



Inference in Heavy-Tailed Nonstationary Multivariate Time Series

Matteo Barigozzi, Giuseppe Cavaliere & Lorenzo Trapani

To cite this article: Matteo Barigozzi, Giuseppe Cavaliere & Lorenzo Trapani (2022): Inference in Heavy-Tailed Nonstationary Multivariate Time Series, Journal of the American Statistical Association, DOI: [10.1080/01621459.2022.2128807](https://doi.org/10.1080/01621459.2022.2128807)

To link to this article: <https://doi.org/10.1080/01621459.2022.2128807>



© 2022 The Author(s). Published with license by Taylor & Francis Group, LLC.



[View supplementary material](#)



Published online: 04 Nov 2022.



[Submit your article to this journal](#)



Article views: 781



[View related articles](#)



[View Crossmark data](#)

Inference in Heavy-Tailed Nonstationary Multivariate Time Series

Matteo Barigozzi^a , Giuseppe Cavaliere^{a,b} , and Lorenzo Trapani^c 

^aUniversity of Bologna, Bologna, Italy; ^bUniversity of Exeter Business School, Exeter, UK; ^cUniversity of Nottingham, Nottingham, UK

ABSTRACT

We study inference on the common stochastic trends in a nonstationary, N -variate time series y_t , in the possible presence of heavy tails. We propose a novel methodology which does not require any knowledge or estimation of the tail index, or even knowledge as to whether certain moments (such as the variance) exist or not, and develop an estimator of the number of stochastic trends m based on the eigenvalues of the sample second moment matrix of y_t . We study the rates of such eigenvalues, showing that the first m ones diverge, as the sample size T passes to infinity, at a rate faster by $O(T)$ than the remaining $N - m$ ones, irrespective of the tail index. We thus exploit this eigen-gap by constructing, for each eigenvalue, a test statistic which diverges to positive infinity or drifts to zero according to whether the relevant eigenvalue belongs to the set of the first m eigenvalues or not. We then construct a randomized statistic based on this, using it as part of a sequential testing procedure, ensuring consistency of the resulting estimator of m . We also discuss an estimator of the common trends based on principal components and show that, up to an invertible linear transformation, such estimator is consistent in the sense that the estimation error is of smaller order than the trend itself. Importantly, we present the case in which we relax the standard assumption of iid innovations, by allowing for heterogeneity of a very general form in the scale of the innovations. Finally, we develop an extension to the large dimensional case. A Monte Carlo study shows that the proposed estimator for m performs particularly well, even in samples of small size. We complete the article by presenting two illustrative applications covering commodity prices and interest rates data. Supplementary materials for this article are available online.

ARTICLE HISTORY

Received August 2021
Accepted September 2022

KEYWORDS

Factor models; Heavy tails;
Nonstationarity; Randomized
tests


1. Introduction

Since the seminal works by Engle and Granger (1987), Stock and Watson (1988), and Johansen (1991), determining the presence and number m of common stochastic trends has become an essential step in the analysis of multivariate time series which are nonstationary over time. Inference on m is of great importance on its own, as it has a “natural” interpretation in many applications: for example, it can provide the number of nonstationary factors in Nelson-Siegel type term structure models, or it can allow to assess the presence of (long-run) integration among financial markets (Kasa 1992). Available estimators are based either on sequential testing (see, e.g., Johansen 1991), or on information criteria (see, e.g., Qu and Perron 2007), and—with few exceptions—strongly rely on the assumption that some moments of the data (the second, or even the fourth) exist. This assumption, however, often lacks empirical support, and data exhibiting heavy tails, which do not have finite second (or even first) moment, are often encountered in many areas: macroeconomics (Ibragimov and Ibragimov 2018), finance (Davis 2010), urban studies (Gabaix 1999), as well as insurance, telecommunication network traffic and meteorology (see, e.g., Embrechts, Klüppelberg, and Mikosch 2013).

Violation of the moment assumptions may result in (possibly severe) incorrect determination of the number of common

trends—see, for example, the simulations in Caner (1998) and the empirical evidence in Falk and Wang (2003). Unfortunately, contributions which explicitly deal with inference on common stochastic trends under infinite variance are rare. Caner (1998) derives the asymptotic distribution of Johansen’s trace test under infinite variance and shows that it depends on the (unknown) tail index of the data. She and Ling (2020) study the (nonstandard) rate of convergence of estimators in nonstationary Vector AutoRegression (VAR) models and show that the limiting distributions depend on the tail index in a nontrivial fashion; similar results are also found in Paulauskas and Rachev (1998), Fasen (2013), and Chan and Zhang (2012) in the contest of least squares estimation of nonstationary autoregressions driven by innovations with heavy tails (see also Davis and Resnick 1985; Davis and Resnick 1986; and Hall, Peng, and Yao 2002). In the univariate case, Jach and Kokoszka (2004) and Cavaliere, Georgiev, and Taylor (2018) show that suitable bootstrap approaches could be used to test whether data are driven by a stochastic trend; knowledge of the tail index is not needed, but extensions of these bootstrap approaches to multiple time series are not available, and are likely to be very hard to develop. Distribution-free approaches could also help overcome this difficulty. Hallin, van den Akker, and Werker (2016) (see also Hallin, Van den Akker, and Werker 2011) apply the rank transformation to the residuals of a Vector Error Correction Model

CONTACT Lorenzo Trapani  lorenzo.trapani@nottingham.ac.uk  University of Nottingham, Nottingham, UK.

 Supplementary materials for this article are available online. Please go to www.tandfonline.com/r/JASA.

© 2022 The Author(s). Published with license by Taylor & Francis Group, LLC.

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

(VECM), obtaining nuisance-free statistics, but this approach requires the correct specification of the VAR.

Our key contribution is the estimator of the number m of common trends of an N -variate time series in the possible presence of heavy tails. Crucially, our procedure does not require any a priori knowledge as to whether the variance is finite or not, or as to how many moments exist, thus, avoiding having to estimate any nuisance parameters or even pre-testing for (in)finite moments. We also show that, in contrast to most of the literature on time series with heavy tails, our methodology also applies to time series with *heterogeneous* innovations. Specifically, we allow for changes in the scale of the innovations of a very general form, which covers, for example, multiple shifts and smooth scale changes. As far as we are aware, this article is the first one where heterogeneity in the scale is allowed under infinite variance.

A heuristic preview of how the methodology works is as follows. The starting point of our analysis is a novel result concerning the properties of the sample second moment matrix of the data in levels (see also Davis, Pfaffel, and Stelzer 2014). We show that the m largest eigenvalues of the matrix diverge to positive infinity, as the sample size T passes to infinity, faster than the remaining eigenvalues by a factor (almost) equal to T . Importantly, this result always holds, irrespective of the variance of the innovations being finite or infinite. Building on this, for each eigenvalue we construct a statistic which diverges to infinity under the null that the eigenvalue is diverging at a “fast” rate, and drifts to zero under the alternative that the relevant eigenvalue diverges at a “slow” rate. Although the limiting distribution of our statistic is bound to depend on nuisance parameters such as the tail index, the relative rate of divergence between the null and the alternative does not depend on any nuisance parameters. Therefore, in order to construct a test, we can rely on rates only, and randomize our statistic using a similar approach to Bandi and Corradi (2014). Thence, our estimator of m is based on running the tests sequentially; in this respect, it mimics the well-known sequential procedure advocated in Johansen (1991) for the determination of the rank of a cointegrated system. Our methodology has at least four desirable features. First, as mentioned above, our technique can be applied to data with infinite variance (and even infinite expectation), with no need to know this a priori. Second, our procedure does not require at any stage the estimation of the tail index of a distribution, which is notoriously delicate. Thirdly, our procedure does not require the correct specification of the lag structure of the underlying VECM model, and it is therefore robust to misspecification of the dynamics. Finally, our results are based only on rates, which makes our procedure extremely easy to implement in practice.

As a final remark, we point out that our approach shares some commonalities with the literature on large dimensional factor models, where the spectrum of the covariance matrix of the data is employed to estimate the number of common factors (see also Zhang, Robinson, and Yao 2019; and Tu, Yao, and Zhang 2020). Whilst the main focus of our paper is on the fixed-dimensional case $N < \infty$, the high-dimensional case $N \rightarrow \infty$ is also relevant and, to the best of our knowledge, the literature is virtually silent on this topic. Inference on common stochastic trends with large N has been developed, either extending the VAR/VECM set-up (see, e.g., Onatski and Wang 2018; Liang and Schienle 2019;

and Bykhovskaya and Gorin 2022), or considering panel factor models (Bai 2004; and Onatski and Wang 2021), but all these contributions assume the existence of higher order moments. To the best of our knowledge, the only papers to deal with heavy tailed observations in the high-dimensional case are the ones by Fan, Liu, and Wang (2018), Yu, He, and Zhang (2019), and He et al. (2022); however, all these papers assume a specific family of distributions (the elliptical distribution family), and consider stationary data only (we also refer to the paper by Chen, Dolado, and Gonzalo 2021, whose estimators are robust in the presence of heavy-tails in the idiosyncratic errors, under the assumption that these are conditionally independent). Extensions to the case of nonstationary data are highly nontrivial; building on our approach, we also consider inference on m in the case of a large factor model, where $N \rightarrow \infty$.

The remainder of the article is organized as follows. Assumptions and preliminary asymptotics are provided in Section 2. The main results on the number of common trends (and the estimation of common trends and loadings) are presented in Section 3; extensions (including the large dimensional case) are in Section 4. We provide Monte Carlo evidence in Section 5, and we validate our methodology through two real data applications in Section 6. Section 7 concludes. Further results and simulations, additional empirical illustrations, technical lemmas and proofs, are in the Supplement.

NOTATION. For a given matrix $A \in \mathbb{R}^{n \times m}$, we denote its element in position (i, j) as A_{ij} ; we use $\|A\|$ to denote its Frobenius norm, that is, $\|A\| = (\sum_{i=1}^n \sum_{j=1}^m A_{ij}^2)^{1/2}$; we also let $\lambda^{(j)}(A)$ denote the j th largest eigenvalue of A . We denote with c_0, c_1, \dots positive, finite constants whose value can change from line to line. The backshift operator for a time series z_t is denoted as L , with $L^k z_t = z_{t-k}$, and Δ denotes the first difference operator, that is, $\Delta z_t = z_t - z_{t-1}$. Given a scalar random variable X , we denote its L_p -norm as $|X|_p$, that is, $|X|_p = (E|X|^p)^{1/p}$. We use $\ln_k x$ to denote the k -iterated logarithm of x truncated at zero—for example, $\ln_2 x = \max\{\ln \ln x, 0\}$; $\lfloor x \rfloor$ denotes the largest integer not greater than x ; and the indicator function is denoted as $I(\cdot)$. Finally, “ \xrightarrow{w} ” denotes weak convergence. Other notation is introduced later on in the article.

2. Theory

Consider an N -dimensional vector y_t with $MA(\infty)$ representation

$$\Delta y_t = C(L) \varepsilon_t, \quad (2.1)$$

where $C(L) = \sum_{j=0}^{\infty} C_j L^j$, and ε_t is a sequence of iid errors. We assume that $y_0 = 0$ and $\varepsilon_t = 0$ for $t \leq 0$, for simplicity and with no loss of generality; a quick inspection of our proofs reveals that all the results derived here can be extended to more general assumptions concerning, for example, the initial value y_0 . Standard arguments based on the multivariate Beveridge-Nelson decomposition of the filter $C(L)$ (see Watson 1994) allow to represent (2.1) as

$$y_t = C \sum_{s=1}^t \varepsilon_s + C^*(L) \varepsilon_t \quad (2.2)$$

where $C = \sum_{j=0}^{\infty} C_j$, $C^*(L) = \sum_{j=0}^{\infty} C_j^* L^j$, $C_j^* = -\sum_{k=j+1}^{\infty} C_k$.

We assume that the $N \times N$ matrix C can have reduced rank, say m . This corresponds to assuming that the long-run behavior of the N -dimensional vector y_t is driven by m nonstationary common factors.

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

By definition, $N - m$ is the rank of cointegration of (2.1)–(2.2). The case $m = 0$ in part (i) corresponds to y_t being (asymptotically) strictly stationary. Conversely, the case $m = N$ implies that y_t is driven by N distinct random walks, and consequently no cointegration between the components of y_t is present. Part (ii) of the assumption requires that the MA coefficients C_j decline geometrically. This is similar to Assumption 1 in Caner (1998), where the C_j s are assumed to decline at a rate which increases as the tail index of the innovations ε_t decreases. Assumption 1 is also implied by Assumption 2.1 in She and Ling (2020), where a finite-order VAR model under the classic $I(1)$ conditions stated, for example, in Ahn and Reinsel (1990) is considered.

Under Assumption 1, on account of the possible rank reduction of C , (2.2) can be given a factor model representation, where the common factors capture the nonstationary behavior of the data. Specifically, since it is always possible to write $C = \Lambda Q'$, where Λ and Q are full rank matrices of dimension $N \times m$, we can define the m -dimensional process $F_t = Q' \sum_{s=1}^t \varepsilon_s$, and using the short-hand notation $u_t = C^*(L) \varepsilon_t$, we can write (2.2) as

$$y_t = \Lambda F_t + u_t, \tag{2.3}$$

see also Stock and Watson (1988), where F_t is a vector ($m \times 1$) of integrated processes and u_t is a serially correlated, zero mean, $I(0)$ process. Hence, the model has strong similarities with factor models with nonstationary dynamic factors—see, for example, Bai (2004), where the latent factors correspond to the set of m common stochastic trends F_t . However, with respect to models with nonstationary factors as in Bai (2004), the factors F_t and the error component u_t do not need to be independent; in addition, no moment restrictions, such as the classic finite variance assumption on u_t , are considered here. The relation between (2.3) and factor models in the large N case is considered in Section 4.3 and Section A.3 in the supplementary materials.

We now make some assumptions on the error term ε_t .

Assumption 2. It holds that: (i) $\{\varepsilon_t, 1 \leq t \leq T\}$ is an iid sequence; (ii) for all nonzero vectors $l \in \mathbb{R}^N$, $l' \varepsilon_t$ has distribution $F_{l\varepsilon}$ with strictly positive density, which is in the domain of attraction of a strictly stable law G with tail index $0 < \eta \leq 2$.

Assumption 2(i) is standard in the analysis of time series with possibly infinite variance. Part (ii) of the assumption implicitly states that the vector ε_t has a multivariate distribution which belongs to the domain of attraction of a strictly stable, multivariate law (see Theorem 2.1.5(a) in Samorodnitsky and Taquq 1994) with common tail index η . This also implies (by Property 1.2.6 in Samorodnitsky and Taquq 1994) that, when $E|\varepsilon_t| < \infty$, $E(\varepsilon_t) = 0$. Further, when $\eta < 2$, it holds that $E|\varepsilon_{i,t}|^p < \infty$

for all $0 \leq p < \eta$, whereas $E|\varepsilon_{i,t}|^\eta = \infty$ (Petrov 1974). Also, by Property 1.2.15 in Samorodnitsky and Taquq (1994), it holds that

$$F_{l\varepsilon}(-x) = \frac{c_{l,1} + o(1)}{x^\eta} L(x), \quad \text{and}$$

$$1 - F_{l\varepsilon}(x) = \frac{c_{l,2} + o(1)}{x^\eta} L(x),$$

as $x \rightarrow \infty$, where $L(x)$ is a slowly varying function in the sense of Karamata (see Seneta 2006), and $c_{l,1}, c_{l,2} \geq 0$, $c_{l,1} + c_{l,2} > 0$. The condition that G is strictly stable entails $c_{l,1} = c_{l,2}$ when $\eta = 1$, thus ruling out asymmetry (see Property 1.2.8 in Samorodnitsky and Taquq 1994).

2.1. Asymptotics

Define

$$S_{11} = \sum_{t=1}^T y_t y_t', \text{ and } S_{00} = \sum_{t=1}^T \Delta y_t \Delta y_t'. \tag{2.4}$$

We report a set of novel results for the eigenvalues of S_{11} and S_{00} , which we require for the construction of the test statistics.

Proposition 1. Let Assumptions 1–2 hold. Then there exists a random variable T_0 such that, for all $T \geq T_0$

$$\lambda^{(j)}(S_{11}) \geq c_0 \frac{T^{1+2/\eta}}{(\ln \ln T)^{2/\eta}}, \text{ for } j \leq m. \tag{2.5}$$

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

$$\lambda^{(j)}(S_{11}) = o_{\text{a.s.}} \left(T^{2/p} (\ln T)^{2(2+\epsilon)/p} \right), \text{ for } j > m, \tag{2.6}$$

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

In (2.6), p should be viewed as “arbitrarily close to η .” The proposition states that the first m eigenvalues of S_{11} diverge at a faster rate than the other ones (faster by an order of “almost” T), thus, entailing that the spectrum of S_{11} has a “spiked” structure. Heuristically, the impact of having heavy tails is apparent in both equations from the $\frac{2}{\eta}$ (and $\frac{2}{p}$) exponent; similarly, nonstationarity (or “integratedness”) impacts (2.5) via the extra T component, which ensures the spikedness of the spectrum of S_{11} .

In order to study S_{00} and its eigenvalues, we need the following assumption, which complements Assumption 1(ii).

Assumption 3. ε_t has density $p_\varepsilon(x)$ such that $\int_{x \in \mathbb{R}^N} |p_\varepsilon(x+y) - p_\varepsilon(x)| dx \leq c_0 \|y\|$.

The integral Lipschitz condition in Assumption 3 is a technical requirement needed for Δy_t to be strong mixing with geometrically declining mixing numbers, and it is a standard requirement in this literature (see, e.g., Pham and Tran 1985).

Proposition 2. Let Assumptions 1–3 hold. Then

$$\lambda^{(1)}(S_{00}) = o_{\text{a.s.}} \left(T^{2/\eta} \left(\prod_{i=1}^n \ln_i T \right)^{2/\eta} (\ln_{n+1} T)^{(2+\epsilon)/\eta} \right), \tag{2.7}$$

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

$$\lambda^{(N)}(S_{00}) \geq c_0 \frac{T^{2/\eta}}{(\ln T)^{(2/\eta-1)(2+\epsilon)}}. \quad (2.8)$$

Similarly to Propositions 1 and 2 provides bounds for the spectrum of S_{00} . Part (2.7) has been shown in Trapani (2014), where it is shown that the bound in (2.7) is almost sharp. The lower bound implied in (2.8) is also almost sharp.

The spectrum of S_{11} (and, in particular, the different rates of divergence of its eigenvalues) can—in principle—be employed in order to determine m . However, S_{11} is unsuitable for direct usage, for two reasons. First, by Proposition 1, its spectrum depends on the nuisance parameter η . Also, it depends on the unit of measurement of the data, and thus it is not scale-free. In order to construct scale-free and nuisance-free statistics, we propose to rescale S_{11} by S_{00} . The rationale for this can be traced back to the use of multivariate KPSS-type statistics, where (with our notation) the null of no stochastic trends would be tested by contrasting S_{11} with S_{00} through the statistic $S_{00}^{-1}S_{11}$ (see Nyblom and Harvey 2000, and Nielsen 2010).¹

Proposition 2 ensures that this is possible: by Equation (2.8), the inverse of S_{00} cannot diverge too fast, and therefore the spectrum of the matrix $S_{00}^{-1}S_{11}$ should still have m eigenvalues that diverge at a faster rate than the others. This is shown in the next theorem.

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

$$\begin{aligned} & \lambda^{(j)}(S_{00}^{-1}S_{11}) \\ & \geq c_0 \frac{T}{(\ln \ln T)^{2/\eta} \left(\prod_{i=1}^n \ln_i T \right)^{2/\eta} (\ln_{n+1} T)^{(2+\epsilon)/\eta}}, \quad \text{for} \\ & 0 \leq j \leq m, \end{aligned} \quad (2.9)$$

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

$$\lambda^{(j)}(S_{00}^{-1}S_{11}) = o_{a.s.}(T^{\epsilon'} (\ln T)^{(2+\epsilon)(2/\eta+2/p-1)}), \quad \text{for } j > m. \quad (2.10)$$

Theorem 1 states that the spectrum of $S_{00}^{-1}S_{11}$ has a similar structure to the spectrum of S_{11} : the first m eigenvalues are spiked and their rate of divergence is faster than that of the remaining eigenvalues by a factor of almost T . More importantly, by normalizing S_{11} by S_{00} , the nuisance parameter η is relegated to the slowly varying (logarithmic) terms. In essence, apart from the slowly varying sequences, Equations (2.9) and (2.10) imply that the rates of divergence of the eigenvalues of $S_{00}^{-1}S_{11}$ are of order (arbitrarily close to) $O(T)$ for the spiked eigenvalues, and (arbitrarily close to) $O(1)$ for the other ones. This is the key property of $\lambda^{(j)}(S_{00}^{-1}S_{11})$: dividing S_{11} by S_{00} washes out the impact of the tail index η , which essentially does not play any role in determining the divergence or not

¹Another possible scaling would be based on $\tilde{S}_{00}^{-1}S_{11}$, with \tilde{S}_{00} a diagonal matrix whose nonzero elements are the same as those of S_{00} . In Section B.3 in the supplementary materials, we report some Monte Carlo evidence on the finite sample performance of this type of scaling. We are grateful to an anonymous referee for suggesting this.

of $\lambda^{(j)}(S_{00}^{-1}S_{11})$. This result is based on rates, but it is possible to find an analogy between the result in Theorem 1 and approaches based on eliminating nuisance parameters using self-normalisation (see, e.g., Shao 2015).

3. Inference on the Common Trends

In this section we collect our main results about estimation and inference on the common trends in the possible presence of heavy tails. In Section 3.1, we report a novel one-shot test about the (minimum) number of common trends. Then, in Section 3.2 we introduce a sequential procedure for the determination of the number of common trends. Estimation of the common trends and associated factor loadings is presented in Section 3.3.

3.1. Testing Hypotheses on the Number of Common Trends

The tests proposed herein will form the basis of our sequential procedure for the determination of the number of common trends—see Section 3.2. We consider the null and the alternative hypotheses

$$\begin{cases} H_0 : m \geq j \\ H_A : m < j \end{cases} \quad (3.1)$$

where $j \in \{1, \dots, N\}$ is a (user-chosen) lower bound on the number of common trends—for example, a test of non-stationarity against the alternative of strict stationarity corresponds to $j = 1$.

Based on Theorem 1, we propose to use

$$\phi_T^{(j)} = \exp \left\{ T^{-\kappa} \lambda^{(j)}(S_{00}^{-1}S_{11}) \right\} - 1, \quad (3.2)$$

where $\kappa \in (0, 1)$; criteria for the choice of κ in applications are discussed in Section 5. Importantly, the $T^{-\kappa}$ term in (3.2) is used in order to exploit the discrepancy in the rates of divergence of the $\lambda^{(j)}(S_{00}^{-1}S_{11})$ under H_0 and under H_A . In particular, it ensures that $T^{-\kappa} \lambda^{(j)}(S_{00}^{-1}S_{11})$ drifts to zero under H_A (i.e., whenever $j > m$), whereas it still passes to infinity under H_0 (i.e., when $j \leq m$). According to (2.10), this only requires a very small value of κ , which would also allow $\lambda^{(j)}(S_{00}^{-1}S_{11})$ to diverge at a rate close to T under H_0 . On account of Theorem 1, it holds that $P(\omega : \lim_{T \rightarrow \infty} \phi_T^{(j)} = \infty) = 1$ for $0 \leq j \leq m$; hence, we can assume that under the null that $m \geq j$, it holds that $\lim_{T \rightarrow \infty} \phi_T^{(j)} = \infty$. Conversely, under the alternative that $j > m$, we have $P(\omega : \lim_{T \rightarrow \infty} \phi_T^{(j)} = 0) = 1$, so that $\lim_{T \rightarrow \infty} \phi_T^{(j)} = 0$. In essence, $\phi_T^{(j)}$ diverges to positive infinity, or converges (to zero), according to whether $\lambda^{(j)}(S_{00}^{-1}S_{11})$ is “large” or “small”.

Since the limiting law of $\phi_T^{(j)}$ under the null is unknown, we propose a randomized version of it. The construction of the test statistic is based on the following three step algorithm, which requires a user-chosen weight function $F(\cdot)$ with support $U \subseteq \mathbb{R}$. The algorithm we propose below has been already used in several contributions—see, for example, Bandi and Corradi (2014), and the discussion therein. In Section B.4 in the supplementary materials, we also consider different randomization schemes.

Step 1. Generate an artificial sample $\{\xi_i^{(j)}, 1 \leq i \leq M\}$, with $\xi_i^{(j)} \sim \text{iid}N(0, 1)$, independent of the original data.

Step 2. For each $u \in U$, define the Bernoulli sequence $\zeta_i^{(j)}(u) = I(\phi_T^{(j)} \xi_i^{(j)} \leq u)$, and let

$$\theta_{T,M}^{(j)}(u) = \frac{2}{\sqrt{M}} \sum_{i=1}^M \left(\zeta_i^{(j)}(u) - \frac{1}{2} \right). \quad (3.3)$$

Step 3. Compute

$$\Theta_{T,M}^{(j)} = \int_U [\theta_{T,M}^{(j)}(u)]^2 dF(u), \quad (3.4)$$

where $F(\cdot)$ is the user-chosen weight function.

In Step 2, the binary variable $\zeta_i^{(j)}(u)$ is created for several values of $u \in U$, and in Step 3, the resulting statistics $\theta_{T,M}^{(j)}(u)$ are averaged across u , through the weight function $F(\cdot)$, thus eliminating the dependence of the test statistic on an arbitrary value u . The following assumption characterizes $F(\cdot)$.

Assumption 4. It holds that (i) $\int_{u \in U} dF(u) = 1$; (ii) $\int_{u \in U} u^2 dF(u) < \infty$.

A possible choice for $F(\cdot)$ could be a distribution function with finite second moment, for example, a Rademacher distribution with $U = \{-c, c\}$ for some $c > 0$, and $F(c) = F(-c) = 1/2$, or the standard normal distribution function.

Let P^* denote probability conditional on the original sample; we use “ $\xrightarrow{D^*}$ ” and “ $\xrightarrow{P^*}$ ” to define conditional convergence in distribution and in probability according to P^* , respectively.

Theorem 2. Let Assumptions 1–4 hold. Under H_0 , as $\min(T, M) \rightarrow \infty$ with

$$M^{1/2} \exp(-T^{1-\kappa-\epsilon}) \rightarrow 0, \quad (3.5)$$

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

$$\Theta_{T,M}^{(j)} \xrightarrow{D^*} \chi_1^2, \quad (3.6)$$

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

$$4M^{-1} \Theta_{T,M}^{(j)} \xrightarrow{P^*} 1, \quad (3.7)$$

for almost all realizations of $\{\varepsilon_t, 0 < t < \infty\}$.

Theorem 2 provides the limiting behavior of $\Theta_{T,M}^{(j)}$, also illustrating the impact of M on the size and power tradeoff. According to (3.7), the larger M the higher the power. Conversely, upon inspecting the proof, it emerges that $\theta_{T,M}^{(j)}(u)$ contains a noncentrality parameter of order $O(M^{1/2} \exp(-T^{1-\kappa-\epsilon}))$, whence the upper bound in (3.5). We discuss the choice of M in Section 5; here, we note that condition (3.5) is, for example, satisfied whenever $M = \lfloor T^k \rfloor$, for all $k > 0$.

The one-shot test developed in this section has at least three advantages compared to existing methods. First, our approach can also be implement to check (asymptotic) strict stationarity.

Indeed, running the test for $j = 1$ corresponds to the null hypothesis that the data are driven by at least one common trend; rejection supports the alternative of stationarity. Second, running the test with $j = N$ corresponds to the null hypothesis that the N variables do not cointegrate, thus, offering a test for the null of no cointegration against the alternative of (at least) one cointegrating relation. Finally, we point out a further advantage over the well-known method of Johansen (1991). Johansen’s likelihood ratio test allows to test the null of rank R (i.e., of $m = N - R$ common trends), where R is user-chosen, versus the alternative of rank greater than R (i.e., less than $N - R$ common trends). However, while the limiting distribution under the null is well-known, if the true rank is lower than R , then the limiting distribution is different (see Bernstein and Nielsen 2019). Hence, Johansen’s test should be used only if the practitioner knows that the rank cannot be lower than R . In contrast, our test does not have this drawback since the null hypothesis is formulated as a minimum bound on the number of common trends.

A final remark on the test is in order. Letting $0 < \alpha < 1$ denote the nominal level of the test, and defining c_α such that $P(\chi_1^2 > c_\alpha) = \alpha$, an immediate consequence of the theorem is that under H_A it holds that $\lim_{\min(T,M) \rightarrow \infty} P^*(\Theta_{T,M}^{(j)} > c_\alpha) = 1$ for almost all realizations of $\{\varepsilon_t, 0 < t < \infty\}$: the test is consistent under the alternative. Conversely, under H_0 we have, for almost all realizations of $\{\varepsilon_t, 0 < t < \infty\}$

$$\lim_{\min(T,M) \rightarrow \infty} P^*(\Theta_{T,M}^{(j)} > c_\alpha) = \alpha. \quad (3.8)$$

Our test is constructed using a randomization which does not vanish asymptotically, and therefore the asymptotics of $\Theta_{T,M}^{(j)}$ is driven by the added randomness. Thus, different researchers using the same data will obtain different values of $\Theta_{T,M}^{(j)}$ and, consequently, different p -values. To ameliorate this, Horváth and Trapani (2019) suggest to compute $\Theta_{T,M}^{(j)}$ for S iterations, using, at each iteration s , an independent sequence $\{\xi_{i,s}^{(j)}\}$ for $1 \leq j \leq M$ and $1 \leq s \leq S$, thence defining

$$Q_{\alpha,S} = \frac{1}{S} \sum_{s=1}^S I\left(\Theta_{T,M,s}^{(j)} \leq c_\alpha\right). \quad (3.9)$$

Based on standard arguments (see Horváth and Trapani 2019), under H_0 the LIL yields

$$\liminf_{S \rightarrow \infty} \lim_{\min(T,M) \rightarrow \infty} \sqrt{\frac{S}{2 \ln \ln S}} \frac{Q_{\alpha,S} - (1 - \alpha)}{\sqrt{\alpha(1 - \alpha)}} = -1. \quad (3.10)$$

Hence, a “strong rule” to decide in favour of H_0 is

$$Q_{\alpha,S} \geq (1 - \alpha) - \sqrt{\alpha(1 - \alpha)} \sqrt{\frac{2 \ln \ln S}{S}}. \quad (3.11)$$

Decisions made on the grounds of (3.11) have vanishing probabilities of Type I and Type II errors, and are the same for all researchers: having $S \rightarrow \infty$ washes out the added randomness.

3.2. Determining m

In order to determine the number of common trends m , we propose to cast the individual one-shot tests discussed above in a sequential procedure, where different values $j = 1, 2, \dots$ for m are tested sequentially (note that the individual tests must be based on artificial random samples independent across j , see below).

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

Algorithm 1.

Step 1. Run the test for $H_0 : m \geq 1$ based on $\Theta_{T,M}^{(1)}$. If the null is rejected, set $\widehat{m} = 0$ and stop, otherwise go to the next step.

Step 2. Starting from $j = 2$, run the test for $H_0 : m \geq j$ based on $\Theta_{T,M}^{(j)}$, constructed using an artificial sample $\{\xi_i^{(j)}\}_{i=1}^M$ generated independently of $\{\xi_i^{(1)}\}_{i=1}^M, \dots, \{\xi_i^{(j-1)}\}_{i=1}^M$. If the null is rejected, set $\widehat{m} = j - 1$ and stop; otherwise, if $j = N$, set $\widehat{m} = N$; otherwise, increase j and repeat Step 2.

Consistency of the proposed procedure is presented in the next theorem.

Theorem 3. Let Assumptions 1–4 hold and define the critical value of each individual test as $c_\alpha = c_\alpha(M)$. As $\min(T, M) \rightarrow \infty$ under (3.5), if $c_\alpha(M) \rightarrow \infty$ with $c_\alpha = o(M)$, then it holds that $P^*(\widehat{m} = m) = 1$ for almost all realizations of $\{\varepsilon_t, -\infty < t < \infty\}$.

Theorem 3 states that \widehat{m} is consistent, as long as the nominal level α of the individual tests is chosen so as to drift to zero. This can be better understood upon inspecting the proof of the theorem: letting α denote the level of each individual test, in (E.9) in the supplementary materials, we show that, $P^*(\widehat{m} = m) \rightarrow (1 - \alpha)^{N-m}$ a.s. conditionally on the sample, whence the requirement $c_\alpha \rightarrow \infty$, which entails $\alpha \rightarrow 0$.

The theorem can also be read in conjunction with Johansen's procedure (Johansen 1991), and its bootstrap implementations (Cavaliere, Rahbek, and Taylor 2012), whose outcome is an estimate of m , say \widetilde{m} , such that, asymptotically, $P(\widetilde{m} = m) \rightarrow 1 - \alpha$ for a given nominal value α for the individual tests. By (E.9), in our case choosing a non-vanishing nominal level α would yield, as mentioned above, that $P^*(\widehat{m} = m) \rightarrow (1 - \alpha)^{N-m}$ a.s. conditionally on the sample, which depends on the unknown m and is, for $m > 1$, worse than Johansen's procedure. A possible way of correcting this is to note that in our procedure the individual tests are independent (conditional on the sample), and therefore one can use a Bonferroni correction with α/N as nominal level for each test, rather than α . In this case, the same calculations as in the proof of Theorem 3 (and Bernoulli's inequality) yield that $P^*(\widehat{m} = m) \rightarrow (1 - \alpha/N)^{N-m}$ a.s. conditionally on the sample, with

$$(1 - \alpha/N)^{N-m} \geq 1 - \frac{N-m}{N}\alpha \geq 1 - \alpha. \quad (3.12)$$

On the other hand, it is well-known that Bonferroni correction may be conservative. A possible way to obtain the same result as Johansen (1991)— that is, an estimator of m (say \widehat{m}^*) such that $P^*(\widehat{m}^* = m) \rightarrow 1 - \alpha$ a.s. conditionally on the sample, is to run

the individual tests with the same randomness across j . In such a case,² the following proposition holds:

Proposition 3. We assume that the assumptions of Theorem 3, and that the individual tests are implemented using $\{\xi_i, 1 \leq i \leq M\}$ with $\xi_i \sim \text{iid}N(0, 1)$, for all $j \geq 1$ in Step 2 of Algorithm 1. Then, when $m = 0$, it holds that $P^*(\widehat{m}^* = 0) = 1$ for almost all realizations of $\{\varepsilon_t, -\infty < t < \infty\}$. When $m > 0$, it holds that $P^*(\widehat{m}^* = m) = 1 - \alpha$ and $P^*(\widehat{m}^* = 0) = \alpha$, for almost all realizations of $\{\varepsilon_t, -\infty < t < \infty\}$.

In Section A.2 in the supplementary materials, we complement Algorithm 1 by proposing a top-down algorithm, as an alternative to Bonferroni correction and to Proposition 3.

As a final remark, we point out that the one-shot tests of Section 3.1 have no power versus local alternatives. This is due to the fact that they are based on rates. In particular, in our case we are unable to discern random-walk type trends from trends with near unit root components.³ However, our procedure is designed to estimate the number of common *nonstationary* factors; hence, the lack of power against nonstationary, near unit root common factors may not be viewed as an issue in our context.

3.3. Estimation of the Common Trends

Recall the common trend representation provided in (2.3),

$$y_t = \Lambda F_t + u_t.$$

After determining m , it is possible to estimate the nonstationary common stochastic trends F_t by using Principal Components (PC), in a similar fashion to Peña and Poncela (2006) and Zhang, Robinson, and Yao (2019). Let \widehat{v}_j denote the eigenvector corresponding to the j th largest eigenvalue of S_{11} under the orthonormalisation restrictions $\|\widehat{v}_j\| = 1$ and $\widehat{v}_i' \widehat{v}_j = 0$ for all $i \neq j$, and such that the first coordinate of each \widehat{v}_j , say $\widehat{v}_{1,j}$, satisfies $\widehat{v}_{1,j} \geq 0$ to avoid sign indeterminacy. Then, defining $\widehat{\Lambda} = (\widehat{v}_1, \dots, \widehat{v}_m)$, the estimator of the common trends F_t is $\widehat{F}_t = \widehat{\Lambda}' y_t$.

The next theorem provides the consistency (up to a transformation) of the estimators of Λ and F_t . Interestingly, the convergence rate of $\widehat{\Lambda}$ is not affected by the tail index (see She and Ling 2020).

Theorem 4. Let Assumptions 1–4 hold. Then there exists an $N \times N$ invertible matrix H such that, for each $1 \leq t \leq T$

$$\|\widehat{\Lambda} - \Lambda H\| = O_p(T^{-1+\epsilon}), \quad (3.13)$$

$$\|\widehat{F}_t - H^{-1} F_t\| = O_p(1) + O_p(T^{-1+1/p}), \quad (3.14)$$

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

²We are grateful to an anonymous Referee for bringing this very interesting point to our attention.

³We report a more in-depth explanation of this in Section A.1 in the supplementary materials.

Theorem 4 states that both $\widehat{\Lambda}$ and \widehat{F}_t are consistent estimators of Λ and F_t – up to an invertible linear transformation, since it is only possible to provide a consistent estimate of the eigenspace, as opposed to the individual eigenvectors. By (3.13), $\widehat{\Lambda}$ is a superconsistent estimator of (a linear combination of the columns of) Λ . This result, which is the same as in the case of finite variance, is a consequence of the fact that F_t is an “integrated” process, and it is related to the eigen-gap found in **Proposition 1**. Equation (3.13) could also be read in conjunction with the literature on large factor models, where—contrary to our case—it is required that $N \rightarrow \infty$. In that context, Bai (2004) obtains the same result as in (3.13) albeit for the case of finite variance: thus, in the presence of integrated processes, the PC estimator is always superconsistent, irrespective of N passing to infinity or not.

According to (3.14), \widehat{F}_t also is a consistent estimator of the space spanned by F_t . The “noise” component does not drift to zero and, when $\eta < 1$, it may even diverge; however, the “signal” F_t is of order $O_P(t^{1/p})$, thus, dominating the estimation error (in fact, when $\eta < 1$, the estimation error is smaller by a factor T). This result can be compared to the estimator proposed by Gonzalo and Granger (1995), which is studied under finite second moment and requires a full specification of the VECM, and with the findings in the large factor models literature (see Lemma 2 in Bai 2004). As far as uniform rates in t are concerned, in the proof of the theorem we also show that $\max_{1 \leq t \leq T} \|\widehat{F}_t - H^{-1}F_t\| = O_P(T^{1/p})$. This arises from the fact that the maximum of a T -dimensional sequence with finite p th moment is bounded by $O_P(T^{1/p})$.

4. Extensions

The framework developed in the previous section does not allow for deterministic terms in the data, and requires ε_t to be identically distributed. We now discuss possible extensions of our set-up, to accommodate for heterogeneous innovations and deterministic terms, showing that our procedure can be used even in these cases, with no modifications required. Moreover, we consider the extension to the large N case.

4.1. Heterogeneous Innovations

We consider a novel framework where we allow for innovation heterogeneity of a very general form. Specifically, we assume that

$$\varepsilon_t = h\left(\frac{t}{T}\right) v_t, \tag{4.1}$$

where v_t satisfies **Assumption 2** and $h(\cdot)$ is a deterministic function. The representation in (4.1) has also been employed in order to deal with heteroscedasticity in data with finite variance (see, e.g., Cavaliere and Taylor 2009; and Patilea and Raïssi 2014).

Assumption 5. $h(\cdot)$ is nontrivial, nonnegative and of bounded variation on $[0, 1]$.

The only requirement on the scale function $h(\cdot)$ is that it has bounded variation on $[0, 1]$. The design in (4.1) includes several potentially interesting cases: $h(\cdot)$ can be piecewise linear, that

is, $h(r) = \sum_{i=1}^n h_i I(c_{i-1} \leq r < c_i)$, with $c_0 = 0$ and $c_n = 1$, thus, considering the possible presence of jumps/regimes in the heterogeneity of ε_t ; or it could be a polynomial function.

Corollary 1. Let **Assumptions 1–5** hold, with **Assumption 2** modified to contain only symmetric stable v_t . Then, as $\min(T, M) \rightarrow \infty$ with (3.5), it holds that, for all j

$$P^*(\Theta_{T,M}^{(j)} > c_\alpha) \rightarrow \alpha, \tag{4.2}$$

under H_0 , with probability tending to 1. Under H_A , (3.7) holds for each j , for almost all realizations of $\{v_t, 0 < t < \infty\}$.

Repeating *verbatim* the proof of **Theorem 3**, the results in **Corollary 1** entail that, using the Algorithm 1 in **Section 3.2**, $P^*(\widehat{m} = m) \rightarrow 1$ with probability tending to 1: \widehat{m} is still a consistent estimator of m .

4.2. Deterministics

We consider the representation

$$y_t = \widetilde{\mu} + C \sum_{s=1}^t \varepsilon_s + C^*(L) \varepsilon_t \tag{4.3}$$

where C and $C^*(L)$ are defined as before. Equation (2.2) is derived from the multivariate Beveridge-Nelson decomposition of $C(L)$, and it can also be obtained from a VECM representation (see She and Ling 2020; Yap and Reinsel 1995)

$$\Delta y_t = \mu + \alpha \beta' y_{t-1} + \sum_{j=1}^{p-1} \Gamma_j \Delta y_{t-j} + \varepsilon_t, \tag{4.4}$$

under the constraint $\mu = \alpha \rho$ with ρ an $(N - m) \times 1$ vector. In this case, our procedure still yields the same results as without the deterministic term.

Corollary 2. Let (4.4) hold. Then, **Theorems 2–4** hold under the same assumptions.

4.3. Large Dimensional Vector-Valued Series

In this section, we extend our analysis by proposing a novel approach to determine m in the large N case. We focus on the case $m < \infty$.

As mentioned in the introduction (see also **Section 2**), in the context of large N , we can make use of the nonstationary factor representation (2.3)

$$y_t = \Lambda F_t + u_t, \tag{4.5}$$

where $\Lambda = (\lambda_1, \dots, \lambda_N)'$ is an $N \times m$ matrix of loadings, F_t is an $m \times 1$ vector of nonstationary factors, and $u_t = (u_{1,t}, \dots, u_{N,t})'$ 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, that is.

$$F_t = F_{t-1} + u_t^F, \text{ and } u_{i,t} = \sum_{j=0}^{\infty} c_{i,j}^u v_{i,t-j}. \tag{4.6}$$

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 .⁴ For completeness, we discuss the case of large dimensional models without the assumption of independence between common factors and idiosyncratics in Section A.3 in the supplementary materials.

Assumption 6. It holds that: (i) both $\{u_t^F\}$ and $\{v_{i,t}\}$ satisfy **Assumption 2**; (ii) $\{u_t^F\}$ and $\{v_{i,t}\}$ are two mutually independent groups, for all $1 \leq i \leq N$; (iii) $|c_{ij}^u| = O(\rho^j)$, $1 \leq i \leq N$, for some $0 < \rho < 1$.

Assumption 7. The loadings λ_i are nonrandom $m \times 1$ vectors such that: (i) $\|\lambda_i\| < \infty$, $1 \leq i \leq N$; (ii) $\lim_{N \rightarrow \infty} N^{-1} \Lambda' \Lambda = \Sigma_\Lambda$, with Σ_Λ an $m \times m$ positive definite matrix; (iii) m is finite and independent of N and T .

Assumption 8. It holds that (i) as $\min(N, T) \rightarrow \infty$, $(NT)^{-2/\eta} \sum_{i=1}^N \sum_{t=1}^T \Delta u_{i,t}^2 \xrightarrow{w} G_{\eta/2}$; and (ii) for all nonzero vectors $l \in \mathbb{R}^m$, as $T \rightarrow \infty$, $T^{-2/\eta} \sum_{t=1}^T (l' \Delta F_t)^2 \xrightarrow{w} G_{\eta/2}^*$, where $G_{\eta/2}$ and $G_{\eta/2}^*$ are two independent strictly stable laws with tail index $\eta/2$.

Assumption 6 states that both common factors and idiosyncratic components have heavy tails, in the same way as (2.2). We allow the idiosyncratic components to be autocorrelated, as is typical in large factor models. Part (ii) of the assumption is also standard in the literature, and it is the same as Assumptions D in Bai (2004). **Assumption 7(ii)** essentially considers only strong, or pervasive, common factors. Further, assuming, as is typical in this literature, that m is finite entails that the rank of cointegration $N - m$ diverges with N . Thus, our results complement the analysis by Onatski and Wang (2018) and Bykhovskaya and Gorin (2022), where the case of $N - m < \infty$ is studied instead. Finally, **Assumption 8** is a high-level assumption, which could be shown under more primitive conditions (see, e.g., McElroy and Politis 2003).

Proposition 4. Let **Assumptions 6–7** hold. Then there exist two random variables N_0 and T_0 such that, for all $N \geq N_0$ and $T \geq T_0$

$$\lambda^{(j)}(S_{11}) \geq c_0 \frac{NT^{1+2/\eta}}{(\ln \ln T)^{2/\eta}}, \text{ for } j \leq m, \quad (4.7)$$

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

$$\lambda^{(j)}(S_{11}) = o_{a.s.} \left((NT)^{2/p} (\ln N \ln T)^{2(2+\epsilon)/p} \right), \text{ for } j > m, \quad (4.8)$$

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

According to **Proposition 4**, there exists a gap between the m largest eigenvalues of S_{11} and the remaining ones as long as

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

⁴We are grateful to a Referee for suggesting this alternative to the setup in **Section 2** to us.

in turn, this is implied by

$$\frac{N^{2/\eta-1-\epsilon}}{T} \rightarrow 0, \quad (4.9)$$

for any $\epsilon > 0$. Condition (4.9) entails that our common trends can be detected, and their number m estimated, as long as either N is not “too large” relatively to T , or that sufficiently many moments exist. For example, when $\eta = 1$, detection is possible only when $N = o(T)$, and as η decreases, the noise introduced by the cross-sectional dimension is more and more likely to drown out the signal associated with the common trends. We note that, when $\eta = 2$, that is, when the variance exists, (4.9) boils down, essentially, to requiring $T \rightarrow \infty$. That is, no restrictions on the relative rates of divergence between N and T are required as they pass to infinity.

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

$$\check{v}_{N,T}^{(j)} = T^{-\kappa} \frac{\lambda^{(j)}(S_{11})}{\sum_{k=1}^N \lambda^{(k)}(S_{00})}, \quad (4.10)$$

where $\kappa > 0$ is user-defined (and arbitrarily small), and use $\check{\phi}_{N,T}^{(j)} = \exp(\check{v}_{N,T}^{(j)}) - 1$ to carry out the test. The rationale for $\check{\phi}_{N,T}^{(j)}$ is similar to that of $\phi_T^{(j)}$ defined in (3.2), and it is based on exploiting the eigen-gap stipulated by **Proposition 4**. Indeed, under (4.9) and under the null that $m \geq j$, $\lambda^{(j)}(S_{11})$ diverges to infinity at a rate (roughly) proportional to $TN^{1-2/\eta}$. Conversely, under the alternative that $m < j$, the $\lambda^{(j)}(S_{11})$ and $\sum_{i=1}^N \sum_{t=1}^T \Delta y_{i,t}^2$ (roughly) have the same rate, and the effect of $T^{-\kappa}$ in (4.10) is to make $\check{v}_{N,T}^{(j)}$ drift to zero. Thus, $\check{\phi}_{N,T}^{(j)}$ has, heuristically, the same rates as $\phi_T^{(j)}$ defined in (3.2), and can be used in the same way.

Algorithm 2.

Step 1. Run the test for $H_0 : m \geq 1$ based on the randomized version of $\check{\phi}_{N,T}^{(1)}$. If the null is rejected, set $\check{m} = 0$ and stop, otherwise go to the next step.

Table 1. Estimation frequencies— $N = 3$.

		$N = 3$								
		$T = 100$				$T = 200$				
		m	3	2	1	0	3	2	1	0
\hat{m}	3									
	2	0.963	0.004	0.000	0.000	0.986	0.001	0.000	0.000	
	1	0.037	0.990	0.011	0.000	0.013	0.994	0.002	0.000	
	0	0.000	0.006	0.989	0.023	0.001	0.003	0.998	0.005	
$\eta = 0.5$	0	0.000	0.000	0.000	0.977	0.000	0.002	0.000	0.995	
	3	0.986	0.001	0.000	0.000	0.995	0.000	0.000	0.000	
	2	0.014	0.995	0.001	0.000	0.004	0.997	0.000	0.000	
	1	0.000	0.004	0.999	0.004	0.001	0.002	1.000	0.003	
$\eta = 1.0$	0	0.000	0.000	0.000	0.996	0.000	0.001	0.000	0.997	
	3	0.991	0.000	0.000	0.000	0.996	0.000	0.000	0.000	
	2	0.009	1.000	0.001	0.000	0.003	0.999	0.000	0.000	
	1	0.000	0.000	0.999	0.001	0.001	0.001	1.000	0.002	
$\eta = 1.5$	0	0.000	0.000	0.000	0.999	0.000	0.000	0.000	0.998	
	3	0.994	0.000	0.000	0.000	0.998	0.000	0.000	0.000	
	2	0.006	1.000	0.001	0.000	0.001	0.999	0.000	0.000	
	1	0.000	0.000	0.999	0.000	0.001	0.001	1.000	0.000	
$\eta = 2$	0	0.000	0.000	0.000	1.000	0.000	0.000	0.000	1.000	

Table 2. Estimation frequencies— $N = 4$.

		$N = 4$									
		$T = 100$					$T = 200$				
m		4	3	2	1	0	4	3	2	1	0
	\hat{m}										
$\eta = 0.5$	4	0.890	0.004	0.000	0.000	0.000	0.976	0.001	0.000	0.000	0.000
	3	0.104	0.964	0.005	0.000	0.000	0.023	0.988	0.002	0.000	0.000
	2	0.003	0.029	0.989	0.015	0.000	0.000	0.011	0.996	0.008	0.000
	1	0.002	0.000	0.006	0.984	0.022	0.000	0.000	0.001	0.992	0.012
	0	0.001	0.003	0.000	0.001	0.978	0.001	0.000	0.001	0.000	0.988
$\eta = 1.0$	4	0.948	0.000	0.000	0.000	0.000	0.995	0.000	0.000	0.000	0.000
	3	0.046	0.990	0.000	0.000	0.000	0.004	0.997	0.001	0.000	0.000
	2	0.003	0.007	0.995	0.003	0.000	0.000	0.003	0.998	0.001	0.000
	1	0.002	0.001	0.004	0.994	0.009	0.000	0.000	0.000	0.999	0.000
	0	0.001	0.002	0.001	0.003	0.991	0.001	0.000	0.001	0.000	1.000
$\eta = 1.5$	4	0.967	0.001	0.000	0.000	0.000	0.998	0.000	0.000	0.000	0.000
	3	0.029	0.990	0.000	0.000	0.000	0.001	0.999	0.001	0.000	0.000
	2	0.001	0.005	0.997	0.000	0.000	0.000	0.000	0.998	0.000	0.000
	1	0.001	0.002	0.003	0.998	0.001	0.000	0.000	0.000	1.000	0.001
	0	0.002	0.002	0.000	0.002	0.999	0.001	0.001	0.001	0.000	0.999
$\eta = 2$	4	0.974	0.000	0.000	0.000	0.000	0.998	0.000	0.000	0.000	0.000
	3	0.022	0.995	0.000	0.000	0.000	0.001	0.999	0.000	0.000	0.000
	2	0.001	0.002	0.999	0.000	0.000	0.000	0.000	0.999	0.000	0.000
	1	0.001	0.001	0.000	0.998	0.001	0.000	0.000	0.000	1.000	0.001
	0	0.002	0.002	0.001	0.002	0.999	0.001	0.001	0.001	0.000	0.999

Table 3. Estimation frequencies— $N = 5$.

		$N = 5$											
		$T = 100$					$T = 200$						
m		5	4	3	2	1	0	5	4	3	2	1	0
	\hat{m}												
$\eta = 0.5$	5	0.787	0.003	0.000	0.000	0.000	0.000	0.944	0.002	0.000	0.000	0.000	0.000
	4	0.201	0.919	0.005	0.000	0.000	0.000	0.055	0.985	0.004	0.000	0.000	0.000
	3	0.009	0.077	0.974	0.007	0.000	0.000	0.000	0.013	0.992	0.004	0.000	0.000
	2	0.000	0.000	0.020	0.993	0.001	0.001	0.000	0.000	0.003	0.996	0.014	0.000
	1	0.003	0.001	0.001	0.000	0.998	0.041	0.001	0.000	0.001	0.000	0.986	0.018
	0	0.000	0.000	0.000	0.000	0.001	0.958	0.000	0.000	0.000	0.000	0.000	0.982
$\eta = 1.0$	5	0.874	0.000	0.000	0.000	0.000	0.000	0.984	0.000	0.000	0.000	0.000	0.000
	4	0.125	0.959	0.002	0.000	0.000	0.000	0.014	0.997	0.000	0.000	0.000	0.000
	3	0.000	0.039	0.990	0.002	0.000	0.000	0.001	0.003	0.999	0.000	0.000	0.000
	2	0.000	0.001	0.007	0.995	0.004	0.000	0.001	0.000	0.000	1.000	0.001	0.000
	1	0.001	0.001	0.001	0.002	0.995	0.019	0.000	0.000	0.000	0.000	0.999	0.004
	0	0.000	0.000	0.000	0.001	0.001	0.981	0.000	0.000	0.001	0.000	0.000	0.996
$\eta = 1.5$	5	0.909	0.000	0.000	0.000	0.000	0.000	0.996	0.000	0.000	0.000	0.000	0.000
	4	0.090	0.974	0.000	0.000	0.000	0.000	0.002	0.998	0.000	0.000	0.000	0.000
	3	0.000	0.024	0.995	0.002	0.000	0.000	0.001	0.001	0.999	0.000	0.000	0.000
	2	0.000	0.000	0.005	0.995	0.001	0.000	0.001	0.001	0.000	1.000	0.000	0.000
	1	0.001	0.002	0.000	0.002	0.999	0.002	0.000	0.000	0.000	0.000	1.000	0.002
	0	0.000	0.000	0.000	0.001	0.000	0.998	0.000	0.000	0.001	0.000	0.000	0.998
$\eta = 2$	5	0.914	0.000	0.000	0.000	0.000	0.000	0.997	0.000	0.000	0.000	0.000	0.000
	4	0.085	0.981	0.000	0.000	0.000	0.000	0.002	0.998	0.000	0.000	0.000	0.000
	3	0.000	0.017	0.993	0.000	0.000	0.000	0.000	0.001	0.999	0.000	0.000	0.000
	2	0.000	0.001	0.005	0.999	0.001	0.000	0.000	0.001	0.000	1.000	0.000	0.000
	1	0.001	0.001	0.001	0.000	0.999	0.002	0.001	0.000	0.000	0.000	1.000	0.002
	0	0.000	0.000	0.001	0.001	0.000	0.998	0.000	0.000	0.001	0.000	0.000	0.998

Step 2. Starting from $j = 2$, run the test for $H_0 : m \geq j$ based on the randomized version of $\check{\phi}_{N,T}^{(1)}$, constructed using an artificial sample $\{\xi_i^{(j)}\}_{i=1}^M$ generated independently of $\{\xi_i^{(1)}\}_{i=1}^M, \dots, \{\xi_i^{(j-1)}\}_{i=1}^M$. If the null is rejected, set $\check{m} = j$ and stop; otherwise, if $j = m_{\max}$, set $\check{m} = m_{\max}$; otherwise, increase j and repeat Step 2.

Theorem 5. Let Assumptions 5–8 and (4.9) hold. As $\min(N, T, M) \rightarrow \infty$ under (3.5), it holds that $P^*(\check{m} = m) \rightarrow 1$ with probability tending to 1.

In Algorithm 2, κ and M need not be the same as in Algorithm 1. We finally note that improved finite sample properties can be obtained by modifying $\check{v}_{N,T}^{(j)}$ as follows (see also Barigozzi

Table 4. Estimation frequencies, large N and $m = 0$.

		N	20				50			
			200 m^*	400 m^*	800 m^*	1600 m^*	200 m^*	400 m^*	800 m^*	1600 m^*
$\eta = 1.9$	m									
	0		0.768	0.890	0.960	0.998	0.932	0.962	0.944	0.956
	1		0.232	0.110	0.040	0.002	0.068	0.038	0.056	0.044
	2		0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
$\eta = 1.5$	m									
	0		0.890	0.884	0.906	0.968	0.892	0.908	0.888	0.926
	1		0.110	0.116	0.094	0.032	0.108	0.092	0.112	0.074
	2		0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
$\eta = 1.1$	m									
	0		0.736	0.764	0.812	0.978	0.710	0.728	0.730	0.888
	1		0.264	0.236	0.188	0.022	0.290	0.272	0.270	0.112
	2		0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
		N	100				200			
		T	200 m^*	400 m^*	800 m^*	1600 m^*	200 m^*	400 m^*	800 m^*	1600 m^*
$\eta = 1.9$	m									
	0		0.968	0.958	0.966	0.958	0.940	0.972	0.968	0.958
	1		0.032	0.042	0.034	0.042	0.060	0.028	0.032	0.042
	2		0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
$\eta = 1.5$	m									
	0		0.894	0.876	0.888	0.892	0.886	0.900	0.876	0.890
	1		0.106	0.124	0.112	0.108	0.114	0.100	0.124	0.110
	2		0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
$\eta = 1.1$	m									
	0		0.716	0.734	0.736	0.876	0.710	0.760	0.744	0.756
	1		0.284	0.266	0.264	0.124	0.290	0.240	0.256	0.244
	2		0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
			0.000	0.000	0.000	0.000	0.000	0.000	0.000	

and Trapani 2022, for the finite variance case)

$$\tilde{v}_{N,T}^{(j)} = T^{-\kappa} \frac{\lambda^{(j)}(S_{11})}{\sum_{k=j+1}^N \lambda^{(k)}(S_{00})}. \quad (4.11)$$

5. Monte Carlo Evidence

In this section, we illustrate the finite sample properties of our procedure through a small scale Monte Carlo exercise. To save space, we report only a limited number of results; further results and details on the implementation of our statistics are in Section B in the supplementary materials.

5.1. The Fixed N Case

As in She and Ling (2020), we simulate the N -variate VAR(1) model

$$y_t = Ay_{t-1} + \varepsilon_t, \quad (5.1)$$

initialized at $y_0 = 0$. We parameterize A as $A = I_N - \Psi\Psi'$, Ψ being an $N \times (N - m)$ matrix with orthonormal columns (i.e., $\Psi'\Psi = I_{N-m}$).⁵ The innovations ε_t in (5.1) are iid and

coordinate-wise independent, from a power law distribution with tail index $\eta \in \{0.5, 1, 1.5, 2\}$. We follow the procedure proposed by Clauset, Shalizi, and Newman (2009) and generate $\varepsilon_{i,t}$ as

$$\varepsilon_{i,t} = (1 - v_{i,t})^{-1/\eta}, \quad (5.2)$$

where $v_{i,t}$ is iid. $U[0, 1]$; $\varepsilon_{i,t}$ is subsequently centered.⁶

First, we note from unreported experiments that our procedure for the determination of the number of common trends is not particularly sensitive to the choice of the various specifications. In our experiments, we have used $M = 100$ to speed up the computational time, but we note that results do not change when setting, for example, $M = T$, $M = T/2$ or $M = T/4$. In (3.2), we have used $\kappa = 10^{-4}$. This is a conservative choice, whose rationale follows from the fact that, in (3.2), dividing by T^κ serves the purpose of making the non-spiked eigenvalues drift to zero. The upper bound provided in (2.10) for such nonspiked eigenvalues is given by slowly varying functions, which suggests that even a very small value of κ

⁵We have created Ψ as $\Psi = D(D'D)^{-1/2}$, where $(D)^{-1/2}$ is the Choleski factor of D . We have set $D \sim \mathbf{1}_{N \times (N-m)} + d_{N \times (N-m)}$, where $\mathbf{1}_{N \times (N-m)}$ is

an $N \times (N - m)$ matrix of ones and $d_{N \times (N-m)}$ is an $N \times (N - m)$ matrix such that $\text{vec}(d_{N \times (N-m)}) \sim N(0, \mathbf{1}_{N \times (N-m)})$. We keep $d_{N \times (N-m)}$ fixed across Monte Carlo iterations.

⁶In unreported experiments, we considered $\varepsilon_t \sim \text{iid}.N(0, I_N)$; results are essentially the same as with $\eta = 2$.

Table 5. Estimation frequencies, large N and $m = 1$.

		N		20								50							
				200		400		800		1600		200		400		800		1600	
		T	m^*	ER	m^*	ER	m^*	ER	m^*	ER	m^*	ER	m^*	ER	m^*	ER	m^*	ER	
$\eta = 1.9$	m																		
	0	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000		
	1	1.000	0.988	1.000	0.992	1.000	1.000	1.000	0.992	1.000	0.984	0.994	0.984	0.998	0.992	0.998	0.996		
	2	0.000	0.008	0.000	0.008	0.000	0.000	0.000	0.008	0.000	0.016	0.006	0.016	0.002	0.008	0.002	0.004		
$\eta = 1.5$	m																		
	0	0.002	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.002	0.000	0.000	0.000	0.000	0.000	0.000	0.000		
	1	0.986	0.946	0.982	0.960	0.978	0.960	0.970	0.970	0.974	0.932	0.980	0.934	0.972	0.952	0.974	0.964		
	2	0.012	0.054	0.018	0.032	0.022	0.032	0.030	0.030	0.024	0.060	0.020	0.058	0.028	0.048	0.026	0.034		
$\eta = 1.1$	m																		
	0	0.002	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.018	0.000	0.008	0.000	0.004	0.000	0.002	0.000		
	1	0.948	0.822	0.950	0.826	0.940	0.856	0.990	0.880	0.916	0.750	0.916	0.788	0.902	0.818	0.860	0.828		
	2	0.005	0.124	0.050	0.138	0.060	0.106	0.010	0.102	0.066	0.178	0.076	0.146	0.094	0.150	0.138	0.144		
$\eta = 1.9$	m																		
	0	0.000	0.000	0.000	0.000	0.000	0.000	0.002	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000		
	1	0.992	0.980	0.998	0.976	0.996	0.992	0.998	0.994	0.999	0.968	1.000	0.988	0.998	0.994	0.998	0.992		
	2	0.008	0.020	0.002	0.024	0.004	0.008	0.000	0.006	0.001	0.032	0.000	0.012	0.002	0.006	0.002	0.008		
$\eta = 1.5$	m																		
	0	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.006	0.000	0.000	0.000	0.000	0.000	0.000	0.000		
	1	0.982	0.914	0.956	0.922	0.962	0.948	0.966	0.974	0.962	0.882	0.964	0.914	0.968	0.936	0.946	0.950		
	2	0.018	0.078	0.044	0.076	0.038	0.050	0.034	0.026	0.032	0.104	0.036	0.078	0.032	0.062	0.054	0.050		
$\eta = 1.1$	m																		
	0	0.036	0.000	0.008	0.000	0.002	0.000	0.004	0.000	0.060	0.000	0.028	0.000	0.010	0.000	0.006	0.000		
	1	0.882	0.702	0.896	0.734	0.888	0.796	0.920	0.800	0.878	0.662	0.904	0.736	0.868	0.736	0.894	0.766		
	2	0.082	0.210	0.096	0.206	0.110	0.152	0.076	0.162	0.062	0.218	0.068	0.178	0.122	0.202	0.100	0.176		
$\eta = 1.9$	m																		
	0	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000		
	1	0.992	0.980	0.998	0.976	0.996	0.992	0.998	0.994	0.999	0.968	1.000	0.988	0.998	0.994	0.998	0.992		
	2	0.008	0.020	0.002	0.024	0.004	0.008	0.000	0.006	0.001	0.032	0.000	0.012	0.002	0.006	0.002	0.008		

should suffice. Indeed, altering the value of κ has virtually no consequence. In order to compute the integral in (3.4), we use the Gauss-Hermite quadrature.⁷ Finally, as far as the family-wise detection procedure is concerned, the level of the individual tests is $\alpha(T) = 0.05/T$; this corresponds to having a critical value c_α which grows logarithmically with T . All routines are based on 1000 iterations and are written using GAUSS 21.

Results are reported in Tables 1–3, where we analyse the properties of our estimator of m with $N \in \{3, 4, 5\}$. The reported frequencies of the estimates of m show that the finite sample properties are largely satisfactory. Our procedure seems to be scarcely affected by the value of m , although, especially for the smaller sample sizes, it appears to be marginally better when $m = 0$ as opposed to the case $m = N$. This difference, however, vanishes as T increases. The impact of N is also very clear: as the VAR dimension increases, the performance of \hat{m} tends to deteriorate, as expected. Inference improves for larger values of T . Indeed, whilst results for $N = 3$ are good even when $\eta = 0.5$ and $T = 100$, when $N = 5$ the estimator \hat{m} requires at least $T = 200$ in order to have a frequency of correctly picking the true value of m higher than 90%. This is, as noted above, more pronounced when $m = N$, and less so when $m = 0$. As it

can also be expected, our procedure improves as η increases; results are anyway very good even in the (very extreme) case $\eta = 0.5$, and the impact of η is less and less important as T increases. Finally, although Tables 1–3 focus only on the iid case, unreported experiments showed that results are essentially the same when allowing for serial dependence.

In the supplementary materials, we report a broader set of results which, in addition to serial dependence in the errors $\varepsilon_{i,t}$, also compare the proposed method with classic information criteria. Results are in Tables B.13–B.18. Broadly speaking, our procedure is very good on average at estimating m —and better than the best performing information criterion, BIC—for all values of N and T (and η). This is true across all values of m , including the stationary case ($m = 0$) and the no cointegration case ($m = N$). Information criteria seem to perform marginally better when $m = 0$, but this is more than offset when considering that they tend to overestimate m in general, especially so when $m = N$ and $m = N - 1$. When errors are serially correlated (see Tables B.15–B.18), results are affected, albeit marginally, but the relative performance of the various methods remains as described above. In Tables B.19–B.21, we investigate the performance of our methodology to determine m in the case of larger values of N ; while not designed for the case $N \rightarrow \infty$, our sequential procedure is, in general, satisfactory, at least when

⁷Details are in Section B.1 of the supplementary materials.

Table 6. Estimation frequencies, large N and $m = 2$.

		N															
		20								50							
		200		400		800		1600		200		400		800		1600	
T	m^*	ER	m^*	ER	m^*	ER	m^*	ER	m^*	ER	m^*	ER	m^*	ER	m^*	ER	
$\eta = 1.9$	m																
	0	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
	1	0.064	0.066	0.038	0.022	0.004	0.030	0.008	0.010	0.056	0.054	0.018	0.036	0.001	0.014	0.004	0.008
	2	0.936	0.922	0.962	0.956	0.994	0.958	0.992	0.974	0.944	0.918	0.980	0.946	0.999	0.966	0.996	0.986
$\eta = 1.5$	m																
	0	0.002	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.002	0.000	0.000	0.000	0.000	0.000	0.000	0.000
	1	0.128	0.168	0.068	0.104	0.038	0.044	0.014	0.032	0.106	0.154	0.066	0.118	0.026	0.052	0.012	0.044
	2	0.866	0.766	0.932	0.830	0.958	0.902	0.978	0.922	0.892	0.762	0.930	0.818	0.962	0.884	0.976	0.902
$\eta = 1.1$	m																
	0	0.004	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.008	0.000	0.004	0.000	0.000	0.000	0.000	0.000
	1	0.290	0.362	0.188	0.290	0.142	0.238	0.074	0.220	0.332	0.412	0.188	0.330	0.160	0.312	0.082	0.238
	2	0.704	0.480	0.804	0.556	0.838	0.620	0.906	0.653	0.656	0.406	0.806	0.498	0.810	0.546	0.868	0.622
$\eta = 1.9$	m																
	0	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
	1	0.050	0.046	0.014	0.012	0.006	0.008	0.002	0.002	0.050	0.060	0.010	0.028	0.008	0.004	0.004	0.006
	2	0.950	0.922	0.986	0.950	0.994	0.972	0.998	0.984	0.948	0.890	0.988	0.948	0.990	0.980	0.994	0.984
$\eta = 1.5$	m																
	0	0.002	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
	1	0.090	0.194	0.064	0.106	0.022	0.078	0.016	0.048	0.120	0.226	0.066	0.130	0.024	0.060	0.016	0.046
	2	0.908	0.700	0.928	0.800	0.964	0.856	0.970	0.894	0.874	0.656	0.928	0.782	0.976	0.862	0.962	0.890
$\eta = 1.1$	m																
	0	0.004	0.000	0.002	0.000	0.000	0.000	0.000	0.000	0.014	0.000	0.004	0.000	0.002	0.000	0.000	0.000
	1	0.396	0.476	0.276	0.414	0.160	0.318	0.070	0.292	0.458	0.468	0.298	0.416	0.186	0.344	0.110	0.306
	2	0.596	0.318	0.716	0.390	0.818	0.522	0.870	0.542	0.528	0.320	0.690	0.386	0.788	0.452	0.822	0.496
$\eta = 1.1$	m																
	0	0.004	0.206	0.006	0.196	0.022	0.160	0.006	0.166	0.000	0.212	0.008	0.198	0.024	0.204	0.068	0.198
	1																
	2																

η is larger than 1 and when T is much larger than N . We have also conducted a small scale experiment in the case of $\eta = 2$ and of Gaussian errors, comparing the performance of our method with Johansen's sequential LR tests (Johansen 1991)—results, in Tables B.22–B.24, show that our method is virtually never outperformed by Johansen's procedure, even in this case. Further, we have compared our approach with using Johansen's sequential LR tests and the critical values in Caner (1998)—results in Tables B.25–B.27 again show that our methodology is, in general, not outperformed by this approach even when using the right critical values.

5.2. The Large N Case

We consider a set of experiments based on the large N setup discussed in Section 4.3. Data are generated according to (4.5), with F_t and $e_{i,t}$ generated independently of each other and according to (5.2), and λ_i generated as iid across i with $\lambda_i \sim N(0, 1)$; we consider $N \in \{20, 50, 100, 200\}$ and $T \in \{200, 400, 800, 1600\}$, and report results for $\eta \in \{1.1, 1.5, 1.9\}$ obtained using $\tilde{v}_{N,T}^{(j)}$ defined in (4.11) (the case $\eta < 1$ follows the same pattern as the reported results, but a larger T , as also predicted by the theory, is required). Results are in Tables 4–6, where we

report the frequencies of estimation of various values of m , also comparing our statistic with an alternative estimator based on the eigenvalue ratio approach when $m > 0$ (see, *inter alia*, Lam and Yao 2012; Ahn and Horenstein 2013; Zhang, Robinson, and Yao 2019). As can be seen, results are broadly in line with the theory: the performance of \tilde{m} deteriorates as—*ceteris paribus*— N increases, η decreases, and conversely improves as T increases.

6. Real Data Examples

We illustrate our methodology through two empirical applications to a small-to-medium scale VAR of $N = 7$ commodity prices (Section 6.1), and to a large ($N = 196$) factor model for the term structure of U.S. interest rate data (Section 6.2). Further empirical evidence for these applications and additional empirical studies are reported in Section C in the supplementary materials.

6.1. Comovements among Commodity Prices

We consider a set of $N = 7$ commodity prices: three oil prices (WTI, Brent crude, and Dubai crude) and the prices of four met-

Table 7. Estimated number of common trends; whole dataset.

Commodity	Tail index	Results and sensitivity analysis						
		Test $H_0 : E X ^2 = \infty$			Nominal level	1%	5%	10%
Copper	1.658 (1.144,2.171)	0.9527 (do not reject H_0)			Johansen's trace test	1	2	2
Gold	1.580 (1.090,2.069)	0.9525 (do not reject H_0)						
Brent crude	2.972 (2.050,3.893)	0.9504 (do not reject H_0)			Johansen's λ_{\max} test	0	0	1
Dubai crude	2.483 (1.171,3.252)	0.9499 (do not reject H_0)						
Nickel	2.063 (1.423,2.702)	0.9548 (do not reject H_0)						
WTI crude	2.532 (1.747,3.316)	0.9502 (do not reject H_0)						
Cobalt	1.691 (1.166,2.215)	0.9504 (do not reject H_0)						
$(\kappa = 10^{-4}, n_S = 2)$	$\frac{0.05}{T}$	$\frac{0.05}{\ln T}$	$\frac{0.05}{N}$	$(\kappa = 10^{-2}, n_S = 2)$	$\frac{0.05}{T}$	$\frac{0.05}{\ln T}$	$\frac{0.05}{N}$	
$T/2$	5	4	4	$T/2$	5	4	4	
T	4	4	4	T	4	4	4	
$2T$	4	4	4	$2T$	4	4	4	
$(\kappa = 10^{-4}, n_S = 4)$	$\frac{0.05}{T}$	$\frac{0.05}{\ln T}$	$\frac{0.05}{N}$	$(\kappa = 10^{-2}, n_S = 4)$	$\frac{0.05}{T}$	$\frac{0.05}{\ln T}$	$\frac{0.05}{N}$	
$T/2$	5	4	4	$T/2$	5	4	4	
T	5	4	4	T	5	4	4	
$2T$	4	4	4	$2T$	4	4	4	

In the top part of the table we report the estimated values of the tail index using the Hill's estimator—the package “ptsuite” in R has been employed, using a number of order statistics equal to $k_T = 40$. We also report, in light of Hill's estimator being inconclusive, the outcome of the strong version of the test by Trapani (2016), developed in Degiannakis et al. (2021); details are in Section C.2 in the supplementary materials.

In the table (top, right part), we also report the number of cointegration relationships found by Johansen's procedure; this has been implemented using $p = 2$ lags in the VAR specification, as suggested using BIC, and constant and restricted linear trend when implementing the test.

In the bottom half of the table, we report results on \hat{m} obtained using different specifications, as written in the table. In particular, in each sub-panel, the columns contain different values of the nominal level of the family-wise procedure, set equal to $\frac{0.05}{T}$, $\frac{0.05}{\ln T}$, and $\frac{0.05}{N}$.

Table 8. Estimated number of common trends; oil prices.

Nominal level	Results and sensitivity analysis						
	1%	5%	10%	Cointegration vectors			
Johansen's trace test	2	2	2	$\hat{\beta}_{1,1}$	1.00	$\hat{\beta}_{2,1}$	0.00
				$\hat{\beta}_{1,2}$	0.00	$\hat{\beta}_{2,2}$	1.00
Johansen's λ_{\max} test	2	2	2	$\hat{\beta}_{1,3}$	-1.08	$\hat{\beta}_{2,3}$	-1.05
$(\kappa = 10^{-4}, n_S = 2)$	$\frac{0.05}{T}$	$\frac{0.05}{\ln T}$	$\frac{0.05}{N}$	$(\kappa = 10^{-2}, n_S = 2)$	$\frac{0.05}{T}$	$\frac{0.05}{\ln T}$	$\frac{0.05}{N}$
$T/2$	2	2	1	$T/2$	2	1	1
T	2	1	1	T	1	1	1
$2T$	1	1	1	$2T$	1	1	1
$(\kappa = 10^{-4}, n_S = 4)$	$\frac{0.05}{T}$	$\frac{0.05}{\ln T}$	$\frac{0.05}{N}$	$(\kappa = 10^{-2}, n_S = 4)$	$\frac{0.05}{T}$	$\frac{0.05}{\ln T}$	$\frac{0.05}{N}$
$T/2$	2	2	1	$T/2$	2	1	1
T	2	1	1	T	1	1	1
$2T$	1	1	1	$2T$	1	1	1

See Table 7 for details; the three series considered are the three crude prices: WTI, Brent, and Dubai.

als (copper, gold, nickel, and cobalt). The presence of common trends can be anticipated due to global demand factors (e.g., growth in emerging Asian countries and especially in China; or changes of preferences toward greener energy sources, which increase demand for copper and decrease demand for oil), and also due to global supply factors (e.g., related to the effect that oil prices have on transportation costs of other commodities; or driven by technological innovations which often require the use of cobalt – see, e.g., Alquist, Bhattacharai, and Coibion 2020). Moreover, the three oil prices should exhibit strong comovements, and similarly should the prices of metals, which are often used in combination in industry (e.g., copper and nickel). In order to study the presence of such common trends, we use a dataset consisting of monthly data from January 1990 to March 2021,

corresponding to a sample of $T = 373$ monthly observations.⁸ We use the logs of the data, which are subsequently demeaned and detrended. We have applied our methodology using the same specifications as described in Section 5, that is, $\kappa = 10^{-4}$, $M = 100$ and $n_S = 2$ in (B.1). In order to assess robustness to these specifications, we have also considered other values of M (including $M = T$) and $n_S = 4$.

We report the results in Table 7. Initially, we report the (Hill's) estimates of the tail indices for the seven series; the associated confidence sets are quite large, but the test by Trapani (2016) supports the hypothesis that all series have infinite variance.

⁸Data have been downloaded from <https://www.imf.org/en/Research/commodity-prices>.

Table 9. Estimated number of common trends; metal prices.

Nominal level	Results and sensitivity analysis				Cointegration vectors		
	1%	5%	10%				
Johansen's trace test	0	1	1		$\hat{\beta}_{1,1}$	1.00	
					$\hat{\beta}_{1,2}$	-0.69	
Johansen's λ_{\max} test	0	1	1		$\hat{\beta}_{1,3}$	-0.50	
					$\hat{\beta}_{1,4}$	-0.13	
$(\kappa = 10^{-4}, n_S = 2)$	$\frac{0.05}{T}$	$\frac{0.05}{\ln T}$	$\frac{0.05}{N}$	$(\kappa = 10^{-2}, n_S = 2)$	$\frac{0.05}{T}$	$\frac{0.05}{\ln T}$	$\frac{0.05}{N}$
$T/2$	1	3	3	$T/2$	3	3	3
T	3	3	3	T	3	3	3
$2T$	3	3	3	$2T$	3	3	3
$(\kappa = 10^{-4}, n_S = 4)$	$\frac{0.05}{T}$	$\frac{0.05}{\ln T}$	$\frac{0.05}{N}$	$(\kappa = 10^{-2}, n_S = 4)$	$\frac{0.05}{T}$	$\frac{0.05}{\ln T}$	$\frac{0.05}{N}$
$T/2$	3	3	3	$T/2$	3	3	3
T	3	3	3	T	3	3	3
$2T$	3	3	3	$2T$	3	3	3

See Table 7 for details; the four series considered are the four metals: copper, nickel, gold, and cobalt.

Table 10. Commodity prices—Estimated loadings \hat{P} .

	$\hat{x}_{1,t}$	$\hat{x}_{2,t}$	$\hat{x}_{3,t}$	$\hat{x}_{4,t}$
WTI	1	0	0	0
Cobalt	0	1	0	0
Gold	0	0	1	0
Copper	0	0	0	1
Brent Crude	1.0409	-0.0252	0.1232	-0.0292
Dubai crude	1.0429	-0.0164	0.1515	-0.0679
Nickel	-0.2133	-0.3436	-1.8855	2.5744

Estimation of m based on Johansen's sequential procedure for the determination of the cointegration rank (using either the trace tests or the maximum eigenvalue tests) provides ambiguous results, with the estimate of m ranging between 5 and 7 (which corresponds to no cointegration). In contrast, through our test we find strong evidence of $m = 4$ common stochastic trends. As shown in the table, our results are broadly robust to different values of M and κ . In (much) fewer cases, we find $m = 5$, which might suggest the presence of a slowly mean reverting component in the data.

In order to shed more light on these findings, we split the series into two sub-groups: one of dimension $N = 3$ (comprising the three crude prices—Brent, Dubai and WTI crude), and one of dimension $N = 4$ (containing the four metal prices). Results for the 3-dimensional series of crude prices are in Table 8. On the one hand, Johansen's tests in this case identifies (at 5% level, and only using the trace test) two common trends ($m = 2$). On the other hand, our methodology provides evidence of a single ($m = 1$) common stochastic trend (and, in some, more rare, cases, of $m = 2$). Results concerning the $N = 4$ metals are in Table 9; in this case, evidence of $m = 3$ common trends emerges from all the procedures considered.

Overall, most of the evidence points toward $m = 4$ common stochastic trends, with much less evidence in support of $m = 5$. We report an estimate of the $m = 4$ common trends, using the results in Section 3.3. In order to identify the trends, based on the results above we propose to order the series as follows: WTI, gold, cobalt, copper, Brent crude, Dubai crude, nickel. Then, we constrain the upper $m \times m$ block of the estimated loadings matrix \hat{P} to be the identity matrix. We report the loadings \hat{P} in

Table 10 (see also Figure C.1 in the supplementary materials, where we plot the estimated common trends \hat{x}_t).

By construction, the first and second trends ($\hat{x}_{1,t}$ and $\hat{x}_{2,t}$) are associated with oil prices and cobalt, respectively. The third one ($\hat{x}_{3,t}$) is associated with gold (by construction), and nickel, with a negative loading; finally, the fourth trend ($\hat{x}_{4,t}$) is associated with copper by construction, and with nickel (with a positive loading). The trends driving metals are also common to oil prices, albeit with smaller loadings.

6.2. The Term Structure of U.S. Interest Rates

Following She and Ling (2020), we evaluate the presence and number of common stochastic trends in the yield curve.⁹ We use monthly data with maturities from 6 months up to 100 years ($N = 196$), spanning the period from January 1985 to September 2018 ($T = 405$). We use the logs of the original data, which are subsequently demeaned. By way of preliminary analysis (reported in Section C.2 in the Supplement), we have applied the test by Trapani (2016) to all series, which lends strong support to the hypothesis of infinite variance, similarly to what found also in She and Ling (2020). We estimate m using Algorithm 2 proposed in Section 4.3; as in the previous section, we assess the robustness of our procedure by computing the test statistics with different specifications, and we also report, for completeness, the estimate of m using the information criterion $IC3$ proposed in Bai (2004) and the $BT1$ and $BT2$ tests proposed by Barigozzi and Trapani (2022). Finally, we also consider using $\sum_{k=j}^N \lambda^{(k)}(S_{00})$ as a rescaling factor in (4.10); this rescaling dampens the eigenvalues of S_{11} more, thus, being bound to result in fewer common stochastic trends detected.

The results in Table 11 can be read from the top-left to the bottom-right of the table as being derived with increasingly "liberal" testing set-ups (i.e., tests become more and more likely to reject the null and consequently find fewer common stochastic

⁹We consider data from the High Quality Market (HQM) Corporate Bond Yield Curve, available from the Federal Reserve Economic Data (FRED)—details on the construction of the yield curves are available from the US Department of Treasury. She and Ling (2020) consider a similar application, based on a VAR with $N = 3$; for completeness, we have also carried out the same exercise, reported in Section C.3 in the supplementary materials.

Table 11. Estimated number of common trends; U.S. interest rates.

Results and sensitivity analysis								
IC	Other estimators							
	5	BT1 3			BT2 3			
Number of common trends estimated using Algorithm 2 with (4.10)								
$(\kappa = 10^{-4}, n_S = 2)$	$\frac{0.05}{T}$	$\frac{0.05}{\ln T}$	$\frac{0.05}{N}$	$(\kappa = 10^{-2}, n_S = 2)$	$\frac{0.05}{T}$	$\frac{0.05}{\ln T}$	$\frac{0.05}{N}$	
$T/2$	5	5	5	$T/2$	5	5	5	
T	5	5	5	T	5	5	5	
$2T$	5	5	5	$2T$	5	5	5	
$(\kappa = 10^{-4}, n_S = 4)$	$\frac{0.05}{T}$	$\frac{0.05}{\ln T}$	$\frac{0.05}{N}$	$(\kappa = 10^{-2}, n_S = 4)$	$\frac{0.05}{T}$	$\frac{0.05}{\ln T}$	$\frac{0.05}{N}$	
$T/2$	5	5	5	$T/2$	5	5	5	
T	5	5	5	T	5	5	5	
$2T$	5	5	5	$2T$	5	5	5	
Number of common trends estimated using Algorithm 2 with $\sum_{k=j}^N \lambda^{(k)}(S_{00})$ in (4.10)								
$(\kappa = 10^{-4}, n_S = 2)$	$\frac{0.05}{T}$	$\frac{0.05}{\ln T}$	$\frac{0.05}{N}$	$(\kappa = 10^{-2}, n_S = 2)$	$\frac{0.05}{T}$	$\frac{0.05}{\ln T}$	$\frac{0.05}{N}$	
$T/2$	5	5	5	$T/2$	5	5	5	
T	5	5	5	T	4	4	4	
$2T$	4	4	4	$2T$	3	3	3	
$(\kappa = 10^{-4}, n_S = 4)$	$\frac{0.05}{T}$	$\frac{0.05}{\ln T}$	$\frac{0.05}{N}$	$(\kappa = 10^{-2}, n_S = 4)$	$\frac{0.05}{T}$	$\frac{0.05}{\ln T}$	$\frac{0.05}{N}$	
$T/2$	5	5	5	$T/2$	5	5	5	
T	5	5	5	T	4	4	4	
$2T$	4	4	4	$2T$	3	3	3	

We report results on \hat{m} obtained using Algorithm 2 under different specifications—each sub-panel, the columns contain different values of the nominal level of the family-wise procedure, set equal to $\frac{0.05}{T}$, $\frac{0.05}{\ln T}$, and $\frac{0.05}{N}$. See also Table 7 for further details.

In the bottom of the table, we report the number of estimated common trends using the tests developed by Barigozzi and Trapani (2022)—referred to as BT1 and BT2, and we refer to that paper for details. We also report the information IC3 proposed in Bai (2004).

trends); the evidence reported in the table suggests that the number of common stochastic trends m is larger than 3. This also confirms that the three common factors spanning the yield curve, namely the level, slope, and curvature (we refer, e.g., to Diebold and Li 2006, and the references therein), are nonstationary. Importantly, and in contrast to the extant literature, our result does not rely on the assumption of finite variance, which is rejected on our data. In addition to the first three factors, we also find substantial evidence of two further common nonstationary factors. These are found also by the criteria proposed by Bai (2004); interestingly, Barigozzi and Trapani (2022) also found evidence of more common factors, although these were found to be “nearly stationary” and thus not found by the BT1 and BT2 tests therein. The plots of the five estimated common trends are reported in Figures C.2–C.6 in the supplementary materials.

7. Conclusions

In this article, we propose a methodology for inference on the common trends in multivariate time series with heavy tailed, heterogeneous innovations. We develop: (i) tests for hypotheses on the number of common trends; (ii) a sequential procedure to consistently estimate the number of common trends; (iii) an estimator of the common trends and of the associated loadings. A key feature of our approach is that estimation of the tail index of the innovations is not needed, and no prior knowledge as to whether the data have finite variance or not is required. Indeed, the procedure can be applied even in the case of finite second moments.

Our method is based on the eigenvalues of the sample second moment matrix of the data in levels, the largest m (m being

the unknown number of common trends) of which are shown to diverge at a higher rate, as T increases, than the remaining ones. Based on such rates, we propose a randomized statistic for testing hypotheses on m ; its limiting distribution is Chi-squared under the null, and diverges under the alternative. Combining these individual tests into a sequence of tests, we prove consistency of the estimator of m by simply letting the nominal level to shrink to zero at a proper rate. We also show that, once m is determined, estimation of the common trends and their loadings can be done using PCA. Our simulations show that our method has good properties, even in samples of small and moderate size. Whilst the main focus of our analysis is a simple case with no deterministic, iid observations and fixed N , we study several extensions, developing also a method to estimate the number of common factors in a large, nonstationary factor model with heavy tails. Further extensions are reported in the accompanying supplementary materials.

Supplementary Materials

Further theoretical results, simulations and empirical results - as well as all technical lemmas and proofs - are in the Online Supplement.

Acknowledgments

We thank the coeditor, the associate editor and three anonymous referees for the many constructive comments and suggestions on earlier versions of the article. We are grateful to Mehmet Caner and Shiqing Ling for helpful feedback. We are also grateful to the participants to the seminar series of the Department of Economics at Lund University (June 9, 2019); to the 23rd Dynamic Econometrics Conference (virtual; March 18–19, 2021); to the New York Camp Econometrics XV (virtual; April 10–11, 2021); and to the seminar series of the SIdE—Italian Econometric Association (virtual; June 7, 2022).

Funding

Matteo Barigozzi and Giuseppe Cavaliere gratefully acknowledge financial support from MIUR (PRIN 2017, grant 2017TA7TYC).

ORCID

Matteo Barigozzi  <http://orcid.org/0000-0001-9090-5708>

Giuseppe Cavaliere  <http://orcid.org/0000-0002-2856-0005>

Lorenzo Trapani  <http://orcid.org/0000-0003-3425-2786>

References

- Ahn, S. C., and Horenstein, A. R. (2013), “Eigenvalue Ratio Test for the Number of Factors,” *Econometrica*, 81, 1203–1227. [12]
- Ahn, S. K., and Reinsel, G. C. (1990), “Estimation for Partially Nonstationary Multivariate Autoregressive Models,” *Journal of the American Statistical Association*, 85, 813–823. [3]
- Alquist, R., Bhattarai, S., and Coibion, O. (2020), “Commodity-Price Comovement and Global Economic Activity,” *Journal of Monetary Economics*, 112, 41–56. [13]
- Bai, J. (2004), “Estimating Cross-Section Common Stochastic Trends in Nonstationary Panel Data,” *Journal of Econometrics*, 122, 137–183. [2,3,7,8,14,15]
- Bandi, F., and Corradi, V. (2014), “Nonparametric Nonstationarity Tests,” *Econometric Theory*, 30, 127–149. [2,4]
- Barigozzi, M., and Trapani, L. (2022), “Testing for Common Trends in Nonstationary Large Datasets,” *Journal of Business and Economic Statistics* (forthcoming). [10,14,15]
- Bernstein, D. H., and Nielsen, B. (2019), “Asymptotic Theory for Cointegration Analysis When the Cointegration Rank is Deficient,” *Econometrics*, 7, 6. [5]
- Bykhovskaya, A., and Gorin, V. (2022), “Cointegration in Large VARs,” *Annals of Statistics* (forthcoming). [2,8]
- Caner, M. (1998), “Tests for Cointegration with Infinite Variance Errors,” *Journal of Econometrics*, 86, 155–175. [1,3,12]
- Cavaliere, G., Georgiev, I., and Taylor, A. R. (2018), “Unit Root Inference for Non-stationary Linear Processes Driven by Infinite Variance Innovations,” *Econometric Theory*, 34, 302–348. [1]
- Cavaliere, G., Rahbek, A., and Taylor, A. R. (2012), “Bootstrap Determination of the Co-integration Rank in Vector Autoregressive Models,” *Econometrica*, 80, 1721–1740. [6]
- Cavaliere, G., and Taylor, A. R. (2009), “Heteroskedastic Time Series with a Unit Root,” *Econometric Theory*, 25, 1228–1276. [7]
- Chan, N. H., and Zhang, R. (2012), “Non-Stationary Autoregressive Processes with Infinite Variance,” *Journal of Time Series Analysis*, 33, 916–934. [1]
- Chen, L., Dolado, J. J., and Gonzalo, J. (2021), “Quantile Factor Models,” *Econometrica*, 89, 875–910. [2]
- Clauset, A., Shalizi, C. R., and Newman, M. E. (2009), “Power-Law Distributions in Empirical Data,” *SIAM Review*, 51, 661–703. [10]
- Davis, R., and Resnick, S. (1985), “Limit Theory for Moving Averages of Random Variables with Regularly Varying Tail Probabilities,” *The Annals of Probability*, 13, 179–195. [1]
- (1986), “Limit Theory for the Sample Covariance and Correlation Functions of Moving Averages,” *The Annals of Statistics*, 14, 533–558. [1]
- Davis, R. A. (2010), “Heavy tails in financial time series,” in *Encyclopedia of Quantitative Finance*, eds. R. Cont, New York: Wiley. [1]
- Davis, R. A., Pfaffel, O., and Stelzer, R. (2014), “Limit Theory for the Largest Eigenvalues of Sample Covariance Matrices with Heavy-Tails,” *Stochastic Processes and their Applications*, 124, 18–50. [2]
- Degiannakis, S., Filis, G., Siourounis, G., and Trapani, L. (2021), “Superkurtosis,” *Journal of Money, Credit and Banking* (forthcoming). [13]
- Diebold, F. X., and Li, C. (2006), “Forecasting the Term Structure of Government Bond Yields,” *Journal of Econometrics*, 130, 337–364. [15]
- Embrechts, P., Klüppelberg, C., and Mikosch, T. (2013), *Modelling Extremal Events: For Insurance and Finance* (Vol. 33), Berlin: Springer. [1]
- Engle, R., and Granger, C. W. (1987), “Co-Integration and Error Correction: Representation, Estimation, and Testing,” *Econometrica*, 55, 251–276. [1]
- Falk, B., and Wang, C.-H. (2003), “Testing Long-Run PPP with Infinite-Variance Returns,” *Journal of Applied Econometrics*, 18, 471–484. [1]
- Fan, J., Liu, H., and Wang, W. (2018), “Large Covariance Estimation through Elliptical Factor Models,” *Annals of Statistics*, 46, 1383–1414. [2]
- Fasen, V. (2013), “Time Series Regression on Integrated Continuous-Time Processes with Heavy and Light Tails,” *Econometric Theory*, 29, 28–67. [1]
- Gabaix, X. (1999), “Zipf’s Law for Cities: An Explanation,” *The Quarterly Journal of Economics*, 114, 739–767. [1]
- Gonzalo, J., and Granger, C. (1995), “Estimation of Common Long-Memory Components in Cointegrated Systems,” *Journal of Business & Economic Statistics*, 13, 27–35. [7]
- Hall, P., Peng, L., and Yao, Q. (2002), “Prediction and Nonparametric Estimation for Time Series with Heavy Tails,” *Journal of Time Series Analysis*, 23, 313–331. [1]
- Hallin, M., Van den Akker, R., and Werker, B. J. (2011), “A Class of Simple Semiparametrically Efficient Rank-based Unit Root Tests,” *Journal of Econometrics*, 163, 200–214. [1]
- Hallin, M., van den Akker, R., and Werker, B. J. (2016), “Semiparametric Error-Correction Models for Cointegration with Trends: Pseudo-Gaussian and Optimal Rank-based Tests of the Cointegration Rank,” *Journal of Econometrics*, 190, 46–61. [1]
- He, Y., Kong, X., Yu, L., and Zhang, X. (2022), “Large-Dimensional Factor Analysis without Moment Constraints,” *Journal of Business & Economic Statistics*, 40, 302–312. [2]
- Horváth, L., and Trapani, L. (2019), “Testing for Randomness in a Random Coefficient Autoregression Model,” *Journal of Econometrics*, 209, 338–352. [5]
- Ibragimov, M., and Ibragimov, R. (2018), “Heavy Tails and Upper-Tail Inequality: The Case of Russia,” *Empirical Economics*, 54, 823–837. [1]
- Jach, A., and Kokoszka, P. (2004), “Subsampling Unit Root Tests for Heavy-Tailed Observations,” *Methodology and Computing in Applied Probability*, 6, 73–97. [1]
- Johansen, S. (1991), “Estimation and Hypothesis Testing of Cointegration Vectors in Gaussian Vector Autoregressive Models,” *Econometrica*, 59, 1551–1580. [1,2,5,6,12]
- Kasa, K. (1992), “Common Stochastic Trends in International Stock Markets,” *Journal of Monetary Economics*, 29, 95–124. [1]
- 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. [12]
- Liang, C., and Schienle, M. (2019), “Determination of Vector Error Correction Models in High Dimensions,” *Journal of Econometrics*, 208, 418–441. [2]
- McElroy, T., and Politis, D. N. (2003), “Limit Theorems for Heavy-Tailed Random Fields with Subsampling Applications,” *Mathematical Methods of Statistics*, 12, 305–328. [8]
- Nielsen, M. Ø. (2010), “Nonparametric Cointegration Analysis of Fractional Systems with Unknown Integration Orders,” *Journal of Econometrics*, 155, 170–187. [4]
- Nyblom, J., and Harvey, A. (2000), “Tests of Common Stochastic Trends,” *Econometric Theory*, 16, 176–199. [4]
- Onatski, A., and Wang, C. (2018), “Alternative Asymptotics for Cointegration Tests in Large VARs,” *Econometrica*, 86, 1465–1478. [2,8]
- (2021), “Spurious Factor Analysis,” *Econometrica*, 89, 591–614. [2]
- Patilea, V., and Raïssi, H. (2014), “Testing Second-Order Dynamics for Autoregressive Processes in Presence of Time-Varying Variance,” *Journal of the American Statistical Association*, 109, 1099–1111. [7]
- Paulauskas, V., and Rachev, S. T. (1998), “Cointegrated Processes with Infinite Variance Innovations,” *Annals of Applied Probability*, 8, 775–792. [1]
- Peña, D., and Poncela, P. (2006), “Nonstationary Dynamic Factor Analysis,” *Journal of Statistical Planning and Inference*, 136, 1237–1257. [6]
- Petrov, V. (1974), “Moments of Distributions Attracted to Stable Laws,” *Theory of Probability & Its Applications*, 18, 569–571. [3]
- Pham, T. D., and Tran, L. T. (1985), “Some Mixing Properties of Time Series Models,” *Stochastic Processes and their Applications*, 19, 297–303. [3]

- Qu, Z., and Perron, P. (2007), "A Modified Information Criterion for Cointegration Tests based on a VAR Approximation," *Econometric Theory*, 23, 638–685. [1]
- Samorodnitsky, G., and Taqqu, M. S. (1994), *Stable Non-Gaussian Random Processes*, London: Chapman & Hall. [3]
- Seneta, E. (2006), *Regularly varying functions* (Vol. 508), Berlin: Springer. [3]
- Shao, X. (2015), "Self-normalization for Time Series: A Review of Recent Developments," *Journal of the American Statistical Association*, 110, 1797–1817. [4]
- She, R., and Ling, S. (2020), "Inference in Heavy-Tailed Vector Error Correction Models," *Journal of Econometrics*, 214, 433–450. [1,3,6,7,10,14]
- Stock, J. H., and Watson, M. W. (1988), "Testing for Common Trends," *Journal of the American Statistical Association*, 83, 1097–1107. [1,3]
- Trapani, L. (2014), "Chover-Type Laws of the k-iterated Logarithm for Weighted Sums of Strongly Mixing Sequences," *Journal of Mathematical Analysis and Applications*, 420, 908–916. [4]
- (2016), "Testing for (In)finite Moments," *Journal of Econometrics*, 191, 57–68. [13,14]
- Tu, Y., Yao, Q., and Zhang, R. (2020), "Error-Correction Factor Models for High-Dimensional Cointegrated Time Series," *Statistica Sinica*, 30, 1463–1484. [2]
- Watson, M. W. (1994), "Vector Autoregressions and Cointegration," *Handbook of Econometrics*, 4, 2843–2915. [2]
- Yap, S. F., and Reinsel, G. C. (1995), "Estimation and Testing for Unit Roots in a Partially Nonstationary Vector Autoregressive Moving Average Model," *Journal of the American Statistical Association*, 90, 253–267. [7]
- Yu, L., He, Y., and Zhang, X. (2019), "Robust Factor Number Specification for Large-Dimensional Elliptical Factor Model," *Journal of Multivariate Analysis*, 174, 104543. [2]
- Zhang, R., Robinson, P., and Yao, Q. (2019), "Identifying Cointegration by Eigenanalysis," *Journal of the American Statistical Association*, 114, 916–927. [2,6,12]