Inference in heavy-tailed non-stationary multivariate time series

Matteo Barigozzi[1] Giuseppe Cavaliere[2] Lorenzo Trapani[3]

[1]University of Bologna [2]University of Bologna [3]University of Nottingham

7th June 2022

Italian Ecoometric Association, Webinar series

Research question

We talk about cointegration and common stochastic trends in the presence of heavy tails.
In essence,

- we study how to estimate the number of common stochastic trends, m, in an N-dimensional time series y_t, where
- N is fixed at the beginning, and of course we can determine the rank of cointegration $N - m$;
- our estimator can be used in the presence of arbitrarily heavy tails, with...
- no need for estimation of nuisance parameters, chiefly the tail index;
- we present an extension to large N.

Literature: standard estimation

The literature on rank of cointegration/common trends:

- very developed - basically no need for references as we all know them;
- but the technology usually requires finite second moments at least (essential),
- as well as the correct specification of the VECM (desirable),
- otherwise using e.g. Johansen's tests results in size distortion, which can be very severe (Caner, 1998).

Literature: heavy tails

This could be an issue:

- there is evidence that some datasets have heavy tails, where the second, or even the first, moment may not exist;
- references are many even here, but e.g. Samorodnitsky and Taqqu (1994) or Embrechts et al. (2013).

There are some contributions on inference in this context:

- Caner (1998, JoE) derives the asymptotics for Johansen's type tests; see also Paulauskas and Rachev (1998, AoAP) and She and Ling (2020, JoE);
  - need to estimate the tail index, a nuisance parameter, to implement the test (= to get the critical values).
- alternatives are
  - bootstrap/resampling: Cavaliere et al. (2018, ET) and Jach and Kokoszka (2004, MCAP): unclear what to do in this case with cointegration though;
  - distribution free: Hallin et al (2011, 2016; JE): they assume finite second moment, funnily enough, but more importantly one needs the correct specification of the VECM.
  - Yao, Zhang and Robinson (2016, JASA) have a series of contribution on using second moment matrices, essentially along similar lines as the literature on factor models, but again finite variance is required.

Our contribution

We:

- not unlike Yao, Zhang and Robinson (2016), use second moment matrices;
- find an eigen-gap result which does not depend on nuisance parameters;
- use such eigen-gap to construct a randomised sequential procedure to determine m, which
- can be used irrespective of having or not heavy tails, of how heavy they are, and of having heteroskedasticity.

Theory/model

Given the N-dimensional vector y_t, consider the $MA(\infty)$ representation

$$\Delta y_t = C(L)\,\varepsilon_t, \tag{0.1}$$

Standard arguments allow to represent (0.1) as

$$y_t = C\sum_{s=1}^{t}\varepsilon_s + C^*(L)\,\varepsilon_t, \tag{0.2}$$

having defined: $C = \sum_{j=0}^{\infty} C_j$, $C^*(L) = \sum_{j=0}^{\infty} C_j^* L^j$ and $C_j^* = \sum_{k=j+1}^{\infty} C_k$.

Assumption

It holds that: (i) $\text{rank}\,(C) = m$, *where* $0 \leq m \leq N$*; (ii)* $\left\| C_j \right\| = O\left(\rho^j\right)$ *for some* $0 < \rho < 1$.

It is always possible to write $C = PQ$, where P and Q are full rank matrices of dimensions $N \times m$ and $m \times N$ respectively.
Defining the m-dimensional process $x_t = Q \sum_{s=1}^{t} \varepsilon_s$, and using the short-hand notation $u_t = C^* (L)\, \varepsilon_t$, we rewrite (0.2) as

$$y_t = Px_t + u_t. \tag{0.3}$$

Theory/"Main" assumption

Assumption

It holds that: (i) ε_t is an i.i.d. sequence; (ii) for all nonzero vectors $l \in \mathbb{R}^N$, $l'\varepsilon_t$ has distribution $F_{l\varepsilon}$ with strictly positive density, which belongs in the domain of attraction of a strictly stable law G with index $0 < \eta \leq 2$.

Note:

- assumption is rather standard;
- tail index is η, as you can see infinite mean is even allowed for;
- we need iid, as is typical in this literature;
- we do not need symmetry, unlike the literature.

Let $S_{11} = \sum_{t=1}^{T} y_t y_t'$,

Proposition

We assume that Assumptions 1-2 are satisfied. Then there exists a random variable T_0 such that, for all $T \geq T_0$

$$\lambda^{(j)}(S_{11}) \geq c_0 \frac{T^{1+2/\eta}}{(\ln \ln T)^{2/\eta}} \text{ for } j \leq m. \tag{0.4}$$

Also, for every $\epsilon > 0$, it holds that

$$\lambda^{(j)}(S_{11}) = o_{a.s.}\left(T^{2/p} (\ln T)^{2(2+\epsilon)/p}\right) \text{ for } j > m, \tag{0.5}$$

for every $0 < p < \eta$ when $\eta \leq 2$ with $E\|\varepsilon_t\|^\eta = \infty$, and $p = 2$ when $\eta = 2$ and $E\|\varepsilon_t\|^\eta < \infty$.

Theory: asymptotics/2

Let $S_{00} = \sum_{t=1}^{T} \Delta y_t \Delta y_t'$.

Assumption

ε_t has density $p_\varepsilon(u)$ such that $\int_{-\infty}^{+\infty} |p_\varepsilon(u+y) - p_\varepsilon(u)|\, du \leq c_0 \|y\|$.

Proposition

We assume that Assumptions 1-3 are satisfied. Then

$$\lambda^{(1)}(S_{00}) = o_{a.s.}\left(T^{2/\eta} \left(\prod_{i=2}^{j} \ln_i T \right)^{2/\eta} \left(\ln_{j+1} T \right)^{(2+\epsilon)/\eta} \right), \tag{0.6}$$

for every $\epsilon > 0$ and every integer $j \geq 2$. Also, there exists a random variable T_0 such that, for all $T \geq T_0$

$$\lambda^{(N)}(S_{00}) \geq c_0 \frac{T^{2/\eta}}{(\ln T)^{(2/\eta - 1)(2+\epsilon)}}, \tag{0.7}$$

for every $\epsilon > 0$.

Theory: asymptotics/3

We would use (the spectrum of) $S_{00}^{-1}S_{11}$ to be scale-free. Putting the two propositions together

Theorem

Let Assumptions 1-3 hold. Then there exists a random variable T_0 such that, for all $T \geq T_0$,

$$\lambda^{(j)}\left(S_{00}^{-1}S_{11}\right) \geq c_0 \frac{T}{(\ln\ln T)^{2/\eta}\left(\prod_{i=1}^{n}\ln_i T\right)^{2/\eta}(\ln_{n+1} T)^{(2+\epsilon)/\eta}}, \text{ for } 0 \leq j \leq m, \tag{0.8}$$

for every $\epsilon > 0$. Moreover, for all $0 < p < \eta$ and every $\epsilon, \epsilon' > 0$,

$$\lambda^{(j)}\left(S_{00}^{-1}S_{11}\right) = o_{a.s.}(T^{\epsilon'}(\ln T)^{(2+\epsilon)(2/\eta+2/p-1)}), \text{ for } j > m. \tag{0.9}$$

The test/1

Based on Theorem 1, we propose to use

$$\phi_T^{(j)} = \exp\left\{T^{-\kappa}\lambda^{(j)}\left(S_{00}^{-1}S_{11}\right)\right\} - 1, \tag{0.10}$$

where $\kappa \in (0,1)$.
On account of Theorem 1, it holds that

$$P\left(\omega : \lim_{T\to\infty} \phi_T^{(j)} = \infty\right) = 1 \text{ for } 0 \leq j \leq m,$$
$$P\left(\omega : \lim_{T\to\infty} \phi_T^{(j)} = 0\right) = 1 \text{ for } j > m,$$

so that we can assume from now on that

$$\lim_{T\to\infty} \phi_T^{(j)} = \infty \text{ for } 0 \leq j \leq m, \tag{0.11}$$
$$\lim_{T\to\infty} \phi_T^{(j)} = 0 \text{ for } j > m. \tag{0.12}$$

The test/2

We propose (a sequence of) tests for

$$\begin{cases} H_0 : m \geq j \\ H_A : m < j \end{cases} \tag{0.13}$$

We present the construction of the test statistic as a three step algorithm.

Step 1 Generate an artificial sample $\left\{\xi_i^{(j)}, 1 \leq i \leq M\right\}$, with $\xi_i^{(j)} \sim i.i.d. N\left(0,1\right)$, independent across j and independent of the original data.
Step 2 For each $u \in U$, define the Bernoulli sequence $\zeta_i^{(j)}\left(u\right) = I\left(\phi_T^{(j)}\xi_i^{(j)} \leq u\right)$, and let

$$\theta_{T,M}^{(j)}\left(u\right) = \frac{2}{\sqrt{M}}\sum_{i=1}^{M}\left(\zeta_i^{(j)}\left(u\right) - \frac{1}{2}\right). \tag{0.14}$$

Step 3 Compute

$$\Theta_{T,M}^{(j)} = \int_U \left[\theta_{T,M}^{(j)}\left(u\right)\right]^2 dF\left(u\right), \tag{0.15}$$

where $F\left(u\right)$ is a user-defined weight function.

The test /3

Let P^* denote the conditional probability with respect to the original sample; we use the notation "$\overset{D^*}{\to}$" and "$\overset{P^*}{\to}$" to define, respectively, conditional convergence in distribution and in probability according to P^*.

Theorem

We assume that Assumptions 1-4 are satisfied. If H_0 holds, then, as $\min(T, M) \to \infty$ with

$$M^{1/2} \exp\left(-T^{1-\kappa-\epsilon}\right) \to 0, \tag{0.16}$$

for any arbitrarily small $\epsilon > 0$, it holds that

$$\Theta_{T,M}^{(j)} \overset{D^*}{\to} \chi_1^2, \tag{0.17}$$

for each j, for almost all realisations of $\{\varepsilon_t, 0 < t < \infty\}$. Under H_A, as $\min(T, M) \to \infty$, it holds that

$$M^{-1}\Theta_{T,M}^{(j)} \overset{P^*}{\to} \frac{1}{4}, \tag{0.18}$$

for each j, for almost all realisations of $\{\varepsilon_t, 0 < t < \infty\}$.

Estimating m

The estimator of m (say $\widehat{m}$) is the output of the following algorithm:

Step 1 Run the test for $H_0 : m \geq 1$ based on $\Theta^{(j)}_{T,M}$. If the null is rejected, set $\widehat{m} = 0$ and stop, otherwise go to the next step.
Step 2 Starting from $j = 1$, run the test for $H_0 : m \geq j$ based on $\Theta^{(j+1)}_{T,M}$, constructed using an artificial sample $\left\{\xi_i^{(j+1)}\right\}_{i=1}^{M}$ generated independently of $\left\{\xi_i^{(1)}\right\}_{i=1}^{M}$, ..., $\left\{\xi_i^{(j)}\right\}_{i=1}^{M}$. If the null is rejected, set $\widehat{m} = j$ and stop; otherwise repeat the step until the null is rejected.

Theorem

We assume that Assumptions 1-4 are satisfied. Define the level of each individual test as $\alpha = \alpha(T)$. *As* $\min(T, M) \to \infty$ *under (0.16), if* $\alpha(T) \to 0$, *then it holds that* $P^*(\widehat{m} = m) = 1$ *for almost all realisations of* $\{\varepsilon_t, -\infty < t < \infty\}$.

Extension: estimation of m in heteroskedastic environments

Considering the case of heterogeneous innovations, viz.

$$\varepsilon_t = h\left(\frac{t}{T}\right) u_t, \tag{0.19}$$

Assumption

The function $h(\cdot)$ is a nontrivial, nonnegative function of bounded variation on $[0, 1]$.

Corollary

We assume that Assumptions 1-5 are satisfied, with Assumption 2 modified to contain only symmetric stable u_t. Then, as $\min(T, M) \to \infty$ with (0.16), it holds that, for all j

$$P^*\left(\Theta_{T,M}^{(j)} > c_\alpha\right) \to \alpha, \tag{0.20}$$

under H_0, with probability tending to 1. Under H_A, (0.18) holds for each j, for almost all realisations of $\{u_t, 0 < t < \infty\}$.

Estimation of common trends

Recalling (0.3)

$$y_t = Px_t + u_t,$$

a "natural" estimator of the common trends x_t can be obtained using Principal Components. Let $\widehat{v}_j$ denote the eigenvector corresponding to the j-th largest eigenvalue of S_{11} under the restrictions $\left\|\widehat{v}_j\right\| = 1$ and $\widehat{v}_i'\widehat{v}_j = 0$ for all $i \neq j$, and constructing $\widehat{P} = [\widehat{v}_1|...|\widehat{v}_m]$, the estimator of the common trends is

$$\widehat{x}_t = \widehat{P}' y_t.$$

Theorem

We assume that Assumptions 1-5 are satisfied. Then it holds that

$$\left\|\widehat{x}_t - H^{-1}x_t\right\| = O_P(1) + O_P\left(T^{-1+1/\eta}\right),$$

where H is an $N \times N$ invertible matrix.

The large N case/1

We extend our analysis by proposing a novel approach to determine m in the large N case. We can make use of the non-stationary factor representation

$$y_t = \Lambda F_t + u_t, \tag{0.21}$$

where $\Lambda = (\lambda_1 | ... | \lambda_N)'$ is an $N \times m$ matrix of loadings, F_t is an $m \times 1$ vector of non-stationary factors (with $m < \infty$), and $u_t = \left(u_{1,t}, ..., u_{N,t}\right)'$ is an N-dimensional vector of idiosyncratic shocks.

As before, F_t is a vector-valued stochastic trend, and we assume an MA structure for the $u_{i,t}$s, i.e.

$$F_t = F_{t-1} + u_t^F, \text{ and } u_{i,t} = \sum_{j=0}^{\infty} c_{i,j}^u v_{i,t-j}.$$

The large N case/3

To deal with the large N case, however, as typical of factor models, we now make the simplifying assumption of independence between the common factors F_t and the idiosyncratic component u_t.

Assumption

It holds that: (i) both $\left\{u_t^F\right\}$ *and* $\left\{u_{i,t}\right\}$ *satisfy Assumption 2; (ii)* $\left\{u_t^F\right\}$ *and* $\left\{u_{i,t}\right\}$ *are two mutually independent groups, for all* $1 \leq i \leq N$.

Assumption

The loadings λ_i *are non-random* $m \times 1$ *vectors with* $m < \infty$, *and such that: (i)* $\|\lambda_i\| < \infty$, $1 \leq i \leq N$; *(ii)* $\lim_{N\to\infty} N^{-1}\Lambda'\Lambda = \Sigma_\Lambda$, *with* Σ_Λ *an* $m \times m$ *positive definite matrix.*

Assumption

It holds that (i) as $\min(N, T) \to \infty$, $(NT)^{-2/\eta} \sum_{i=1}^{N} \sum_{t=1}^{T} \Delta u_{i,t}^2 \xrightarrow{w} G_{\eta/2}$; *and (ii) for all nonzero vectors* $l \in \mathbb{R}^m$, *as* $T \to \infty$, $T^{-2/\eta} \sum_{t=1}^{T} (l'\Delta F_t)^2 \xrightarrow{w} G^*_{\eta/2}$.

The large N case/4

Proposition

Let Assumptions 5-6 hold. Then there exist two random variables N_0 and T_0 such that, for all $N \geq N_0$ and $T \geq T_0$

$$\lambda^{(j)}(S_{11}) \geq c_0 \frac{NT^{1+2/\eta}}{(\ln \ln T)^{2/\eta}}, \text{ for } j \leq m, \tag{0.22}$$

Also, for every $\epsilon > 0$, it holds that

$$\lambda^{(j)}(S_{11}) = o_{a.s.}\left((NT)^{2/p} (\ln N \ln T)^{2(2+\epsilon)/p}\right), \text{ for } j > m, \tag{0.23}$$

for every $0 < p < \eta$ when $\eta \leq 2$ with $E\left|\varepsilon_{i,t}\right|^{\eta} = \infty$, and $p = 2$ when $\eta = 2$ with $E\left|\varepsilon_{i,t}\right|^{2} < \infty$.

The large N case/5

So there exists a gap between the m largest eigenvalues of S_{11} and the remaining ones as long as

$$\lim_{\min(N,T)\to\infty} \frac{(NT)^{2/p} (\ln N \ln T)^{2(2+\epsilon)/p} (\ln \ln T)^{2/\eta}}{NT^{1+2/\eta}} = 0;$$

in turn, this is implied by

$$\frac{N^{2/\eta-1-\epsilon}}{T} \to 0, \tag{0.24}$$

for any $\epsilon > 0$.

A "natural" statistic to test for $H_0 : m \geq j$ based on rescaling $\lambda^{(j)}(S_{11})$ by the trace of S_{00}, viz.

$$\check{\nu}_{N,T}^{(j)} = T^{-\kappa} \frac{\lambda^{(j)}(S_{11})}{\sum_{k=1}^{N} \lambda^{(k)}(S_{00})}, \tag{0.25}$$

where $\kappa > 0$ is user-defined (and arbitrarily small), and use $\check{\phi}_{N,T}^{(j)} = \exp\left(\check{\nu}_{N,T}^{(j)}\right) - 1$ to carry out the test.

Theorem

Let Assumptions 4-7 and (0.24) hold. As $\min(N, T, M) \to \infty$ *under (0.16), it holds that* $P^*(\check{m} = m) \to 1$ *with probability tending to* 1.

Just comments:

- works very well, irrespective of η
- works very well, even when $m = 0$ and $m = N$
- works well when N increases, but needs T bigger and bigger in that case.

We consider a set of $N = 7$ commodity prices: three oil prices (WTI, Brent crude, and Dubai crude) and the prices of four metals (copper, gold, nickel, and cobalt). But why don't I show you some numbers....

We evaluate the presence and number of common stochastic trends in the yield curve.
We use monthly data with maturities from 6 months up to 100 years ($N = 196$), spanning the period from January 1985 to September 2018 ($T = 405$).
But why don't I show you some numbers....

Thank you!