FACTORS AND HIGH-DIMENSIONAL TIME SERIES: THE *Dynamic*, THE *Static*, AND THE *Weak*

Marc Hallin
ECARES and Department of Mathematics
Université libre de Bruxelles

based on joint work with Matteo Barigozzi (University of Bologna, Italy)

BIRS workshop
"Analysis of Complex Data: Tensors, Networks, and Dynamic Systems"
Banff, May 13–17, 2024

- After many interesting talks about high-dimensional matrix- and tensor-valued time series ...

- ... for the closing talk, let us go back to simpler traditional real-valued time series!

- Everything in this talk remains valid in the matrix- and tensor-valued context, though.

- Work in progress (with Matteo Barigozzi).

The family tree of Factor Models is rooted in early-twentieth-century psychometrics: the concept of (unobserved) factor first appears a hundred and twenty years ago in

Spearman, C. (1904). General intelligence, objectively determined and measured, *The American Journal of Psychology* , Vol. 15, pp. 201–292.

But the story of Factor Models in High-Dimensional Time Series and Time-Series Econometrics takes off with

Chamberlain, G. (1983). Funds, factors and diversification in arbitrage pricing models, *Econometrica*, Vol. 51, pp. 1305–1323

and

Chamberlain, G. and Rothschild, M. (1983). Arbitrage, factor structure, and mean-variance analysis on large asset markets, *Econometrica*, Vol. 51, pp. 1281–1323

and their approximate static factor model.

2. *The* **Static***: Chamberlain and Rothschild*

The model proposed by Chamberlain and Rothschild is

- an *approximate* factor model (the n-dimensional idiosyncratic component is not required to be an i.i.d. process of mutually orthogonal zero-mean variables with finite diagonal covariance matrix)
- with *static* loadings (the unobserved factors are loaded contemporeanously via a matrix of loadings).

Their factor model decomposition is

$$X_{it} = \chi_{it}^{\text{stat}} + \xi_{it}^{\text{stat}} := \mathbf{B}_i \mathbf{f}_t + \xi_{it}^{\text{stat}} \quad i = 1, \ldots, n, \; t = 1, \ldots T$$

with a crucial novel (wrt to previous literature) feature: high-dimensional asymptotics are considered, with both n and T going to infinity.

Assumptions are quite mild:

(a) $X_{11}, \dots X_{nT}$ is the observed finite-(n, T) realization of a second-order stationary process $\mathbf{X} = \{X_{it} | i \in \mathbb{N},\ t \in \mathbb{Z}\}$ with (for ease of exposition and without loss of generality) zero-mean, strictly positive variances, and finite second-order moments;

(b) $\boldsymbol{\chi}_t^{\text{stat}} = (\chi_{1t}^{\text{stat}}, \dots, \chi_{nt}^{\text{stat}})'$ (the statically common component) is the value at time t of the unobserved finite-(n, T) realization of a second-order stationary process $\{\chi_{it}^{\text{stat}}\}$ with $n \times n$ covariance matrix $\boldsymbol{\Sigma}_{\boldsymbol{\chi}}^{(n)}$; as $n \to \infty$, the r nonzero eigenvalues of $\boldsymbol{\Sigma}_{\boldsymbol{\chi}}^{(n)}$ tend to infinity;

(c) $\mathbf{B}_i$ is a $1 \times r$ row vector of scalar *loadings*;

(d) $\mathbf{f}_t = (f_{1t}, \dots, f_{rt})'$ is a second-order stationary latent r-dimensional process of *factors* with $\mathrm{E}[\mathbf{f}_t] = \mathbf{0}$, $\mathrm{E}[\mathbf{f}_t \mathbf{f}_t'] = \mathbf{I}_{r \times r}$, and $\mathrm{E}[f_{kt}\xi_{jt}] = 0$ for $1 \leq k \leq r$, $1 \leq j \leq n$, and $t \in \mathbb{Z}$;

(e) $\boldsymbol{\xi}_t^{\text{stat}} = (\xi_{1t}^{\text{stat}}, \dots, \xi_{nt}^{\text{stat}})'$ (the *statically idiosyncratic component*) is the value at time t of the unobserved finite-(n, T) realization of a second-order stationary process $\{\xi_{it}^{\text{stat}}\}$ with $n \times n$ covariance matrix $\boldsymbol{\Sigma}_{\boldsymbol{\xi}}^{(n)}$; as $n \to \infty$, the eigenvalues of $\boldsymbol{\Sigma}_{\boldsymbol{\xi}}^{(n)}$ remain bounded.

The $\xi_{it}^{\text{stat}(n)}$'s need not be white noise and their covariance matrix $\boldsymbol{\Sigma}_{\boldsymbol{\xi}}^{(n)}$ needs not be diagonal, whence the terminology "approximate" factor model.

• This is a semiparametric statistical model, with parameters $\mathbf{B}_i$, $i = 1, \ldots, n$ and nuisance the unspecified distributions of the idiosyncratic process $\boldsymbol{\xi}$ and the factors. Most people interpret the decomposition into common and idiosyncratic and the particular form of the common as describing a data-generating process; some (Bai and Li 2012; Onatski 2012; ...) even consider the values of the factors as parameters.

• Chamberlain and Rothschild do not consider the estimation problem for their approximate static factor model. They do not impose any rate on the divergence of the eigenvalues of the covariance matrix $\boldsymbol{\Sigma}_{\boldsymbol{\xi}}^{(n)}$.

• Estimation, with additional assumptions on the rate of divergence of the eigenvalues of covariance matrix $\mathbf{\Sigma}_{\boldsymbol{\xi}}^{(n)}$ and the idiosyncratic cross-correlations, is considered in Stock and Watson (2002), Bai and Ng (2002), Bai (2003), Bai and Li (2012), ... who provide a rigorous treatment of the asymptotic properties of PCA-based estimators for the loadings and the factors of the Chamberlain and Rothschild model, and show, as expected, that if both n and T tend to infinity, consistency (up to orthogonal transformations, as usual) is achieved.

• Typically, once factors are extracted via PCA from an n-dimensional (n large) time series $\mathbf{X}_t$, they are used in a second step to predict a given set of target variables. This approach, in general, offers sizeable improvements over univariate or small-n forecasting models.

• To this day, this approximate static factor model remains the most popular tool in the analysis of high-dimensional time series; it has been used in countless applications.

1.2 Geweke (1977), Sargent and Sims (1977), and dynamic loadings

Five years before Chamberlain and Rothschild, Geweke

Geweke, J. (1977), The dynamic factor analysis of economic time series, in: D.J. Aigner and A.S. Goldberger, Eds., *Latent Variables in Socio-Economic Models,* pp. 365–383, Amsterdam: North Holland

shortly followed by Sargent and Sims

Sargent, T.J. and Sims, C.A. (977). Business cycle modeling without pretending to have too much a priori economic theory, in: C.A. Sims, Ed., *New Methods in Business Cycle Research*, pp. 45–109, Federal Reserve Bank of Minneapolis

had understood that, if factor models were to be used in econometrics, the time-series nature of econometric data could not be ignored.

Their model is

- an *exact* factor model (idiosyncratic components are mutually orthogonal white noise)
- with *dynamic* loadings (the unobserved factors are loaded via filters).

“Dynamic loadings” here means that the unobserved value f_{kt} of a factor at time t may be loaded by the observation with some lag (e.g., at time $t+k$): instead of contemporaneous loadings

$$\boldsymbol{\chi}_t^{\text{stat}} = \mathbf{B}\mathbf{f}_t$$

via a $n \times r$ loading matrix $\mathbf{B}$ (*static loadings*), Geweke considers *dynamic loadings* via *loading filters* $\mathbf{B}(L)$

$$\boldsymbol{\chi}_t^{\text{dyn}} = \mathbf{B}(L)\mathbf{f}_t \quad \text{with} \quad \mathbf{B}(L) = \sum_{\nu=0}^{\infty} \mathbf{B}_\nu L^\nu$$

where $\mathbf{B}(L)$ is an $n \times r$ matrix of one-sided filters with square-summable entries.

This was an extremely innovative idea.

On the other hand, Geweke, Sargent, and Sims do not go all the way with taking into account the time-series nature of the data, as they still assume an *exact factor model*, with i.i.d. and mutually orthogonal idiosyncratic components.

This is a terribly strong assumption, which cannot be expected to hold in econometric data. However, thanks to that assumption, their model is identified (up to an orthogonal transformation of the factors) and traditional fixed-n asymptotics can be considered.

- The scope of their exact dynamic factor model, thus, is not high-dimensional.
- As in Chamberlain and Rothschild, the approach is a (semiparametric) statistical modeling one.

One also should mention two groups of earlier contributions which considered extensions of the exact factor model apt to capture specific aspects of the observed time series:

(I) Engle and Watson (1981), Shumway and Stoffer (1982), Watson and Engle (1983), and Quah and Sargent (1993) adopted a "state-space approach" where a parametric dynamic equation for the factors, e.g., a VAR specification, is added to the static factor model, decomposition;

(II) Peña and Box (1987) and Tiao and Tsay (1989) revisited the exact static factor model (still assuming the idiosyncratic components to be a second-order stationary white noise process).

- Approach (I) was extended to the high-dimensional setting $n \to \infty$ by Doz et al. (2011) who considered the use of the Kalman filter combined with Gaussian maximum likelihood estimation via the Expectation Maximization algorithm; see also Barigozzi and Luciani (2022). This approach is among of the most frequently used in macroeconomic policy analysis; it is employed for now-casting (Giannone et al. (2008)) and for building indicators of economic activity (Barigozzi and Luciani (2022)). See also Poncela et al. (2021) for a survey.

- Approach (II) was extended to the high-dimensional setting $n \to \infty$ by Lam et al., (2011) and Lam and Yao (2012) who still consider principal-component-based estimation but based on a sum of autocovariances—under an assumption of white noise idiosyncratic components which, again, is unlikely to hold in practice.

- All these approaches are semiparametric statistical modeling ones.

2. *The* **Dynamic**: *Forni et al. (2000)*

The General or *Generalized* Factor Model (henceforth GDFM) proposed by Forni et al. (2000) and Forni and Lippi (2001) is combining the dynamic loadings idea of Geweke (1977) and Sargent and Sims (1977) with the high-dimensional asymptotics $(n, T \to \infty)$ of Chamberlain (1983) and Chamberlain and Rothschild (1983).

The following presentation is inspired from the time-domain exposition of Hallin and Lippi (2013), which avoids the spectral-domain approach originally used by Forni and Lippi (2001) (and considered in Gersing et al. (2024) in the definition of the GDFM.

Avoiding the spectral -based definition of the GDFM also takes care of the case of spectral eigenvalues diverging only on a subset of frequencies.

2.1. *The General Dynamic Factor Model*

The approach here is entirely nonparametric. The only fundamental assumption is that the observation, an $n \times T$ panel, is the finite realization, for $1 \leq i \leq n$ and $1 \leq t \leq T$ (n and T large), of a double-indexed second-order time-stationary stochastic process

$$\mathbf{X} := \{X_{it} | i \in \mathbb{N},\ t \in \mathbb{Z}\},$$

with (for convenience) mean zero—that is, a collection of n centered observed time series of length T, related to n individuals or "cross-sectional items" or, equivalently, one single time series in dimension n, observed for $t = 1, \ldots, T$. Instead of being part of the model specification, the factor model decomposition into common and idiosyncratic is "endogenous;" no parameters.

Notation

- $\mathbf{X}_t^{(n)}$ the n-dimensional vector $(X_{1t}^{(n)}, \ldots, X_{nt}^{(n)})'$,
- $\mathbf{X}_t$ the fixed-t collection $\{X_{it} | i \in \mathbb{N}\}$, and by $\mathbf{X}^{(n)}$ the n-dimensional process $\{X_{it} | 1 \leq i \leq n,\ t \in \mathbb{Z}\}$;

Denote by $\mathcal{H}^{\mathbf{X}}$ the Hilbert space spanned by $\mathbf{X}$, equipped with the L_2 covariance scalar product, that is, the set of all L_2-convergent linear combinations of X_{it}'s and limits of L_2-convergent sequences thereof.

Definition 1. A random variable ζ with values in $\mathcal{H}^{\mathbf{X}}$ and variance σ^2_ζ is called *dynamically common* if either

(i) $\sigma^2_\zeta > 0$ and ζ/σ_ζ is the limit in quadratic mean, as $n \to \infty$, of a sequence of standardized elements of $\mathcal{H}^{\mathbf{X}}$ of the form $\dfrac{w^{(n)}_{\mathbf{X}}}{(\mathrm{Var}(w^{(n)}_{\mathbf{X}}))^{1/2}}$, where

$$w^{(n)}_{\mathbf{X}} := \sum_{i=1}^{n} \sum_{k=-\infty}^{\infty} a^{(n)}_{ik} X_{i,t-k} \quad \text{with} \sum_{i=1}^{n} \sum_{k=-\infty}^{\infty} (a^{(n)}_{ik})^2 = 1$$

is such that $\lim_{n\to\infty} \mathrm{Var}(w^{(n)}_{\mathbf{X}}) = \infty$

or

(ii) $\sigma^2_\zeta = 0$ (hence $\zeta = 0$ a.s.)

Note that this definition does not depend on the choice of t

Definition 2. Call *dynamically common space* the Hilbert space $\mathcal{H}^{\mathbf{X}}_{\text{dyn com}}$ spanned by the collection of all dynamically common variables in $\mathcal{H}^{\mathbf{X}}$; call *dynamically idiosyncratic space* its orthogonal complement (with respect to $\mathcal{H}^{\mathbf{X}}$) $\mathcal{H}^{\mathbf{X}}_{\text{dyn idio}} := (\mathcal{H}^{\mathbf{X}}_{\text{dyn com}})^{\perp}$.

Projecting each X_{it} onto $\mathcal{H}^{\mathbf{X}}_{\text{dyn com}}$ and its orthogonal complement $\mathcal{H}^{\mathbf{X}}_{\text{dyn idio}}$ yields the canonical decomposition

$$X_{it} = \chi^{\text{dyn}}_{it} + \xi^{\text{dyn}}_{it}, \quad i \in \mathbb{N}, \ \ t \in \mathbb{Z}$$

of X_{it} into a *dynamically common component* χ^{dyn}_{it} and a *dynamically idiosyncratic* component ξ^{dyn}_{it}, respectively, which are mutually orthogonal at all leads and lags: the *General Dynamic Factor (GDFM) decomosition* of $\mathbf{X}$.

• Note that this decompositions is "endogenous," always exists, and does not impose any restriction (beyond second-order stationarity) on the data-generating process of $\mathbf{X}$. In that sense, it is not a statistical *model* involving *parameters*, but a canonical representation result. Whether it constitutes the description of a data-generating process or not is irrelevant.

• This representation result nature of the GDFM decomposition (as opposed to the statistical model nature, in Chamberlain, Rothschild, Stock, Watson, Bai, ... of the static factor model decomposition) was first emphasized in Forni et al. (2000) and Forni and Lippi (2001) where, however, a frequency domain approach is adopted.

So far, indeed, no assumption has been imposed on the second-order stationary process $\mathbf{X}$. Adding the requirement that, for any $n \in \mathbb{N}$, $\mathbf{X}^{(n)}$ admits a *spectral density matrix* $\theta \mapsto \mathbf{\Sigma}^{(n)}(\theta)$, $\theta \in (-\pi, \pi]$ with eigenvalues

$$\lambda_{X;1}^{(n)}(\theta) \geq \lambda_{X;2}^{(n)}(\theta) \geq \ldots \geq \lambda_{X;n}^{(n)}(\theta)$$

such that

$$\lim_{n\to\infty} \lambda_{X;q}^{(n)}(\theta) = \infty \quad \text{and} \quad \lim_{n\to\infty} \lambda_{X;q+1}^{(n)}(\theta) < \infty, \quad \theta\text{-a.e. in } (-\pi, \pi],$$

for some finite $q \in \mathbb{N}$ independent of n, it can be shown (see Hallin and Lippi (2013)) that

- all $\{\chi_{it}^{\text{dyn}} | t \in \mathbb{Z}\}$'s are driven by a q-dimensional orthonormal white noise process $\{\mathbf{u}_t = (u_{1t}, \ldots, u_{qt})' | t \in \mathbb{Z}\}$ of *common shocks*. The GDFM decomposition, in that case, takes the form

$$X_{it} = \chi_{it}^{\text{dyn}} + \xi_{it}^{\text{dyn}} =: \sum_{i=1}^{q} \mathbf{B}_i(L)\mathbf{u}_t + \xi_{it}^{\text{dyn}} \quad i \in \mathbb{N}, \ \ t \in \mathbb{Z}$$

for some collection $\mathbf{B}_i(L) := (B_{i1}(L), \ldots, B_{iq}(L))$ of linear $1 \times q$ square-summable filters $B_{ij}(L)$, $i \in \mathbb{N}, \ j = 1, \ldots, q$. See Hallin and Lippi (2013) for the relation between the eigenvectors/eigenvalues of the spectral density matrices $\mathbf{\Sigma}^{(n)}(\theta)$ and the loading filters $\mathbf{B}_i(L)$.

- It follows from the above results that the GDFM (loading filters and factors) is asymptotically identified as $n \to \infty$.

- Forni et al. (2000) show that the common and idiosyncratic components χ_{it} and ξ_{it} can be consistently estimated, as $n, T \to \infty$, via dynamic (spectral) PCA, a technique introduced by Brillinger (2001) which, unfortunately, involves two-sided filters, hence performs poorly at the ends of the observation period—making it unsuitable in the context of prediction problems.

- Under the very mild additional assumption of rational spectral densities, Forni, Hallin, Lippi, and Zaffaroni (2017) show that also the loadings and the factors can be consistently estimated, as $n, T \to \infty$, via a two-step approach based on an equivalent autoregressive representation (not described here) derived from the results by Anderson and Deistler (2008) and Forni, Hallin, Lippi, and Zaffaroni (2015) on singular stochastic processes. See also the recent results by Barigozzi, Hallin, Luciani, and Zaffaroni (2023).

- This latter approach only involves one-sided filters, hence allows for constructing GDFM-based forecasts.

- Forni, Giovannelli, Lippi, and Soccorsi (2018) show that such forecasts improve over the Stock and Watson (2002a) ones based on the static factor model even when the assumptions of the static factor model are satisfied.

2.2. *The* **Static** *revisited*

A representation-based characterization of the approximate static factor decomposition of Chamberlain and Rothschild also can be obtained instead of the classical model-based one.

Denoting by $\mathcal{H}^{\mathbf{X}_t}$ the Hilbert space spanned by $\mathbf{X}_t$, consider the following decomposition of $\mathcal{H}^{\mathbf{X}_t}$ into a statically *(at time t)* common subspace $\mathcal{H}^{\mathbf{X}_t}_{\text{stat com}}$ and its orthogonal complement (within $\mathcal{H}^{\mathbf{X}_t}$) $\mathcal{H}^{\mathbf{X}_t}_{\text{stat idio}} := (\mathcal{H}^{\mathbf{X}_t}_{\text{stat com}})^{\perp}$.

Definition 1'. A random variable ζ with values in $\mathcal{H}^{\mathbf{X}_t}$ is called *statically (at time t) common* if

(i) its variance σ^2_ζ is strictly positive, and ζ/σ_ζ is the limit in quadratic mean, as $n \to \infty$, of a sequence of standardized elements of $\mathcal{H}^{\mathbf{X}_t}$ of the form $\dfrac{w^{(n)}_{\mathbf{X}_t}}{(\text{Var}(w^{(n)}_{\mathbf{X}_t}))^{1/2}}$, where

$$w^{(n)}_{\mathbf{X}_t} := \sum_{i=1}^{n} b^{(n)}_i X_{it} \quad \text{with} \ \sum_{i=1}^{n} (b^{(n)}_i)^2 = 1$$

is such that $\lim_{n\to\infty} \text{Var}(w^{(n)}_{\mathbf{X}_t}) = \infty$

or

(ii) $\zeta = 0$ a.s.

Definition 2'. Call *statically (at time t) common space* the Hilbert space $\mathcal{H}^{\mathbf{X}_t}_{\text{stat com}}$ spanned by the collection of all statically (at time t) common variables in $\mathcal{H}^{\mathbf{X}_t}$; call *statically (at time t) idiosyncratic space* its orthogonal complement (with respect to $\mathcal{H}^{\mathbf{X}_t}$) $\mathcal{H}^{\mathbf{X}_t}_{\text{stat idio}} := \left(\mathcal{H}^{\mathbf{X}_t}_{\text{stat com}}\right)^{\perp}$.

We then have the canonical decomposition

$$X_{it} = \chi^{\text{stat}}_{it} + \xi^{\text{stat}}_{it}$$

where χ^{stat}_{it} and ξ^{stat}_{it} are the projections of X_{it} on $\mathcal{H}^{\mathbf{X}_t}_{\text{stat com}}$ and $\mathcal{H}^{\mathbf{X}_t}_{\text{stat idio}}$, respectively.

- Note that these definitions ($\mathcal{H}^{\mathbf{X}_t}$, $\mathcal{H}^{\mathbf{X}_t}_{\text{stat com}}$, $\mathcal{H}^{\mathbf{X}_t}_{\text{stat idio}}$, ...), unlike their GDFM counterparts, depend on t. Orthogonality between the common and the idiosyncratic commponents, thus, only holds at time t, while in the dynamic case it holds for all leads and lags.

Adding the requirement that the eigenvalues

$$\lambda^{(n)}_{X;1} \geq \lambda^{(n)}_{X;2} \geq \ldots \geq \lambda^{(n)}_{X;n}$$

of the $n \times n$ covariance matrix of $\mathbf{X}^{(n)}$ are such that

$$\lim_{n\to\infty} \lambda^{(n)}_{X;r} = \infty \quad \text{and} \quad \lim_{n\to\infty} \lambda^{(n)}_{X;r+1} < \infty, \quad \theta\text{-a.e. in } (-\pi, \pi]$$

for some finite $r \in \mathbb{N}$ independent of n, it can be shown (see Hallin and Lippi (2013)) that $\mathcal{H}^{\mathbf{X}_t}_{\text{stat com}}$ admits r-dimensional orthonormal bases $\mathbf{f}_t = (f_{1t}, \ldots, f_{qt})'$, any of which can be considered as a r-tuple of factors, yielding a Chamberlain and Rothschild approximate static factor decomposition with $\chi^{\text{stat}}_{it} = \mathbf{B}_i \mathbf{f}_t$.

• These factors, in turn, are driven by a q-dimensional orthonormal white noise process $\{\mathbf{u}_t = (u_{1t}, \ldots, u_{qt})' | t \in \mathbb{Z}\}$ of $q \leq r$ *common shocks*.

• If, however, consistent PC-based estimation is to be performed, additional constraints on ξ^{stat}_{it} are necessary (see Stock, Watson, Bai, Bai and Ng, etc.)—in sharp contrast with the GDFM case where no additional assumption is required.

3. *The* **Weak**

The concept of *weak factors* appear in various places in the literature with, however, diverse meanings.

(a) *Statically rate-weak factors* (De Mol et al. (2008); Onatski (2012); ...) in static factor models are related with covariance eigenvalues diverging at sublinear rate (usually, n^α with $\alpha < 1$). Superstrong factors could be defined similarly in case of superlinear divergence rates. Although they have not been considered in the literature so far, *dynamically rate-weak factors* can be defined similarly from the eigenvalues of spectral density matrices.

(b) The same terminology is used by Hallin and Liška (2011) in the context of panels divided into subpanels or blocks where weak factors are common in some block(s) and idiosyncratic in some other(s) (in a dynamic approach). Such weak factors can be rate-weak, -strong, or -superstrong.

(c) Finally, a recent arXiv post ("Weak factors are everywhere") by Gersing, Rust, and Deistler (2023) defines *weakly common components* as the difference (at time t) between the dynamically common and the statically common components of X_{it}, and *weakly common factors*. Any (possibly infinite-dimensional) orthonormal basis of the space they are spanning can be considered a weak factor but this bears no relation to (a) and (b) above.

3.1. Weak factors: De Mol, Giannone, and Reichlin (2008); Onatski (2012)

A (static) factor is weak in the sense of (a)—"rate-weak" or "weakly influential" if the corresponding loadings, as $n \to \infty$, are too small, or too sparse, for the corresponding eigenvalues to diverge at rate n.

As a consequence, such weak factors are not consistently detected by the PC-based identification methods (Bai and Ng 2002; Alessi, Barigozzi, and Capasso 2010, ...) and are not consistently recovered by the classical PC-based estimation methods. These factors, thus, as pointed out by

Onatski, A. (2012). Asymptotics of the principal components estimator of large factor models with weakly influential factors. *Journal of Econometrics* 168, 244–258,

are an unpleasant grain of sand in the gears of static factor model methods. The same concept had been considered, in a different perspective, by

De Mol, Giannone, and Reichlin (2008). Forecasting using a large number of predictors: Is Bayesian shrinkage a valid alternative to principal components?" *Journal of Econometrics* 146, 318–328.

Since then, an abundant literature has considered the case of divergence rates n^{α_k} with $0 < \alpha_k < 1$ for the kth factor: see Bai and Ng (2021), Freyaldenhoven (2022), etc.

3.2. *"Weak factors are everywhere"*: Gersing, Rust, and Deistler (2023)

With an intriguing title *"Weak factors are everywhere"* (arXiv:2307.10067v2), Gersing, Rust, and Deistler (2023) are aiming at *"reconcile"* the Static and the Dynamic by showing that the static factor space $\mathcal{H}^{\mathbf{X}_t}_{\text{stat com}}$, for any t, is a subspace of the dynamic factor space $\mathcal{H}^{\mathbf{X}}_{\text{dyn com}}$.

Accordingly, they propose a new canonical factor decomposition of the form

$$X_{it} = \underbrace{\chi^{\text{stat}}_{it} + \overbrace{\chi^{\text{weak}}_{it}}_{\chi^{\text{dyn}}_{it}} + \xi^{\text{dyn}}_{it}}^{\xi^{\text{stat}}_{it}}$$

where χ^{weak}_{it} is the difference between χ^{dyn}_{it} and χ^{stat}_{it}, which they call the *weakly common component* of $\mathbf{X}$ at time t.

This is a very ingenious idea, which shows how the Dynamic factor model is overarching the Static one.

$$X_{it} = \underbrace{\chi_{it}^{\text{stat}} + \overbrace{\chi_{it}^{\text{weak}}}_{\chi_{it}^{\text{dyn}}} + \xi_{it}^{\text{dyn}}}^{\xi_{it}^{\text{stat}}}$$

- In this new canonical decomposition, χ_{it}^{weak}, being in the dynamically common space $\mathcal{H}_{\text{dyn com}}^{\mathbf{X}}$, is orthogonal, all leads and lags, to the dynamically idiosyncratic component ξ_{it}^{dyn}; being in the statically idiosyncratic space at time t, it is also orthogonal to the statically common component at time t χ_{it}^{stat} (but not its leads and lags).

- Since χ_{it}^{weak} belongs to the dynamically common space $\mathcal{H}_{\text{dyn com}}^{\mathbf{X}}$ but not to the statically common space $\mathcal{H}_{\text{stat com}}^{\mathbf{X}}$, it consists of non-pervasive lagged values of the dynamic factors, or combinations thereof.

- This concept of weakly common component, however, is unrelated to the concept of "statically rate-weak factor" as initially developed in De Mol et al. (2008) and Onatski (2012) (actually, χ_{it}^{weak} cannot be a static factor, neither strong nor weak). Although the Gersing et al. weak common components *"are everywhere,"* they are not not "rate-weak factors" in the sense of (a), nor linear combinations thereof ...

3.3. Weak factors in panels with block structure.

Panels with block structure are $(n \times T)$ panels divided into K $(n_k \times T)$ subpanels with $n = n_1 + \ldots, n_K$: see Hallin and Liška (2011).

There, a (dynamic/static) factor which is rate-strong in a block k_0 can be globally rate-weak if the dimension n_{k_0} of the block is such that $n_{k_0}/n \to 0$.

There is no obvious reason, however, to adopt an asymptotic scenario in which this happens. That asymptotic scenario, indeed, is a mathematical fiction aiming at a good approximation of the actual finite-dimensional observation.

3.3. *Weak factors are nowhere: exchangeability*

While the Gersing-et-al.-weak components χ_{it}^{weak} are taken care of by considering the dynamic rather than the static approach, the problem with (statically) rate-weak factors remains and, in practice, they typically will be swallowed into the idiosyncratic components.

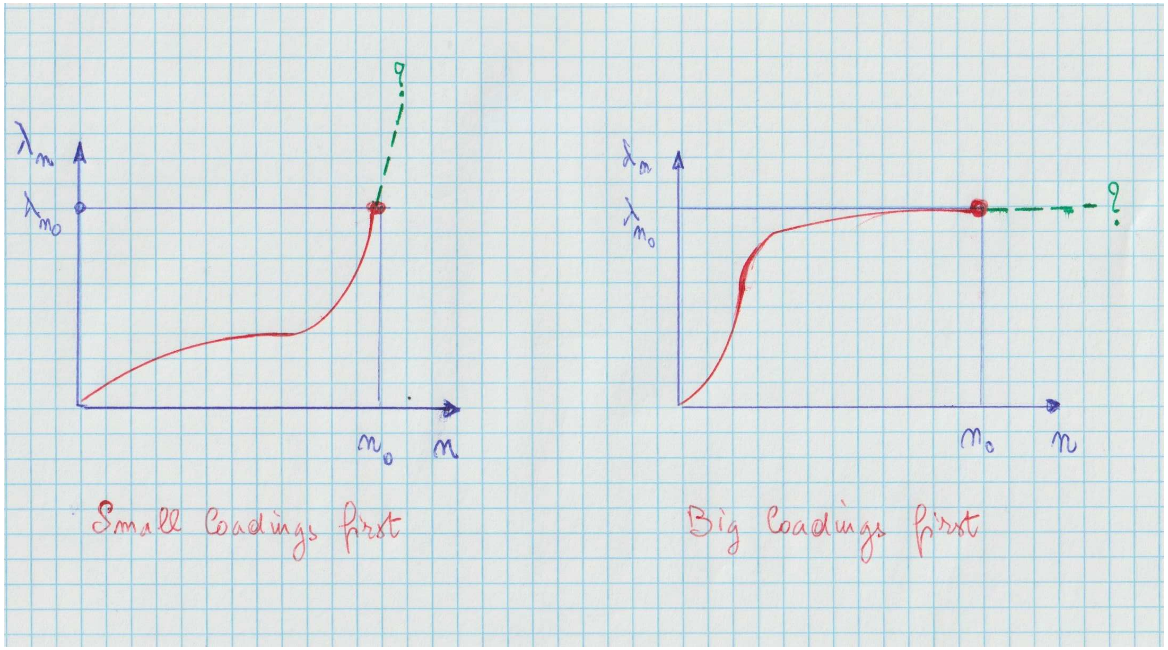
But, **do we really need that concept of rate-weak factors?**

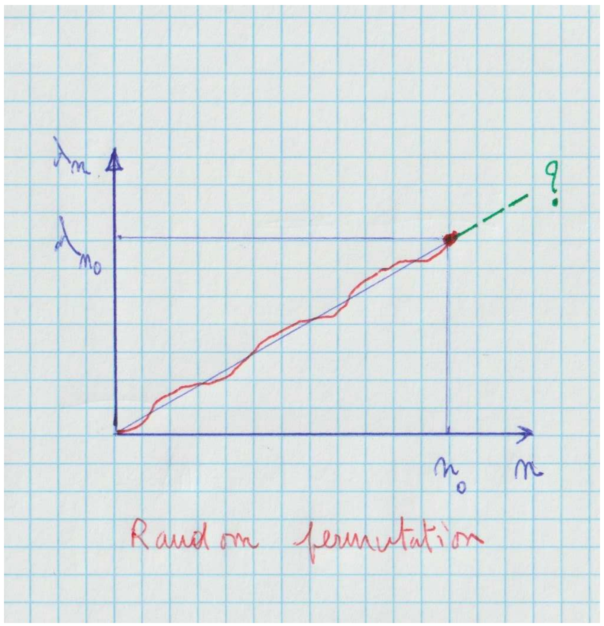
An all too often neglected property of panel data is that the cross-sectional ordering is entirely arbitrary (e.g., alphabetical ordering of cross-sectional items) and should not have any impact on the analysis.

• A panel, actually, is the equivalence class of all its $n!$ cross-sectional permutations.

• Let us show that this rules out the presence of rate-weak (but also rate-superstrong) factors, be they static or dynamic: rate-weak factors are nowhere!

One panel, three cross-sectional orderings





Recall that, under our fully nonparametric approach (no parameters), the only assumption we need in order to obtain a factor model decomposition is that the observed panel is the finite $(n \times T)$ realization of a second-order stationary process of the form $\{X_{it} : i \in \mathbb{N}, t \in \mathbb{Z}\}$.

The mathematical translation of the irrelevance of the cross-sectional ordering is cross-sectional exchangeability.

The stochastic process $\{X_{it} : i \in \mathbb{N}, t \in \mathbb{Z}\}$ is *cross-sectionallly exchangeable* if for any $k \in \mathbb{N}$, any k-tuple $i_1, \ldots, i_k$, and any permutation π of the integers $(1, \ldots, k)$, the k-dimensional stochastic processes

$$\{(X_{i_1 t}, \ldots, X_{i_k t}) : t \in \mathbb{Z}\} \quad \text{and} \quad \{(X_{i_{\pi(1)} t}, \ldots, X_{i_{\pi(k)} t}) : t \in \mathbb{Z}\}$$

are equal in distribution.

That assumption of exchangeability has been used in Barigozzi, Hallin, Luciani, and Zaffaroni (2023) (Inferential theory for generalized dynamic factor models, *Journal of Econometrics*) in order to obtain asymptotic distributional results for the GDFM estimators.

Under this additional assumption of exchangeability, the data-generating process can be seen as a two-step process:

(i) a finite $(n \times T)$ realization of $\{X_{it}\}$, arbitrarily cross-sectionally ordered (following some arbitrary but fixed "alphabetical order"), yielding a sequence of increments $\delta_\nu := \lambda_1^{(\nu)} - \lambda_1^{(\nu-1)}$, $\nu = 1, \ldots, n$ (letting $\lambda_1^{(0)} := 0$) of the first eigenvalue of the nested $\nu \times \nu$ covariance matrices $\mathbf{\Sigma}_{X_1,\ldots,X_\nu}$;

(ii) a random cross-sectional permutation π of the same (π uniform over the $n!$ possible permutations of $(1, \ldots, n)$), with nested $\nu \times \nu$ covariance matrices $\mathbf{\Sigma}_{(X_{\pi^{-1}(1)},\ldots,X_{\pi^{-1}(\nu)})}$, yielding a sequence $\delta_\nu^\pi = \delta_{\pi^{-1}(\nu)}$ of their first eigenvalue increments.

- δ_ν is the contribution of cross-sectional item ν to $\lambda_1^{(n)}$
- the sequence δ_ν^π is a permutation of the sequence δ_ν and

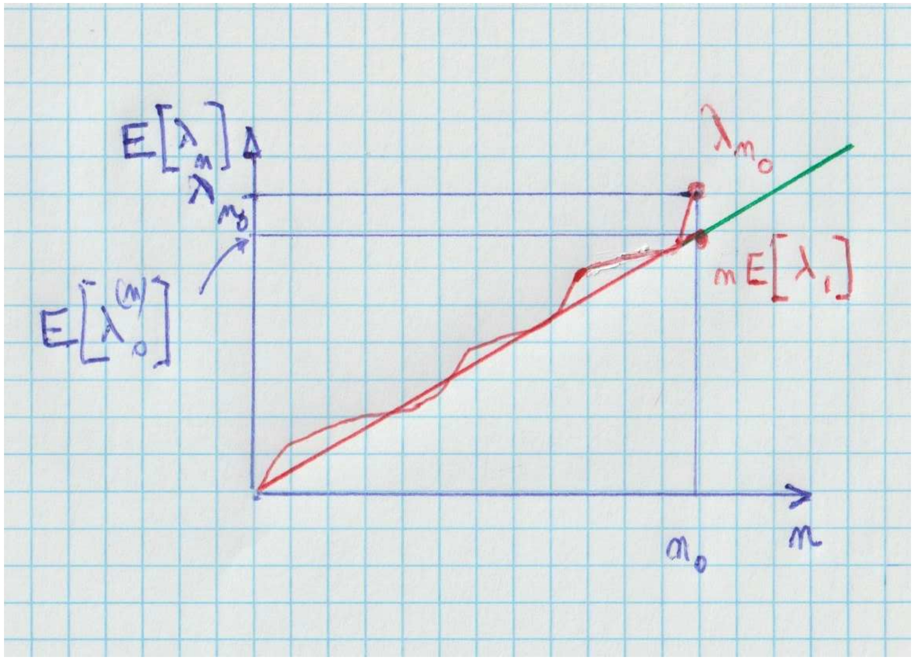
$$\sum_{\nu=1}^{n} \delta_{\pi^{-1}(\nu)} = \sum_{\nu=1}^{n} \delta_\nu = \lambda_1^{(n)}$$

- due to exchangeability, $\mathrm{E}[\delta_\nu] =: \delta$ does not depend on ν; hence,

$$\mathrm{E}[\lambda_1^{(n)}] = n\delta,$$

which precludes non-linear growth of $[\lambda_1^{(n)}$.

- The same resoning holds for $\lambda_1^{(n)}, \lambda_2^{(n)}$, etc., as well as for dynamic eigenvalues.



Under cross-sectional exchangeability, if λ_n is a diverging eigenvalue,

$\mathrm{E}[\lambda_n] = n\mathrm{E}[\lambda_1]$ is linear in n

Note that if λ_n is a bounded eigenvalue, $\mathrm{E}[\lambda_n] = \mathrm{E}[\lambda_1]$ is a constant.

3.4. Undetected strong factors: *"don't throw out the baby with the bathwater!"*

Onatski (2012)'s justification for considering rate-weak factors, however, is more subtle. His claim is that weak-factor asymptotics provide a better approximation in a finite-(n, T) situation where the smallest linearly divergent eigenvalues do not separate well from the bulk of bounded eigenvalues. He does not require, thus, the presence of a "genuinely rate-weak factor," but uses rate-weak asymptotics as a tool for the detection and estimation of these hardly detectable strong factors.

One may wonder whether this is worth the effort. Identifying such hardly detectable factors and incorporating them in the common space, indeed, is of limited importance—provided, however, that factor models are not considered as a dimension-reduction technique.

Whether static or dynamic, undetected strong factors, indeed, are not lost, but wrongly left in the idiosyncratic space.

In a dimension-reduction perspective, the common component χ_{it} is considered a (more tractable, being reduced rank) approximation of the high-dimensional observation X_{it}. The idiosyncratic ξ_{it} then is discarded as if it were an *error term* (a regrettable terminology used by many authors—not to mention those who impose white noise assumptions on it).

But ξ_{it}, typically, is *not* an error term! It need not be small, may be strongly autocorrelated, and may have high predictive value for X_{it}. And its empirical version may contain undetected (strong, but weakly influential) factors.

Rather than a dimension reduction technique, factor models, thus, should be considered a *"divide and conquer"* procedure where the common and the idiosyncratic are analyzed via distinct appropriate methods. The resulting analyses, at the end of the day, are to be brought back together: for instance, once forecasts $\hat{\chi}_{it}$ and $\hat{\xi}_{it}$ of the common and the idiosyncratic have been obtained, their sum $\hat{\chi}_{it} + \hat{\xi}_{it}$ should be used in forecasting X_{it}.

Because cross-correlations in the idiosyncratic component, are mild, little is lost if componentwise or sparse time-series techniques are implemented instead of (untractable) high-dimensional ones.

Now,

• Undetected factors remain undetected because their empirical finite-(n, T) cross-correlations are small.

• the advantage of detecting such factors and incorporating them into the common component is that their, t cross-correlationshere, would be exploited—which they are not in a componentwise or sparse analysis of the idiosyncratic.

• Since these cross-correlations are small, that advantage is small, too.

Provided that the idiosyncratic is not discarded, thus, the problem of undetected factors is not a crucial one.

4. *Conclusions*

(i) The Dynamic approach is nesting the Static one;

(ii) under the (natural) assumption of exchangeability, rate-weak factors, whether static or dynamic, "are nowhere"

(iii) *Gersing-Rust-Deistler -weak factors "are everywhere"*, but are taken into account in the Dynamic approach—which explains the empirical finding that the Dynamic approach outperforms the Static one even when the assumptions of the static model are satisfied;

(iv) factor models are not a dimension-reduction technique and the idiosyncratic component (which is not an error term) should not be discarded, as it may be large and have high predictive power; in particular, it may contain undetected factors; componentwise or sparse analyses are in order;

(v) undetected factors, then, are not a crucial problem.

(vi) These conclusions are likely to extend to matrix- and tensor-valued factor models, and to spatio-temporal ones, where, however, a dynamic approach still needs to be developed.