**Dealing with Large Covariance Matrices: Some Lessons from Large Factor Models**

**Lorenzo Trapani, University of Nottingham**
**lorenzo.l.trapani@gmail.com**

# 1 Introduction

We'll be talking about (determining, but not only, the number of factors in) static approximate factor models, viz.

$$X_{i,t} = \phi_i' F_t + u_{i,t}, \qquad 1 \leq i \leq N, 1 \leq t \leq T, \tag{1}$$

where $\phi_i$ and $F_t$ are column vectors of dimension $k$.

It is an important topic:

- Factor models are a *dimension reduction technique* that has been around in the statistical literature since almost one hundred years
  - they were originally introduced in the field of psychometrics;
  - recent availability of large datasets made them increasingly popular;
  - nowadays commonly used by public and private institutions as central banks and investment banks for analysing large panels of time series;
  - you can think of several situations in which a factor model may be useful in economics: just think of a case in which variables are driven by few common factors, representing comovements, and idiosyncratic terms, representing e.g. measurement errors or individual/local features:
    - GDP or inflation could be driven by few factors representing the business cycle plus some measurement errors;
    - equity returns which are driven by few factors representing the market effect plus some factors specific of a given company or sector;
    - asset pricing, where extra returns are driven by some factors which the literature has tried to identify;
    - forecasting, when you have too many predictors and want to squeeze them into few ones - e.g. with inflation;
- starting from the seminal contribution by Chamberlain and Rothschild (1983), many contributions have focused on the case of panel data, where both $N$ and $T$ are large - see, *inter alia*, the review of Bai and Ng (2008).
- inference
  - certainly one needs to estimate $\phi_i$ and $F_t$: there are techniques (e.g. PCA, Kalman filtering, etc...) and I am happy to provide references;
  - but, the first step in the analysis of (1) is, arguably, the determination of the number of common factors, $k$. All techniques are usually based on a well-established fact: the first $k$ eigenvalues of the covariance matrix of the $X_{i,t}$s diverge to infinity whereas the other ones stay bounded. Two main approaches have been developed.

* the first one is based on finding a threshold for the eigenvalues of the covariance matrix of the $X_{i,t}$s, which can be used to decide which eigenvalues are finite and which ones are not; the information criteria proposed by Bai and Ng (2002) belong in this category;
* the second possible approach is based on computing the ratio of adjacent eigenvalues, again exploiting the fact that such ratio eventually diverges: this is the rationale employed by Onatski (2009, 2012) and Ahn and Horenstein (2013), *inter alia*.

Neither approach is free from problems.

* eigenvalue thresholding requires the choice of a penalty function, as is customary in the context of information criteria (see Bai and Ng, 2002). However, such choice is not unique:
  - it affects at least the finite sample properties of the estimated $k$; note however that, building on an idea in Hallin and Liska (2007), Alessi, Barigozzi and Capasso (2010) propose a robust, data-driven methodology to tune the choice of the penalty function which works very well in simulations;
  - existing techniques also require comparing the goodness of fit of different versions of (1), for $1 \leq k \leq k_{\max}$; results seem to be rather sensitive to the specification of the upper bound $k_{\max}$ for at least some of the proposed approaches (see the Monte Carlo evidence in Ahn and Horenstein, 2013);
* the use of the eigenvalues ratio ameliorates such arbitrariness
  - nonetheless, existing contributions make extensive use of (large) random matrix theory (see Bai, 1999, for a complete and insightful review), which requires several constraints on the form and amount of serial and cross sectional dependence;
  - a standard requirement is that the sample sizes $N$ and $T$ are not too different from each other, usually assuming that, as $\min\{N,T\} \to \infty$, $\frac{N}{T} \to c \in (0,\infty)$. This is not always desirable, and examples include such diverse fields as

  **a.** accounting (where data are often recorded on an annual basis and are available for many companies, but for a limited number of years),

  **b.** finance (where e.g. data on hedge funds performance are available for thousands of funds, which are live for a relatively short span),

  **c.** microeconometrics with firm level data,

  **d.** marketing studies (where revealed preferences are often recorded over a limited period of time for many consumers),

  **e.** genomics (where usually thousands of genomes are observed for tens of patients)

Some vocabulary is needed here:

- static vs. dynamic factor models: $F_t$ affects $X_t$ only contemporaneously, as opposed to
$$X_{i,t} = \phi_i' (1 - \rho L) F_t + u_{i,t};$$
- exact vs approximate: the covariance structure of the idiosyncratic needs not be diagonal:
$$E(u_{i,t} u_{j,s}) \neq 0;$$
- curse vs blessing of dimensionality: unless we are in an exact factor model, we cannot estimate (1) unless $N \to \infty$.

# 2 Determining $k$: stationary data, but a more general technique

The idea is always the same.

Let $X_t \equiv [X_{1,t}, ..., X_{N,t}]'$, and make it zero mean for simplicity. Then it is well known that, if you take the $p$-th eigenvalue (say $\lambda^{(p)}$) of $E(X_t X_t')$

$$\lambda^{(p)} \to \begin{array}{c} \infty \\ < \infty \end{array} \text{ according as } \begin{array}{c} p \leq k \\ p > k \end{array}.$$

So, in principle, one could test, for $p = 1, ...$, for the null that the $p$-th eigenvalue (say $\lambda^{(p)}$) of $E(X_t X_t')$ diverges to positive infinity, versus the alternative that it is bounded:

$$\begin{cases} H_0 : \lambda^{(p)} \to \infty \\ H_A : \lambda^{(p)} < \infty \end{cases}. \tag{2}$$

The tests can then be employed as part of a "sequential" procedure to determine $k$.

## 2.1 Assumptions and preliminary theory

Consider the matrix form of (1)

$$X_t = \Phi F_t + u_t; \tag{3}$$

in (54), $u_t \equiv [u_{1,t}, ..., u_{N,t}]'$ and $\Phi \equiv [\phi_1 | ... | \phi_N]'$.

As promised, we assume, without loss of generality

- that the data have mean zero,
- and also that common factors and idiosyncratic errors are orthogonal (typical in this literature).

**Assumption 1**. It holds that *(i)* $E\left(X_{i,t}\right)=0$ for $1\leq i\leq N$ and $1\leq t\leq T$; *(ii)* $E\left(F_{j,t}u_{i,t}'\right)=0$ for $1\leq j\leq k$, $1\leq i\leq N$ and $1\leq t\leq T$.

By Assumption 1, $T^{-1}\sum_{t=1}^{T}E\left(X_{t}X_{t}'\right)\equiv\Sigma_{X}=\Phi\Sigma_{F}\Phi'+\Sigma_{u}$, having defined $\Sigma_{F}\equiv T^{-1}\sum_{t=1}^{T}E\left(F_{t}F_{t}'\right)$ and $\Sigma_{u}\equiv T^{-1}\sum_{t=1}^{T}E\left(u_{t}u_{t}'\right)$. The following notation will also be used extensively henceforth: the $p$-th largest eigenvalue of $\Sigma_{X}$ is denoted as $\lambda^{(p)}$; the $p$-th eigenvalue of $\Phi\Sigma_{F}\Phi'$ as $\gamma^{(p)}$; and, finally, the $p$-th eigenvalue of $\Sigma_{u}$ as $\omega^{(p)}$.

**Assumption 2**. It holds that *(i)* $\gamma^{(p)}=m_{p}N$ for $1\leq p\leq k$ and some $m_{p}>0$; *(ii)* $\omega^{(p)}\leq M$ for all $1\leq p\leq N$; *(iii)* $N^{-1}\sum_{i=1}^{N}\gamma^{(p)}\leq M$ for all $N$.

Assumption 2 adds some structure to the spectra of $\Phi\Sigma_{F}\Phi'$ and $\Sigma_{u}$:

- similar, in spirit, to Assumptions 4, 5 and 8 in Forni, Giannone, Lippi and Reichlin (2009);
- the $\omega^{(p)}$s must be finite;
  - however, they do not need to be distinct or bounded away from zero, and some or all of them could indeed be zero.
- lots of the $\gamma^{(p)}$s are zero of course. The ones who are not
  - diverge as $N\to\infty$, proportionally to $N$;
  - this requires that $N^{-1}\Phi'\Phi$ tends to a positive definite matrix;
  - often called "strong" or "pervasive" factors. We could anyway study the more general case of $\gamma^{(p)}=m_{p}N^{1-\nu_{p}}$ with $\nu_{p}\in[0,1)$.
- Assumption 2 does not require that the $\lambda^{(p)}$s be distinct, or that the diverging eigenvalues be well-separated, which are typical requirement in this literature (see e.g. Wang and Fan, 2016, and also Forni, Giannone, Lippi and Reichlin, 2009).

The following well-known result characterizes the eigenvalues of $\Sigma_{X}$.

**Lemma 1.** *Let $c^{(p)}$ be a set of nonnegative finite numbers, which are strictly positive for $p\leq k$. Then, under Assumptions 1 and 2(i)-(ii), it holds that, as $N\to\infty$*

$$\begin{cases}\frac{\lambda^{(p)}}{N}\to c^{(p)}\ \textit{for}\ 1\leq p\leq k\\ \lambda^{(p)}\to c^{(p)}\ \textit{for}\ k+1\leq p\leq N\end{cases}. \tag{4}$$

*Further, define*

$$\overline{\lambda}_{N}\equiv\frac{1}{N}\sum_{p=1}^{N}\lambda^{(p)}; \tag{5}$$

*under Assumptions 1 and 2, it holds that*

$$\begin{cases} \limsup_{N\to\infty} \overline{\lambda}_N = \overline{\lambda}^{\sup} < \infty \\ \liminf_{N\to\infty} \overline{\lambda}_N = \overline{\lambda}^{\inf} > 0 \end{cases} . \tag{6}$$

According to Lemma 1, $\lambda^{(p)}$ either diverges at a rate $O(N)$, or it converges to a finite constant (which may well be equal to zero) according as $p \leq k$ or not. Basically, the behaviour of the eigenvalues of $\Sigma_X$ as $N$ passes to infinity is the same as that of the eigenvalues of $\Phi\Sigma_F\Phi'$.

## 2.2 Estimation of $\lambda^{(p)}$

Consider $\widehat{\Sigma}_X \equiv \frac{1}{T}\sum_{t=1}^{T} X_t X_t'$, and let $\widehat{\lambda}^{(p)}$ denote the $p$-th largest eigenvalue of $\widehat{\Sigma}_X$. In order to derive the asymptotics of $\widehat{\lambda}^{(p)}$, we need the following assumption.

**Assumption 3.** It holds that *(i)* $E|X_{i,t}|^{4+\epsilon} \leq M$ for $1 \leq i \leq N$, $1 \leq t \leq T$ and some $\epsilon > 0$; *(ii)*

$$E \max_{1\leq \tilde{t}\leq T} \left[ \sum_{t=1}^{\tilde{t}} X_{h,t}X_{j,t} - E\left(X_{h,t}X_{j,t}\right) \right]^2 \leq MT$$

for $1 \leq h, j \leq N$.

Assumption 3

- requires 4-th moments (lots but not that much after all);
- it constrains the amount of serial correlation that one can have in the process $\{X_{h,t}X_{j,t}\}_{t=1}^{T}$ - and therefore, albeit indirectly, in $X_{i,t}$.
  - so, $X_{i,t}$ does not need to be independent across $t$, which is a requirement in "classical" Random Matrix Theory (see Bai, 1999).
- it is on the $X_{i,t}$s, as opposed to considering the unobservable quantities $F_t$ and $u_{i,t}$ (see however Forni, Giannone, Lippi and Reichlin, 2009). In this respect, Assumption 3*(ii)*, on account of its involving observable quantities only, should be easier to understand and verify;
- examples could be
  - the data are independent;
  - assume that $X_{i,t}$ is a stationary process with the representation $X_{i,t} = f_i(\varepsilon_{i,t}, \varepsilon_{i,t-1}, ...)$ for some measurable function $f_i : \mathbb{R}^\infty \to \mathbb{R}$, an *i.i.d.* sequence $\{\varepsilon_{i,t}\}$, with

$$\left\| X_{i,t} - E\left(X_{i,t}|F_i^{s,t}\right) \right\|_2 \leq c_{i,t}s^{-\varrho_i}, \tag{7}$$

where $\mathrm{F}_{i}^{s,t}$ is the $\sigma$-field generated by $\{\varepsilon_{i,t}, \varepsilon_{i,t-1}, ..., \varepsilon_{i,t-s}\}$, $\|\cdot\|_2$ denotes the $L_2$-norm and $c_{i,t}$ is a sequence of non-negative numbers. Examples: linear processes, ARCH and GARCH processes and data from dynamical systems and Volterra series (see Davidson, 2002, *inter alia*).

The rate of convergence of $\widehat{\lambda}^{(p)}$ is in the following lemma.

**Lemma 2.** *Under Assumptions 1 and 3, it holds that*

$$\widehat{\lambda}^{(p)} = \lambda^{(p)} + O_{a.s.}\left[\frac{N}{\sqrt{T}}\left(\ln^{1+\epsilon} N\right)\left(\ln^{\frac{1+\epsilon}{2}} T\right)\right], \tag{8}$$

*for* $1 \leq p \leq \min\{N, T\}$, *where* $\epsilon > 0$.

Lemma 9

- contains a strong rate for $\widehat{\lambda}^{(p)} - \lambda^{(p)}$ - almost sure is needed here, not just a technical sophistication;
- valid for any combination of $N$ and $T$, and for all estimated eigenvalues;
- does not require any assumptions on the $\lambda^{(p)}$s: these do not need to be distinct or (when they diverge) well-separated; some of the eigenvalues may be equal to zero; and the eigenvalues that diverge do not need to do it at any special rate.

Compare with classical Random Matrix Theory

- where it has been shown that, under the assumptions that $X_{i,t}$ is *i.i.d.* across $i$ and $t$ and that $\frac{N}{T} \to c \in (0, \infty)$, it holds that $\widehat{\lambda}^{(p)} - \lambda^{(p)} = O_{a.s.}(1)$ - see Bai and Yin (1993).
- Lemma 9 illustrates what happens in the presence of common factors, which introduce dependence, and loads thereof
- Wang and Fan (2016; see Theorem 3.1) derive the limiting distribution of $\widehat{\lambda}^{(p)}$ for $1 \leq p \leq k$, showing asymptotic normality at a rate $\frac{N}{\sqrt{T}}$. This suggests that the strong rate in (8) should be optimal, modulo the logarithmic terms, at least for $1 \leq p \leq k$;
  - however that Lemma 9 holds for all eigenvalues, not only for the spiked ones.

## 2.3 The test

We test for

$$\begin{cases} H_0 : \lambda^{(p)} = m_p N \\ H_A : \lambda^{(p)} = m_p < \infty \end{cases},$$

for some $0 < m_p < \infty$ and finite. The more general case of weak factors is also studied.

Let $\beta \equiv \frac{\ln N}{\ln T}$, and define $\delta \in [0, 1)$ such that

$$\delta \begin{cases} > 0 \\ > 1 - \frac{1}{2\beta} \end{cases} \quad \text{according as} \quad \begin{matrix} \beta \leq \frac{1}{2} \\ \beta > \frac{1}{2} \end{matrix} . \tag{9}$$

Finally, consider the following estimator of $\overline{\lambda}_N$

$$\widehat{\overline{\lambda}}_N \equiv \frac{1}{N} \sum_{p=1}^{N} \widehat{\lambda}^{(p)}. \tag{10}$$

We are now ready to introduce the test. Define

$$\varphi^{(p)} \equiv \exp\left\{ N^{-\delta} \frac{\widehat{\lambda}^{(p)}}{\widehat{\overline{\lambda}}_N} \right\}. \tag{11}$$

Then

- under the null that $\lambda^{(p)} = m_p N$, $\varphi^{(p)} \to \infty$ at a rate $\exp\left\{N^{1-\delta}\right\}$;
- conversely, $\varphi^{(p)}$ converges to a finite number under the alternative that $\lambda^{(p)} < \infty$. Let's try to see why...
- note also that $\widehat{\overline{\lambda}}_N$ makes the argument of the exponential scale-free; in principle, any statistic that ensures scale invariance may also be used.

Given that $\varphi^{(p)} \to \infty$ under the null, we cannot use it directly and we instead propose a randomised version of it. We present the construction of the test statistic as a four step algorithm.

**Step 1** Generate an artificial sample $\left\{\xi_j^{(p)}\right\}_{j=1}^{R}$ as *i.i.d.* $N(0, 1)$, and define the sequence $\sqrt{\varphi^{(p)}} \times \xi_j^{(p)}$, $1 \leq j \leq R$;

**Step 2** Define the sample $\left\{\zeta_j^{(p)}(u)\right\}_{j=1}^{R}$ as

$$\zeta_j^{(p)}(u) \equiv I\left[\sqrt{\varphi^{(p)}} \times \xi_j^{(p)} \leq u\right], \tag{12}$$

with $u$ extracted from a distribution $F(u)$ with support $U \subset R \backslash \{0\}$;

**Step 3** Compute

$$\vartheta^{(p)}(u) \equiv \frac{2}{\sqrt{R}} \sum_{j=1}^{R} \left[ \zeta_j^{(p)}(u) - \frac{1}{2} \right]; \tag{13}$$

**Step 4** Define the test statistic

$$\Theta^{(p)} \equiv \int_U \left[ \vartheta^{(p)}(u) \right]^2 dF(u). \tag{14}$$

Here is a heuristic preview

- Under the null,
  - $\varphi^{(p)}$ passes to infinity, so...
  - ... the variance of $\sqrt{\varphi^{(p)}} \times \xi_j^{(p)}$ should be $\infty$, so...
  - ... the *i.i.d.* sequence $\left\{ \zeta_j^{(p)}(u) \right\}_{j=1}^{R}$ follows a Bernoulli distribution with $E\left[ \zeta_j^{(p)}(u) \right] = \frac{1}{2}$, so...
  - ... in (13) a CLT should hold whereby, as $R \to \infty$, $\vartheta^{(p)}(u)$ should be $N(0,1)$.
- Under the alternative,
  - $\varphi^{(p)}$ should remain finite, so...
  - ... it can be expected that, for any $u \neq 0$, $E\left[ \zeta_j^{(p)}(u) \right] \neq \frac{1}{2}$, so...
  - ... in (13), there is a sum of *i.i.d.* random variables with nonzero mean, which by the LLN diverges to positive infinity at a speed $\sqrt{R}$.

Let's formalise.

Notation: $P^*$ is the probability law of $\left\{ \zeta_j^{(p)}(u) \right\}_{j=1}^{R}$ conditional on the sample, and "$\xrightarrow{D^*}$" denotes convergence in distribution according to $P^*$.

**Theorem 3.** *Let Assumptions 1-3 hold. Then, under* $H_0 : \lambda^{(p)} = m_p N$, *as* $\min\{N, T, R\} \to \infty$ *with*

$$R \exp\left\{ -\epsilon N^{1-\delta} \right\} \to 0, \tag{15}$$

*for some* $0 < \epsilon < \frac{m_p}{\overline{\lambda}_N}$, *it holds that* $\Theta^{(p)} \xrightarrow{D^*} \chi_1^2$ *a.s.-*$P^*$ *conditionally on the sample.*

Define $c_\alpha$ such that, as $\min\{N, T, R\} \to \infty$, it holds that $P\left[ \Theta^{(p)} \leq c_\alpha \right] = \alpha$ under $H_0$. The following theorem states that the test is consistent versus the alternative $H_A : \lambda^{(p)} \leq m_p$.

**Theorem 4.** *Let Assumptions 1-3 hold. Under $H_A$, as $\min\{N,T,R\} \to \infty$, it holds that $P\left[\Theta^{(p)} > c_\alpha\right] = 1$ a.s.-$P^*$ conditionally on the sample.*

All very well, but this is a test on the single $p$: how do we determine $k$? Well, we keep testing until we stop (=reject). Take a look

**Step 1** Run the test for $H_0 : \lambda^{(1)} = \infty$ based on $\Theta^{(1)}$. If the null is rejected, set $\widehat{k} = 0$ and stop, otherwise go to the next step.

**Step 2** Starting from $p = 1$, run the test for $H_0 : \lambda^{(p+1)} = \infty$ based on $\Theta^{(p+1)}$, constructed using an artificial sample $\left\{\xi_j^{(p+1)}\right\}_{j=1}^R$ generated independently of $\left\{\xi_j^{(1)}\right\}_{j=1}^R$, ..., $\left\{\xi_j^{(p)}\right\}_{j=1}^R$. If the null is rejected, set $\widehat{k} = p$ and stop; otherwise repeat the step until the null is rejected (or until a pre-specified maximum number, say $k_{\max}$, is reached).

As can be expected, in this context a pivotal role is played by the level of the individual tests, $\alpha$, which should be chosen so that $\widehat{k}$ is a good approximation of $k$, at least asymptotically.

**Theorem 5.** *Let Assumptions 1-3 hold, and define the level of each individual test as $\alpha = \alpha(N,T)$. As $\min\{N,R,T\} \to \infty$ under (15), if $k_{\max} \geq k$ and $\alpha(N,T) \to 0$, then it holds that $P\left[\widehat{k} = k\right] = 1$ a.s.-$P^*$ conditionally on the sample.*

Theorem 5:

- states that $\widehat{k}$ is consistent, as long as the level $\alpha$ of the individual tests is chosen so as to converge to zero: no specific rates are required;
- no requirement on any special choice of $k_{\max}$: as long as this value is "large enough" (that is, as long as $k_{\max} \geq k$), the theorem holds
  - usually the literature uses the Schwert's rule (Schwert, 1989; see also the comments in Bai and Ng, 2002, p. 203), although other choices are also possible. In our context, even the choice $k_{\max} = \min\{N,T\}$ is allowed; indeed, simulations show that the estimation procedure is not sensitive to the choice of $k_{\max}$.

### 2.3.1 Final comments, etc...

But an extremely important side-note. We'll go fast, but it is really important anyway. What if I have weak factors:

- either because they are weak, or
- because only some units are affected?

Formally, eigenvalues no longer diverge as fast as $N$. We can anyway readily study

$$\begin{cases} H_0 : \lambda^{(p)} = m_p N^{1-\nu_p} \\ H_A : \lambda^{(p)} \leq m_p < \infty \end{cases}, \tag{16}$$

for some $\nu_p \in [0, 1)$.

Consider the following extension of Assumption 2.

**Assumption 4**. It holds that *(i)* $\gamma^{(p)} = m_p N^{1-\nu_p}$ with $\nu_p \in [0, 1)$ for $1 \leq p \leq k$ and some $m_p > 0$; *(ii)* $\omega^{(p)} \leq M$ for all $1 \leq p \leq N$; *(iii)* (a) $N^{-1} \sum_{i=1}^{N} \gamma^{(p)} \leq M$ for all $N$, and (b) either $\omega^{(N)} > 0$ or $N^{-1} \sum_{i=1}^{N} \gamma^{(p)} \geq M'$ for all $N$.

Assumption 4 is the same as Assumption 2. The only difference is part *(iii)*(b), which implies a lower bound on $\frac{1}{N} \sum_{p=1}^{N} \lambda^{(p)}$. This holds, for example, if $\Sigma_u$ has full rank.

**Lemma 6.** *Let $c^{(p)}$, $1 \leq p \leq N$, be a sequence of nonnegative finite numbers, which are strictly positive for $p \leq k$. Then, under Assumption 4(i)-(ii), it holds that, as $N \to \infty$*

$$\begin{cases} \frac{\lambda^{(p)}}{N^{1-\nu_p}} \to c^{(p)} \text{ for } 1 \leq p \leq k \\ \lambda^{(p)} \to c^{(p)} \text{ for } k+1 \leq p \leq N \end{cases}. \tag{17}$$

*Further, under Assumptions 1 and 4(i)-(iii), it holds that*

$$\begin{cases} \limsup_{N\to\infty} \overline{\lambda}_N = \overline{\lambda}^{\sup} < \infty \\ \liminf_{N\to\infty} \overline{\lambda}_N = \overline{\lambda}^{\inf} > 0 \end{cases}. \tag{18}$$

Let $\chi^2_{1,1-\alpha}$ denote the $1-\alpha$ upper percentile of the $\chi^2_1$ distribution. The following corollary extends the results in Theorems 3 and 78 to the case of weak factors.

**Corollary 7.** *Let Assumptions 3 and 4 hold, and let* $\min\{N, T, R\} \to \infty$. *Under $H_0 : \lambda^{(p)} = m_p N^{1-\nu_p}$, if it holds that*

$$R \exp\left\{-\epsilon N^{[1-(\nu_p+\delta)]}\right\} \to 0, \tag{19}$$

*for some $0 < \epsilon < \frac{m_p}{\overline{\lambda}_N}$, then it holds that $P\left[\Theta^{(p)} \leq \chi^2_{1,1-\alpha}\right] = \alpha$ a.s.-$P^*$ conditionally on the sample. Further, under $H_A : \lambda^{(p)} \leq M < \infty$, if (15) holds, then $P\left[\Theta^{(p)} > \chi^2_{1,1-\alpha}\right] = 1$ a.s.-$P^*$ conditionally on the sample.*

We find that

- the test is able to accept $H_0$ in this context, with a probability of a Type I error equal to a given level $\alpha$;

- however, this depends on $\nu_p$, and on the relative rate of divergence between $N$ and $T$ as they pass to infinity:
  - able to detect the presence of weak factors, as long as they are not "too weak"; otherwise, they would be indistinguishable from an idiosyncratic component with some degree of cross-correlation.
  - At a minimum, we need $\nu_p + \delta < 1$.
    - case $\beta \leq \frac{1}{2}$: it is required that $\nu_p < 1$. Thus, when $\beta$ approaches zero (and, therefore, when $N$ is much smaller than $T$), this entails that the test is able to detect even very weak factors. This result entails that when there is enough signal ($T$), the dimensionality of the covariance matrix $N$ is less of a problem.
    - case $\beta > \frac{1}{2}$: it is required that $\nu_p < 1 - \frac{1}{2\beta}$: as $\beta$ increases (and, therefore, as $N$ becomes much bigger than $T$), the test is less and less able to detect weak factors.
    - It is interesting to consider the case of $\beta = 1$, which indicates $N$ and $T$ being of comparable magnitude: in such case, (19) requires that $\nu_p < \frac{1}{2}$. Thus, the test can detect weak factors, but only if the corresponding eigenvalue diverges a bit faster than $N^{1/2}$.

Let's look at some numbers. I compared against the methodologies suggested in Bai and Ng (2002; referred to as IC1, IC2, PC1, PC2 below), also considering the refinements developed by Alessi, Barigozzi and Capasso (2010); Onatski (2010; referred to as ON), and Ahn and Horenstein (2013; referred to as ER and GR):

$$
\begin{aligned}
IC1 &= \arg \min_{0 \leq k \leq k_{\max}} \left[ \ln V(k) + C_0 k \frac{N+T}{NT} \ln \left( \frac{NT}{N+T} \right) \right] \\
IC2 &= \arg \min_{0 \leq k \leq k_{\max}} \left[ \ln V(k) + C_0 k \frac{N+T}{NT} \ln \left( \min \{N, T\} \right) \right] \\
PC1 &= \arg \min_{0 \leq k \leq k_{\max}} \left[ V(k) + C_0 \hat{\sigma}^2 k \frac{N+T}{NT} \ln \left( \frac{NT}{N+T} \right) \right] \\
PC2 &= \arg \min_{0 \leq k \leq k_{\max}} \left[ V(k) + C_0 \hat{\sigma}^2 k \frac{N+T}{NT} \ln \left( \min \{N, T\} \right) \right] \\
ON &= \arg \max_{0 \leq k \leq k_{\max}} \left[ k | \widehat{\lambda}^{(k)} > \left( 1 + N^{-1/3} \right) \hat{u} \right] \\
ER &= \arg \max_{0 \leq k \leq k_{\max}} \frac{\widehat{\lambda}^{(k)}}{\widehat{\lambda}^{(k+1)}} \\
GR &= \arg \max_{0 \leq k \leq k_{\max}} \frac{\ln \left[ 1 + \widehat{\lambda}^{(k)} / v(k) \right]}{\ln \left[ 1 + \widehat{\lambda}^{(k+1)} / v(k+1) \right]}
\end{aligned}
$$

where

$$V(k) = \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\left(X_{i,t} - \widehat{\phi}_i'\widehat{F}_t\right)^2,$$

with $\widehat{\phi}_i$ and $\widehat{F}_t$ the estimators of $\phi_i$ and $F_t$ studied in Bai (2003) under exactly $k$ factors. We define $\widehat{\sigma}^2 = V(k_{\max})$, $\widehat{u} = 2.7\widehat{\lambda}^{(k_{\max}+1)} - 1.7\widehat{\lambda}^{(2k_{\max}+1)}$ and $v(k) = \sum_{j=k+1}^{\min\{N,T\}}\widehat{\lambda}^{(j)}$. Finally, considering $IC1$, $IC2$, $PC1$ and $PC2$, we report the best result out of the four criteria (that is, the one which corresponds to the lowest value of $ME$) for the case $C_0 = 1$ (which corresponds to the criteria studied by Bai and Ng, 2002). In their contribution, Alessi, Barigozzi and Capasso (2010) recommend to employ different values of the tuning constant $C_0$, and to evaluate the estimated number of factors over a whole range of values of $C_0$, thereby selecting the optimal one, identified as the value which yields a stable estimate of $k$. We implemented this procedure by searching for the optimal value of $C_0$ over the grid $[0, 13]$, using intervals of width 0.005; results are reported for $PC1$, which was the best performing criterion across all exercises.

Let's look at the numbers, but the messages will be

- no criterion can be the absolute winner;
- the test has excellent power for all cases considered, being able to detect whether $k = 0$ or not;
- otherwise
  - results are good all across the board when there is only serial dependence; $\widehat{k}$ is the best criterion when $N$ is small;
  - when there is cross sectional dependence, conclusions become more mixed:
    - $\widehat{k}$ fares better than the other criteria when $N$ is quite small
    - all estimators fare worse as $k$ increases.

# 3 Determining the dimension of factor structures with nonstationary data

This topic could be useful per se, or as a revision. Sadly, I must change the notation (to be consistent with our paper).

We will consider a methodology to estimate the dimension of the space spanned by the common (nonstationary) factors in a large approximate factor model

$$X_t = \Lambda \mathcal{F}_t + u_t, \tag{20}$$

where $\Lambda$ is an $N \times r$ matrix and $\mathcal{F}_t$ is an $r$-dimensional vector. We will also make use of the scalar version of (20)

$$X_{i,t} = \lambda_i' \mathcal{F}_t + u_{i,t}, \tag{21}$$

with $1 \leq i \leq N$ and $1 \leq t \leq T$.

Although the relevant assumptions are spelt out later on, in (20) we are assuming that there are three possible categories of common factors in the vector $\mathcal{F}_t$:

- factors with a linear trend and an additional, either an $I(1)$ or $I(0)$, zero mean component;
- pure, zero mean $I(1)$ factors with no trends;
- and, finally, stationary common factors.

Each group may well have dimension zero, which means that, for example, no $I(1)$ factors with drift exist, etc. We also assume, throughout the paper, that $u_{i,t} \sim I(0)$ for each $i$.

Equation (20) could be useful:

- in applications:
  - employment fluctuations across 60 industries in the US - Bai (2004);
  - interest rates at different maturities in the US and Canada - Moon and Perron (2007);
  - real macroeconomic data in US - Maciejowska (2010);
  - bilateral exchange rates with US dollar - Engel, Mark, and West (2015).
- also, models with common linear trends have also been used for temperature data in US - Chen and Wu (2017);
- Testing for the number of $I(1)$ factors allows to test for:

- the presence of unit roots in large panels - Bai and Ng (2004);
- the rank of cointegration in large panels - Onatski and Wang (2017).

As we saw before, there are contributions to determine the number of common factors. Most methodologies focus on the case of stationary datasets, though: there is little on nonstationary data:

- we could certinaly use methods for differenced data:
  - information criteria: Bai and Ng (2002);
    - ... but, remember, thresholds are set arbitrarily;
    - ... and cannot test for no-factors;
    - tests: Onatski (2010), Trapani (2017);
      - ... but the nature of the factors is not understandable (differencing destroys all differences).
- we could analyse the level data directly:
  - information criteria: Bai (2004), Maciejowska (2010);
    - ... but same limitations as above;
  - except we can distinguish stationary from non-stationary factors;
  - no test available.

Some ideas on what we do.

We build on the usual idea that eigenvalues of the second moment matrix of the data are tested to verify whether they diverge to infinity as $\min(N,T) \to \infty$, or whether they are bounded. Analyis is based on the sample second moment matrix of $X_t$

$$\sum_{t=1}^{T} X_t X_t'.$$

A heuristic preview of how the procedure works:

- if I have **linear trends**, it can be expected that
  - the sample second moment matrix of $X_t$ will diverge as fast as $T^3$;
  - due to the well known eigenvalue separation property of large factor models, it can be expected that the eigenvalues corresponding to common factors should diverge as fast as $N$

– This suggests to consider the eigenvalues of $T^{-3}\sum_t X_t X_t'$ (denoted as, say, $\nu_1^{(p)}$), and test for

$$\begin{cases} H_0 : \nu_1^{(p)} \to \infty \\ H_A : \nu_1^{(p)} < \infty \end{cases}.$$

as $\min(N,T) \to \infty$; the test can be carried out for $p = 1, 2, ...$, stopping as soon as the null is rejected.

- if I have zero mean, $I(1)$ **common factors**
  - the FCLT suggests that the second moment matrix of $X_t$ will diverge as fast as $T^2$,
  - again the eigenvalues corresponding to the common factors diverge as fast as $N$.
  - Thus, one could study the eigenvalues of $T^{-2}\sum_t X_t X_t'$ (denoted as, say, $\nu_2^{(p)}$), and test for

$$\begin{cases} H_0 : \nu_2^{(p)} \to \infty \\ H_A : \nu_2^{(p)} < \infty \end{cases},$$

as above

These two steps should provide an estimate of the number of common factors which, respectively, have a linear trend and are genuinely zero mean $I(1)$ processes.

- then, of course you can use

$$T^{-1}\sum_t \Delta X_t \Delta X_t'$$

and the previous procedure, to determine the number of total common factors.

## 3.1 Model and assumptions

Recall (20)

$$X_t = \Lambda \mathcal{F}_t + u_t.$$

How many linear trends can I have, first of all?

We show that the number of common factors with a linear trend can be either zero - no common factors with linear trends - or 1. We assume that

$$\mathcal{F}_t = A(d_1 t) + B\psi_t, \tag{22}$$

where $A$ is a nonzero $r \times 1$ vector and $B$ an $r \times r$ matrix; importantly, in (22) $d_1$ is a dummy variable. The $r$-dimensional vector $\psi_t$ has components which are a mixture of $I(0)$ and $I(1)$, with no linear trends.

We consider the following assumption, which ensures that the factors $\mathcal{F}_t$ are fully identified.

**Assumption 1**. It holds that: (i) $A$ is nonzero; (ii) $rank(B) = r$; (iii) the vector $\psi_t$ can be rearranged and partitioned as $[\psi'_{at}, \psi'_{bt}]'$, where $\psi_{at} \sim I(1)$ has dimension $r_2 + d_2$ and $\psi_{bt} \sim I(0)$ has dimension $r_3 + (1 - d_2)$, where $d_2$ is a dummy variable.

**Lemma 8.** *Under Assumption 1, (21) can be equivalently represented as*

$$X_{i,t} = \lambda_i^{(1)} f_t^{(1)} + \lambda_i^{(2)\prime} f_t^{(2)} + \lambda_i^{(3)\prime} f_t^{(3)} + u_{i,t}, \tag{23}$$

*where the common factors are defined by the following equations*

$$f_t^{(1)} = d_1 t + d_2 f_t^{(1)\dagger} + (1 - d_2) g_t, \tag{24}$$

$$f_t^{(1)\dagger} = f_0^{(1)\dagger} + \sum_{j=1}^{t} e_t^{(1)}, \tag{25}$$

$$f_t^{(2)} = f_0^{(2)} + \sum_{j=1}^{t} e_t^{(2)}; \tag{26}$$

*in (24)-(26): $f_t^{(1)}$ is an $r_1 \times 1$ vector, $f_t^{(2)}$ is an $r_2 \times 1$ vector, $f_t^{(3)}$ is an $r_3 \times 1$ vector, and $e_t^{(1)}$, $e_t^{(2)}$, $g_t$ and $f_t^{(3)}$ are $I(0)$.*

Therefore, the number of common factors in $X_{i,t}$ is summarised in the table below.

| Factor type | Number |
|---|---|
| With linear trend | $r_1 d_1$ |
| Zero mean, $I(1)$ | $r_2 + r_1 (1 - d_1) d_2$ |
| Zero mean, $I(0)$ | $r_3 + r_1 (1 - d_1) (1 - d_2)$ |
| Total nonstationary | $r^* = r_1 d_1 + r_2 + r_1 (1 - d_1) d_2$ |
| Total number of common factors | $r = r^* + r_3 + r_1 (1 - d_1) (1 - d_2)$ $= r_1 + r_2 + r_3$ |

We now spell out the main assumptions. Consider the $(r_2 + d_2)$-dimensional vectors $f_t^*$, where $f_t^* = \left[f_t^{(1)\dagger}, f_t^{(2)\prime}\right]'$ or $f_t^* = f_t^{(2)}$, according as $d_2 = 1$ or 0; and

$e_t$, where $e_t = \left[e_t^{(1)}, e_t^{(2)\prime}\right]'$ or $e_t = e_t^{(2)}$, according as $d_2 = 1$ or 0. We define the long-run covariance matrix associated with $f_t^*$ as

$$\Sigma_{\Delta f} = \lim_{T\to\infty} Var\left(T^{-1/2}\sum_{t=1}^{T} e_t\right). \tag{27}$$

**Assumption 2.** Let $\kappa > 0$. It holds that (i) $E\left\|e_t\right\|^{2+\kappa} < \infty$ for all $t$; (ii) $E\left|f_0^{\dagger}\right|^{2+\kappa} < \infty$; (iii) $\Sigma_{\Delta f}$ is positive definite; (iv) there exists, on a suitably enlarged probability space, an $(r_2 + d_2)$-dimensional standard Wiener process $W(t)$ such that

$$\sup_{1\le j\le t}\left\|f_j^* - \Sigma_{\Delta f}^{1/2} W(j)\right\| = O_{a.s.}\left(t^{1/2-\epsilon}\right),$$

for some $\epsilon > 0$; (v) $E\left\|\sum_{t=1}^{T} e_t\right\|^{2+\kappa} \le C_0\left(\sum_{t=1}^{T} E\left\|e_t\right\|^2\right)^{\frac{2+\kappa}{2}}$.

Assumption 2 says that

- we require the existence of at least the second moment of the innovation $e_t$ and of the initial condition $f_0^*$ respectively;
- a strong approximation exists for the partial sums process $f_t$;
- a Burkholder-type inequality holds;
- so, $e_t$ is allowed to be (weakly) dependent over time
  - e.g. I could assume such processes as iid, linear processes, ARCH/GARCH, Volterra series and data generated by dynamical systems.

**Assumption 3** It holds that: (i) (a) $\max_{1\le i\le N, 1\le t\le T} E\left|u_{i,t}\right|^4 < \infty$; (b) $\max_{1\le t\le T} E\left\|f_t^{(3)}\right\|^4 < \infty$; and (c) $\max_{1\le t\le T} E\left|g_t\right|^4 < \infty$; (ii) (a) $\max_{1\le i\le N} E\left\|\sum_{t=1}^{T} f_t^* u_{i,t}\right\|^2 \le C_0 T^2$; (b) $E\left\|\sum_{t=1}^{T} f_t^* f_t^{(3)\prime}\right\|^2 \le C_0 T^2$; and (c) $E\left\|\sum_{t=1}^{T} f_t^* g_t\right\|^2 \le C_0 T^2$; (iii) $E\left\|\sum_{t=1}^{T} t f_t^*\right\|^2 \le C_0 T^5$; (iv) (a) $\max_{1\le i\le N} E\left|\sum_{t=1}^{T} t u_{i,t}\right|^2 \le C_0 T^3$; (b) $E\left\|\sum_{t=1}^{T} t f_t^{(3)}\right\|^2 \le C_0 T^3$; and (c) $E\left|\sum_{t=1}^{T} t g_t\right|^2 \le C_0 T^3$; (v) $E\left\|\sum_{t=1}^{T} f_t^* f_t^{*\prime}\right\|^2 \le C_0 T^4$.

Assumption 3 says that

- we need the existence of the 4-th moments, which is a milder assumption than the customary 8-th moment existence requirement;

- other parts are very high-level.

**Assumption 4**. The loadings $\Lambda$ are nonstochastic with (i) $\max_{1\leq i\leq N}\|\lambda_i\| < \infty$; (ii) $\lim_{N\to\infty}\frac{\Lambda'\Lambda}{N}\to\Sigma_\Lambda$, where the matrix $\Sigma_\Lambda$ is positive definite.

Assumption 4 is standard

- every diagonal block of $\Sigma_\Lambda$ is also positive definite;
- the common factors belonging in each category are "strong" or "pervasive".

## 3.2 Preliminary results

We base inference on the two matrices

$$\Sigma_1 = \frac{1}{T^3}\sum_{t=1}^{T} X_t X_t', \tag{28}$$

$$\Sigma_2 = \frac{1}{T^2}\sum_{t=1}^{T} X_t X_t'; \tag{29}$$

we denote the $p$-th largest eigenvalues of $\Sigma_1$ and $\Sigma_2$ as $\nu_1^{(p)}$and $\nu_2^{(p)}$ respectively. We will also make use of the (slowly varying) sequence

$$l_{N,T} = (\ln N)^{1+\epsilon}(\ln T)^{\frac{3}{2}+\epsilon},$$

where $\epsilon > 0$.

**Theorem 9.** *Under Assumptions 1-4, it holds that, for every positive, bounded constants $C_p$, there are some random $N_0$ and $T_0$ such that, for all $N \geq N_0$ and $T \geq T_0$*

$$\nu_1^{(p)} \geq C_p N \text{ for } p \leq r_1, \tag{30}$$

$$\nu_1^{(p)} = O_{a.s.}\left(\frac{N}{\sqrt{T}}l_{N,T}\right) \text{ for } p > r_1, \tag{31}$$

*and*

$$\nu_2^{(p)} \geq C_p\frac{N}{\ln\ln T} \text{ for } 1 \leq p \leq r_2 + \max\{r_1, d_2\}, \tag{32}$$

$$\nu_2^{(p)} = O_{a.s.}\left(\frac{N}{\sqrt{T}}l_{N,T}\right) \text{ for } p > r_2 + \max\{r_1, d_2\}. \tag{33}$$

Theorem :

- is a separation result for the eigenvalues corresponding to common factors in $\Sigma_1$ and $\Sigma_2$.

- (30) and (31) refer to $\Sigma_1$:
  - we normalise by $T^{-3}$: because, in essence, $T^{-3}\sum_{t=1}^{T} t^2 < \infty$;
  - the first $r_1$ eigenvalues diverge to infinity at a rate $N$; the others don't, they are bounded;
  - no restrictions on the relative rate of divergence between $N$ and $T$ as they pass to infinity;
  - $\nu_1^{(p)}$, when $p > r_1$, may be very large;
    - it is however smaller - by a factor $T^{-1/2}$ - compared to that of $\nu_1^{(p)}$ when $p \leq r_1$!
- (32) and (33) refer to $\Sigma_2$:
  - we normalise by $T^{-2}$:
    - because the partial sums of $f_t^* f_t^{*\prime}$ should grow at least as fast as $T^2$ by the FCLT in functional spaces (note, this is only an intuition: we need an a.s. rate, so we use the Law of the Iterated Logarithm);
  - the remaining eigenvalues may also diverge, but this will happen at a slower rate.
- the theorem provides only rates: no distributional results are available
  - when data are stationary, an asymptotic distribution for the estimates of the diverging eigenvalues exists, but...
  - ... we do not know if this can also be done for the $\nu_1^{(p)}$s and the $\nu_2^{(p)}$s.
  - Hence, in what follows we will rely only on rates!

Finally, in order to construct the relevant test statistics, we will also make use of the first differenced version of (23):

$$\Delta X_t = \Lambda \Delta f_t + \Delta u_t. \tag{34}$$

**Assumption 5** It holds that: (i) $E(\Delta f_{j,t} \Delta u_{i,t}) = 0$ for $1 \leq j \leq r$ and $1 \leq i \leq N$; (ii) $\max_{1\leq i\leq N, 1\leq t\leq T} E|\Delta X_{i,t}|^4 \leq C_0$; (iii)

$$E \max_{1\leq \widetilde{t}\leq T} \left| \sum_{t=1}^{\widetilde{t}} \Delta X_{h,t} \Delta X_{j,t} - E(\Delta X_{h,t} \Delta X_{j,t}) \right|^2 \leq C_0;$$

(iv) (a) $T^{-1}\sum_{t=1}^{T} E(\Delta f_t \Delta f_t')$ is a positive definite matrix; (b) the largest eigenvalue of $T^{-1}\sum_{t=1}^{T} E(\Delta u_t \Delta u_t')$ is finite; (c) $T^{-1}\sum_{t=1}^{T} E(\Delta u_t \Delta u_t')$ is a positive definite matrix.

As we will see, we consider the matrix $T^{-1}\sum_{t=1}^{T} E(\Delta X_t \Delta X_t')$; with the same notation as before, we denote the $p$-th estimate of the largest eigenvalue of $T^{-1}\sum_{t=1}^{T} E(\Delta X_t \Delta X_t')$ as $\nu_3^{(p)}$.

### 3.3 The test(s)

We now present our algorithm to estimate the dimension of the factor space. We begin by determining $r_1$, based on $\nu_1^{(p)}$, and we then determine $r_2$, using $\nu_2^{(p)}$.

Consider the notation $\beta = \frac{\ln N}{\ln T}$, and define

$$\delta \begin{cases} > 0 & \text{when } \beta < \frac{1}{2} \\ > 1 - \frac{1}{2\beta} & \text{when } \beta \geq \frac{1}{2} \end{cases} . \tag{35}$$

The role played by $\delta$ is the same as before:

- the largest eigenvalues are (modulo some slowly varying functions) proportional to $N$;
- the others, to $\frac{N}{T^{1/2}}$;
- so, when premultiplying eigenvalues by $N^{-\delta}$, the former will be proportional to $N^{1-\delta}$, thereby still diverging; the latter will be proportional to $\frac{N^{1-\delta}}{T^{1/2}}$, which, by construction, will drift to zero.

We will also extensively use the quantity

$$\widehat{\overline{\nu}}_{3,j} = \frac{1}{4\left[N-(j+1)\right]} \sum_{p=j}^{N} \nu_3^{(p)}; \tag{36}$$

this will be employed in order to rescale the estimated eigenvalues, so as to render the test statistic scale invariant. Note the division by 4; this is done, heuristically, since it is possible that $\Delta X_{i,t}$ could inflate the variance by overdifferencing, and the factor 4 represents the largest inflation factor possible.

*Determining the presence of factors with linear trends*

Recall the number of common factors with one linear trend is denoted as $r_1$; we know that either $r_1 = 0$, or $r_1 = 1$.

Consider first $T^{-3}\sum_{t=1}^{T} X_t X_t'$, and its eigenvalues $\nu_1^{(p)}$:

- based on (30)-(31), the first $r_1$ eigenvalues of $T^{-3}\sum_{t=1}^{T} X_t X_t'$ should diverge to positive infinity, as $(N,T) \to \infty$, and...
- ...at a faster rate than the remaining ones.

Thus, the cornerstone of the algorithm to determine $r_1$ is based on checking whether $\nu_1^{(p)}$ diverges sufficiently fast. Consider the following transformation of $\nu_1^{(p)}$

$$\phi_1^{(p)} = \exp\left\{ N^{-\delta} \frac{\nu_1^{(p)}}{\widehat{\overline{\nu}}_{3,p}} \right\}. \tag{37}$$

Based on equations (30) and (31), it can be assumed that

$$\begin{array}{ll} \lim_{\min(N,T)\to\infty} \phi_1^{(p)} = \infty & \text{for } p \leq r_1, \\ \lim_{\min(N,T)\to\infty} \phi_1^{(p)} = 1 & \text{for } p > r_1, \end{array}$$

hold true. The first step is to construct a test for

$$\left\{ \begin{array}{l} H_0^{(1)} : \nu_1^{(p)} \geq C_p N \\ H_1^{(1)} : \nu_1^{(p)} \leq C_p \frac{N}{\sqrt{T}} l_{N,T} \end{array} \right. ; \tag{38}$$

since $\phi_1^{(p)}$ either diverges to infinity or not, but does not have any randomness, we propose to use randomisation.

**Step 1** Generate an *i.i.d.* sample $\left\{\xi_{1,j}^{(p)}\right\}_{j=1}^{R_1}$ from a common distribution $G_1$, independently across $p$, and define the sequence $\sqrt{\phi_1^{(p)}} \times \xi_j^{(p)}$, $1 \leq j \leq R$;

**Step 2** For any $u$ drawn from a distribution $F_1(u)$, define

$$\zeta_{1,j}^{(p)}(u) = I\left[\phi_1^{(p)} \times \xi_{1,j}^{(p)} \leq u\right].$$

**Step 3** Compute

$$\vartheta_1^{(p)}(u) = \frac{1}{\sqrt{R_1}} \sum_{j=1}^{R_1} \frac{\zeta_{1,j}^{(p)}(u) - G_1(0)}{\sqrt{G_1(0)\left[1 - G_1(0)\right]}}.$$

**Step 4** Compute

$$\Theta_1^{(p)} = \int_{-\infty}^{+\infty} \left|\vartheta_1^{(p)}(u)\right|^2 dF_1(u).$$

Same comments as before - it is getting repetitive now:

- the only novelty is that we now want to generalise, and use some distribution $G_1$, in Step 1, instead of the normal
- no real reason to do it, but the maths becomes more elegant.

**Assumption 6.** It holds that: (i) (a) $G_1$ has a bounded density function and (b) $G_1(0) \neq 0$ or 1; (ii) (a) $\int u^2 dF_1(u) < \infty$.

It holds that

**Theorem 10.** *Consider (38). Under Assumptions 1-6, if*

$$\lim_{\min\{N,R_1\}\to\infty} \sqrt{R_1}\exp\left\{-N^{1-\delta}\right\} = 0, \tag{39}$$

*then it holds that*

$$\Theta_1^{(p)} \overset{D^*}{\to} \chi_1^2 \text{ under } H_0^{(1)}, \tag{40}$$

$$\frac{1}{R_1}\frac{\int_{-\infty}^{\infty}\left(G_1(u)-G_1(0)\right)^2 dF_1(u)}{G_1(0)\left(1-G_1(0)\right)}\Theta_1^{(p)} \overset{P^*}{\to} 1 \text{ under } H_1^{(1)}, \tag{41}$$

*for almost all realisations of* $\{e_t, u_{i,t}, 1\le i\le N, 1\le t\le T\}$ *and for all* $p$.

The determination of $r_1$ follows from

**Step 1** Run the test for $H_0 : \lambda_1^{(1)} = \infty$ based on $\Theta_1^{(1)}$. If the null is rejected, set $\widehat{r}_1 = 0$ and stop, otherwise set $\widehat{r}_1 = 1$.

This step consists of a single test. As can be expected, in order to ensure that $\widehat{r}_1$ is consistent, a pivotal role is played by the level of the individual tests, $\alpha_1$, through the relevant critical value denoted as $c_{\alpha,1}$.

**Lemma 11.** *Let Assumptions 1-6 and (39) hold. As* $\min\{N, R_1, T\}\to\infty$, *if* $c_{\alpha,1}\to\infty$ *with* $c_{\alpha,1} = o(R_1)$, *then it holds that* $P[\widehat{r}_1 = r_1] = 1$ *for almost all realisations of* $\{e_t, u_{i,t}, 1\le i\le N, 1\le t\le T\}$.

Note that

- the output of this step is $\widehat{r}_1$, which is an estimate of $r_1$;
- now, $r_1$ can only be either 0 or 1, whence the test being stopped at $H_0 : \nu_1^{(2)} = \infty$
- so, in essence, this step of the procedure is just a test for a common factor with a linear trend.

*Determining the number of* $I(1)$ *common factors*

Consider the matrix $T^{-2}\sum_{t=1}^{T} X_t X_t'$, and let its $p$-th largest eigenvalue $\nu_2^{(p)}$. Based on Theorem 9, the first $r^*$ eigenvalues of $T^{-2}\sum_{t=1}^{T} X_t X_t'$ should diverge to positive infinity, as $(N,T)\to\infty$, at a faster rate than the remaining ones. We exploit this fact, as in the above, by considering the following transformation of $\nu_2^{(p)}$

$$\phi_2^{(p)} = \exp\left\{N^{-\delta}\left(\ln\ln T\right)\frac{\nu_2^{(p)}}{\widehat{\overline{\nu}}_{3,p}}\right\}, \tag{42}$$

which is very similar to (37) except for the presence of the $\ln\ln T$ term - this is based on (32).

Based on (32)-(33), it can be assumed that

$$\begin{array}{ll} \lim_{\min(N,T)\to\infty} \phi_2^{(p)} = \infty & \text{for } p \leq r^*, \\ \lim_{\min(N,T)\to\infty} \phi_2^{(p)} = 1 & \text{for } p > r^*, \end{array}$$

hold true. We propose the following test for

$$\begin{cases} H_0^{(2)} : \nu_2^{(p)} \geq C_p \frac{N}{\ln \ln T} \\ H_1^{(2)} : \nu_2^{(p)} \leq C_p \frac{N}{\sqrt{T}} l_{N,T} \end{cases} . \tag{43}$$

**Step 1** Generate an *i.i.d.* sample $\left\{\xi_{2,j}^{(p)}\right\}_{j=1}^{R_2}$ from a common distribution $G_2$, independently across $p$, and define the sequence $\sqrt{\phi_2^{(p)}} \times \xi_j^{(p)}$, $1 \leq j \leq R_2$;

**Step 2** For any $u$ drawn from a distribution $F_2(u)$, define

$$\zeta_{2,j}^{(p)}(u) = I\left[\phi_2^{(p)} \times \xi_{2,j}^{(p)} \leq u\right].$$

**Step 3** Compute

$$\vartheta_2^{(p)}(u) = \frac{1}{\sqrt{R_2}} \sum_{j=1}^{R_2} \frac{\zeta_{2,j}^{(p)}(u) - G_2(0)}{\sqrt{G_2(0)[1 - G_2(0)]}}.$$

**Step 4** Compute

$$\Theta_2^{(p)} = \int_{-\infty}^{+\infty} \left|\vartheta_2^{(p)}(u)\right|^2 dF_2(u).$$

**Assumption 7**. It holds that: (i) (a) $G_2$ has a bounded density function and (b) $G_2(0) \neq 0$ or 1; (ii) (a) $\int u^2 dF_2(u) < \infty$.

It holds that

**Theorem 12.** *Consider (43). Under Assumptions 1-7, if*

$$\lim_{\min\{N,R_2\}\to\infty} \sqrt{R_2} \exp\left\{-N^{1-\delta}\right\} = 0, \tag{44}$$

*then it holds that*

$$\Theta_2^{(p)} \xrightarrow{D^*} \chi_1^2 \text{ under } H_0^{(2)}, \tag{45}$$

$$\frac{1}{R_2} \frac{\int_{-\infty}^{\infty} \left(G_2(u) - G_2(0)\right)^2 dF_1(u)}{G_2(0)(1 - G_2(0))} \Theta_2^{(p)} \xrightarrow{P^*} 1 \text{ under } H_1^{(2)}, \tag{46}$$

*for almost all realisations of* $\{e_t, u_{i,t}, 1 \leq i \leq N, 1 \leq t \leq T\}$ *and for all* $p$.

As noted before, conditionally on the sample the sequence $\Theta_2^{(p)}$, $p \geq 1$, is independent across $p$. We recommend the following algorithm for the determination of $r^*$:

**Step 1** Run the test for $H_0 : \nu_2^{(1)} = \infty$ based on $\Theta_2^{(1)}$. If the null is rejected, set $\widehat{r}^* = 0$ and stop, otherwise go to the next step.

**Step 2** Starting from $p = 1$, run the test for $H_0 : \nu_2^{(p+1)} = \infty$ based on $\Theta_2^{(p+1)}$, constructed using an artificial sample $\left\{\xi_{2,j}^{(p+1)}\right\}_{j=1}^{R_2}$ generated independently of $\left\{\xi_{2,j}^{(1)}\right\}_{j=1}^{R_2}$, ..., $\left\{\xi_{2,j}^{(p)}\right\}_{j=1}^{R_2}$. If the null is rejected, set $\widehat{r}^* = p$ and stop; otherwise repeat the step until the null is rejected (or until a pre-specified maximum number, say $r_{\max}^*$, is reached).

As can be expected, in this context a pivotal role is played by the level of the individual tests, which should be chosen so that $\widehat{r}^*$ is a good approximation of $r^*$, at least asymptotically. Similarly to the previous case, let $c_{\alpha,2}$ denote the critical value of the test at each step.

**Lemma 13.** *Let Assumptions 1-7 and (44) hold. As* $\min\{N, R_2, T\} \to \infty$*, if* $r_{2,\max} \geq r_2$ *and* $c_{\alpha,2} \to \infty$ *with* $c_{\alpha,2} = o(R_2)$*, then it holds that* $P[\widehat{r}^* = r^*] = 1$ *for almost all realisations of* $\{e_t, u_{i,t}, 1 \leq i \leq N, 1 \leq t \leq T\}$.

Note

- we can go quickly here: same interpretation as before;
- after estimating $r^*$, it is possible to estimate the number of common $I(1)$ factors which have mean zero as $\widehat{r}^* - \widehat{r}_1$;
- indeed, Lemma 13 ensures that

$$P[\widehat{r}^* - \widehat{r}_1 = r_2 + r_1(1 - d_1)d_2] = 1,$$

  under the conditions of Lemmas 11 and 13.

- for completeness, after estimating $r^*$, you can apply the very first algorithm to first differenced data, in order to estimate the total number of common factors:
  - denoting such estimate as $\widehat{r}$, it follows that an estimator of the number of common stationary factors is $\widehat{r} - \widehat{r}^*$.

## 3.4 Comment: on weak factors again

All common factors are assumed to be strong. This is a direct consequence of having $\|\Lambda\|^2 = O(N)$. It is however possible to imagine a situation in which the common factors are "weak", or "less pervasive", for the same well known reasons:

- either because they are weak genuinely;
- or because they only affect some units.

For the sake of simplicity, but with no loss of generality, we consider the case where there are $r$ $I(1)$ factors in total, no stationary factors and, in

$$X_t = \Lambda F_t + u_t,$$

the matrix $\Lambda'\Lambda$ is diagonal, with diagonal elements $c_p$ given by

$$c_p = \begin{cases} N & \text{for } 1 \leq p \leq p' \\ N^{1-\kappa_p} & \text{for } p' < p \leq r \end{cases}.$$

Allowing for $\kappa_p \in (0,1)$ corresponds to the case of having weak factors, and the larger $\kappa_p$ the weaker the corresponding factor. Suppose that the researcher is using $\Sigma_2$ and its eigenvalues in order to determine the number of $I(1)$ common factors. Repeating exactly the same arguments in the proof of Theorem 9, it can be shown that

$$\nu_2^{(p)} \geq C_0 \frac{c_p}{\ln \ln T}. \tag{47}$$

Equation (47) entails that, whenever $p' < p \leq r$,

$$\nu_2^{(p)} \geq C_0 \frac{N^{1-\kappa_p}}{\ln \ln T}. \tag{48}$$

Our procedure, essentially, is based on testing whether, as $\min\{N,T\} \to \infty$

$$\begin{cases} H_0^{(2)} : N^{-\delta}\nu_2^{(p)} \to \infty \\ H_1^{(2)} : N^{-\delta}\nu_2^{(p)} \to 0 \end{cases},$$

with $\delta$ selected as per (61). Thus, based on (48), weak factors can be determined if

$$\lim_{\min\{N,T\}\to\infty} \frac{N^{1-\kappa_p-\delta}}{\ln \ln T} \to \infty,$$

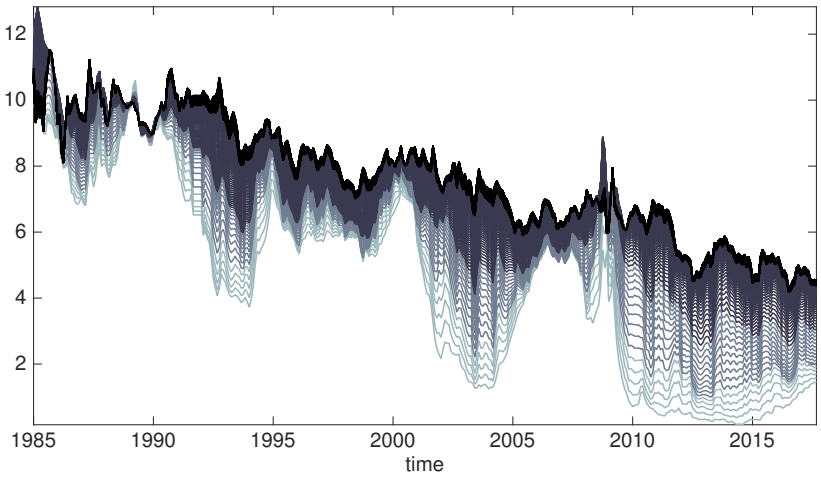
which requires

$$\kappa_p < 1 - \delta. \tag{49}$$

It is the same as before! Take a look

- when $\beta \leq \frac{1}{2}$, $\delta = 0$; thus, it is required that $\kappa_p < 1$:
  - when $N$ is smaller than $T$, our procedure is able to detect even very weak factors
- when $\beta > \frac{1}{2}$ (and therefore $N$ is bigger than $T$), it is required that $\kappa_p < 1 - \frac{1}{2\beta}$:
  - as $\beta$ increases the test is less and less able to detect weak factors;
  - when $N$ and $T$ have the same order of magnitude, and thus $\beta = 1$, weak factors can be detected as long as $\kappa_p < \frac{1}{2}$ - that is, when the eigenvalues associated with that factor diverge to infinity a bit faster than $N$.

Figure 1: HQM Corporate Bond Yield Curve




# 4 Empirical evidence

We have some numbers from the Monte Carlo; available upon request, but the main message is

- it works a dream!
- the exception, if you will, is when there is no linear trend but a common $I(1)$ factor and $T$ is small - in this case, the test is occasionally fooled and thinks there is a linear trend. This vanishes when $T$ increases.

Let's take a look at an application. It's not the main point, it is not the only thing you can do, but let us see.

We illustrate our methodology through an application to the High Quality Market (HQM) Corporate Bond Yield Curve, available from the Federal Reserve Economic Data (FRED)

We use

- monthly data on HQM Corporate Bonds with maturities from 6 months up to 100 years ($N = 196$),
- spanning the period from January 1985 to September 2017 ($T = 393$).

The data are shown in Figure 1, which shows evidence of non-stationarity and co-movements both cross-sectionally and across time.

Results:

Table 1: Estimated number of factors in the HQM Corporate Bond Yield Curve

| | | $BT1$ | $BT2$ | $BT3$ | $IC$ |
|---|---|---|---|---|---|
| with linear trend | $\widehat{r}_1$ | 0 | 0 | 1 | n.a. |
| non-stationary | $\widehat{r}^*$ | 1 | 3 | 5 | 5 |
| zero-mean, $I(1)$ | $\widehat{r}_2$ | 1 | 3 | 4 | n.a. |
| all factors | $\widehat{r}$ | 1 | 5 | 5 | 5 |
| zero-mean, $I(0)$ | $\widehat{r}_3$ | 0 | 2 | 0 | 0 |

- borderline evidence of a common factor with a linear trend - indeed, this is picked up by $BT3$;
- there could be up to 4 zero mean $I(1)$ common factors
  - based on the discrepancy between these two criteria, it may be argued that two of such factors may be only borderline non-stationary
- to sum up, five common factors, which we estimate as the principal components of $X_t$, using the covariance $T^{-2}\sum_{t=1}^{T} X_t X_t'$ and imposing the identifying constraint $\Lambda'\Lambda = NI_r$.

The estimated factors are shown in Figure 2 (solid red lines).

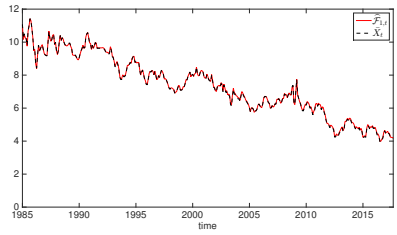
Let us see, especially in light of the literature which usually identifies 3 common factors (level, slope and curvature):

- in order to identify $\widehat{F}_{1,t}$, we consider the proxy $\bar{X}_t = N^{-1}\sum_{i=1}^{N} X_{i,t}$; we found that $corr(\bar{X}_t, \widehat{F}_{1,t}) \simeq 1$, which strongly suggests that $\widehat{F}_{1,t}$ can be viewed as the *level* of the curve;
- turning to $\widehat{F}_{2,t}$, we use, as a proxy for the slope, $dX_t = N^{-1}\sum_{i=2}^{N}(\ln X_{i,t} - \ln X_{i-1,t}) = N^{-1}(\ln X_{N,t} - \ln X_{1,t})$. We find that $corr(dX_t, \widehat{F}_{2,t}) = .82$, which suggests that $\widehat{F}_{2,t}$ can be interpreted as the *slope* of the term structure;
- we compare $\widehat{F}_{3,t}$ to $d^2X_t = (N-2)^{-1}\sum_{i=2}^{N-1}(X_{i+1,t} - 2X_{i,t} + X_{i-1,t})$ as a proxy for the curvature; we find $corr(d^2X_t, \widehat{F}_{3,t}) = .53$, which shows some evidence that $\widehat{F}_{3,t}$ can be interpreted as the *curvature*.

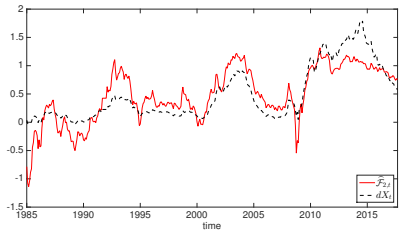
As far as the remaining two estimated common factors are concerned

- we evaluate the correlation between $S_t$ - the spread between the 10 years HQM bond rate and the Federal Funds rate - and the fourth factor finding that $corr(S_t, \widehat{F}_{4,t}) = .51$, whence we propose to interpret $\widehat{F}_{4,t}$ as the *spread factor*.

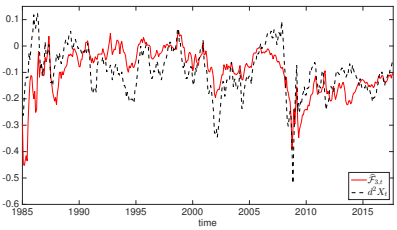
Figure 2: Estimated and identified common factors $\widehat{F}_{j,t}$ with proxies.



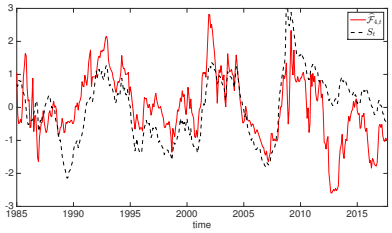

$\widehat{F}_{1,t}$



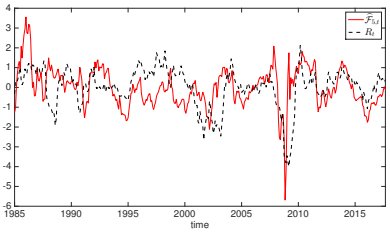

$\widehat{F}_{2,t}$




$\widehat{F}_{3,t}$



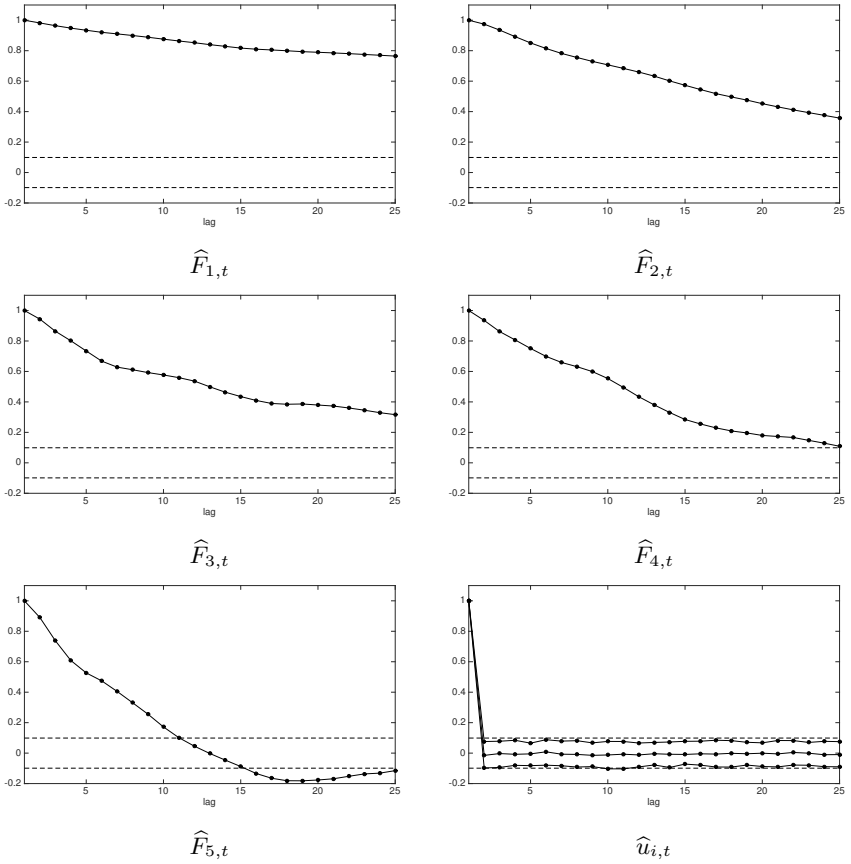

$\widehat{F}_{4,t}$




$\widehat{F}_{5,t}$

- letting $R_t$ be the yearly returns of the Standard & Poor's index, we have $corr(R_t, \widehat{F}_{5,t}) = .30$; this seems to suggest that $\widehat{F}_{5,t}$ may be viewed as a *financial factor*, or that, at a minimum, $\widehat{F}_{5,t}$ is intimately related to the financial market.

Figure 3: Autocorrelation of estimated common factors $\widehat{F}_{j,t}$ and idiosyncratic errors $\widehat{u}_{i,t}$.

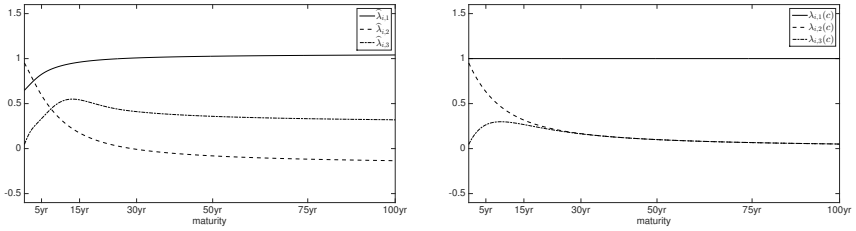



$\widehat{F}_{1,t}$ $\widehat{F}_{2,t}$

$\widehat{F}_{3,t}$ $\widehat{F}_{4,t}$

$\widehat{F}_{5,t}$ $\widehat{u}_{i,t}$

Figure 4: Estimated and theoretical factor loadings.




$\widehat{\lambda}_{i,j}$, $j = 1, 2, 3$ $\lambda_{i,j}(c)$, $j = 1, 2, 3$ and $c = 0.2$

# 5 Monitoring for structural breaks

We investigate the issue of testing for the stability of

$$X_{i,t} = a_i' f_t + u_{i,t}, \tag{50}$$

where $\{X_{i,t},\ 1 \leq i \leq N,\ 1 \leq t \leq T\}$ is a panel of $N$ observed time series observed for $T$ periods and $a_i$ and $f_t$ are latent vectors of loadings and factors, respectively, both of dimension $r < N$ and representing the "signal" component of the data.

In particular, we focus on the *sequential monitoring* - that is, we propose a test to check whether there are any breaks as new data come in.

Sequential detection of breaks in (1) is important for at least four reasons

- it is important to verify whether a model, which has been valid thus far, is still capable of adequately approximate the behaviour of new data;
- there is empirical evidence that factor structures do tend to change over time, especially in presence of a crisis;
- inference on factor models can be severely marred by the presence of a break; think about forecasting;
- it is generally true that on-line monitoring might be costly, but nowadays, at least in economics and finance where data are collected and made available automatically, such cost is almost negligible, especially if compared with the potential costs of employing a model which is no longer valid.

There are several possible ways in which a factor model may undergo a change at a point in time $\tau$, but

- in all cases it may be argued that a change in the factor structure of the data will result in a change in the covariance matrix of $\{X_{i,t}\}_{i=1}^N$;
- since common factors determine the presence and number of spiked eigenvalue of the covariance of $\{X_{i,t}\}_{i=1}^N$...
- ... it is natural to investigate whether a change has occurred in a factor structure by verifying whether changes have occurred in the spectrum of the covariance matrix.

Formally, we test for the null hypothesis that the factor structure does not change, viz.:

$$H_0 : X_{i,t} = \sum_{j=1}^{r} a_{ij} f_{j,t} + u_{i,t}, \quad 1 \leq t \leq T.$$

As far as alternatives are concerned, we focus on two different possible breaks at a point in time $\tau$:

- changes in the loadings attached to one or more common factor:

$$H_{A,1} : \begin{cases} X_{i,t} = \sum_{j=1}^{r} a_{ij} f_{j,t} + u_{i,t} & \text{for} \quad 1 \leq t < \tau \\ X_{i,t} = \sum_{j=1}^{r} \widetilde{a}_{ij} f_{j,t} + u_{i,t} & \quad\quad\;\; \tau \leq t \leq T \end{cases}, \tag{51}$$

  where $\widetilde{a}_{ij} \neq a_{ij}$ for some $i$ and $j$;

- the appearance of $q \geq 1$ new factors:

$$H_{A,2} : \begin{cases} X_{i,t} = \sum_{j=1}^{r} a_{ij} f_{j,t} + u_{i,t} & \text{for} \quad 1 \leq t < \tau \\ X_{i,t} = \sum_{j=1}^{r} a_{ij} f_{j,t} + \sum_{j=1}^{q} b_{ij} g_{j,t} + u_{i,t} & \quad\quad\;\; \tau \leq t \leq T \end{cases}. \tag{52}$$

Now:

- Hypothesis $H_{A,1}$ is the typical case considered in all the above cited literature on change points in factor models;
- $H_{A,2}$ has received less attention from the literature;
- other alternatives can also be accommodated in our framework.

Main result

- we show that, under both $H_{A,1}$ and $H_{A,2}$, the $(r+1)$-*th* largest eigenvalue of the covariance matrix of $\{X_{i,t}\}_{i=1}^{N}$
  - becomes unbounded at time $\tau + 1$, passing to infinity as fast as the sample size $N$ if there is a break;
  - it stays bounded under the null of no break.
- Thus, we base our test on the estimated $(r+1)$-*th* eigenvalue of the sample covariance matrix of $\{X_{i,t}\}_{i=1}^{N}$.
- This has its difficulties:
  - under the null of no break, the $(r+1)$-*th* sample eigenvalue does not have a known distribution;
  - indeed it cannot even be estimated consistently;
  - we have no idea about the dependence structure of the estimators, as a technical (but crucial) difficulty.

# 6 Assumptions and preliminary theory

Recall the model

$$X_{i,t} = a_i'(t) f_t + u_{i,t}, \quad 1 \leq i \leq N,\ 1 \leq t \leq T. \tag{53}$$

We use the notation $r\,(t)$ to denote the number of factors at a given time $t$, i.e. the vectors of loadings $a_i(t)$ and of factors $f_t$ have dimension $r(t)$.

Consider also the matrix form of (53):

$$X_t = A(t) f_t + u_t, \quad 1 \leq t \leq T, \tag{54}$$

where, $A(t) = [a_1(t)|...|a_N(t)]'$ is the loadings matrix and $u_t = [u_{1,t}, ..., u_{N,t}]'$ is called "idiosyncratic" component.

We define the covariance matrix of the data at time $t$ as $\Sigma_X\,(t) = E\,(X_t X_t')$. Consider the (population) rolling covariance matrix

$$\Sigma_m\,(t) = \frac{1}{m} \sum_{k=t-m+1}^{t} \Sigma_X\,(k)\,, \quad m \leq t \leq T, \tag{55}$$

and its sample counterpart

$$\widehat{\Sigma}_m\,(t) = \frac{1}{m} \sum_{k=t-m+1}^{t} X_k X_k', \quad m \leq t \leq T. \tag{56}$$

Very important notation

- $m$ will then denote our sample size when estimating the model
  - hence our asymptotic results are defined for $m \to \infty$;
- we assume that for the first $m$ periods no change-point is present and we have $r$ factors;
- our monitoring procedure will last until $T > m$:
  - therefore, the total number of observations $T$ includes both the estimation and the monitoring period.
  - Note that, in real applications, the monitoring may be expected to go on indefinitely, so that $T \to \infty$.

We start with the following assumption.

**Assumption 1** It holds that *(i)* $E\,(X_{i,t}) = 0$ for all $1 \leq i \leq N$ and $1 \leq t \leq T$; *(ii)* $E(f_{j,t} u_{i,t}) = 0$ for all $i, j, t$; (iii) $r(t) = r$ for $1 \leq t \leq m$; (iv) $r(t) < N$ for $1 \leq t \leq T$ and for any $N$.

Notes:

- "classical" assumption...
- by part *(iii)*, we have that, in presence of breaks, the change-point location $\tau$ is such that $\tau > m$;
- the covariance is decomposed as

$$\Sigma_X(t) = A(t)\,\Sigma_F(t)\,A(t)' + \Sigma_u(t),$$

having defined $\Sigma_F(t) = E(f_t f_t')$ and $\Sigma_u(t) = E(u_t u_t')$.

Similarly to the above, we denote the *k-th* largest eigenvalue of $\Sigma_m(t)$ as $\lambda^{(k)}(t)$, the *k-th* eigenvalue of $A(t)\,\Sigma_F(t)\,A(t)'$ as $\gamma^{(k)}(t)$; and, finally, the *k-th* eigenvalue of $\Sigma_u(t)$ as $\omega^{(k)}(t)$; similarly, we denote the *k-th* largest eigenvalue of $\widehat{\Sigma}_m(t)$ as $\widehat{\lambda}^{(k)}(t)$. In order to derive our results on the population and sample eigenvalues, we make the following assumptions.

**Assumption 2** It holds that *(i)* $\gamma^{(k)}(t) = C_k(t)\,N$ for all $1 \leq k \leq r(t)$, some finite $C_k(t) > 0$ and for $m \leq t \leq T$; *(ii)* $\omega^{(k)}(t) \leq C_0$ for all $1 \leq k \leq N$ and $m \leq t \leq T$.

**Assumption 3** It holds that *(i)* $E|X_{i,t}|^{4+\epsilon} \leq C_0$ for $1 \leq i \leq N$, $1 \leq t \leq T$ and some $\epsilon > 0$; *(ii)*

$$E \max_{t_0 \leq \tilde{t} \leq t_0+m-1} \left| \sum_{t=1}^{\tilde{t}} X_{h,t}X_{j,t} - E(X_{h,t}X_{j,t}) \right|^2 \leq C_1 m$$

for $1 \leq h, j \leq N$ and $1 \leq t_0 \leq T - m$.

Notes

- same as the previous assumptions in the first paper presented, and therefore...
- ... we constrain the amount of serial correlation that one can have in the process $\{X_{h,t}X_{j,t}\}_{t=1}^{T}$ and therefore, albeit indirectly, in $\{X_{i,t}\}_{t=1}^{T}$:
  - recall that this assumption is satisfied by any linear process with summable fourth cumulants. Some examples
    - Volterra series;
    - ARCH/GARCH processes, thus allowing for the case of conditional heteroskedasticity.

The following result characterizes the behaviour of the population eigenvalues of $\Sigma_m(t)$.

**Lemma 14.** *Under Assumptions 1 and 2, it holds that*

$$\lambda^{(r+1)}(t) \leq C_0, \quad m \leq t \leq T, \text{ under } H_0. \tag{57}$$

*Further, it holds that*

$$\lambda^{(r+1)}(t) \begin{cases} \leq C_0 & m \leq t < \tau, \\ \geq C_1 \frac{t-\tau+1}{m} N & \tau \leq t < \tau + m, \\ \leq C_0 & \tau + m \leq t \leq T, \end{cases} \quad \text{under } H_{A,1}, \tag{58}$$

$$\lambda^{(r+1)}(t) \begin{cases} \leq C_0 & m \leq t < \tau, \\ \geq C_1 \frac{t-\tau+1}{m} N & \tau \leq t < \tau + m, \\ \geq C_1 N & \tau + m \leq t \leq T, \end{cases} \quad \text{under } H_{A,2}. \tag{59}$$

Consider now the slowly varying sequence at infinity:

$$l(m, N) = (\ln N)^{1+\epsilon} (\ln m)^{\frac{1+\epsilon}{2}},$$

where $\epsilon > 0$. The sample counterpart to Lemma 14 is the following.

**Lemma 15.** *Under Assumptions 1 and 3, it holds that*

$$\widehat{\lambda}^{(r+1)}(t) = \lambda^{(r+1)}(t) + O_{a.s.}\left(\frac{N}{m^{1/2}} l(m, N)\right), \quad m \leq t \leq T. \tag{60}$$

# 7 Testing procedure and asymptotics

Define $\delta \in [0, 1)$ such that

$$\delta \begin{cases} > 0 & \\ > 1 - \frac{1}{2}\frac{\ln m}{\ln N} & \end{cases} \quad \text{according as} \quad \begin{matrix} N \leq m^{1/2} \\ N > m^{1/2} \end{matrix} ; \tag{61}$$

thus, the choice of $\delta$ is uniquely determined by $N$ and $m$, with no need to estimate anything.

We consider the statistic

$$\phi_{N,m}(t) = g\left(N^{-\delta} \frac{\widehat{\lambda}^{(r+1)}(t)}{\frac{1}{N}\sum_{k=1}^{N} \widehat{\lambda}^{(k)}(t)}\right), \quad m \leq t \leq T, \tag{62}$$

where

- $g(\cdot)$ is a monotonically increasing function such that $g(0) = 0$ and $\lim_{x\to\infty} g(x) = \infty$;
- we use $g(a) = a$, but other choices are also possible;
- the denominator in (62) makes the argument of $g(\cdot)$ scale invariant.

The results above imply - now it is too repetitious for our own good - that

$$\lim_{N,m\to\infty}\phi_{N,m}(t)=g(0)=0,\ \text{ w.p. 1, when }\ N^{-\delta}\widehat{\lambda}^{(r+1)}(t)\to 0,$$
$$\lim_{N,m\to\infty}\phi_{N,m}(t)=g(\infty)=\infty,\ \text{ w.p. 1, when }\ N^{-\delta}\widehat{\lambda}^{(r+1)}(t)\to \infty.$$

Letting $t^*_{N,m}$ be a point in time such that

$$\lim_{N,m\to\infty}\frac{N^{1-\delta}}{m}\left(t^*_{N,m}-\tau+1\right)=\infty,$$

we therefore have that

$$\lim_{N,m\to\infty}\phi_{N,m}(t)=0,\quad m\le t\le T,\quad \text{under } H_0,$$

while

$$\lim_{N,m\to\infty}\phi_{N,m}(t)=\begin{cases}0 & m\le t<\tau,\\ \infty & t^*_{N,m}\le t<\tau+m,\\ 0 & \tau+m\le t\le T,\end{cases},\quad \text{under } H_{A,1},$$

or

$$\lim_{N,m\to\infty}\phi_{N,m}(t)=\begin{cases}0 & m\le t<\tau,\\ \infty & t^*_{N,m}\le t<\tau+m,\end{cases},\quad \text{under } H_{A,2};$$

between $\tau$ and $t^*_{N,m}-\tau+1$, $\phi_{N,m}(t)$ is growing from 0 to $\infty$.

Given that the results above entail that we only have rates for $\phi_{N,m}(t)$, we propose a to use a randomised version of it, built according to the following steps.

*Step A1.* At each given $t\ge m$, generate an *i.i.d.* sample $\left\{\xi_j(t)\right\}_{j=1}^{R}$ with common distribution $G_\phi$ such that $G_\phi(0)\neq 0$ or 1.

*Step A2.* For any $u$ drawn from a distribution $F_\phi(u)$, define

$$\zeta_j(u;t)=I\left[\xi_j(t)\le u\phi^{-1}_{N,m}(t)\right].$$

*Step A3.* Compute

$$\vartheta(u;t)=\frac{1}{\sqrt{R}}\sum_{j=1}^{R}\frac{\zeta_j(u;t)-G_\phi(0)}{\sqrt{G_\phi(0)\left[1-G_\phi(0)\right]}}.$$

*Step A4.* Compute

$$\Theta_t=\int_{-\infty}^{+\infty}\left|\vartheta(u;t)\right|^2 dF_\phi(u).$$

Although the details of the behaviour of $\Theta_t$ under the null and the alternative are spelt out later on, a heuristic preview of the main argument may be helpful. In essence, under the alternative the Bernoulli random variable $\zeta_j(u;t)$ should be equal to 1 or 0 with probability $G_\phi(0)$ and $1-G_\phi(0)$ respectively, and thus have mean $G_\phi(0)$. In this case, when constructing $\vartheta(u;t)$, a Central Limit Theorem holds and therefore we expect $\Theta_t$ to have a chi-square distribution. On the other hand, under the null $\zeta_j(u;t)$ should be (heuristically) 0 or 1 with probability 0 or 1 (depending on the sign of $u$) - thus its mean should be different than $G_\phi(0)$ (and equal to 0 or 1 depending on the sign of $u$) and a Law of Large Numbers should hold. Note that, by construction, conditionally on the sample the sequence $\{\Theta_t\}_{t=m}^{T}$ is independent across $t$. In order to study $\Theta_t$, we need the following assumptions.

**Assumption 4**. It holds that: *(i)* $G_\phi(\cdot)$ has a bounded density; *(ii)* $\int_{-\infty}^{+\infty} u^2 dF_\phi(u) < \infty$.

**Assumption 5**. It holds that, as $\min(N,m,R)\to\infty$: (i)

$$R^{1/2}\left[g\left(N^{1-\delta}\frac{t-\tau+1}{m}\right)\right]^{-1}\to 0, \quad \begin{array}{ll}\text{under } H_{A,1}, & \text{for } t^*_{N,m}\le t<\tau+m,\\ \text{under } H_{A,2}, & \text{for } t^*_{N,m}\le t\le T;\end{array}$$

(ii) $R^{1/2}\left[g\left(N^{1-\delta}\right)\right]^{-1}\to 0$ under $H_{A,1}$, for $\tau+m\le t\le T$.

Assumption 5 looks horrible. It is. But it implies, for example, that

- choosing $R=N$ is valid, in principle,
- and it works very well in simulations.

Let now $P^*$ represent the conditional probability with respect to $\{X_{i,t}, 1\le i\le N, 1\le t\le T\}$; "$\overset{\mathcal{D}^*}{\to}$" and "$\overset{P^*}{\to}$" denote, respectively, conditional convergence in distribution and in probability according to $P^*$.

**Theorem 16.** *Under Assumptions 1-5, as* $\min(N,m,R)\to\infty$*, it holds that*

$$\Theta_t \overset{\mathcal{D}^*}{\to} \chi^2_1, \quad \begin{array}{ll}\text{under } H_{A,1}, & \text{for } t^*_{N,m}\le t<\tau+m,\\ \text{under } H_{A,2}, & \text{for } t^*_{N,m}\le t\le T,\end{array} \tag{63}$$

*for almost all realisations of* $\{X_{i,t},\ 1\le i\le N,\ 1\le t\le T\}$.
*Under Assumptions 1-4, as* $\min(N,m,R)\to\infty$*, it holds that*

$$\frac{1}{R}\Theta_t \overset{P^*}{\to} \frac{\int_{-\infty}^{+\infty}\left|I_{[0,\infty)}(u)-G_\phi(0)\right|^2 dF_\phi(u)}{G_\phi(0)\left[1-G_\phi(0)\right]}, \quad \begin{array}{ll}\text{under } H_0, & \text{for } m\le t\le T,\\ \text{under } H_{A,1}, & \text{for } m\le t<\tau,\\ & \text{and } \tau+m\le t\le T,\\ \text{under } H_{A,2}, & \text{for } m\le t<\tau,\end{array} \tag{64}$$

*for almost all realisations of* $\{X_{i,t},\ 1\le i\le N,\ 1\le t\le T\}$.

The theorem

- is an intermediate result;
- the fact that $\Theta_t$ converges to a chi-square distribution would be potentially useful if one wished to test for the (rather unusual) null of having a break
- but, it is necessary to have a statistic which diverges under the null and is bounded under the alternative.

We therefore propose to randomise $\Theta_t$, with a second randomisation based on

$$\psi_{N,m,R}(t) = h\left(\frac{\Theta_t}{\widetilde{l}(N,m,R)}\right), \quad m \leq t \leq T, \tag{65}$$

where

$$\widetilde{l}(N,m,R) = (\ln N)^{2+\epsilon} (\ln m)^{2+\epsilon} (\ln R)^{2+\epsilon},$$

for some $\epsilon > 0$ - in practice, any small value of $\epsilon$ works well.

In (65), the function $h(\cdot)$, similarly to $g(\cdot)$ in (62), is a monotonically increasing function such that $h(0) = 0$ and $\lim_{x\to\infty} h(x) = \infty$; again, we use $h(a) = a$.

Similarly to the case of $\phi_{N,m}(t)$, Theorem 16 entails that

$$\lim_{N,m,R\to\infty} \psi_{N,m,R}(t) = \infty, \quad m \leq t \leq T, \quad \text{under } H_0,$$

and

$$\lim_{N,m,R\to\infty} \psi_{N,m,R}(t) = \begin{cases} 0 & m \leq t < \tau, \\ \infty & t^*_{N,m} \leq t < \tau + m, \\ 0 & \tau + m \leq t \leq T, \end{cases}, \quad \text{under } H_{A,1},$$

while

$$\lim_{N,m,R\to\infty} \psi_{N,m,R}(t) = \begin{cases} 0 & m \leq t < \tau, \\ \infty & t^*_{N,m} \leq t < \tau + m, \end{cases}, \quad \text{under } H_{A,2};$$

Consider now the second randomisation.

*Step B1.* At each given $t \geq m$, generate an *i.i.d.* sample $\left\{\widetilde{\xi}_j(t)\right\}_{j=1}^{W}$ with common distribution $G_\psi$ such that $G_\psi(0) \neq 0$ or 1.

*Step B2.* For any $u$ drawn from a distribution $F_\psi(u)$, define

$$\widetilde{\zeta}_j(u;t) = I\left[\widetilde{\xi}_j(t) \leq u\psi^{-1}_{N,m,R}(t)\right].$$

*Step B3.* Compute

$$\gamma(u;t) = \frac{1}{\sqrt{W}} \sum_{j=1}^{W} \frac{\widetilde{\zeta}_j(u;t) - G_\psi(0)}{\sqrt{G_\psi(0)\left[1 - G_\psi(0)\right]}}.$$

*Step B4.* Compute

$$\Gamma_t = \int_{-\infty}^{+\infty} |\gamma(u;t)|^2 \, dF_\psi(u).$$

The following assumptions are needed in order to study the asymptotic behavior of $\Gamma_t$; note their similarity with Assumptions 4 and 5.

**Assumption 6**. It holds that: *(i)* $G_\psi(\cdot)$ has a bounded density; *(ii)* $\int_{-\infty}^{+\infty} u^4 dF_\psi(u) < \infty$.

**Assumption 7**. It holds that, as $\min(N, m, R, W) \to \infty$

$$W^{1/2} \left[ h \left( \frac{R}{\widetilde{l}(N, m, R)} \right) \right]^{-1} \to 0.$$

Let $P^\dagger$ represent the conditional probability with respect to $\{X_{i,t}, 1 \leq i \leq N,\ 1 \leq t \leq T\}$ and $\{\xi_j(t), 1 \leq j \leq R,\ m \leq t \leq T\}$; we use the notation "$\overset{\mathcal{D}^\dagger}{\to}$" and "$\overset{P^\dagger}{\to}$" to define, respectively, conditional convergence in distribution and in probability according to $P^\dagger$.

**Theorem 17.** *Under Assumptions 1-7, as* $\min(N, m, R, W) \to \infty$*, it holds that*

$$\Gamma_t \overset{\mathcal{D}^\dagger}{\to} \chi^2_1, \quad \begin{array}{ll} \textit{under } H_0, & \textit{for } \ m \leq t \leq T, \\ \textit{under } H_{A,1}, & \textit{for } \ m \leq t < \tau \textit{ and } \tau + m \leq t \leq T, \\ \textit{under } H_{A,2}, & \textit{for } \ m \leq t < \tau, \end{array} \tag{66}$$

*for almost all realisations of* $\{X_{i,t},\ 1 \leq i \leq N,\ 1 \leq t \leq T\}$ *and* $\{\xi_j(t), 1 \leq j \leq R,\ m \leq t \leq T\}$.
*Under Assumptions 1-6, as* $\min(N, m, R, W) \to \infty$*, it holds that*

$$\frac{1}{W}\Gamma_t \overset{P^\dagger}{\to} \begin{array}{ll} \dfrac{\int_{-\infty}^{+\infty} \left| I_{[0,\infty)}(u) - G_\psi(0) \right|^2 dF_\psi(u)}{G_\psi(0)\left[1 - G_\psi(0)\right]}, & \textit{under } H_{A,1}, \quad \textit{for } \ t^*_{N,m} \leq t < \tau + m, \\ & \textit{under } H_{A,2}, \quad \textit{for } \ t^*_{N,m} \leq t \leq T, \end{array} \tag{67}$$

*for almost all realisations of* $\{X_{i,t},\ 1 \leq i \leq N,\ 1 \leq t \leq T\}$ *and* $\{\xi_j(t),\ 1 \leq j \leq R,\ m \leq t \leq T\}$.

Theorem 17 is, again, an intermediate result:

- $\Gamma_t$ has (asymptotically) a chi-square distribution under the null of no breaks;
- further, by construction the sequence $\Gamma_t$, $m \leq t \leq T$ is independent across $t$ conditional on the sample.
- from these two basic facts we propose a monitoring scheme for the on-line detection of breaks in the factor structure.

## 7.1 Sequential monitoring of factor models

Recall that, after collecting $m$ observations, we monitor our model over the period $m+1 \leq t \leq T$, which has size denoted as $T_m = T - m$. We then consider a monitoring procedure based on the detector

$$d(k;m) = \left| \sum_{t=m+1}^{m+k} \frac{\Gamma_t - 1}{\sqrt{2}} \right|, \quad 1 \leq k \leq T_m, \tag{68}$$

which covers the entire monitoring period. In other words our detector is made of the cumulative sum of the centered and standardized version of the sequence $\{\Gamma_t\}_{t=m}^{T}$, obtained by double randomisation.

Given the stopping rule

$$\widehat{k}_m = \begin{cases} \inf\{1 \leq k \leq T_m \text{ such that } d(k;m) \geq \nu(k;m)\}, \\ T_m \text{ if the above does not hold in the monitoring period.} \end{cases} \tag{69}$$

we define the estimated change-point location as $\widehat{\tau}_m = \widehat{k}_m + m$. The threshold function in (69) is defined as (see **?** and **?**)

$$\nu(k;m) = c_{\alpha,m}\nu^*(k;m), \tag{70}$$

$$\nu^*(k;m) = m^{1/2}\left(1 + \frac{k}{m}\right)\left(\frac{k}{k+m}\right)^{\eta}, \ \eta \in \left[0, \frac{1}{2}\right], \tag{71}$$

where $c_{\alpha,m}$ is a critical value corresponding to a pre-specified level $\alpha$. Depending on the choice of $\eta$, the critical value is defined as

$$P\left(\sup_{0\leq t\leq 1} \frac{|B(t)|}{t^{\eta}} \leq c_{\alpha,m}\right) = 1 - \alpha, \ \text{ for } \eta \in \left[0, \frac{1}{2}\right), \tag{72}$$

where $\{B(t),\ 0 \leq t \leq 1\}$ denotes a standard Wiener process, or

$$c_{\alpha,m} = \frac{D_m - \ln[-\ln(1-\alpha)]}{A_m}, \ \text{ for } \eta = \frac{1}{2}, \tag{73}$$

with $A_m = (2\ln\ln m)^{1/2}$ and $D_m = 2\ln\ln m + \frac{1}{2}\ln\ln\ln m - \frac{1}{2}\ln\pi$. Note that in (72) $c_{\alpha,m}$ does not depend on $m$, whilst it does in (73).

In order to derive our main theorem, we also need the following assumptions.

**Assumption 8**. It holds that (i) $T_m = O(m^{\varkappa})$ for some $\varkappa \geq 1$; (ii) $\liminf_{m\to\infty} \frac{T_m}{m} > 0$; (iii) $T_m > \tau + C_0 m^{1/2+\epsilon}$ for $\epsilon > 0$ such that $\frac{N^{1-\delta}}{m^{1/2-\epsilon}} \to C_1$.

**Assumption 9**. It holds that (i) $\int_{-\infty}^{+\infty} |u|^{4+2\delta}\, dF_{\psi}(u) < \infty$; (ii)

$$m^{1/2+\epsilon}\left\{W^{-1} + W\left[h\left(\frac{R}{\widetilde{l}(N,m,R)}\right)\right]^{-2} + \left[h\left(\frac{R}{\widetilde{l}(N,m,R)}\right)\right]^{-1}\right\} \to 0,$$

for some $\epsilon > 0$.

**Theorem 18.** *Let Assumptions 1-9 hold. Under $H_0$ it holds that, as* $\min(N, m, R, W) \to \infty$

$$P^{\dagger}\left(\max_{1\leq k\leq T_m}\frac{d(k;m)}{\nu^*(k;m)}\leq x\right)\to P\left(\sup_{0\leq t\leq 1}\frac{|B(t)|}{t^{\eta}}\leq x\right),\quad \textit{for } \eta\in[0,\frac{1}{2}), \tag{74}$$

$$P^{\dagger}\left(\max_{1\leq k\leq T_m}\frac{d(k;m)}{\nu^*(k;m)}\leq \frac{x+D_m}{A_m}\right)\to e^{-e^{-x}},\quad \textit{for } \eta=\frac{1}{2}, \tag{75}$$

*for almost all realisations of $\{X_{i,t},\ 1\leq i\leq N,\ 1\leq t\leq T\}$ and $\{\xi_j(t),\ 1\leq j\leq R,\ m\leq t\leq T\}$ and for $x\in(-\infty,\infty)$.*
*Under $H_{A,1}$ and $H_{A,2}$, as* $\min(N, m, R, W) \to \infty$*, and for a given significance level $\alpha$, it holds that*

$$c_{\alpha,m}^{-1}\max_{1\leq k\leq T_m}\frac{d(k;m)}{\nu^*(k;m)}\xrightarrow{P^{\dagger}}\infty,\quad \textit{for all } \eta\in\left[0,\frac{1}{2}\right], \tag{76}$$

*where $c_{\alpha,m}$ is defined in* (72) *when $\eta<\frac{1}{2}$ and in* (73) *when $\eta=\frac{1}{2}$.*

The main implication of the theorem is summarized in the following result:

**Corollary 19.** *Under the assumptions of Theorem 18 it holds that:*

$$\lim_{\min(N,m,R,W)\to\infty}P^{\dagger}(\widehat{\tau}_m<T)\leq\alpha,\quad \textit{under } H_0, \tag{77}$$

$$\lim_{\min(N,m,R,W)\to\infty}P^{\dagger}(\widehat{\tau}_m<T)=1,\quad \textit{under } H_{A,1} \textit{ and } H_{A,2}, \tag{78}$$

*for almost all realisations of $\{X_{i,t},\ 1\leq i\leq N,\ 1\leq t\leq T\}$ and $\{\xi_j(t),\ 1\leq j\leq R,\ m\leq t\leq T\}$.*

Notes:

- the notion of size implied by (77), in this context, is very different from the one usually considered in the literature;
- we wish to keep the false rejection probability as little as possible, and therefore (at a minimum) below the threshold $\alpha$, rather than making it close to $\alpha$.
- This makes the monitoring procedure different from the standard Neyman-Pearson paradigm:
  - given that the monitoring horizon keeps expanding, the purpose of $c_{\alpha,m}$ is to ensure that the chance of a false break detection is as little as possible.

**Tables**

| | $(N,T)$ | | **DGP features** $\rho=0,\ b=0$ | $\rho=.5,\ b=0$ | $\rho=.5,\ b=.5$ |
|---|---|---|---|---|---|
| 0.1 | $(50,50)$ | $\widehat{k}$-average | 0.000 | 0.000 | 0.000 |
| | | (% *wrong*) | (0) | (0) | (0) |
| | | $BN$-average | 1.000 | 1.000 | 4.018 |
| | | (% *wrong*) | (100.0) | (100.0) | (100.0) |
| | | $ABC$-average | 0.026 | 0.160 | 0.468 |
| | | (% *wrong*) | (2.40) | (12.0) | (28.20) |
| | | $ON$-average | 0.030 | 0.942 | 2.312 |
| | | (% *wrong*) | (3.00) | (69.00) | (100.0) |
| | | $ER$-average | 0.000 | 0.854 | 3.540 |
| | | (% *wrong*) | (0) | (36.40) | (100.0) |
| | | $GR$-average | 0.000 | 0.940 | 3.714 |
| | | (% *wrong*) | (0) | (39.80) | (100.0) |
| | $(25,100)$ | $\widehat{k}$-average | 0.000 | 0.000 | 0.058 |
| | | (% *wrong*) | (0) | (0) | (5.60) |
| | | $BN$-average | 1.000 | 1.000 | 5.024 |
| | | (% *wrong*) | (100.0) | (100.0) | (100.0) |
| | | $ABC$-average | 0.580 | 0.786 | 0.432 |
| | | (% *wrong*) | (18.0) | (37.60) | (28.40) |
| | | $ON$-average | 0.004 | 0.200 | 0.750 |
| | | (% *wrong*) | (0.40) | (19.00) | (62.00) |
| | | $ER$-average | 0.000 | 0.002 | 3.868 |
| | | (% *wrong*) | (0) | (0.20) | (100.0) |
| | | $GR$-average | 0.000 | 0.006 | 3.388 |
| | | (% *wrong*) | (0) | (0.40) | (100.0) |
| | $(25,200)$ | $\widehat{k}$-average | 0.000 | 0.000 | 0.034 |
| | | (% *wrong*) | (0) | (0) | (3.40) |
| | | $BN$-average | 1.000 | 1.000 | 5.672 |
| | | (% *wrong*) | (100.0) | (100.0) | (100.0) |
| | | $ABC$-average | 0.188 | 0.716 | 0.270 |
| | | (% *wrong*) | (4.60) | (26.60) | (19.40) |
| | | $ON$-average | 0.000 | 0.008 | 0.070 |
| | | (% *wrong*) | (0) | (0.80) | (6.80) |
| | | $ER$-average | 0.000 | 0.000 | 4.236 |
| | | (% *wrong*) | (0) | (0) | (100.0) |
| | | $GR$-average | 0.000 | 0.000 | 3.446 |
| | | (% *wrong*) | (0) | (0) | (100.0) |
| | $(50,100)$ | $\widehat{k}$-average | 0.000 | 0.000 | 0.000 |
| | | (% *wrong*) | (0) | (0) | (0) |
| | | $BN$-average | 1.000 | 1.000 | 7.538 |
| | | (% *wrong*) | (100.0) | (100.0) | (100.0) |
| | | $ABC$-average | 0.020 | 0.078 | 0.106 |
| | | (% *wrong*) | (1.40) | (6.40) | (8.80) |
| | | $ON$-average | 0.000 | 0.152 | 1.040 |
| | | (% *wrong*) | (0) | (14.00) | (76.40) |
| | | $ER$-average | 0.000 | 0.000 | 3.598 |
| | | (% *wrong*) | (0) | (0) | (100.0) |
| | | $GR$-average | 0.000 | 0.000 | 3.622 |
| | | (% *wrong*) | (0) | (0) | (100.0) |
| | $(50,200)$ | $\widehat{k}$-average | 0.000 | 0.000 | 0.002 |
| | | (% *wrong*) | (0) | (0) | (0.20) |
| | | $BN$-average | 1.000 | 1.000 | 8.000 |
| | | (% *wrong*) | (100.0) | (100.0) | (100.0) |

**Determination of $k$: stationary case, $k = 0$.**

| | | **DGP features** | | | | |
|---|---|---|---|---|---|---|
| | | $k=1$ | | $k=3$ | | |
| $(N,T)$ | | $\rho=.5,\ b=0$ | $\rho=.5,\ b=.5$ | $\rho=.5,\ b=0$ | $\rho=.5,\ b=.5$ | $\rho=.5$ |
| $(50,50)$ | $\widehat{k}$-average | 0.996 | 1.002 | 3.012 | 2.938 | 4. |
| | (% *wrong*) | (0.40) | (0.20) | (1.20) | (6.20) | (1 |
| | $BN$-average | 1.002 | 4.870 | 5.868 | 8.000 | 5. |
| | (% *wrong*) | (0.20) | (93.40) | (92.0) | (100.0) | (0 |
| | $ABC$-average | 1.144 | 1.378 | 3.106 | 3.232 | 5. |
| | (% *wrong*) | (9.60) | (25.40) | (9.0) | (20.60) | (4 |
| | $ON$-average | 1.418 | 2.452 | 3.399 | 3.396 | 5. |
| | (% *wrong*) | 37.60 | (90.20) | (37.2) | (37.20) | ( |
| | $ER$-average | 1.004 | 1.008 | 2.999 | 2.816 | 4. |
| | (% *wrong*) | (0.40) | (0.80) | (0.50) | (16.40) | (0 |
| | $GR$-average | 1.003 | 1.006 | 2.998 | 2.902 | 4. |
| | (% *wrong*) | (0.30) | (0.60) | (0.10) | (10.60) | (0 |
| $(25,100)$ | $\widehat{k}$-average | 1.000 | 1.002 | 2.992 | 2.988 | 5. |
| | (% *wrong*) | (0) | (0.20) | (0.80) | (3.40) | (0 |
| | $BN$-average | 1.000 | 4.516 | 4.912 | 8.000 | 5. |
| | (% *wrong*) | (0) | (69.80) | (72.80) | (100.0) | (0 |
| | $ABC$-average | 1.706 | 1.350 | 3.440 | 3.068 | 5. |
| | (% *wrong*) | (36.80) | (26.60) | (27.20) | (16.0) | (14 |
| | $ON$-average | 1.062 | 1.266 | 3.136 | 3.478 | 5. |
| | (% *wrong*) | (6.20) | (25.20) | (13.00) | (43.60) | ( |
| | $ER$-average | 1.000 | 1.004 | 3.000 | 2.936 | 5. |
| | (% *wrong*) | (0) | (0.40) | (0) | (11.40) | ( |
| | $GR$-average | 1.000 | 1.002 | 3.000 | 3.010 | 5. |
| | (% *wrong*) | (0) | (0.20) | (0) | (9.00) | ( |
| $(25,200)$ | $\widehat{k}$-average | 1.000 | 1.002 | 3.000 | 2.998 | 5. |
| | (% *wrong*) | (0) | (0.20) | (0) | (0.20) | ( |
| | $BN$-average | 1.000 | 5.200 | 3.044 | 6.494 | 5. |
| | (% *wrong*) | (0) | (63.20) | (4.00) | (77.80) | ( |
| | $ABC$-average | 1.626 | 1.192 | 3.384 | 3.060 | 5. |
| | (% *wrong*) | (27.80) | (15.20) | (24.20) | (11.40) | (16 |
| | $ON$-average | 1.000 | 1.014 | 3.006 | 3.000 | 5. |
| | (% *wrong*) | (0) | (1.40) | (0.60) | (0) | ( |
| | $ER$-average | 1.000 | 1.010 | 3.000 | 3.000 | 5. |
| | (% *wrong*) | (0) | (0.90) | (0) | (0) | ( |
| | $GR$-average | 1.000 | 1.008 | 3.000 | 3.000 | 5. |
| | (% *wrong*) | (0) | (0.80) | (0) | (0) | ( |
| $(50,100)$ | $\widehat{k}$-average | 1.000 | 1.000 | 2.998 | 3.018 | 4. |
| | (% *wrong*) | (0) | (0) | (0.20) | (4.20) | (1 |
| | $BN$-average | 1.000 | 7.648 | 5.102 | 7.842 | 5. |
| | (% *wrong*) | (0) | (99.60) | (83.4) | (100.0) | ( |
| | $ABC$-average | 1.062 | 1.090 | 3.048 | 3.054 | 5. |
| | (% *wrong*) | (5.20) | (8.60) | (4.20) | (5.0) | (3 |
| | $ON$-average | 1.046 | 1.474 | 3.046 | 3.048 | 5. |
| | (% *wrong*) | (4.60) | (43.40) | (4.60) | (4.80) | ( |
| | $ER$-average | 1.000 | 1.005 | 3.000 | 2.934 | 5. |
| | (% *wrong*) | (0) | (0.50) | (0) | (3.80) | ( |
| | $GR$-average | 1.000 | 1.008 | 3.000 | 2.996 | 5. |
| | (% *wrong*) | (0) | (0.80) | (0) | (0.40) | ( |
| | $\widehat{k}$-average | 1.000 | 1.000 | 3.000 | 3.000 | 5. |
| | (% *wrong*) | (0) | (0) | (0) | (0) | ( |
| | $BN$-average | 1.000 | 8.000 | 3.028 | 8.000 | 5. |

**Determining $k$; stationary case.**

**References**

*For the first chapter*


Ahn, S.C., Horenstein, A.R., 2013. Eigenvalue ratio test for the number of factors. Econometrica, 81, 1203–1227.

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

Bai, J., Ng, S., 2002. Determining the number of factors in approximate factor models. Econometrica, 70, 191–221.

Bai, J., Ng, S., 2007. Determining the number of primitive shocks in factor models. Journal of Business and Economic Statistics, 25, 52–60.

Bai, J., Ng, S., 2008. Large dimensional factor models. Foundations and Trends in Econometrics, 3, 89–163.

Bai, J., 2009. Panel data models with interactive fixed effects. Econometrica, 77, 1229–1279.

Bai, Z.D., 1999. Methodologies in spectral analysis of large-dimensional random matrices, a review. Statistica Sinica, 9, 611–677.

Bai, Z.D., Yin, Y.Q., 1993. Limit of the smallest eigenvalue of a large dimensional sample covariance matrix. Annals of Probability, 21, 1275–1294.

Cai, G., 2006. Chover-type laws of the iterated logarithm for weighted sums of $\tilde{\rho}$-mixing sequences, Journal of Applied Mathematics and Stochastic Analysis 2006, Article ID 65023.

Castagnetti, C., Rossi, E., Trapani, L., 2015a. Inference on factor structures in heterogeneous panels. Journal of Econometrics, 184, 145–157.

Castagnetti, C., Rossi, E., Trapani, L., 2015b. Testing for no factor structures: on the use of Hausman-type statistics. Economics Letters, 130, 66–68.

Chamberlain, G., Rothschild, M., 1983. Arbitrage, factor structure, and mean-variance analysis on large asset markets. Econometrica, 51, 1281–1304.

Corradi, V., Swanson, N.R., 2006. The effects of data transformation on common cycle, cointegration, and unit root tests: Monte Carlo and a simple test. Journal of Econometrics, 132, 195–229.

Davidson, 1994. *Stochastic Limit Theory.* Oxford: Oxford University Press.

Davidson, J., 2002. Establishing conditions for the functional central limit theorem in nonlinear and semiparametric time series processes. Journal of Econometrics 106, 243–269.

Forni, M., Giannone D., Lippi M., Reichlin L., 2009. Opening the black box: structural factor models with large cross-sections. Econometric Theory, 25, 1319–1347.

Hallin, M., Liska, R., 2007. Determining the number of factors in the generalized dynamic factor model. Journal of the American Statistical Association, 102, 603–617.

Horn, R.A., Johnson, C.R., 1985. *Matrix Analysis.* Cambridge: Cambridge University Press.

Lin Z., Bai Z., 2010. *Probability Inequalities*. Science Press, Beijing and Springer-Verlag, Berlin.

Ling, S., 2007. Testing for change points in time series models and limiting theorems for NED sequences. Annals of Statistics 35, 1213–1237.

McLeish, D.L., 1975. A maximal inequality and dependent strong laws. Annals of Probability, 3, 829–839.

Onatski, A., 2009. A formal statistical test for the number of factors in the approximate factor models. Econometrica, 77, 1447–1479.

Onatski, A., 2010. Determining the number of factors from empirical distribution of eigenvalues. Review of Economic and Statistics, 92, 1004–1016.

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

Onatski, A., 2015. Asymptotic analysis of the squared estimation error in misspecified factor models. Journal of Econometrics, 186, 388–406.

Pearson, E.S., 1950. On questions raised by the combination of tests based on discontinuous distributions. Biometrika, 37, 383–398.

Schwert, W., 1989. Test for unit roots: a Monte Carlo investigation. Journal of Business and Economic Statistics, 7, 147–159.

Serfling, R.J., 1970. Moment inequalities for the maximum cumulative sum. The Annals of Mathematical Statistics, 41, 1227–1234.

Tao, T., 2012. *Topics in Random Matrix Theory*. American Mathematical Society.

Trzcinka, C., 1986. On the number of factors in the arbitrage pricing model. Journal of Finance, XLI, 347–368.

Wang, W., Fan, J., 2016. Asymptotics of empirical eigen-structure for high dimensional spiked covariance. Forthcoming at the Annals of Statistics.

*For the second chapter*

Ahn, S. C. and A. R. Horenstein (2013). Eigenvalue ratio test for the number of factors. Econometrica 81, 1203-1227.

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

Ang, A. and M. Piazzesi (2003). A no-arbitrage vector autoregression of term structure dynamics with macroeconomic and latent variables. Journal of Monetary Economics 50, 745-787.

Bai, J. (2004). Estimating cross-section common stochastic trends in non-stationary panel data. Journal of Econometrics 122, 137-183.

Bai, J. and C. Kao (2006). On the estimation and inference of a panel cointegration model with cross-sectional dependence. Contributions to economic analysis 274, 3-30.

Bai, J., C. Kao, and S. Ng (2009). Panel cointegration with global stochastic trends. Journal of Econometrics 149, 82-99.

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

Bai, J. and S. Ng (2004). A panic attack on unit roots and cointegration. Econometrica 72 (4), 1127-1177.

Bandi, F. and V. Corradi (2014). Nonparametric nonstationarity tests. Econometric Theory 30, 127-149.

Berkes, I. and W. Philipp (1979). Approximation thorems for independent and weakly dependent random vectors. The Annals of Probability, 29-54.

Chamberlain, G. and M. Rothschild (1983). Arbitrage, factor structure, and mean-variance analysis on large asset markets. Econometrica, 1281-1304.

Chen, L. and W. B. Wu (2018). Testing for trends in high-dimensional time series. Journal of the American Statistical Association. forthcoming.

Christensen, J. H. E. and J. A. Lopez (2008). Common risk factors in the US treasury and corporate bond markets: An arbitrage-free dynamic Nelson-Siegel modeling approach. Technical report, Federal Reserve Bank of San Francisco.

Coroneo, L., D. Giannone, and M. Modugno (2016). Unspanned macroeconomic factors in the yield curve. Journal of Business & Economic Statistics 34, 472-485.

Corradi, V. and N. R. Swanson (2006). The e ects of data transformation on common cycle, cointegration, and unit root tests: Monte Carlo and a simple test. Journal of Econometrics 132, 195-229.

Dai, Q. and K. J. Singleton (2000). Speci cation analysis of a new term structure models. The Journal of Finance 55, 1943-1978.

Diebold, F. X. and C. Li (2006). Forecasting the term structure of government bond yields. Journal of Econometrics 130, 337-364.

Diebold, F. X., G. D. Rudebusch, and S. B. Aruoba (2006). The macroeconomy and the yield curve: a dynamic latent factor approach. Journal of Econometrics 131, 309-338.

Donsker, M. and S. Varadhan (1977). On laws of the iterated logarithm for local times. Communications on Pure and Applied Mathematics 30 (6), 707-753.

Duffie, D., L. Saita, and K. Wang (2007). Multi-period corporate default prediction with stochastic covariates. Journal of Financial Economics 83, 635-665.

Duffie, D. and K. J. Singleton (1999). Modeling term structures of defaultable bonds. The Review of Financial Studies 12, 687-720.

Dungey, M., V. L. Martin, and A. R. Pagan (2000). A multivariate latent factor decomposition of international bond yield spreads. Journal of Applied Econometrics 15 (6), 697-715.

Engel, C., N. C. Mark, and K. D. West (2015). Factor model forecasts of exchange rates. Econometric Reviews 34 (1-2), 32-55.

Escribano, A. and D. Pe˜na (1994). Cointegration and common factors. Journal of Time Series Analysis 15 (6), 577-586.

Estrella, A. and F. S. Mishkin (1998). Predicting us recessions: Financial variables as leading indicators. The Review of Economics and Statistics 80, 45-61.

Gengenbach, C., F. C. Palm, and J.-P. Urbain (2009). Panel unit root tests in the presence of cross-sectional dependencies: comparison and implications for modelling. Econometric Reviews 29 (2), 111-145.

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

Kapetanios, G., M. H. Pesaran, and T. Yamagata (2011). Panels with non-stationary multifactor error structures. Journal of Econometrics 160 (2), 326-348.

Lam, C. and Q. Yao (2012). Factor modeling for high-dimensional time series: inference for the number of factors. The Annals of Statistics 40 (2), 694-726.

Lin, Z. and Z. Bai (2010). Probability inequalities of random variables. In Probability Inequalities, pp. 37-50. Springer.

Liu, W. and Z. Lin (2009). Strong approximation for a class of stationary processes.Stochastic Processes and their Applications 119 (1), 249-280.

Maciejowska, K. (2010). Common factors in nonstationary panel data with a deterministic trend - estimation and distribution theory.

Moon, H. R. and B. Perron (2004). Testing for a unit root in panels with dynamic factors. Journal of Econometrics 122 (1), 81-126.

Moon, H. R. and B. Perron (2007). An empirical analysis of nonstationarity in a panel of interest rates with factors. Journal of Applied Econometrics 22 (2), 383-400.

Nelson, C. R. and A. F. Siegel (1987). Parsimonious modeling of yield curves. Journal of Business 60, 473-489.

Onatski, A. (2009). Testing hypotheses about the number of factors in large factor models. Econometrica 77, 1447-1479.

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

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

Onatski, A. and C. Wang (2016). Alternative asymptotics for cointegration tests in large VARs. Forthcoming at Econometrica.

Pearson, E. (1950). On questions raised by the combination of tests based on discontinuous distributions. Biometrika 37, 383-398.

Pena, D. and P. Poncela (2006). Nonstationary dynamic factor analysis. Journal of Statistical Planning and Inference 136 (4), 1237-1257.

Pesaran, M. H., L. V. Smith, and T. Yamagata (2013). Panel unit root tests in the presence of a multifactor error structure. Journal of Econometrics 175 (2), 94-115.

Stock, J. H. and M. W.Watson (1988). Testing for common trends. Journal of the American Statistical Association 83 (404), 1097-1107.

Trapani, L. (2017). A randomised testing procedure to determine the number of factors.Journal of the American Statistical Association. forthcoming.

Wang, W. and J. Fan (2016). Asymptotics of empirical eigen-structure for high dimensional spiked covariance. The Annals of Statistics 45 (3), 1342-1374.

Wu, W. B. (2007). Strong invariance principles for dependent random variables. The Annals of Probability 35, 2294-2320.

Zhang, B., G. Pan, and J. Gao (2017). CLT for largest eigenvalues and unit root tests for high-dimensional nonstationary time series. The Annals of Statistics. forthcoming.

Zhang, R., P. Robinson, and Q. Yao (2018). Identifying cointegration by eigenanalysis. Journal of the American Statistical Association. forthcoming.

*For the third chapter*

Ahn, S. C. and Horenstein, A. R. (2013). Eigenvalue ratio test for the number of factors. Econometrica 81, 1203-1227.

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

Andreou, E. and Ghysels, E. (2006). Monitoring disruptions in nancial markets. Journal of Econometrics 135, 77-124.

Aue, A., Hormann, S., Horvath, L., Huskova, M. and Steinebach, J. G. (2012). Sequential testing for the stability of high-frequency portfolio betas. Econometric Theory 28, 804-837.

Aue, A. and Horvath, L. (2004). Delay time in sequential detection of change. Statistics & Probability Letters 67(3), 221-231.

Bai, J. (2003). Inferential theory for factor models of large dimensions. Econometrica 71, 135-171.

Bai, J. (2004). Estimating cross-section common stochastic trends in nonstationary panel data. Journal of Econometrics 122, 137-183.

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

Bai, Z. and Yao, J. (2012). On sample eigenvalues in a generalized spiked population model. Journal of Multivariate Analysis 106, 167-177.

Baltagi, B. H., Kao, C. and Wang, F. (2017). Identi cation and estimation of a large factor model with structural instability. Journal of Econometrics 197, 87-100.

Bandi, F. and Corradi, V. (2014). Nonparametric nonstationarity tests. Econometric Theory 30, 127-149.

Barigozzi, M., Cho, H. and Fryzlewicz, P. (2017). Simultaneous multiple change-point and factor analysis for high-dimensional time series. Forthcoming at the Journal of Econometrics.

Bates, B. J., Plagborg-M ller, M., Stock, J. H. and Watson, M. W. (2013). Consistent factor estimation in dynamic factor models with structural instability. Journal of Econometrics 177, 289-304.

Breitung, J. and Eickmeier, S. (2011). Testing for structural breaks in dynamic factor models. Journal of Econometrics 163, 71-84.

Brodsky, B. (2010). Sequential detection and estimation of change-points', Sequential Analysis 29, 217-233.

Chamberlain, G. and Rothschild, M. (1983). Arbitrage, factor structure, and mean-variance analysis on large asset markets. Econometrica 51, 1281-1304.

Chen, L., Dolado, J. J. and Gonzalo, J. (2014). Detecting big structural breaks in large factor models. Journal of Econometrics 180, 30-48.

Chen, L. H. Y. (1978). A short note on the conditional Borel-Cantelli lemma. The Annals of Probability 6, 699-700.

Cheng, X., Liao, Z. and Schorfheide, F. (2016). Shrinkage estimation of high-dimensional factor models with structural instabilities. The Review of Economic Studies 83, 1511-1543.

Chu, C., Stinchcombe, M. and White, H. (1996). Monitoring structural change. Econometrica 64, 1045-1066.

Corradi, V. and Swanson, N. R. (2006). The effects of data transformation on common cycle, cointegration, and unit root tests: Monte Carlo and a simple test. Journal of Econometrics 132, 195-229.

Corradi, V. and Swanson, N. R. (2014). Testing for structural stability of factor augmented forecasting models. Journal of Econometrics 182, 100-118.

Csorgo, M. and Horvath, L. (1997), Limit theorems in change-point analysis, Vol. 18, John Wiley & Sons Inc.

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

El Karoui, N. (2007). Tracy-Widom limit for the largest eigenvalue of a large class of complex sample covariance matrices. The Annals of Probability 35, 663-714.

Fan, J., Liao, Y. and Mincheva, M. (2013). Large covariance estimation by thresholding principal orthogonal complements. Journal of the Royal Statistical Society: Series B 75, 603-680.

Forni, M., Giannone, D., Lippi, M. and Reichlin, L. (2009). Opening the black box: Structural factor models versus structural VARs. Econometric Theory 25, 1319-1347.

Forni, M., Hallin, M., Lippi, M. and Reichlin, L. (2000). The Generalized Dynamic Factor Model: Identification and estimation. The Review of Economics and Statistics 82, 540-554.

Groen, J. J., Kapetanios, G. and Price, S. (2013). Multivariate methods for monitoring structural change. Journal of Applied Econometrics 28, 250-274.

Han, X. and Inoue, A. (2015). Tests for parameter instability in dynamic factor models. Econometric Theory 31, 1117-1152.

Hannan, E. J. (1970), Multiple Time Series, John Wiley & Sons.

Horvath, L., Huskova, M., Kokoszka, P. and Steinebach, J. (2004). Monitoring changes in linear models. Journal of Statistical Planning and Inference 126, 225-251.

Horvath, L., Kokoszka, P. and Steinebach, J. (2007). On sequential detection of parameter changes in linear regression. Statistics and Probability Letters 80, 1806-1813.

Horvath, L. and Trapani, L. (2017). Testing for randomness in a random coefficient autoregression. Technical report.

Johnstone, I. M. and Lu, A. Y. (2009). On consistency and sparsity for principal components analysis in high dimensions. Journal of the American Statistical Association 104, 682-693.

Jung, S. and Marron, J. S. (2009). PCA consistency in high dimension, low sample size context. The Annals of Statistics 37, 4104-4130.

Kirch, C. and Tadjuidje Kamgaing, J. (2015). On the use of estimating functions in monitoring time series for change points. Journal of Statistical Planning and Inference 161, 25-49.

Kirch, C. and Weber, S. (2017), Modified sequential change point procedures based on estimating functions. Technical report.

Komlos, J., Major, P. and Tusnady, G. (1975). An approximation of partial sums of independent R.V.'s and the sample DF.I. Z. Wahrscheinlichkeitstheorie und verwandte Gebiete 32, 111-131.

Komlos, J., Major, P. and Tusnady, G. (1976). An approximation of partial sums of independent R.V.'s and the sample DF.II. Z. Wahrscheinlichkeitstheorie und verwandte Gebiete 34, 33-58.

Lai, T. L. (1995). Sequential changepoint detection in quality control and dynamical systems. Journal of the Royal Statistical Society: Series B 57, 613-658.

Lam, C. and Yao, Q. (2012). Factor modeling for high-dimensional time series: inference for the number of factors. The Annals of Statistics 40, 694-726.

Li, J., Todorov, V., Tauchen, G. and Lin, H. (2017). Rank tests at jump events. Journal of Business & Economic Statistics, forthcoming.

Massacci, D. (2017). Least squares estimation of large dimensional threshold factor models. Journal of Econometrics 197, 101-129.

Moench, E., Ng, S. and Potter, S. (2013). Dynamic hierarchical factor models. The Review of Economics and Statistics 95, 1811-1817.

Moricz, F. (1983). A general moment inequality for the maximum of the rectangular partial sums of multiple series. Acta Mathematica Hungarica 41, 337-346.

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

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

Onatski, A., Moreira, M. J. and Hallin, M. (2014). Signal detection in high dimension: The multispiked case. The Annals of Statistics 42, 225-254.

Paul, D. (2007). Asymptotics of sample eigenstructure for a large dimensional spiked covariance model. Statistica Sinica 17, 1617-1642.

Pearson, E. (1950). On questions raised by the combination of tests based on discontinuous distributions. Biometrika 37, 383-398.

Sen, P. K. (1981), Sequential nonparametrics: invariance principles and statistical inference, Probability and mathematical statistics, Wiley New York.

Stock, J. H. and Watson, M. W. (2002). Forecasting using principal components from a large number of predictors. Journal of the American Statistical Association 97, 1167-1179.

Stock, J. H. and Watson, M. W. (2009), Forecasting in dynamic factor models subject to structural instability, in J. Castle and N. Shephard, eds, 'The Methodology and Practice of Econometrics: A Festschrift in Honour of David F. Hendry. Oxford University Press.

Stock, J. H. and Watson, M. W. (2011). Dynamic factor models, in M. Clements and D. Hendry, eds, 'Oxford Handbook of Economic Forecasting', Oxford University Press.

Tao, T. (2012). Topics in Random Matrix Theory. Graduate studies in mathematics, American Mathematical Society.

Trapani, L. (2017). A randomised sequential procedure to determine the number of factors. Journal of the American Statistical Association, forthcoming.

Wang, W. and Fan, J. (2017). Asymptotics of empirical eigen-structure for high dimensional spiked covariance. The Annals of Statistics, forthcoming.

Wu, W. B. (2005). Nonlinear system theory: Another look at dependence. Proceedings of the National Academy of Sciences of the United States of America 102, 14150-14154.

Yamamoto, Y. and Tanaka, S. (2015). Testing for factor loading structural change under common breaks. Journal of Econometrics 189, 187-206.