimension reduction (Pearson, 1902).

- Assume $r = 1$. To reduce the dimension of $\boldsymbol{X}$ we look to minimize the distances between the observations and their projections onto a one dimensional subspace (line).
- the linear projection of $\mathbf{x}_t = (x_{1t} \cdots x_{nt})'$ onto $\mathbf{a} = (a_1 \cdots a_n)'$ with $\|\mathbf{a}\| = \mathbf{a}'\mathbf{a} = 1$ is $\mathbf{a}\mathbf{a}'\mathbf{x}_t$.
- We want to minimize the sum of distances between all $\mathbf{x}_t$ and their projections

$$\min_{\mathbf{a}:\mathbf{a}'\mathbf{a}=1} \sum_{t=1}^{T} \|\mathbf{x}_t - \mathbf{a}\mathbf{a}'\mathbf{x}_t\|^2 = \min_{a_i:\sum_{i=1}^{n} a_i^2=1} \sum_{t=1}^{T} \sum_{i=1}^{n} (x_{it} - a_i \mathbf{a}'\mathbf{x}_t)^2$$

- This is different from LS where we have a dependent variable, say x_{1t} and $n-1$ independent variables and we solve $\min_{b_i} \sum_{t=1}^{T}(x_{1t} - \sum_{i=2}^{n} b_i x_{it})^2$.
- In PC we minimize Euclidean distance in $\mathbb{R}^n$ in LS we minimize a distance in $\mathbb{R}$ in the subspace of the dependent variable.



Here $n = 2$ and $r = 1$.

PC for dimension reduction (cont.)

- Now, by Pythagora theorem $(\mathbf{x}_t - \mathbf{a}\mathbf{a}'\mathbf{x}_t)'\mathbf{a}\mathbf{a}'\mathbf{x}_t = 0$ (the error is orthogonal to the projection)

$$\begin{aligned}\sum_{t=1}^{T} \|\mathbf{x}_t - \mathbf{a}\mathbf{a}'\mathbf{x}_t\|^2 &= \sum_{t=1}^{T}(\mathbf{x}_t - \mathbf{a}\mathbf{a}'\mathbf{x}_t)'(\mathbf{x}_t - \mathbf{a}\mathbf{a}'\mathbf{x}_t) = \sum_{t=1}^{T}(\mathbf{x}_t - \mathbf{a}\mathbf{a}'\mathbf{x}_t)'\mathbf{x}_t \\ &= \sum_{t=1}^{T}\mathbf{x}_t'\mathbf{x}_t - \sum_{t=1}^{T}\mathbf{x}_t'\mathbf{a}\mathbf{a}'\mathbf{x}_t = \sum_{t=1}^{T}\mathbf{x}_t'\mathbf{x}_t - \sum_{t=1}^{T}\mathbf{a}'\mathbf{x}_t\mathbf{x}_t'\mathbf{a}\end{aligned}$$

- It follows that

$$\arg\min_{\mathbf{a}:\mathbf{a}'\mathbf{a}=1} \sum_{t=1}^{T} \|\mathbf{x}_t - \mathbf{a}\mathbf{a}'\mathbf{x}_t\|^2 = \arg\max_{\mathbf{a}:\mathbf{a}'\mathbf{a}=1} \sum_{t=1}^{T} \mathbf{a}'\mathbf{x}_t\mathbf{x}_t'\mathbf{a}$$

PC in high-dimensions.

- We can rewrite the maximization problem as

$$\arg \max_{\mathbf{a}:\mathbf{a}'\mathbf{a}=1} \frac{1}{nT}\mathbf{a}'\boldsymbol{X}'\boldsymbol{X}\mathbf{a}$$

- The solution is $\widehat{\mathbf{a}} = \widehat{\mathbf{V}}^x$ the leading eigenvector of $(nT)^{-1}\boldsymbol{X}'\boldsymbol{X}$ which is the same as the leading eigenvector of $T^{-1}\boldsymbol{X}'\boldsymbol{X}$ and of $\boldsymbol{X}'\boldsymbol{X}$.
- The value of the objective function at its max is $n^{-1}\widehat{\mu}_1^x$ which is finite since we rescale by n.
- The optimal linear projection $\widehat{\mathbf{V}}^{x\prime}\mathbf{x}_t$ is the 1st PC of $\boldsymbol{X}'\boldsymbol{X}$ which has variance $\widehat{\mu}_1^x$, so the 1st normalized PC is $(\widehat{\mu}_1^x)^{-1/2}\widehat{\mathbf{V}}^{x\prime}\mathbf{x}_t$.
- Note that algebraically we could exchange n and T and solve finding PCs for $\boldsymbol{X}\boldsymbol{X}'$, but this is not natural since in time series T is the sample size, not n!
- In population the PCs are defined in the same way but now the norm is a variance, so as a result we have for the weights the eigenvectors of $\boldsymbol{\Gamma}^x = \mathsf{E}[\mathbf{x}_t\mathbf{x}_t']$.

Principal components representation vs. static factor model.

- Since the eigenvectors are an orthonormal basis in $\mathbb{R}^n$, for a given r

$$x_{it} = \sum_{j=1}^{n} V_{ij}^{x} \underbrace{\left(\mathbf{V}_j^{x'} \mathbf{x}_t\right)}_{i \text{ th PC}} = \underbrace{\sum_{j=1}^{r} V_{ij}^{x} \left(\mathbf{V}_j^{x'} \mathbf{x}_t\right)}_{x_{it,[r]}} + \underbrace{\sum_{j=r+1}^{n} V_{ij}^{x} \left(\mathbf{V}_j^{x'} \mathbf{x}_t\right)}_{e_{it}}$$

- $x_{it,[r]}$ is the optimal linear r-dimensional representation of x_{it}, it is such that $\sum_{i=1}^{n} \mathsf{E}[e_{it}^2] = \text{tr}(\boldsymbol{\Gamma}^e)$ is minimum. It minimizes the sum of covariances since $(nT)^{-1} \sum_{i,j=1}^{n} \mathsf{E}[e_{it}e_{jt}] \leq \mu_1^e \leq \text{tr}(\boldsymbol{\Gamma}^e)$, but $\boldsymbol{\Gamma}^e$ is not necessarily diagonal.
- PC is a representation since no assumption is made on e_{it}.
- A static r-factor model is $x_{it} = \underbrace{\sum_{j=1}^{r} \Lambda_{ij} F_{jt}}_{\chi_{it}} + \xi_{it}$
- If the model is exact $\boldsymbol{\Gamma}^\xi$ is diagonal, and χ_{it} accounts for all covariances, but this depends on the assumptions we make. This is a statistical model.
- Under an approximate factor model the two approaches are reconciled, provided $n \to \infty$.

PC estimation of factors.

- PCs are linear combinations of the data with optimal weights. This is what we are looking for when retrieving the factors.
- Considering the weights $\boldsymbol{w}_F$ defined above such that $\boldsymbol{w}_F'\boldsymbol{w}_F = n$ the PC maximization becomes

$$\arg \max_{\boldsymbol{w}:\boldsymbol{w}_F'\boldsymbol{w}_F=n} \frac{1}{n^2T}\boldsymbol{w}_F'\boldsymbol{X}'\boldsymbol{X}\boldsymbol{w}_F$$

  so that one solution is $\widehat{\boldsymbol{w}}_F = \sqrt{n}\widehat{\mathbf{V}}^x$ and the value of the objective function at its max is still $n^{-1}\widehat{\mu}_1^x$.
- Since $\widehat{\boldsymbol{w}}_F$ are the optimal weights, they are an estimator of the unfeasible optimal weights $n(\boldsymbol{\Lambda}'\boldsymbol{\Lambda})^{-1}\boldsymbol{\Lambda}'$ so we can write $\widehat{\boldsymbol{w}}_F = n(\widehat{\boldsymbol{\Lambda}}'\widehat{\boldsymbol{\Lambda}})^{-1}\widehat{\boldsymbol{\Lambda}}'$.

PC estimation of factors (cont.).

- An estimator of the factor is the 1st normalized PC

$$\widehat{F}_t^{\mathsf{PC}} = \frac{\widehat{\mathbf{V}}^{x\prime}\mathbf{x}_t}{\sqrt{\widehat{\mu}_1^x}} = \frac{\sqrt{n}\widehat{\boldsymbol{w}}_F'\mathbf{x}_t}{\sqrt{n}\sqrt{n}\sqrt{\widehat{\mu}_1^x}} = \sqrt{\frac{n}{\widehat{\mu}_1^x}}\frac{\widehat{\boldsymbol{w}}_F'\boldsymbol{\Lambda}F_t}{n} + \sqrt{\frac{n}{\widehat{\mu}_1^x}}\frac{\widehat{\boldsymbol{w}}_F'\boldsymbol{\xi}_t}{n}$$
$$= \underbrace{\sqrt{\frac{n}{\widehat{\mu}_1^x}}(\widehat{\boldsymbol{\Lambda}}'\widehat{\boldsymbol{\Lambda}})^{-1}\widehat{\boldsymbol{\Lambda}}'\boldsymbol{\Lambda}}_{\widehat{K}} F_t + O_p\left(\frac{1}{\sqrt{n}}\right),$$

since $n^{-1}|\widehat{\mu}_1^x - \mu_1^\chi| = o_p(1)$ and $\mu_1^\chi = O(n)$ by assumption.

- If we choose $\widehat{\boldsymbol{\Lambda}} = \widehat{\mathbf{V}}^x\sqrt{\widehat{\mu}_1^x}$ then given that $\boldsymbol{\Lambda} = \mathbf{V}^\chi\sqrt{\mu_1^\chi}$,

$$\widehat{K} = \sqrt{n}(\widehat{\mu}_1^x)^{-1}\widehat{\mathbf{V}}^{x\prime}\mathbf{V}^\chi\sqrt{\mu_1^\chi} = \frac{n}{\widehat{\mu}_1^x}\widehat{\mathbf{V}}^{x\prime}\mathbf{V}^\chi\sqrt{\frac{\mu_1^\chi}{n}} = \pm 1 + o_p(1),$$

since $n^{-1}|\widehat{\mu}_1^x - \mu_1^\chi| = o_p(1)$ and $|\widehat{\mathbf{V}}^{x\prime}\mathbf{V}^\chi \pm 1| = o_p(1)$ (Davis & Kahan, 1970).

- The 1st normalized PC is a consistent estimator of F_t (the $o_p(1)$ are all $O_p(n^{-1/2}) + O_p(T^{-1/2})$).
- The common component is estimated as $\widehat{\boldsymbol{\chi}}_t = \widehat{\mathbf{V}}^x\widehat{\mathbf{V}}^{x\prime}\mathbf{x}_t$.

Least squares estimation of a static factor model:

$$\left(\widehat{\boldsymbol{\Lambda}}, \widehat{\boldsymbol{F}}\right) = \arg\min_{\underline{\boldsymbol{\Lambda}}, \underline{\boldsymbol{F}}} \frac{1}{nT} \sum_{i=1}^{n} \sum_{t=1}^{T} (x_{it} - \underline{\boldsymbol{\lambda}}_i' \underline{\mathbf{F}}_t)^2,$$

which is equivalent to

$$\min_{\underline{\boldsymbol{\Lambda}}, \underline{\boldsymbol{F}}} \frac{1}{nT} \mathrm{tr}\left\{\left(\boldsymbol{X} - \underline{\boldsymbol{F}}\,\underline{\boldsymbol{\Lambda}}'\right)\left(\boldsymbol{X} - \underline{\boldsymbol{F}}\,\underline{\boldsymbol{\Lambda}}'\right)'\right\},$$

or

$$\min_{\underline{\boldsymbol{\Lambda}}, \underline{\boldsymbol{F}}} \frac{1}{nT} \mathrm{tr}\left\{\left(\boldsymbol{X} - \underline{\boldsymbol{F}}\,\underline{\boldsymbol{\Lambda}}'\right)'\left(\boldsymbol{X} - \underline{\boldsymbol{F}}\,\underline{\boldsymbol{\Lambda}}'\right)\right\}.$$

We need to impose r^2 constraints to identify the minimum. Two choices:

(1) $\frac{\underline{\underline{\boldsymbol{\Lambda}}'\underline{\boldsymbol{\Lambda}}}}{n}$ diagonal and $\frac{\underline{\underline{\boldsymbol{F}}'\underline{\boldsymbol{F}}}}{T} = \mathbf{I}_r$;

(2) $\frac{\underline{\underline{\boldsymbol{\Lambda}}'\underline{\boldsymbol{\Lambda}}}}{n} = \mathbf{I}_r$ and $\frac{\underline{\underline{\boldsymbol{F}}'\underline{\boldsymbol{F}}}}{T}$ diagonal.

Then,

(a) solve for $\widehat{\boldsymbol{\Lambda}}$ with constraints 1 or 2 and then we get $\widehat{\boldsymbol{F}}$ by linear projection;

(b) solve for $\widehat{\boldsymbol{F}}$ with constraints 1 or 2 and then we get $\widehat{\boldsymbol{\Lambda}}$ by linear projection.

Sample covariance matrix. Define:

- $\widehat{\mathbf{\Gamma}}^x = \frac{\boldsymbol{X}'\boldsymbol{X}}{T}$ which is $n \times n$ with
  - $\widehat{\mathbf{M}}^x$ $r \times r$ diagonal with r largest evals of $\widehat{\mathbf{\Gamma}}^x$;
  - $\widehat{\mathbf{V}}^x$ $n \times r$ with as columns the r corresponding normalized evecs.
- $\widetilde{\mathbf{\Gamma}}^x = \frac{\boldsymbol{X}\boldsymbol{X}'}{n}$ which is $T \times T$ with
  - $\widetilde{\mathbf{M}}^x$ $r \times r$ diagonal with r largest evals of $\widetilde{\mathbf{\Gamma}}^x$;
  - $\widetilde{\mathbf{V}}^x$ $T \times r$ with as columns the r corresponding normalized evecs.
- Notice that, provided $r < \min(n, T)$,

$$\frac{\widehat{\mathbf{M}}^x}{n} = \frac{\widetilde{\mathbf{M}}^x}{T}$$

since the non-zero evals of $\frac{\boldsymbol{X}'\boldsymbol{X}}{nT}$ and of $\frac{\boldsymbol{X}\boldsymbol{X}'}{nT}$ coincide.

Four solutions. Normalized PCs of $\boldsymbol{X}$ (Forni, Giannone, Lippi & Reichlin, 2009).

(1a) Minimize wrt $\underline{\boldsymbol{\Lambda}}$ under the constraint $\frac{\underline{\boldsymbol{\Lambda}}'\underline{\boldsymbol{\Lambda}}}{n}$ is diagonal which gives

$$\widehat{\boldsymbol{\Lambda}} = \widehat{\mathbf{V}}^x(\widehat{\mathbf{M}}^x)^{1/2}.$$

Then:

$$\frac{\widehat{\boldsymbol{\Lambda}}'\widehat{\boldsymbol{\Lambda}}}{n} = \frac{\widehat{\mathbf{M}}^x}{n}$$

and

$$\widehat{\boldsymbol{F}} = \boldsymbol{X}\widehat{\boldsymbol{\Lambda}}(\widehat{\boldsymbol{\Lambda}}'\widehat{\boldsymbol{\Lambda}})^{-1} = \boldsymbol{X}\widehat{\mathbf{V}}^x(\widehat{\mathbf{M}}^x)^{-1/2}.$$

This solution is such that, as required:

$$\begin{aligned}\frac{\widehat{\boldsymbol{F}}'\widehat{\boldsymbol{F}}}{T} &= (\widehat{\mathbf{M}}^x)^{-1/2}\widehat{\mathbf{V}}^{x\prime}\frac{\boldsymbol{X}'\boldsymbol{X}}{T}\widehat{\mathbf{V}}^x(\widehat{\mathbf{M}}^x)^{-1/2}\\ &= (\widehat{\mathbf{M}}^x)^{-1/2}\widehat{\mathbf{V}}^{x\prime}\left(\widehat{\mathbf{V}}^x\widehat{\mathbf{M}}^x\widehat{\mathbf{V}}^{x\prime} + \widehat{\mathbf{V}}^x_{n-r}\widehat{\mathbf{M}}^x_{n-r}\widehat{\mathbf{V}}^{x\prime}_{n-r}\right)\widehat{\mathbf{V}}^x(\widehat{\mathbf{M}}^x)^{-1/2}\\ &= (\widehat{\mathbf{M}}^x)^{-1/2}\widehat{\mathbf{V}}^{x\prime}\widehat{\mathbf{V}}^x\widehat{\mathbf{M}}^x\widehat{\mathbf{V}}^{x\prime}\widehat{\mathbf{V}}^x(\widehat{\mathbf{M}}^x)^{-1/2} = \mathbf{I}_r.\end{aligned}$$

The common component is estimated as:

$$\widehat{\boldsymbol{C}} = \widehat{\boldsymbol{F}}\widehat{\boldsymbol{\Lambda}}' = \boldsymbol{X}\widehat{\mathbf{V}}^x\widehat{\mathbf{V}}^{x\prime}.$$

Four solutions (Bai, 2003).

(1b) Minimize wrt $\underline{\boldsymbol{F}}$ under the constraint $\frac{\underline{\boldsymbol{F}}'\underline{\boldsymbol{F}}}{T} = \mathbf{I}_r$

$$\widetilde{\boldsymbol{F}} = \sqrt{T}\,\widetilde{\mathbf{V}}^x.$$

Then, obviously $\frac{\widetilde{\boldsymbol{F}}'\widetilde{\boldsymbol{F}}}{T} = \mathbf{I}_r$ and

$$\widetilde{\boldsymbol{\Lambda}} = \boldsymbol{X}'\widehat{\boldsymbol{F}}(\widehat{\boldsymbol{F}}'\widehat{\boldsymbol{F}})^{-1} = \frac{\boldsymbol{X}'\widetilde{\mathbf{V}}^x}{\sqrt{T}}.$$

This solution is such that, as required:

$$\begin{aligned}\frac{\widetilde{\boldsymbol{\Lambda}}'\widetilde{\boldsymbol{\Lambda}}}{n} &= \widetilde{\mathbf{V}}^{x\prime}\frac{\boldsymbol{X}\boldsymbol{X}'}{nT}\widetilde{\mathbf{V}}^x\\ &= \widetilde{\mathbf{V}}^{x\prime}\frac{\left(\widetilde{\mathbf{V}}^x\widetilde{\mathbf{M}}^x\widetilde{\mathbf{V}}^{x\prime} + \widetilde{\mathbf{V}}^x_{n-r}\widetilde{\mathbf{M}}^x_{n-r}\widetilde{\mathbf{V}}^{x\prime}_{n-r}\right)}{T}\widetilde{\mathbf{V}}^x = \frac{\widetilde{\mathbf{M}}^x}{T}.\end{aligned}$$

The common component is estimated as:

$$\widehat{\boldsymbol{C}} = \widetilde{\boldsymbol{F}}\widetilde{\boldsymbol{\Lambda}}' = \widetilde{\mathbf{V}}^x\widetilde{\mathbf{V}}^{x\prime}\boldsymbol{X}.$$

Four solutions (Stock and Watson, 2002).

(2a) Minimize wrt $\underline{\boldsymbol{\Lambda}}$ under the constraint $\frac{\underline{\boldsymbol{\Lambda}}'\underline{\boldsymbol{\Lambda}}}{T} = \mathbf{I}_r$

$$\widetilde{\boldsymbol{\Lambda}} = \sqrt{n}\,\widehat{\mathbf{V}}^x.$$

Then, obviously $\frac{\widetilde{\boldsymbol{\Lambda}}'\widetilde{\boldsymbol{\Lambda}}}{n} = \mathbf{I}_r$ and

$$\widetilde{\boldsymbol{F}} = \boldsymbol{X}\widehat{\boldsymbol{\Lambda}}(\widehat{\boldsymbol{\Lambda}}'\widehat{\boldsymbol{\Lambda}})^{-1} = \frac{\boldsymbol{X}\widehat{\mathbf{V}}^x}{\sqrt{n}}.$$

This solution is such that, as required:

$$\begin{aligned}\frac{\widetilde{\boldsymbol{F}}'\widetilde{\boldsymbol{F}}}{T} &= \widehat{\mathbf{V}}^{x\prime}\,\frac{\boldsymbol{X}'\boldsymbol{X}}{nT}\widehat{\mathbf{V}}^x \\ &= \widehat{\mathbf{V}}^{x\prime}\,\frac{\left(\widehat{\mathbf{V}}^x\widehat{\mathbf{M}}^x\widehat{\mathbf{V}}^{x\prime} + \widehat{\mathbf{V}}^x_{n-r}\widehat{\mathbf{M}}^x_{n-r}\widehat{\mathbf{V}}^{x\prime}_{n-r}\right)}{n}\widehat{\mathbf{V}}^x = \frac{\widehat{\mathbf{M}}^x}{n}.\end{aligned}$$

The common component is estimated as:

$$\widehat{\boldsymbol{C}} = \widehat{\boldsymbol{F}}\widehat{\boldsymbol{\Lambda}}' = \boldsymbol{X}\widehat{\mathbf{V}}^x\widehat{\mathbf{V}}^{x\prime}.$$

Four solutions. Normalized PCs of $\boldsymbol{X}'$.

(2b) Minimize wrt $\underline{\boldsymbol{F}}$ under the constraint $\frac{\boldsymbol{F}'\boldsymbol{F}}{T}$ diagonal

$$\widetilde{\boldsymbol{F}} = \widetilde{\mathbf{V}}^x(\widetilde{\mathbf{M}}^x)^{1/2}.$$

Then,

$$\frac{\widetilde{\boldsymbol{F}}'\widetilde{\boldsymbol{F}}}{T} = \frac{\widetilde{\mathbf{M}}^x}{T}.$$

and

$$\widetilde{\boldsymbol{\Lambda}} = \boldsymbol{X}'\widehat{\boldsymbol{F}}(\widehat{\boldsymbol{F}}'\widehat{\boldsymbol{F}})^{-1} = \boldsymbol{X}'\widetilde{\mathbf{V}}^x(\widetilde{\mathbf{M}}^x)^{-1/2}.$$

This solution is such that, as required:

$$\begin{aligned}
\frac{\widehat{\boldsymbol{\Lambda}}'\widehat{\boldsymbol{\Lambda}}}{n} &= (\widetilde{\mathbf{M}}^x)^{-1/2}\widetilde{\mathbf{V}}^{x\prime}\frac{\boldsymbol{X}\boldsymbol{X}'}{n}\widetilde{\mathbf{V}}^x(\widetilde{\mathbf{M}}^x)^{-1/2}\\
&= (\widetilde{\mathbf{M}}^x)^{-1/2}\widetilde{\mathbf{V}}^{x\prime}\left(\widetilde{\mathbf{V}}^x\widetilde{\mathbf{M}}^x\widetilde{\mathbf{V}}^{x\prime} + \widetilde{\mathbf{V}}^x_{n-r}\widetilde{\mathbf{M}}^x_{n-r}\widetilde{\mathbf{V}}^{x\prime}_{n-r}\right)\widetilde{\mathbf{V}}^x(\widetilde{\mathbf{M}}^x)^{-1/2}\\
&= (\widetilde{\mathbf{M}}^x)^{-1/2}\widetilde{\mathbf{V}}^{x\prime}\widetilde{\mathbf{V}}^x\widetilde{\mathbf{M}}^x\widetilde{\mathbf{V}}^{x\prime}\widetilde{\mathbf{V}}^x(\widetilde{\mathbf{M}}^x)^{-1/2} = \mathbf{I}_r.
\end{aligned}$$

The common component is estimated as:

$$\widehat{\boldsymbol{C}} = \widetilde{\boldsymbol{F}}\widetilde{\boldsymbol{\Lambda}}' = \widetilde{\mathbf{V}}^x\widetilde{\mathbf{V}}^{x\prime}\boldsymbol{X}.$$

- All solutions give some form of PC and equivalent and have the same asymptotic properties.
- So PC is the least squares estimator of a factor model.
- We focus on solution (1a):

$$\widehat{\boldsymbol{\lambda}}_i^{\mathsf{PC}\prime} = \widehat{\mathbf{v}}_i^{x\prime}(\widehat{\mathbf{M}}^x)^{1/2}, \qquad \widehat{\mathbf{F}}_t^{\mathsf{PC}} = (\widehat{\mathbf{M}}^x)^{-1/2}\widehat{\mathbf{V}}^{x\prime}\mathbf{x}_t.$$

- This is the classical solution (Pearson, 1902; Hotelling, 1933; Mardia, Kent & Bibby, 1979; Jolliffe, 2002; Peña, 2002).
- Indeed, dynamic factor models are about time series, so we treat $\boldsymbol{\Lambda}$ as deterministic while $\{\mathbf{F}_t\}$ are r-dimensional stochastic processes, weighted averages of the n dimensional stochastic process $\{\mathbf{x}_t\}$.
- It is then natural to consider solutions based on the $n \times n$ covariance matrix $\widehat{\boldsymbol{\Gamma}}^x$ and not those on the $T \times T$ covariance matrix $\widetilde{\boldsymbol{\Gamma}}^x$.
- Notice that it is not necessary to have a consistent estimator of the whole sample covariance. So $\widehat{\boldsymbol{\Gamma}}^x$ does not have to be consistent, indeed it cannot be consistent if $n > T$, we just need $n^{-1}\|\widehat{\boldsymbol{\Gamma}}^x - \boldsymbol{\Gamma}^x\| = o_p(1)$.
- Reversing n and T requires less natural assumptions to prove consistency.

Asymptotic properties. Loadings.

(Bai, 2003; Barigozzi, 2022).

- For any given $i = 1, \dots, n$

$$\begin{aligned}\sqrt{T}(\widehat{\boldsymbol{\lambda}}_i^{\mathsf{PC}} - \widehat{\mathbf{H}}'\boldsymbol{\lambda}_i) &= \widehat{\mathbf{H}}' \left(\frac{1}{T} \sum_{t=1}^{T} \mathbf{F}_t \mathbf{F}_t' \right)^{-1} \left(\frac{1}{\sqrt{T}} \sum_{t=1}^{T} \mathbf{F}_t \xi_{it} \right) + o_p(1) \\ &= \left(\frac{1}{T} \sum_{t=1}^{T} \widehat{\mathbf{H}}^{-1} \mathbf{F}_t \mathbf{F}_t' \widehat{\mathbf{H}}^{-1'} \right)^{-1} \left(\frac{1}{\sqrt{T}} \sum_{t=1}^{T} \widehat{\mathbf{H}}^{-1} \mathbf{F}_t \xi_{it} \right) + o_p(1).\end{aligned}$$

  This is OLS when, for a fixed i, we regress x_{it} onto $\widehat{\mathbf{H}}^{-1}\mathbf{F}_t$.

- So if $\frac{\sqrt{T}}{n} \to 0$ then

$$\sqrt{T}(\widehat{\boldsymbol{\lambda}}_i^{\mathsf{PC}} - \widehat{\mathbf{H}}'\boldsymbol{\lambda}_i) \to_d \mathcal{N}\left(\mathbf{0}_r, \boldsymbol{\mathcal{V}}_i^{\mathsf{PC}}\right).$$

Asymptotic covariance of loadings.

$$
\begin{aligned}
\boldsymbol{\mathcal{V}}_i^{\mathsf{PC}} &= \boldsymbol{V}_0^{-1}\boldsymbol{Q}_0\boldsymbol{\Phi}_i\boldsymbol{Q}_0'\boldsymbol{V}_0^{-1},\\
\boldsymbol{\Phi}_i &= \lim_{T\to\infty}\frac{1}{T}\sum_{t=1}^{T}\sum_{s=1}^{T}\mathsf{E}[\mathbf{F}_t\mathbf{F}_s'\xi_{it}\xi_{is}] = \lim_{T\to\infty}\frac{\mathsf{E}[\boldsymbol{F}'\zeta_i\zeta_i'\boldsymbol{F}]}{T},\\
\mathbf{Q}_0 &= \boldsymbol{V}_0\boldsymbol{\Upsilon}_0'(\boldsymbol{\Gamma}^F)^{-1/2}
\end{aligned}
$$

such that $\boldsymbol{\Upsilon}_0$ are evec of $(\boldsymbol{\Gamma}^F)^{1/2}\boldsymbol{\Sigma}_\Lambda(\boldsymbol{\Gamma}^F)^{1/2}$ with evals $\boldsymbol{V}_0$.

Cfr. Bai (2003) where

$$
\begin{aligned}
\boldsymbol{\mathcal{V}}_i^{\mathsf{PC,B}} &= (\boldsymbol{\mathcal{Q}}^{-1})'\boldsymbol{\Phi}_i(\boldsymbol{\mathcal{Q}})^{-1},\\
\boldsymbol{\mathcal{Q}}^{-1} &= (\boldsymbol{\Sigma}_\Lambda)^{1/2}\boldsymbol{\Upsilon}_1(\mathbf{V}_0)^{-1/2}
\end{aligned}
$$

such that $\boldsymbol{\Upsilon}_1$ are evec of $\boldsymbol{\Sigma}_\Lambda^{1/2}\boldsymbol{\Gamma}^F\boldsymbol{\Sigma}_\Lambda^{1/2}$ with evals $\boldsymbol{V}_0$.

Notice that,

$$
\mathsf{tr}(\boldsymbol{\mathcal{V}}_i^{\mathsf{PC}}) = \mathsf{tr}(\boldsymbol{\mathcal{V}}_i^{\mathsf{PC,B}}).
$$

Asymptotic properties. Factors.

(Bai, 2003; Barigozzi, 2022).

- For any given $t = 1, \dots, T$

$$\sqrt{n}(\widehat{\mathbf{F}}_t^{\mathsf{PC}} - \widehat{\mathbf{H}}^{-1}\mathbf{F}_t) = \widehat{\mathbf{H}}^{-1}\left(\frac{1}{n}\sum_{i=1}^n \boldsymbol{\lambda}_i\boldsymbol{\lambda}_i'\right)^{-1}\left(\frac{1}{\sqrt{n}}\sum_{i=1}^n \boldsymbol{\lambda}_i\xi_{it}\right) + o_p(1)$$
$$= \left(\frac{1}{n}\sum_{i=1}^n \widehat{\mathbf{H}}'\boldsymbol{\lambda}_i\boldsymbol{\lambda}_i'\widehat{\mathbf{H}}\right)^{-1}\left(\frac{1}{\sqrt{n}}\sum_{i=1}^n \widehat{\mathbf{H}}'\boldsymbol{\lambda}_i\xi_{it}\right) + o_p(1).$$

  This is OLS when, for a fixed t, we regress x_{it} onto $\widehat{\mathbf{H}}'\boldsymbol{\lambda}_i$.

- So if $\frac{\sqrt{n}}{T} \to 0$ then

$$\sqrt{n}(\widehat{\mathbf{F}}_t^{\mathsf{PC}} - \widehat{\mathbf{H}}^{-1}\mathbf{F}_t) \to_d \mathcal{N}\left(\mathbf{0}_r, \boldsymbol{\mathcal{W}}_t^{\mathsf{PC}}\right).$$

Asymptotic covariance of factors.

$$\begin{aligned}
\boldsymbol{\mathcal{W}}_t^{\mathsf{PC}} &= (\mathbf{Q}_0')^{-1}\boldsymbol{\Gamma}_t(\mathbf{Q}_0)^{-1},\\
\boldsymbol{\Gamma}_t &= \lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{n}\boldsymbol{\lambda}_i\boldsymbol{\lambda}_j'\mathsf{E}[\xi_{it}\xi_{jt}] = \lim_{n\to\infty}\frac{\boldsymbol{\Lambda}'\mathsf{E}[\boldsymbol{\xi}_t\boldsymbol{\xi}_t']\boldsymbol{\Lambda}}{n},\\
(\mathbf{Q}_0)^{-1} &= (\boldsymbol{\Gamma}^F)^{1/2}\boldsymbol{\Upsilon}_0(\boldsymbol{V}_0)^{-1}
\end{aligned}$$

such that $\boldsymbol{\Upsilon}_0$ are evec of $(\boldsymbol{\Gamma}^F)^{1/2}\boldsymbol{\Sigma}_\Lambda(\boldsymbol{\Gamma}^F)^{1/2}$ with evals $\boldsymbol{V}_0$.

Cfr. Bai (2003) where

$$\begin{aligned}
\boldsymbol{\mathcal{W}}_t^{\mathsf{PC,B}} &= (\boldsymbol{V}_0)^{-1}\boldsymbol{\mathcal{Q}}\boldsymbol{\Gamma}_t\boldsymbol{\mathcal{Q}}'(\boldsymbol{V}_0)^{-1}\\
\boldsymbol{\mathcal{Q}} &= (\mathbf{V}_0)^{1/2}\boldsymbol{\Upsilon}_1'(\boldsymbol{\Sigma}_\Lambda)^{-1/2}
\end{aligned}$$

such that $\boldsymbol{\Upsilon}_1$ are evec of $\boldsymbol{\Sigma}_\Lambda^{1/2}\boldsymbol{\Gamma}^F\boldsymbol{\Sigma}_\Lambda^{1/2}$ with evals $\boldsymbol{V}_0$.

Notice that,

$$\mathsf{tr}(\boldsymbol{\mathcal{W}}_t^{\mathsf{PC}}) = \mathsf{tr}(\boldsymbol{\mathcal{W}}_t^{\mathsf{PC,B}}).$$

Asymptotic properties. Common component.

(Bai, 2003; Barigozzi, 2022).

- For any given $i = 1, \ldots, n$ and $t = 1, \ldots, T$

$$|\widehat{\chi}_{it}^{\mathsf{PC}} - \chi_{it}| = O_p\left(\frac{1}{\sqrt{T}}\right) + O_p\left(\frac{1}{\sqrt{n}}\right)$$

  with $\widehat{\chi}_{it}^{\mathsf{PC}} = \widehat{\boldsymbol{\lambda}}_i^{\mathsf{PC}\prime} \widehat{\mathbf{F}}_t^{\mathsf{PC}} = \widehat{\mathbf{v}}_i^{x\prime} \widehat{\mathbf{V}}^{x\prime} \mathbf{x}_t$.

- And, as $n, T \to \infty$,

$$\frac{(\widehat{\chi}_{it}^{\mathsf{PC}} - \chi_{it})}{\left(\frac{\boldsymbol{\lambda}_i' \boldsymbol{\mathcal{W}}_t^{\mathsf{PC}} \boldsymbol{\lambda}_i}{n} + \frac{\mathbf{F}_t' \boldsymbol{\mathcal{V}}_i^{\mathsf{PC}} \mathbf{F}_t}{T}\right)^{1/2}} \to_d \mathcal{N}(0, 1).$$

- It does not depend on the chosen identification.

The above results depend on $\widehat{\mathbf{H}} = \left(\frac{\boldsymbol{F}'\boldsymbol{F}}{T}\right)\left(\frac{\boldsymbol{\Lambda}'\widehat{\boldsymbol{\Lambda}}}{n}\right)\left(\frac{\widehat{\mathbf{M}}^x}{n}\right)^{-1}$ which is unknown.

Under the classical identification conditions used in exploratory factor analysis (Bai & Ng, 2013; Barigozzi, 2022).

$$\widehat{\mathbf{H}} = \boldsymbol{J} + o_p\left(\max\left(\frac{1}{\sqrt{n}}, \frac{1}{\sqrt{T}}\right)\right),$$

where $\boldsymbol{J}$ is an $r \times r$ diagonal matrix with entries ± 1.

Under global identification $\boldsymbol{J} = \mathbf{I}_r$.

Asymptotic properties of PC under global identification - Loadings

(Bai & Ng, 2013; Barigozzi, 2022).

- for any given $i = 1, \dots, n$ as $n, T \to \infty$

$$\|\widehat{\boldsymbol{\lambda}}_i^{\mathsf{PC}} - \boldsymbol{\lambda}_i^{\mathsf{OLS}}\| = O_p\left(\frac{1}{n}\right) + O_p\left(\frac{1}{\sqrt{nT}}\right);$$

- if $\frac{\sqrt{T}}{n} \to 0$ then

$$\sqrt{T}(\widehat{\boldsymbol{\lambda}}_i^{\mathsf{PC}} - \boldsymbol{\lambda}_i) \to_d \mathcal{N}\left(\mathbf{0}_r, \boldsymbol{\mathcal{V}}_i^{\mathsf{OLS}}\right)$$

  with

$$\boldsymbol{\mathcal{V}}_i^{\mathsf{OLS}} = (\boldsymbol{\Gamma}^F)^{-1}\left\{\lim_{T\to\infty} \frac{\mathsf{E}[\boldsymbol{F}'\mathsf{E}[\zeta_i\zeta_i']\boldsymbol{F}]}{T}\right\}(\boldsymbol{\Gamma}^F)^{-1} = \lim_{T\to\infty} \frac{\mathsf{E}[\boldsymbol{F}'\mathsf{E}[\zeta_i\zeta_i']\boldsymbol{F}]}{T},$$

- PC is asymptotically equivalent to OLS.
- $\boldsymbol{\mathcal{V}}_i^{\mathsf{OLS}}$ has sandwich form due to the fact that we do not take into account idiosyncratic serial correlations since PC is non parametric.

Asymptotic properties of PC under global identification - Factors

(Bai & Ng, 2013; Barigozzi, 2022).

- for any given $t = 1, \ldots, T$ as $n, T \to \infty$

$$\|\widehat{\mathbf{F}}_t^{\mathsf{PC}} - \mathbf{F}_t^{\mathsf{OLS}}\| = O_p\left(\frac{1}{T}\right) + O_p\left(\frac{1}{\sqrt{nT}}\right);$$

- if $\frac{\sqrt{n}}{T} \to 0$ then

$$\sqrt{n}(\widehat{\mathbf{F}}_t^{\mathsf{PC}} - \mathbf{F}_t) \to_d \mathcal{N}\left(\mathbf{0}_r, \boldsymbol{\mathcal{W}}_t^{\mathsf{OLS}}\right)$$

  with

$$\boldsymbol{\mathcal{W}}_t^{\mathsf{OLS}} = (\boldsymbol{\Sigma}_\Lambda)^{-1}\left\{\lim_{n\to\infty}\frac{\boldsymbol{\Lambda}'\mathsf{E}[\boldsymbol{\xi}_t\boldsymbol{\xi}_t']\boldsymbol{\Lambda}}{n}\right\}(\boldsymbol{\Sigma}_\Lambda)^{-1}.$$

- PC is asymptotically equivalent to OLS.
- $\boldsymbol{\mathcal{W}}_t^{\mathsf{OLS}}$ has sandwich form due to the fact that we do not take into account idiosyncratic cross-sectional correlations and heteroskedasticity since PC is non parametric.

Is PC the best we can do? We could use ML and GLS.

- PC is nonparametric (no assumption on idiosyncratic distribution), ML is fully parametric.
- GLS is better than OLS for factors when idiosyncratic is heteroskedastic across i.
- GLS is better than OLS for loadings when idiosyncratic is heteroskedastic across t (but we assume stationarity).
- ML/GLS coincides with PC in the case of i.i.d. idiosyncratic components.

- **The Likelihood**

Consider the stacked version of the model

$$\boldsymbol{\mathcal{X}} = \underbrace{(\boldsymbol{\Lambda} \otimes \mathbf{I}_T)}_{\boldsymbol{\mathcal{L}}} \boldsymbol{\mathcal{F}} + \boldsymbol{\mathcal{E}}.$$

Let:

$$\boldsymbol{\Omega}^x = \mathsf{E}[\boldsymbol{\mathcal{X}}\boldsymbol{\mathcal{X}}'], \quad \boldsymbol{\Omega}^F = \mathsf{E}[\boldsymbol{\mathcal{F}}\boldsymbol{\mathcal{F}}'], \quad \boldsymbol{\Omega}^\xi = \mathsf{E}[\boldsymbol{\mathcal{E}}\boldsymbol{\mathcal{E}}'].$$

Gaussian quasi log-likelihood:

$$\begin{aligned}
\ell(\boldsymbol{\mathcal{X}}, \underline{\boldsymbol{\varphi}}) &= -\frac{nT}{2} - \frac{1}{2} \log \det \underline{\boldsymbol{\Omega}}^x - \frac{1}{2} \mathsf{tr} \left(\boldsymbol{\mathcal{X}}\boldsymbol{\mathcal{X}}' (\underline{\boldsymbol{\Omega}}^x)^{-1} \right) \\
&\simeq -\frac{1}{2} \log \det \left(\underline{\boldsymbol{\mathcal{L}}}\, \underline{\boldsymbol{\Omega}}^F \underline{\boldsymbol{\mathcal{L}}}' + \underline{\boldsymbol{\Omega}}^\xi \right) - \frac{1}{2} \left(\boldsymbol{\mathcal{X}}' (\underline{\boldsymbol{\mathcal{L}}}\, \underline{\boldsymbol{\Omega}}^F \underline{\boldsymbol{\mathcal{L}}}' + \underline{\boldsymbol{\Omega}}^\xi)^{-1} \boldsymbol{\mathcal{X}} \right).
\end{aligned}$$

The parameters to be estimated are $\boldsymbol{\varphi} = (\boldsymbol{\Lambda}, \boldsymbol{\Omega}^F, \boldsymbol{\Omega}^\xi)$.

ML is in general unfeasible:

- too many parameters not enough degrees of freedom:
  - the ML estimator of $\boldsymbol{\Omega}^\xi$ cannot be positive definite;
  - for time series $\boldsymbol{\Omega}^F$ is a full matrix.

We introduce some mis-specifications:

1. we treat the idiosyncratic components as if they were uncorrelated

$\Rightarrow \boldsymbol{\Omega}^{\xi}$ is replaced by $\mathbf{I}_T \otimes \boldsymbol{\Sigma}^{\xi}$ where $\boldsymbol{\Sigma}^{\xi}$ is diagonal with entries $\sigma_i^2 = \mathsf{E}[\xi_{it}^2]$.

We always work with the log-likelihood:

$$\ell_0(\boldsymbol{\mathcal{X}}, \underline{\boldsymbol{\varphi}}) \simeq -\frac{1}{2}\log\det\left(\underline{\boldsymbol{\mathcal{L}}}\,\underline{\boldsymbol{\Omega}}^F \underline{\boldsymbol{\mathcal{L}}}' + \mathbf{I}_T \otimes \underline{\boldsymbol{\Sigma}}^{\xi}\right) \\ -\frac{1}{2}\left(\boldsymbol{\mathcal{X}}'(\underline{\boldsymbol{\mathcal{L}}}\,\underline{\boldsymbol{\Omega}}^F \underline{\boldsymbol{\mathcal{L}}}' + \mathbf{I}_T \otimes \underline{\boldsymbol{\Sigma}}^{\xi})^{-1}\boldsymbol{\mathcal{X}}\right).$$

We are doing QML rather than ML!

Moreover,

2a. for static model we consider the factors as if they are serially uncorrelated and $\boldsymbol{\Omega}^F$ is replaced by $\mathbf{I}_T \otimes \boldsymbol{\Gamma}^F = \mathbf{I}_{rT}$;

2b. for dynamic model we assume a parametric model for factor dynamics and parametrize $\boldsymbol{\Omega}^F$ accordingly.

- **Approximate Static Factor Model - Quasi Maximum Likelihood**

The log-likelihood is

$$\ell_{0,S}(\boldsymbol{\mathcal{X}}, \underline{\varphi}) \simeq -\frac{T}{2} \log\det\left(\underline{\mathbf{\Lambda}}\, \underline{\mathbf{\Lambda}}' + \underline{\mathbf{\Sigma}}^{\xi}\right) - \frac{1}{2}\sum_{t=1}^{T}\left(\mathbf{x}_t'(\underline{\mathbf{\Lambda}}\, \underline{\mathbf{\Lambda}}' + \underline{\mathbf{\Sigma}}^{\xi})^{-1}\mathbf{x}_t\right),$$

The parameters to be estimated are $\varphi = (\mathbf{\Lambda}, \mathbf{\Sigma}^{\xi})$.

We work under the global identification assumptions.

Issues

1. No closed form solution for QML estimator exists, we need numerical approaches, e.g., EM algorithm
   (Rubin & Thayer, 1982; Bai & Li, 2012, 2016; Ng, Yau & Chan, 2015; Sundberg and Feldmann, 2016).
2. How to estimate the factors which are not appearing in the log-likelihood? Least-squares or regression estimators
   (Thomson, 1951; Bartlett, 1937).

Asymptotic properties QML estimator - Loadings

(Bai & Li, 2016; Barigozzi, 2023).

- for any given $i = 1, \dots, n$ as $n, T \to \infty$

$$\|\widehat{\boldsymbol{\lambda}}_i^{\text{QML,S}} - \widehat{\boldsymbol{\lambda}}_i^{\text{PC}}\| = O_p\left(\frac{1}{n}\right), \quad \|\widehat{\boldsymbol{\lambda}}_i^{\text{PC}} - \boldsymbol{\lambda}_i^{\text{OLS}}\| = O_p\left(\frac{1}{n}\right) + O_p\left(\frac{1}{\sqrt{nT}}\right);$$

- if $\frac{\sqrt{T}}{n} \to 0$ then

$$\sqrt{T}(\widehat{\boldsymbol{\lambda}}_i^{\text{QML,S}} - \boldsymbol{\lambda}_i) \to_d \mathcal{N}\left(\mathbf{0}_r, \boldsymbol{\mathcal{V}}_i^{\text{OLS}}\right)$$

$$\boldsymbol{\mathcal{V}}_i^{\text{OLS}} = (\boldsymbol{\Gamma}^F)^{-1} \left\{\lim_{T\to\infty} \frac{\mathsf{E}[\boldsymbol{F}'\mathsf{E}[\boldsymbol{\zeta}_i\boldsymbol{\zeta}_i']\boldsymbol{F}]}{T}\right\} (\boldsymbol{\Gamma}^F)^{-1} = \lim_{T\to\infty} \frac{\mathsf{E}[\boldsymbol{F}'\mathsf{E}[\boldsymbol{\zeta}_i\boldsymbol{\zeta}_i']\boldsymbol{F}]}{T}.$$

- QML is asymptotically equivalent to PC and OLS.
- $\boldsymbol{\mathcal{V}}_i^{\text{OLS}}$ has sandwich form due to neglected serial idiosyncratic correlation since likelihood is misspecified.
- Neglecting cross-sectional idiosyncratic correlation has no impact but, in practice, QML estimation of $\boldsymbol{\Gamma}^\xi$ is unfeasible.
- Treating factors as serially uncorrelated does not affect the result since autocorrelation of regressors does not affect OLS.

- Consistency of loadings requires $n \to \infty$, otherwise we cannot identify the model.
- The mis-specification error, which we introduce by using a mis-specified log-likelihood, vanishes asymptotically only if $n \to \infty$.
- The QML estimator does not suffer of the curse of dimensionality, but, in fact, it produces consistent estimates only in a high-dimensional setting, i.e., it enjoys a blessing of dimensionality.

Special cases.

- Exact not autocorrelated heteroskedastic case, $\boldsymbol{\Omega}^{\xi} = \mathbf{I}_T \otimes \boldsymbol{\Sigma}^{\xi}$. The estimated loadings are the same as before, so have no closed form but now are $\sqrt{T}$-consistent and asymptotically normal (Anderson & Rubin, 1956).
- Exact not autocorrelated homoskedastic case, $\boldsymbol{\Omega}^{\xi} = \sigma^2 \mathbf{I}_{nT}$. The estimated loadings are given by $\widehat{\boldsymbol{\lambda}}_i^{\mathsf{QML},0} = \left(\widehat{\mathbf{M}}^x - \widehat{\sigma}^{2\mathsf{QML},0}\mathbf{I}_r\right)^{1/2} \widehat{\mathbf{v}}_i^x$ they are $\sqrt{T}$-consistent and asymptotically normal (Tipping & Bishop, 1999).
- In both cases (Bai & Li, 2012)

$$\|\widehat{\boldsymbol{\lambda}}_i^{\mathsf{QML}} - \boldsymbol{\lambda}_i^{\mathsf{OLS}}\| = O_p\left(\frac{1}{\sqrt{nT}}\right). \qquad (*)$$

- if n fixed the asymptotic covariance is very complicated because $(*)$ is not negligible, this is the classical case (Amemyia, Fuller & Pantula, 1987).
- if $n \to \infty$ then $(*)$ is negligible so the asymptotic covariance is $\boldsymbol{\mathcal{V}}_i^{\mathsf{OLS},*} = \sigma_i^2(\boldsymbol{\Gamma}^F)^{-1} = \sigma_i^2 \mathbf{I}_r$ or $\boldsymbol{\mathcal{V}}_i^{\mathsf{OLS},0} = \sigma^2(\boldsymbol{\Gamma}^F)^{-1} = \sigma^2 \mathbf{I}_r$, since now the likelihood is correctly specified (Bai & Li, 2012).

idiosyncratic	PC		QML	
1. $\mathbf{\Omega}^\xi$ full	$\min(n, \sqrt{T})$	$\mathcal{V}_i^{\textsf{OLS}}$	$\min(n, \sqrt{T})$	$\mathcal{V}_i^{\textsf{OLS}}$
2. $\mathbf{\Omega}^\xi = \mathbf{I}_T \otimes \mathbf{\Gamma}^\xi$	$\min(n, \sqrt{T})$	$\mathcal{V}_i^{\textsf{OLS,*}}$	$\min(n, \sqrt{T})$	$\mathcal{V}_i^{\textsf{OLS,*}}$
3. $\mathbf{\Omega}^\xi = \mathbf{I}_T \otimes \mathbf{\Sigma}^\xi$	$\min(n, \sqrt{T})$	$\mathcal{V}_i^{\textsf{OLS,*}}$	$\sqrt{T}$	$\mathcal{V}_i^{\textsf{OLS,*}}$ (if $n \to \infty$) too complex (if n fixed)
4. $\mathbf{\Omega}^\xi = \sigma^2 \mathbf{I}_{nT}$	$\min(n, \sqrt{T})$	$\mathcal{V}_i^{\textsf{OLS,0}}$	$\sqrt{T}$	$\mathcal{V}_i^{\textsf{OLS,0}}$ (if $n \to \infty$) too complex (if n fixed)

Asymptotic covariances

$\mathcal{V}_i^{\textsf{OLS}} = (\mathbf{\Gamma}^F)^{-1} \left\{ \lim_{T\to\infty} \frac{\mathsf{E}[\boldsymbol{F}' \mathsf{E}[\boldsymbol{\zeta}_i \boldsymbol{\zeta}_i'] \boldsymbol{F}]}{T} \right\} (\mathbf{\Gamma}^F)^{-1}$, $\mathcal{V}_i^{\textsf{OLS,*}} = \sigma_i^2 (\mathbf{\Gamma}^F)^{-1}$, $\mathcal{V}_i^{\textsf{OLS,0}} = \sigma^2 (\mathbf{\Gamma}^F)^{-1}$

$\mathbf{\Gamma}^F = \lim_{T\to\infty} \frac{\boldsymbol{F}'\boldsymbol{F}}{T}$, here $\mathbf{\Gamma}^F = \mathbf{I}_r$ by assumption

Estimators

PC $\widehat{\boldsymbol{\lambda}}_i^{\textsf{PC}} = (\mathrm{M}^x)^{1/2} \widehat{\mathbf{v}}_i^x$ cases 1, 2, 3, 4;

QML $\widehat{\boldsymbol{\lambda}}_i^{\textsf{QML,S}}$ no closed form, case 1, 2, 3; $\widehat{\boldsymbol{\lambda}}_i^{\textsf{QML,0}} = (\mathrm{M}^x - \widehat{\sigma}^{2\textsf{QML,0}})^{1/2} \widehat{\mathbf{v}}_i^x$, case 4

How to estimate factors given ML estimator of the parameters?

- If factors are treated as parameters, the log-likelihood can be written as (Anderson & Rubin, 1956; Anderson, 2003)

$$\ell_{0,S}(\boldsymbol{\mathcal{X}},\underline{\boldsymbol{\varphi}},\underline{\boldsymbol{\mathcal{F}}}) \simeq -\frac{T}{2}\log\det(\underline{\boldsymbol{\Sigma}}^{\xi}) - \frac{1}{2}\sum_{t=1}^{T}\Big((\mathbf{x}_t - \underline{\boldsymbol{\Lambda}}\,\underline{\mathbf{F}}_t)'(\underline{\boldsymbol{\Sigma}}^{\xi})^{-1}(\mathbf{x}_t - \underline{\boldsymbol{\Lambda}}\,\underline{\mathbf{F}}_t)\Big).$$

  For given $\boldsymbol{\varphi} = (\boldsymbol{\Lambda}, \boldsymbol{\Sigma}^{\xi})$ and any given t the ML estimator of the factors is

$$\mathbf{F}_t^{\mathsf{WLS}} = \Big(\boldsymbol{\Lambda}'(\boldsymbol{\Sigma}^{\xi})^{-1}\boldsymbol{\Lambda}\Big)^{-1}\boldsymbol{\Lambda}'(\boldsymbol{\Sigma}^{\xi})^{-1}\mathbf{x}_t,$$

- When we compute the WLS using the QML estimator of the parameters we have the classical "least-squares estimator" $\widehat{\mathbf{F}}_t^{\mathsf{WLS}}$ (Bartlett, 1937).
- $\boldsymbol{\mathcal{F}} = (\mathbf{F}_1' \cdots \mathbf{F}_T')'$ are additional rT parameters to be estimated, and this is possible only if $n \to \infty \Rightarrow$ blessing of dimensionality!
- Both the log-likelihood and its maximum WLS need $\boldsymbol{\Sigma}^{\xi}$ positive definite.

How to estimate factors given ML estimator of the parameters?

- If we treat the factors as random variables, but we do not model their dynamics, then their optimal (in mean-squared sense) linear estimator is the linear projection of the true factors onto the observed data:

$$\mathbf{F}_t^{\mathsf{LP}} = \mathbf{\Gamma}^F \mathbf{\Lambda}' \left(\mathbf{\Lambda} \mathbf{\Gamma}^F \mathbf{\Lambda}' + \mathbf{\Sigma}^\xi \right)^{-1} \mathbf{x}_t = \left(\mathbf{\Lambda}' (\mathbf{\Sigma}^\xi)^{-1} \mathbf{\Lambda} + \mathbf{I}_r \right)^{-1} \mathbf{\Lambda}' (\mathbf{\Sigma}^\xi)^{-1} \mathbf{x}_t$$

- When we compute the LP using the QML estimator of the parameters we have the classical "regression estimator" $\widehat{\mathbf{F}}_t^{\mathsf{LP}}$ (Thomson, 1951).
- The LP in its first formulation does not need $\mathbf{\Sigma}^\xi$ positive definite.
- For finite n the LP has always a smaller MSE than the WLS.
- For any given $t = 1, \dots, T$ as $n \to \infty$,

$$\|\mathbf{F}_t^{\mathsf{WLS}} - \mathbf{F}_t^{\mathsf{LP}}\| = O_p\left(\frac{1}{n}\right).$$

since $(\mathbf{\Lambda}'(\mathbf{\Sigma}^\xi)^{-1}\mathbf{\Lambda} + \mathbf{I}_r)^{-1} = (\mathbf{\Lambda}'(\mathbf{\Sigma}^\xi)^{-1}\mathbf{\Lambda})^{-1} + O(n^{-1})$ (Taylor expansion).

Asymptotic properties WLS and LP estimators - Factors

(Bai & Li, 2016).

- for any given $t = 1, \dots, T$ as $n, T \to \infty$

$$\|\widehat{\mathbf{F}}_t^{\mathsf{WLS}} - \mathbf{F}_t^{\mathsf{WLS}}\| = O_p\left(\frac{1}{T}\right) + O_p\left(\frac{1}{\sqrt{nT}}\right), \qquad \|\mathbf{F}_t^{\mathsf{WLS}} - \mathbf{F}_t\| = O_p\left(\frac{1}{\sqrt{n}}\right).$$

- if $\frac{\sqrt{n}}{T} \to 0$ then

$$\sqrt{T}(\widehat{\mathbf{F}}_t^{\mathsf{WLS}} - \mathbf{F}_t) \to_d \mathcal{N}\left(\mathbf{0}_r, \mathcal{W}_t^{\mathsf{WLS}}\right)$$

$$\mathcal{W}_t^{\mathsf{WLS}} = (\boldsymbol{\Sigma}_{\Lambda\xi\Lambda})^{-1}\left\{\lim_{n\to\infty} \frac{\boldsymbol{\Lambda}'(\boldsymbol{\Sigma}^{\xi})^{-1}\mathsf{E}[\boldsymbol{\xi}_t\boldsymbol{\xi}_t'](\boldsymbol{\Sigma}^{\xi})^{-1}\boldsymbol{\Lambda}}{n}\right\}(\boldsymbol{\Sigma}_{\Lambda\xi\Lambda})^{-1},$$
$$\boldsymbol{\Sigma}_{\Lambda\xi\Lambda} = \lim_{n\to\infty} n^{-1}\boldsymbol{\Lambda}'(\boldsymbol{\Sigma}^{\xi})^{-1}\boldsymbol{\Lambda}.$$

- The same properties hold for the LP estimator.
- $\mathcal{W}_t^{\mathsf{WLS}}$ has sandwich form due to neglected cross-sectional idiosyncratic correlation when implementing WLS or LP, as GLS which requires estimating $(\boldsymbol{\Gamma}^{\xi})^{-1}$ is unfeasible.
- Serial correlation has no impact for $\widehat{\mathbf{F}}_t^{\mathsf{WLS}}$ and serial heteroskedasticity is ruled out by assumption.

Efficiency of WLS/LP (Barigozzi & Luciani, 2019)

If $\sum_{i=1,i\neq j}^{n} |[\mathbf{\Gamma}^{\xi}]_{ij}| = o(n)$, then

$$\mathcal{W}_t^{\mathsf{OLS}} \succ \mathcal{W}_t^{\mathsf{WLS}}$$

WLS is more efficient than PC.

The assumption on $\mathbf{\Gamma}^{\xi}$ implies some form of sparsity (Bai & Liao, 2016).

Special cases.

- Exact heteroskedastic case $\mathbf{\Gamma}^{\xi} = \mathbf{\Sigma}^{\xi}$. WLS/LP and PC are $\min(\sqrt{n}, T)$-consistent and the asymptotic covariances are
  - for WLS/LP: $\boldsymbol{\mathcal{W}}_t^{\mathsf{WLS},*} = (\mathbf{\Sigma}_{\Lambda\xi\Lambda})^{-1}$.
  - for PC: $\boldsymbol{\mathcal{W}}_t^{\mathsf{OLS},*} = (\mathbf{\Sigma}_{\Lambda})^{-1} \left\{\lim_{n\to\infty} \frac{\mathbf{\Lambda}'\mathbf{\Sigma}^{\xi}\mathbf{\Lambda}}{n}\right\} (\mathbf{\Sigma}_{\Lambda})^{-1}$.
  - So $\boldsymbol{\mathcal{W}}_t^{\mathsf{OLS},*} \succ \boldsymbol{\mathcal{W}}_t^{\mathsf{WLS},*}$, WLS is more efficient than OLS.
- Exact homoskedastic case, $\mathbf{\Gamma}^{\xi} = \sigma^2 \mathbf{I}_n$.
  - OLS and WLS coincide

$$\mathbf{F}_t^{\mathsf{WLS}} = \left(\mathbf{\Lambda}'(\sigma^2\mathbf{I}_n)^{-1}\mathbf{\Lambda}\right)^{-1} \mathbf{\Lambda}'(\sigma^2\mathbf{I}_n)^{-1}\mathbf{x}_t = (\mathbf{\Lambda}'\mathbf{\Lambda})^{-1}\mathbf{\Lambda}'\mathbf{x}_t = \mathbf{F}_t^{\mathsf{OLS}}.$$

  - OLS and LP are asymptotically equivalent as $n \to \infty$.
  - WLS/LP and PC are $\min(\sqrt{n}, T)$-consistent and the asymptotic covariance is $\boldsymbol{\mathcal{W}}_t^{\mathsf{OLS},0} = \sigma^2(\mathbf{\Sigma}_{\Lambda})^{-1}$.

idiosyncratic	PC		WLS/LP	
1. $\boldsymbol{\Omega}^\xi$ full	$\min(\sqrt{n}, T)$	$\mathcal{W}_t^{\textsf{OLS}}$	$\min(\sqrt{n}, T)$	$\mathcal{W}_t^{\textsf{WLS}}$
2. $\boldsymbol{\Omega}^\xi = \mathbf{I}_T \otimes \boldsymbol{\Gamma}^\xi$	$\min(\sqrt{n}, T)$	$\mathcal{W}_t^{\textsf{OLS}}$	$\min(\sqrt{n}, T)$	$\mathcal{W}_t^{\textsf{WLS}}$
3. $\boldsymbol{\Omega}^\xi = \mathbf{I}_T \otimes \boldsymbol{\Sigma}^\xi$	$\min(\sqrt{n}, T)$	$\mathcal{W}_t^{\textsf{OLS,*}}$	$\min(\sqrt{n}, T)$	$\mathcal{W}_t^{\textsf{WLS,*}}$
4. $\boldsymbol{\Omega}^\xi = \sigma^2 \mathbf{I}_{nT}$	$\min(\sqrt{n}, T)$	$\mathcal{W}_t^{\textsf{OLS,0}}$	$\min(\sqrt{n}, T)$	$\mathcal{W}_t^{\textsf{OLS,0}}$

Asymptotic covariances

PC $\mathcal{W}_t^{\textsf{OLS}} = (\boldsymbol{\Sigma}_\Lambda)^{-1} \left\{ \lim_{n\to\infty} \frac{\mathsf{E}[\boldsymbol{\Lambda}' \mathsf{E}[\boldsymbol{\xi}_t \boldsymbol{\xi}_t'] \boldsymbol{\Lambda}]}{n} \right\} (\boldsymbol{\Sigma}_\Lambda)^{-1}$,

$\mathcal{W}_t^{\textsf{OLS,*}} = (\boldsymbol{\Sigma}_\Lambda)^{-1} \left\{ \lim_{n\to\infty} \frac{\mathsf{E}[\boldsymbol{\Lambda}' \boldsymbol{\Sigma}^\xi \boldsymbol{\Lambda}]}{n} \right\} (\boldsymbol{\Sigma}_\Lambda)^{-1}$, $\mathcal{W}_t^{\textsf{OLS,0}} = \sigma^2 (\boldsymbol{\Sigma}_\Lambda)^{-1}$

WLS/LP $\mathcal{W}_t^{\textsf{WLS}} = (\boldsymbol{\Sigma}_{\Lambda\xi\Lambda})^{-1} \left\{ \lim_{n\to\infty} \frac{\boldsymbol{\Lambda}' (\boldsymbol{\Sigma}^\xi)^{-1} \mathsf{E}[\boldsymbol{\xi}_t \boldsymbol{\xi}_t'] (\boldsymbol{\Sigma}^\xi)^{-1} \boldsymbol{\Lambda}}{n} \right\} (\boldsymbol{\Sigma}_{\Lambda\xi\Lambda})^{-1}$, $\mathcal{W}_t^{\textsf{WLS,*}} = (\boldsymbol{\Sigma}_{\Lambda\xi\Lambda})^{-1}$

$\boldsymbol{\Sigma}_\Lambda = \lim_{n\to\infty} \frac{\boldsymbol{\Lambda}'\boldsymbol{\Lambda}}{n}$, $\boldsymbol{\Sigma}_{\Lambda\xi\Lambda} = \lim_{n\to\infty} \frac{\boldsymbol{\Lambda}' (\boldsymbol{\Sigma}^\xi)^{-1} \boldsymbol{\Lambda}}{n}$, here either $\boldsymbol{\Sigma}_\Lambda$ or $\boldsymbol{\Sigma}_{\Lambda\xi\Lambda}$ are diagonal.

Estimators

PC $\widehat{\mathbf{F}}_t^{\textsf{PC}} = (\widehat{\boldsymbol{\Lambda}}^{\textsf{PC}'} \widehat{\boldsymbol{\Lambda}}^{\textsf{PC}})^{-1} \widehat{\boldsymbol{\Lambda}}^{\textsf{PC}'} \mathbf{x}_t$, case 1, 2, 3, 4;

WLS $\widehat{\mathbf{F}}_t^{\textsf{WLS}} = (\widehat{\boldsymbol{\Lambda}}^{\textsf{QML,S}'} (\widehat{\boldsymbol{\Sigma}}^{\xi,\textsf{QML,S}})^{-1} \widehat{\boldsymbol{\Lambda}}^{\textsf{QML,S}})^{-1} \widehat{\boldsymbol{\Lambda}}^{\textsf{QML,S}'} (\widehat{\boldsymbol{\Sigma}}^{\xi,\textsf{QML,S}})^{-1} \mathbf{x}_t$, case 1, 2, 3;

$\widehat{\mathbf{F}}_t^{\textsf{WLS}} = \widehat{\mathbf{F}}_t^{\textsf{PC}}$, case 4;

LP $\widehat{\mathbf{F}}_t^{\textsf{LP}} = (\widehat{\boldsymbol{\Lambda}}^{\textsf{QML,S}'} (\widehat{\boldsymbol{\Sigma}}^{\xi,\textsf{QML,S}})^{-1} \widehat{\boldsymbol{\Lambda}}^{\textsf{QML,S}} + \mathbf{I}_r)^{-1} \widehat{\boldsymbol{\Lambda}}^{\textsf{QML,S}'} (\widehat{\boldsymbol{\Sigma}}^{\xi,\textsf{QML,S}})^{-1} \mathbf{x}_t$, case 1, 2, 3;

$\widehat{\mathbf{F}}_t^{\textsf{LP}} = (\widehat{\boldsymbol{\Lambda}}^{\textsf{QML,0}'} \widehat{\boldsymbol{\Lambda}}^{\textsf{QML,0}} + \widehat{\sigma}^{2,\textsf{QML,0}} \mathbf{I}_r)^{-1} \widehat{\boldsymbol{\Lambda}}^{\textsf{QML,0}'} \mathbf{x}_t$

Can we do better than ML plus WLS/LP?

- In time series we could and should exploit the autocorrelation of the data.
- Factors are autocorrelated.
- Factors can have a lagged effect on the data.
- PC does not account for dynamics.
- ML is hard as it requires numerical maximization.

- **Approximate Dynamic Factor Model - Expectation Maximization**

For simplicity assume a VAR(1) dynamics:

$$\begin{aligned} x_{it} &= \boldsymbol{\lambda}_i'\mathbf{F}_t + \xi_{it}, \\ \mathbf{F}_t &= \mathbf{A}\mathbf{F}_{t-1} + \mathbf{v}_t, \\ \mathbf{v}_t &= \mathbf{H}\mathbf{u}_t. \end{aligned}$$

Same assumptions plus:

8 stable VAR, eigenvalues of $\mathbf{A}$ inside the unit circle;

9 $\text{rk}(\mathbf{H}) = q \leq r$;

10 $\{\mathbf{u}_t\}$ is i.i.d. with $\mathsf{E}[\mathbf{u}_t] = \mathbf{0}_r$, $\boldsymbol{\Gamma}^u = \mathbf{I}_q$, finite 4th order moments.

For simplicity hereafter we consider $r = q$ so $\boldsymbol{\Gamma}^v = \mathbf{H}\mathbf{H}' \succ 0$.

Since we are explicitly modeling the dynamics in the factors $\mathbf{\Omega}^F \equiv \mathbf{\Omega}^F(\mathbf{A}, \mathbf{\Gamma}^v)$, e.g, if $r = 1$,

$$\mathbf{\Omega}^F = \begin{pmatrix} \frac{\Gamma^v}{1-A^2} & \frac{A\Gamma^v}{1-A^2} & \cdots & \frac{\Gamma^v A^{T-1}}{1-A^2} \\ \frac{A\Gamma^v}{1-A^2} & \frac{\Gamma^v}{1-A^2} & \cdots & \frac{\Gamma^v A^{T-2}}{1-A^2} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{A^{T-1}\Gamma^v}{1-A^2} & \frac{A^{T-2}\Gamma^v}{1-A^2} & \cdots & \frac{\Gamma^v}{1-A^2} \end{pmatrix},$$

and we cannot assume it to be diagonal.

Gaussian quasi log-likelihood with mis-specified idiosyncratic correlations:

$$\begin{aligned}\ell_{0,D}(\boldsymbol{\mathcal{X}},\underline{\boldsymbol{\varphi}}) \simeq & -\frac{1}{2}\log\det\left(\underline{\boldsymbol{\mathcal{L}}}\,\underline{\boldsymbol{\Omega}}^F(\underline{\mathbf{A}},\underline{\boldsymbol{\Gamma}}^v)\underline{\boldsymbol{\mathcal{L}}}' + \mathbf{I}_T\otimes\underline{\boldsymbol{\Sigma}}^\xi\right)\\ & -\frac{1}{2}\left(\boldsymbol{\mathcal{X}}'(\underline{\boldsymbol{\mathcal{L}}}\,\underline{\boldsymbol{\Omega}}^F(\underline{\mathbf{A}},\underline{\boldsymbol{\Gamma}}^v)\underline{\boldsymbol{\mathcal{L}}}' + \mathbf{I}_T\otimes\underline{\boldsymbol{\Sigma}}^\xi)^{-1}\boldsymbol{\mathcal{X}}\right).\end{aligned}$$

The parameters to be estimated are $\boldsymbol{\varphi} = (\boldsymbol{\Lambda},\mathbf{A},\boldsymbol{\Gamma}^v,\boldsymbol{\Sigma}^\xi)$.

We work under the global identification assumptions.

Issues

1. How to estimate the factors? Kalman filter or Kalman smoother.
2. The likelihood is intractable, we need the factors as input and alternative maximization approaches.
   - Newton-Raphson maximization of the prediction error log-likelihood based on the Kalman filter. No closed form solution. Unfeasible in high-dimensions. (Harvey, 1990; Stock & Watson, 1989, 1991; Hannan & Deistler, 2012).
   - Multi-step approaches, but they do not exploit the feedback from factors to loadings.
     - PC+VAR (Forni, Giannone, Lippi & Reichlin, 2009);
     - PC+VAR+Kalman smoother (Doz, Giannone & Reichlin, 2011);
     - QML+WLS+VAR+Kalman smoother (Bai & Li, 2016).
   - Kalman smoother plus EM algorithm: fast, easy, and has closed form solution (Quah & Sargent, 1993; Doz, Giannone & Reichlin, 2012; Barigozzi & Luciani, 20xx).

Estimation of the factors.

- They are autocorrelated so cannot be treated as parameters.
- The optimal predictor is $\mathsf{E}_{\varphi}[\boldsymbol{\mathcal{F}}|\boldsymbol{\mathcal{X}}]$ which under Gaussianity is the linear projection

$$\begin{aligned}\mathbf{F}_t^{\mathsf{WK}} &= (\boldsymbol{\iota}_t' \otimes \mathbf{I}_r)\boldsymbol{\Omega}^F \boldsymbol{\mathcal{L}}'(\boldsymbol{\mathcal{L}}\boldsymbol{\Omega}^F\boldsymbol{\mathcal{L}}' + \mathbf{I}_T \otimes \boldsymbol{\Sigma}^\xi)^{-1}\boldsymbol{\mathcal{X}}\\ &= (\boldsymbol{\iota}_t' \otimes \mathbf{I}_r)\left(\mathbf{I}_T \otimes \left(\boldsymbol{\Lambda}'(\boldsymbol{\Sigma}^\xi)^{-1}\boldsymbol{\Lambda}\right) + (\boldsymbol{\Omega}^F)^{-1}\right)^{-1}\left(\mathbf{I}_T \otimes \boldsymbol{\Lambda}'(\boldsymbol{\Sigma}^\xi)^{-1}\right)\boldsymbol{\mathcal{X}}\end{aligned}$$

- This is the unfeasible estimator obtained by taking the inverse Fourier transform of the Wiener-Kolmogorov smoother.
- At a given t we compute a weighted average of the elements of $\boldsymbol{\mathcal{X}}$ which are all T present, past, and future values of all n time series

  $\Rightarrow$ **cross-sectional and dynamic weighted average!**

Estimation of the factors.

- $\mathbf{F}_t^{\mathsf{WK}}$ can be computed recursively by means of the Kalman smoother.
- The Kalman smoother is computed with a backward recursion from T to 1 after the Kalman filter which is a forward recursion from 1 to T.
- After these recursions we get the estimates:
  - one-step ahead $\mathbf{F}_{t|t-1}$ and its associated MSE $\mathbf{P}_{t|t-1}$;
  - Kalman filter $\mathbf{F}_{t|t}$ and its associated MSE $\mathbf{P}_{t|t}$;
  - Kalman smoother $\mathbf{F}_{t|T}$ and its associated MSE $\mathbf{P}_{t|T}$.

Estimation of the factors.

- The Kalman filter is

$$\mathbf{F}_{t|t} = \mathbf{F}_{t|t-1} + \underbrace{\mathbf{P}_{t|t-1}\boldsymbol{\Lambda}'(\boldsymbol{\Lambda}\mathbf{P}_{t|t-1}\boldsymbol{\Lambda} + \boldsymbol{\Sigma}^{\xi})^{-1}}_{\text{Kalman gain}}\underbrace{(\mathbf{x}_t - \boldsymbol{\Lambda}\mathbf{F}_{t|t-1})}_{\text{prediction error}}$$

$$= \mathbf{F}_{t|t-1} + (\boldsymbol{\Lambda}'(\boldsymbol{\Sigma}^{\xi})^{-1}\boldsymbol{\Lambda} + \mathbf{P}_{t|t-1}^{-1})^{-1}\boldsymbol{\Lambda}'(\boldsymbol{\Sigma}^{\xi})^{-1}(\mathbf{x}_t - \boldsymbol{\Lambda}\mathbf{F}_{t|t-1})$$

  with

  - $\mathbf{F}_{t|t-1} = \mathbf{A}\mathbf{F}_{t-1|t-1}$;
  - $\mathbf{P}_{t|t-1} = \mathbf{A}\mathbf{P}_{t-1|t-1}\mathbf{A}' + \boldsymbol{\Gamma}^{v}$;
  - $\mathbf{P}_{t|t} = \mathbf{P}_{t|t-1} - \mathbf{P}_{t|t-1}\boldsymbol{\Lambda}'(\boldsymbol{\Lambda}\mathbf{P}_{t|t-1}\boldsymbol{\Lambda} + \boldsymbol{\Sigma}^{\xi})^{-1}\boldsymbol{\Lambda}\mathbf{P}_{t|t-1}$.

- The Kalman smoother is

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

- Notice that we must use $\boldsymbol{\Sigma}^{\xi}$ since inverting $\boldsymbol{\Gamma}^{\xi}$ might not be feasible in high-dimensions. Mis-specified Kalman filter and smoother.

Prediction error log-likelihood

(Harvey, 1990; Stock & Watson, 1989, 1991; Hannan & Deistler, 2012).

$$\begin{aligned}\ell_{0,D}(\boldsymbol{\mathcal{X}},\underline{\boldsymbol{\varphi}}) = &-\frac{1}{2}\sum_{t=1}^{T}\log\det\mathbf{P}_{t|t-1}(\underline{\boldsymbol{\varphi}})\\ &-\frac{1}{2}\sum_{t=1}^{T}(\mathbf{x}_t-\underline{\boldsymbol{\Lambda}}\mathbf{F}_{t|t-1}(\underline{\boldsymbol{\varphi}}))'(\mathbf{P}_{t|t-1}(\underline{\boldsymbol{\varphi}}))^{-1}(\mathbf{x}_t-\underline{\boldsymbol{\Lambda}}\mathbf{F}_{t|t-1}(\underline{\boldsymbol{\varphi}}))\end{aligned}$$

Unfeasible to maximize in high-dimensions. No closed form solution.

By Bayes' law the log-likelihood is decomposed as

$$\ell_{0,D}(\boldsymbol{\mathcal{X}}, \underline{\boldsymbol{\varphi}}) = \ell_{0,D}(\boldsymbol{\mathcal{X}}|\boldsymbol{\mathcal{F}}, \underline{\boldsymbol{\varphi}}) + \ell_{0,D}(\boldsymbol{\mathcal{F}}, \underline{\boldsymbol{\varphi}}) - \ell_{0,D}(\boldsymbol{\mathcal{F}}|\boldsymbol{\mathcal{X}}, \underline{\boldsymbol{\varphi}}).$$

where

$$\ell_{0,D}(\boldsymbol{\mathcal{X}}|\boldsymbol{\mathcal{F}}, \underline{\boldsymbol{\varphi}}) \simeq -\frac{T}{2}\log\det(\underline{\boldsymbol{\Sigma}}^{\xi}) - \frac{1}{2}\sum_{t=1}^{T}\Big((\mathbf{x}_t - \underline{\boldsymbol{\Lambda}}\mathbf{F}_t)'(\underline{\boldsymbol{\Sigma}}^{\xi})^{-1}(\mathbf{x}_t - \underline{\boldsymbol{\Lambda}}\mathbf{F}_t)\Big),$$

$$\ell_{0,D}(\boldsymbol{\mathcal{F}}, \underline{\boldsymbol{\varphi}}) \simeq -\frac{T}{2}\log\det(\underline{\boldsymbol{\Gamma}}^{v}) - \frac{1}{2}\sum_{t=1}^{T}\big((\mathbf{F}_t - \underline{\mathbf{A}}\mathbf{F}_{t-1})'(\underline{\boldsymbol{\Gamma}}^{v})^{-1}(\mathbf{F}_t - \underline{\mathbf{A}}\mathbf{F}_{t-1})\big).$$

Easy to maximize if $\mathbf{F}_t$ is known.

The hard part would be to maximize $\ell_{0,D}(\boldsymbol{\mathcal{F}}|\boldsymbol{\mathcal{X}}, \underline{\boldsymbol{\varphi}})$ but it is not needed.

EM algorithm.

$$\ell_{0,D}(\boldsymbol{\mathcal{X}},\underline{\boldsymbol{\varphi}}) = \underbrace{\mathsf{E}_{\boldsymbol{\varphi}}\left[\ell_{0,D}(\boldsymbol{\mathcal{X}}|\boldsymbol{\mathcal{F}},\underline{\boldsymbol{\varphi}}) + \ell_{0,D}(\boldsymbol{\mathcal{F}},\underline{\boldsymbol{\varphi}})|\boldsymbol{\mathcal{X}}\right]}_{\mathcal{Q}(\underline{\boldsymbol{\varphi}},\boldsymbol{\varphi})} - \underbrace{\mathsf{E}_{\boldsymbol{\varphi}}\left[\ell_{0,D}(\boldsymbol{\mathcal{F}}|\boldsymbol{\mathcal{X}},\underline{\boldsymbol{\varphi}})|\boldsymbol{\mathcal{X}}\right]}_{\mathcal{H}(\underline{\boldsymbol{\varphi}},\boldsymbol{\varphi})}.$$

For any $k \geq 0$, given an estimator of the parameters $\widehat{\boldsymbol{\varphi}}^{(k)}$.

E Compute $\mathcal{Q}(\underline{\boldsymbol{\varphi}},\widehat{\boldsymbol{\varphi}}^{(k)})$.

M Solve $\widehat{\boldsymbol{\varphi}}^{(k+1)} = \arg\max_{\underline{\boldsymbol{\varphi}}} \mathcal{Q}(\underline{\boldsymbol{\varphi}},\widehat{\boldsymbol{\varphi}}^{(k)})$.

Start with PCA, e.g. $\widehat{\boldsymbol{\Lambda}}^{(0)} = \widehat{\boldsymbol{\Lambda}}^{\mathsf{PC}}$.

Stop at k^* s.t. $|\ell_{0,D}(\boldsymbol{\mathcal{X}},\widehat{\boldsymbol{\varphi}}^{(k^*+1)}) - \ell_{0,D}(\boldsymbol{\mathcal{X}},\widehat{\boldsymbol{\varphi}}^{(k^*)})|$ is small.

The EM estimator is $\widehat{\boldsymbol{\varphi}}^{\mathsf{EM}} = \widehat{\boldsymbol{\varphi}}^{(k^*+1)}$.

Main intuition

By construction $\mathcal{H}(\widehat{\boldsymbol{\varphi}}^{(k)},\widehat{\boldsymbol{\varphi}}^{(k)}) \leq \mathcal{H}(\underline{\boldsymbol{\varphi}},\widehat{\boldsymbol{\varphi}}^{(k)})$ for any $\underline{\boldsymbol{\varphi}}$, so

$$\ell_{0,D}(\boldsymbol{\mathcal{X}},\widehat{\boldsymbol{\varphi}}^{(k+1)}) \geq \ell_{0,D}(\boldsymbol{\mathcal{X}},\widehat{\boldsymbol{\varphi}}^{(k)}).$$

EM estimators.

- The EM estimator of the loadings is:

$$\widehat{\boldsymbol{\lambda}}_i^{\mathsf{EM}} = \left(\sum_{t=1}^{T} \mathbf{F}_{t|T}^{(k^*)} \mathbf{F}_{t|T}^{(k^*)'} + \mathbf{P}_{t|T}^{(k^*)}\right)^{-1} \left(\sum_{t=1}^{T} \mathbf{F}_{t|T}^{(k)^*} x_{it}\right),$$

  where $\mathbf{F}_{t|T}^{(k^*)}$ and $\mathbf{P}_{t|T}^{(k^*)}$ are obtained from Kalman smoother when using $\widehat{\boldsymbol{\varphi}}^{(k^*)}$.

- The EM estimator of the factors is $\widehat{\mathbf{F}}_t^{\mathsf{EM}} = \mathbf{F}_{t|T}^{(k^*+1)}$.

- Both have a closed form expression!

Asymptotic properties EM estimator - Loadings

(Barigozzi & Luciani, 20xx).

- for any given $i = 1, \dots, n$ as $n, T \to \infty$

$$\|\widehat{\boldsymbol{\lambda}}_i^{\text{EM}} - \widehat{\boldsymbol{\lambda}}_i^{\text{QML,D}}\| = O_p\left(\frac{1}{n}\right) + O_p\left(\frac{1}{T}\right) + O_p\left(\frac{1}{\sqrt{nT}}\right)$$

$$\|\widehat{\boldsymbol{\lambda}}_i^{\text{QML,D}} - \widehat{\boldsymbol{\lambda}}_i^{\text{QML,S}}\| = O_p\left(\frac{1}{n}\right) + O_p\left(\frac{1}{T}\right) + O_p\left(\frac{1}{\sqrt{nT}}\right)$$

- if $\frac{\sqrt{T}}{n} \to 0$, then

$$\sqrt{T}(\widehat{\boldsymbol{\lambda}}_i^{\text{EM}} - \boldsymbol{\lambda}_i) \to_d \mathcal{N}\left(\mathbf{0}_r, \boldsymbol{\mathcal{V}}_i^{\text{OLS}}\right),$$

$$\boldsymbol{\mathcal{V}}_i^{\text{OLS}} = (\boldsymbol{\Gamma}^F)^{-1}\left\{\lim_{T\to\infty} \frac{\mathsf{E}[\boldsymbol{F}'\boldsymbol{\zeta}_i\boldsymbol{\zeta}_i'\boldsymbol{F}]}{T}\right\}(\boldsymbol{\Gamma}^F)^{-1} = \lim_{T\to\infty} \frac{\mathsf{E}[\boldsymbol{F}'\boldsymbol{\zeta}_i\boldsymbol{\zeta}_i'\boldsymbol{F}]}{T}.$$

- EM is asymptotically equivalent to QML of a dynamic as well as of a static model and to PC and OLS.
- Since the EM is initialized with PC then the loadings estimator is similar to a one step estimator (Lehmann & Casella, 2006).

Asymptotic properties EM estimator - Factors

(Barigozzi & Luciani, 20xx).

- for any given $t = 1, \dots, T$ as $n, T \to \infty$

$$\|\widehat{\mathbf{F}}_t^{\mathsf{EM}} - \widehat{\mathbf{F}}_{t|t}\| = O_p\left(\frac{1}{n}\right), \qquad \|\widehat{\mathbf{F}}_{t|t} - \widehat{\mathbf{F}}_t^{\mathsf{WLS}}\| = O_p\left(\frac{1}{n}\right)$$

- if $\frac{\sqrt{n}}{T} \to 0$, then

$$\sqrt{n}(\widehat{\mathbf{F}}_t^{\mathsf{EM}} - \mathbf{F}_t) \to_d \mathcal{N}\left(\mathbf{0}_r, \boldsymbol{\mathcal{W}}^{\mathsf{WLS}}\right),$$

$$\boldsymbol{\mathcal{W}}^{\mathsf{WLS}} = \boldsymbol{\Sigma}_{\Lambda\xi\Lambda}^{-1}\left(\lim_{n\to\infty} \frac{\boldsymbol{\Lambda}'(\boldsymbol{\Sigma}^{\xi})^{-1}\mathsf{E}[\boldsymbol{\xi}_t\boldsymbol{\xi}_t'](\boldsymbol{\Sigma}^{\xi})^{-1}\boldsymbol{\Lambda}}{n}\right)\boldsymbol{\Sigma}_{\Lambda\xi\Lambda}^{-1}.$$

- EM, which is the Kalman smoother, is asymptotically equivalent to the Kalman filter and to the WLS and LP.
- It can be more efficient than PC if $\boldsymbol{\Gamma}^{\xi}$ is sparse.

Asymptotic properties. Common component.

(Barigozzi & Luciani, 20xx).

- For any given $i = 1, \dots, n$ and $t = 1, \dots, T$

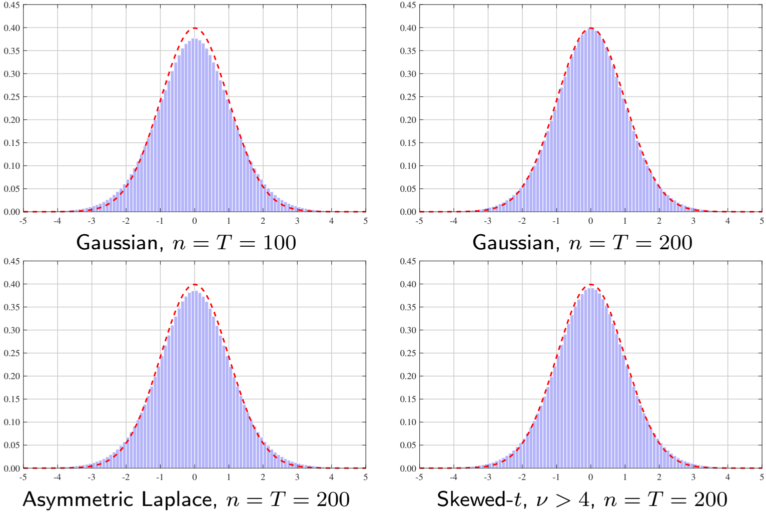
$$|\widehat{\chi}_{it}^{\mathsf{EM}} - \chi_{it}| = O_p\left(\frac{1}{\sqrt{T}}\right) + O_p\left(\frac{1}{\sqrt{n}}\right)$$

  with $\widehat{\chi}_{it}^{\mathsf{EM}} = \widehat{\boldsymbol{\lambda}}_i^{\mathsf{EM}\prime} \widehat{\mathbf{F}}_t^{\mathsf{EM}}$.

- And, as $n, T \to \infty$,

$$\frac{(\widehat{\chi}_{it}^{\mathsf{EM}} - \chi_{it})}{\left(\frac{\boldsymbol{\lambda}_i' \boldsymbol{\mathcal{W}}_t^{\mathsf{WLS}} \boldsymbol{\lambda}_i}{n} + \frac{\mathbf{F}_t' \boldsymbol{\mathcal{V}}_i^{\mathsf{OLS}} \mathbf{F}_t}{T}\right)^{1/2}} \to_d \mathcal{N}(0, 1).$$

Asymptotic distribution of common component
Serially and cross-correlated idiosyncratic components - Robust covariances



Kalman smoother and WLS.

- In the case $r = 1$ (Ruiz & Poncela, 2022).

$$F_{t|T} = \frac{2A}{2+B}\left(F_{t-1|t-1} + F_{t+1|T} - F_{t+1|t}\right) + \frac{B}{2+B}F_t^{\mathsf{WLS}},$$

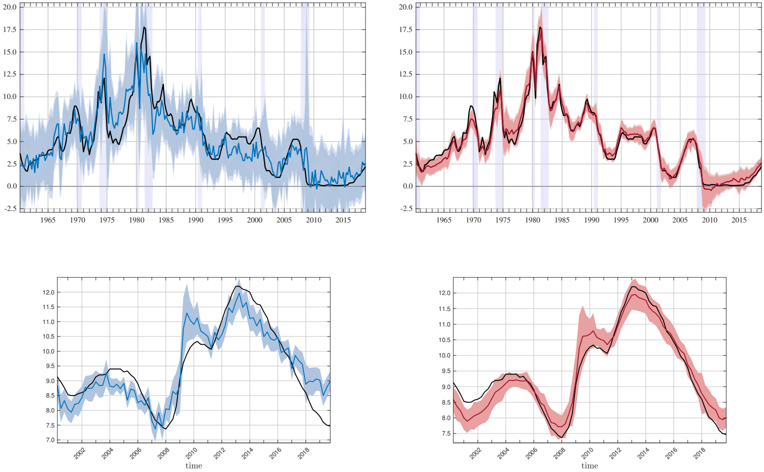
with $B = 2(\mathbf{\Lambda}'(\mathbf{\Gamma}^{\xi})^{-1}\mathbf{\Lambda})P$ and $P \simeq P_{t|t-1}$ for all $t \geq \bar{t}$ finite.

- By assumption $B \asymp n$ and $|P - \Gamma^v| = o(1)$, so as $n \to \infty$, $|F_{t|T} - F_t^{\mathsf{WLS}}| \to 0$.
- But if factors are persistent $A \lesssim 1$ and do not fluctuate much $\Gamma^v \gtrsim 0$, then, at least in finite samples there might be considerable differences between the Kalman smoother and the WLS.

- EM for loadings is as good as PC.
- Kalman smoother for factors is equivalent to WLS which might be more efficient than PC.
- Why not PC or just QML+WLS?
- EM+Kalman smoother is the most used method in institutions because it allows for:
  - missing data and mixed frequency, e.g., for now-casting;
  - imposing constraints, e.g., for identification.
- Kalman smoother might have better finite sample performance than WLS in presence of small deviations for stationarity.

US Fed Funds rate

EA Unemployment rate



QML+WLS

EM+Kalman smoother

- **Generalized Dynamic Factor Model**

Define the spectral density matrix of $\{\mathbf{x}_t\}$ (Discrete Fourier Transform, DFT):

$$\mathbf{\Sigma}^x(\theta) = \frac{1}{2\pi} \sum_{k=-\infty}^{\infty} \mathbf{\Gamma}_k^x e^{-\iota\theta k}, \quad \theta \in [-\pi, \pi],$$

where $\iota = \sqrt{-1}$ and $\mathbf{\Gamma}_k^x = \mathsf{E}[\mathbf{x}_t \mathbf{x}_{t-k}]$ (recall $\mathbf{\Gamma}_{-k}^x = \mathbf{\Gamma}_k^{x'}$), such that (Inverse Fourier Transform, IFT):

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

The eigenvalues of $\mathbf{\Sigma}^x(\theta)$ denoted as $\mu_j^x(\theta)$ are called dynamic eigenvalues.

The GDFM is:

$$x_{it} = \underbrace{\boldsymbol{\lambda}_i^{*'}(L)\boldsymbol{f}_t}_{\chi_{it}} + \xi_{it}, \quad \boldsymbol{f}_t = \mathbf{G}(L)\mathbf{u}_t$$

$$x_{it} = \boldsymbol{\lambda}_i^{*'}(L)\mathbf{G}(L)\mathbf{u}_t + \xi_{it} = \underbrace{\mathbf{b}_i'(L)\mathbf{u}_t}_{\chi_{it}} + \xi_{it}$$

Then, the vector of factors is an orthonormal white noise $\mathbf{u}_t$.

Same assumptions as the approximate factor model plus:

A $\mathbf{b}_i(L)$ has square-summable coefficients;

B $\boldsymbol{\Sigma}^{\chi}(\theta)$ is rational;

C $\underline{c}_j(\theta) \leq \liminf_{n\to\infty} \frac{\mu_j^{\chi}(\theta)}{n} \leq \limsup_{n\to\infty} \frac{\mu_j^{\chi}(\theta)}{n} \leq \bar{c}_j(\theta),\ j = 1, \dots, q, \theta$-a.e.;

D $\sup_{n\in\mathbb{N}} \sup_{\theta\in[-\pi,\pi]} \mu_1^{\xi}(\theta) \leq M$.

Recall that

- if order of $\boldsymbol{\lambda}_i^{*'}(L)$ is $s < \infty$ restricted GDFM;
- if order of $\boldsymbol{\lambda}_i^{*'}(L)$ is $s = \infty$ unrestricted GDFM or GDFM.

Representation Theorem (Forni & Lippi, 2001).

$\mathbf{x}_t$ admits a Generalized Dynamic Factor representation with

$$\lim_{n\to\infty} \mu_q^{\chi}(\theta) = \infty, \quad \theta\text{-a.e. in } [-\pi, \pi],$$
$$\sup_{n\in\mathbb{N}} \sup_{\theta\in[-\pi,\pi]} \mu_1^{\xi}(\theta) \le M.$$

$\Updownarrow$ if and only if

$$\lim_{n\to\infty} \mu_q^{x}(\theta) = \infty, \quad \theta\text{-a.e. in } [-\pi, \pi],$$
$$\sup_{n\in\mathbb{N}} \sup_{\theta\in[-\pi,\pi]} \mu_{q+1}^{x}(\theta) \le M.$$

- By Weyl's inequality we easily prove $\Downarrow$.
- To prove $\Uparrow$ is much more difficult and in general is not true for eigenvalues of a covariance matrix, so the static factor model is not a representation result.
- As $n \to \infty$ we identify the number of factors!

Representation Theorem (Hallin & Lippi, 2013).

- $\mathcal{H}^{\mathbf{x}}$ the Hilbert space of all L_2-convergent linear combinations of x_{it}'s and limits of L_2-convergent sequences thereof.
- Let $w_{\mathbf{x}}^{(n)} \in \mathcal{H}^{\mathbf{X}}$ such that

$$w_{\mathbf{X}}^{(n)} = \sum_{i=1}^{n} \sum_{k=-\infty}^{\infty} a_{ik}^{(n)} x_{i,t-k}, \quad \lim_{n\to\infty} \mathrm{Var}(w_{\mathbf{X}}^{(n)}) = \infty,$$

with $\sum_{i=1}^{n} \sum_{k=-\infty}^{\infty} (a_{ik}^{(n)})^2 = 1$.

- $\zeta \in \mathcal{H}_{com}^{\mathbf{X}}$ if with $\mathrm{Var}(\zeta) > 0$ and

$$\lim_{n\to\infty} \mathsf{E}\left[\left(\frac{w_{\mathbf{X}}^{(n)}}{\sqrt{\mathrm{Var}(w_{\mathbf{X}}^{(n)})}} - \frac{\zeta}{\sqrt{\mathrm{Var}(\zeta)}}\right)^2\right] = 0.$$

a common r.v. is recovered as $n \to \infty$ by dynamic aggregation.

- Let also $\mathcal{H}_{idio}^{\mathbf{X}} = \mathcal{H}_{com,\perp}^{\mathbf{X}}$
- So $\mathcal{H}^{\mathbf{X}} = \mathcal{H}_{com}^{\mathbf{X}} \oplus \mathcal{H}_{idio}^{\mathbf{X}}$.

Dynamic weighted averages. Large n to recover factors.

- Take any $n \times r$ filter matrix $\boldsymbol{W}_u(L) = (\boldsymbol{w}_{u,1}(L) \cdots \boldsymbol{w}_{u,n}(L))'$ and such that

$$n^{-1}\boldsymbol{W}_u(L)'\mathbf{B}(L) = \mathbf{K}(L) \succ 0, \qquad n^{-1}\sum_{i=1}^{n}\sum_{k=-\infty}^{\infty}\boldsymbol{w}_{u,ik}\boldsymbol{w}_{u,ik}' = \mathbf{I}_r$$

and with coefficients $\|\boldsymbol{w}_{u,ik}\| \leq c$ for some $c > 0$ independent of i.

- For any given t an estimator of a linear dynamic combination of the factors is

$$\begin{aligned}\check{\mathbf{u}}_t = \frac{\boldsymbol{W}_u(L)'\mathbf{x}_t}{n} &= \frac{\boldsymbol{W}_u(L)'\mathbf{B}(L)\mathbf{u}_t}{n} + \frac{\boldsymbol{W}_u(L)'\boldsymbol{\xi}_t}{n} \\ &= \mathbf{K}(L)\mathbf{u}_t + \frac{1}{n}\sum_{i=1}^{n}\sum_{k=-\infty}^{\infty}\boldsymbol{w}_{u,ik}\xi_{i,t-k}.\end{aligned}$$

- By dynamic averaging we do not recover white noise factors, but in general we obtain autocorrelated factors.

- Then we have $\sqrt{n}$-consistency if as $n \to \infty$ (assume $q = 1$ for simplicity):

$$\mathsf{E}\left[\left|\frac{1}{n}\sum_{i=1}^{n}\sum_{k=-\infty}^{\infty} w_{u,ik}\xi_{i,t-k}\right|^2\right] \le \frac{c^2}{n}\frac{\boldsymbol{\iota}'\boldsymbol{\Sigma}^{\xi}(0)\boldsymbol{\iota}}{n} \le \frac{c^2}{n}\mu_1^{\xi}(0) = O\left(\frac{1}{n}\right),$$

or

$$\mathsf{E}\left[\left|\frac{1}{n}\sum_{i=1}^{n}\sum_{k=-\infty}^{\infty} w_{u,ik}\xi_{i,t-k}\right|^2\right] \le \frac{c^2}{n^2}\sum_{i,j=1}^{n}\sum_{k,h=-\infty}^{\infty} |\mathsf{E}[\xi_{i,t-k}\xi_{j,t-h}]| = O\left(\frac{1}{n}\right).$$

if we assume summability of cross-covariances and standard summability of cross-autocovariances.

Dynamic PC - Population

- Consider the case of one factor, $q = 1$.
- In the static case we know that the optimal weights are given by the solution of PCs, which in population are such that we solve $\max_{\mathbf{a}:\mathbf{a}'\mathbf{a}=1} \frac{\mathbf{a}'\mathbf{\Gamma}^x\mathbf{a}}{n}$.
- In the dynamic case to find the optimal weights we have to maximize the variance of $\boldsymbol{a}'(L)\mathbf{x}_t = \sum_{k=-\infty}^{\infty} \boldsymbol{a}_k \mathbf{x}_{t-k}$ such that the coefficients $\boldsymbol{a}_k$ are the solution of

$$\max_{\boldsymbol{a}_k:\boldsymbol{a}'(e^{\iota\theta})\boldsymbol{a}(e^{-\iota\theta})=1} \frac{\boldsymbol{a}'(e^{\iota\theta})\mathbf{\Sigma}^x(\theta)\boldsymbol{a}(e^{-\iota\theta})}{n}$$

  where $\mathbf{a}(e^{-\iota\theta}) = \sum_{k=-\infty}^{\infty} \boldsymbol{a}_k e^{-k\iota\theta}$.
- The solution is given by $\mathbf{P}^x(\theta)$ the leading eigenvector of $\mathbf{\Sigma}^x(\theta)$ and the value of the objective function is $n^{-1}\mu_1^x(\theta)$.
- The common component is the IFT of the linear projection onto the 1st PC:

$$\widetilde{\boldsymbol{\chi}}_t = \left\{ \sum_{k=-\infty}^{\infty} \left[\int_{-\pi}^{\pi} \mathbf{P}^x(\theta)\mathbf{P}^{x\dagger}(\theta)e^{\iota\theta k}\mathrm{d}\theta \right] L^k \right\} \mathbf{x}_t = \mathbf{K}'(L)\mathbf{x}_t$$

- By dynamic averaging we do not recover one-sided filters (dynamic loadings), but in general we obtain two-sided filters.

Estimation of unrestricted GDFM - Dynamic PC

(Forni, Hallin, Lippi & Recihlin, 2000).

- Consider the smoothed periodogram estimator of the spectral density matrix:

$$\widehat{\mathbf{\Sigma}}(\theta_h) = \frac{1}{2\pi} \sum_{k=-B_T}^{B_T} \left(1 - \frac{|k|}{B_T}\right) \widehat{\mathbf{\Gamma}}_k^x e^{-\iota \theta_h k}, \quad \theta_h = \frac{\pi h}{B_T}, \quad |h| \leq B_T,$$

  where $\iota = \sqrt{-1}$ and (recall $\widehat{\mathbf{\Gamma}}_{-k}^x = \widehat{\mathbf{\Gamma}}_k^{x\prime}$) $\widehat{\mathbf{\Gamma}}_k^x = \frac{1}{T-k} \sum_{t=k+1}^{T} \mathbf{x}_t \mathbf{x}_{t-k}$. Let,
  - $\widehat{\mathbf{L}}(\theta_h)$ be the $q \times q$ diagonal matrix of q largest eigenvalues of $\widehat{\mathbf{\Sigma}}(\theta_h)$;
  - $\widehat{\mathbf{P}}(\theta_h)$ be the $n \times q$ matrix of normalized eigenvectors of $\widehat{\mathbf{\Sigma}}(\theta_h)$.
- The common component is estimated as

$$\widehat{\boldsymbol{\chi}}_t^{\mathsf{DPC}} = \sum_{k=-M_T}^{M_T} \left[\sum_{h=-B_T}^{B_T} \widehat{\mathbf{P}}^x(\theta_h) \widehat{\mathbf{P}}^{x\dagger}(\theta_h) e^{\iota \theta_h k} \right] \mathbf{x}_{t-k} = \widehat{\mathbf{K}}(L) \mathbf{x}_t,$$

  for some truncation integer M_T.

Asymptotic properties of dynamic PC estimator - Common component.

(Barigozzi, La Vecchia & Liu, 2023).

- For any given $i = 1, \ldots, n$ and $t = 1, \ldots, T$

$$|\widehat{\chi}_{it}^{\mathsf{DPC}} - \chi_{it}| = O_p\left(\frac{M_T}{\sqrt{n}}\right) + O_p\left(\sqrt{\frac{M_T^2 B_T \log B_T}{T}}\right) + O_p\left(\frac{M_T}{B_T}\right)$$

- The optimal bandwidth is $B_T \simeq T^{1/3}$.
- If we use smoother kernel we can get better rates. E.g., with quadratic kernel $B_T \simeq T^{2/5}$.
- It depends on the truncation M_T.
- No asymptotic distribution is available.

Estimation of restricted GDFM - Dynamic + static PC

(Forni, Hallin, Lippi & Recihlin, 2005).

- From dynamic PC we also get

$$\widehat{\mathbf{\Sigma}}^{\chi}(\theta_h) = \widehat{\mathbf{P}}(\theta_h)\widehat{\mathbf{L}}(\theta_h)\widehat{\mathbf{P}}^{\dagger}(\theta_h), \quad \theta_h = \frac{\pi h}{B_T}, \quad |h| \leq B_T$$

and $\widehat{\mathbf{\Sigma}}^{\xi}(\theta_h) = \widehat{\mathbf{\Sigma}}^{x}(\theta_h) - \widehat{\mathbf{\Sigma}}^{\chi}(\theta_h)$.

- By IFT

$$\widehat{\mathbf{\Gamma}}_k^{\chi} = \sum_{h=-B_T}^{B_T} \widehat{\mathbf{\Sigma}}^{\chi}(\theta_h)e^{\iota\theta_h k}, \quad \widehat{\mathbf{\Gamma}}_k^{\xi} = \sum_{h=-B_T}^{B_T} \widehat{\mathbf{\Sigma}}^{\xi}(\theta_h)e^{\iota\theta_h k}, \quad |k| \leq B_T.$$

- In restricted GDFM: $\boldsymbol{\chi}_t = \mathbf{\Lambda}\mathbf{F}_t$ with $\mathbf{F}_t = (\mathbf{u}_t \cdots \mathbf{u}_{t-s})'$ and $q(s+1) = r$.
- Use r PCs on $\widehat{\mathbf{\Gamma}}_0^{\chi}$ having as r leading eigenvectors $\widehat{\mathbf{V}}^{\chi}$

$$\widehat{\boldsymbol{\chi}}_t^{\text{FHLR}} = \widehat{\mathbf{V}}^{\chi}\widehat{\mathbf{V}}^{\chi\prime}\mathbf{x}_t$$

- It accounts for dynamic loadings since in the first step we use dynamic PC.
- To account for heteroskedasticity use the eigenvectors of $\widehat{\mathbf{\Gamma}}_0^{\chi}(\widehat{\mathbf{\Sigma}}^{\xi})^{-1}$, with $\widehat{\mathbf{\Sigma}}^{\xi}$ the diagonal of $\widehat{\mathbf{\Gamma}}_0^{\xi}$.

Asymptotic properties of dynamic + static PC estimator - Common component.
(Barigozzi, Cho & Owens, 2023).

- For any given $i = 1, \ldots, n$ and $t = 1, \ldots, T$

$$|\widehat{\chi}_{it}^{\mathsf{FHLR}} - \chi_{it}| = O_p\left(\frac{1}{\sqrt{n}}\right) + O_p\left(\sqrt{\frac{B_T \log B_T}{T}}\right) + O_p\left(\frac{1}{B_T}\right)$$

- The optimal bandwidth is $B_T \simeq T^{1/3}$.
- If we use smoother kernel we can get better rates. E.g., with quadratic kernel $B_T \simeq T^{2/5}$.
- No asymptotic distribution is available.

Unrestricted GDFM - one-sided representation

(Anderson & Deistler, 2008; Forni, Hallin, Lippi & Zaffaroni, 2015).

- The unrestricted GDFM has an equivalent representation

$$\mathbf{A}(L)\mathbf{x}_t = \mathbf{R}\mathbf{u}_t + \mathbf{A}(L)\boldsymbol{\xi}_t$$

  where

  - $\mathbf{A}(L)$ has finite lag, is block diagonal, with blocks of size at least $q+1$;
  - $\mathbf{R}$ is $n \times q$ full rank;
  - $\mathbf{A}(L)\boldsymbol{\xi}_t$ is still idiosyncratic.

- We can assume that the q largest eigenvalues of $\mathbf{R}\mathbf{R}'$ diverging with n.

Estimation of unrestricted GDFM - Dynamic PC + VAR + static PC

(Forni, Hallin, Lippi & Zaffaroni, 2017).

- From dynamic PC and IFT we get $\widehat{\mathbf{\Gamma}}_k^{\chi}$, for $|k| \leq B_T$.
- Estimate VAR(p) on each block by Yule-Walker, e.g., for $p = 1$, $\widehat{\mathbf{A}} = (\widehat{\mathbf{\Gamma}}_0^{\chi})^{-1}\widehat{\mathbf{\Gamma}}_1^{\chi}$.
- Compute the q-largest PCs for the filtered process $\widehat{\mathbf{v}}_t = \widehat{\mathbf{A}}(L)\mathbf{x}_t$ which is now a white noise with covariance $\widehat{\mathbf{\Gamma}}^v$ having the q leading eigenvectors $\widehat{\mathbf{V}}^v$ and eigenvalues $\widehat{\mathbf{M}}^v$

$$\widehat{\mathbf{R}} = \widehat{\mathbf{V}}^v(\widehat{\mathbf{M}}^v)^{1/2}, \qquad \widehat{\mathbf{u}}_t = (\widehat{\mathbf{M}}^v)^{-1/2}\widehat{\mathbf{V}}^{v\prime}\widehat{\mathbf{v}}_t.$$

- The common component is estimated as (say $p = 1$ for simplicity)

$$\widehat{\boldsymbol{\chi}}_t^{\mathsf{FHLZ}} = \sum_{k=0}^{M_T} \widehat{\mathbf{A}}^k\widehat{\mathbf{R}}\widehat{\mathbf{u}}_{t-k}$$

for some truncation integer M_T.

Asymptotic properties of dynamic PC + VAR + static PC estimator - Common component - Consistency.

(Barigozzi, Cho & Owens, 2023).

- For any given $i = 1, \ldots, n$ and $t = 1, \ldots, T$

$$|\widehat{\chi}_{it}^{\mathsf{FHLZ}} - \chi_{it}| = O_p\left(\frac{M_T}{\sqrt{n}}\right) + O_p\left(\sqrt{\frac{M_T^2 B_T \log B_T}{T}}\right) + O_p\left(\frac{M_T}{B_T}\right).$$

- The optimal bandwidth is $B_T \simeq T^{1/3}$.
- If we use smoother kernel we can get better rates. E.g., with quadratic kernel $B_T \simeq T^{2/5}$.
- It depends on the truncation M_T.

Estimation of unrestricted GDFM - Dynamic PC + VAR + static PC

(Barigozzi, Hallin, Luciani & Zaffaroni, 2023).

- Let: $\zeta_{nT} = \min\left(\frac{\sqrt{n}}{M_T}, \sqrt{\frac{T}{M_T^2 B_T \log B_T}}, \frac{B_T}{M_T}\right)$, such that $\zeta_{nT} \to \infty$, as $n, T \to \infty$.
- Let $\bar{n} = \frac{\zeta_{nT}^2}{L_1(\zeta_{nT})}$ and $\bar{T} = \frac{\zeta_{nT}^2}{L_2(\zeta_{nT})}$ for some functions $L_1(\cdot)$ and $L_2(\cdot)$ slowly varing at infinity.
- In the last step consider the PC estimators $\check{\mathbf{R}}$ and $\check{\mathbf{u}}_{t-k}$ obtained from

$$\check{\mathbf{\Gamma}}^v = \frac{1}{\bar{T}} \sum_{t=T-\bar{T}+1}^{T} (\widehat{v}_{s(1),t} \cdots \widehat{v}_{s(\bar{n}),t})'(\widehat{v}_{s(1),t} \cdots \widehat{v}_{s(\bar{n}),t}),$$

  for some $\{s(1), \dots, s(\bar{n})\} \subset \{1, \dots, n\}$.
- Consider the resulting estimated common component (say $p = 1$ for simplicity)

$$\check{\boldsymbol{\chi}}_t^{\mathsf{FHLZ}} = \sum_{k=0}^{M_T} \check{\mathbf{A}}^k \check{\mathbf{R}} \check{\mathbf{u}}_{t-k}$$

  where $\check{\mathbf{A}}$ is $\bar{n} \times \bar{n}$ using only the rows and columns $\{s(1), \dots, s(\bar{n})\}$.

Asymptotic properties of dynamic PC + VAR + static PC estimator - Common component - Asymptotic distribution.

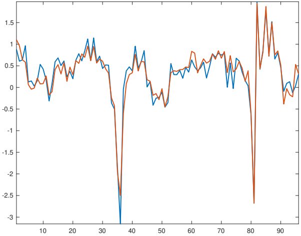
(Barigozzi, Hallin, Luciani & Zaffaroni, 2023).

For any given $i \in \{s(1), \ldots, s(\bar{n})\}$ and $t = T - \bar{T} + 1, \ldots, T$, as $n, T \to \infty$ we can neglect the error of the first two steps

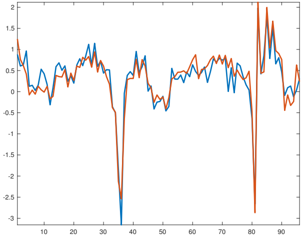
$$\frac{(\check{\chi}_{it}^{\mathsf{FLHZ}} - \chi_{it})}{\left(\frac{\boldsymbol{r}_i' \boldsymbol{\mathcal{W}}_t^{\mathsf{PC}} \boldsymbol{r}_i}{\bar{n}} + \frac{\mathbf{u}_t' \boldsymbol{\mathcal{V}}_i^{\mathsf{PC}} \mathbf{u}_t}{\bar{T}}\right)^{1/2}} \to_d \mathcal{N}(0, 1),$$

with obvious definitions of $\boldsymbol{\mathcal{W}}_t^{\mathsf{PC}}$ and $\boldsymbol{\mathcal{V}}_i^{\mathsf{PC}}$.

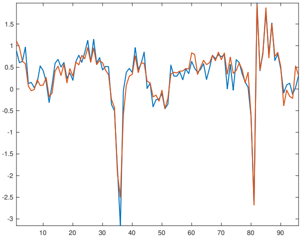
Common component (red) of EA GDP growth rate (blue)



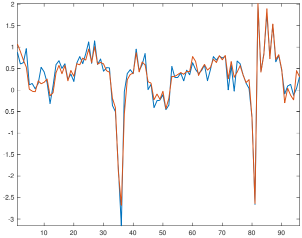
PC



dynamic PC



dynamic + static PC



dynamic PC + VAR + static PC

- **Applications and Extensions**

- Forecasting
- Coincident indicators
- IRFs
- The case of unit roots

Direct forecasts

- Let y_t be a target variable and let the predictors be $\mathbf{z}_t = \boldsymbol{\mu}_z + \boldsymbol{\Lambda}_z \mathbf{F}_t + \boldsymbol{\xi}_{zt}$.
- Instead of regressing y_{t+h} onto $\mathbf{z}_t$ we can use the factors $\mathbf{F}_t$ as proxies of the predictors.
- In fact we can also have $y_t = \mu_y + \boldsymbol{\lambda}_y' \mathbf{F}_t + \xi_{yt}$ so y_t is also driven by the same factors.
- Let $\mathbf{x}_t = (y_t \ \mathbf{z}_t')'$, then
$$\mathbf{x}_t = \boldsymbol{\mu} + \boldsymbol{\Lambda} \mathbf{F}_t + \boldsymbol{\xi}_t$$
- We can regress $\mathbf{x}_{t+h}$ onto the factors
$$\mathbf{x}_{t+h} = \boldsymbol{\alpha}_h + \boldsymbol{B}_h \mathbf{F}_t + \boldsymbol{e}_{t+h}$$
and compute direct forecasts.

Direct forecasts

- Direct forecast from a static factor model (Stock & Watson, 2002; Bai & Ng, 2006; De Mol, Giannone & Reichlin, 2008).

$$\widehat{\mathbf{x}}^{\text{PC}}_{T+h|T} = \widehat{\boldsymbol{\alpha}}^{\text{OLS}}_h + \widehat{\boldsymbol{B}}^{\text{OLS}}_h \widehat{\mathbf{F}}^{\text{PC}}_T = \bar{\mathbf{x}} + \widehat{\boldsymbol{\Gamma}}^x_{-h} \widehat{\mathbf{V}}^x (\widehat{\mathbf{V}}^{x\prime} \widehat{\boldsymbol{\Gamma}}^x_0 \widehat{\mathbf{V}}^x)^{-1} \widehat{\mathbf{V}}^{x\prime} (\widehat{\mathbf{x}}_T - \bar{\mathbf{x}})$$

using OLS and $\widehat{\mathbf{F}}^{\text{PC}}_t = (\widehat{\mathbf{M}}^x)^{-1/2} \widehat{\mathbf{V}}^{x\prime} (\widehat{\mathbf{x}}_T - \bar{\mathbf{x}})$.

- Direct forecast from a restricted GDFM (Forni, Hallin, Lippi & Reichlin, 2005).

$$\widehat{\mathbf{x}}^{\text{FHLR}}_{T+h|T} = \widehat{\boldsymbol{\alpha}}^{\text{OLS}}_h + \widehat{\boldsymbol{B}}^{\text{OLS}}_h \widehat{\mathbf{F}}^{\text{FHLR}}_T = \bar{\mathbf{x}} + \widehat{\boldsymbol{\Gamma}}^\chi_{-h} \widehat{\mathbf{V}}^\chi (\widehat{\mathbf{V}}^{\chi\prime} \widehat{\boldsymbol{\Gamma}}^\chi_0 \widehat{\mathbf{V}}^\chi)^{-1} \widehat{\mathbf{V}}^{\chi\prime} (\widehat{\mathbf{x}}_T - \bar{\mathbf{x}})$$

using OLS and $\widehat{\mathbf{F}}^{\text{FHLR}}_t = (\widehat{\mathbf{M}}^\chi)^{-1/2} \widehat{\mathbf{V}}^{\chi\prime} (\widehat{\mathbf{x}}_T - \bar{\mathbf{x}})$.

- Comparison:
  - $\widehat{\mathbf{x}}^{\text{PC}}_{T+h|T}$ does not require factors, it is the standard PC regression.
  - $\widehat{\mathbf{x}}^{\text{FHLR}}_{T+h|T}$ exploits the dynamic factor structure.

Recursive forecasts

- Recursive forecast from a dynamic factor model with VAR(1) for the factors
  - Use the EM algorithm

$$\widehat{\mathbf{x}}^{\mathsf{EM}}_{T+h|T} = \bar{\mathbf{x}} + \widehat{\mathbf{\Lambda}}^{\mathsf{EM}}(\widehat{\mathbf{A}}^{\mathsf{EM}})^h \widehat{\mathbf{F}}^{\mathsf{EM}}_T$$

    with $\widehat{\mathbf{F}}^{\mathsf{EM}}_T$ from the Kalman filter which at $t = T$ is also the smoother.
  - Since the Kalman filter can deal with missing data (just predicting and not updating), this is the method to be used for nowcasting.
  - Alternatively use PC and fit VAR on estimated factors

$$\widehat{\mathbf{x}}^{\mathsf{PC}}_{T+h|T} = \bar{\mathbf{x}} + \widehat{\mathbf{\Lambda}}^{\mathsf{PC}}(\widehat{\mathbf{A}}^{\mathsf{PC}})^h \widehat{\mathbf{F}}^{\mathsf{PC}}_T$$

    with $\widehat{\mathbf{A}}^{\mathsf{PC}} = (\sum_{t=2}^T \widehat{\mathbf{F}}^{\mathsf{PC}}_{t-1} \widehat{\mathbf{F}}^{\mathsf{PC}\prime}_{t-1})^{-1} (\sum_{t=2}^T \widehat{\mathbf{F}}^{\mathsf{PC}}_{t-1} \widehat{\mathbf{F}}^{\mathsf{PC}\prime}_{t})$.
- Recursive forecast from an unrestricted GDFM

$$\widehat{\mathbf{x}}^{\mathsf{FHLZ}}_{T+h|T} = \bar{\mathbf{x}} + \sum_{k=0}^{M_T} \widehat{\mathbf{A}}^{k+h} \widehat{\mathbf{R}} \widehat{\mathbf{u}}_{T-k}.$$

The role of idiosyncratic components.

- The optimal one-step ahead forecast of series i is

$$\begin{aligned}
\mathsf{E}[x_{it+1}|\boldsymbol{X}_t] &= \mathsf{E}[\boldsymbol{\lambda}_i^{*'}(L)\mathbf{f}_{t+1} + \xi_{it+1}|\boldsymbol{X}_t] \\
&= \mathsf{E}[\boldsymbol{\lambda}_i^{*'}(L)\mathbf{f}_{t+1}|\boldsymbol{X}_t] + \mathsf{E}[\xi_{it+1}|\boldsymbol{X}_t] \\
&= \underbrace{\mathsf{E}[\boldsymbol{\lambda}_i^{*'}(L)\mathbf{f}_{t+1}|\boldsymbol{F}_t]}_{\chi_{i,T+1|T}} + \underbrace{\mathsf{E}[\xi_{it+1}|\boldsymbol{\Xi}_t]}_{\xi_{i,T+1|T}}
\end{aligned}$$

- Previous forecasting methods are for computing linear estimates of $\chi_{i,T+1|T}$.
- Adding one series to the dataset does not increase complexity for $\chi_{i,T+1|T}$, term which is driven by $\simeq q$ parameters only.
- Adding forecast for the idiosyncratic components might help.
  - exact factor model: add univariate forecasts, e.g., AR;
  - approximate factor model: add multivariate sparse forecasts, e.g., lasso.
- For macroeconomic variables this is seldom the case (Boivin & Ng, 2005; Bai & Ng, 2008; Luciani, 2014).

Factor plus sparse.

- FarmPredict - AR + PC + VAR lasso (Fan, Masini & Medeiros, 2023).

$$(1 - a_i L)x_{it} = c_i + \underbrace{\boldsymbol{\lambda}_i' \mathbf{F}_t}_{\chi_{it}} + \underbrace{\sum_{j=1}^{n} \rho_{ij}\xi_{j,t-1} + u_{it}}_{\xi_{it}}.$$

- Forecast:

$$x_{i,T+1|T} = \bar{x}_i + \widehat{a}_i^{\mathsf{OLS}} x_{iT} + \widehat{\chi}_{i,T+1|T}^{\mathsf{PC}} + \sum_{j=1}^{n} \widehat{\rho}_{ij}^{\mathsf{LASSO}} \widehat{\xi}_{j,T}$$

with $\widehat{\boldsymbol{P}}^{\mathsf{LASSO}} = \{\widehat{\rho}_{ij}^{\mathsf{LASSO}}, i,j = 1, \ldots, n\}$ such that

  - $\widehat{\boldsymbol{P}}^{\mathsf{LASSO}} = \arg\min \sum_{t=1}^{T} \left(\widehat{\boldsymbol{\xi}}_t - \boldsymbol{P}\widehat{\boldsymbol{\xi}}_{t-1}\right)^2 + \gamma \|\boldsymbol{P}\|_1$;
  - $\widehat{\xi}_{it} = \widehat{e}_{it} - \widehat{\chi}_{it}^{\mathsf{PC}}$, $\widehat{e}_{it} = (1 - \widehat{a}_i^{\mathsf{OLS}})x_{it}$, and $\widehat{\chi}_{it}^{\mathsf{PC}}$ obtained by PC from $(\widehat{e}_{1t} \cdots \widehat{e}_{nt})'$.

Factor plus sparse.

- `fnets` - GDFM + VAR lasso (Barigozzi, Cho & Owens, 2023).

$$x_{it} = c_i + \underbrace{\boldsymbol{b}_i'(L)\mathbf{u}_t}_{\chi_{it}} + \underbrace{\sum_{j=1}^{n} a_{ij}\xi_{j,t-1} + \nu_{it}}_{\xi_{it}}.$$

- Forecast:

$$x_{i,T+1|T} = \bar{x}_i + \widehat{\chi}^{\mathsf{FHLR}}_{i,T+1|T} + \sum_{j=1}^{n} \widehat{a}^{\mathsf{LASSO}}_{ij} \widehat{\xi}_{j,T}$$

with $\widehat{\boldsymbol{A}}^{\mathsf{LASSO}} = \{\widehat{a}^{\mathsf{LASSO}}_{ij}, i,j = 1,\ldots,n\}$ such that

- $\widehat{\boldsymbol{A}}^{\mathsf{LASSO}} = \arg\min \mathrm{tr}\left\{\mathbf{A}\widehat{\mathbf{\Gamma}}^{\xi}_0\mathbf{A}' - 2\mathbf{A}\widehat{\mathbf{\Gamma}}^{\xi}_1\right\} + \gamma\|\boldsymbol{A}\|_1;$
- $\widehat{\mathbf{\Gamma}}^{\xi}_k$ from dynamic PC and IFT;
- $\widehat{\xi}_{it} = x_{it} - \widehat{\chi}^{\mathsf{FHLR}}_{it}$, and $\widehat{\chi}^{\mathsf{FHLR}}_{it}$ obtained by dynamic + static PC.

Comparison `FarmPredict` vs. `fnets`

High-low range measures of US financial companies - $n = 46$.

Rolling window out-of-sample 2012 using as sample the $T = 252$ previous days.

		fnets	AR	FarmPredict
FE^{avg}	Mean	**0.7258**	0.7572	0.7616
	Median	**0.6029**	0.6511	0.6243
FE^{max}	Mean	**0.8433**	0.879	0.8745
	Median	**0.7925**	0.8437	0.8259

$$\text{FE}^{\text{avg}}_{T+1} = \frac{\sum_i (x_{i,T+1} - \widehat{x}_{i,T+1|T})^2}{\sum_i x^2_{i,T+1}} \quad \text{and} \quad \text{FE}^{\text{max}}_{T+1} = \frac{\max_i |x_{i,T+1} - \widehat{x}_{i,T+1|T}|}{\max_i |x_{i,T+1}|}.$$

Coincident indicators

Eurocoin (Altissimo, Cristadoro, Forni, Lippi & Veronese, 2010)

Core inflation (Cristadoro, Forni, Reichlin & Veronese, 2005)

- $\mathbf{x}_t$ are monthly stationary predictors such that

$$\mathbf{x}_t = \boldsymbol{\mu} + \boldsymbol{\Lambda}\mathbf{F}_t^M + \boldsymbol{\xi}_t.$$

- Y_t is log of monthly GDP or Inflation in month t such that

$$y_t^Q = Y_t - Y_{t-3} = \mu_y + \boldsymbol{\lambda}_y'\mathbf{F}_t^Q + \xi_{y,t}$$

- Notice that Y_t is observed only at lower frequency (quarterly).
- If we assume the approximation for levels $Y_t^Q = \sum_{k=0}^{2} Y_{t-k}$ then

$$\begin{aligned} y_t^Q = Y_t^Q - Y_{t-3}^Q &= (Y_t + Y_{t-1} + Y_{t-2}) - (Y_{t-3} + Y_{t-4} + Y_{t-5}) \\ &= y_t^M + 2y_{t-1}^M + 3y_{t-2}^M + 2y_{t-3}^M + y_{t-4}^M \\ &= (1 + L + L^2)^2 y_t^M \end{aligned}$$

- The monthly and quarterly factors are such that (Mariano & Murasawa, 2003)

$$\mathbf{F}_t^Q = \mathbf{F}_t^M + 2\mathbf{F}_{t-1}^M + 3\mathbf{F}_{t-2}^M + 2\mathbf{F}_{t-3}^M + \mathbf{F}_{t-4}^M = (1 + L + L^2)^2\mathbf{F}_t^M$$

Coincident indicators

Eurocoin (Altissimo, Cristadoro, Forni, Lippi & Veronese, 2010)

Core inflation (Cristadoro, Forni, Reichlin & Veronese, 2005)

- Consider a smoothed version of y_t^Q at yearly frequency

$$c_t = (1 + 2L + 3L^2 + 4L^3 + 3L^4 + 2L^5 + L^6)^2 y_t^Q$$

- A long-run indicator is given by the projection of c_t onto estimated $\mathbf{F}_t^Q$

$$\widehat{e}_t^{\mathsf{FHLR}} = \mu_y + (c_t - \bar{c})\widehat{\mathbf{F}}_t^{Q,\mathsf{FHLR}'} \left(\sum_{t=1}^{T} \widehat{\mathbf{F}}_t^{Q,\mathsf{FHLR}} \widehat{\mathbf{F}}_t^{Q,\mathsf{FHLR}'} \right)^{-1} \widehat{\mathbf{F}}_t^{Q,\mathsf{FHLR}}$$

or

$$\widehat{e}_t^{\mathsf{PC}} = \mu_y + (c_t - \bar{c})\widehat{\mathbf{F}}_t^{Q,\mathsf{PC}'} \left(\sum_{t=1}^{T} \widehat{\mathbf{F}}_t^{Q,\mathsf{PC}} \widehat{\mathbf{F}}_t^{Q,\mathsf{PC}'} \right)^{-1} \widehat{\mathbf{F}}_t^{Q,\mathsf{PC}}$$

EA GDP growth rate

EA HICP inflation

$\widehat{e}_t^{\textsf{FHLR}}$ (red), $\widehat{e}_t^{\textsf{PC}}$ (yellow)

Impulse response functions (Forni, Giannone, Lippi & Reichlin, 2010)

- From the dynamic factor model

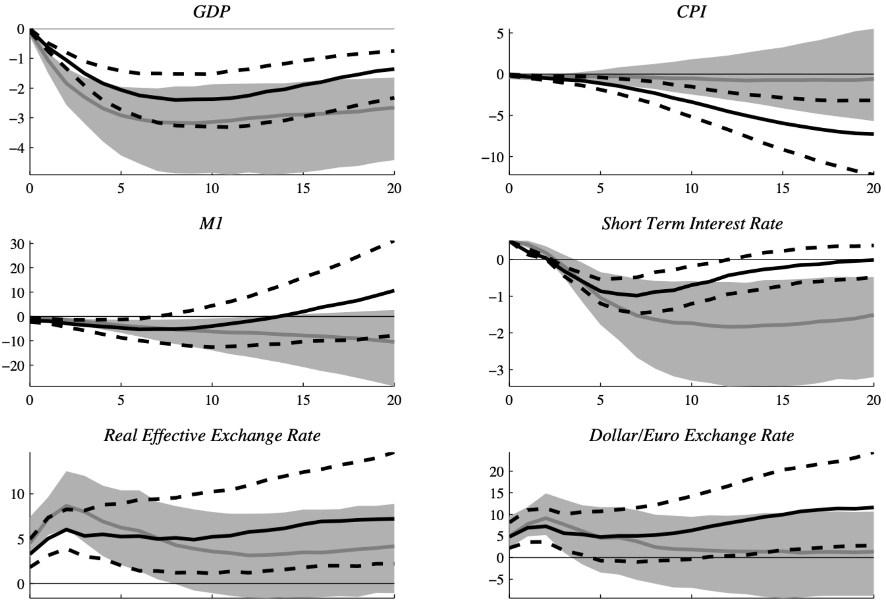
$$x_{it} = \boldsymbol{\lambda}_i'\mathbf{F}_t + \xi_{it}, \quad \mathbf{F}_t = \mathbf{A}\mathbf{F}_{t-1} + \mathbf{H}\mathbf{u}_t$$

- Once estimated via PC + VAR the reduced form IRFs and shocks are

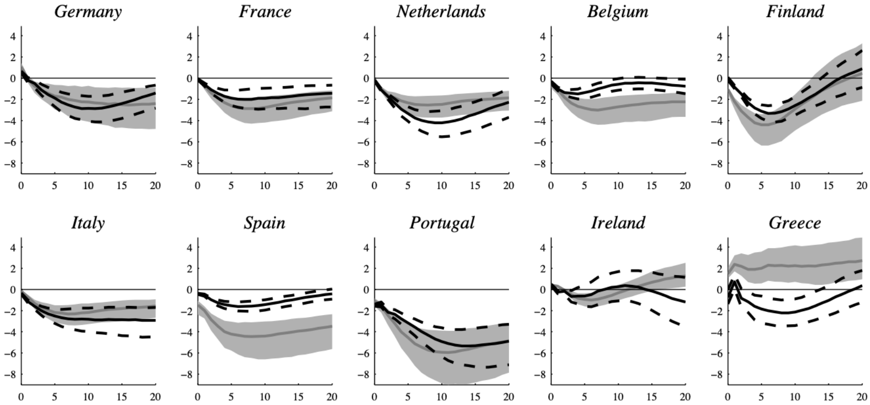
$$\widehat{\mathbf{c}}_i^{\mathsf{PC}'}(L)\widehat{\mathbf{u}}_t^{\mathsf{PC}} = \widehat{\boldsymbol{\lambda}}_i^{\mathsf{PC}'}\sum_{k=0}^{K}(\widehat{\mathbf{A}}^{\mathsf{PC}})^k\widehat{\mathbf{H}}^{\mathsf{PC}}\widehat{\mathbf{u}}_{t-k}^{\mathsf{PC}}$$

- However, we can just prove $|\widehat{\mathbf{u}}_t^{\mathsf{PC}} - \mathbf{R}\mathbf{u}_t| = o_p(1)$, with $\mathbf{R}$ invertible unless further restrictions are imposed:
    - statistical: $T^{-1}\sum_{t=1}^T \mathbf{u}_t\mathbf{u}_t' = \mathbf{I}_q \Rightarrow \mathbf{R}$ is orthogonal;
    - statistical: $T^{-1}\sum_{t=1}^T \mathbf{u}_t\mathbf{u}_t' = \mathbf{I}_q$ plus $\mathbf{H}'\mathbf{H}$ diagonal $\Rightarrow \mathbf{R}$ diagonal ± 1;
    - economic: $T^{-1}\sum_{t=1}^T \mathbf{u}_t\mathbf{u}_t' = \mathbf{I}_q$ plus structure on some $\mathbf{c}_i(L)$ (sign, recursive, long-run) ;
    - economic: identify $\mathbf{u}_t$ via external proxies (IV).

Effects of EA monetary policy - PC + sign restrictions (Barigozzi, Conti & Luciani, 2014).

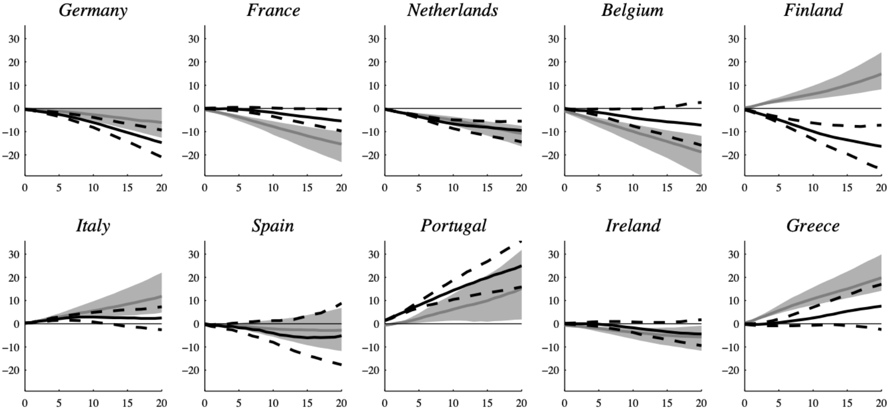


Effects of EA monetary policy - PC + sign restrictions (Barigozzi, Conti & Luciani, 2014).



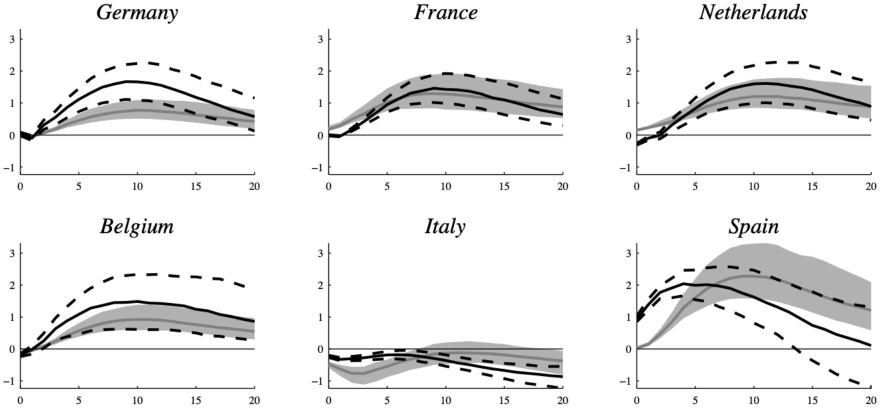
GDP

Effects of EA monetary policy - PC + sign restrictions (Barigozzi, Conti & Luciani, 2014).



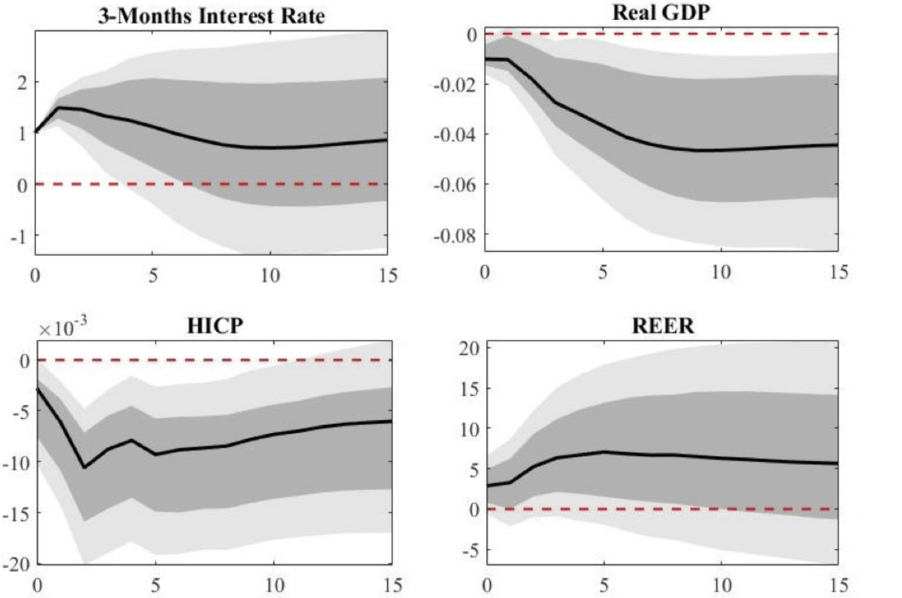
CPI

Effects of EA monetary policy - PC + sign restrictions (Barigozzi, Conti & Luciani, 2014).

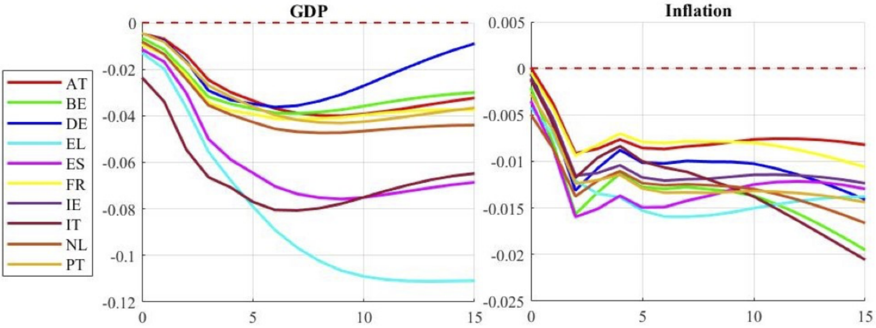


Unemployment rate

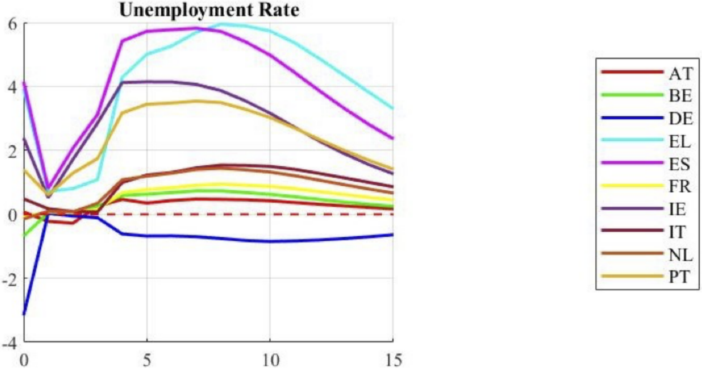
Effects of EA monetary policy - PC + IV identification (Barigozzi, Lissona & Tonni, 2024).



Effects of EA monetary policy - PC + IV identification (Barigozzi, Lissona & Tonni, 2024).



Effects of EA monetary policy - PC + IV identification (Barigozzi, Lissona & Tonni, 2024).



Lon-run impulse response functions (Barigozzi, Lippi & Luciani, 2021)

- To estimate the long-run effects we must account for unit roots and cointegration.
- We need a dynamic factor model for $I(1)$ data.
- The factors are $I(1)$ but cointegrated, so their dynamics is either a VECM or a VAR in levels.
- The idiosyncratic components are $I(1)$.
- There are deterministic trends.

Lon-run impulse response functions (Barigozzi, Lippi & Luciani, 2021)

- The model is

$$y_{it} = a_i + b_i t + \boldsymbol{\lambda}_i' \mathbf{F}_t + \xi_{it}$$
$$\mathbf{F}_t = \mathbf{A}\mathbf{F}_{t-1} + \mathbf{H}\mathbf{u}_t, \qquad \xi_{it} = \rho_i \xi_{i,t-1} + e_{it}.$$

  where $b_i \neq 0$ for $n_b = o(n)$ series and $\rho_{it} = 1$ for $n_I = o(n)$ series or $\rho_{it} = 0$ otherwise.

- Estimation:
  1. De-trend via OLS $\widehat{x}_{it} = y_{it} - \widehat{a}_i^{OLS} - \widehat{b}_i^{OLS} t$;
  2. Loadings by PC on $\Delta \widehat{x}_{it} \Rightarrow \widehat{\boldsymbol{\Lambda}}^{\mathsf{PC}}$;
  3. Factors $\widehat{\mathbf{F}}_t^{\mathsf{PC}} = (\widehat{\boldsymbol{\Lambda}}^{\mathsf{PC}\prime} \widehat{\boldsymbol{\Lambda}}^{\mathsf{PC}})^{-1} \widehat{\boldsymbol{\Lambda}}^{\mathsf{PC}\prime} \widehat{\mathbf{x}}_t$;
  4. VAR (or VECM) by OLS on $\widehat{\mathbf{F}}_t^{\mathsf{PC}} \Rightarrow \widehat{\mathbf{A}}^{\mathsf{PC}}$ and $\widehat{\mathbf{H}}^{\mathsf{PC}}$.

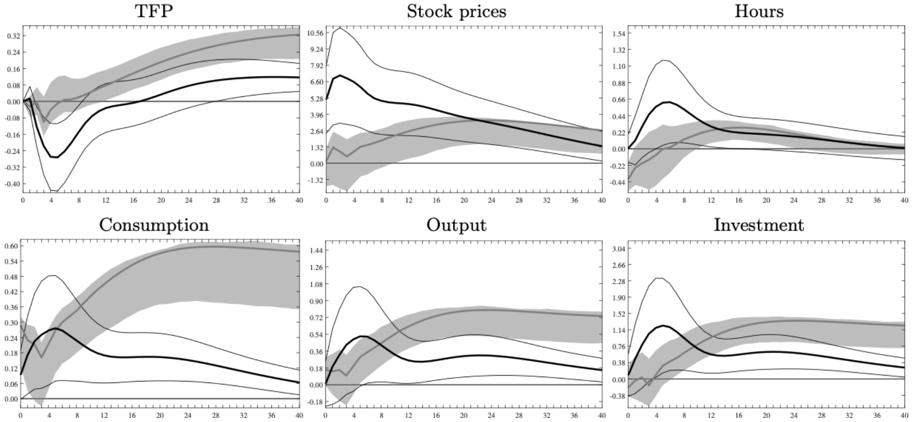
- The reduced form IRFs and shocks are

$$\widehat{\mathbf{c}}_i^{\mathsf{PC}\prime}(L) \widehat{\mathbf{u}}_t^{\mathsf{PC}} = \widehat{\boldsymbol{\lambda}}_i^{\mathsf{PC}\prime} \sum_{k=0}^{K} \sum_{h=0}^{k} (\widehat{\mathbf{A}}^{\mathsf{PC}})^h \widehat{\mathbf{H}}^{\mathsf{PC}} \widehat{\mathbf{u}}_{t-h}^{\mathsf{PC}}.$$

- This estimator is consistent as $n, T \to \infty$. The rate depends on n_b and n_I.
- If $n_b = n_I = 0$ the consistency rate is $\min(\sqrt{n}, \sqrt{T})$.

Effects of news shocks - Stationary vs $I(1)$ factor model

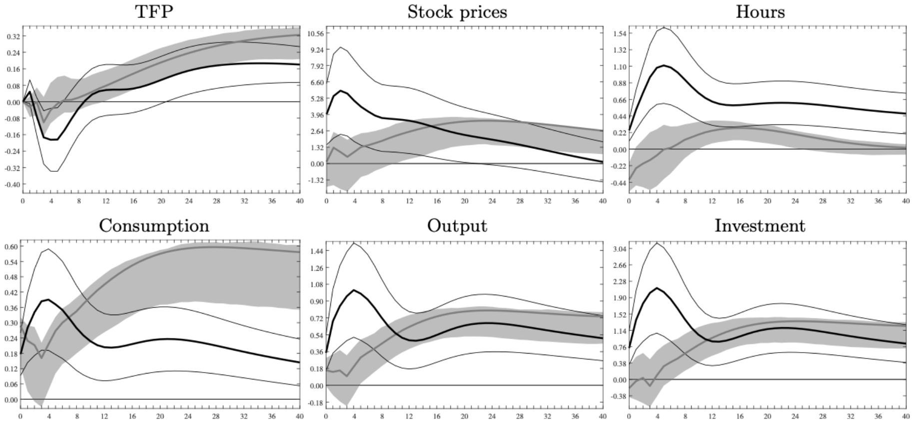
(Forni, Gambetti & Sala, 2014; Barigozzi, Lippi & Luciani, 2021).



VAR in levels

Effects of news shocks - Stationary vs $I(1)$ factor model

(Forni, Gambetti & Sala, 2014; Barigozzi, Lippi & Luciani, 2021).



VECM

Coincident indicators - Output gap (Barigozzi & Luciani, 2023; Barigozzi &Lissona, 2024).

- Identification can be made on the factors instead of the impulse responses.
- Given an $I(1)$ dynamic factor model, we can identify a common trend is identified from

$$\mathbf{F}_t = \boldsymbol{\Psi}\tau_t + \boldsymbol{\omega}_t, \qquad \tau_t = \tau_{t-1} + \nu_t.$$

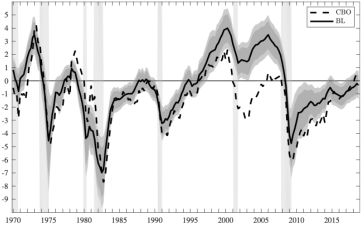
- For GDP we have

$$y_{it} = a_i + b_i t + \boldsymbol{\lambda}_i' \mathbf{F}_t + \xi_{it} = \underbrace{a_i + b_i t + \boldsymbol{\lambda}_i' \boldsymbol{\Psi}\tau_t}_{\text{Potential output}} + \underbrace{\boldsymbol{\lambda}_i' \boldsymbol{\omega}_t}_{\text{Output gap}} + \xi_{it}$$

- We can estimate the model using the EM algorithm twice.

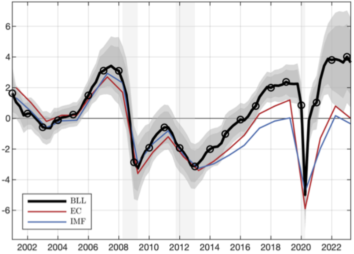
Output gap

(Barigozzi & Luciani, 2023.)



US

(Barigozzi &Lissona, 2024).



EA

Other applications and extensions

- Breaks (Breitung & Eickmeier, 2011; Cheng, Liao & Schorfeide, 2016; Corradi & Swanson, 2014; Barigozzi, Cho & Fryzlewicz, 2018; Barigozzi & Trapani, 2021; Bai, Duan & Han, 2021, 2022; Barigozzi, Cho & Trapani, 20xx).
- Volatility (Barigozzi & Hallin, 2016, 2017, 2020).
- Networks (Barigozzi & Hallin, 2017; Barigozzi, Cho & Owens, 2023).
- Local stationarity (Motta, Hafner & von Sachs, 2011; Barigozzi, Hallin, Soccorsi & von Sachs, 2021).
- Random fields (Barigozzi, La Vecchia & Liu, 2023).
- Matrix time series (Yu, He, Kong & Zhang, 2022; He, Kong, Trapani & Yu, 2023; Barigozzi & Trapin, 20xx).
- Tensor time series (Barigozzi, He, Li & Trapani, 2023).
- Tail robust estimators (Barigozzi, He, Li & Trapani, 2023; Barigozzi, Cho & Maeng, 20xx).

Thank you!