Returns and Volatilities in High Dimension
a General Factor Model Approach

Marc Hallin
ECARES, Université Libre de Bruxelles

2017 Hermann Otto Hirschfeld Lecture Series

Berlin, 20–21 November 2017

Parallel to the outburst of big data in data sciences and statistics, the analysis of high-dimensional time series in the past few years has become one of the most active subjects of econometrics and financial econometrics.

Parallel to the outburst of big data in data sciences and statistics, the analysis of high-dimensional time series in the past few years has become one of the most active subjects of econometrics and financial econometrics.

The reason is that data in those fields generally take the form of n time series observed over a time period T, where n is quite large—often, latger than T.

"Traditional asymptotics" (fixed n; $T \to \infty$) then are inadequate, or infeasible, due to the usual problems related with the "curse of dimensionality"

Factor models, under their various forms (static/dynamic, observed/unobserved factors, ...) arguably are the most successful methods in that context—possibly, the only succesful ones ...

The situation is even more problematic with volatilities ...

The literature is huge, and develops at a fast pace, basically in all domains of applications: genetics, chemometrics, environmental studies, image analysis, finance, econometrics ...

Much of that literature deals with sampling models: methods are developed for high-dimensional independent and identically distributed observations. At best, those methods are shown to resist some degree of serial dependence.

In econometrics and financial econometrics (but also in environmetrics), serial dependence is ubiquitous, and the time series aspects of the problem should not be ignored—quite on the contrary, they MUST be exploited.

Econometricians in that respect were somewhat ahead of statisticians[a].

Statisticians indeed still restrict much of their analysis to i.i.d. (often Gaussian) data. Econometricians for several decades have been struggling with *time series* in high dimension where, out of the need for applicable methods, they had to come up with feasible solutions. Prototypes are the pioneering contributions by

- Sargent, T.J. and C.A. Sims (1977). Business cycle modelling without pretending to have too much a priori economic theory, in C.A. Sims, Ed., *New Methods in Business Cycle Research*, Federal Reserve Bank of Minneapolis, Minneapolis, pp. 45-109
- Chamberlain, G. (1983). Funds, factors, and diversification in arbitrage pricing models. *Econometrica* 51, 1281-1304
- Chamberlain, G. and M. Rothschild (1983). Arbitrage, factor structure and mean-variance analysis in large asset markets. *Econometrica* 51, 1305-1324.

---

[a] Statisticians mostly resort to the mathematically beautiful theory of large random matrices—spiked models, etc.—which is poorly fit to statistical applications.

Those papers can be considered as early forerunners of the present literature on factor models—a literature that only started some 20 tears later, in the early 2000's, essentially with the contributions by

- Stock and Watson (2002) *Journal of Business and Economic Statistics* 20, 147–162 ("static" factor models),
- Forni, Hallin, Lippi and Reichlin (2000) *The Review of Economics and Statistics* 82, 540–554 (the "general dynamic factor model"), and
- Bai and Ng (2002) *Econometrica* 70, 191–221 ("static" factor models).

Those papers triggered the modern developments of factor models, and hundreds of papers have followed.

Outline of the Lectures

In these four lectures, we will show how factor models allow us to conduct, in a nonparametric and basically model-free way, a combined study of returns and volatilities, based on the onservation of a large panel of observed stock returns.

1. The general dynamic factor method
2. Identifying the number of factors
3. Dynamic factors in the presence of blocks
4. Dynamic factors and volatilities: extracting the market volatility shocks.

Returns and Volatilities in High Dimension
a General Factor Model Approach

1. The General Dynamic Factor Model

Marc Hallin

ECARES, Université Libre de Bruxelles

Hermann Otto Hirschfeld Lecture Series 2017

Berlin, 20–21 November 2017

1. Panel data and the Factor Model Strategy

Observations in econometrics, increasingly often, take the form of an $n \times T$ double-indexed array

$$\{X_{it};\ i = 1, \ldots, n, \quad t = 1, \ldots, T\}$$

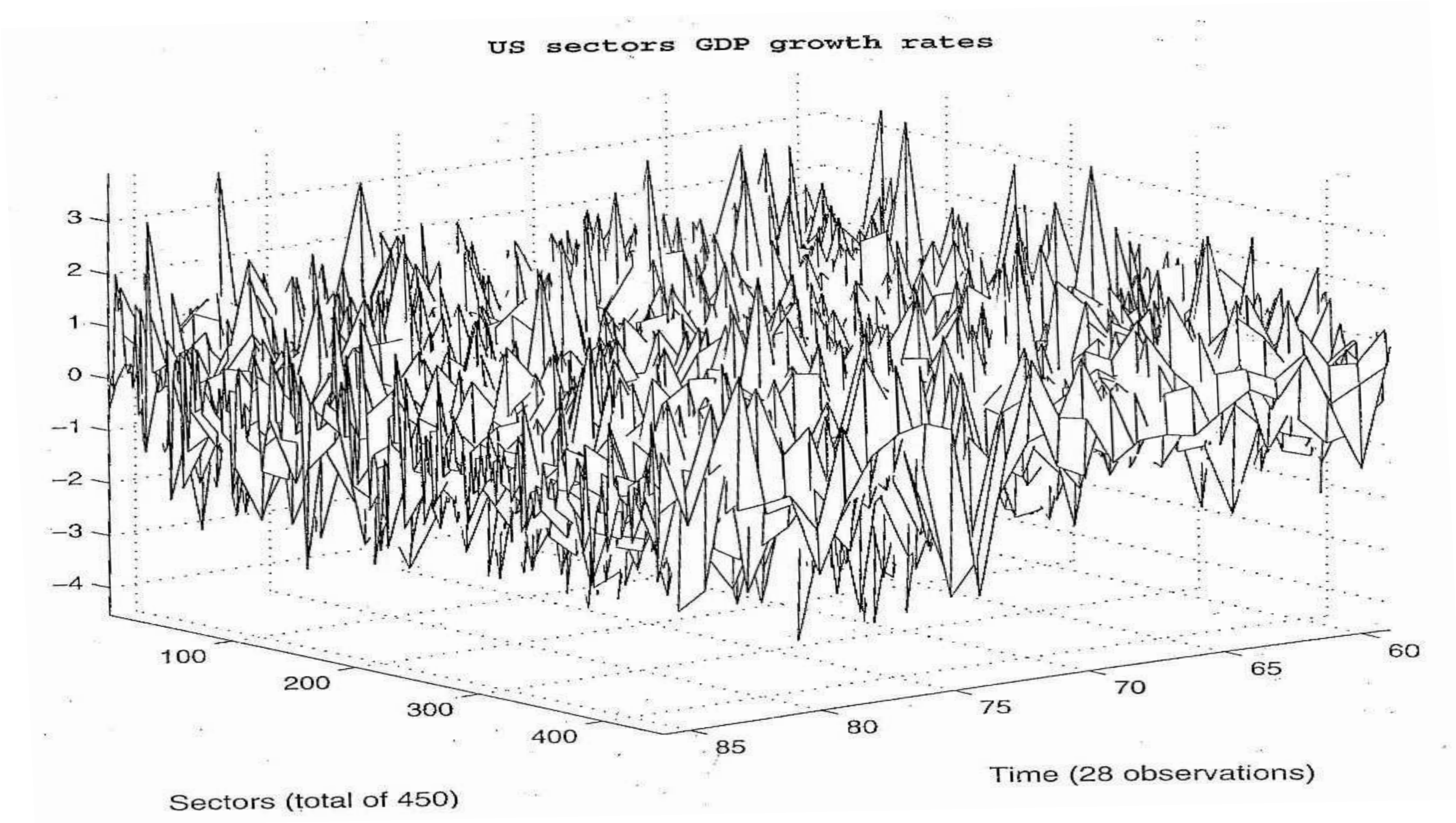
of observed random variables, where i is a cross-sectional index and t stands for time. Such an array is called a *panel*.

Each row in the panel is a univariate time series of length T, while each column can be considered the observation, at time t, of a n-dimensional time series.

The dimension n, in econometric applications, is often of the order of several hundreds or a thousand. All those series exhibit complex (lagged) cross-correlations, a sensible parametrization of which is infeasible (even the simplest VAR(1) model would involve n^2 autoregression parameters, plus $n(n+1)/2$ innovation covariances: for $n = 1000$, this means a number of parameters of the order of 10^6!).

a real-life example

evolution of US sectors GDP growth rates (450 series), over 28 years



Other real-life datasets examples include

- the daily returns of the Standard & Poor 100 stocks over ten years: $n = 90$, $T = 3457$ in M. Barigozzi and M. Hallin (2017) Generalized dynamic factor models and volatilities: estimation and forecasting, *Journal of Econometrics* **201**, 307–321;
- (Eurozone macroeconomics): monthly macroeconomic indicators (industrial production, prices, money aggregates, interest rates (nominal and real), spreads, etc.) over 10 years: $n = 447$, $T = 120$ in M. Forni, M. Hallin, M. Lippi, and L. Reichlin (2001) Coincident and leading indicators for the Euro area, *The Economic Journal* **111**, 62-85;
- (joint analysis of three international stock markets): daily returns, over 15 years, of Standard & Poor 500 for the US, Standard & Poor Europe 350 for Europe, and Nikkei 225 for Japan: $n = 830$ stocks, $T = 4000$ days in M. Barigozzi, M. Hallin and S. Soccorsi (2017) Identification of global and national shocks in international financial markets via general dynamic factor models, Submitted.
- ...

Traditional multivariate time series methods in such datasets are totally helpless

The *factor model* approach in the analysis of such datasets consists in decomposing the observation X_{it} into a sum

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

of two mutually orthogonal (uncorrelated, all leads and lags) unobserved components: the *common component* χ_{it} and the *idiosyncratic component* ξ_{it}.

Various characterizations of *commonality/idiosyncrasy* lead to various *factor models*, *exact* or *approximate*, *static* or *dynamic*, etc.

All those characterizations share one common point, though, which is required for their "technical success":

- χ_{it} has "reduced rank": although n-dimensional, it is driven by a small number q of mutually orthogonal shocks $(\varepsilon_{1t}, \ldots, \varepsilon_{qt})'$: the "common shocks"
- ξ_{it} is only "mildly" cross-correlated

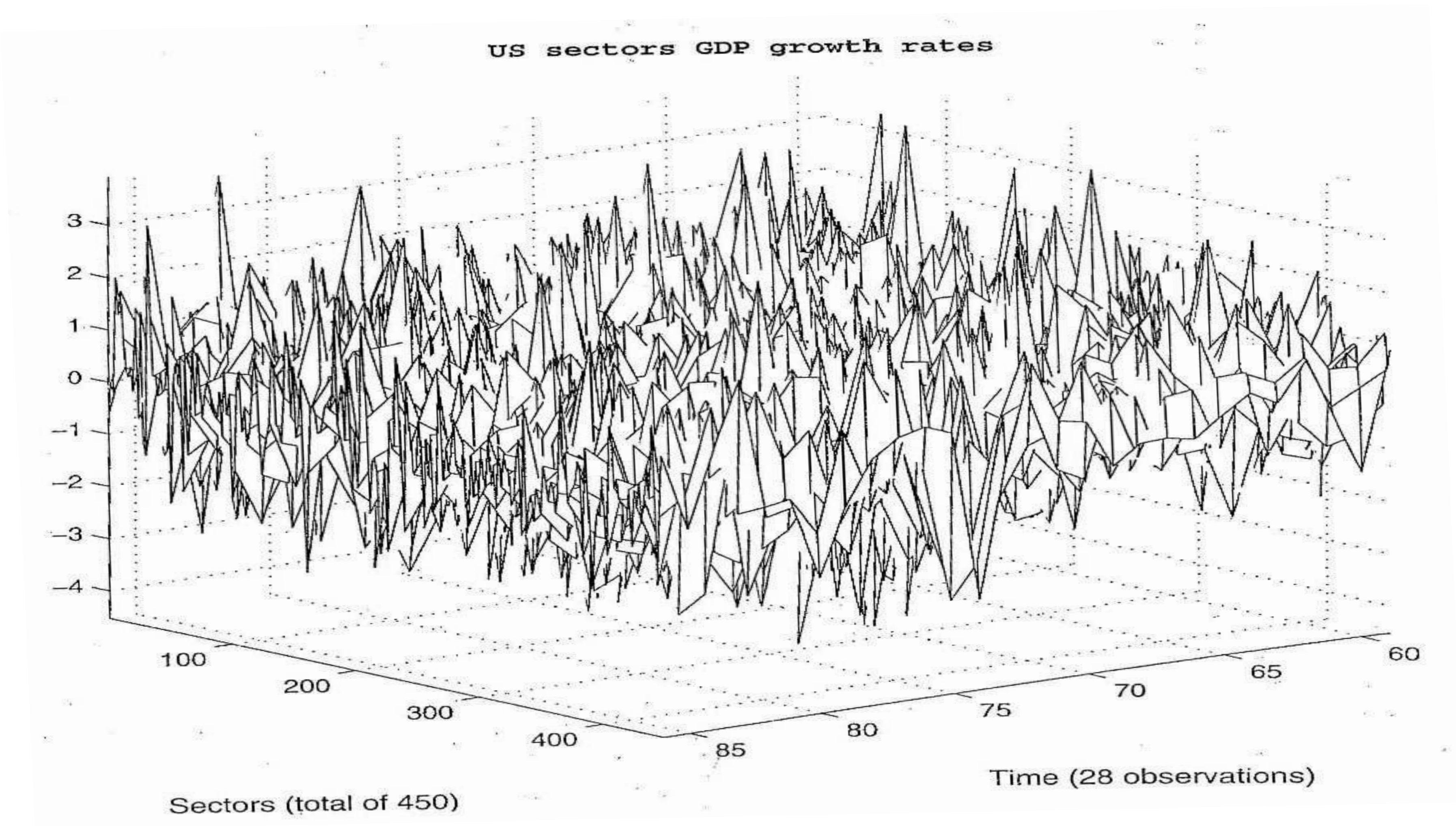
factor models are a "divide and rule" strategy

If we manage to disentangle χ_{it} and ξ_{it}, and to recover the common shocks, $(\varepsilon_{1t}, \ldots, \varepsilon_{qt})'$, then

- we can handle (e.g., forecast) χ_{it};
- as for ξ_{it}, being "mildly" cross-correlated, it can be analyzed (e.g., forecast) componentwise without much loss.

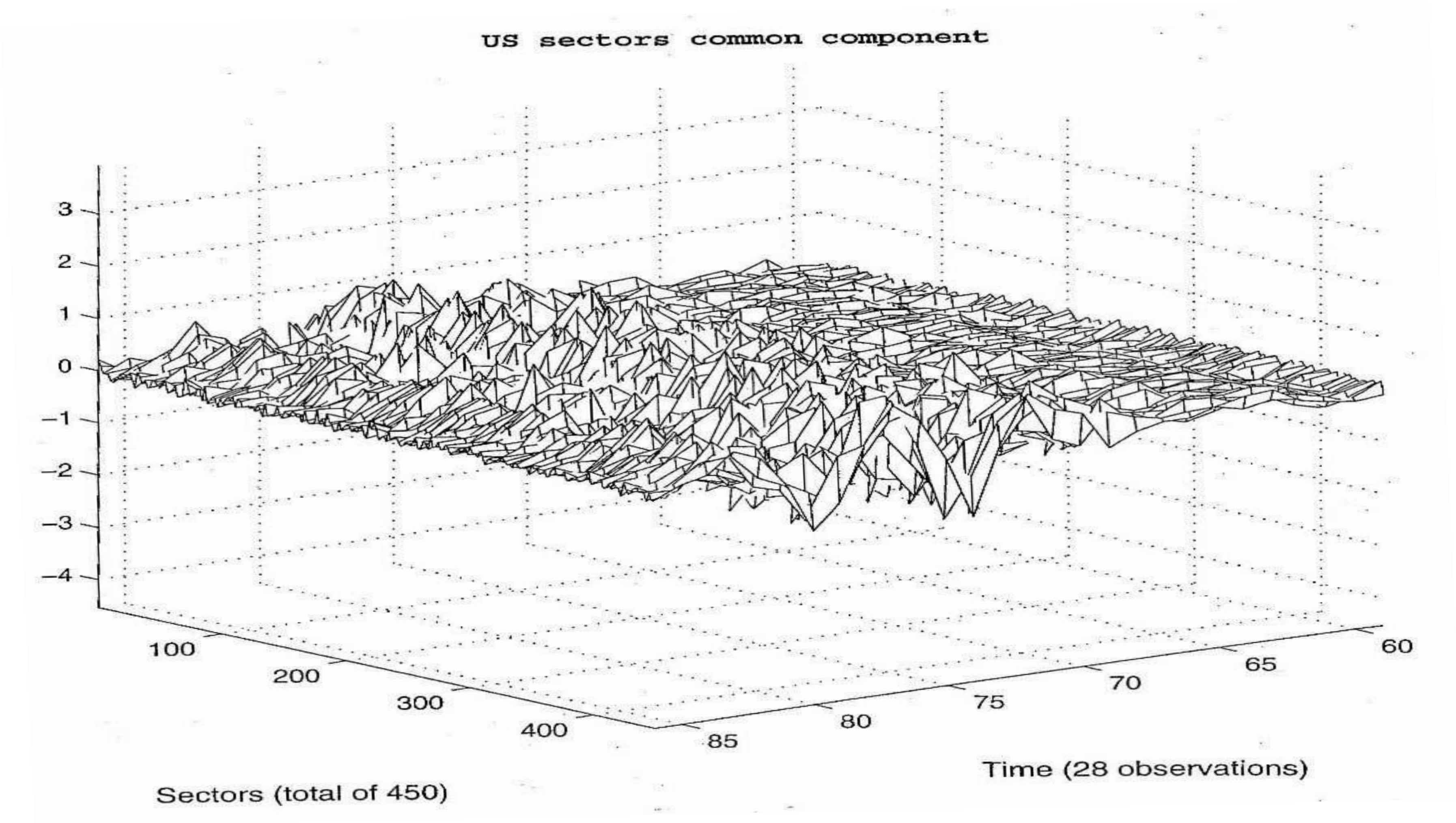
Putting χ_{it} and ξ_{it} together again, we have defeated the curse of dimensionality!

US sectors GDP growth rates (450 series), over 28 years



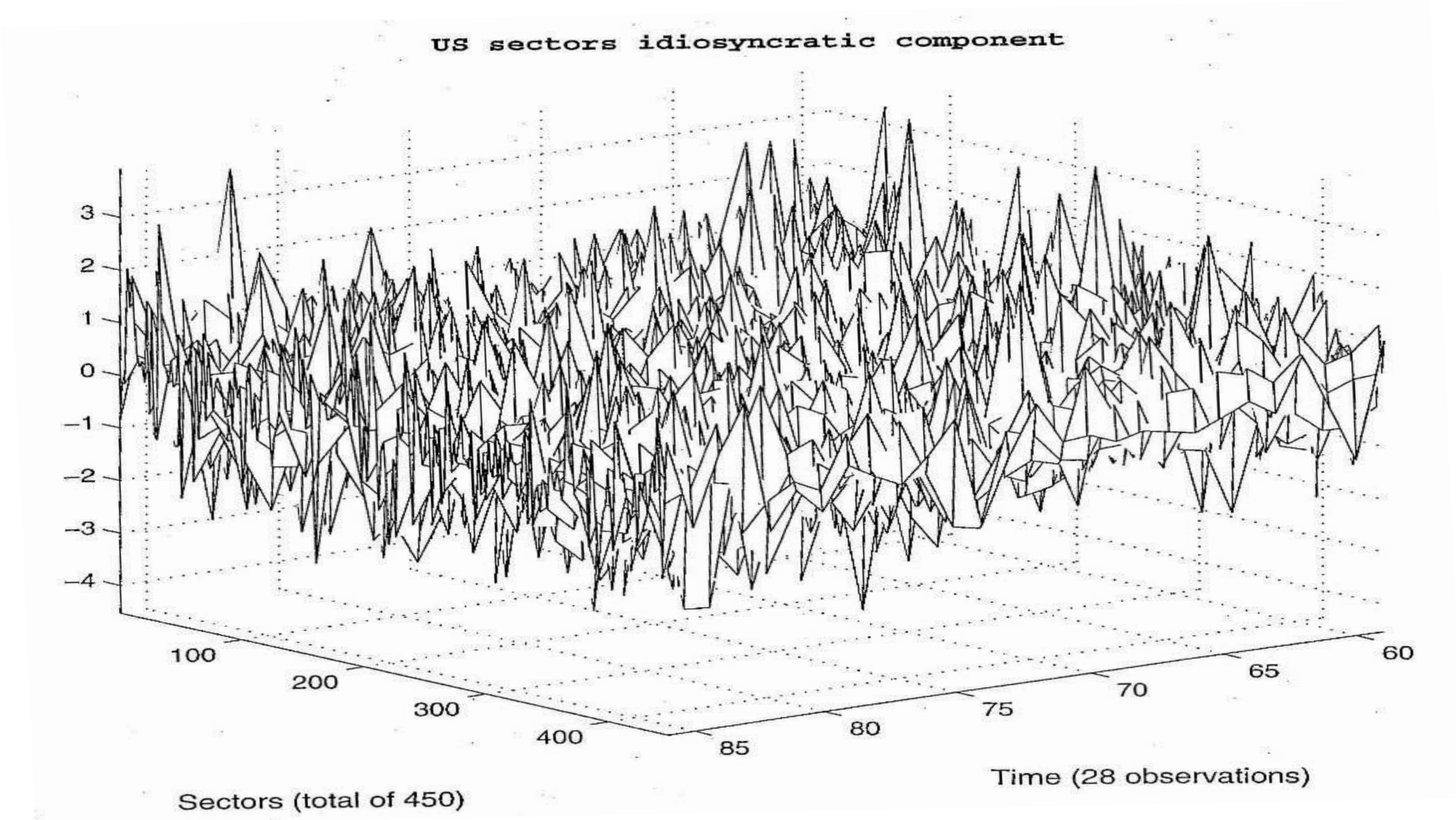
a real-life example

decompose into a common component (with $q = 2$)



a real-life example

and an idiosyncratic component



Since their introduction, factor model methods have been quite successful in the analysis of large panels of econometric data, and have entered daily practice in most national statistical institutes, central banks and business cycle analysis institutions.

Questions to be answered:

- How to define "common" and "idiosyncratic" (technical and interpretational issues)?
- Definition/Existence of a decomposition into "common + idiosyncratic"?
- How to disentangle "common" and "idiosyncratic": estimation, consistency, rates, identification of q, etc.
- How to analyze volatilities (based on an observed panel of returns)?

2. Panel data: common versus idiosyncratic

An $n \times T$ panel

$$\begin{array}{cccc} X_{11}, & X_{12}, & \ldots, & X_{1T} \\ \vdots & \vdots & & \vdots \\ X_{n1}, & X_{n2}, & \ldots, & X_{nT} \end{array}$$

is a finite realization of some double-indexed stochastic process of the form

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

hence, a collection of n observed time series of length T, related to n individuals or "cross-sectional items", or, equivalently, one single time series in dimension n.

The following assumption will be made throughout.

ASSUMPTION A1(i). The process $\mathbf{X}$ is second-order time-stationary, that is, for all i, i', i'', t and k, the variances Var(X_{it}) and covariances Cov$(X_{i't}X_{i'',t-k})$ exist, are finite, and do not depend on t.

For simplicity, we henceforth also assume that all X_{it}'s are centered and, in order to avoid trivialities, nondegenerate:

ASSUMPTION A1(ii). For all $i \in \mathbb{N}$ and $t \in \mathbb{Z}$, $\mathrm{E}[X_{it}] = 0$ and $0 < \mathrm{E}[X_{it}^2]$.

... and, whenever needed,

ASSUMPTION A1(iii). For all $n \in \mathbb{N}$ the n-dimensional process

$$\mathbf{X}^{(n)} := \{X_{it} |\, 1 \leq i \leq n,\ t \in \mathbb{Z}\}$$

admits a spectral density.

Panel data: common versus idiosyncratic

Let Assumption A1 hold, and 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.

Similarly, we use the notation $\mathcal{H}_t^{\mathbf{X}}$, $\mathcal{H}^{\mathbf{X}^{(n)}}$, and $\mathcal{H}_t^{\mathbf{X}^{(n)}}$ for the subspaces of $\mathcal{H}^{\mathbf{X}}$ spanned by $\{X_{is} | \ i \in \mathbb{N}, \ s \leq t\}$, $\{X_{is} | \ 1 \leq i \leq n, \ t \in \mathbb{Z}\}$, and $\{X_{is} | \ 1 \leq i \leq n, \ s \leq t\}$, respectively.

Panel data: common versus Idiosyncratic

A panel is not a “natural” object, but an artificial construction.

Panel data: common versus Idiosyncratic

A panel is not a “natural” object, but an artificial construction.

The n time series constituting the panel have been put together by someone, who did it on purpose—usually, for the reason that those series all carry, or are expected to carry, some information about some unobservable feature or latent process of interest.

Panel data: common versus Idiosyncratic

A panel is not a "natural" object, but an artificial construction.

The n time series constituting the panel have been put together by someone, who did it on purpose—usually, for the reason that those series all carry, or are expected to carry, some information about some unobservable feature or latent process of interest.

Those unobserved common features, in general, are the most relevant issue of the analysis. However, their exact relation to the observed X_{it} prior to the analysis is not known, and "commonness" is the only way to identify it:

Panel data: common versus Idiosyncratic

A panel is not a "natural" object, but an artificial construction.

The n time series constituting the panel have been put together by someone, who did it on purpose—usually, for the reason that those series all carry, or are expected to carry, some information about some unobservable feature or latent process of interest.

Those unobserved common features, in general, are the most relevant issue of the analysis. However, their exact relation to the observed X_{it} prior to the analysis is not known, and "commonness" is the only way to identify it:

Example: the *business cycle*, which is *common* to all variables describing an economy, but remains otherwise undefined;

Example: the *market liquidity*, which is *common* to a market-wide panel of liquidity measurements, but remains otherwise undefined; etc.

Panel data: common versus Idiosyncratic

A panel is not a “natural” object, but an artificial construction.

The n time series constituting the panel have been put together by someone, who did it on purpose—usually, for the reason that those series all carry, or are expected to carry, some information about some unobservable feature or latent process of interest.

Those unobserved common features, in general, are the most relevant issue of the analysis. However, their exact relation to the observed X_{it} prior to the analysis is not known, and “commonness” is the only way to identify it:

Example: the *business cycle*, which is *common* to all variables describing an economy, but remains otherwise undefined;

Example: the *market liquidity*, which is *common* to a market-wide panel of liquidity measurements, but remains otherwise undefined; etc.

Variables orthogonal to all “common variables” will be called “idiosyncratic”.

Moreover, the cross-sectional ordering of the panel, which in principle, is arbitrary, should play no role in the characterization of “commonness” and “idiosyncrasy”. Sensible concepts and sensible statistical procedures should be invariant under permutation of cross-sectional items.

More precise definitions are needed, though.

How should we define "common"? "idiosyncratic"? What should we impose on χ_{it} and ξ_{it} in order for the decomposition

$$X_{it} = \chi_{it} + \xi_{it}, \text{that is, "common"}_{it} + \text{"idiosyncratic"}_{it}, \quad i \in \mathbb{N},\ t \in \mathbb{Z}$$

to make sense, to exist, and to enjoy the properties required for the success of the "factor model strategy"?

A variety of definitions of "idiosyncratic", hence a variety of "factor model" can be found in the literature.

Idiosyncratic?

Idiosyncratic components are expected to be item-specific, or nearly so. One therefore might be tempted to call idiosyncratic those processes $\{\zeta_{it}\}$ in $\mathcal{H}^{\mathbf{X}}$ that do not exhibit *any* cross-correlation at all: ζ_{it} and $\zeta_{i',t-k}$ mutually orthogonal for all $i' \neq i$ and all $k \in \mathbb{Z}$).

Imposing such a condition leads to the so-called *exact* or *strong factor model* considered, for instance, by Sargent and Sims (1997).

Idiosyncratic?

Idiosyncratic components are expected to be item-specific, or nearly so. One therefore might be tempted to call idiosyncratic those processes $\{\zeta_{it}\}$ in $\mathcal{H}^{\mathbf{X}}$ that do not exhibit *any* cross-correlation at all: ζ_{it} and $\zeta_{i',t-k}$ mutually orthogonal for all $i' \neq i$ and all $k \in \mathbb{Z}$).

Imposing such a condition leads to the so-called *exact* or *strong factor model* considered, for instance, by Sargent and Sims (1997).

That requirement, however, is too restrictive for most practical purpose. It is unpleasantly sensitive, in particular, to the possible presence in the panel of two closely related series: if, for instance, X_{2t} is of the form $X_{2t} = a(L)X_{1t}$ for some linear filter $a(L)$, it automatically gets treated as fully "common", although it could be strictly orthogonal to X_{it} for all t and $i > 2$.

The requirement that the idiosyncratic components ξ_{it} be cross-sectionally *strictly* orthogonal to each other at all leads and lags therefore has to be weakened into a milder requirement of "limited cross-correlation", yielding an *approximate* or *weak factor model*.

Idiosyncratic?

In order to introduce a more precise definition of that idea of “mild cross-correlation”, let us consider two examples of extreme idiosyncrasy/extreme commonness.

Idiosyncratic?

In order to introduce a more precise definition of that idea of "mild cross-correlation", let us consider two examples of extreme idiosyncrasy/extreme commonness.

Let

$$X_{it} = \phi_{it} + \psi_{it}, \quad i \in \mathbb{N}, \ t \in \mathbb{Z} \text{ with } \phi_{it} = \phi_t \text{ i.i.d. } \mathcal{N}(0, \sigma^2_\phi), \text{ and } \psi_{it} \text{ i.i.d. } \mathcal{N}(0, \sigma^2_\psi)$$

where ϕ_t and $\psi_{i,t-k}$ are orthogonal for all i, t, and k.

Clearly (Law of Large Numbers),

$$\phi_{it} = \lim_{n\to\infty} n^{-1} \sum_{j=1}^{n} X_{jt} \quad \text{and} \quad \psi_{it} = X_{it} - \lim_{n\to\infty} n^{-1} \sum_{j=1}^{n} X_{jt}$$

where convergence holds in quadratic mean, so that the processes $\{\phi_{it}, \ i \in \mathbb{N}, \ t \in \mathbb{Z}\}$ and $\{\psi_{it}, \ i \in \mathbb{N}, \ t \in \mathbb{Z}\}$ both are in $\mathcal{H}^{\mathbf{X}}$.

Since it has no cross-correlations at all, $\{\psi_{it}\}$ is an example of extreme idiosyncrasy, whereas $\phi_{it} = \phi_t$, which appears in all cross-sectional items, qualifies as an example of extreme "commonness".

Idiosyncratic?

Now, consider arbitrary sequences

$$\sum_{i=1}^{n} \sum_{k=-\infty}^{\infty} a_{ik}^{(n)} \psi_{i,t-k}$$

of *normed* (i.e., with coefficients satisfying $\sum_{i=1}^{n} \sum_{k=-\infty}^{\infty} (a_{ik}^{(n)})^2 = 1$ for all n) linear combinations of the ψ's. Their variances are

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

It follows that the maximal variance, for given n, over all normed linear combinations, of the ψ_{it}'s, remains bounded as $n \to \infty$.

Common?

The situation is entirely different for the linear combinations

$$w^{(n)} := \sum_{i=1}^{n} \sum_{k=-\infty}^{\infty} a_{ik}^{(n)} \phi_{i,t-k} = \sum_{i=1}^{n} \sum_{k=-\infty}^{\infty} a_{ik}^{(n)} \phi_{t-k}$$

involving the ϕ's. We have

$$\sigma^2_{w^{(n)}} = \sum_{k=-\infty}^{\infty} (\sum_{i=1}^{n} a_{ik}^{(n)})^2 \sigma^2_{\phi}.$$

Choosing, for instance, $a_{ik}^{(n)} = a_k^{(n)}$ ($a_{ik}^{(n)}$'s that do not depend on i, so that $n \sum_{k=-\infty}^{\infty} (a_k^{(n)})^2 = 1$), we obtain

$$\sigma^2_{w^{(n)}} = \sum_{k=-\infty}^{\infty} (n a_k^{(n)})^2 \sigma^2_{\phi} = n^2 \sum_{k=-\infty}^{\infty} (a_k^{(n)})^2 \sigma^2_{\phi} = n \sigma^2_{\phi}.$$

It immediately follows that the maximal variance, for given n, over all normed linear combinations, of the ϕ_{it}'s, tends to infinity as $n \to \infty$.

Idiosyncratic?

Another example is

$X_{it} = \phi_{it} + \psi_{it}, \quad i \in \mathbb{N}, \ t \in \mathbb{Z}$ with $\phi_{it} = (1 - \rho_i L)^{-1} \varepsilon_t, \ \varepsilon_t$ i.i.d. $\mathcal{N}(0, 1)$, and ψ_{it} i.i.d. $\mathcal{N}(0, \sigma^2_{\psi})$

where ε_t and $\psi_{i,t-k}$ are orthogonal for all i, t, and k:

the maximal variance, for given n, over all normed linear combinations, of the ϕ_{it}'s, tends to infinity as $n \to \infty$while the same maximum for the ψ_{it}'s remains bounded.

- The family of "exploding" sequences of linear combinations of the X_{it}'s is spanning a sequence of subspaces of $\mathcal{H}^{(n)}$: denote them by $\mathcal{H}^{(n)}_{\text{common}}$;

- The family of "bounded" sequences of linear combinations of the X_{it}'s is spanning a sequence of subspaces of $\mathcal{H}^{(n)}$: denote them by $\mathcal{H}^{(n)}_{\text{idio}}$.

... this construction has a principal components flavor *dynamic principal components*

- The normed linear combination $W_{1;t}^{(n)}$ of present, past and future values of $\mathbf{X}_t^{(n)} := (X_{1t}, \ldots, X_{nt})'$ maximizing the variance is called $\mathbf{X}_t^{(n)}$'s first dynamic principal component, its variance $\lambda_1^{(n)}$ is $\mathbf{X}_t^{(n)}$'s first integrated dynamic eigenvalue
- the normed linear combination $W_{2;t}^{(n)}$ of present, past and future values of $\mathbf{X}_t^{(n)} := (X_{1t}, \ldots, X_{nt})'$ maximizing the variance subject to being orthogonal, at all leads and lags, to $W_{1;t}^{(n)}$, is called $\mathbf{X}_t^{(n)}$'s second dynamic principal component, its variance $\lambda_2^{(n)}$ is $\mathbf{X}_t^{(n)}$'s second integrated dynamic eigenvalue
- ...
- the normed linear combination $W_{n;t}^{(n)}$ of ...

(As we shall see, those dynamic principal components and integrated dynamic eigenvalues are related to the eigenvectors and eigenvalues of $\mathbf{X}^{(n)}$'s spectral density matrix.)

common? Idiosyncratic?

Let q be such that $\lambda_q^{(n)} \to \infty$ and $\lambda_{q+1}^{(n)}$ bounded as $n \to \infty$; assume $q < \infty$. In practice, q is very small compared to n.

Then,

- $\mathcal{H}_{\text{common}}^{(n)}$ is spanned by $W_{1;t}^{(n)}, \dots, W_{q;t}^{(n)}$
- $\mathcal{H}_{\text{idio}}^{(n)}$ is spanned by $W_{q+1;t}^{(n)}, \dots, W_{n;t}^{(n)}$
- $\mathcal{H}_{\text{common}}^{(n)}$ and $\mathcal{H}_{\text{idio}}^{(n)}$ are mutually orthogonal (all leads and lags)

DEFINITION 1. A random variable ζ in $\mathcal{H}^{\mathbf{X}}$, with variance $0 < \sigma_\zeta^2$, is called *common* (relative to the process $\mathbf{X}$) if ζ/σ_ζ is the limit in quadratic mean of a sequence of elements of $\mathcal{H}^{(n)}_{\text{common}}$.

DEFINITION 2. The closed space spanned by the common variables is called the *common space* $\mathcal{H}_{\text{common}}$. Its orthogonal complement (in $\mathcal{H}^{\mathbf{X}}$) is called the *idiosyncratic space* $\mathcal{H}_{\text{idio}}$.

The following definition then can be adopted:

define χ_{it} and ξ_{it} as the projections of X_{it} onto $\mathcal{H}_{\text{common}}$ and $\mathcal{H}_{\text{idio}}$, respectively.

The following assumption makes sense (constructing counterexamples is quite difficult, and they are weird)

ASSUMPTION A2 The common space $\mathcal{H}_{\text{common}}$ has "limited complexity", that is, is driven by a finite (but unspecified) number q of mutually orthogonal white noises—denote them as $\{\mathbf{U}_t\} = \{(U_{1t}, \ldots, U_{qt})' |\ t \in \mathbb{Z}\}$.

The above definitions straightforwardly lead to the following extremely general representation result.

THEOREM 1. *Under the assumptions made, there exist two uniquely defined mutually orthogonal processes* $\boldsymbol{\chi} = \{\chi_{it}\}$ *and* $\boldsymbol{\xi} = \{\xi_{it}\}$ *in* $\mathcal{H}^{\mathbf{X}}_{\text{com}}$ *and* $\mathcal{H}^{\mathbf{X}}_{\text{idio}}$*, respectively, such that*

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

Moreover, there exist a q*-tuple* $\{\mathbf{U}_t\} = \{(U_{1t}, \ldots, U_{qt})'\}$ *of mutually orthogonal white noises (namely,* Var$(U_{jt}) = 1$ *and* Cov$(U_{jt}, U_{j't'}) = 0$ *unless* $j' = j$ *and* $t' = t$*), and a collection of one-sided square-summable filters* $B_{ij}(L)$*,* $i \in \mathbb{N},\ j = 1, \ldots, q$ *such that this decomposition takes the form*

$$X_{it} = \chi_{it} + \xi_{it} \quad \text{with} \quad \chi_{it} = \sum_{j=1}^{q} B_{ij}(L) U_{jt} \qquad i \in \mathbb{N},\ t \in \mathbb{Z}$$

and $\mathcal{H}^{\mathbf{U}}_t = \mathcal{H}^{\boldsymbol{\chi}}_t$ *for all* $t \in \mathbb{Z}$*.*

DEFINITION. The above decomposition is called the *general dynamic factor model representation* of $\mathbf{X}$—in short, a *general dynamic factor model* for $\mathbf{X}$.

• No “factors”, thus, but a common component driven by a finite (and unspecified) number q of white noises: the common shocks of the panel. Those shocks, are loaded, with lags, by X_{it} via the one-sided filters $B_{ij}(L)$, $\in \mathbb{N}$, $j = 1, \ldots, q$.

• Common shocks remain largely undetermined, but the space they are spanning at time t is uniquely characterized as the *innovation space* of the (unobserved, but unique) common component $\boldsymbol{\chi}$.

• The existence and uniqueness of the general dynamic factor model representation of $\mathbf{X}$ do not require (beyond second-order stationarity and a finite q) any constraint on the data-generating process. Therefore, it does not constitute a statistical “model” in the usual sense.

• The properties required by the “factor model strategy” are satisfied: χ_{it} is driven by q-dimensional white noise, and ξ_{it} is only mildly cross-correlated (no pervasive cross-correlations).

• Let $\chi_{it}^{(n)}$ and $\xi_{it}^{(n)}$ denote the projections of X_{it} onto $\mathcal{H}^{(n)}_{\text{common}}$ and $\mathcal{H}^{(n)}_{\text{idio}}$, respectively. Under mild assumptions, it can be shown (Forni, Hallin, Lippi and Reichlin 2000) that $\chi_{it}^{(n)}$ and $\xi_{it}^{(n)}$ consistently recover χ_{it} and ξ_{it} as both n and T tend to infinity,.

The most popular way of turning the decomposition of Theorem 2 into a statistically tractable model, however, is not the general factor model one, and consists in imposing on the common component χ_{it} a simple linear structure, of the form

$$\begin{cases} \chi_{it} & = & b_{i1}F_{1t} + \ldots + b_{ir}F_{rt} \\ \mathbf{F}_t & := & (F_{1t}, \ldots, F_{rt})' = \mathbf{A}_1\mathbf{F}_{t-1} + \ldots + \mathbf{A}_p\mathbf{F}_{t-p} + \mathbf{R}\mathbf{U}_t \quad i \in \mathbb{N},\ t \in \mathbb{Z} \end{cases}$$

where $\mathbf{U}_t := (U_{1t}, \ldots, U_{qt})'$ is a q-tuple of white noises, the $r \times r$ matrices $\mathbf{A}_1, \ldots, \mathbf{A}_p$ define some stationary VAR, and the $r \times q$ matrix $\mathbf{R}$ has rank q.

Expressing $\mathbf{F}_t$ in terms of the $\mathbf{U}_t$'s and placing restrictions on the cross-covariances of idiosyncratics yield a particular form of the General Dynamic Factor where the common space at time t has finite dimension r.

Here, X_{it} depends in a static way (instantaneous dependence: loading constants instead of loading filters) on the unobserved r-dimensional latent process $\{\mathbf{F}_t\}$: the VAR(p) dynamics of the r *factors* $(F_{1t}, \ldots, F_{rt})$, driven by the q-dimensional *common shocks* $\mathbf{U}_t$ ($r \geq q$), are to account for the dynamics of the common components throughout the panel.

The dimension of the space spanned by the χ_{it}'s for given t is $r < \infty$ (since all χ_{it}'s, for fixed t, are linear combinations of $F_{1t}, \ldots, F_{rt}$). This is a very restrictive assumption, as we shall see.

But the dynamic dimension of the doubly indexed process $\{\chi_{it}\}$ is q, since it is driven by q white noises $(q \leq r)$.

Denote by

$$\mathbf{\Lambda}_r^{(n)} := \begin{pmatrix} \lambda_1^{(n)} & \dots & 0 \\ & \ddots & \\ 0 & \dots & \lambda_r^{(n)} \end{pmatrix} \quad \text{and} \quad \mathbf{P}_r^{(n)} := \begin{pmatrix} \mathbf{p}_1^{(n)\prime} \\ \vdots \\ \mathbf{p}_r^{(n)\prime} \end{pmatrix} := \begin{pmatrix} p_{11}^{(n)} & \dots & p_{1n}^{(n)} \\ \vdots & \ddots & \vdots \\ p_{r1}^{(n)} & \dots & p_{rn}^{(n)} \end{pmatrix}$$

the $r \times r$ diagonal matrix containing the r largest eigenvalues of the $n \times n$ covariance matrix E$[\mathbf{X}_t^{(n)} \mathbf{X}_t^{(n)\prime}]$ and the corresponding $r \times n$ matrix of row eigenvectors.

Then

$$(\mathbf{\Lambda}_r^{(n)})^{-1/2} \mathbf{P}_r^{(n)} \mathbf{X}_t^{(n)}$$

is the standardized projection of $\mathbf{X}_t^{(n)}$ onto the r-dimensional space spanned by $\mathbf{X}_t^{(n)}$'s r first principal components (standard, static PCA).

It can be shown (see Bai and Ng (2002) or Stock and Watson (2002a and b) and many others) that, provided that

- the assumptions of the static factor model hold (hard to check for!), and
- further assumptions on the cross-correlations of the static principal components (which are othogonal at lag 0 only) are satisfied (hard to check for!),

those projections converge, as both n and T tend to infinity, to the space spanned by $\mathbf{F}_t$, the only identified feature of the model.

However, the static factor model is placing severe restrictions on the data-generating process.

Those restrictions at first sight may look quite innocuous, but they are not.

Consider, for instance, the very simple case under which $q = 1$ (one single common shock U_t) and the elementary AR(1) loading scheme

$$\chi_{it} = \rho_i \chi_{i,t-1} + U_t \qquad t \in \mathbb{Z},\ \rho_i \in (-1, 1),\ i \in \mathbb{N}.$$

with ρ_i drawn from a uniform distribution between, say, -.8 and .8

Here the stochastic variables χ_{it} for fixed t and $i \in \mathbb{N}$ are spanning an infinite-dimensional space. This very simple case does not admit a finite-r static representation, as each lag of U_t has to be counted as one distinct factor.

This very simple case does not admit a finite-r static representation, as each lag of U_t has to be counted as one distinct factor.

7. Dynamic PCA and Dynamic Factor Models

When either the assumptions of static factor models or those on the cross-correlations of the static principal components fail to hold, static PCA does not provide a consistent estimation of the common/idiosyncratic decomposition.

Fortunately, if projections onto the static principal components are replaced by projections onto the so-called dynamic principal components, consistency holds (Forni et al. 2000) without any assumptions (but Assumptions A1–A3).

Brillinger's Dynamic Principal Components

The problem with traditional (static) principal components in a time series context is that serial dependencies are overlooked/unexploited.

A static principal component $\sum_{i=1}^{n} a_i \zeta_{i,t}$ associated with a small static eigenvalue may have a negligible instantaneous impact on $\boldsymbol{\zeta}_t$. But the same linear combination may have a high covariance with $\boldsymbol{\zeta}_{t+1}$, hence a high predictive value: discarding it then results in a significant loss of information.

As a result, static principal components in general do not provide any reasonable solution to the dimension reduction problem in the presence of serial dependence. Besides this conceptual failure, static principal components, when computed from serially dependent observations, also run into technical problems: while cross-sectionally uncorrelated at fixed time t, they typically still exhibit lagged cross-correlations.

Brillinger's Dynamic Principal Components

The concept of dynamic principal component was introduced by Brillinger (1981).

Brillinger's Dynamic Principal Components

The concept of dynamic principal component was introduced by Brillinger (1981).

Dynamic principal components are based on a factorization of spectral density matrices. Therefore, let us make the additional assumption that all $\{X_t, t \in \mathbb{Z}\}$ processes in the panel admit spectral densities. More precisely, we assume the following.

Brillinger's Dynamic Principal Components

The concept of dynamic principal component was introduced by Brillinger (1981).

Dynamic principal components are based on a factorization of spectral density matrices. Therefore, let us make the additional assumption that all $\{X_t, t \in \mathbb{Z}\}$ processes in the panel admit spectral densities. More precisely, we assume the following.

ASSUMPTION A3. For all $n \in \mathbb{N}$, the spectral measure of $\{\mathbf{X}_t^{(n)}, t \in \mathbb{Z}\}$ is absolutely continuous with respect to the Lebesgue measure on $[-\pi, \pi]$, that is, $\{\mathbf{X}_t^{(n)}\}$ has a spectral density matrix $\mathbf{\Sigma}^{(n)}(\theta)$, with entries $(\sigma_{ij}(\theta))$, $\theta \in [-\pi, \pi]$.

- The matrices $\mathbf{\Sigma}^{(n)}(\theta)$ are nested for all θ (so that $\sigma_{ij}(\theta)$ needs no n superscript), Hermitian and positive semidefinite.

- Spectral densities always are defined up to a set of θ values with Lebesgue measure zero; rather than functions, we are dealing with equivalence classes of a.e. equal functions, thus.

- By $\mathbf{\Sigma}^{(n)}(\theta)$, in the sequel, we tacitly mean a representative of such a class; the same comment applies to $\mathbf{\Sigma}^{(n)}(\theta)$'s eigenvalues and eigenvectors.

Brillinger's Dynamic Principal Components

Since each $\mathbf{\Sigma}^{(n)}(\theta)$ is Hermitian positive semidefinite, it has n nonnegative eigenvalues, associated with n eigenvectors

$$\lambda_1^{(n)}(\theta) \geq \lambda_2^{(n)}(\theta) \geq \ldots \geq \lambda_n^{(n)}(\theta) \quad \text{and} \quad \mathbf{p}_1^{(n)}(\theta), \mathbf{p}_2^{(n)}(\theta), \ldots, \mathbf{p}_n^{(n)}(\theta);$$

call them the *dynamic* eigenvalues and *dynamic* eigenvectors of $\{\mathbf{X}_t^{(n)}\}$, respectively.

Brillinger's Dynamic Principal Components

It can be shown that the univariate process

$$\{W_{jt}^{(n)} := \underline{\mathbf{p}}_j^{(n)\prime}(L)\mathbf{X}_t^{(n)} |\ t \in \mathbb{Z}\}$$

with variance (the integrated jth dynamic eigenvalue)

$$\lambda_j^{(n)} := \int_{-\pi}^{\pi} \lambda_j^{(n)}(\theta)\mathrm{d}\theta$$

is the normed linear combination of past, present and future values of the n-dimensional process $\mathbf{X}^{(n)}$ maximizing the variance subject to being orthogonal (all leads and lags) to $W_{1t}^{(n)} ... W_{j-1,t}^{(n)}$.

Call it $\mathbf{X}^{(n)}$'s jth *dynamic principal component* ($j = 1, \ldots, n$)

It is easily seen that A2, under A1-A3, implies that

- the qth eigenvalue $\lambda_q^{(n)}(\theta)$ diverges to infinity as $n \to \infty$ θ-a.e. in $[-\pi,\ \pi]$, and
- the $(q+1)$thth eigenvalue $\lambda_{q+1}^{(n)}(\theta)$ is bounded as $n \to \infty$ θ-a.e. in $[-\pi,\ \pi]$.

Dynamic PCA and dynamic factor models

Denote by $\chi_{it}^{(n)}$ the projection of X_{it} onto the space spanned by the first q dynamic components $W_{1t}^{(n)}, \ldots, W_{qt}^{(n)}$. That projection takes the form

$$\chi_{it}^{(n)} := \underline{\mathbf{K}}_i^{(n)}(L)\mathbf{X}_t^{(n)},$$

where

$$\mathbf{K}_i^{(n)}(\theta) := p_{1,i}^{(n)*}(\theta)\mathbf{p}_1^{(n)\prime}(\theta) + \ldots + p_{q,i}^{(n)*}(\theta)\mathbf{p}_q^{(n)\prime}(\theta) \quad \theta \in [-\pi, \pi].$$

The following result shows how the sequence $\chi_{it}^{(n)}$ yields a consistent reconstruction of the general dynamic factor decomposition: basically,

$$\chi_{it}^{(n)} = \chi_{it} + o_{\mathrm{P}}(1) \quad \xi_{it}^{(n)} := X_{it}^{(n)} - \chi_{it}^{(n)} = \xi_{it} + o_{\mathrm{P}}(1) \quad \text{as} \quad n \to \infty,\ i \in \mathbb{N},\ t \in \mathbb{N}.$$

In practice, spectral densities (hence the filters $\underline{\mathbf{K}}_i^{(n)}(L)$) are to be estimated from the data (finite n and T); we will not enter into the details of the estimation of spectral densities (the estimators we are using involve lag-window estimates of the cross-spectrum of X_{it} and X_{jt}).

This yields consistent estimators $\chi_{it}^{n,T}$ and $\xi_{it}^{n,T}$ of χ_{it} and ξ_{it}. More precisely,

THEOREM. (Forni, Hallin, Lippi, and Reichlin 2000). *Under very general assumptions on the lag-window estimates of spectral densities, for all $\varepsilon > 0$ and $\eta > 0$, there exists a $N_0(\varepsilon, \eta)$ such that*

$$\mathrm{P}\left[\left|\chi_{it}^{n,T} - \chi_{it}\right| > \varepsilon\right] \leq \eta \quad \textit{and} \quad \mathrm{P}\left[\left|\xi_{it}^{n,T} - \xi_{it}\right| > \varepsilon\right] \leq \eta$$

for all sequence $t = t^*(T)$ satisfying

$$0 < a \leq \liminf_{T\to\infty} \frac{t^*(T)}{T} \leq \limsup_{T\to\infty} \frac{t^*(T)}{T} \leq b < 1$$

for some (fixed) a and b, all $n \geq N_0$, and all T larger than some $T_0(n, \varepsilon, \eta)$.

Dynamic PCA and dynamic factor models

This is fine in the "center" of the observation period, but not quite useful at the end of it (for $i \sim n$), hence for forecasting ...
The reason is that , unlike the projections on static principal components, the projection onto Brillinger's dynamic principal components typically yields two-sided filters $\mathbf{K}_i^{(n)}(L)$.

As a consequence, practitioners have been favoring the static factor model despite of its more restrictive (and hardly checkable) assumptions and lack of parsimony

That two-sidedness issue has been solved later in Forni, Hallin, Lippi and Zaffaroni (2015 and 2017), where a relation between dynamic and static facto models is established.

8. One-sidedness issues: back to the Static Model!

A solution indeed was made possible, thanks to a result in

Anderson, B.D.O. and M. Deistler (2008). Properties of zero-free transfer function matrices, *SICE Journal of Control, Measurement and System Integration* **1**, 1-9.

on “tall” vectors.

tall vectors

We say that an n-dimensional vector process is tall if its dynamic rank (the dimension of the white noise it is driven by) is strictly less than n.

For example, any $(q+1)$-dimensional vector of common components

$$(\chi_{1t} \ \chi_{2t} \ \cdots \ \chi_{q+1,t}),$$

has dimension $q+1$ while its dynamic rank cannot exceed q: such a $(q+1)$-tuple is *tall*.

Anderson and Deistler prove the following result.

Let $\mathbf{w}_t$ be q-dimensional white noise. Consider the tall vector

$$\mathbf{Y}_t = \mathbf{D}(L)\mathbf{w}_t$$

with dimension $n > q$ and rank q. Assume that $\mathbf{D}(L)$ (a $(n \times q)$ filter) is rational, that is, there exist matrix filters $(E_{ij}(L))$ and $(F_{ij}(L))$ such that

$$\mathbf{D}(L) = \left(\frac{E_{ij}(L)}{F_{ij}(L)} \right), \quad i = 1, \ldots, n, \;\; j = 1, \ldots q$$

where $F_{ij}(0) = 1$, $(E_{ij}(L))$ has degree m, and $(F_{ij}(L))$ has degree p: $\mathbf{D}(L)$ thus involves $P = nq(m + p + 1)$ parameters.

For generic values of the parameters in $\mathbb{R}^P$, $\mathbf{Y}_t$ admits a finite autoregressive representation

$$\mathbf{A}(L)\mathbf{Y}_t = \mathbf{R}\mathbf{w}_t$$

where $\mathbf{A}(L)$ is $n \times n$ and $\mathbf{R}$ is $n \times q$.
Moreover it can be proved that, for $n = q + 1$, this autoregressive representation (of minimum order) is unique

Tall vectors and the dynamic factor model

For almost all values of $i_1, \dots, i_{q+1}$, it holds that the $(q+1)$-tuple $(\chi_{i_1,t}, \dots, \chi_{i_{q+1},t})$ has dynamic rank q, hence is tall with dimension $(q+1)$ and rank q. Assume for convenience that $n = m(q+1)$, where $m \in \mathbb{N}$, and write $(\chi_{1,t}, \dots, \chi_{n,t})$ as $(\boldsymbol{\chi}_t^{1\prime}, \dots, \boldsymbol{\chi}_t^{m\prime})'$.

The Anderson-Deistler result applies, so that

$$\begin{pmatrix} \mathbf{A}^1(L) & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{A}^2(L) & \cdots & \mathbf{0} \\ & & \ddots & \\ \mathbf{0} & \mathbf{0} & \cdots & \mathbf{A}^m(L) \end{pmatrix} \begin{pmatrix} \boldsymbol{\chi}_t^1 \\ \boldsymbol{\chi}_t^2 \\ \vdots \\ \boldsymbol{\chi}_t^m \end{pmatrix} = \begin{pmatrix} \mathbf{R}^1 \\ \mathbf{R}^2 \\ \vdots \\ \mathbf{R}^m \end{pmatrix} \mathbf{v}_t,$$

that is, for some $\mathbf{A}_n(L)$, $\mathbf{R}_n$ and $\mathbf{v}_t$, we have the AR representation

$$\mathbf{A}_n(L)\boldsymbol{\chi}_{nt} = \mathbf{R}_n\mathbf{v}_t$$

where the vectors $\boldsymbol{\chi}_t^k$ are non overlapping $(q+1)$-dimensional subvectors of $(\chi_{1,t}, \dots, \chi_{n,t})$, the blocks $\mathbf{A}^k(L)$ are $(q+1) \times (q+1)$, the matrix $\mathbf{R}_n$ is $n \times q$ and $\mathbf{v}_t$ is q-dimensional white noise.

From general dynamic to static factors!

Since

$$\mathbf{A}_n(L)\boldsymbol{\chi}_{nt} = \mathbf{R}_n\mathbf{v}_t$$

we have

$$\mathbf{A}_n(L)\mathbf{X}_{nt} = \mathbf{A}_n(L)(\boldsymbol{\chi}_{nt} + \boldsymbol{\xi}_{nt}) = \mathbf{R}_n\mathbf{v}_t + \mathbf{A}_n(L)\boldsymbol{\xi}_{nt}$$

where, being a linear transformation of idiosyncratic components, $\boldsymbol{A}_n(L)\boldsymbol{\xi}_{nt}$ itself is idiosyncratic:

THIS IS A STATIC FACTOR MODEL
FOR THE FILTERED PANEL $\mathbf{A}_n(L)\mathbf{X}_{nt}$.

From this, Forni, Hallin, Lippi and Zaffaroni (*Journal of Econometrics,* 2015, 2017) deduce a winning strategy for a one-sided reconstruction of the common components χ_{it} (hence, also for the ξ_{it}'s),

yielding consistency rates comparable to those obtained (Bai 2003) for the static method—-the validity of which, however, requires the much more stringent assumptions of the static model.

Numerical exercises (both Monte-Carlo and empirical: see Forni, Giovannelli, Lippi, and Soccorsi (2016) "Dynamic Factor model with infinite dimensional factor space: forecasting," CEPR Discussion Papers 11161) demonstrate the forecasting superiority of the resulting method, which seems to outperform all other methods proposed in the literature while remaining valid under much milder and more general assumptions on the data-generating process.

9. Some conclusions, and future perspectives

The General Dynamic Factor Model method

- successfully can handle high-dimensional time series data without placing (much) restrictions on the data-generating process—a model-free approach besed on a general representation result

- contains all other factor models as particular cases

- is computationally comparable to its static counterpart (essentially, only requires additional $(q+1)$-dimensional VAR fitting)

- is based on a decomposition "reduced rank" + "idiosyncratic" that admits an intuitive interpretation in terms of common shocks on top of its "operational justification" (contrary to other similarly "operational" decompositions such as "reduced rank" + "sparse" ...)

- outperforms its competitors even in case the assumptions required for their validity are satisfied

All codes for general dynamic factor model methods (with comments) can be downloaded from Matteo Barigozzi's website at the LSE.

Returns and Volatilities in High Dimension
a General Factor Model Approach

2. Determining the Number of Factors

Marc Hallin

ECARES, Université Libre de Bruxelles

Hermann Otto Hirschfeld Lecture Series 2017

Berlin, 20–21 November 2017

Summing up, under very general assumptions A1 (second-order stationarity), A2 (unspecified but finite number q of common shocks), and A3 (existence of spectral densities), we have established

- the characteristic behavior of the dynamic eigenvalues of $\mathbf{X}^{(n)}$:

$$\lambda_q^{(n)}(\theta) \text{ diverges to } \infty \quad \theta\text{-a.e. as } n \to \infty;$$

$$\lambda_{q+1}^{(n)}(\theta) \quad \theta\text{-a.e. bounded as } n \to \infty.$$

- that the (two-sided) estimation methods by Forni et al. (2000) projecting X_{it} onto its estimated q first dynamic principal components provides a consistent, as n and $T \to \infty$, reconstruction of the common component χ_{it}

- that the (one-sided) estimation methods by Forni et al. (2015, 2017) projecting an autoregressive transform of X_{it} onto its estimated q first static principal components provides a consistent, as n and $T \to \infty$, reconstruction of the common component χ_{it}

So far, however, it has been assumed that q is known. In practice, q has to be identified from the data.

One possibility is eye-inspection

Example: evolution of US sectors GDP growth rates (450 series), over 28 years

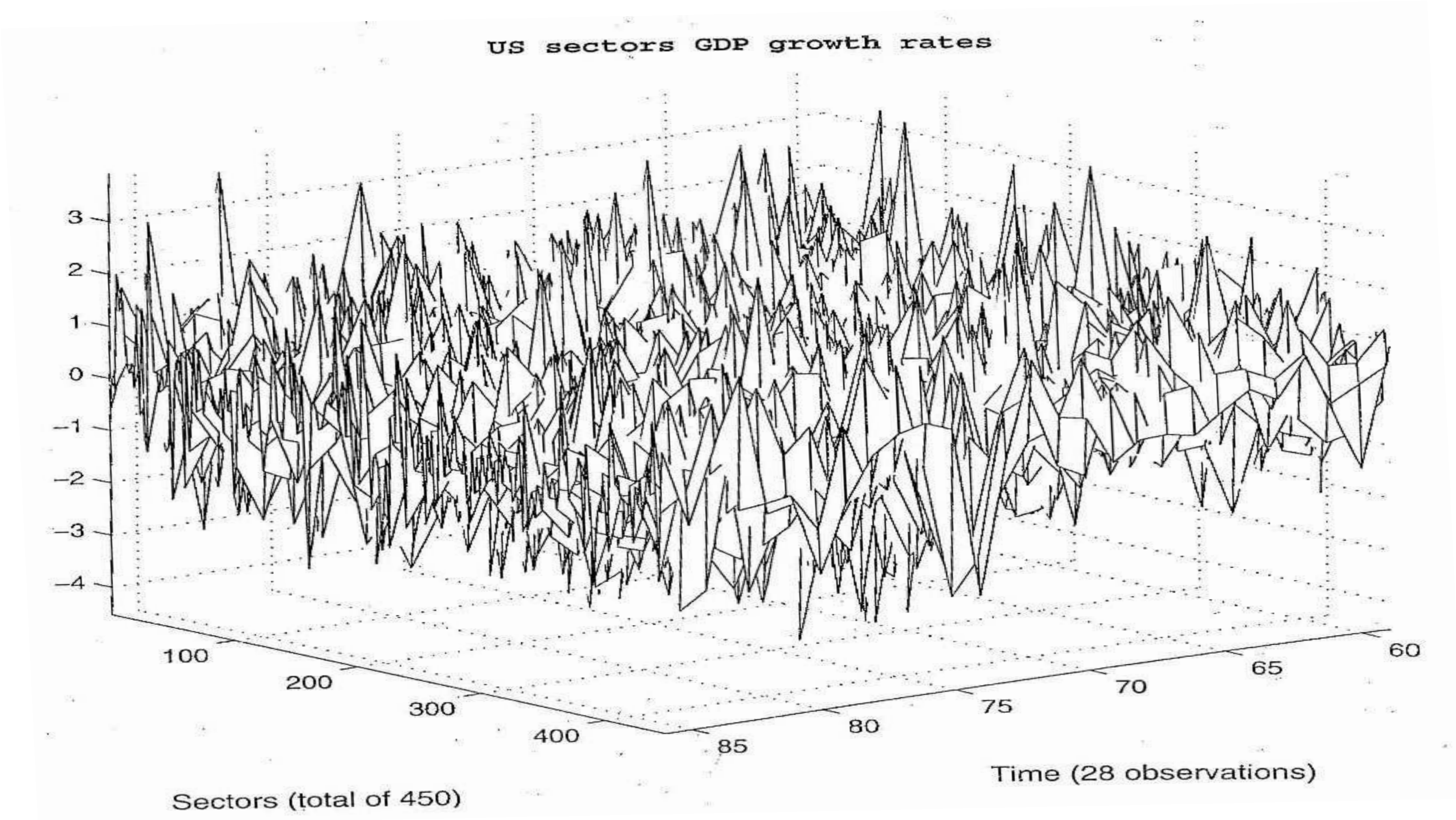
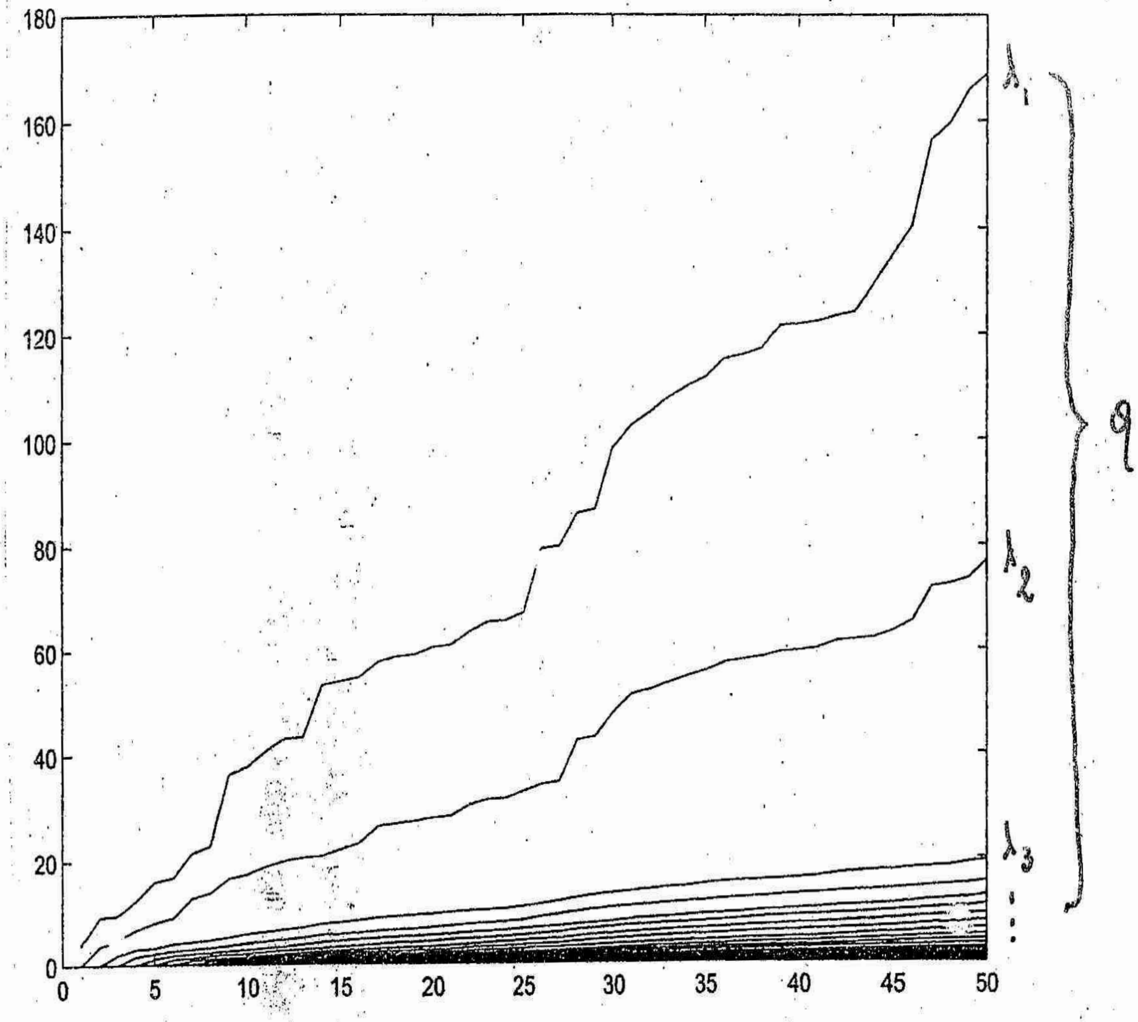


FIGURE 1.—DYNAMIC EIGENVALUES AVERAGED OVER FREQUENCIES



Horizontal axis: ~~n~~ m, vertical axis: variance/2π.

Evolution of US sectors GDP growth rates (450 series), over 28 years : integrated spectral eigenvalues $\int_{-\pi}^{\pi} \lambda_j^{(n)}(\theta)\mathrm{d}\theta$

In practice, eye-inspection is not that bad ... however, more sophisticated identification methods are desirable.

- An information-theoretic identification procedure has been first proposed in

Bai, J. and S. Ng (2002) Determining the number of factors in approximate factor models. *Econometrica* **70**, 191-221.

for the static case. As we shall see, it is not without drawbacks (and, in practice, it performs poorly).

- The dynamic case has been considered in

Hallin M. and R. Liška (2007). Determining the number of factors in the general dynamic factor model. *Journal of the American Statistical Association* **102**, 603-617.

- Based on the ideas developed there, the drawbacks of Bai and Ng have been fixed in

Alessi, L., M. Barigozzi, and M. Capasso (2010). Improved penalization for determining the number of factors in approximate factor models. *Statistics & Probability Letters* **80**, 1806-1813.

- a totally different method, based on the theory of random matrices and hypothesis testing ideas, has been proposed in

Onatski, A. (2010). Determining the number of factors from empirical distribution of eigenvalues. *The Review of Economics and Statistics* **92**, 1004-1016.

but for "exact" static factor models

We will focus on Hallin and Liška (2007) and the general dynamic case (Alessi et al. being a particular case).

1. Population level

In order to fix the ideas, first assume that the spectral density matrices $\mathbf{\Sigma}^{(n)}(\theta)$, with eigenvalues $\lambda_j^{(n)}(\theta)$, are known

1. Population level

In order to fix the ideas, first assume that the spectral density matrices $\mathbf{\Sigma}^{(n)}(\theta)$, with eigenvalues $\lambda_j^{(n)}(\theta)$, are known

Let

$$\hat{q}_n := \mathrm{Argmin}_{0 \leq k \leq k_{\max}} L^{(n)}(k)$$

where $k_{\max}$ is some maximal number of factors we are willing to consider, and

$$L^{(n)}(k) := \frac{1}{n} \sum_{j=k+1}^{n} \int_{-\pi}^{\pi} \lambda_j^{(n)}(\theta) d\theta + kp(n)$$

with

- $\frac{1}{n} \sum_{j=k+1}^{n} \int_{-\pi}^{\pi} \lambda_j^{(n)}(\theta) d\theta$ is the average contribution of idiosyncratics, would k factors be selected ("unexplained variance", to be minimized)
- $kp(n)$ a penalty (or we would automatically select $\hat{q}_n = n$, which yields $L^{(n)}(k) = 0$)

1. Population level

In order to fix the ideas, first assume that the spectral density matrices $\mathbf{\Sigma}^{(n)}(\theta)$, with eigenvalues $\lambda_j^{(n)}(\theta)$, are known

Let

$$\hat{q}_n := \text{Argmin}_{0 \leq k \leq k_{\max}} L^{(n)}(k)$$

where $k_{\max}$ is some maximal number of factors we are willing to consider, and

$$L^{(n)}(k) := \frac{1}{n} \sum_{j=k+1}^{n} \int_{-\pi}^{\pi} \lambda_j^{(n)}(\theta) d\theta + kp(n)$$

with

- $\frac{1}{n} \sum_{j=k+1}^{n} \int_{-\pi}^{\pi} \lambda_j^{(n)}(\theta) d\theta$ is the average contribution of idiosyncratics, would k factors be selected ("unexplained variance", to be minimized)
- $kp(n)$ a penalty (or we would automatically select $\hat{q}_n = n$, which yields $L^{(n)}(k) = 0$)

Note that $\hat{q}_n$ here is not random (since the λ's aren't)

the choice of the penalty factor $p(n)$ is crucial:

the choice of the penalty factor $p(n)$ is crucial:

- it has to go to zero as $n \to \infty$, or a slowly diverging eigenvalue will go undetected (overpenalization)

the choice of the penalty factor $p(n)$ is crucial:

- it has to go to zero as $n \to \infty$, or a slowly diverging eigenvalue will go undetected (overpenalization)
- it cannot not go to zero too fast, or $\hat{q}_n \to \infty$ as $n \to \infty$ (underpenalization).

the choice of the penalty factor $p(n)$ is crucial:

- it has to go to zero as $n \to \infty$, or a slowly diverging eigenvalue will go undetected (overpenalization)
- it cannot not go to zero too fast, or $\hat{q}_n \to \infty$ as $n \to \infty$ (underpenalization).

The divergence rate of the diverging eigenvalues $\lambda_1^{(n)}, \ldots, \lambda_q^{(n)}$ plays a fundamental role here

Assumption (A4) The smallest diverging eigenvalue of $\mathbf{\Sigma}^{(n)}$ diverges linearly in n, that is, there exist constants $0 < c_q^- \leq c_q^+$ such that

$$c_q^- \leq \liminf_{n\to\infty} n^{-1}\lambda_q^{(n)}(\theta) \leq \limsup_{n\to\infty} n^{-1}\lambda_q^{(n)}(\theta) \leq c_q^+$$

(a natural assumption if some "cross-sectional stability" of the influence of each factor, implying a linear divergence of all diverging $\lambda_j^{(n)}(\theta)$'s, is assumed), and the only rate invariant under cross-sectional permutations.

However, only $c_q^- \leq \liminf_{n\to\infty} n^{-1}\lambda_q^{(n)}(\theta)$ is needed below.

Lemma. *Let the penalty* $p(n)$ *be such that*

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

Then, $\lim_{n\to\infty} \hat{q}_n = q$ *(*q *is consistently identified).*

consistency: population level

Lemma. *Let the penalty $p(n)$ be such that*

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

Then, $\lim_{n\to\infty} \hat{q}_n = q$ (q is consistently identified).

Important, but distressing Remark. The function $n \mapsto p(n)$ is an adequate penalty (yielding consistent $\hat{q}_n$)

iff

$n \mapsto cp(n)$ also is, where $c > 0$ is an arbitrary constant!

proof ...

The proof is extremely simple.

proof ...

The proof is extremely simple.

Let us show that

$$\lim_{n\to\infty}[L^{(n)}(k) - L^{(n)}(q)] > 0$$

for all $k \neq q$ (where q is the "true" number of factors—so that, for n large enough, q is a minimizer of $L^{(n)}(k)$)

proof ...

The proof is extremely simple.

Let us show that

$$\lim_{n\to\infty}[L^{(n)}(k) - L^{(n)}(q)] > 0$$

for all $k \neq q$ (where q is the "true" number of factors—so that, for n large enough, q is a minimizer of $L^{(n)}(k)$)

that is, let us show that there exists a finite n_0 such that for all $n > n_0$ and $k \neq q$,

$$\frac{1}{n}\sum_{j=k+1}^{n}\left\{\int_{-\pi}^{\pi}\lambda_j^{(n)}(\theta)d\theta\right\} + kp(n) > \frac{1}{n}\sum_{j=q+1}^{n}\left\{\int_{-\pi}^{\pi}\lambda_j^{(n)}(\theta)d\theta\right\} + qp(n)$$

proof ...

- either $k > q$; the claim takes the form

$$\sum_{j=k+1}^{n}\left\{\int_{-\pi}^{\pi}\lambda_j^{(n)}(\theta)d\theta\right\}+nkp(n)>\sum_{j=q+1}^{k}\left\{\int_{-\pi}^{\pi}\lambda_j^{(n)}(\theta)d\theta\right\}+\sum_{j=k+1}^{n}\left\{\int_{-\pi}^{\pi}\lambda_j^{(n)}(\theta)d\theta\right\}+nqp(n),$$

proof ...

- either $k > q$; the claim takes the form

$$\sum_{j=k+1}^{n}\left\{\int_{-\pi}^{\pi}\lambda_j^{(n)}(\theta)d\theta\right\}+nkp(n) > \sum_{j=q+1}^{k}\left\{\int_{-\pi}^{\pi}\lambda_j^{(n)}(\theta)d\theta\right\}+\sum_{j=k+1}^{n}\left\{\int_{-\pi}^{\pi}\lambda_j^{(n)}(\theta)d\theta\right\}+nqp(n),$$

that is,

$$np(n) > \frac{1}{k-q}\sum_{j=q+1}^{k}\left\{\int_{-\pi}^{\pi}\lambda_j^{(n)}(\theta)d\theta\right\}$$

which follows, for n large enough, from the fact that $np(n) \to \infty$, whereas

$$\frac{1}{k-q}\sum_{j=q+1}^{k}\left\{\int_{-\pi}^{\pi}\lambda_j^{(n)}(\theta)d\theta\right\}$$

remains bounded.

proof ...

- or $k < q$; the claim takes the form

$$\sum_{j=k+1}^{q}\left\{\int_{-\pi}^{\pi}\lambda_j^{(n)}(\theta)d\theta\right\}+\sum_{j=q+1}^{n}\left\{\int_{-\pi}^{\pi}\lambda_j^{(n)}(\theta)d\theta\right\}+nkp(n)>\sum_{j=q+1}^{n}\left\{\int_{-\pi}^{\pi}\lambda_j^{(n)}(\theta)d\theta\right\}+nqp(n),$$

that is,

$$\frac{1}{n}\sum_{j=k+1}^{q}\left\{\int_{-\pi}^{\pi}\lambda_j^{(n)}(\theta)d\theta\right\}>(q-k)p(n)$$

proof ...

- or $k < q$; the claim takes the form

$$\sum_{j=k+1}^{q}\left\{\int_{-\pi}^{\pi}\lambda_j^{(n)}(\theta)d\theta\right\}+\sum_{j=q+1}^{n}\left\{\int_{-\pi}^{\pi}\lambda_j^{(n)}(\theta)d\theta\right\}+nkp(n)>\sum_{j=q+1}^{n}\left\{\int_{-\pi}^{\pi}\lambda_j^{(n)}(\theta)d\theta\right\}+nqp(n),$$

that is,

$$\frac{1}{n}\sum_{j=k+1}^{q}\left\{\int_{-\pi}^{\pi}\lambda_j^{(n)}(\theta)d\theta\right\}>(q-k)p(n)$$

which follows, for n large enough, from the fact that $p(n)\to 0$, while $\dfrac{1}{n}\lambda_j^{(n)}(\theta)$ is bounded from below for $j \leq q$

2. Sample Level

In practice, $\mathbf{\Sigma}^{(n)}(\theta)$ is unknown, and has to be estimated: denote by $\mathbf{\Sigma}^{(n,T)}(\theta)$ the lag-window spectral estimate

$$\mathbf{\Sigma}^{(n,T)}(\theta) := \frac{1}{\pi} \sum_{u=-M_T}^{M_T} w(M_T^{-1}u)\mathbf{\Gamma}_u^{(n,T)} e^{-iu\theta}$$

where $\mathbf{\Gamma}_u^{(n,T)}$ stands for the sample cross-covariance matrix of $\mathbf{X}_{nt}$ and $\mathbf{X}_{n,t-u}$ based on T observations, $\alpha \mapsto w(\alpha)$ is a positive even weight function, and M_T a truncation parameter

denote by $\lambda_i^{(n,T)}(\theta)$ the corresponding eigenvalues

Under adequate assumptions on M_T and w, the following uniform consistency property holds for $\mathbf{\Sigma}^{(n,T)}(\theta)$:

there exists constants L_1, L_2, and T_0 such that

$$\sup_n \max_{1\leq i,j\leq n} \sup_\theta \mathrm{E}\left[\left|\mathbf{\Sigma}^{(n,T)}(\theta) - \mathbf{\Sigma}^{(n)}(\theta)\right|^2_{ij}\right] \leq L_1 M_T T^{-1} + L_2 M_T^{-4}$$

for any $T > T_0$ (uniform version of Parzen 1957)

information criterion

The information criterion now takes the form

$$IC_n^T(k) := \frac{1}{n}\sum_{j=k+1}^{n}\frac{1}{2M_T+1}\sum_{\ell=-M_T}^{M_T}\lambda_j^{(n,T)}(\theta_\ell) + kp(n,T)$$

where $\theta_\ell := \pi\ell/(2M_T+1)$ for $\ell = -M_T, \ldots, M_T$, and $p(n,T)$ is a penalty now depending on both n and T

The information criterion now takes the form

$$IC_n^T(k) := \frac{1}{n}\sum_{j=k+1}^{n}\frac{1}{2M_T+1}\sum_{\ell=-M_T}^{M_T}\lambda_j^{(n,T)}(\theta_\ell) + kp(n,T)$$

where $\theta_\ell := \pi\ell/(2M_T+1)$ for $\ell = -M_T, \ldots, M_T$, and $p(n,T)$ is a penalty now depending on both n and T

The selected $\hat{q}_n^T$ is the (now, random) quantity

$$\hat{q}_n^T := \mathrm{Argmin}_{0\leq k\leq q_{\max}} IC_n^T(k)$$

N.B. The information criterion also can be taken under logarithmic form

$$IC_n^T(k) := \log\left[\frac{1}{n}\sum_{j=k+1}^{n}\frac{1}{2M_T+1}\sum_{\ell=-M_T}^{M_T}\lambda_j^{(n,T)}(\theta_\ell)\right] + kp(n,T)$$

Proposition. *Under assumptions A1-A4 and conditions*

(a) $p(n,T) \to 0$,

(b) min$(n, M_T^2, M_T^{-1/2}T^{1/2})p(n,T) \to \infty$,

(c) the entries $\sigma_{ij}(\theta)$ of $\mathbf{\Sigma}^{(n)}(\theta)$ are uniformly (in n and θ) bounded, and

(d) have uniformly bounded (in n and θ) derivatives up to the order two,

$\mathrm{P}\left[\hat{q}_n^T = q\right] \to 1$ as n and T tend to infinity

Proposition. *Under assumptions A1-A4 and conditions*

- *(a)* $p(n,T) \to 0$*,*
- *(b)* *min*$(n, M_T^2, M_T^{-1/2}T^{1/2})p(n,T) \to \infty$*,*
- *(c)* *the entries* $\sigma_{ij}(\theta)$ *of* $\mathbf{\Sigma}^{(n)}(\theta)$ *are uniformly (in* n *and* θ*) bounded, and*
- *(d)* *have uniformly bounded (in* n *and* θ*) derivatives up to the order two,*

$\mathrm{P}\left[\hat{q}_n^T = q\right] \to 1$ *as* n *and* T *tend to infinity*

Important Remark. Here again, $p(n,T)$ is thus an adequate penalty (yielding consistent $\hat{q}_n^T$) iff $cp(n,T)$ also is, where $c > 0$ is an arbitrary constant!

limited practical value of consistency

Due to the role of that arbitrary positive constant c, the above consistency result at first sight does not have any practical value: by varying c, one can obtain (for fixed n and T) any value of q between 0 and $q_{\max}$!!!!

limited practical value of consistency

Due to the role of that arbitrary positive constant c, the above consistency result at first sight does not have any practical value: by varying c, one can obtain (for fixed n and T) any value of q between 0 and $q_{\max}$!!!!

The same remark holds for most information criterion methods (certainly, for Bai and Ng's) ...

limited practical value of consistency

Due to the role of that arbitrary positive constant c, the above consistency result at first sight does not have any practical value: by varying c, one can obtain (for fixed n and T) any value of q between 0 and $q_{\max}$!!!!

The same remark holds for most information criterion methods (certainly, for Bai and Ng's) ...

Let us show, however, how this interplay between the constant c and the dependence of $\hat{q}_n^T$ on n and T can be exploited in a clever, cross-validation spirit

Typical behavior of $c \mapsto \hat{q}^T_{c;n}$

Let $\hat{q}^T_{c;n}$ denote the selected value of q based on the penalty function $cp(n, T)$

Typical behavior of $c \mapsto \hat{q}^T_{c;n}$

Let $\hat{q}^T_{c;n}$ denote the selected value of q based on the penalty function $cp(n, T)$

Empirical findings:

- there exist intervals of c values such that $\hat{q}^T_{c;n} = 0,\ 1, \ldots,\ q-1,\ q$ irrespective of n and T ("stability intervals")

Typical behavior of $c \mapsto \hat{q}^T_{c;n}$

Let $\hat{q}^T_{c;n}$ denote the selected value of q based on the penalty function $cp(n,T)$

Empirical findings:

- there exist intervals of c values such that $\hat{q}^T_{c;n} = 0,\ 1, \ldots,\ q-1,\ q$ irrespective of n and T ("stability intervals")
- for $c \simeq 0$, $\hat{q}^T_{c;n} \simeq q_{\max}$ (essentially, no penalty); for c "large", $\hat{q}^T_{c;n} = 0$ (gross overpenalization)

Typical behavior of $c \mapsto \hat{q}^T_{c;n}$

Let $\hat{q}^T_{c;n}$ denote the selected value of q based on the penalty function $cp(n, T)$

Empirical findings:

- there exist intervals of c values such that $\hat{q}^T_{c;n} = 0,\ 1, \ldots,\ q-1,\ q$ irrespective of n and T ("stability intervals")
- for $c \simeq 0$, $\hat{q}^T_{c;n} \simeq q_{\max}$ (essentially, no penalty); for c "large", $\hat{q}^T_{c;n} = 0$ (gross overpenalization)
- in between, consistency implies that $\hat{q}^T_{c;n}$ be an eventually redescending function of n, T

Typical behavior of $c \mapsto \hat{q}^T_{c;n}$

Let $\hat{q}^T_{c;n}$ denote the selected value of q based on the penalty function $cp(n, T)$

Empirical findings:

- there exist intervals of c values such that $\hat{q}^T_{c;n} = 0,\ 1, \ldots,\ q-1,\ q$ irrespective of n and T ("stability intervals")
- for $c \simeq 0$, $\hat{q}^T_{c;n} \simeq q_{\max}$ (essentially, no penalty); for c "large", $\hat{q}^T_{c;n} = 0$ (gross overpenalization)
- in between, consistency implies that $\hat{q}^T_{c;n}$ be an eventually redescending function of n, T

Explanation: for various c values, the values of $\hat{q}^T_{c;n}$ are well ordered for n (when plotted against n, graphs do not cross)

Typical behavior of $c \mapsto \hat{q}^T_{c;n}$

Let $\hat{q}^T_{c;n}$ denote the selected value of q based on the penalty function $cp(n, T)$

Empirical findings:

- there exist intervals of c values such that $\hat{q}^T_{c;n} = 0,\ 1, \ldots,\ q-1,\ q$ irrespective of n and T ("stability intervals")
- for $c \simeq 0$, $\hat{q}^T_{c;n} \simeq q_{\max}$ (essentially, no penalty); for c "large", $\hat{q}^T_{c;n} = 0$ (gross overpenalization)
- in between, consistency implies that $\hat{q}^T_{c;n}$ be an eventually redescending function of n, T

Explanation: for various c values, the values of $\hat{q}^T_{c;n}$ are well ordered for n (when plotted against n, graphs do not cross)

- for $c \sim 0$, $\hat{q}^T_{c;n} = q_{\max}$ (no penalty)
- for c small (underpenalization), $\hat{q}^T_{c;n}$ is too large, hence (consistency) tends to q from above
- for c neither too small nor to big, $\hat{q}^T_{c;n}$ is neither too small nor to big, hence is stable (in view of consistency, at q)
- for c big (overpenalization), $\hat{q}^T_{c;n}$ is too small, hence (consistency) tends to q from below
- for c very big (overpenalization), $\hat{q}^T_{c;n} = 0$

A numerical example

As an illustration, a panel of size $n = 300$ and length $T = 300$ was generated

- The common part was modelled with $q = 3$ factors and MA loadings
- The truncation parameter was set as $M_T = [0.75\sqrt{T}]$
- A triangular window was used, q_{max} was set to 19, and
- the penalty function
  $p_3(n, T) = \left(\min\left[n, M_T^2, M_T^{-1/2} T^{1/2}\right]\right)^{-1} \log\left(\min\left[n, M_T^2, M_T^{-1/2} T^{1/2}\right]\right)$
  was chosen.

A numerical example

As an illustration, a panel of size $n = 300$ and length $T = 300$ was generated

- The common part was modelled with $q = 3$ factors and MA loadings
- The truncation parameter was set as $M_T = [0.75\sqrt{T}]$
- A triangular window was used, q_{max} was set to 19, and
- the penalty function $p_3(n,T) = \left(\min\left[n, M_T^2, M_T^{-1/2}T^{1/2}\right]\right)^{-1} \log\left(\min\left[n, M_T^2, M_T^{-1/2}T^{1/2}\right]\right)$ was chosen.

A numerical example

As an illustration, a panel of size $n = 300$ and length $T = 300$ was generated

- The common part was modelled with $q = 3$ factors and MA loadings
- The truncation parameter was set as $M_T = [0.75\sqrt{T}]$
- A triangular window was used, q_{max} was set to 19, and
- the penalty function
$p_3(n, T) = \left(\min\left[n, M_T^2, M_T^{-1/2}T^{1/2}\right]\right)^{-1} \log\left(\min\left[n, M_T^2, M_T^{-1/2}T^{1/2}\right]\right)$
was chosen.

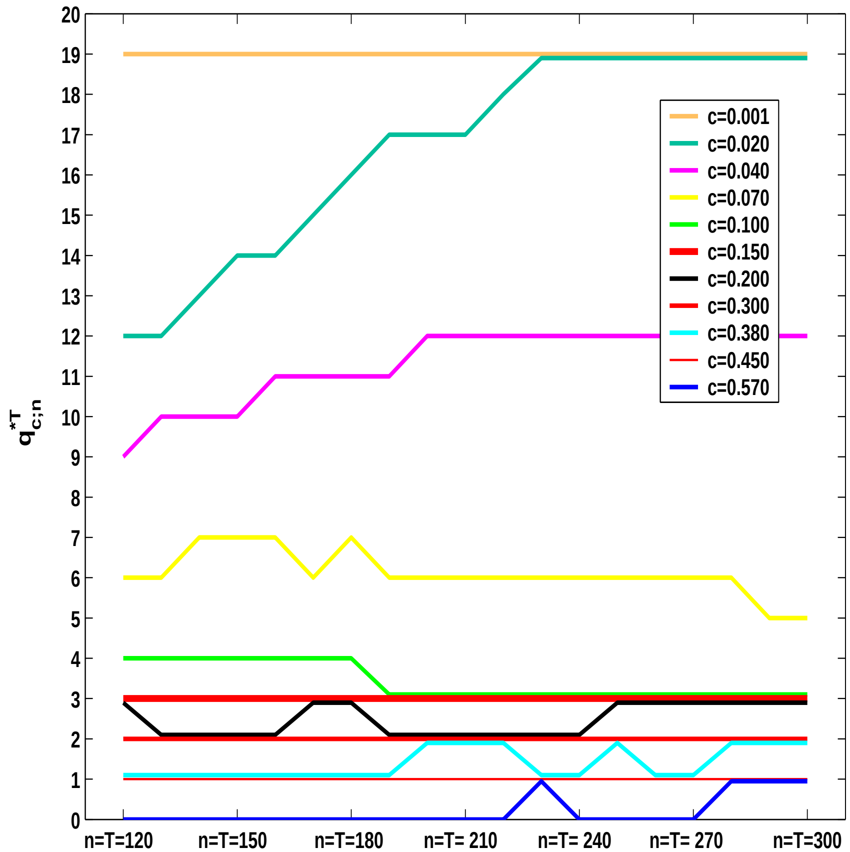
A numerical example

As an illustration, a panel of size $n = 300$ and length $T = 300$ was generated

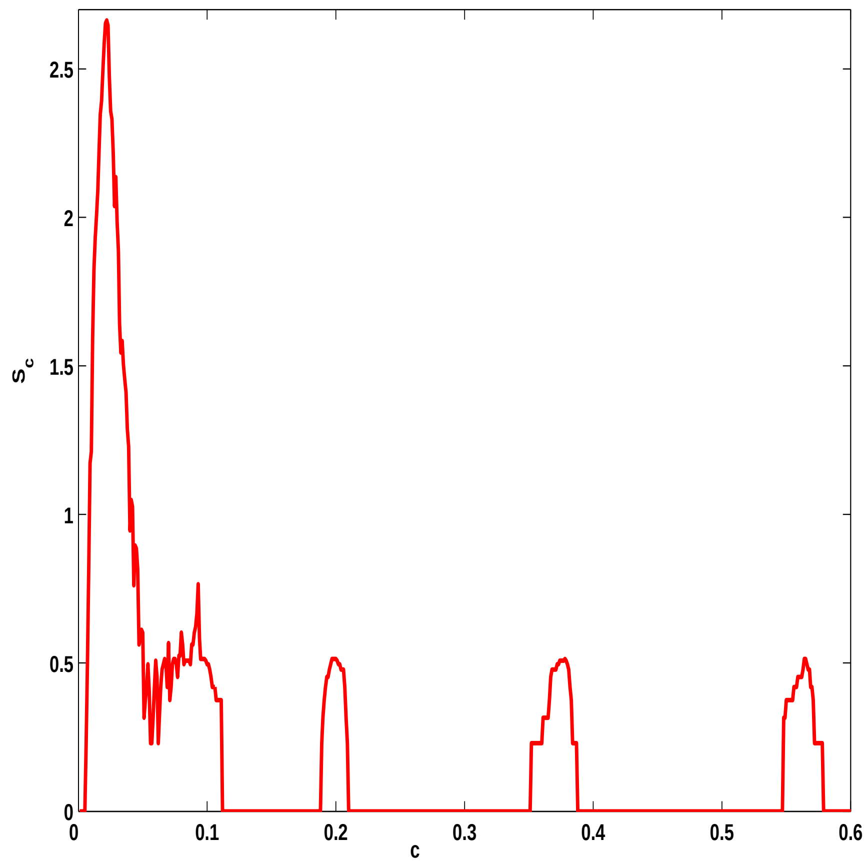
- The common part was modelled with $q = 3$ factors and MA loadings
- The truncation parameter was set as $M_T = [0.75\sqrt{T}]$
- A triangular window was used, q_{max} was set to 19, and
- the penalty function $p_3(n, T) = \left(\min\left[n, M_T^2, M_T^{-1/2}T^{1/2}\right]\right)^{-1} \log\left(\min\left[n, M_T^2, M_T^{-1/2}T^{1/2}\right]\right)$ was chosen.

MA loadings

(a1)



(a2)

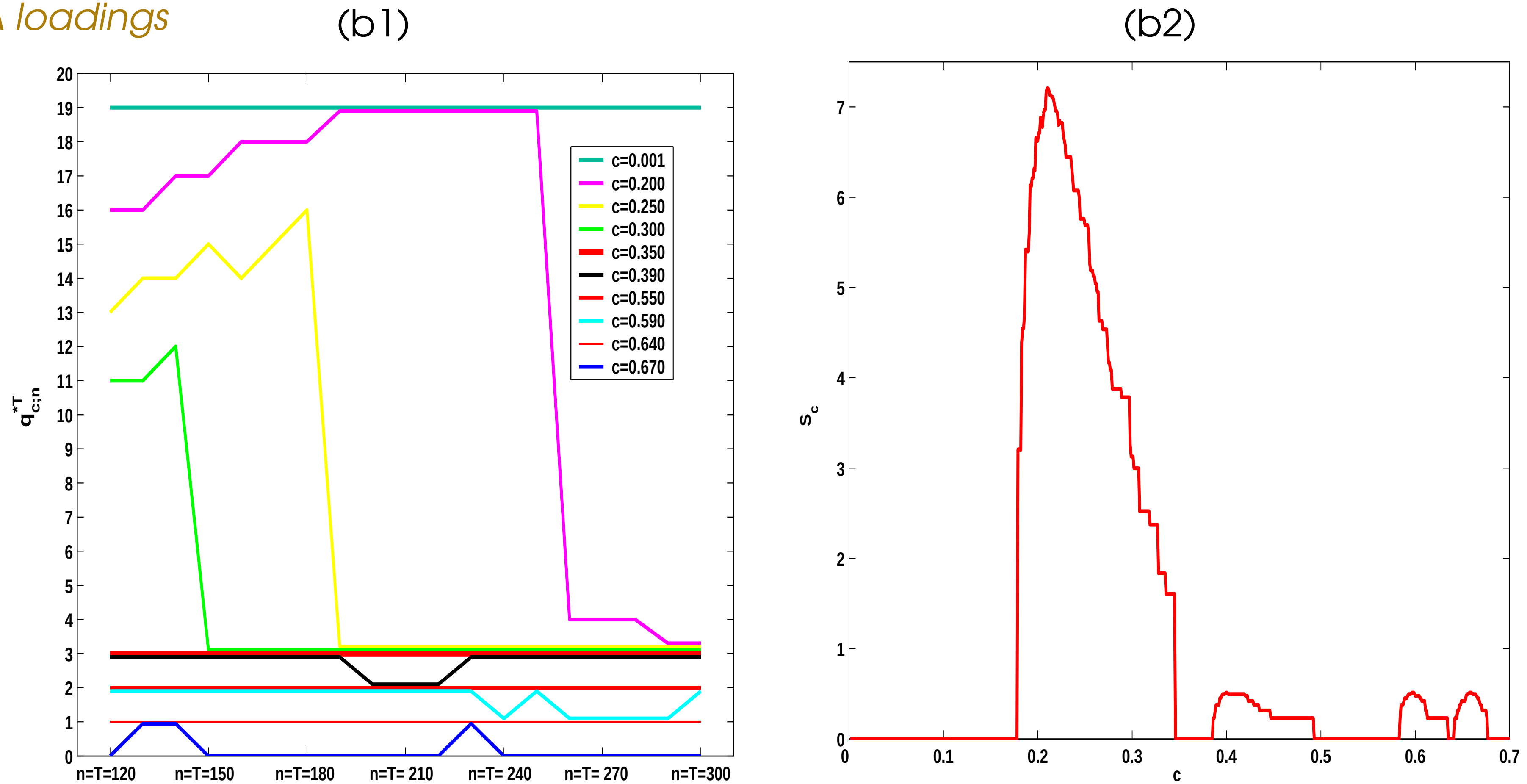


MA loadings, $q = 3$, $n = T = 300$, $M_T = [0.75\sqrt{T}]$. Graphs of $(n_j, T_j) \mapsto q^{T_j}_{c;n_j}$ and

$$c \mapsto S_c := \left[\frac{1}{IJ} \sum_{i,j} \left(\hat{q}^{T_j}_{c;n_i} - \frac{1}{IJ} \sum_{i,j} \hat{q}^{T_j}_{c;n_i} \right)^2 \right]^{1/2}$$ for

$(n_j, T_j) = (120, 120), (130, 130), \ldots, (300, 300)$ and various values of c, using penalty function p_1 $q_{\text{max}} = 19$, and $IC^T_{2;n}$

MA loadings



MA loadings, $q = 3$, $n = T = 300$, $M_T = [0.75\sqrt{T}]$. Graphs of $(n_j, T_j) \mapsto q^{T_j}_{c;n_j}$ and

$$c \mapsto S_c := \left[\frac{1}{IJ} \sum_{i,j} \left(\hat{q}^{T_j}_{c;n_i} - \frac{1}{IJ} \sum_{i,j} \hat{q}^{T_j}_{c;n_i} \right)^2 \right]^{1/2} \text{ for}$$

$(n_j, T_j) = (120, 120), (130, 130), \ldots, (300, 300)$ and various values of c, using penalty function p_1, $q_{\max} = 19$, and $IC^T_{2;n}$

The most informative picture (also, for automatic detection) is a joint plot of

$$c \mapsto S_c := \left[\frac{1}{IJ} \sum_{i,j} \left(\hat{q}^{T_j}_{c;n_i} - \frac{1}{IJ} \sum_{i,j} \hat{q}^{T_j}_{c;n_i} \right)^2 \right]^{1/2}$$

(for I values of n_i and J values of T_j) and

$$c \mapsto \hat{q}^{T}_{c;n}$$

(the “final” selection for c)

The most informative picture (also, for automatic detection) is a joint plot of

$$c \mapsto S_c := \left[\frac{1}{IJ} \sum_{i,j} \left(\hat{q}^{T_j}_{c;n_i} - \frac{1}{IJ} \sum_{i,j} \hat{q}^{T_j}_{c;n_i} \right)^2 \right]^{1/2}$$

(for I values of n_i and J values of T_j) and

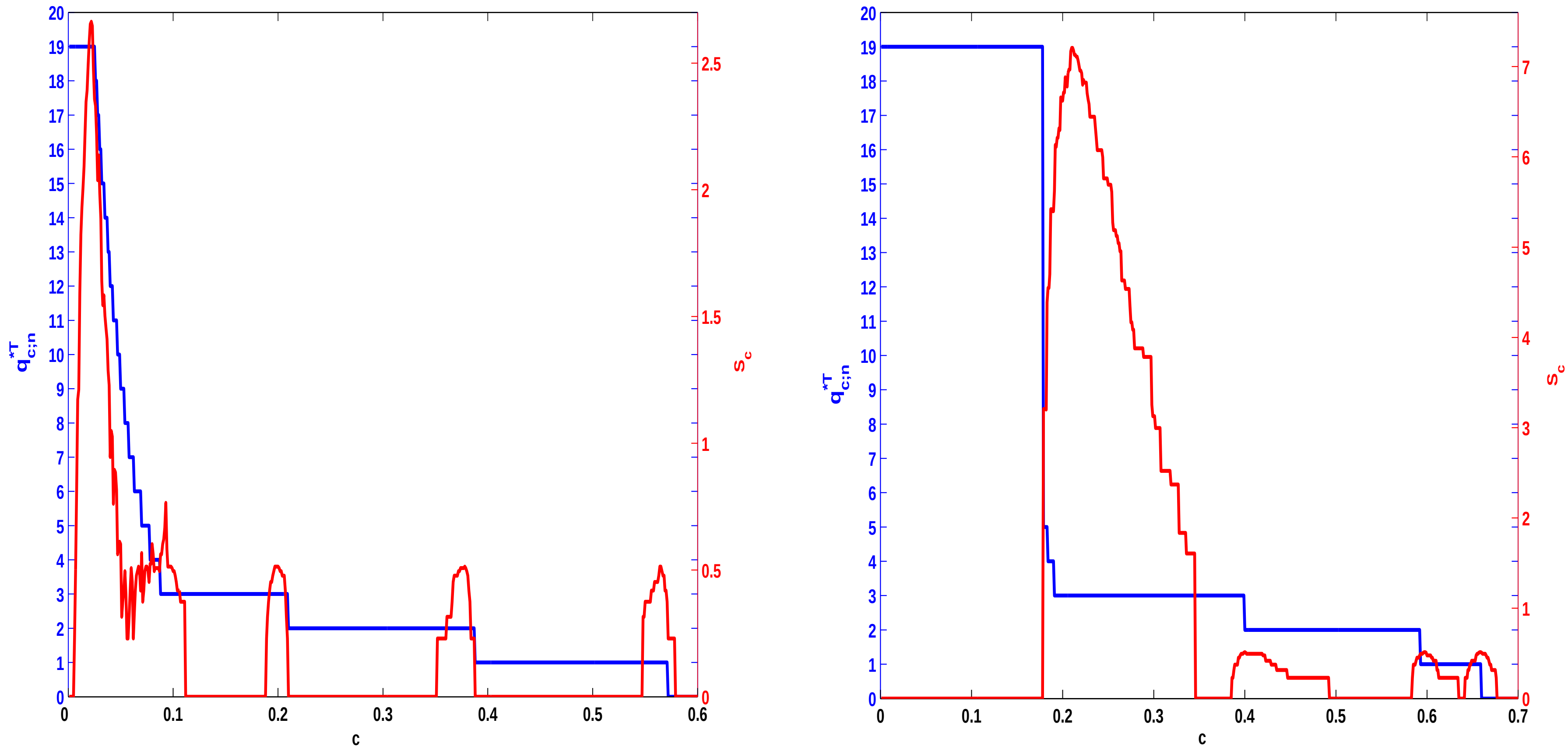
$$c \mapsto \hat{q}^{T}_{c;n}$$

(the “final” selection for c)

- a successful automatic selection procedure consists in picking the value of $\hat{q}^{T}_{c;n}$ associated with the second stability interval of $c \mapsto S_c$

MA loadings

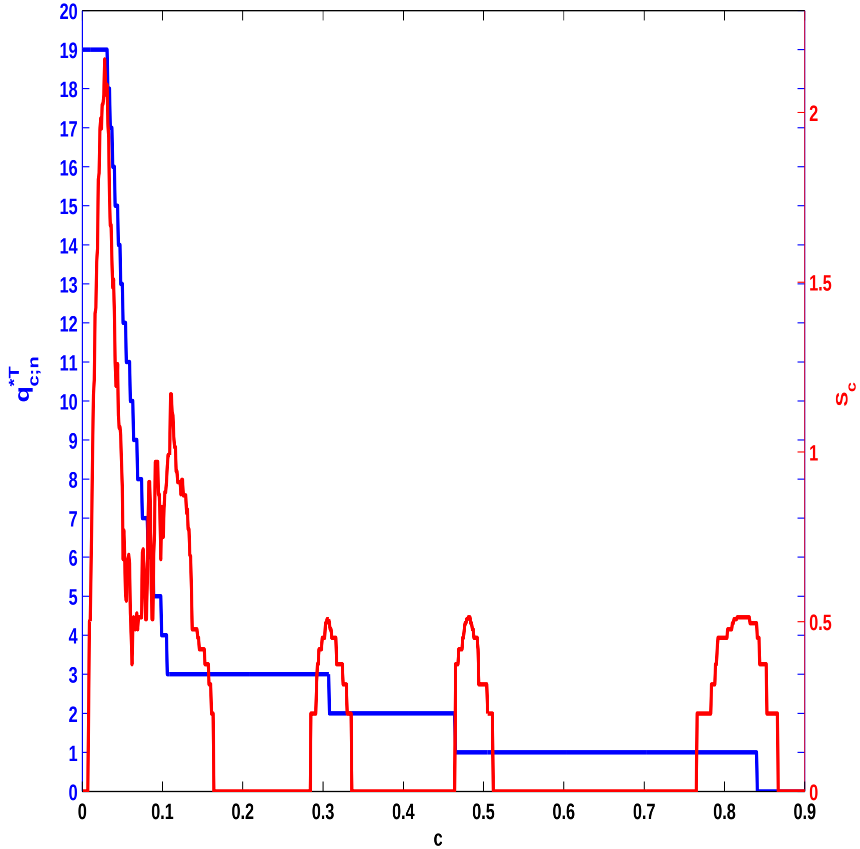
The relevant plots are thus
(c1)



MA loadings, $q = 3$, $n = T = 300$; $M_T = [0.75\sqrt{T}]$. Simultaneous plots of $c \mapsto S_c$ and $c \mapsto q^T_{c,n}$, using penalty function p_1, $q_{\text{max}} = 19$.

AR loadings

(d1)



Another example: AR loadings,

$q = 3$, $n = T = 300$, $M_T = [0.75\sqrt{T}]$. Simultaneous plots of $c \mapsto S_c$ and $c \mapsto q^T_{c,n}$, using penalty function p_1, $q_{\text{max}} = 19$.

heuristics of the $c \mapsto \hat{q}_c$ behavior 1/3

- for simplicity, assume again that $\mathbf{\Sigma}(\theta)$ is known
- $\hat{q}_c^n$ is characterized by the fact that
  - $$\frac{1}{(k-\hat{q})} \sum_{j=\hat{q}+1}^{k} \int \lambda_j^{(n)}(\theta) d\theta < c\, np(n) \quad k = \hat{q}+1, \hat{q}+2, \ldots,$$

  that is, $\frac{1}{\ell} \sum_{j=\hat{q}+1}^{\hat{q}+\ell} \int \lambda_j^{(n)}(\theta) d\theta < c\, np(n)$ for all $\ell = 1, 2, \ldots,$
  which holds iff it holds for $\ell = 1$

  and
  - $$\frac{1}{(\hat{q}-k)} \sum_{j=k+1}^{\hat{q}} \int \lambda_j^{(n)}(\theta) d\theta > c\, np(n) \quad k = \hat{q}-1, \hat{q}-2, \ldots$$

  that is, $\frac{1}{\ell+1} \sum_{j=\hat{q}-\ell}^{\hat{q}} \int \lambda_j^{(n)}(\theta) d\theta > c\, np(n)$ for all $\ell = 0, 1, 2, \ldots,$
  which holds iff it holds for $\ell = 0$

heuristics of the $c \mapsto \hat{q}_c$ behavior 1/3

- for simplicity, assume again that $\mathbf{\Sigma}(\theta)$ is known
- $\hat{q}_c^n$ is characterized by the fact that
  - $$\frac{1}{(k-\hat{q})} \sum_{j=\hat{q}+1}^{k} \int \lambda_j^{(n)}(\theta) d\theta < c\, np(n) \quad k = \hat{q}+1, \hat{q}+2, \ldots,$$

  that is, $\frac{1}{\ell} \sum_{j=\hat{q}+1}^{\hat{q}+\ell} \int \lambda_j^{(n)}(\theta) d\theta < c\, np(n)$ for all $\ell = 1, 2, \ldots,$
  which holds iff it holds for $\ell = 1$

  and

  - $$\frac{1}{(\hat{q}-k)} \sum_{j=k+1}^{\hat{q}} \int \lambda_j^{(n)}(\theta) d\theta > c\, np(n) \quad k = \hat{q}-1, \hat{q}-2, \ldots$$

  that is, $\frac{1}{\ell+1} \sum_{j=\hat{q}-\ell}^{\hat{q}} \int \lambda_j^{(n)}(\theta) d\theta > c\, np(n)$ for all $\ell = 0, 1, 2, \ldots,$
  which holds iff it holds for $\ell = 0$

heuristics of the $c \mapsto \hat{q}_c$ *behavior 2/3*

- $\hat{q}$ thus is fully identified by

$$\frac{1}{n}\int \lambda_{\hat{q}+1}{}^{(n)}(\theta)d\theta < c\,p(n) < \quad \frac{1}{n}\int \lambda_{\hat{q}}{}^{(n)}(\theta)d\theta$$

heuristics of the $c \mapsto \hat{q}_c$ behavior 2/3

- $\hat{q}$ thus is fully identified by

$$\frac{1}{n}\int \lambda_{\hat{q}+1}{}^{(n)}(\theta)d\theta < c\,p(n) < \quad \frac{1}{n}\int \lambda_{\hat{q}}{}^{(n)}(\theta)d\theta$$

- that is, once c is chosen, the criterion identifies the number of factors as the (unique, in view of monotonocity in c) $\hat{q}_c^n$ such that

$$c\,p(n) \text{ “separates” } \frac{1}{n}\int \lambda_{\hat{q}+1}{}^{(n)}(\theta)d\theta \quad \text{and} \quad \frac{1}{n}\int \lambda_{\hat{q}}{}^{(n)}(\theta)d\theta$$

heuristics of the $c \mapsto \hat{q}_c$ behavior 2/3

- $\hat{q}$ thus is fully identified by

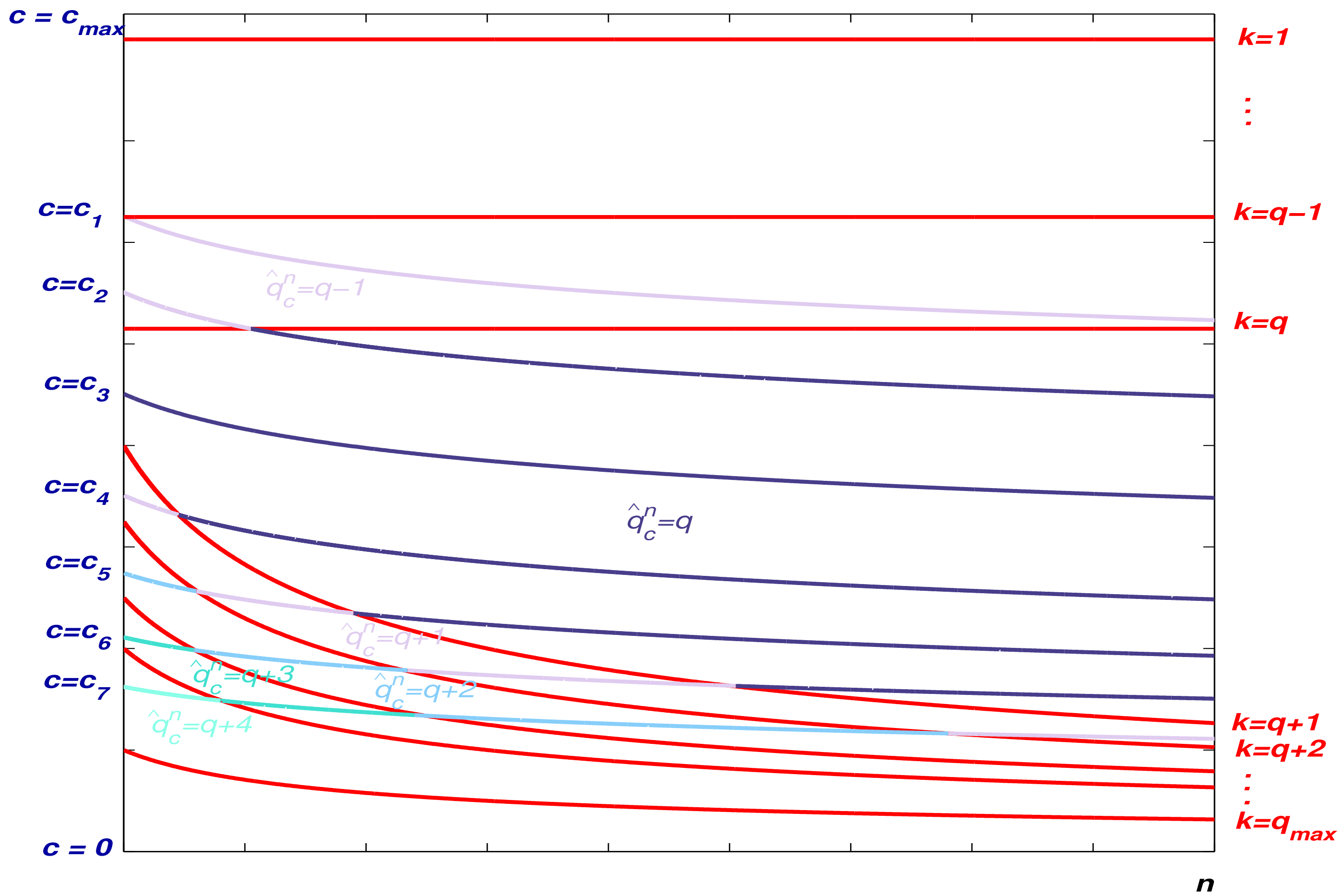
$$\frac{1}{n}\int \lambda_{\hat{q}+1}{}^{(n)}(\theta)d\theta < c\,p(n) < \quad \frac{1}{n}\int \lambda_{\hat{q}}{}^{(n)}(\theta)d\theta$$

- that is, once c is chosen, the criterion identifies the number of factors as the (unique, in view of monotonocity in c) $\hat{q}_c^n$ such that

$$c\,p(n) \text{ "separates" } \frac{1}{n}\int \lambda_{\hat{q}+1}{}^{(n)}(\theta)d\theta \quad \text{and} \quad \frac{1}{n}\int \lambda_{\hat{q}}{}^{(n)}(\theta)d\theta$$

- graphical evidence ...

heuristics of the $c \mapsto \hat{q}_c$ behavior 3/3



Heuristic behavior of $c \mapsto S_c$. Graphs of $n \mapsto \frac{1}{n}\int \lambda_k^{(n)}(\theta)d\theta$, $k = 1, \ldots, q_{\text{max}}$ (in red: $O(n^{-1})$ for $k > q$, ~constant for $k \leq q$), and $n \mapsto cp(n)$, $c = 0, \ldots, c_{\text{max}}$ (in blue: go to zero at rate slower than n^{-1}), along with the corresponding $\hat{q}_c^{(n)}$'s. Note that $S_{c_1} = S_{c_3} = 0$, whereas $S_{c_2}, S_{c_4}, \ldots, S_{c_7}$ are strictly positive.

4. An application

A panel of $n = 132$ monthly time series observed from January 1960 through December 2003 (T = 528).

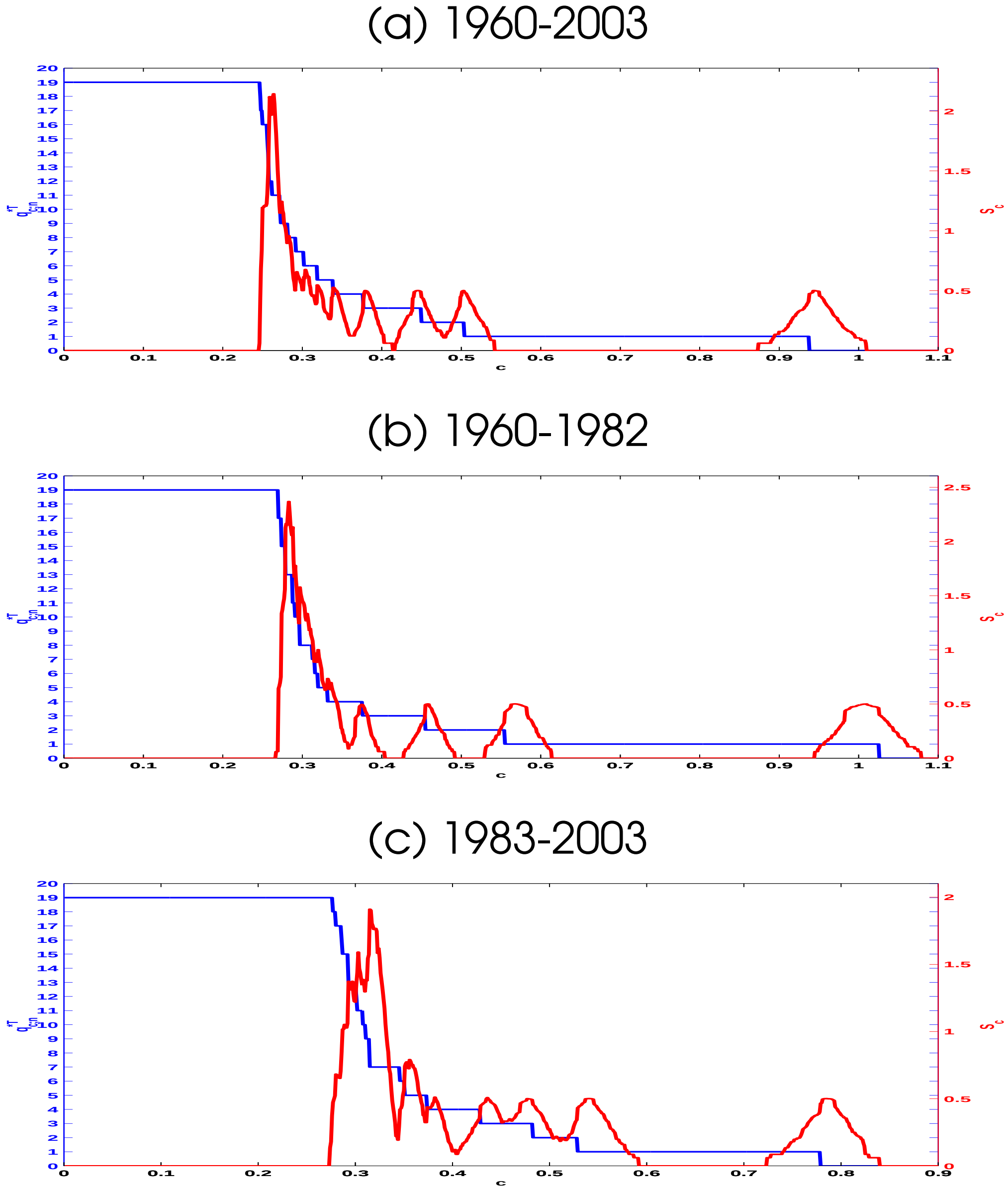
These series are considered by economists to be a representative summary of the U.S. economy and have been studied by Stock and Watson (2005), Giannone et al. (2005a,b), and Bai and Ng (2005) with surprisingly divergent conclusions

Stock and Watson found 7 static factors

Giannone et al. (2005a,b) (based on a different methodology and restricting to a carefully selected subset of 12 series) arrived at 2 factors

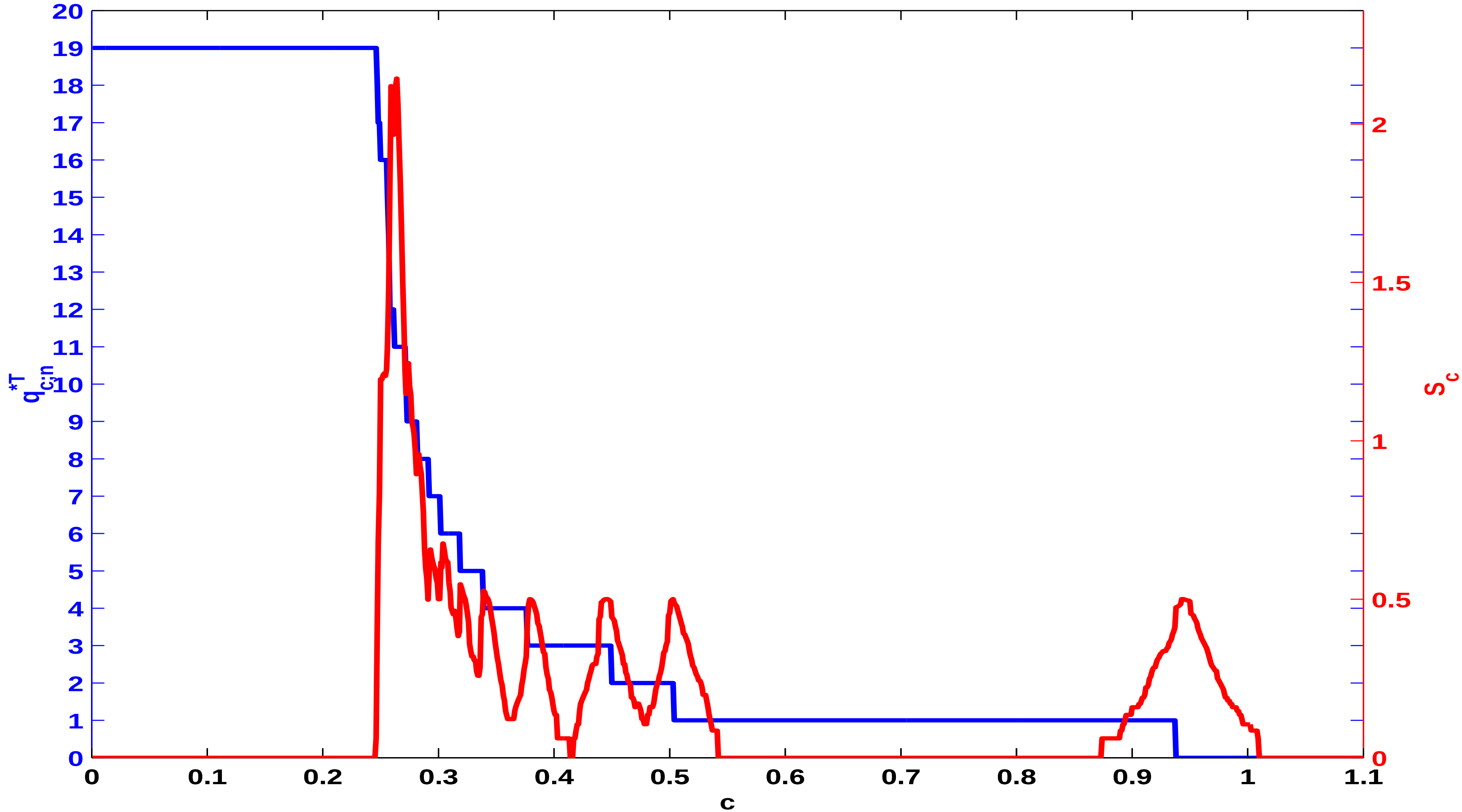
Bai and Ng do not give a clear-cut conclusion; in an early version of their article, they mention up to 10 static factors, and 7 dynamic ones, but in their final version, they conclude in favor of 4 dynamic factors spanning 7 static ones

Much instability, thus—either due to the fact that the assumptions of the model considered (which is not the general dynamic factor model) do not hold, or to the fact that they (Bai and Ng) "put $c = 1$", or both.



An analysis of the US Economy dataset (1960-2003). Simultaneous plots of $c \mapsto S_c$ and $c \mapsto q^T_{c,n}$, using penalty function p_1, $q$max $= 19$, and $IC^T_{2;n}$ criterion, over the periods 1960-2003, 1960-1982, and 1983-2003.

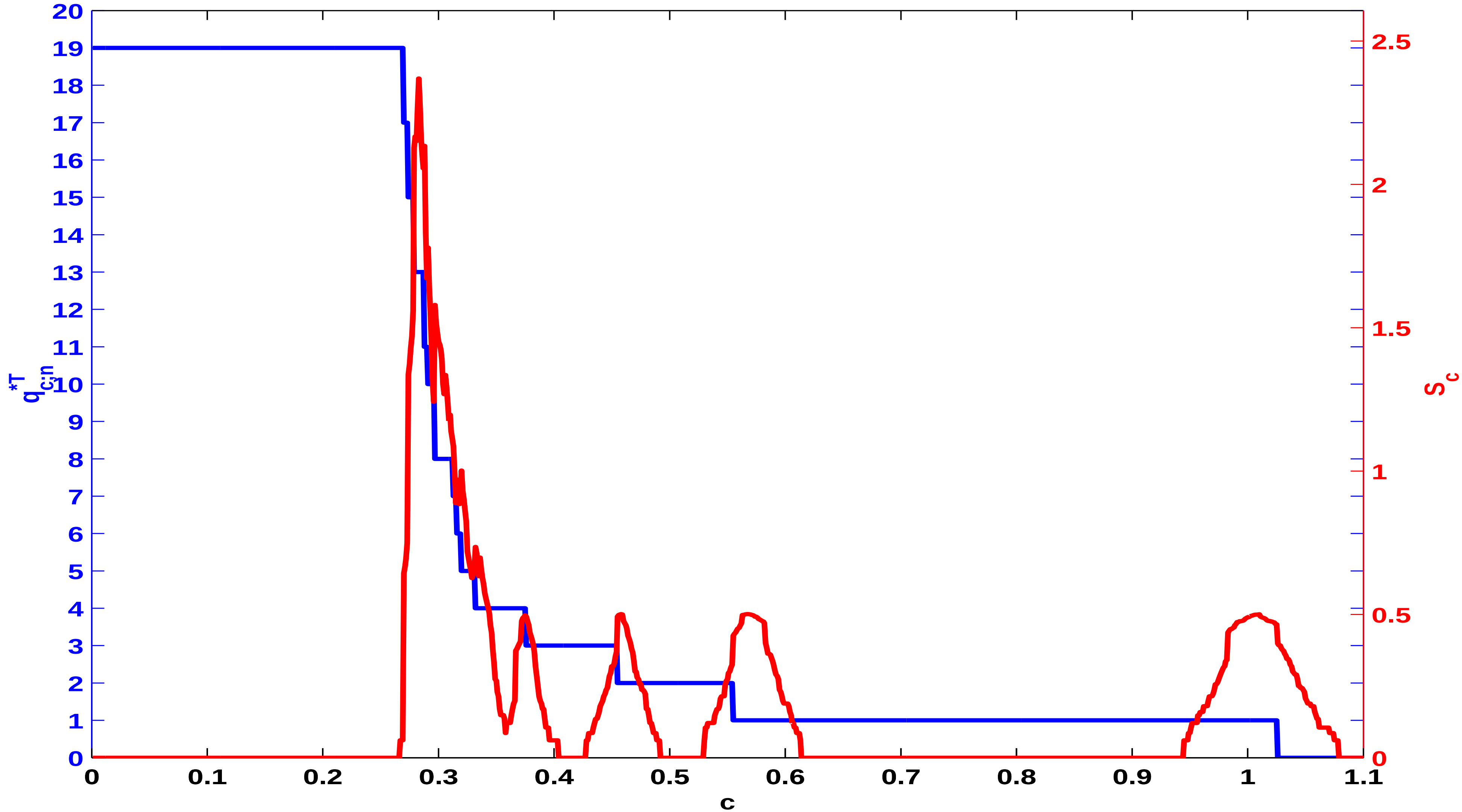
(a) 1960-2003



An analysis of the US Economy dataset (1960-2003). Simultaneous plots of $c \mapsto S_c$ and $c \mapsto q^T_{c,n}$, using penalty function p_1, $q_{\text{max}} = 19$, and $IC^T_{2;n}$ criterion, over the period 1960-2003.

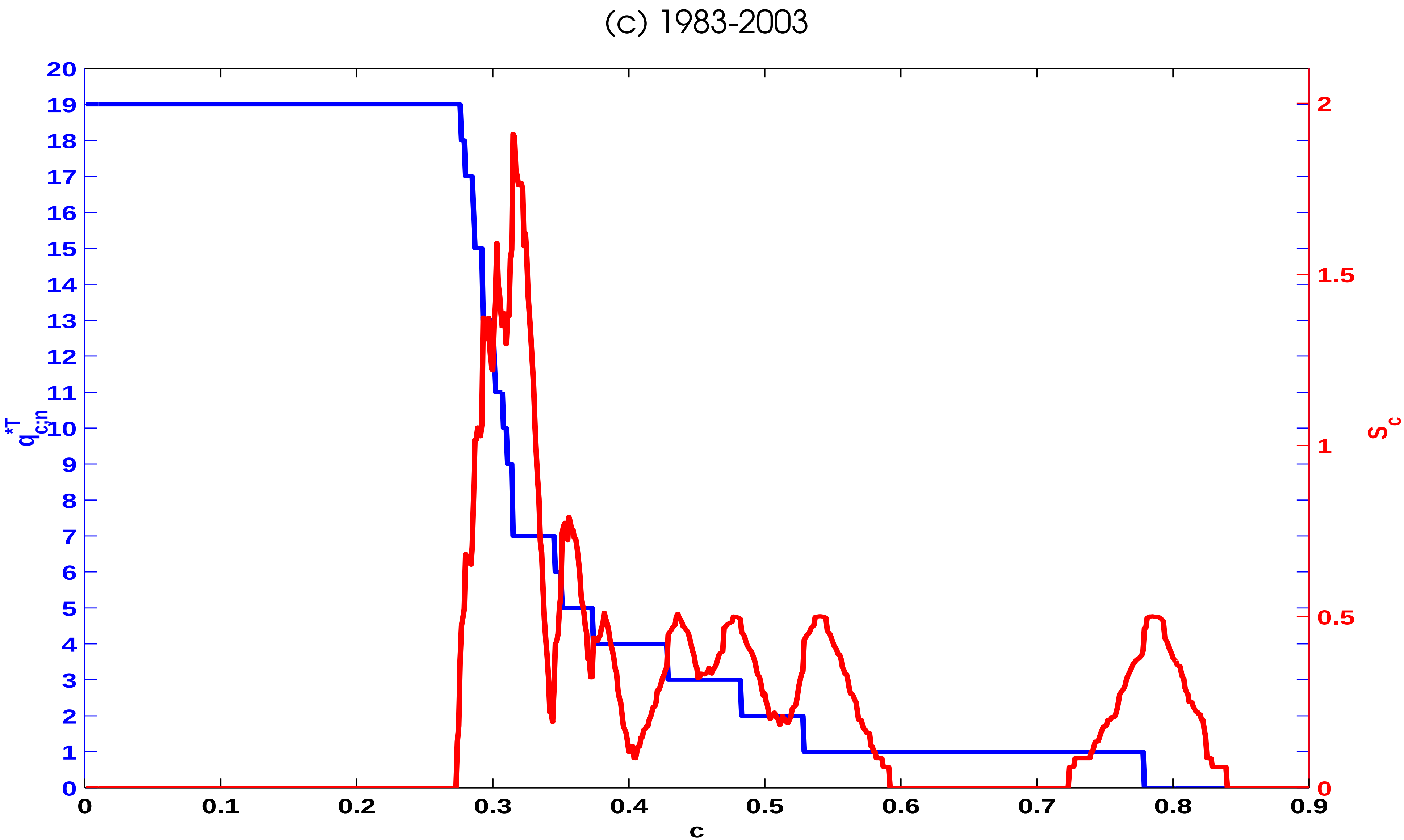
$q = 3$? 1? The conclusion is not very clear ...

(b) 1960-1982



An analysis of the US Economy dataset (1960-2003). Simultaneous plots of $c \mapsto S_c$ and $c \mapsto q^T_{c,n}$, using penalty function p_1, $q_{\text{max}} = 19$, and $IC^T_{2;n}$ criterion, over the period 1960-1982.

$q = 3$ clearly emerges



An analysis of the US Economy dataset (1960-2003). Simultaneous plots of $c \mapsto S_c$ and $c \mapsto q^T_{c,n}$, using penalty function p_1, $q_{\max} = 19$, and $IC^T_{2;n}$ criterion, over the period 1983-2003.

the conclusion is very clearly $q = 1$.

Returns and Volatilities in High Dimension
a General Factor Model Approach

3. Dynamic Factors in the Presence of Blocks

Marc Hallin

ECARES, Université Libre de Bruxelles

Hermann Otto Hirschfeld Lecture Series 2017

Berlin, 20–21 November 2017

based on

Hallin, M. and R. Liška (2011) Dynamic factors in the presence of block structure. *Journal of Econometrics* **163**, 29-41.

Hallin, M., CH. Mathias, H. Pirotte, and D. Veredas (2011) Market liquidity as dynamic factors. *Journal of Econometrics* **163**, 42-50.

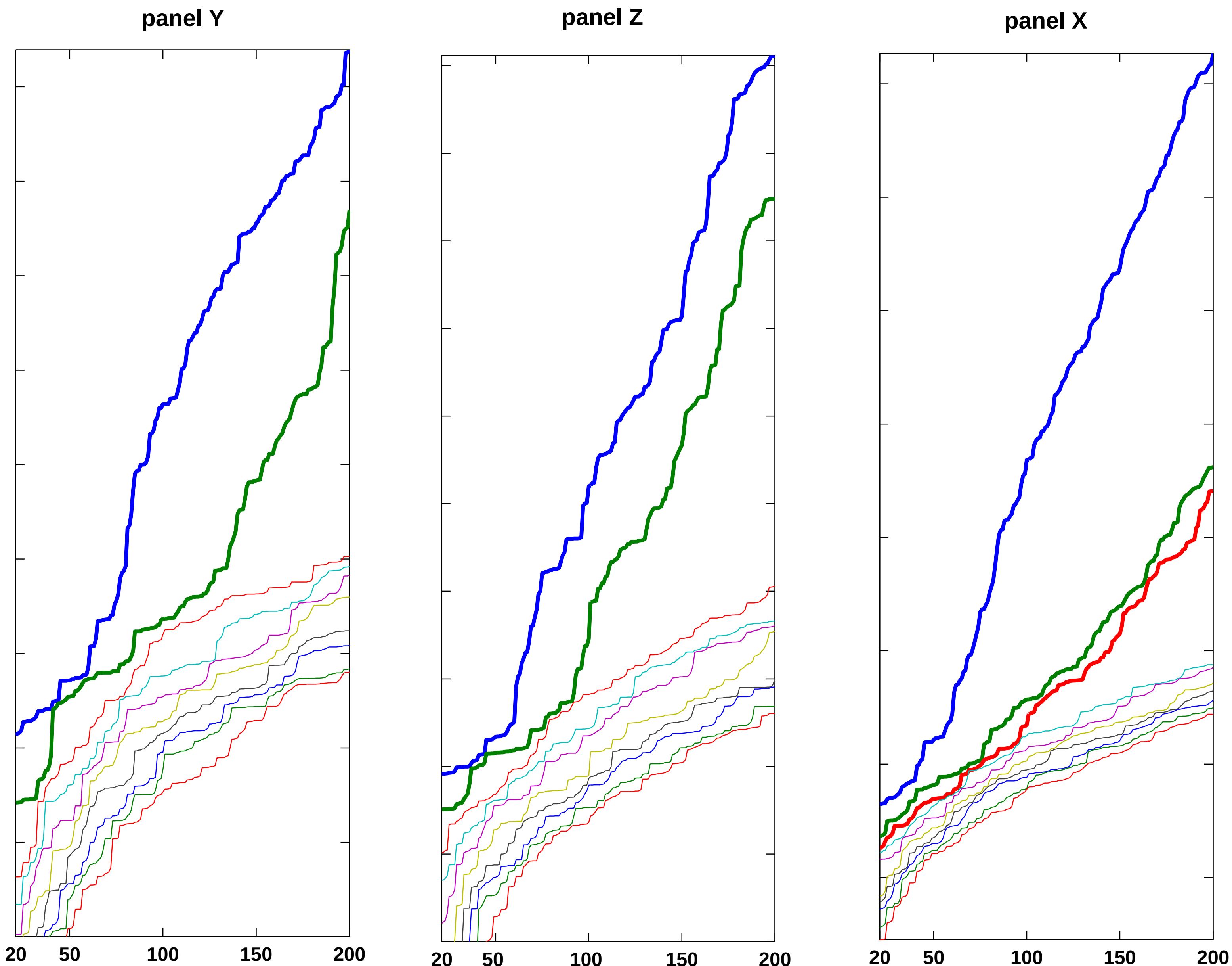
- What happens in the presence of blocks (a panel consisting of two or several subpanels)?

example:

(a) a panel of n_F=96 French economic series $\{X_{it}^F\}$
(b) a panel of n_G=114 German economic series $\{X_{jt}^G\}$
(c) the joint panel (n= 210) $\{X_{it}^F,\ X_{jt}^G\}$

Subpanel spectral eigenvalues (two subpanels)

Behavior of 10 largest dynamic eigenvalues (averaged over frequencies):
(a) France; (b) Germany; (c) France and Germany.



Panels with two blocks

Three distinct analyses can be conducted, based on

- two *marginal factor models*, with q_F and q_G common shocks, respectively

$$
\begin{aligned}
X_{it}^F &= \chi_{it}^F + \xi_{it}^F \\
X_{jt}^G &= \chi_{jt}^G + \xi_{jt}^G
\end{aligned}
$$

- a *global factor model*, with q common shocks

$$
\begin{aligned}
X_{it}^F &= \chi_{it}^{FG} + \xi_{it}^{FG} \\
X_{jt}^G &= \chi_{jt}^{FG} + \xi_{jt}^{FG}
\end{aligned}
$$

This provides three decompositions of the whole Hilbert space $\mathcal{H}$ spanned by the panel into

- an F-common space $\mathcal{H}_F^\chi$ and an F-idiosyncratic space $\mathcal{H}_F^\xi := (\mathcal{H}_F^\chi)^\perp$
- a G-common space $\mathcal{H}_G^\chi$ and a G-idiosyncratic space $\mathcal{H}_G^\xi := (\mathcal{H}_G^\chi)^\perp$
- an FG-common space $\mathcal{H}_{FG}^\chi$ and an FG-idiosyncratic space $\mathcal{H}_{FG}^\xi := (\mathcal{H}_{FG}^\chi)^\perp$

Panels with two blocks

Clearly,

$$\mathcal{H}_F^\chi \subseteq \mathcal{H}_{FG}^\chi \quad \text{and} \quad \mathcal{H}_G^\chi \subseteq \mathcal{H}_{FG}^\chi$$

so that

$$\max(q_F, q_G) \leq q \leq q_F + q_G.$$

Panels with two blocks

More formally, consider two double-indexed sequences

$$\mathbf{Y} := \{Y_{it}, i \in \mathbb{N}, t \in \mathbb{Z}\}$$

and

$$\mathbf{Z} := \{Z_{jt}, j \in \mathbb{N}, t \in \mathbb{Z}\},$$

where t stands for time and i, j are cross-sectional indices, of observed random variables.

Let $\mathbf{Y}_{n_y} := \{\mathbf{Y}_{n_y,t}, t \in \mathbb{Z}\}$ and $\mathbf{Z}_{n_z} := \{\mathbf{Z}_{n_z,t}, t \in \mathbb{Z}\}$ be the n_y- and n_z-dimensional subprocesses of $\mathbf{Y}$ and $\mathbf{Z}$, respectively, where $\mathbf{Y}_{n_y,t} := (Y_{1t} \ \ldots, Y_{n_y t})'$ and $\mathbf{Z}_{n_z,t} := (Z_{1t} \ \ldots, Z_{n_z t})'$, and write

$$\mathbf{X}_{\mathbf{n},t} := (Y_{1t} \ \ldots, Y_{n_y t}, Z_{1t} \ \ldots, Z_{n_z t})' := (\mathbf{Y}'_{n_y,t} \, \mathbf{Z}'_{n_z,t})'$$

with $\mathbf{n} := (n_y, n_z)$ and $n := n_y + n_z$.

Panels with two blocks

We throughout assume that $\mathbf{X}$ satisfies our assumptions A1, A2, and A3. All stochastic variables in this paper belong to the Hilbert space $L_2(\Omega, \mathcal{F}, \mathrm{P})$, where $(\Omega, \mathcal{F}, \mathrm{P})$ is some given probability space.

The Hilbert subspaces spanned by the processes $\mathbf{Y}$, $\mathbf{Z}$ and $\mathbf{X}$ are denoted by $\mathcal{H}_y$, $\mathcal{H}_z$ and $\mathcal{H}$, respectively.

Panels with two blocks

- Denoting by $\mathbf{\Sigma}_{y;n_y}(\theta)$ and $\mathbf{\Sigma}_{z;n_z}(\theta)$ the $(n_y \times n_y)$ and $(n_z \times n_z)$ spectral density matrices of $\mathbf{Y}_{n_y,t}$ and $\mathbf{Z}_{n_z,t}$, and by $\mathbf{\Sigma}_{yz;\mathbf{n}}(\theta) = \mathbf{\Sigma}^*_{zy;\mathbf{n}}(\theta)$ their $(n_y \times n_z)$ cross-spectrum matrix, write

$$\mathbf{\Sigma}_{\mathbf{n}}(\theta) =: \begin{pmatrix} \mathbf{\Sigma}_{y;n_y}(\theta) & \mathbf{\Sigma}_{yz;\mathbf{n}}(\theta) \\ \mathbf{\Sigma}_{zy;\mathbf{n}}(\theta) & \mathbf{\Sigma}_{z;n_z}(\theta) \end{pmatrix}, \quad \theta \in [-\pi, \pi]$$

for the $(n \times n)$ spectral density matrix of $\mathbf{X}_{\mathbf{n},t}$, with elements $\sigma_{i_1 i_2}(\theta)$, $\sigma_{j_1 j_2}(\theta)$ or $\sigma_{k_1 k_2}(\theta),\ k_1,\ k_2 = 1, \ldots, n,\ i_1,\ i_2 = 1, \ldots, n_y,\ j_1,\ j_2 = 1, \ldots, n_z.$

- For any $\theta \in [-\pi, \pi]$, let $\lambda_{y;n_y,i}(\theta)$ be $\mathbf{\Sigma}_{y;n_y}(\theta)$'s i-th eigenvalue (in decreasing order of magnitude). The function $\theta \mapsto \lambda_{y;n_y,i}(\theta)$ is called $\mathbf{\Sigma}_{y;n_y}(\theta)$'s i-th *dynamic eigenvalue*.

- The notation $\theta \mapsto \lambda_{z;n_z,j}(\theta)$ and $\theta \mapsto \lambda_{\mathbf{n},k}(\theta)$ is used in an obvious way for the dynamic eigenvalues of $\mathbf{\Sigma}_{z;n_z}(\theta)$ and $\mathbf{\Sigma}_{\mathbf{n}}(\theta)$, respectively.

- The corresponding *dynamic eigenvectors*, of dimensions $(n_y \times 1)$, $(n_z \times 1)$, and $(n \times 1)$, are denoted by $\mathbf{p}_{y;n_y,i}(\theta)$, $\mathbf{p}_{z;n_z,j}(\theta)$, and $\mathbf{p}_{\mathbf{n},k}(\theta)$, respectively.

Panels with two blocks

Note on the asymptotics.

Let n_y, n_z, $n = n_y + n_z$. When considering asymptotics, both n_y and n_z tend to infinity. But the joint panel $\mathbf{X_n}$ should not be taken as piling a block of Y's on top of a block of Z's. Rather, for each n, the n_y and n_z cross-sectional indices should be assigned "at random" to the two blocks. Such allocation clearly does not perturb spectral eigenvalues etc.

Panels with two blocks

Under Assumptions A1-A3, those subpanel dynamic eigenvalues have the typical asymptotic behavior:

For some $q_y, q_z \in \mathbb{N}$,

(i) the q_y*-th dynamic eigenvalue of* $\mathbf{\Sigma}_{y;n_y}(\theta)$, $\lambda_{y;n_y,q_y}(\theta)$, *diverges as* $n_y \to \infty$, *a.e. in* $[-\pi, \pi]$, *while the* $(q_y + 1)$*-th one,* $\lambda_{y;n_y,q_y+1}(\theta)$, *is* θ*-a.e. bounded;*

(ii) the q_z*-th dynamic eigenvalue of* $\mathbf{\Sigma}_{z;n_z}(\theta)$, $\lambda_{z;n_z,q_z}(\theta)$, *diverges as* $n_z \to \infty$, *a.e. in* $[-\pi, \pi]$, *while the* $(q_z + 1)$*-th one,* $\lambda_{z;n_z,q_z+1}(\theta)$, *is* θ*-a.e. bounded.*

That behavior entails a similar behavior for the dynamic eigenvalues $\lambda_{\mathbf{n},k}(\theta)$ of $\mathbf{\Sigma}_{\mathbf{n}}(\theta)$ of the joint panel.

For some $q \in \mathbb{N}$, *with* $\max(q_y, q_z) \leq q \leq q_y + q_z$,

(iii) $\mathbf{\Sigma}_{\mathbf{n}}(\theta)$*'s* q*-th dynamic eigenvalue* $\lambda_{\mathbf{n},q}(\theta)$ *diverges as* $\min(n_y, n_z) \to \infty$, *a.e. in* $[-\pi, \pi]$, *while the* $(q + 1)$*-th one,* $\lambda_{\mathbf{n},q+1}(\theta)$, *is* θ*-a.e. bounded.*

Panels with two blocks

We then have the following factor model representations.

(a) $\mathbf{Y}$ has a dynamic factor representation (q_y shocks—call them the $y-$fcommon shocks, spanning the y-common space $\mathcal{H}_y^{\chi}$, with orthogonal complement $\mathcal{H}_y^{\xi}$)

$$Y_{it} = \chi_{y;it} + \xi_{y;it} = \sum_{l=1}^{q_y} b_{y;il}(L)u_{y;lt} + \xi_{y;it}\ , \quad i \in \mathbb{N},\ t \in \mathbb{Z};$$

(b) $\mathbf{Z}$ has a dynamic factor representation (q_z shocks—call them the $z-$common shocks, spanning the z-common space $\mathcal{H}_z^{\chi}$, with orthogonal complement $\mathcal{H}_z^{\xi}$)

$$Z_{jt} = \chi_{z;jt} + \xi_{z;jt} = \sum_{l=1}^{q_z} b_{z;jl}(L)u_{z;lt} + \xi_{z;jt}\ , \quad j \in \mathbb{N},\ t \in \mathbb{Z};$$

(c) $\mathbf{X}$ has a dynamic factor representation (q shocks—call them the *joint* common shocks, spanning the joint common space $\mathcal{H}_x^{\chi}$, with orthogonal complement $\mathcal{H}_x^{\xi}$)

$$X_{kt} = \begin{cases} Y_{it} & = \chi_{xy;it} + \xi_{xy;it} = \sum_{l=1}^{q} b_{xy;il}(L)u_{lt} + \xi_{xy;it}, \quad k \in \mathbb{N},\ t \in \mathbb{Z} \qquad \text{if } X_{kt} = Y_{it} \\ Z_{jt} & = \chi_{xz;jt} + \xi_{xz;jt} = \sum_{l=1}^{q} b_{xz;jl}(L)u_{lt} + \xi_{xz;jt}, \quad k \in \mathbb{N},\ t \in \mathbb{Z} \qquad \text{if } X_{kt} = Z_{jt} \end{cases}$$

All filters involved have square-summable coefficients.

Panels with two blocks

Letting $\mathcal{H}^{\chi}_{F\cap G} := \mathcal{H}^{\chi}_{F} \bigcap \mathcal{H}^{\chi}_{G}$,

- either $\mathcal{H}^{\chi}_{F\cap G} = \{0\}$, and $q_F + q_G = q$
- or $\mathcal{H}^{\chi}_{F\cap G} \neq \{0\}$, and it is driven by $q_{F\cap G} := q_F + q_G - q$ mutually orthogonal white noises

The elements of $\mathcal{H}^{\chi}_{F\cap G}$ being both F- and G-common, can be called "strongly common". Each common component decomposes into a "strongly common component ϕ" (its projection onto $\mathcal{H}^{\chi}_{F\cap G}$) and a residual ψ, which we call "weakly common":

$$\chi^F_{it} = \phi_{F;it} + \psi_{F;it} \quad \text{and} \quad \chi^G_{jt} = \phi_{G;jt} + \psi_{G;jt}.$$

Panels with two blocks

Similarly,

$$(\mathcal{H}^{\chi}_{FG})^{\perp} \subseteq (\mathcal{H}^{\chi}_{F})^{\perp} \quad \text{and} \quad (\mathcal{H}^{\chi}_{FG})^{\perp} \subseteq (\mathcal{H}^{\chi}_{G})^{\perp}.$$

Being both F- and G- idiosyncratic, the space $(\mathcal{H}^{\chi}_{FG})^{\perp}$ can be called "strongly idiosyncratic". Each marginally idiosyncratic component decomposes into a "strongly idiosyncratic component ξ" and a residual ν, which we call "weakly idiosyncratic":

$$\xi^{F}_{it} = \nu_{F;it} + \xi^{FG}_{it} \quad \text{and} \quad \xi^{G}_{jt} = \nu_{G;jt} + \xi^{FG}_{jt}.$$

Panels with two blocks

We thus have two decompositions into four mutually orthogonal components:

$$X^F_{it} = \underbrace{\overbrace{\phi_{F;it} + \psi_{F;it}}^{}}_{\chi^F_{it}}\! $$

$$X^F_{it} = \overbrace{\underbrace{\phi_{F;it} + \psi_{F;it}}_{\chi^F_{it}} + \nu_{F;it}}^{\chi^{FG}_{it}} + \xi^{FG}_{it} \quad \text{and} \quad X^G_{jt} = \overbrace{\underbrace{\phi_{G;jt} + \psi_{G;jt}}_{\chi^G_{jt}} + \nu_{G;jt}}^{\chi^{FG}_{jt}} + \xi^{FG}_{jt}, \quad i, j \in \mathbb{N},\ t \in \mathbb{Z}.$$

with $\xi^F_{it} = \nu_{F;it} + \xi^{FG}_{it}$ and $\xi^G_{jt} = \nu_{G;jt} + \xi^{FG}_{jt}$ marked by underbraces.

Panels with two blocks

More generally, with obvious notation, we obtain the decompositions

$$Y_{it} = \overbrace{\underbrace{\phi_{y;it} + \psi_{y;it}}_{\chi_{y;it}} + \nu_{y;it}}^{\chi_{xy;it}} + \xi_{xy;it} \quad \text{and} \quad Z_{jt} = \overbrace{\underbrace{\phi_{z;jt} + \psi_{z;jt}}_{\chi_{z;jt}} + \nu_{z;jt}}^{\chi_{xz;jt}} + \xi_{xz;jt}\,, \quad i,j \in \mathbb{N},\ t \in \mathbb{Z}.$$

of the original observations into four mutually orthogonal components.

These decompositions induce additive decompositions of the variances of the observations into a sum of four terms indicating the relative contributions of each component.

identification of factor numbers

Before turning to a data-driven reconstruction of those decompositions, the dynamic dimensions q_F, q_G and q are to be estimated via the Hallin and Liška method, the application and conclusions of which we briefly summarize.

Consistency of the Hallin and Liška method requires the additional assumption

ASSUMPTION 4. Linear divergence of dynamic eigenvalues: letting $n := n_F + n_G$ and $\mathbf{n} := (n_F, n_G)$, $\lambda_{\mathbf{n}q}(\theta)$ is $O(n)$ (and not $o(n)$) as $\mathbf{n} \to \infty$ (meaning both n_F and $n_G \to \infty$)

Identification of factor numbers

- The lag window method (Bartlett lag window of size M_T) provides estimations $\boldsymbol{\Sigma}_{\mathbf{n}}^T(\theta_l)$ of $\boldsymbol{\Sigma}_{\mathbf{n}}(\theta)$ at frequencies $\theta_l := \pi l/(M_T + 1/2)$ for $l = -M_T, \ldots, M_T$.

- Based on these estimations, consider the information criterion

$$IC_{\mathbf{n};c}^T(k) := \log\left[\frac{1}{n-k}\sum_{i=k+1}^{n}\frac{1}{2M_T+1}\sum_{l=-M_T}^{M_T}\lambda_{\mathbf{n}i}^T(\theta_l)\right] + kcp(n,T), \quad 0 \le k \le q_{\max}, \quad c \in \mathbb{R}_0^+,$$

where $p(n,T)$ is $o(1)$ and $p^{-1}(n,T) = o\left(\min(n, M_T^2, M_T^{-1/2}T^{1/2})\right)$ as both n and T tend to infinity

$q_{\max}$ is some predetermined upper bound; $c > 0$ is arbitrary

(eigenvalues $\lambda_{\mathbf{n}i}^T(\theta_l)$ are those of $\boldsymbol{\Sigma}_{\mathbf{n}}^T(\theta_l)$)

Identification of factor numbers

- The lag window method (Bartlett lag window of size M_T) provides estimations $\mathbf{\Sigma}^T_{\mathbf{n}}(\theta_l)$ of $\mathbf{\Sigma}_{\mathbf{n}}(\theta)$ at frequencies $\theta_l := \pi l/(M_T + 1/2)$ for $l = -M_T, \ldots, M_T$.

- Based on these estimations, consider the information criterion

$$IC^T_{\mathbf{n};c}(k) := \log \left[\frac{1}{n-k} \sum_{i=k+1}^{n} \frac{1}{2M_T+1} \sum_{l=-M_T}^{M_T} \lambda^T_{\mathbf{n}i}(\theta_l) \right] + kcp(n,T), \quad 0 \leq k \leq q_{\max}, \quad c \in \mathbb{R}_0^+,$$

where $p(n,T)$ is $o(1)$ and $p^{-1}(n,T) = o\left(\min(n, M_T^2, M_T^{-1/2}T^{1/2})\right)$ as both n and T tend to infinity

$q_{\max}$ is some predetermined upper bound; $c > 0$ is arbitrary

(eigenvalues $\lambda^T_{\mathbf{n}i}(\theta_l)$ are those of $\mathbf{\Sigma}^T_{\mathbf{n}}(\theta_l)$)

For any $c > 0$, a consistent identification of the number of factors is

$$q^T_{\mathbf{n};c} := \mathrm{argmin}_{0 \leq k \leq q_{\max}} IC^T_{\mathbf{n};c}(k).$$

- but of course, $q^T_{\mathbf{n};c}$ heavily depends on c!!

The Hallin-Liška procedure

- Consider a $J-$tuple $q_{c,\mathbf{n}_j}^{T_j}$, $j = 1, \ldots, J$, where $\mathbf{n}_j = (n_{y;j}, n_{z;j})$ with $0 < n_{y;1} < \ldots < n_{y;J} = n_y$, $0 < n_{z;1} < \ldots < n_{z;J} = n_z$,
and $0 < T_1 \leq \ldots \leq T_J = T$ (possibly, $T_1 = T_J = T$)

and the corresponding "history" $q_{c;\mathbf{n}_j}^{T_j}$, $j = 1, \ldots, J$ of the selection.

- The procedure is based on a joint plot of $c \mapsto S_c$, where

$$S_c^2 := J^{-1} \sum_{j=1}^{J} (q_{\mathbf{n}_j;c}^{T_j} - J^{-1} \sum_{j=1}^{J} q_{\mathbf{n}_j;c}^{T_j})^2$$

and $c \mapsto \hat{q}_{c;n}^{T}$ (the "final" identification for given c).

The Hallin-Liška procedure

- Consider a $J-$tuple $q_{c,\mathbf{n}_j}^{T_j}$, $j = 1, \ldots, J$, where $\mathbf{n}_j = (n_{y;j}, n_{z;j})$ with $0 < n_{y;1} < \ldots < n_{y;J} = n_y$, $0 < n_{z;1} < \ldots < n_{z;J} = n_z$, and $0 < T_1 \leq \ldots \leq T_J = T$ (possibly, $T_1 = T_J = T$)

and the corresponding "history" $q_{c;\mathbf{n}_j}^{T_j}$, $j = 1, \ldots, J$ of the selection.

- The procedure is based on a joint plot of $c \mapsto S_c$, where

$$S_c^2 := J^{-1} \sum_{j=1}^{J} (q_{\mathbf{n}_j;c}^{T_j} - J^{-1} \sum_{j=1}^{J} q_{\mathbf{n}_j;c}^{T_j})^2$$

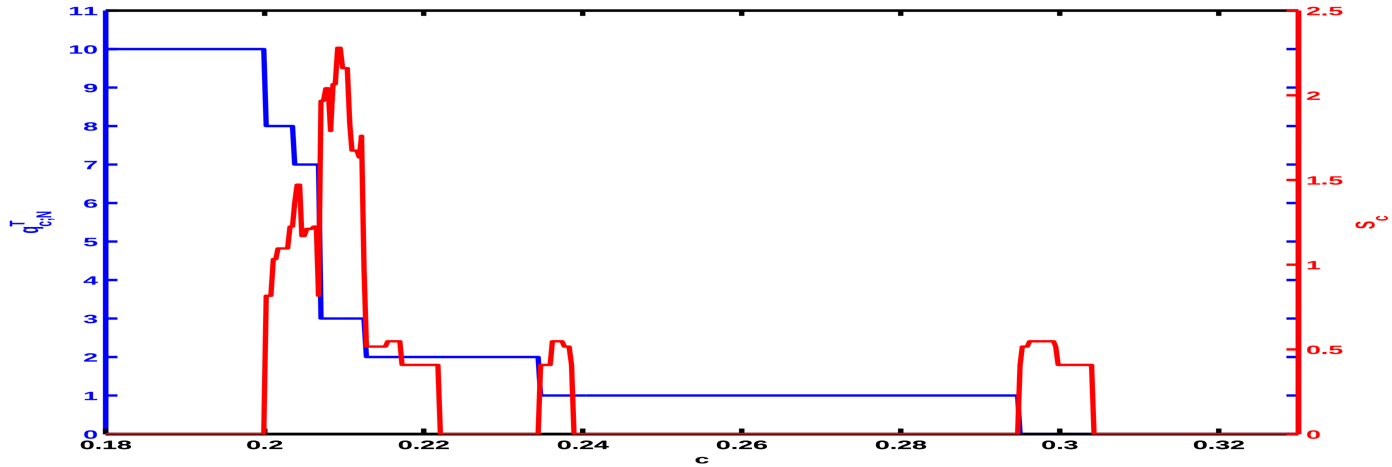
and $c \mapsto \hat{q}_{c;n}^{T}$ (the "final" identification for given c).

The procedure consists in picking the value of $\hat{q}_{c;n}^{T}$ associated with the second stability interval of $c \mapsto S_c$

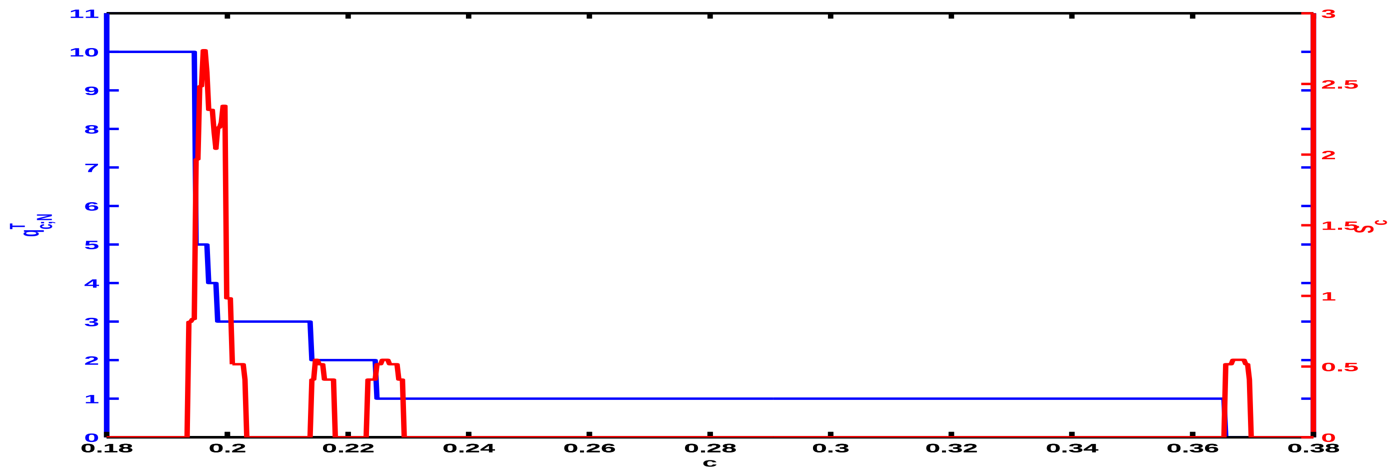
- The Hallin-Liška method is applied to the three estimated spectral density matrices, yielding estimated values of the numbers q, q_F, q_G and $q_{F\cap G} = q_F + q_G - q$ of factors

France and Germany

(a) France

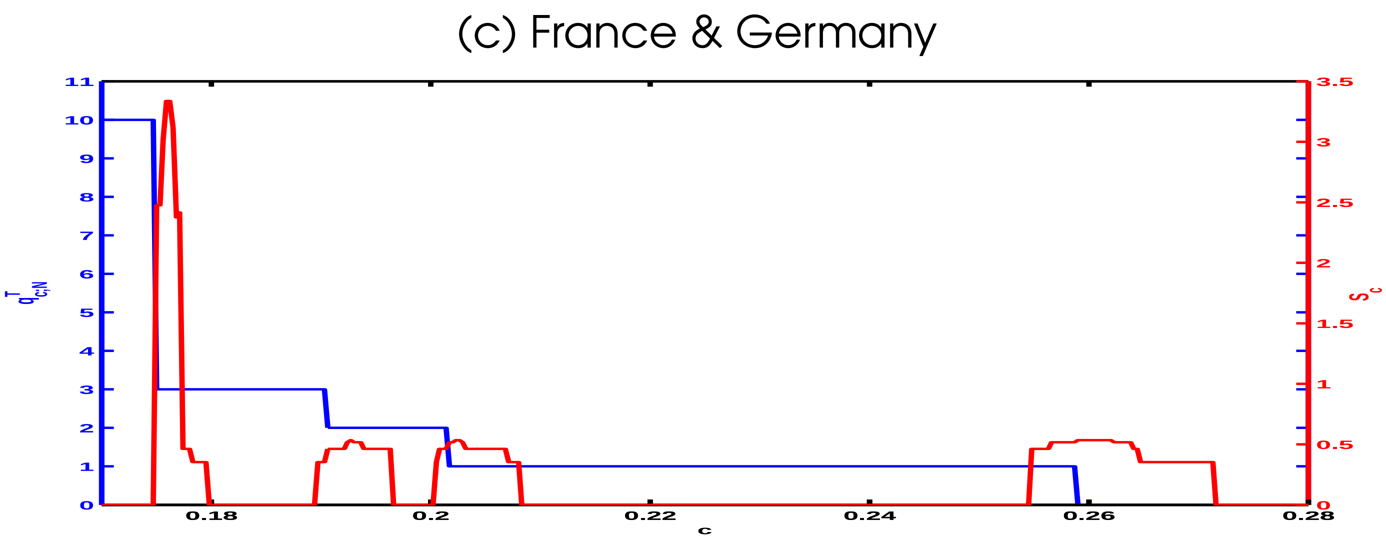


(b) Germany



$q_F = 2$ factors are identified for France, $q_G = 3$ for Germany

France and Germany



$q = 3$ joint factors are identified

This implies that France's common space is a proper subspace of Germany's common space, and coincides with the strongly common subspace.

Next,

• applying the Forni-Lippi method to the global panel easily yields consistent reconstructions of χ_{it}^{FG} and ξ_{it}^{FG} (France), χ_{jt}^{FG} and ξ_{jt}^{FG} (Germany);

• applying the Forni-Lippi method to the marginal panels similarly would yield consistent reconstructions of $\chi_{F;it}$, $\xi_{F;it}$ and $\chi_{G;jt}$, $\xi_{G;jt}$, hence, by difference,

$$\nu_{F;it} = \chi_{it}^{FG} - \chi_{F;it} = \xi_{F;it} - \xi_{it}^{FG}$$

and

$$\nu_{G;jt} = \chi_{jt}^{FG} - \chi_{G;jt} = \xi_{G;jt} - \xi_{jt}^{FG}.$$

Note that $\chi_{y;it}$ also is the common component of $\chi_{xy;it}$ (y_{it}'s common component in the joint factor model decomposition), so that the same result can be obtained from factorizing the joint common spectral density matrices.

As a reconstruction of $\chi_{y;it}$ we therefore rather consider the projection $\chi^{\mathbf{n}}_{y;it}$ of $\chi^{\mathbf{n}}_{xy;it}$ onto the space spanned by the first q_y dynamic principal components

$$\mathbf{V}^{\mathbf{n}}_{y;t} := \left(V^{\mathbf{n}}_{y;1t}, \ldots, V^{\mathbf{n}}_{y;q_y t}\right)', \quad \text{with} \quad V^{\mathbf{n}}_{y;kt} := \underline{\mathbf{p}}^{*}_{\chi_{xy};\mathbf{n},k}(L)\boldsymbol{\chi}^{\mathbf{n}}_{xy,;t},$$

of the spectral density matrix $\mathbf{\Sigma}_{\chi_{xy};\mathbf{n}}(\theta)$ of $\boldsymbol{\chi}^{\mathbf{n}}_{xy;t} = \left(\chi^{\mathbf{n}}_{xy;1t}, \ldots, \chi^{\mathbf{n}}_{xy;t}\right)'$; $\mathbf{p}_{\chi_{xy};\mathbf{n},k}(\theta)$ here denotes the dynamic eigenvector associated with $\mathbf{\Sigma}_{\chi_{xy};\mathbf{n}}(\theta)$'s k-th dynamic eigenvalue $\lambda_{\chi_{xy};\mathbf{n},k}(\theta)$.

Similar results, with obvious notational adjustments, hold for the $\mathbf{Z}$'s.

- Disentangling the strongly common component ϕ and the weakly common component ψ is more tricky, though ...

Disentangling the strongly and weakly common components ...

- By definition, $\phi_{F;it}$ is obtained as the projection of χ^F_{it} onto $\mathcal{H}^{\chi}_{F\cap G}$, and $\psi_{F;it}$ follows as the residual $\chi^F_{it} - \phi_{F;it}$.

- Unlike $\mathcal{H}^{\chi}_{y}$ and $\mathcal{H}^{\chi}_{z}$, however, $\mathcal{H}^{\chi}_{F\cap G}$ is not characterized via an explicit sequence of orthonormal bases (it is not related to any set of dynamic PC's).

- The previous Forni-Hallin-Lippi-Reichlin methods, thus, do not apply unless some sequence of orthonormal bases can be computed from some different approach.

Preparation (1)

The following proposition is adapted from Theorem 8.3.1 in Brillinger (1981).

Assume that the $(r+s)$-dimensional second-order mean zero stationary process $\{(\boldsymbol{\zeta}_t', \boldsymbol{\eta}_t')',\ t \in \mathbb{Z}\}$ is such that the spectral density matrix $\boldsymbol{\Sigma}_{\eta\eta}(\theta)$ of $\boldsymbol{\eta}_t$ is nonsingular.

PROPOSITION 1. *The projection of $\boldsymbol{\zeta}_t$ onto the closed space $\mathcal{H}_\eta$ spanned by $\{\boldsymbol{\eta}_t\ , t \in \mathbb{Z}\}$—that is, the r-tuple $\mathbf{A}^*(L)\boldsymbol{\eta}_t$ of square summable linear combinations of the present, past and future of $\boldsymbol{\eta}_t$ minimizing*

$$\mathrm{E}[(\boldsymbol{\zeta}_t - \mathbf{A}^*(L)\boldsymbol{\eta}_t)'(\boldsymbol{\zeta}_t - \mathbf{A}^*(L)\boldsymbol{\eta}_t)]$$

is

$$\underline{\boldsymbol{\Sigma}}_{\zeta\eta}(L)\underline{\boldsymbol{\Sigma}}_{\eta\eta}^{-1}(L)\boldsymbol{\eta}_t,$$

where $\boldsymbol{\Sigma}_{\zeta\eta}(\theta)$ denotes the cross-spectrum of $\boldsymbol{\zeta}_t$ and $\boldsymbol{\eta}_t$.

Actually, Brillinger also requires $(\boldsymbol{\zeta}_t', \boldsymbol{\eta}_t')'$ to have absolutely summable autocovariances, so that the filter $\underline{\boldsymbol{\Sigma}}_{\zeta\eta}(L)\underline{\boldsymbol{\Sigma}}_{\eta\eta}^{-1}(L)$ also is absolutely summable. We, however, do not need this here.

Preparation (2)

The following result, to the best of our knowledge, is new.

Let $\mathcal{H}_1$, $\mathcal{H}_2$ and $\mathcal{H}_{12}$ be the Hilbert spaces spanned by $\{\mathbf{V}_{1;t}, t \in \mathbb{Z}\}$, $\{\mathbf{V}_{2;t}, t \in \mathbb{Z}\}$, and $\{(\mathbf{V}'_{1;t}, \mathbf{V}'_{2;t})', t \in \mathbb{Z}\}$, respectively, where $\mathbf{V}_{1;t} := (V_{1;1,t}, \ldots, V_{1;q_1,t})'$ is a q_1-tuple (resp. $\mathbf{V}_{2;t} := (V_{2;1,t}, \ldots, V_{2;q_2,t})'$ a q_2-tuple) of mutually orthogonal (at all leads and lags) nondegenerate stochastic processes: the dynamic dimensions of $\mathcal{H}_1$ and $\mathcal{H}_2$ are thus q_1 and q_2, respectively. Denoting by

$$\mathbf{\Sigma}(\theta) =: \begin{pmatrix} \mathbf{\Sigma}_{11}(\theta) & \mathbf{\Sigma}_{12}(\theta) \\ \mathbf{\Sigma}_{21}(\theta) & \mathbf{\Sigma}_{22}(\theta) \end{pmatrix}, \quad \theta \in [-\pi, \pi],$$

the spectral density matrix of $(\mathbf{V}'_{1;t}, \mathbf{V}'_{2;t})'$, with

$$\mathbf{\Sigma}_{11}(\theta) = \text{diag}(\lambda_{1;1}(\theta), \ldots, \lambda_{1;q_1}(\theta))$$

and

$$\mathbf{\Sigma}_{22}(\theta) = \text{diag}(\lambda_{2;1}(\theta), \ldots, \lambda_{2;q_2}(\theta)),$$

assume that $\mathbf{\Sigma}(\theta)$ has rank q_{12} θ-a.e., so that $\mathcal{H}_{12}$ has dynamic dimension q_{12}, and the intersection $\mathcal{H}_{1\cap 2} := \mathcal{H}_1 \cap \mathcal{H}_2$ dynamic dimension $q_{1\cap 2} = q_1 + q_2 - q_{12}$.

We have the following result.

Preparation (2)

PROPOSITION 2. *(i) The spectral density*

$$\mathbf{\Sigma}_{22}^{-1/2}(\theta)\mathbf{\Sigma}_{21}(\theta)\text{diag}(\lambda_{1;1}^{-1}(\theta), \dots, \lambda_{1;q_1}^{-1}(\theta))\mathbf{\Sigma}_{12}(\theta)\mathbf{\Sigma}_{22}^{-1/2}(\theta), \quad \theta \in [-\pi, \pi]$$

of the q_2-dimensional process $\underline{\mathbf{\Sigma}}_{22}^{-1/2}(L)\underline{\mathbf{\Sigma}}_{21}(L)\underline{\mathbf{\Sigma}}_{11}^{-1}(L)\mathbf{V}_{1;t}$ has a maximal eigenvalue equal to one, with multiplicity $q_{1\cap 2}$.

(ii) Denoting by $\mathbf{p}_{1\cap 2;1}(\theta), \dots, \mathbf{p}_{1\cap 2;q_{1\cap 2}}(\theta)$ an arbitrary orthonormal basis of the corresponding $q_{1\cap 2}$-dimensional eigenspace, the process $\{\mathbf{\Upsilon}_t := (\Upsilon_{1,t}, \dots, \Upsilon_{q_{1\cap 2},t})', t \in \mathbb{Z}\}$, with

$$\Upsilon_{k,t} := \underline{\mathbf{p}}_{1\cap 2;k}^{*}(L)\underline{\mathbf{\Sigma}}_{22}^{-1/2}(L)\mathbf{V}_{2;t}, \quad k = 1, \dots, q_{1\cap 2},$$

provides an orthonormal basis for $\mathcal{H}_{1\cap 2}$.

The intuition is that a random variable $\Upsilon \in \mathcal{H}_2$, that is, of the form $\Upsilon = \underline{\mathbf{a}}_{\Upsilon}^{*}(L)\mathbf{V}_{2;t}$, belongs to $\mathcal{H}_{1\cap 2}$ iff it coincides with its projection onto the space $\mathcal{H}_1$ spanned by the $\mathbf{V}_{1;t}$'s. That projection, on $\mathcal{H}_{1\cap 2}$, is thus an identity—so that its $q_{1\cap 2}$ largest eigenvalues are ones.

Now, in view of Proposition 1, $\underline{\mathbf{\Sigma}}_{22}^{-1/2}(L)\underline{\mathbf{\Sigma}}_{21}(L)\underline{\mathbf{\Sigma}}_{11}^{-1}(L)\mathbf{V}_{1;t}$ is the projection of the standardized random vector $\mathbf{V}_{1;t}$ onto the space spanned by the standardized process $\{\mathbf{V}_{2;t}\}$; the spectral density in part (i) is the spectral density of that projection.

Preparation (2)

PROOF. A random variable $\Upsilon \in \mathcal{H}_2$, that is, a variable of the form $\Upsilon = \underline{\mathbf{a}}^*_\Upsilon(L)\mathbf{V}_{2;t}$ with $\underline{\mathbf{a}}^*_\Upsilon(L) := (\underline{a}_{\Upsilon,1}(L), \ldots, \underline{a}_{\Upsilon,q_2}(L))$ belongs to $\mathcal{H}_{1\cap 2}$ iff it coincides with its projection onto the space $\mathcal{H}_1$ spanned by the $\mathbf{V}_{1;t}$'s.

In view of Proposition 1, that projection is

$$\underline{\mathbf{a}}^*_\Upsilon(L)\underline{\mathbf{\Sigma}}_{21}(L)\text{diag}(\underline{\lambda}^{-1}_{1;1}(L), \ldots, \underline{\lambda}^{-1}_{1;q_1}(L))\mathbf{V}_{1;t}.$$

The variance of a projection being less than or equal to the variance of the projected variable, the variance of that projection is less than or equal to the variance of Υ itself:

$$\int_{-\pi}^{\pi} \mathbf{a}^*_\Upsilon(\theta)\mathbf{\Sigma}_{21}(\theta)\text{diag}(\lambda^{-1}_{1;1}(\theta), \ldots, \lambda^{-1}_{1;q_1}(\theta))\mathbf{\Sigma}_{12}(\theta)\mathbf{a}_\Upsilon(\theta)d\theta$$

$$= \int_{-\pi}^{\pi} \mathbf{a}^*_\Upsilon(\theta)\mathbf{\Sigma}^{1/2}_{22}(\theta)\left[\mathbf{\Sigma}^{-1/2}_{22}(\theta)\mathbf{\Sigma}_{21}(\theta)\text{diag}(\lambda^{-1}_{1;1}(\theta), \ldots, \lambda^{-1}_{1;q_1}(\theta))\mathbf{\Sigma}_{12}(\theta)\mathbf{\Sigma}^{-1/2}_{22}(\theta)\right]\mathbf{\Sigma}^{1/2}_{22}(\theta)\mathbf{a}_\Upsilon(\theta)d\theta$$

$$\leq \int_{-\pi}^{\pi} \mathbf{a}^*_\Upsilon(\theta)\mathbf{\Sigma}_{22}(\theta)\mathbf{a}_\Upsilon(\theta)d\theta,$$

irrespective of $\mathbf{\Sigma}^{1/2}_{22}(\theta)\mathbf{a}_\Upsilon(\theta)$.

Preparation (2)

It follows that

- the spectral density matrix

$$\mathbf{\Sigma}_{22}^{-1/2}(\theta)\mathbf{\Sigma}_{21}(\theta)\text{diag}(\lambda_{1;1}^{-1}(\theta),\ldots,\lambda_{1;q_1}^{-1}(\theta))\mathbf{\Sigma}_{12}(\theta)\mathbf{\Sigma}_{22}^{-1/2}(\theta), \quad \theta \in [-\pi,\pi]$$

has eigenvalues less than or equal to one (θ-a.e.), and that

- Υ is in $\mathcal{H}_{1\cap 2}$ iff $\mathbf{\Sigma}_{22}^{1/2}(\theta)\mathbf{a}_{\Upsilon}(\theta)$ belongs to the eigenspace of that matrix associated with eigenvalue one.

The $q_{1\cap 2}$ random variables

$$\Upsilon_{k,t} := \underline{\mathbf{p}}_{1\cap 2;k}^{*}(L)\underline{\mathbf{\Sigma}}_{22}^{-1/2}(L)\mathbf{V}_{2;t}, \quad k = 1,\ldots,q_{1\cap 2},$$

clearly satisfy that condition, and it is easy to check that the spectral density of $\{\mathbf{\Upsilon}_t, t \in \mathbb{Z}\}$ moreover is the $q_{1\cap 2} \times q_{1\cap 2}$ identity matrix.

The result follows. $\square$

Disentangling the strongly weakly common components ...

- Proposition 2, with $\mathcal{H}_1 = \mathcal{H}^{\chi}_{y;\mathbf{n}}$ (spanned by $\mathbf{V}^{\mathbf{n}}_{y;t}$) and $\mathcal{H}_2 = \mathcal{H}^{\chi}_{z;\mathbf{n}}$ (spanned by $\mathbf{V}^{\mathbf{n}}_{z;t}$), hence $q_1 = q_y$, $q_2 = q_z$ and $q_{1\cap 2} = q_{y\cap z}$, provides an orthonormal basis $\mathbf{V}^{\mathbf{n}}_{F\cap G;t}$ for the intersection;

- Proposition 1 then provides the desired projection $\phi^{\mathbf{n}}_{y;it}$ of $\chi^{\mathbf{n}}_{y;it}$ onto that intersection.

More precisely, denote by

$$\mathbf{\Sigma}^{\mathbf{n}}_{\mathbf{V}}(\theta) := \left(\begin{array}{cc} \mathbf{\Sigma}^{\mathbf{n}}_{\mathbf{V}_y\mathbf{V}_y}(\theta) & \mathbf{\Sigma}^{\mathbf{n}}_{\mathbf{V}_y\mathbf{V}_z}(\theta) \\ \mathbf{\Sigma}^{\mathbf{n}}_{\mathbf{V}_z\mathbf{V}_y}(\theta) & \mathbf{\Sigma}^{\mathbf{n}}_{\mathbf{V}_z\mathbf{V}_z}(\theta) \end{array} \right)$$

the spectrum of $(\mathbf{V}^{\mathbf{n}\prime}_{y;t}, \mathbf{V}^{\mathbf{n}\prime}_{z;t})'$.

The matrix in part (i) of Proposition 2 (namely, the spectral density of the projection onto $\mathcal{H}^{\chi}_{z;\mathbf{n}}$ (spanned by $\mathbf{V}^{\mathbf{n}}_{z;t}$ of the standardized version of $\mathbf{V}^{\mathbf{n}}_{y;t}$ here takes the form

$$[\mathbf{\Sigma}^{\mathbf{n}}_{\mathbf{V}_z\mathbf{V}_z}(\theta)]^{-1/2}\mathbf{\Sigma}^{\mathbf{n}}_{\mathbf{V}_z\mathbf{V}_y}(\theta)[\mathbf{\Sigma}^{\mathbf{n}}_{\mathbf{V}_y\mathbf{V}_y}(\theta)]^{-1}\mathbf{\Sigma}^{\mathbf{n}}_{\mathbf{V}_y\mathbf{V}_z}(\theta)[\mathbf{\Sigma}^{\mathbf{n}}_{\mathbf{V}_z\mathbf{V}_z}(\theta)]^{-1/2};$$

denote by $\mathbf{p}_{y\cap z;\mathbf{n},1}(\theta), \ldots, \mathbf{p}_{y\cap z;\mathbf{n},q_{y\cap z}}(\theta)$ its $q_{y\cap z}$ first eigenvectors.

Disentangling the strongly weakly common components ...

An orthonormal basis of $\mathcal{H}^{\chi}_{y;\mathbf{n}} \cap \mathcal{H}^{\chi}_{z;\mathbf{n}}$ is $\mathbf{V}^{\mathbf{n}}_{y\cap z;t} := (V^{\mathbf{n}}_{y\cap z;1,t}, \dots, V^{\mathbf{n}}_{y\cap z;q_{y\cap z},t})'$ where, in view of part (ii) of the same proposition,

$$V^{\mathbf{n}}_{y\cap z;k,t} := \underline{\mathbf{p}}^{*}_{y\cap z;\mathbf{n},k}(L) \left(\underline{\mathbf{\Sigma}}^{\mathbf{n}}_{\mathbf{V}_z \mathbf{V}_z}\right)^{-1/2}(L)\mathbf{V}^{\mathbf{n}}_{z;t}, \quad k = 1, \dots, q_{y\cap z}.$$

Since

$$\chi^{\mathbf{n}}_{y;it} = \sum_{k=1}^{q_y} \underline{p}_{\chi_{xy};\mathbf{n},k,i}(L) V^{\mathbf{n}}_{y:k,t} = (\underline{p}_{\chi_{xy};\mathbf{n},1,i}(L), \dots, \underline{p}_{\chi_{xy};\mathbf{n},q_y,i}(L))\mathbf{V}^{\mathbf{n}}_{y:t}$$

(the projection onto the y-dynamic principal components of $\chi^{\mathbf{n}}_{xy;it}$), we first compute the projection onto $\mathcal{H}^{\chi}_{y;\mathbf{n}} \cap \mathcal{H}^{\chi}_{z;\mathbf{n}}$ of $\mathbf{V}^{\mathbf{n}}_{y:t}$.

That projection is obtained by applying Proposition 1 to the $(q_y + q_{y\cap z})$-dimensional random vector $(\mathbf{V}^{\mathbf{n}\prime}_{y:t}, \mathbf{V}^{\mathbf{n}\prime}_{y\cap z;t})'$, with spectral density

$$\begin{pmatrix} \boldsymbol{\Sigma}^{\mathbf{n}}_{\mathbf{V}_y\mathbf{V}_y}(\theta) & \boldsymbol{\Sigma}^{\mathbf{n}}_{\mathbf{V}_y\mathbf{V}_{y\cap z}}(\theta) \\ \boldsymbol{\Sigma}^{\mathbf{n}}_{\mathbf{V}_{y\cap z}\mathbf{V}_y}(\theta) & \boldsymbol{\Sigma}^{\mathbf{n}}_{\mathbf{V}_{y\cap z}\mathbf{V}_{y\cap z}}(\theta) \end{pmatrix} = \begin{pmatrix} \text{diag}(\lambda_{\chi_{xy};\mathbf{n},1}(\theta), \ldots, \lambda_{\chi_{xy};\mathbf{n},q_y}(\theta)) & \boldsymbol{\Sigma}^{\mathbf{n}}_{\mathbf{V}_y\mathbf{V}_{y\cap z}}(\theta) \\ \boldsymbol{\Sigma}^{\mathbf{n}}_{\mathbf{V}_{y\cap z}\mathbf{V}_y}(\theta) & \mathbf{I}^{\mathbf{n}}_{y\cap z\times y\cap z}(\theta) \end{pmatrix},$$

where $\boldsymbol{\Sigma}^{\mathbf{n}}_{\mathbf{V}_y\mathbf{V}_{y\cap z}}(\theta)$ follows from classical spectral algebra:

Disentangling the strongly weakly common components ...

Write

$$\mathbf{P}_{y\cap z;\mathbf{n}}(\theta) := \text{diag}(\lambda^{-1/2}_{\chi_{xz};\mathbf{n},1}(\theta), \ldots, \lambda^{-1/2}_{\chi_{xz};\mathbf{n},q_z}(\theta)) \left(\mathbf{p}_{y\cap z;\mathbf{n},1}(\theta), \ldots, \mathbf{p}_{y\cap z;\mathbf{n},q_{y\cap z}}(\theta)\right),$$

$$\mathbf{P}_{\chi_{xy};\mathbf{n}}(\theta) := \left(\mathbf{p}_{\chi_{xy};\mathbf{n},1}(\theta), \ldots, \mathbf{p}_{\chi_{xy};\mathbf{n},q_y}(\theta)\right),$$

$$\mathbf{P}_{\chi_{xz};\mathbf{n}}(\theta) := \left(\mathbf{p}_{\chi_{xz};\mathbf{n},1}(\theta), \ldots, \mathbf{p}_{\chi_{xy};\mathbf{n},q_z}(\theta)\right),$$

$$\mathbf{P}_{\mathbf{n}(y)}(\theta) := \left(\mathbf{p}_{\mathbf{n},1(y)}(\theta), \ldots, \mathbf{p}_{\mathbf{n},q,(y)}(\theta)\right),$$

and

$$\mathbf{P}_{\mathbf{n}(z)}(\theta) := \left(\mathbf{p}_{\mathbf{n},1,(z)}(\theta), \ldots, \mathbf{p}_{\mathbf{n},q,(z)}(\theta)\right),$$

with $\mathbf{p}_{\mathbf{n},k(y)}(\theta)$ collecting the components $p_{\mathbf{n},k,i}(\theta)$ of $\mathbf{p}_{\mathbf{n},k}(\theta)$ such that X_{it} belongs to the y-subpanel

and $\mathbf{p}_{\mathbf{n},k,(z)}(\theta)$ collecting the components $p_{\mathbf{n},k,j}(\theta)$ such that X_{jt} belongs to the z-subpanel).

Then,

Disentangling the strongly weakly common components ...

we have

$$\mathbf{\Sigma}^{\mathbf{n}}_{\mathbf{V}_y\mathbf{V}_{y\cap z}}(\theta) \quad = \quad \mathbf{P}^*_{\chi xy;\mathbf{n}}(\theta)\mathbf{P}_{\mathbf{n}(y)}(\theta)\text{diag}(\lambda_{\mathbf{n},1}(\theta),\ldots,\lambda_{\mathbf{n},q}(\theta))\mathbf{P}^*_{\mathbf{n}(z)}(\theta)\mathbf{P}_{\chi xz;\mathbf{n}}(\theta)\mathbf{P}_{y\cap z;\mathbf{n}}(\theta).$$

The desired projection of $\mathbf{V}^{\mathbf{n}}_{y:t}$, in view of Proposition 1, is $\underline{\mathbf{\Sigma}}^{\mathbf{n}}_{\mathbf{V}_y\mathbf{V}_{y\cap z}}(L)\mathbf{V}^{\mathbf{n}}_{y\cap z;t}$.

Hence, the reconstructions we are proposing are, for the strongly common component $\phi_{y;it}$,

$$\begin{aligned}
\phi^{\mathbf{n}}_{y;it} := &\ (\underline{p}_{\chi xy;\mathbf{n},1,i}(L),\ldots,\underline{p}_{\chi xy;\mathbf{n},q_y,i}(L))\underline{\mathbf{\Sigma}}^{\mathbf{n}}_{\mathbf{V}_y\mathbf{V}_{y\cap z}}(L)\mathbf{V}^{\mathbf{n}}_{y\cap z;t} \\
= &\ (\underline{p}_{\chi xy;\mathbf{n},1,i}(L),\ldots,\underline{p}_{\chi xy;\mathbf{n},q_y,i}(L))\underline{\mathbf{P}}^*_{\chi xy;\mathbf{n}}(L)\underline{\mathbf{P}}_{\mathbf{n}(y)}(L) \\
&\qquad \times\text{diag}(\underline{\lambda}_{\mathbf{n},1}(L),\ldots,\underline{\lambda}_{\mathbf{n},q}(L))\underline{\mathbf{P}}^*_{\mathbf{n}(z)}(L)\underline{\mathbf{P}}_{\chi xz;\mathbf{n}}(L)\underline{\mathbf{P}}_{y\cap z;\mathbf{n}}(L)\underline{\mathbf{P}}^*_{y\cap z;\mathbf{n}}(L)\mathbf{V}^{\mathbf{n}}_{z;t} \\
=: &\ \underline{\mathbf{H}}^*_{y;\mathbf{n},i}(L)\mathbf{V}^{\mathbf{n}}_{z;t},
\end{aligned}$$

and, for the weakly common one $\psi_{y;it}$, $\psi^{\mathbf{n}}_{y;it} := \chi^{\mathbf{n}}_{y;it} - \phi^{\mathbf{n}}_{y;it}$.

With obvious changes, we similarly define $\phi^{\mathbf{n}}_{z;jt}$ and $\psi^{\mathbf{n}}_{z;jt}$.

Wee then have the following consistency result for $\phi^{\mathbf{n}}_{y;it}$ and $\phi^{\mathbf{n}}_{z;jt}$ (hence $\psi^{\mathbf{n}}_{y;it}$ and $\psi^{\mathbf{n}}_{z;jt}$).

PROPOSITION. *Let Assumptions A1, A2, A3, and and A4 hold. Then*

$$\lim_{\min(n_y,n_z)\to\infty} \phi^{\mathbf{n}}_{y;it} = \phi_{y;it} \quad \textit{and} \quad \lim_{\min(n_y,n_z)\to\infty} \phi^{\mathbf{n}}_{z;jt} = \phi_{z;jt}$$

in quadratic mean, for any i, j, and t.

PROOF. The proof still follows from Proposition 2 of Forni *et al.* (2000), and the fact that all spectral densities involved, for given n, are locally continuous functions of $\mathbf{\Sigma}_{\mathbf{n}}(\theta)$. $\square$

Further results

We refer to the JoE paper for results on estimation, and the more complicated case of $K > 2$ blocks

Applications

Dataset of monthly Industrial Production Indexes for France, Germany, and Italy, observed from January 1995 through December 2006 ($T = 143$ throughout).

All data were preadjusted by taking a log-difference transformation, then centered and normalized using their sample means and standard errors.

Two blocks: France, Germany

Y_{it} : French data, $n_y = n_F = 96$

Z_{jt} : German data, $n_z = n_G = 114$ (hence $n = n_{FG} = 210$)

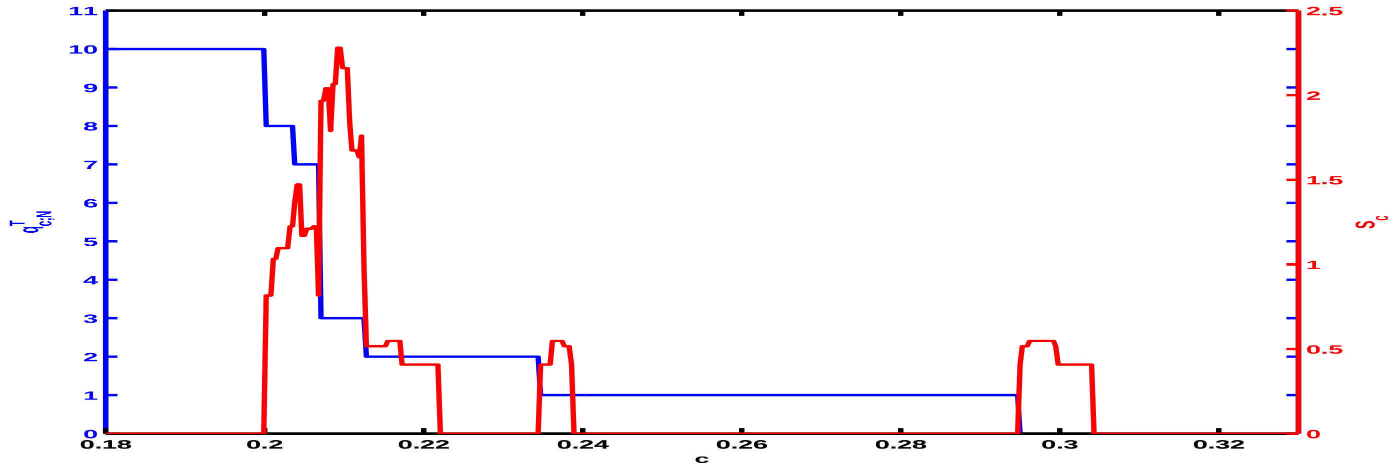
- We ran the identification method on the French and German subpanels, with sequences $n_{F,j} = 96 - 2j$, $j = 1, \ldots, 5$ and $n_{G,j} = 96 - 2j$, $j = 1, \ldots, 5$, respectively, then on the pooled panel, with sequence $n_{FG,j} = 210 - 2j$, $j = 1, \ldots, 8$ and an "almost constant " proportion $96/210$, $114/210$ of French and German observations (namely, $\lceil 96 n_{FG,j}/210 \rceil$ French observations, and $\lfloor 114 n_{FG,j}/210 \rfloor$ German ones. In all cases, we put $T_j = T = 143$, $j = 1, \ldots, 5$.

- The range for c values, after some preliminary exploration, was taken as $[0, 0.0002, 0.0004, \ldots, 0.5]$, and $q_{\max}$ was set to 10.

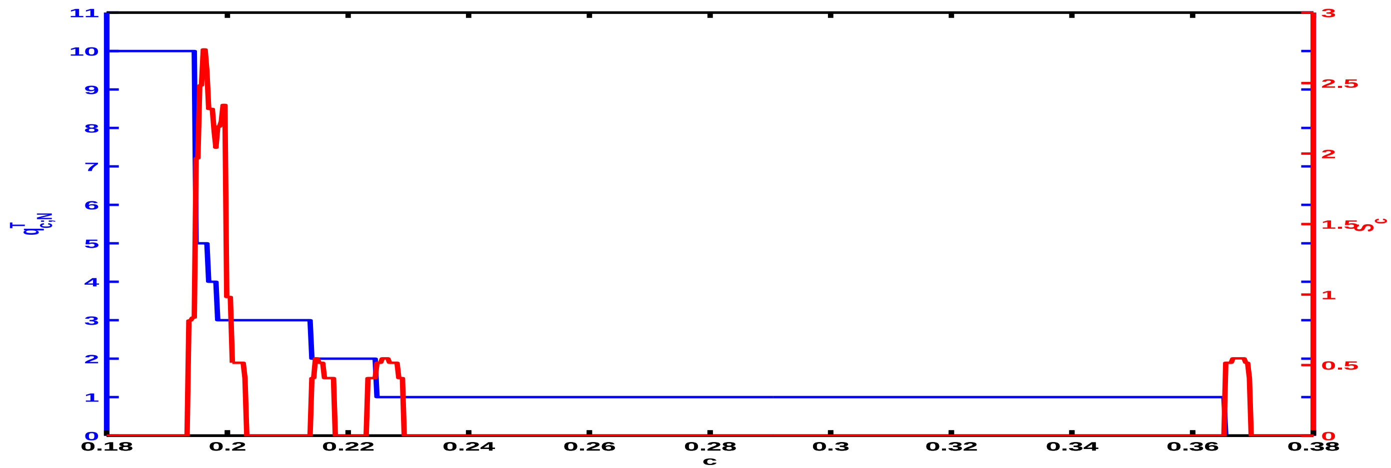
- Panels were randomly ordered prior to the analysis. The penalty function was $p(n, T) = \left(\min\left[n, M_T^2, M_T^{-1/2} T^{1/2}\right]\right)^{-1/2}$.

France and Germany

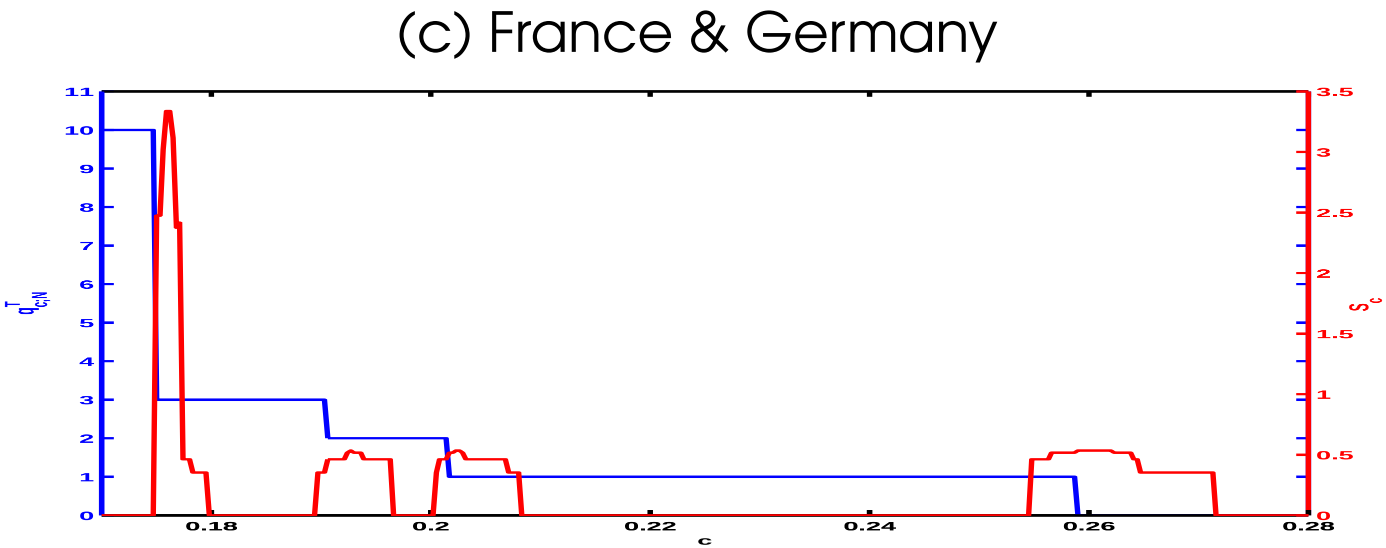
(a) France



(b) Germany

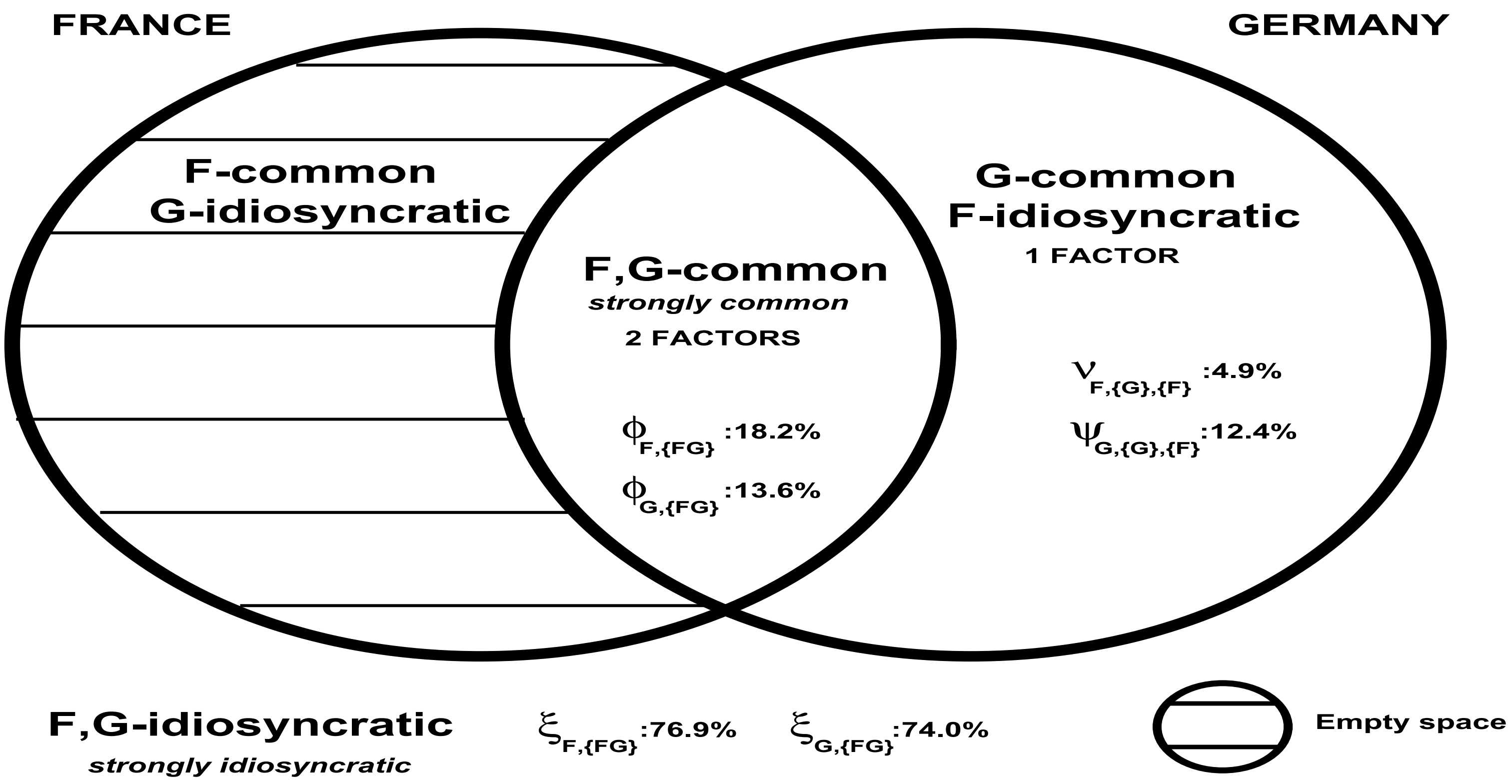


$q_F = 2$ factors are identified for France, $q_G = 3$ for Germany



$q = 3$ joint factors are identified

France and Germany



Decomposition of the France-Germany panel data into four mutually orthogonal components, with the corresponding percentages of explained variation.

- The French-common factors thus are strongly common (no weakly common space), whereas one German-common factor is French-idiosyncratic, and explains almost as much (12.4 %) of German variation as the two strongly common ones (13.6 %).

Three blocks : France, Germany and Italy

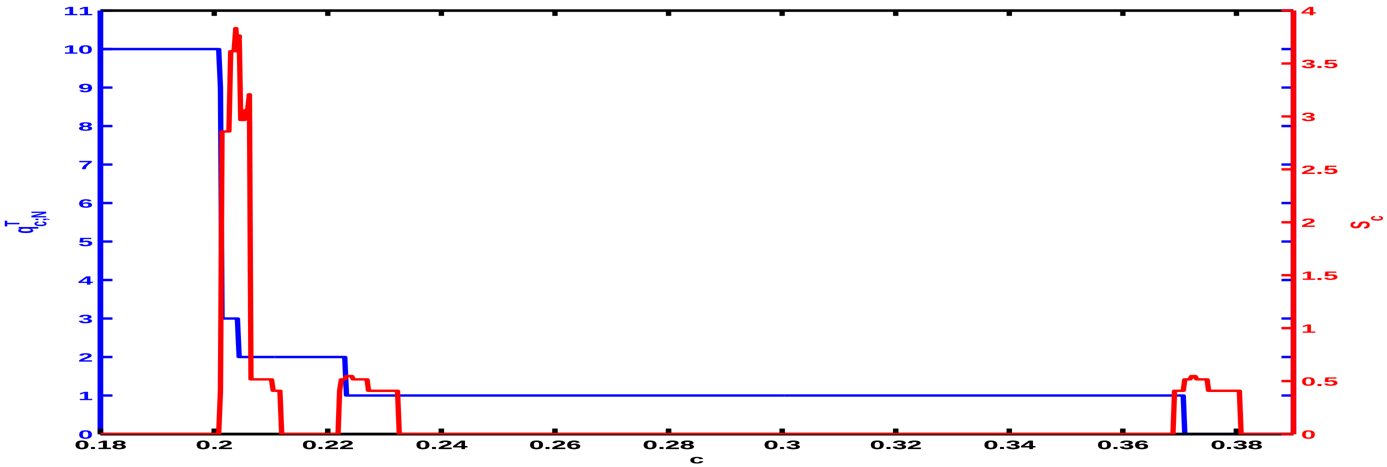
- Additional block : Italy, $n_I = 91$, $T = 143$; now $n = n_{FGI} = 301$

- From the resulting panel ($n = n_{FGI} = 301$ series), we can extract six subpanels—the three panels we already analyzed in Section 6.1 (the two-block French-German panel, the French and the German one-block subpanels), one new one-block subpanel (the marginal Italian one, with $n_I = 91$), two new two-block subpanels (the French-Italian one, with $n_{FI} = 187$ and the German-Italian one, with $n_{GI} = 205$, respectively).

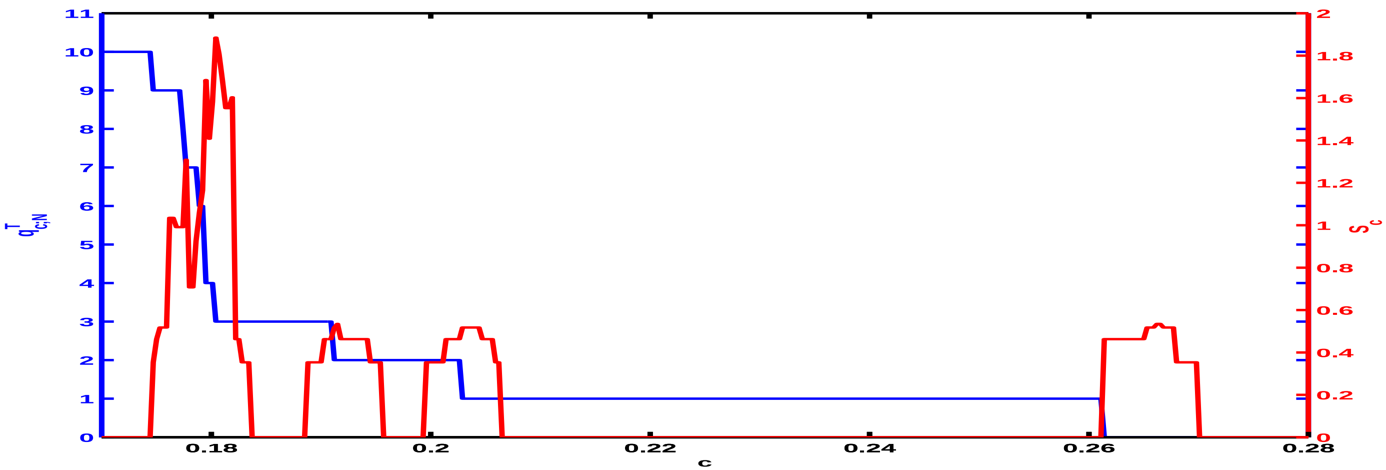
- Analyzing these new subpanels along the same lines as before (with, using obvious notation, $n_{I,j} = 91 - 2j$, $j = 1, \ldots,$ 5, $n_{GI,j} = 191 - 2j$, $j = 1, \ldots,$ 8, $n_{FI,j} = 187 - 2j$, $j = 1, \ldots,$ 8, and $n_{FGI,j} = 301 - 2j$, $j = 1, \ldots,$ 15), still with $M_T = 0.5\sqrt{T} = 5$, the same penalty function and the same $q_{\text{max}} = 10$ as before, we obtain the identification results shown in the following four graphs.

Three blocks : France, Germany and Italy

(d) Italy

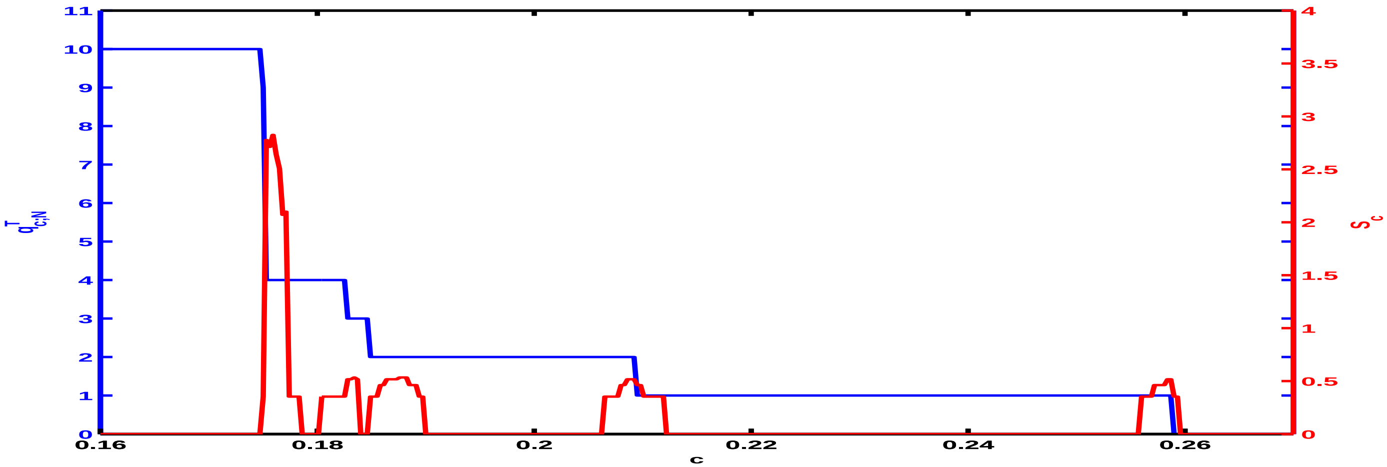


(e) France & Italy

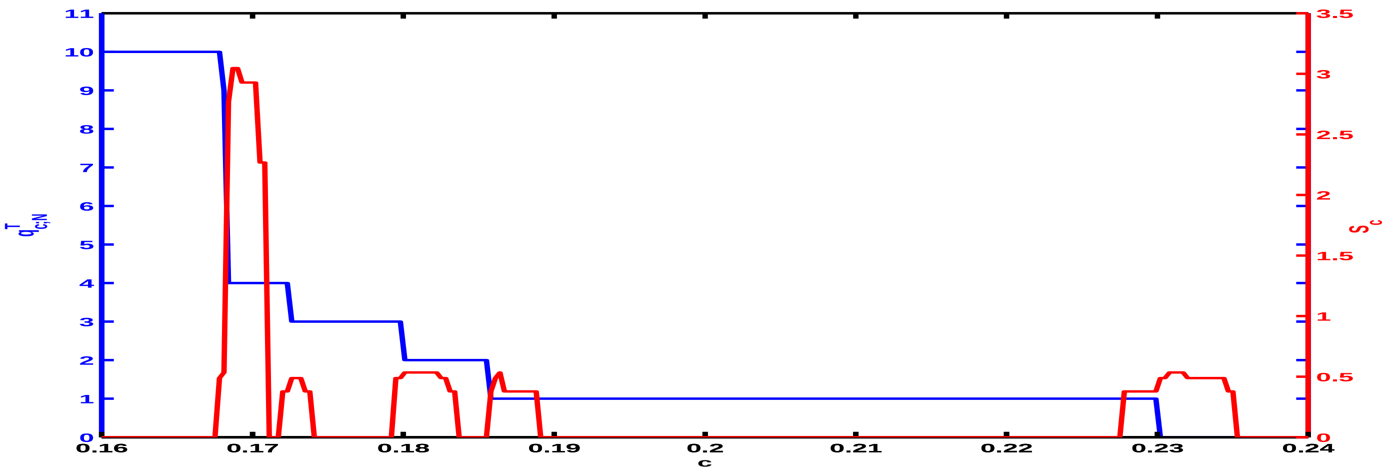


Identified marginal number of factors for Italy alone is 2; together with France, 3

(f) Germany & Italy



(g) France & Germany & Italy

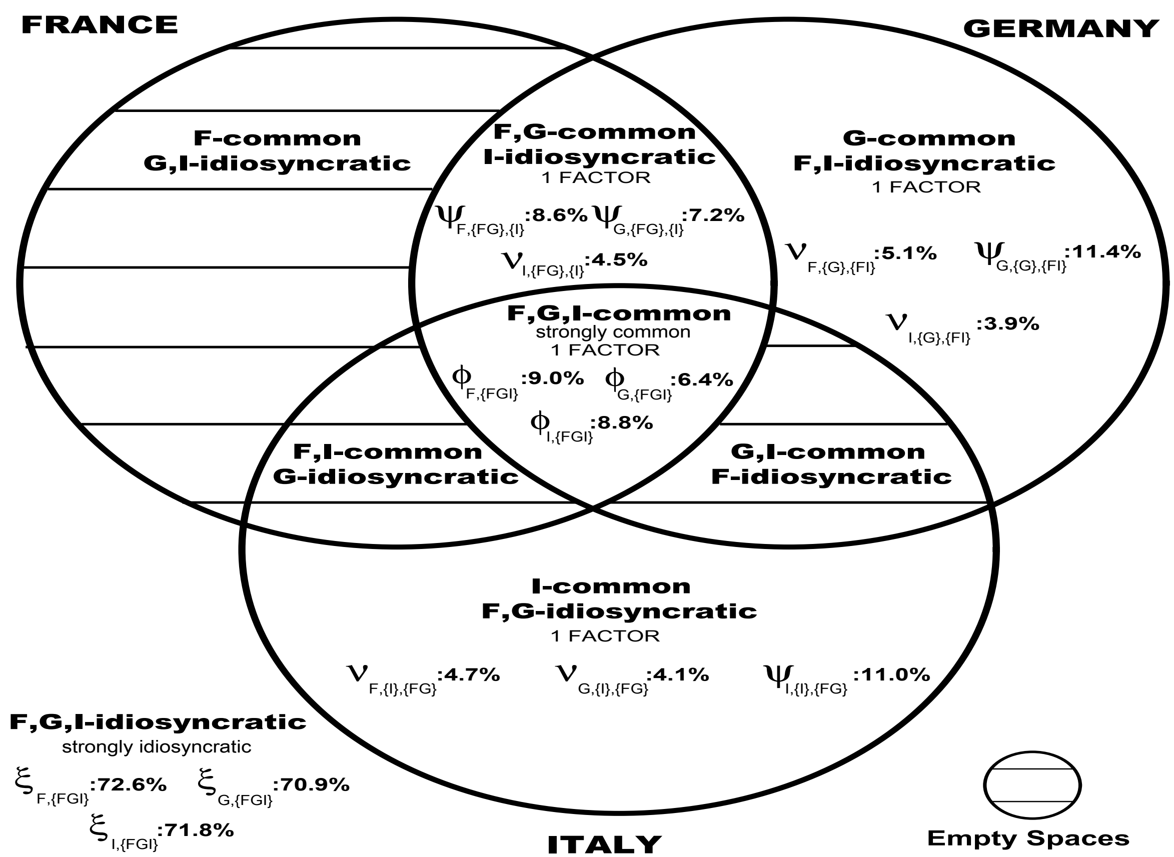


Italy and Germany: four identified factors; the whole panel also has 4 factors

- These graphs again very clearly identify a total number of $q^T_{\mathbf{n},FGI} = 4$ joint common factors (for $c \in [0.1710, 0.1718]$), $q^T_{(nF,n_I),FI} = 3$ (for $c \in [0.1838, 0.1886]$) French-Italian, and $q^T_{(n_G,n_I),GI} = 4$ (for $c \in [0.1786, 0.1800]$) German-Italian marginal "binational" factors, and $q^T_{n_I,I} = 2$ (for $c \in [0.2118, 0.22218]$) marginal Italian factors. Along with the figures obtained in Section 6.1 for France and Germany, this implies that $\mathcal{H}^\chi_F \subset \mathcal{H}^\chi_G$, hence $\mathcal{H}^\chi_F \cap \mathcal{H}^\xi_G = \{0\} = \mathcal{H}^\chi_{GI} \cap \mathcal{H}^\xi_F$. The relations between those various (dynamic) dimensions are easily obtained; for instance, $q_{(n_F,n_G),FG} = q_{n_F,F} + q_{n_G,G} - q_{(n_F,n_G)}$, a relation we already used in Section 6.1, or

$$q_{(n_F,n_G),FG} = q_{n_F,F} + q_{n_G,G} + q_{n_I,I} - q_{(n_F,n_G),FG} - q_{(n_F,n_I),FI} - q_{(n_G,n_I),GI} + q_{(n_F,n_G,n_I),FGI}.$$

These relations imply that the three countries are sharing one strongly common factor. As already noted, France (two factors) has no specific common factor, but one (the strongly common one) shared with Germany and Italy, and one shared withGermany alone. Both Italy (two factors) and Germany (three factors) have a "national" factor. Italy's "non national" factor is the strongly common one; Germany's "non national" factors are those shared with France, and include the strongly common one. The Italian and German "national" factors need not be mutually orthogonal.



Decomposition of the France-Germany-Italy panel data into eight components, with the corresponding percentages of explained variation.

- the three countries all exhibit a high percentage of about 70 % of strongly idiosyncratic variation
- France has no common components but those shared with Germany and Italy (one), and with Germany alone (one)
- both Italy and Germany have a "national common component, which in the case of Italy induces a nonnegligible percentage of about 3.9 % of weakly idiosyncratic variation in the other two countries (Germany and France)
- Italy's only "non national" common factor is the strongly common one, which is common to the three countries under study

Liquidity

- An asset is liquid if it is easily convertible into cash, the reference asset with perfect liquidity.

- This definition is often rephrased in terms of time, volume, and cost: when people think about liquidity, they may think about trading quickly, about trading large size, or about trading at low cost.

- The time dimension refers to *resiliency*—the speed with which pricing errors caused by uninformative order-flow shocks are corrected or neutralized in the market. Cost refers to *tightness*—the accepted price for immediacy in resolving the trade. Last, volume refers to *depth*—the volume that can be traded without price variations.

• Though the concept of liquidity is qualitatively clear, its quantitative evaluation poses a major problem.

• Liquidity is an unobserved variable, hence is evaluated via liquidity-related proxies—call them *liquidity measures*.

• measuring liquidity is a delicate task because of the difficulty to capture the three dimensions of liquidity in a single measure yet reaching a consensus on the liquidity measures to be taken into account.

This double difficulty seriously challenges the objectivity of any final assessment.

• Daily close or open bid-ask spreads—the difference between the lowest ask and highest bid prices for an asset at some given point in time—measure liquidity effects, but mainly cover tightness.

• Daily realized volumes only cover transacted depth ...

Market Liquidity

Market Liquidity aims at measuring “commonness in liquidity across securities”.

- There is empirical evidence that a “common” or “market” component is significantly present in various liquidity measures taken over a large cross-section of stocks.

Two typical contributions :

- Hasbrouck and Seppi (*Journal of Financial Economics*, 2001) perform a (classical, hence static) Principal Component Analysis on liquidity measures, out of which they consider up to three principal components.

- Korajczyk and Sadka (*Journal of Financial Economics*, 2008) study a very large sample (more than 4000 stocks followed during 18 years), they use via a static factor model method, for eight distinct measures, and extract up to the third principal component for each liquidity measure.

- None of these approaches fully exploits the time series nature of the data.

- They all overlook the leading/lagging phenomenons that may exist among the various liquidity measures and are particularly relevant here, since liquidity-related data are highly autocorrelated.

- • Dynamic factor models quite naturally enter into the picture: they allow for disentangling commonness (market components) and idiosyncrasy (stock-specific components), not only across panels consisting of some given liquidity measure observed over a large number of stocks, but also across panels juxtaposing several such measures.

- • General Dynamic Factor Model methods do not impose any restriction (beyond the usual assumptions of second-order stationarity, etc.) on the actual data generating process.

Dataset (two blocks):

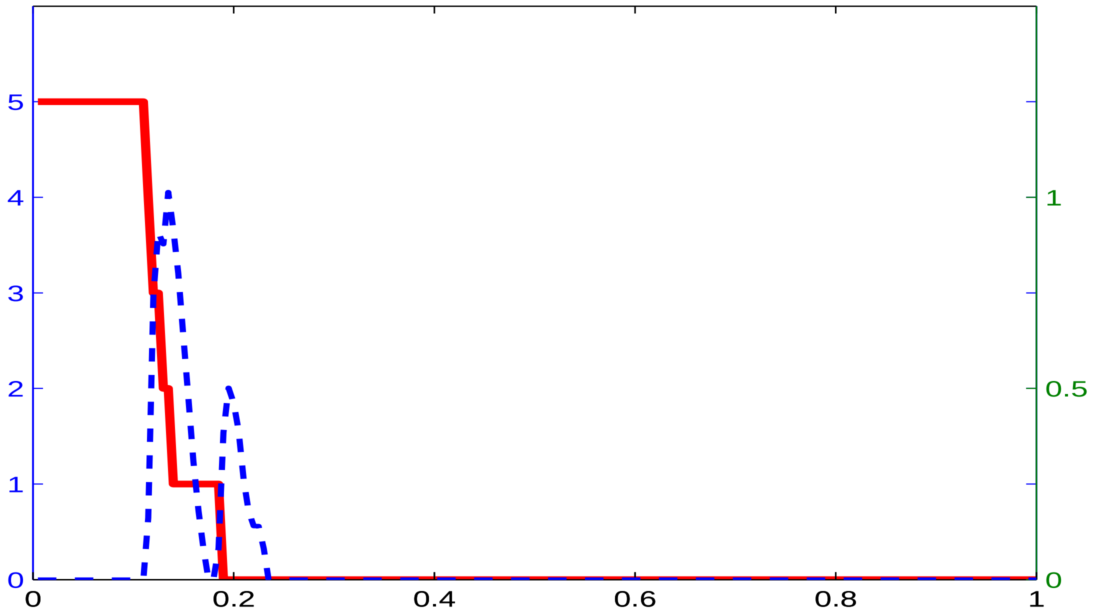
- daily close relative bid-ask spread and
- daily realized dollar volume,

for 426 S&P500 listed stocks from January 2004 till December 2006—a period considered as a “normal” state of market liquidity.

Number of Factors

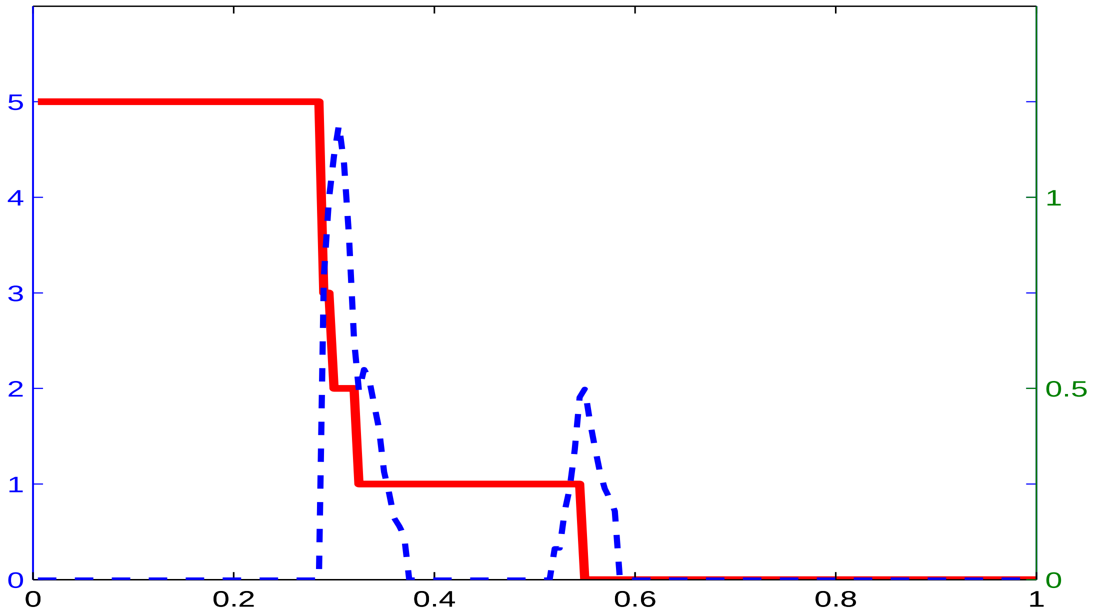
Joint panel

SPR$_n$



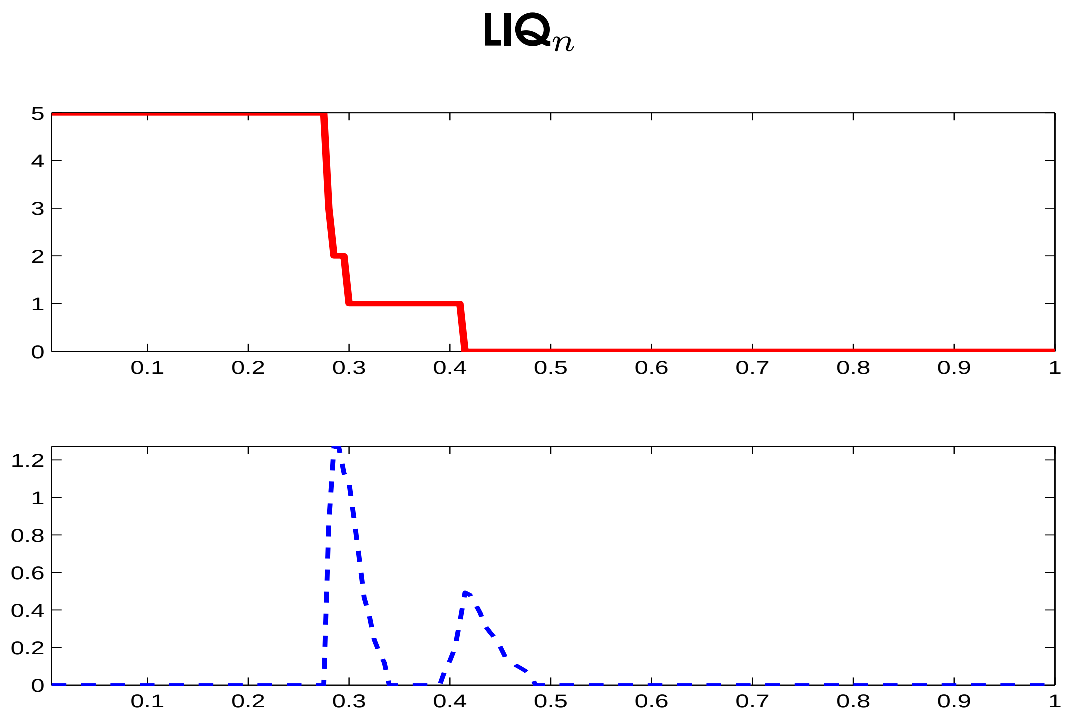
Daily realized dollar volumes

VOL$_n$

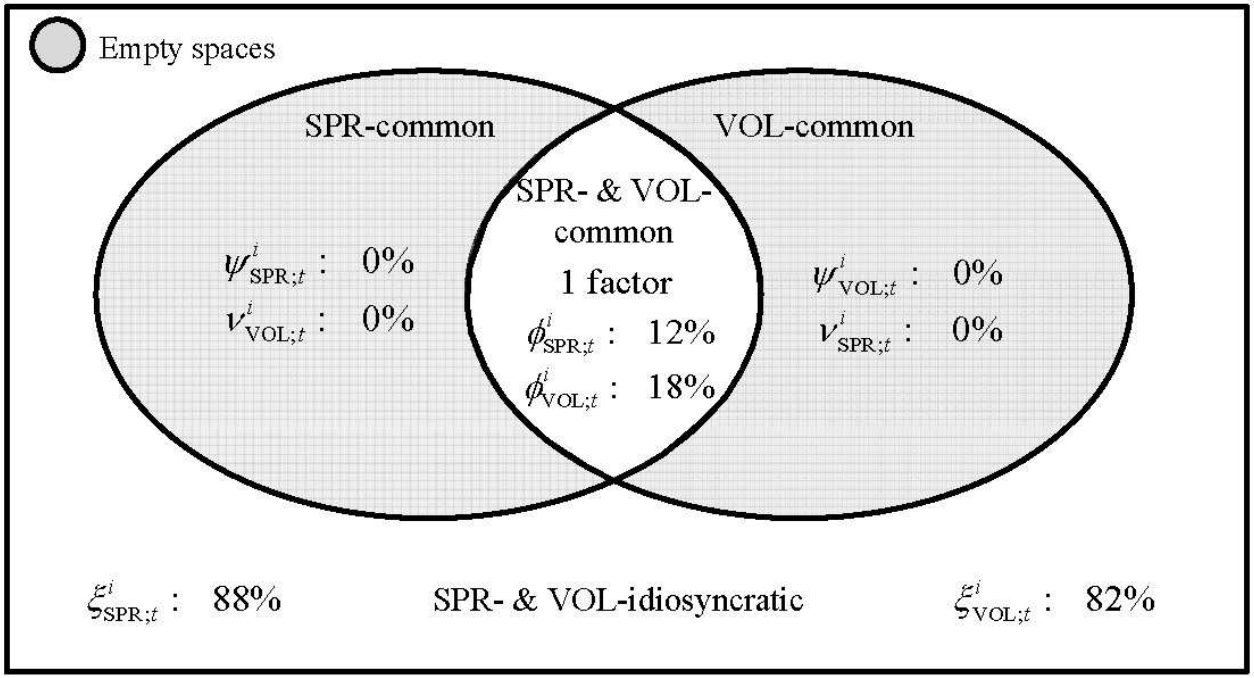


Number of Factors

Daily close relative bid-ask spreads



Number of Factors



Conclusions

- The general dynamic factor model, along with the Forni et al. estimation method and the Hallin-Liška identification procedure, allows for a subtle analysis of (weakly, strongly, etc.) "common" and "idiosyncratic" variations in the presence of blocks;

- when several conflicting proxies are considered for unobserved quantities, the same method allows for summarizing those proxies by extracting their strongly common shocks.

- For details on the K-block case, we refere to Hallin and Liška (2011).

Returns and Volatilities in High Dimension
a General Factor Model Approach

4. Dynamic Factors and Volatilities:

Extracting the Market Volatility Shocks

Marc Hallin

ECARES, Université Libre de Bruxelles

Hermann Otto Hirschfeld Lecture Series 2017

Berlin, 20–21 November 2017

1. Introduction

Slight change in notation:

let

$$\begin{array}{cccc} Y_{11}, & Y_{12}, & \ldots, & Y_{1T} \\ \vdots & \vdots & & \vdots \\ Y_{n1}, & Y_{n2}, & \ldots, & Y_{nT} \end{array}$$

denote a $n \times T$ panel of returns, that is, the finite realization of a stochastic process of the form

$$\{Y_{it} | i \in \mathbb{N},\ t \in \mathbb{Z}\}.$$

- Hence, a collection of n time series of returns, observed over a period of length T, related to n individuals stocks

Both n and T are "large", and (n, T)-asymptotics are considered throughout.

These n observed series are exposed, in general, to the influence of the same covariables, which are not recorded, inducing complex interrelations that are not statistically tractable, or that would involve uncomfortably many parameters ...

... parametric methods, as a rule, are helpless or unrealistic (or both) ...

... factor model methods (under their various forms) in this context appear to be the most successful tools.

But they never were applied in the analysis of volatilities.

Recall that, when applied to $\{Y_{it} | i \in \mathbb{N},\ t \in \mathbb{Z}\}$, factor model methods aim at identifying and reconstructing a decomposition of Y_{it} (actually, a decomposition of the Hilbert space spanned by the panel) into two mutually orthogonal parts

$$Y_{it} = X_{it} + Z_{it} = \text{“common”}_{it} + \text{“idiosyncratic”}_{it}$$

where

1. Introduction

$$Y_{it} = X_{it} + Z_{it} = \text{“common”}_{it} + \text{“idiosyncratic”}_{it}$$

where (the so-called “general dynamic factor model” or GDFM)

- the common component $\{X_{it} | i \in \mathbb{N},\ t \in \mathbb{Z}\}$ is driven by a small number q (q unspecified) of mutually orthogonal white noises—the “common” or “market shocks” $\{u_{1t}\}$, ..., $\{u_{qt}\}$:

$$X_{it} = \sum_{k=1}^{q} b_{ik}(L) u_{kt} \qquad i \in \mathbb{N},\ t \in \mathbb{Z};$$

yielding a reduced-rank linear process (with rank q)

- the idiosyncratic component $\{Z_{it} | i \in \mathbb{N},\ t \in \mathbb{Z}\}$ is only mildly cross-correlated—here (GDFM), such that all normed linear combinations of the form

$$\sum_{i=1}^{n} \sum_{j=-\infty}^{\infty} a_{ij} Z_{i,t-j}$$

have bounded variance as $n \to \infty$. $i \in \mathbb{N}$, $j = 1, \ldots, q$; equivalently, the eigenvalues of Z_{it}'s $n \times n$ spectral density matrices are bounded as $n \to \infty$.

- the filters $b_{ij}(L)$ are one-sided and square-summable.

... a "divide and rule" strategy:

- Being reduced rank, the series of common components somehow can be handled as a low-dimensional series—in particular, the (low-dimensional) common shocks can be recovered and fundamental representations of the X_{it}'s can be estimated

- Being only mildly cross-correlated, the n-dimensional series of idiosyncratic components Z_{it} can be handled, without much loss, as n univariate (auto-correlated but not cross-correlated) series. In particular, univariate AR fits and a global VAR fit roughly produce the same residuals

- Decomposing stock returns into a market-driven and an idiosyncratic or stock-specific component is certainly an important issue ...

- In financial econometrics, risk management, and portfolio optimization, risks and volatilities are at least as important: decomposing volatilities into a *common*, market-driven contribution and an individual, *idiosyncratic* one is perhaps the main issues. Market-driven risks indeed cannot be diversified away, while individual ones can be eliminated through clever portfolio diversification ....

- **being entirely based on the covariance structure of levels or returns, however, the above factor model decomposition cannot tell us anything about a decomposition of volatilities.**

It may seem natural to define "market risk" or "market volatility" as the risk associated with the market-driven component of returns:

$$\text{market risk} := \text{risk of the market-driven component of returns}$$

This is the approach adopted, for instance, in Fan, Liao, and Shi (2013) where the risk associated with a (second-order stationary) panel is measured by its unconditional marginal covariance matrix, and "market risk" is defined as the covariance matrix $\mathbf{\Sigma}_{\text{common}} = \text{Cov}(\mathbf{X})$ of the common component (in a low rank + sparse [a] context).

[a] Would Fan et al. consider a general dynamic factor model approach, their assumption of sparsity could be dropped.

“Natural” as it is, that idea is not a good idea or, at least, it is ways too simplistic.

The decomposition between level-common and level-idiosyncratic indeed has been based on the (unconditional) (auto)covariance structure of levels only, which carries no information on volatilities (which deal with conditional variances or scales). Therefore, this cannot be a good approach:

market volatility shocks are likely to affect the volatility of level-idiosyncratic components as well as they affect the volatility of level-icommon components

Building on that remark, we propose a two-step general dynamic factor procedure to extract the market volatility shocks from a large panel of stock returns.

- The method is entirely nonparametric and model-free.
- As a by-product, it provides considerable insight on the way market volatility shocks are loaded and propagate across the panel.

2. the General Dynamic Factor Model

The factor model decomposition we are adopting here is the so-called **General Dynamic Factor Model** one (Forni, Hallin, Lippi, and Reichlin (2000)), which encompasses all others, and is based on an idea of *dynamic (non-) pervasiveness*.

More precisely, assume that $\{Y_{it} | i = 1, \ldots, n; t \in \mathbb{Z}\}$ is zero-mean, second-order stationary, and has a spectral density matrix $\mathbf{\Sigma}_n(\theta)$, $\theta \in [-\pi, \pi]$, with eigenvalues $\lambda_{\mathbf{Y};n,i}(\theta)$, $i = 1, \ldots, n$ (the *dynamic eigenvalues*, in decreasing order of magnitude)

Then, $\{Y_{it}\}$ admits a *general dynamic factor representation*

$$Y_{it} = X_{it} + Z_{it} = \sum_{k=1}^{q} b_{ik}(L) u_{kt} + Z_{it} \qquad i \in \mathbb{N},\ t \in \mathbb{Z}$$

iff for some finite $q \in \mathbb{N}$,

- the qth dynamic eigenvalue $\lambda_{\mathbf{Y};n,q}(\theta)$, diverges as $n \to \infty$, a.e. in $[-\pi, \pi]$, while
- the $(q+1)$th one, $\lambda_{\mathbf{Y};n,q+1}(\theta)$, is θ-a.e. bounded.

Such behavior is quite typical; although "$q = \infty$" (an infinite number of diverging dynamic eigenvalues) is mathematically possible, it only happens under very weird and contrived data-generating mechanisms: see Hallin and Lippi (*Stochastic Processes and their Applications*, 2013).

Methods have been proposed to estimate and analyze that General Dynamic Factor Model decomposition:

- based on Brillinger's concept of dynamic principal components, hence two-sided filters, see Forni, Hallin, Lippi, and Reichlin (2000)
- for a more sophisticated one, involving one-sided filters only, see Forni, Hallin, Lippi and Zaffaroni (2015, 2017).

That factor model analysis, applied to the levels (returns) Y_{it}, is **Step 1** of our two-step procedure.

So far, nothing about volatilities ... the analysis just performed is entirely based on the levels' covariance structure, **which does not contain any information on volatilities.**

So far, nothing about volatilities ...

Traditional (univariate) analyses of volatilities are based on some nonlinear transformation of innovations or residuals e_t (for instance, the (centered) squared residuals e_t^2, or their (centered) logarithms).

The problem is: how can we obtain such residuals?

Typically, residuals are obtained by fitting parametric time series models to the observed returns which, because of the curse of dimensionality, is impossible here (or, the factor model approach would not be necessary)

The general dynamic factor method actually allows us to recover two collections of residuals: one for the common components, and another one for the idiosyncratic components.

2. the General Dynamic Factor Model

Forni, Hallin, Lippi, and Zaffaroni (*Journal of Econometrics* 2015, 2017) show that, under an additional (but mild) assumption of a rational spectrum, there exist a (block-diagonal) $n \times n$ matrix $\mathbf{A}_n(L)$ of one-sided filters (that can be estimated via classical $(q+1)$-dimensional VAR fitting) such that

$$(\mathbf{I}_n - \mathbf{A}_n(L))\,\mathbf{Y}_{n,t} = \mathbf{H}_n\mathbf{u}_t + (\mathbf{I}_n - \mathbf{A}_n(L))\,\mathbf{Z}_{n,t} =: \mathbf{e}_{n,t} + \widetilde{\mathbf{Z}}_{n,t}, \quad t \in \mathbb{Z},$$

where

- $\widetilde{\mathbf{Z}}_{n,t} := (\mathbf{I}_n - \mathbf{A}_n(L))\,\mathbf{Z}_{n,t}$ is idiosyncratic (its largest dynamic eigenvalue is θ–a.e. bounded as $n \to \infty$) and
- $\mathbf{e}_{n,t} := \mathbf{H}_n\mathbf{u}_t$ is n-dimensional white noise (the fundamental shocks of $\mathbf{Y}_{n,t}$'s common component $\mathbf{X}_{n,t}$, with reduced rank q).

(Without loss of generality, we can impose the identifying assumption that $\mathbf{H}_n'\mathbf{H}_n = n\mathbf{I}_q$).

2. the General Dynamic Factor Model

This decomposition

$$(\mathbf{I}_n - \mathbf{A}_n(L))\,\mathbf{Y}_{n,t} =: \mathbf{e}_{n,t} + \widetilde{\mathbf{Z}}_{n,t}, \quad t \in \mathbb{Z},$$

is a static factor model representation of $\mathbf{Y}_{n,t}$

hence can be analyzed via the static factor methods developed by Stock, Watson, Bai, Ng, etc. ... without imposing, however, the restrictive assumptions they need), yielding a consistent **one-sided**, hence feasible, reconstruction of $\mathbf{e}_{n,t}$ and $\widetilde{\mathbf{Z}}_{n,t}$.

2. the General Dynamic Factor Model

This decomposition

$$(\mathbf{I}_n - \mathbf{A}_n(L))\,\mathbf{Y}_{n,t} =: \mathbf{e}_{n,t} + \widetilde{\mathbf{Z}}_{n,t}, \quad t \in \mathbb{Z},$$

is a static factor model representation of $\mathbf{Y}_{n,t}$

hence can be analyzed via the static factor methods developed by Stock, Watson, Bai, Ng, etc. ... without imposing, however, the restrictive assumptions they need), yielding a consistent **one-sided**, hence feasible, reconstruction of $\mathbf{e}_{n,t}$ and $\widetilde{\mathbf{Z}}_{n,t}$.

Those $\mathbf{e}_{n,t}$'s are the residuals we need for an analysis of the volatility of the level-common components. They are a reduced-rank process (dimension n, driven by q-dimensional noise).

As for the idiosyncratic $\widetilde{\mathbf{Z}}_{n,t}$'s, since they are only mildly cross-correlated, residuals v_{it} can be obtained via individual AR fitting: **those v_{it}'s are the residuals we need for an analysis of the volatility of the level-idiosyncratic components**.

3. Common and idiosyncratic volatility/Market volatility

The general dynamic factor model analysis of returns thus points at two sources of information on market volatility shocks:

- volatility of the components e_{it} of $\mathbf{e}_{n,t}$ (**volatility of level-common components**)
- volatility of the shocks v_{it} driving the components $\widetilde{Z}_{n,it}$ of $\widetilde{\mathbf{Z}}_{n,t}$ (**volatility of level-idiosyncratic components**).

... two panels of residuals, thus, affected by, and hence containing information on, the same market volatility shocks we are inerested in.

Those two (large) panels of residuals have to be analyzed jointly (a two-block-panel analysis).

That two-block analysis is the object of **Step 2** of our two-step factor model analysis.

3. Common and idiosyncratic volatility/Market volatility

Before turning to the problem of analyzing panels in the presence of blocks, we need choosing a volatility proxy. Denoting by e_t some univariate residual (the estimated value of some uncorrelated zero-mean white noise), any monotone increasing function s_t of e_t^2, in principle, could serve as a proxy.

For simplicity, we followed the literature and chose

$$s_{it} := \log(e_{it}^2) \quad \text{and} \quad w_{it} := \log(v_{it}^2)$$

as proposed by Engle and Marcucci (2006).

Feel free, however, to use your favorite one; we tried several ones on empirical data, and the results are only marginally affected.

- After due centering, the $\{s_{n,it}\}$'s and $\{w_{n,it}\}$'s then constitute the two $n \times T$ subpanels of a $2n \times T$ panel with block structure.

4. Panels with block structure (two blocks)

Recall the general decomposition into four mutually orthogonal components:

$$X_{it}^{F} = \underbrace{\overbrace{\phi_{F;it} + \psi_{F;it}}^{}}_{\chi_{it}^{F}}\!\!\overbrace{{} + \zeta_{F;it}}^{\chi_{it}^{FG}} \ldots$$

$$X_{it}^{F} = \overbrace{\phi_{F;it} + \psi_{F;it} + \zeta_{F;it}}^{\chi_{it}^{FG}} + \xi_{it}^{FG}, \quad \chi_{it}^{F} = \phi_{F;it} + \psi_{F;it}, \quad \xi_{it}^{F} = \zeta_{F;it} + \xi_{it}^{FG}$$

and

$$X_{jt}^{G} = \overbrace{\phi_{G;jt} + \psi_{G;jt} + \zeta_{G;jt}}^{\chi_{jt}^{FG}} + \xi_{jt}^{FG}, \quad \chi_{jt}^{G} = \phi_{G;jt} + \psi_{G;jt}, \quad \xi_{jt}^{G} = \zeta_{G;jt} + \xi_{jt}^{FG}, \qquad i, j \in \mathbb{N},\ t \in \mathbb{Z}.$$

$\phi_{F;it}$ is F- and G-common: strongly common

$\psi_{F;it}$ is F-common but G-idiosyncratic: weakly F-common

$\zeta_{F;it}$ is F-idiosyncratic but G-common: weakly F-idiosyncratic

ξ_{it}^{FG} is FG-idiosyncratic: strongly idiosyncratic

etc. (same for Germany)

4. Panels with block structure (two blocks)

Statistical analysis:

- identification of q_F, q_G and q via the Hallin-Liška (*JASA* 2007) method

- consistent reconstruction of $\phi_{F;it}$, $\psi_{F;it}$, etc. and estimation of their contributions to the total sum of squares as in Hallin and Liška (*Journal of Econometrics* 2011)

In our case,

the role of France is played by the $\{s_{n,it}\}$'s (originating from the level-common shocks),

the role of Germany by the $\{w_{n,it}\}$'s originating from the level-idiosyncratic shocks).

The strongly common, weakly common and weakly idiosyncratic components all qualify as "market-driven volatilities". In the S&P100 case below, $Q = q_s = q_w = 1$ is identified. Then, the decomposition only has strongly common components and strongly idiosyncratic ones:

$$X_{it}^F = \underbrace{\overbrace{\phi_{F;it} + 0 + 0}^{\chi_{it}^{FG}}}_{\chi_{it}^F} + \underbrace{\xi_{it}^{FG}}_{\xi_{it}^F} \quad \text{and} \quad X_{jt}^G = \underbrace{\overbrace{\phi_{G;jt} + 0 + 0}^{\chi_{jt}^{FG}}}_{\chi_{jt}^G} + \underbrace{\xi_{jt}^{FG}}_{\xi_{jt}^G}, \quad i, j \in \mathbb{N},\ t \in \mathbb{Z}.$$

That decomposition follows from the joint analysis—no need to consider the blocks separately anymore.

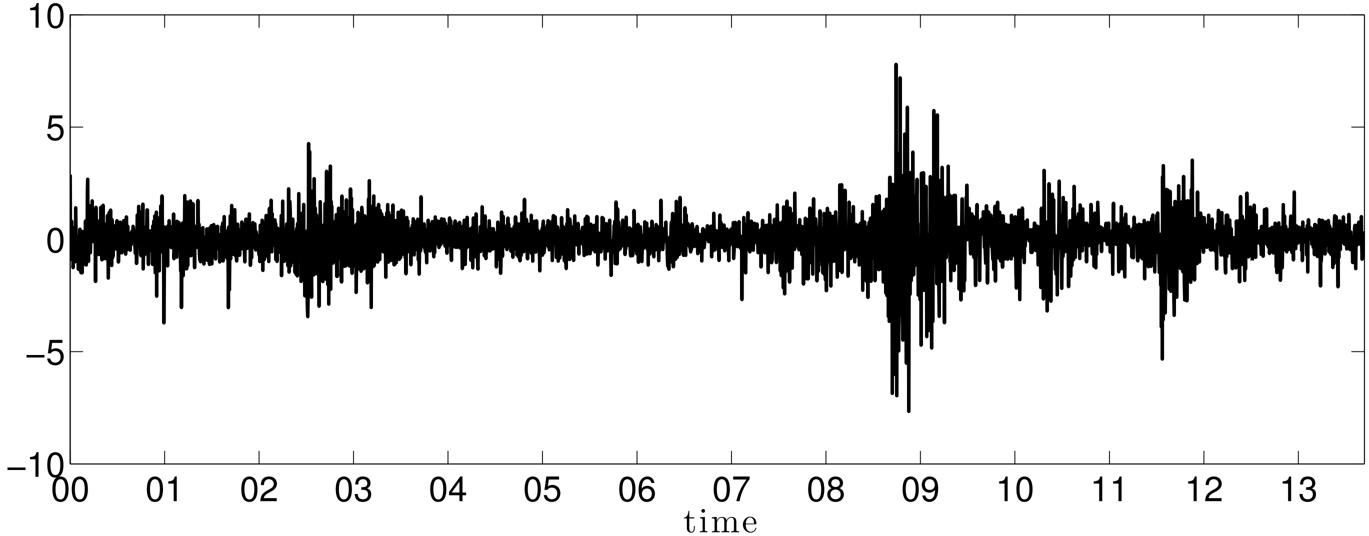
5. Application: S&P100

We illustrate the method by an application to the S&P100 series : $n = 90$ series (some stocks were not traded, and were removed from the analysis) of daily log-returns observed between January 2000 and September 2013 (T =3457).

Step 1. A FACTOR MODEL ANALYSIS OF THE LEVELS Y_{it}

- a number $q = 1$ of dynamic factors is identified via the Hallin-Liška (JASA 2007) method
- the one-sided method of Forni-Hallin-Lippi-Zaffaroni (*Journal of Econometrics* 2015, 2017) yields (a reconstruction of) the level-common components X_{it}, their shocks e_{it}, and the level-idiosyncratic $\tilde{Z}_{it}$
- univariate AR models (orders selected via AIC or BIC) are fitted to the $\tilde{Z}_{it}$'s, yielding residuals v_{it}
- the volatility proxies $\{s_{n,it}\}$ are computed from the level-common shocks e_{it}
- the volatility proxies $\{w_{n,it}\}$ are computed from the level-idiosyncratic shocks v_{it}

The market shocks u_t^T on returns, period 2000–2013.



See

- the dot–com bubble, the Enron (late 2001) and Worldcom (mid–2002) scandals
- the 2003 Iraq war
- the Great 2008–2009 Financial crisis starting with Lehman Brothers bankruptcy (September 2008);
- the 2010–2012 euro sovereign bond crisis.

The largest shocks over the period, by far, are those related with the 2008–2009 financial crisis.

It is interesting to compute the ratios between the sum of the (empirical) variances of the estimated common components $\mathbf{X}_t^T$ to the sum of the (empirical) variances of the observed returns:

$$R^2_{Y.\text{market}} := \frac{\sum_{i=1}^n \sum_{t=1}^T (X_{it}^T)^2}{\sum_{i=1}^n \sum_{t=1}^T (Y_{it})^2}. \approx 0.36$$

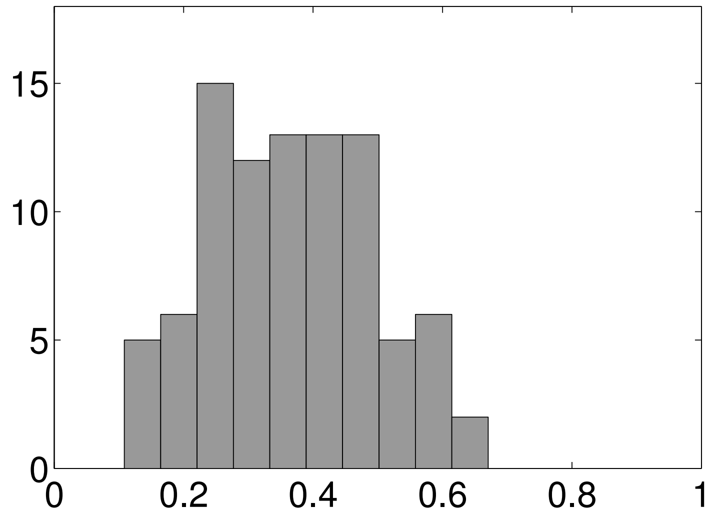
(an average (over the panel) of the relative contribution of the common shocks) and also

$$R^2_{Y_i.\text{market}} := \frac{\sum_{t=1}^T (X_{it}^T)^2}{\sum_{t=1}^T (Y_{it})^2}, \quad i = 1, \ldots, n$$

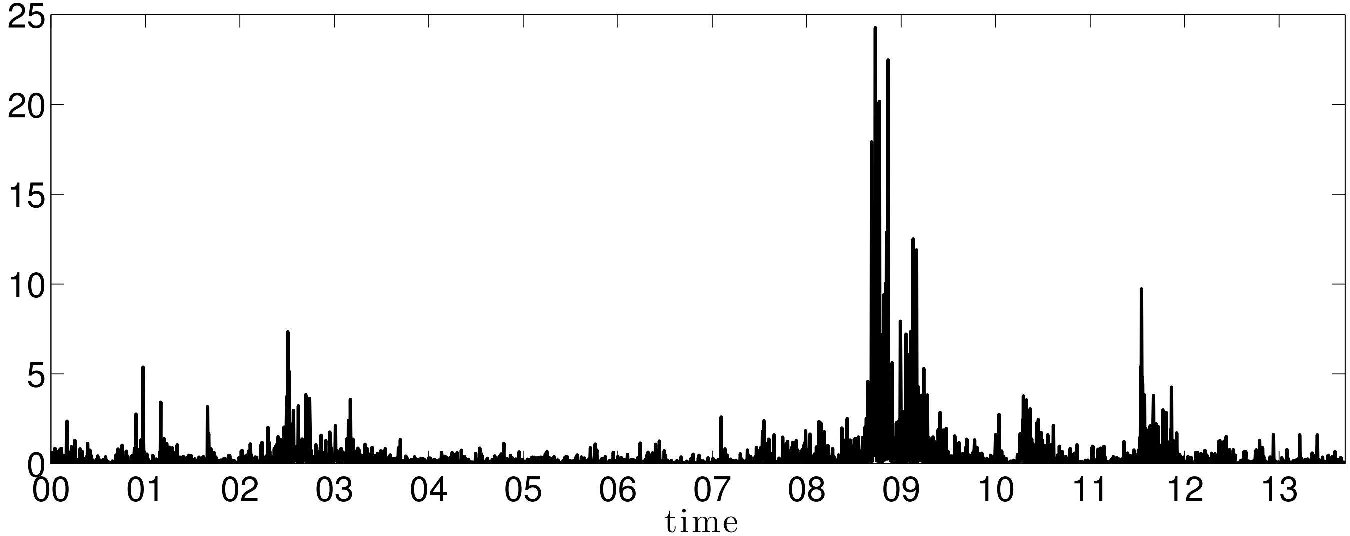
a stock-specific measure of the relative contribution of the common shocks (averaged over time), and

$$R^2_{Y_t.\text{market}} := \frac{\sum_{i=1}^n (X_{it}^T)^2}{\sum_{i=1}^n (Y_{it})^2}, \quad t = 1, \ldots, T,$$

measuring the evolution through time of the average (over the panel) contribution of common shocks.



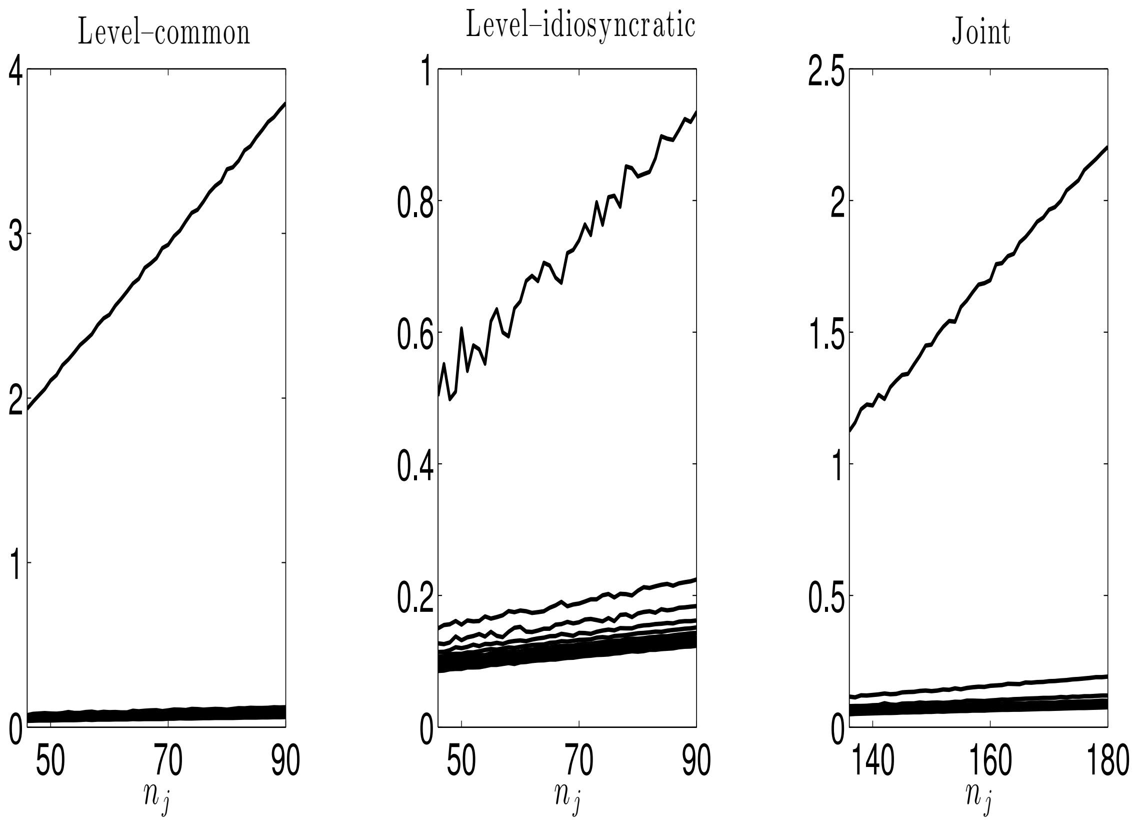
Histogram for the proportions $R^2_{Y_i.\text{market}}$ of variance explained by the market shocks to returns across the panel.



Time series of the proportions $R^2_{Y_t.\text{market}}$ of variance explained by the market shocks to returns at time t.

Step 2. A 2-BLOCK FACTOR MODEL ANALYSIS OF THE VOLATILITY PROXIES $\{s_{n,it}\}$ AND $\{w_{n,it}\}$

Evidence of factor structure in the volatility proxy panels.



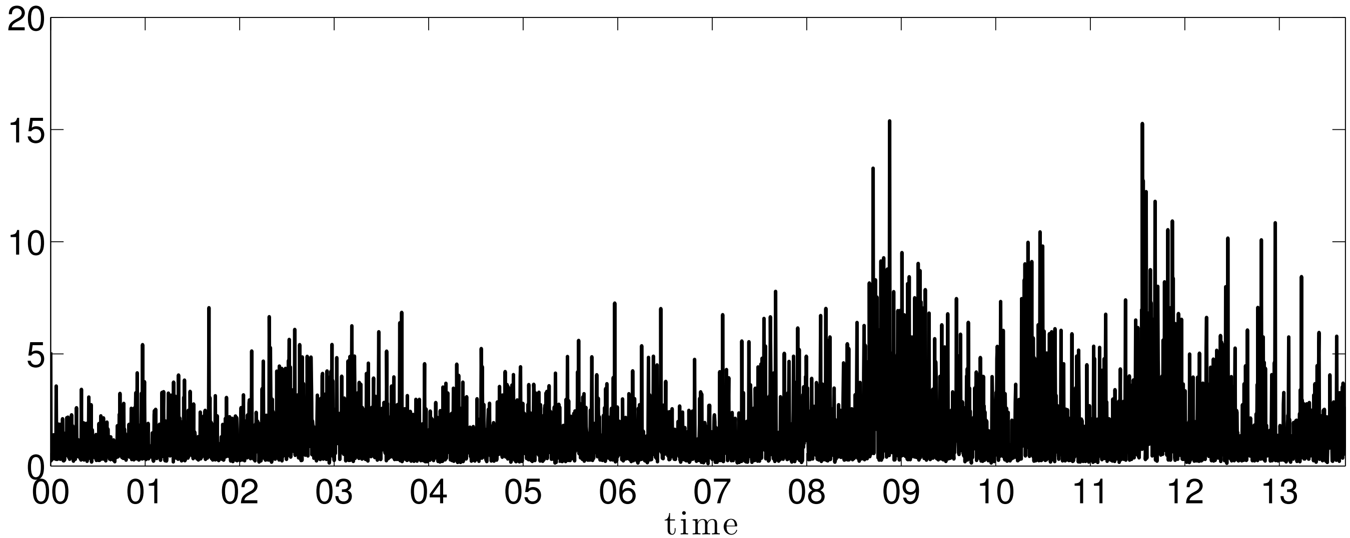
Ten largest dynamic eigenvalues, averaged over frequencies, computed for panels of increasing sizes: $45 \leq n_j \leq n = 90$ for the level–common and level–idiosyncratic volatility panels, and $135 \leq n_j \leq 2n = 180$ for the joint volatility panel.

Step 2. A 2-BLOCK FACTOR MODEL ANALYSIS OF THE VOLATILITIES $\{s_{n,it}\}$ AND $\{w_{n,it}\}$

• the following numbers of dynamic factors are identified via the Hallin-Liška (JASA 2007) method: $q_s = 1$, $q_w = 1$, $q_{sw} = 1$.

• This again implies that a unique volatility-strongly-common shock is driving both the level-common s_{it}'s and the level-idiosyncratic w_{it}'s: no weakly common nor weakly idiosyncratic components here, which greatly simplifies the analysis (a standard FHLZ approach to the $2n$-dimensional panel is sufficient)

• That common shock thus qualifies as *the* market volatility shock, impacting both the level-common and level-idiosyncratic components of the S&P100 panel, with different strengths, though



Estimated market shock $\exp(\varepsilon_t^T)$ on volatilities, period 2000–2013.

Note

- 01 the dotcom bubble
- 03 Iraq war
- 09 is the Great Financial Crisis (which started in 2008)
- 11-12 is the Eurocrisis

5. Application: S&P100

The overall contribution of market shocks to the variances of the volatility proxies s_{it} and w_{it} can be evaluated by means of the ratios

$$R^2_{s.\text{market}} := \frac{\sum_{t=1}^{T} \sum_{i=1}^{n} (\phi^T_{\mathbf{s;it}})^2}{\sum_{t=1}^{T} \sum_{i=1}^{n} (s^T_{it})^2} \approx 0.60$$

and

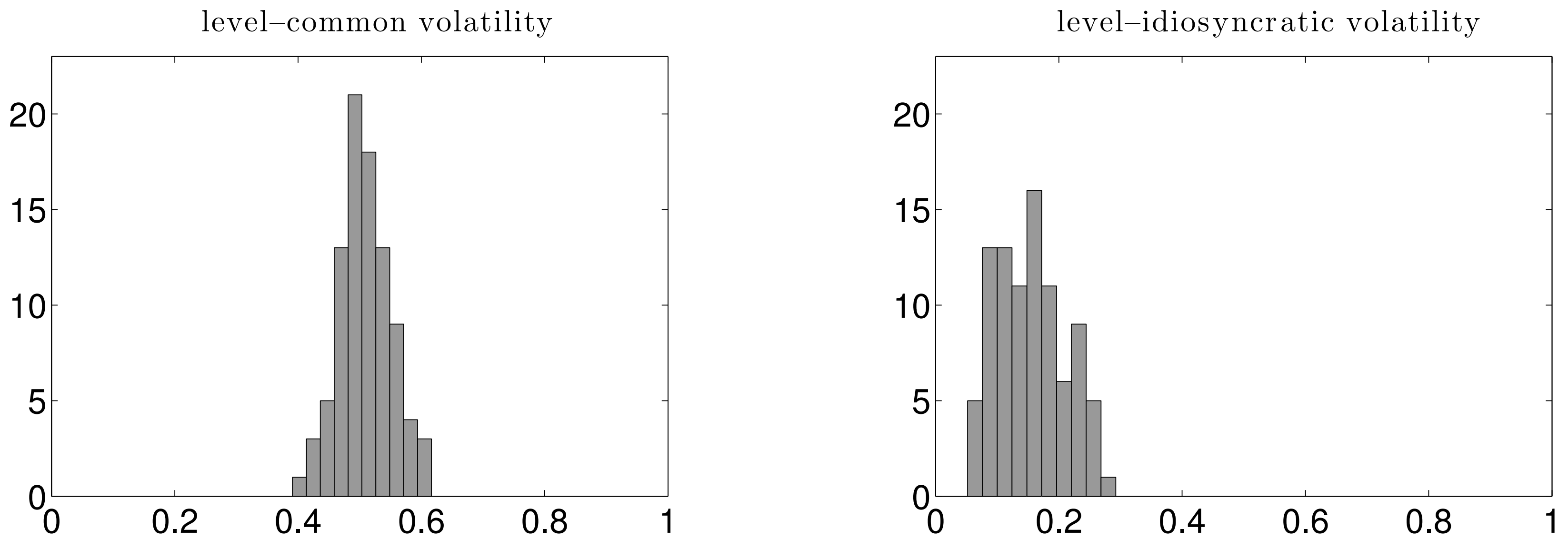
$$R^2_{w.\text{market}} := \frac{\sum_{t=1}^{T} \sum_{i=1}^{n} (\phi^T_{\mathbf{w;it}})^2}{\sum_{t=1}^{T} \sum_{i=1}^{n} (s^T_{it})^2} \approx 0.17$$

For each individual stock i, a measure of the same impact is

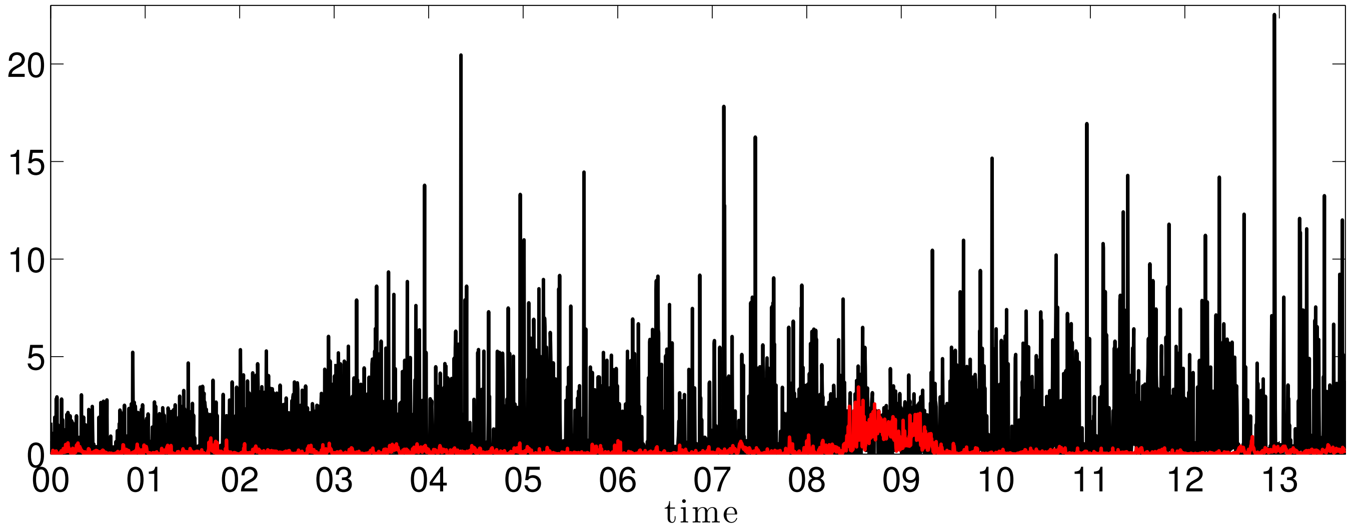
$$R^2_{s_i.\text{market}} := \frac{\sum_{t=1}^{T} (\phi^T_{\mathbf{s;it}})^2}{\sum_{t=1}^{T} (s^T_{it})^2} \quad \text{and} \quad R^2_{w_i.\text{market}} := \frac{\sum_{t=1}^{T} (\phi^T_{\mathbf{w;it}})^2}{\sum_{t=1}^{T} (s^T_{it})^2}, \quad i = 1, \ldots, n;$$

while their evolution through time is captured by

$$R^2_{s_t.\text{market}} := \frac{\sum_{i=1}^{n} (\phi^T_{\mathbf{s;it}})^2}{\sum_{i=1}^{n} (s^T_{it})^2} \quad \text{and} \quad R^2_{w_t.\text{market}} := \frac{\sum_{i=1}^{n} (\phi^T_{\mathbf{w;it}})^2}{\sum_{i=1}^{n} (s^T_{it})^2}, \quad t = 1, \ldots, T.$$



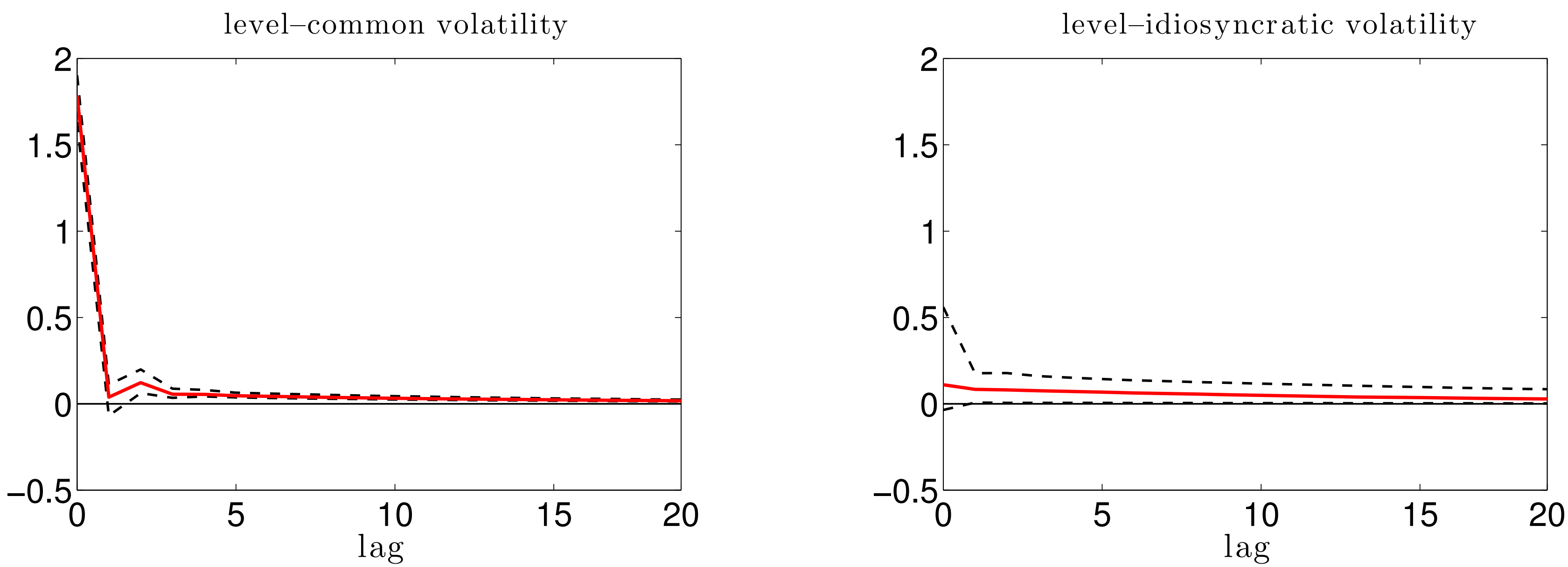
Histograms for the proportions of variances explained by the market volatility shocks across the panel: $R^2_{s_i.\text{market}}$ (left) and $R^2_{w_i.\text{market}}$ (right).



Time series of the estimated proportions $R^2_{s_t.\text{market}}$ (black) and $R^2_{w_t.\text{market}}$ (red) of variances explained by the market volatility shocks.

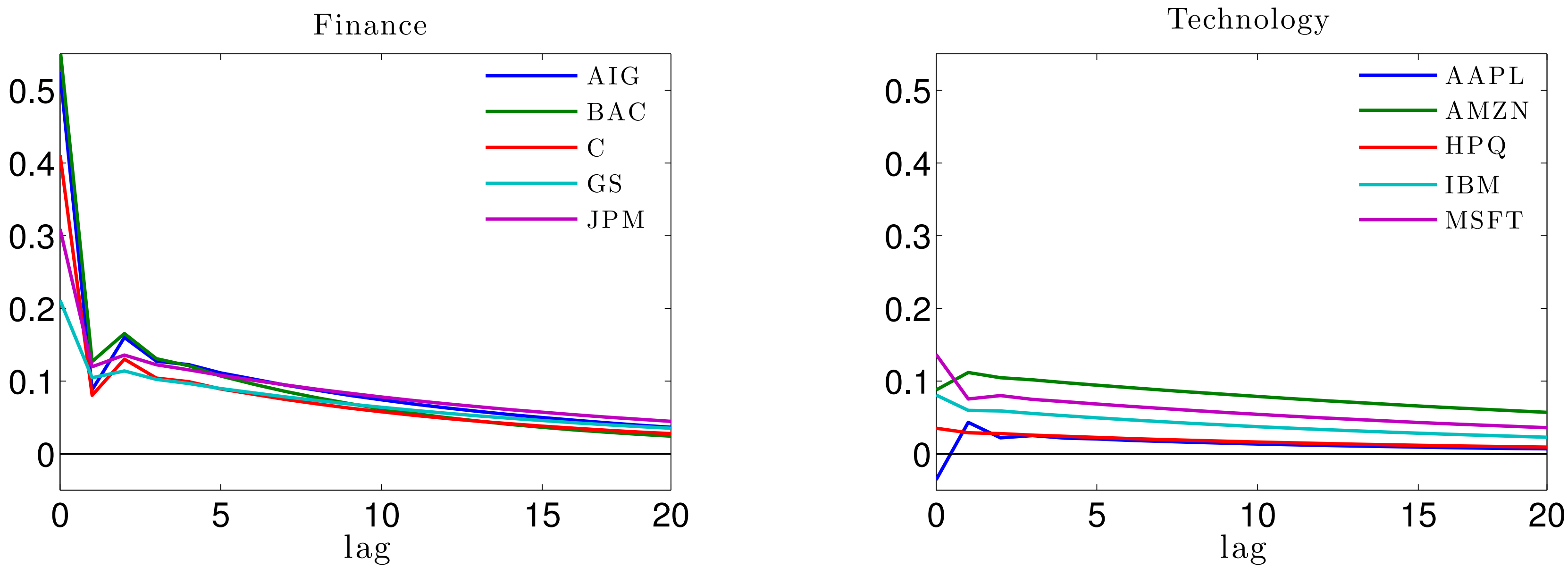
5. Application: S&P100

The transfer or impulse–response functions describing the dynamic loading, by the volatility proxies, of the market volatility shocks. For each stock i, those functions take the form of scalar filters (one for s_{it}, another one for w_{it}), plotted as sequences of coefficients associated with the various lags.



Median, maximum, and minimum of the distribution of impulse–response functions of volatilities to a one–standard-deviation market volatility shock, that is, the sequence of loading coefficients divided by the standard error of the shocks, for level–common (left) and level–idiosyncratic (right) volatilities, respectively.

The right-hand figure, showing a distinctive differentiation of stock behaviors, is the interesting one (the volatility risks associated with stocks yielding low impulse-response coefficients can be diversified away).



Impulse–response functions of volatilities to a one–standard-deviation market volatility shock, that is, the sequence of loading coefficients divided by the standard error of the shocks, for level–idiosyncratic volatilities of selected stocks from the Financial (left) and Technology (right) sectors, respectively.

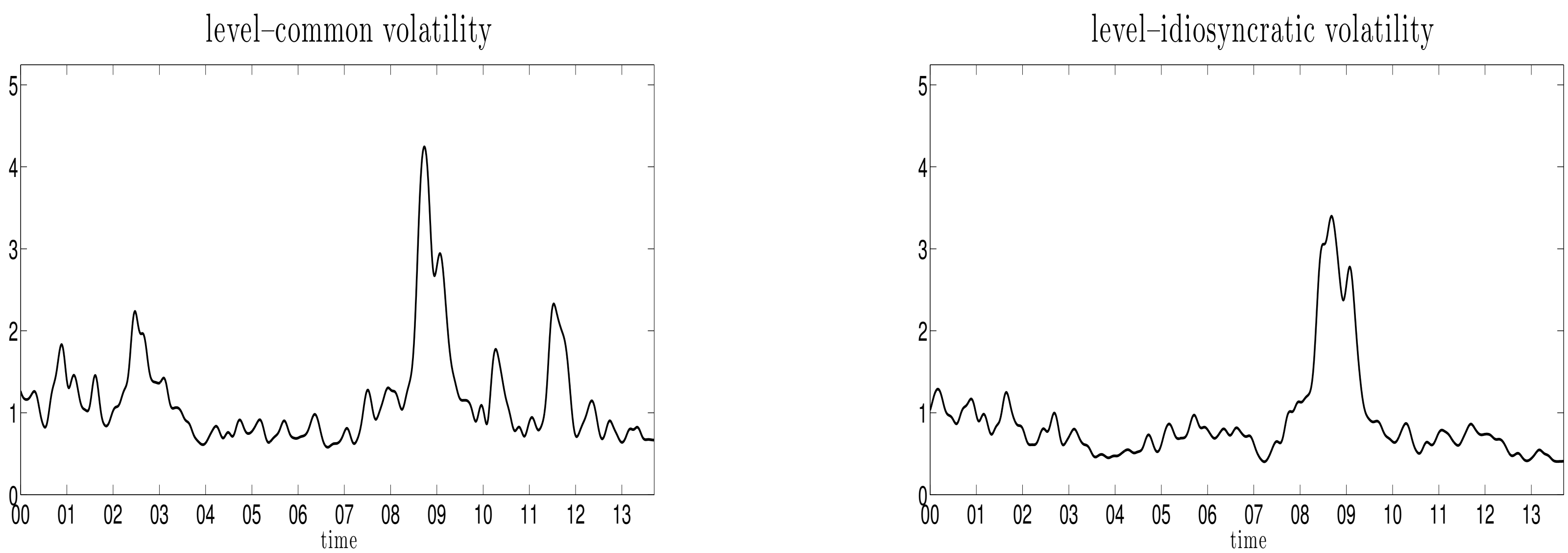
The Technology sector offers more opportunities for diversification.

5. Application: S&P100

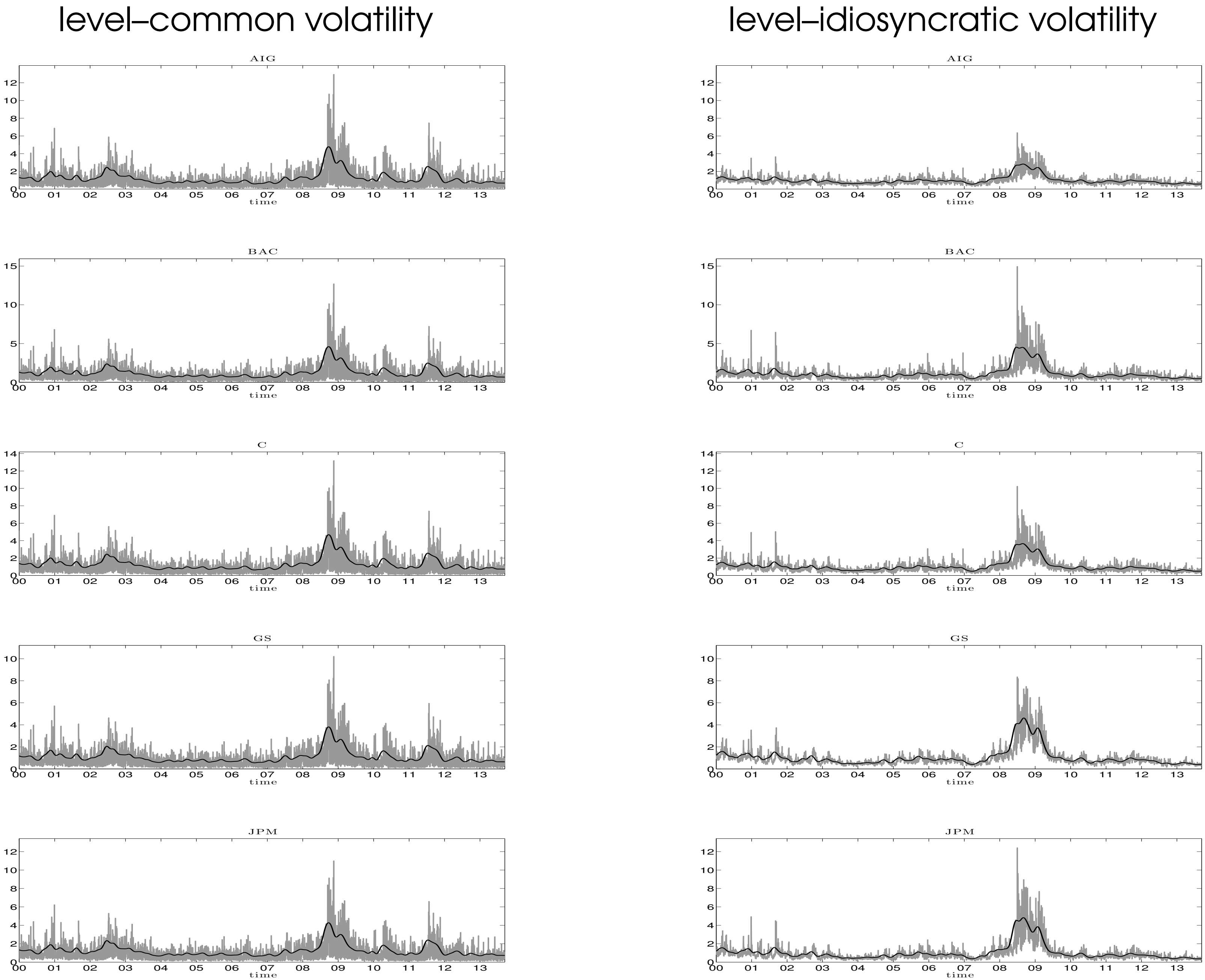
Finally, to conclude, we turn to the analysis, for a few selected stocks, of the market-driven volatilities, which we define as

$$\chi^T_{\mathbf{e^2};\mathbf{it}} := \exp(\phi^T_{\mathbf{s};\mathbf{it}} + \bar{s}^T_{it}), \quad \chi^T_{\mathbf{v^2};\mathbf{it}} := \exp(\phi^T_{\mathbf{w};\mathbf{it}} + \bar{w}^T_{it}), \quad i = 1, \ldots, n, \quad t = 1, \ldots, T, \mathbb{N}$$

where $\bar{s}^T_{it}$ and $\bar{w}^T_{it}$ stand for empirical means.



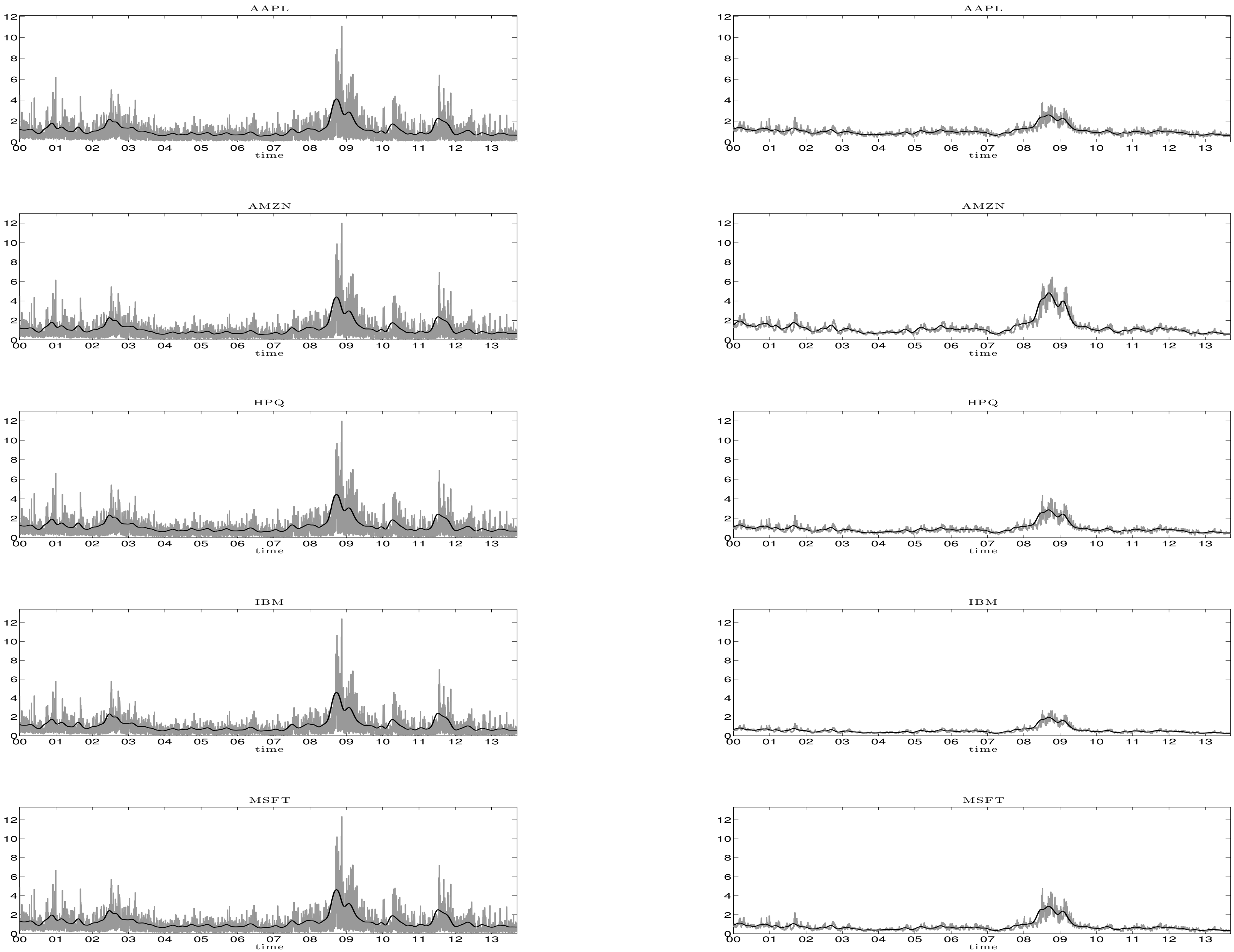
Kernel-smoothed cross–sectional averages of market volatilities. The bandwidth used corresponds to 3 weeks of trading (15 days).



Estimated market volatilities for five selected stocks from the Financial sector, along with their smoothed versions (black solid line).

level–common volatility

level–idiosyncratic volatility



Estimated market volatilities for five selected stocks from the Technological sector, along with their smoothed versions (black solid line).

6. Conclusions and perspectives

- Dynamic factor methods can be applied to volatilities in high-dimensional time series (in large panels of stocks)

- contrary to most existing methods for the analysis of volatility, they are fully nonparametric and model-free: curse of dimensionality turns into a blessing!

- the decompositions between "level-common" and "level-idiosyncratic" on one hand, between "volatility-common" and "volatility-idiosyncratic" in general do not coincide: common volatility shocks do affect level-idiosyncratic components as well as the level-common one;

- dynamic portfolio optimization—where diversification is the main issue—should take into account the market impact on the volatilities of the level-idiosyncratic components as well as their impact on the level-common ones;

- the particular case of a single market shock ($q_s = 1$, $q_w = 1$, $q_{sw} = 1$) is particularly simple, and might well be the rule in financial panels: to be checked against other datasets;

- our approach opens the door to volatility prediction and portfolio optimization without curse of dimensionality in large panels of stock returns.

References

Barigozzi, M. and M. Hallin (2016). Generalized dynamic factor models and volatilities: recovering the market volatility shocks. *Econometrics Journal* **19**, 33-60.

Barigozzi, M. and M. Hallin (2017). Generalized dynamic factor models and volatilities: estimation and forecasting, *Journal of Econometrics* **201**, 307–321.

Barigozzi, M., M. Hallin, and S. Soccorsi (2017). Identification of global and national shocks in international financial markets via general dynamic factor models. Available at SSRN: https://ssrn.com/abstract=2937243 or http://dx.doi.org/10.2139/ssrn.2937243 .