Inference in heavy-tailed non-stationary multivariate time series*

Matteo Barigozzi* **Giuseppe Cavaliere†** **Lorenzo Trapani‡**

*University of Bologna
e-mail: matteo.barigozzi@unibo.it

†*University of Bologna and University of Exeter Business School*
e-mail: giuseppe.cavaliere@unibo.it

‡*University of Nottingham*
e-mail: lorenzo.trapani@nottingham.ac.uk

Abstract:

We study inference on the common stochastic trends in a non-stationary, 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 randomised 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 a 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 *i.i.d.* 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 paper by presenting two illustrative applications covering commodity prices and interest rates data.

Keywords and phrases: non-stationarity, heavy tails, randomized tests, factor models.

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 essen-

*We thank the co-editor, the associate editor and three anonymous referees for the many constructive comments and suggestions on earlier versions of the paper. We are also grateful to the participants to the seminar series of the Department of Economics at Lund University (9 June 2019); to the 23rd Dynamic Econometrics Conference (virtual; 18-19 March, 2021); to the New York Camp Econometrics XV (virtual; 10-11 April 2021); and to the seminar series of the SIdE - Italian Econometric Association (virtual; 7 June 2022). Matteo Barigozzi and Giuseppe Cavaliere gratefully acknowledge financial support from MIUR (PRIN 2017, Grant 2017TA7TYC).

tial step in the analysis of multivariate time series which are non-stationary 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 non-stationary 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 et al., 2013).

Violation of the moment assumptions may result in (possibly severe) incorrect determination of the number of common trends - see, e.g., 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 (non-standard) rate of convergence of estimators in non-stationary Vector AutoRegression (VAR) models and show that the limiting distributions depend on the tail index in a non-trivial fashion; similar results are also found in Paulauskas and Rachev (1998), Fasen (2013), and in Chan and Zhang (2012) in the contest of least squares estimation of non-stationary autoregressions driven by innovations with heavy tails (see also Davis and Resnick, 1985; Davis and Resnick, 1986; and Hall et al., 2002). In the univariate case, Jach and Kokoszka (2004) and Cavaliere et al. (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 et al. (2016) (see also Hallin et al., 2011) apply the rank transformation to the residuals of a Vector Error Correction Model (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, e.g., multiple shifts and smooth scale changes. As far as we are aware, this paper 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 et al., 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 randomise 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. Firstly, 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*. Secondly, 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 et al., 2019; and Tu et al., 2020). Whilst the main focus of our paper is on the fixed-dimensional case $N < \infty$, the high-dimensional case $N \to \infty$ 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 et al. (2018), Yu et al. (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 et al., 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 $N \to \infty$.

The remainder of the paper is organised 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 $A_{i,j}$; we use $\|A\|$ to denote its Frobenius norm, i.e., $\|A\| = (\sum_{i=1}^{n} \sum_{j=1}^{m} A_{i,j}^2)^{1/2}$; we also let $\lambda^{(j)}(A)$ denote the j-th largest eigenvalue of A. We denote with $c_0, c_1, ...$ 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, i.e., $\Delta z_t = z_t - z_{t-1}$. Given a scalar random variable X, we denote its L_p-norm as $|X|_p$, i.e., $|X|_p = (E|X|^p)^{1/p}$. We use $\ln_k x$ to denote the k-iterated logarithm of x truncated at zero - e.g., $\ln_2 x = \max\{\ln \ln x, 0\}$; $\lfloor x \rfloor$ 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 paper.

2. Theory

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

$$\Delta y_t = C(L)\varepsilon_t, \tag{2.1}$$

where $C(L) = \sum_{j=0}^{\infty} C_j L^j$, and ε_t is a sequence of *i.i.d.* 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, e.g., 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

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

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 behaviour of the N-dimensional vector y_t is driven by m non-stationary common factors.

Assumption 1. *It holds that: (i) 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_js 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, e.g., 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 behaviour 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 shorthand 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 non-stationary dynamic factors - see, e.g., Bai (2004), where the latent factors correspond to the set of m common stochastic trends F_t. However, with respect to models with non-stationary 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 Supplement.

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 i.i.d. sequence; (ii) for all nonzero vectors* $l \in \mathbb{R}^N$, $l'\varepsilon_t$ *has distribution* $F_{l\varepsilon}$ *with strictly positive density, which 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 Taqqu, 1994) with common tail index η. This also implies (by Property 1.2.6

in Samorodnitsky and Taqqu, 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 Taqqu (1994), it holds that

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

as $x \to \infty$, 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 Taqqu, 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

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

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*

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

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 $\epsilon > 0$.

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

Similarly to Proposition 1, Proposition 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]

[1]Another possible scaling would be based on $\widetilde{S}_{00}^{-1}S_{11}$, with $\widetilde{S}_{00}$ a diagonal matrix whose nonzero elements are the same as those of S_{00}. In Section B.3 in the Supplement, 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.

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 $T \geq T_0$,*

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

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

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

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 normalising 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 $\lambda^{(j)}\left(S_{00}^{-1}S_{11}\right)$: 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 of $\lambda^{(j)}\left(S_{00}^{-1}S_{11}\right)$. 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

$$\begin{cases} H_0 : m \geq j \\ H_A : m < j \end{cases} \tag{3.1}$$

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

Based on Theorem 1, we propose to use

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

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 $\lambda^{(j)}\left(S_{00}^{-1}S_{11}\right)$ under H_0 and under H_A. In particular, it ensures that $T^{-\kappa}\lambda^{(j)}\left(S_{00}^{-1}S_{11}\right)$ 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 $\lambda^{(j)}\left(S_{00}^{-1}S_{11}\right)$ to diverge at a rate close to T under H_0. On account of Theorem 1, it holds that $P(\omega : \lim_{T\to\infty}\phi_T^{(j)} = \infty) = 1$ for $0 \leq j \leq m$; hence, we can assume that under the null that $m \geq j$, it holds that $\lim_{T\to\infty}\phi_T^{(j)} = \infty$. Conversely, under the alternative that $j > m$, we have $P(\omega : \lim_{T\to\infty}\phi_T^{(j)} = 0) = 1$, so that $\lim_{T\to\infty}\phi_T^{(j)} = 0$. In essence, $\phi_T^{(j)}$ diverges to positive infinity, or converges (to zero), according to whether $\lambda^{(j)}\left(S_{00}^{-1}S_{11}\right)$ is "large" or "small".

Since the limiting law of $\phi_T^{(j)}$ under the null is unknown, we propose a randomised 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, e.g., Bandi and Corradi (2014), and the discussion therein. In Section B.4 in the Supplement, we also consider different randomisation schemes.

Step 1 Generate an artificial sample $\{\xi_i^{(j)}, 1 \leq i \leq M\}$, with $\xi_i^{(j)} \sim i.i.d.N(0, 1)$, 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

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

Step 3 Compute

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

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, e.g., 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 the 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* $\min(T, M) \to \infty$ *with*

$$M^{1/2} \exp\left(-T^{1-\kappa-\epsilon}\right) \to 0, \tag{3.5}$$

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

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

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

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

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

Theorem 2 provides the limiting behaviour of $\Theta_{T,M}^{(j)}$, also illustrating the impact of M on the size and power trade-off. 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 non-centrality parameter of order $O\left(M^{1/2}\exp\left(-T^{1-\kappa-\epsilon}\right)\right)$, whence the upper bound in (3.5). We discuss the choice of M in Section 5; here, we note that condition (3.5) is, e.g., 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, whilst 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 $P\left(\chi_1^2 > c_\alpha\right) = \alpha$, an immediate consequence of the theorem is that under H_A it holds that $\lim_{\min(T,M)\to\infty} P^*(\Theta_{T,M}^{(j)} > c_\alpha) = 1$ for almost all realisations of $\{\varepsilon_t, 0 < t < \infty\}$: the test is consistent under the alternative. Conversely, under H_0 we have, for almost all realisations of $\{\varepsilon_t, 0 < t < \infty\}$

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

Our test is constructed using a randomisation 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

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

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

$$\lim\inf_{S\to\infty} \lim_{\min(T,M)\to\infty} \sqrt{\frac{S}{2\ln\ln S}} \frac{Q_{\alpha,S} - (1-\alpha)}{\sqrt{\alpha\,(1-\alpha)}} = -1. \tag{3.10}$$

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

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

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 $S \to \infty$ 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 $j = 1, 2, ...$ 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 $\{\xi_i^{(1)}\}_{i=1}^M$, ..., $\{\xi_i^{(j-1)}\}_{i=1}^M$. 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* $\min(T, M) \to \infty$ *under (3.5), if* $c_\alpha(M) \to \infty$ *with* $c_\alpha = o(M)$, *then it holds that* $P^*(\widehat{m} = m) = 1$ *for almost all realisations 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), we show that, $P^*(\widehat{m} = m) \to (1-\alpha)^{N-m}$ a.s. conditionally on the sample, whence the requirement $c_\alpha \to \infty$, which entails $\alpha \to 0$.

The theorem can also be read in conjunction with Johansen's procedure (Johansen, 1991), and its bootstrap implementations (Cavaliere et al., 2012), whose outcome is an estimate of m, say $\widetilde{m}$, such that, asymptotically, $P(\widetilde{m} = m) \to 1-\alpha$ 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 $P^*(\widehat{m} = m) \to (1-\alpha)^{N-m}$ 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 $P^*(\widehat{m} = m) \to (1-\alpha/N)^{N-m}$ a.s. conditionally on the sample, with

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

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) - i.e., an estimator of m (say $\widehat{m}^*$) such that $P^*(\widehat{m}^* = m) \to 1-\alpha$ 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 hold, and that the individual tests are implemented using* $\{\xi_i, 1 \leq i \leq M\}$ *with* $\xi_i \sim i.i.d.N(0,1)$, *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 realisations 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 realisations*

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

of $\{\varepsilon_t, -\infty < t < \infty\}$.

In Section A.2 in the Supplement, 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 *non-stationary* factors; hence, the lack of power against non-stationary, 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 non-stationary common stochastic trends F_t by using Principal Components (PC), in a similar fashion to Peña and Poncela (2006) and Zhang et al. (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 $\widehat{\Lambda} = (\widehat{v}_1, ..., \widehat{v}_m)$, 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* $1 \leq t \leq T$

$$\left\| \widehat{\Lambda} - \Lambda H \right\| = O_P\left(T^{-1+\epsilon}\right), \tag{3.13}$$

$$\left\| \widehat{F}_t - H^{-1} F_t \right\| = O_P(1) + O_P\left(T^{-1+1/p}\right), \tag{3.14}$$

[3] We report a more in-depth explanation of this in Section A.1 in the Supplement.

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* $E|\varepsilon_{i,t}|^2 < \infty$*.*

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 $N \to \infty$. 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 $O_P\left(t^{1/p}\right)$, 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 $\max_{1\leq t\leq T}\left\|\widehat{F}_t - H^{-1}F_t\right\| = O_P\left(T^{1/p}\right)$. This arises from the fact that the maximum of a T-dimensional sequence with finite p-th moment is bounded by $O_P\left(T^{1/p}\right)$.

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 deterministics, 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

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

where υ_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 heteroskedasticity 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, i.e., $h(r) = \sum_{i=1}^{n} h_i I(c_{i-1} \leq r < c_i)$, 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 υ_t. Then, as $\min(T,M) \to \infty$ with (3.5), it holds that, for all j*

$$P^*(\Theta^{(j)}_{T,M} > c_\alpha) \to \alpha, \tag{4.2}$$

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

Repeating *verbatim* the proof of Theorem 3, the results in Corollary 1 entail that, using the Algorithm 1 in Section 3.2, $P^*(\widehat{m} = m) \to 1$ 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; and 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, 3 and 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 non-stationary factor representation (2.3)

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

where $\Lambda = (\lambda_1, ..., \lambda_N)'$ is an $N \times m$ matrix of loadings, F_t is an $m \times 1$ vector of non-stationary factors, and $u_t = (u_{1,t}, ..., u_{N,t})'$ 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, i.e.

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

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 Supplement.

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 $1 \le i \le N$; (iii) $|c_{i,j}^u| = O(\rho^j)$, $1 \le i \le N$, for some $0 < \rho < 1$.*

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

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

Assumption 8. *It holds that (i) as $\min(N,T) \to \infty$, $(NT)^{-2/\eta}\sum_{i=1}^{N}\sum_{t=1}^{T}\Delta u_{i,t}^2 \xrightarrow{w} G_{\eta/2}$; and (ii) for all nonzero vectors $l \in \mathbb{R}^m$, as $T \to \infty$, $T^{-2/\eta}\sum_{t=1}^{T}(l'\Delta F_t)^2 \xrightarrow{w} G^*_{\eta/2}$, 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 $T \geq T_0$*

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

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

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

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

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

in turn, this is implied by

$$\frac{N^{2/\eta-1-\epsilon}}{T} \to 0, \tag{4.9}$$

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 $\eta = 2$, i.e., when the variance exists, (4.9) boils down, essentially, to requiring $T \to \infty$ - i.e., 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.

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

where $\kappa > 0$ is user-defined (and arbitrarily small), and use $\check{\phi}_{N,T}^{(j)} = \exp(\check{\nu}_{N,T}^{(j)}) - 1$ 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 $\check{\nu}_{N,T}^{(j)}$ 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 randomised version of $\check{\phi}_{N,T}^{(1)}$. If the null is rejected, set $\check{m} = 0$ 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 the randomised version of $\check{\phi}_{N,T}^{(1)}$, constructed using an artificial sample $\{\xi_i^{(j)}\}_{i=1}^{M}$ generated independently of $\{\xi_i^{(1)}\}_{i=1}^{M}$, ..., $\{\xi_i^{(j-1)}\}_{i=1}^{M}$. 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* $\min(N,T,M) \to \infty$ *under (3.5), it holds that* $P^*(\check{m} = m) \to 1$ *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 $\check{\nu}_{N,T}^{(j)}$ as follows (see also Barigozzi and Trapani, 2022, for the finite variance case)

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

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 Supplement.

5.1. The fixed N case

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

$$y_t = Ay_{t-1} + \varepsilon_t, \tag{5.1}$$

initialized at $y_0 = 0$. We parameterise 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 *i.i.d.* 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 et al. (2009) and generate $\varepsilon_{i,t}$ as

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

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

[5] We 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 $vec\left(d_{N\times(N-m)}\right) \sim N\left(0, \mathbf{1}_{N(N-m)}\right)$. We keep $d_{N\times(N-m)}$ fixed across Monte Carlo iterations.

[6] In unreported experiments, we considered $\varepsilon_t \sim i.i.d.N(0, I_N)$; results are essentially the same as with $\eta = 2$.

TABLE 1
Estimation frequencies - $N = 3$

		$N = 3$							
		$T = 100$				$T = 200$			
	m	3	2	1	0	3	2	1	0
	$\widehat{m}$								
	3	0.963	0.004	0.000	0.000	0.986	0.001	0.000	0.000
$\eta = 0.5$	2	0.037	0.990	0.011	0.000	0.013	0.994	0.002	0.000
	1	0.000	0.006	0.989	0.023	0.001	0.003	0.998	0.005
	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
$\eta = 1.0$	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
	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
$\eta = 1.5$	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
	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
$\eta = 2$	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
	0	0.000	0.000	0.000	1.000	0.000	0.000	0.000	1.000

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, e.g., $M = T$, $M = T/2$ 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 non-spiked eigenvalues is given by slowly varying functions, which suggests that even a very small value of κ 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 $1,000$ 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

[7] Details are in Section B.1 of the Supplement.

TABLE 2
Estimation frequencies - $N = 4$

		$N = 4$									
		$T = 100$					$T = 200$				
	m	4	3	2	1	0	4	3	2	1	0
	$\widehat{m}$										
	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
$\eta = 0.5$	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
	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
$\eta = 1.0$	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
	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
$\eta = 1.5$	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
	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
$\eta = 2$	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

clear: as the VAR dimension increases, the performance of $\widehat{m}$ 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 $\widehat{m}$ 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 *i.i.d.* case, unreported experiments showed that results are essentially the same when allowing for serial dependence.

In the Supplement, 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,

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
	$\widehat{m}$												
	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
$\eta = 0.5$	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
	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
$\eta = 1.0$	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
	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
$\eta = 1.5$	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
	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
$\eta = 2$	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

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; whilst not designed for the case $N \to \infty$, our sequential procedure is, in general, satisfactory, at least when 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 *i.i.d.* across i with $\lambda_i \sim N(0,1)$; we consider $N \in \{20, 50, 100, 200\}$

TABLE 4
Estimation frequencies, large N and $m = 0$

	N	20				50			
	T	200	400	800	1600	200	400	800	1600
		m^*	m^*	m^*	m^*	m^*	m^*	m^*	m^*
	m								
	0	0.768	0.890	0.960	0.998	0.932	0.962	0.944	0.956
$\eta = 1.9$	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
	3	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
	m								
	0	0.890	0.884	0.906	0.968	0.892	0.908	0.888	0.926
$\eta = 1.5$	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
	3	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
	m								
	0	0.736	0.764	0.812	0.978	0.710	0.728	0.730	0.888
$\eta = 1.1$	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
	3	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000

	N	100				200			
	T	200	400	800	1600	200	400	800	1600
		m^*	m^*	m^*	m^*	m^*	m^*	m^*	m^*
	m								
	0	0.968	0.958	0.966	0.958	0.940	0.972	0.968	0.958
$\eta = 1.9$	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
	3	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
	m								
	0	0.894	0.876	0.888	0.892	0.886	0.900	0.876	0.890
$\eta = 1.5$	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
	3	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
	m								
	0	0.716	0.734	0.736	0.876	0.710	0.760	0.744	0.756
$\eta = 1.1$	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
	3	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000

and $T \in \{200, 400, 800, 1600\}$, and report results for $\eta \in \{1.1, 1.5, 1.9\}$ obtained using $\widetilde{\nu}^{(j)}_{N,T}$ 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 et al., 2019). As can be seen, results are broadly in line with the theory: the performance of $\breve{m}$ 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

TABLE 5

Estimation frequencies, large N and $m = 1$

	N	20								50							
	T	200		400		800		1600		200		400		800		1600	
		m^*	ER	m^*	ER	m^*	ER	m^*	ER	m^*	ER	m^*	ER	m^*	ER	m^*	ER
	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
$\eta = 1.9$	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
	3	0.000	0.004	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
	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
$\eta = 1.5$	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
	3	0.000	0.000	0.000	0.008	0.000	0.008	0.000	0.000	0.000	0.008	0.000	0.008	0.000	0.000	0.000	0.002
	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
$\eta = 1.1$	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
	3	0.000	0.054	0.000	0.036	0.000	0.039	0.000	0.018	0.000	0.072	0.000	0.066	0.000	0.032	0.000	0.028

	N	100								200							
	T	200		400		800		1600		200		400		800		1600	
		m^*	ER	m^*	ER	m^*	ER	m^*	ER	m^*	ER	m^*	ER	m^*	ER	m^*	ER
	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
$\eta = 1.9$	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
	3	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
	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
$\eta = 1.5$	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
	3	0.000	0.008	0.000	0.002	0.000	0.002	0.000	0.000	0.000	0.014	0.000	0.008	0.000	0.002	0.000	0.000
	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
$\eta = 1.1$	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
	3	0.000	0.088	0.000	0.060	0.000	0.052	0.000	0.038	0.000	0.120	0.000	0.086	0.000	0.062	0.000	0.058

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 Supplement.

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 metals (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 towards 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 et al., 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

TABLE 6

Estimation frequencies, large N and $m = 2$

	N	20								50							
	T	200		400		800		1600		200		400		800		1600	
		m^*	ER	m^*	ER	m^*	ER	m^*	ER	m^*	ER	m^*	ER	m^*	ER	m^*	ER
	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
$\eta = 1.9$	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
	3	0.000	0.012	0.000	0.022	0.002	0.012	0.000	0.016	0.000	0.028	0.002	0.018	0.000	0.020	0.000	0.006
	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
$\eta = 1.5$	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
	3	0.004	0.066	0.000	0.066	0.004	0.054	0.008	0.046	0.000	0.084	0.004	0.064	0.012	0.064	0.012	0.054
	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
$\eta = 1.1$	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
	3	0.002	0.158	0.000	0.154	0.020	0.142	0.020	0.118	0.004	0.182	0.002	0.172	0.030	0.142	0.050	0.140
	N	100								200							
	T	200		400		800		1600		200		400		800		1600	
		m^*	ER	m^*	ER	m^*	ER	m^*	ER	m^*	ER	m^*	ER	m^*	ER	m^*	ER
	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
$\eta = 1.9$	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
	3	0.000	0.032	0.000	0.038	0.000	0.020	0.000	0.014	0.002	0.050	0.002	0.024	0.002	0.016	0.002	0.010
	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
$\eta = 1.5$	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
	3	0.000	0.106	0.008	0.094	0.014	0.066	0.014	0.058	0.006	0.118	0.006	0.088	0.000	0.078	0.022	0.064
	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
$\eta = 1.1$	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
	3	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

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, i.e., $\kappa = 10^{-4}$, $M = 100$ 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. 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

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

TABLE 7
Estimated number of common trends; whole dataset

Results and sensitivity analysis

Commodity	*Tail index*	*Test* $H_0 : E\|X\|^2 = \infty$
Copper	1.658 (1.144,2.171)	0.9527 (do not reject H_0)
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)
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)

nominal level	1%	5%	10%
Johansen's trace test	1	2	2
Johansen's $\lambda_{\max}$ test	0	0	1

$\left(\kappa = 10^{-4}, n_S = 2\right)$	$\frac{0.05}{T}$	$\frac{0.05}{\ln T}$	$\frac{0.05}{N}$
$T/2$	5	4	4
T	4	4	4
$2T$	4	4	4

$\left(\kappa = 10^{-2}, n_S = 2\right)$	$\frac{0.05}{T}$	$\frac{0.05}{\ln T}$	$\frac{0.05}{N}$
$T/2$	5	4	4
T	4	4	4
$2T$	4	4	4

$\left(\kappa = 10^{-4}, n_S = 4\right)$	$\frac{0.05}{T}$	$\frac{0.05}{\ln T}$	$\frac{0.05}{N}$
$T/2$	5	4	4
T	5	4	4
$2T$	4	4	4

$\left(\kappa = 10^{-2}, n_S = 4\right)$	$\frac{0.05}{T}$	$\frac{0.05}{\ln T}$	$\frac{0.05}{N}$
$T/2$	5	4	4
T	5	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 Supplement.
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 $\widehat{m}$ 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 $\frac{0.05}{N}$.

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 towards $m = 4$ 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

TABLE 8
Estimated number of common trends; oil prices

Results and sensitivity analysis

nominal level	1%	5%	10%
Johansen's trace test	2	2	2
Johansen's $\lambda_{\max}$ test	2	2	2

Cointegration vectors

$\widehat{\beta}_{1,1}$	1.00	$\widehat{\beta}_{2,1}$	0.00
$\widehat{\beta}_{1,2}$	0.00	$\widehat{\beta}_{2,2}$	1.00
$\widehat{\beta}_{1,3}$	−1.08	$\widehat{\beta}_{2,3}$	−1.05

$\left(\kappa = 10^{-4}, n_S = 2\right)$	$\frac{0.05}{T}$	$\frac{0.05}{\ln T}$	$\frac{0.05}{N}$
$T/2$	2	2	1
T	2	1	1
$2T$	1	1	1

$\left(\kappa = 10^{-2}, n_S = 2\right)$	$\frac{0.05}{T}$	$\frac{0.05}{\ln T}$	$\frac{0.05}{N}$
$T/2$	2	1	1
T	1	1	1
$2T$	1	1	1

$\left(\kappa = 10^{-4}, n_S = 4\right)$	$\frac{0.05}{T}$	$\frac{0.05}{\ln T}$	$\frac{0.05}{N}$
$T/2$	2	2	1
T	2	1	1
$2T$	1	1	1

$\left(\kappa = 10^{-2}, n_S = 4\right)$	$\frac{0.05}{T}$	$\frac{0.05}{\ln T}$	$\frac{0.05}{N}$
$T/2$	2	2	1
T	2	1	1
$2T$	1	1	1

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

constrain the upper $m \times m$ block of the estimated loadings matrix $\widehat{P}$ to be the identity matrix. We report the loadings $\widehat{P}$ in Table 10 (see also Figure C.1 in the Supplement, where we plot the estimated common trends $\widehat{x}_t$).

By construction, the first and second trends ($\widehat{x}_{1,t}$ and $\widehat{x}_{2,t}$) are associated with oil prices and cobalt respectively. The third one ($\widehat{x}_{3,t}$) is associated with gold (by construction), and nickel, with a negative loading; finally, the fourth trend ($\widehat{x}_{4,t}$) is associate 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 US 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 ($T = 405$). demeaned. By way of preliminary analysis (reported in Section C.2 in the Supplement), we have applied the test by

[9]We 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 Supplement.

TABLE 9
Estimated number of common trends; metal prices

Results and sensitivity analysis				Cointegration vectors	
nominal level	1%	5%	10%		
				$\widehat{\beta}_{1,1}$	1.00
Johansen's trace test	0	1	1	$\widehat{\beta}_{1,2}$	−0.69
				$\widehat{\beta}_{1,3}$	−0.50
Johansen's $\lambda_{\max}$ test	0	1	1	$\widehat{\beta}_{1,4}$	−0.13

$\left(\kappa = 10^{-4}, n_S = 2\right)$	$\frac{0.05}{T}$	$\frac{0.05}{\ln T}$	$\frac{0.05}{N}$	$\left(\kappa = 10^{-2}, n_S = 2\right)$	$\frac{0.05}{T}$	$\frac{0.05}{\ln T}$	$\frac{0.05}{N}$
$T/2$	1	3	3	$T/2$	3	3	3
T	3	3	3	T	3	3	3
$2T$	3	3	3	$2T$	3	3	3
$\left(\kappa = 10^{-4}, n_S = 4\right)$	$\frac{0.05}{T}$	$\frac{0.05}{\ln T}$	$\frac{0.05}{N}$	$\left(\kappa = 10^{-2}, n_S = 4\right)$	$\frac{0.05}{T}$	$\frac{0.05}{\ln T}$	$\frac{0.05}{N}$
$T/2$	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 $\widehat{P}$

	$\widehat{x}_{1,t}$	$\widehat{x}_{2,t}$	$\widehat{x}_{3,t}$	$\widehat{x}_{4,t}$
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

Trapani (2016) to all series, which lends strong support to the 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 trends); the evidence reported in the table suggests that the number of common stochastic trends m is larger than 3. This also confirms that

TABLE 11
Estimated number of common trends; US interest rates

Results and sensitivity analysis

Other estimators

IC	$BT1$	$BT2$
5	3	3

Number of common trends estimated using Algorithm 2 with (4.10)

$\left(\kappa = 10^{-4}, n_S = 2\right)$	$\frac{0.05}{T}$	$\frac{0.05}{\ln T}$	$\frac{0.05}{N}$	$\left(\kappa = 10^{-2}, n_S = 2\right)$	$\frac{0.05}{T}$	$\frac{0.05}{\ln T}$	$\frac{0.05}{N}$
$T/2$	5	5	5	$T/2$	5	5	5
T	5	5	5	T	5	5	5
$2T$	5	5	5	$2T$	5	5	5
$\left(\kappa = 10^{-4}, n_S = 4\right)$	$\frac{0.05}{T}$	$\frac{0.05}{\ln T}$	$\frac{0.05}{N}$	$\left(\kappa = 10^{-2}, n_S = 4\right)$	$\frac{0.05}{T}$	$\frac{0.05}{\ln T}$	$\frac{0.05}{N}$
$T/2$	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)

$\left(\kappa = 10^{-4}, n_S = 2\right)$	$\frac{0.05}{T}$	$\frac{0.05}{\ln T}$	$\frac{0.05}{N}$	$\left(\kappa = 10^{-2}, n_S = 2\right)$	$\frac{0.05}{T}$	$\frac{0.05}{\ln T}$	$\frac{0.05}{N}$
$T/2$	5	5	5	$T/2$	5	5	5
T	5	5	5	T	4	4	4
$2T$	4	4	4	$2T$	3	3	3
$\left(\kappa = 10^{-4}, n_S = 4\right)$	$\frac{0.05}{T}$	$\frac{0.05}{\ln T}$	$\frac{0.05}{N}$	$\left(\kappa = 10^{-2}, n_S = 4\right)$	$\frac{0.05}{T}$	$\frac{0.05}{\ln T}$	$\frac{0.05}{N}$
$T/2$	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 $\check{m}$ obtained using Algorithm 2 under different specifications - 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 $\frac{0.05}{N}$. 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).

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 non-stationary. 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 non-stationary 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 Supplement.

7. Conclusions

In this paper, 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 randomised 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 deterministics, *i.i.d.* 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 supplement.

References

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

Ahn, S. K. and G. C. Reinsel (1990). Estimation for partially nonstationary multivariate autoregressive models. *Journal of the American Statistical Association 85*, 813–823.

Alquist, R., S. Bhattarai, and O. Coibion (2020). Commodity-price comovement and global economic activity. *Journal of Monetary Economics 112*, 41–56.

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

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

Barigozzi, M. and L. Trapani (2022). Testing for common trends in nonstationary large datasets. *Journal of Business and Economic Statistics*. forthcoming.

Bernstein, D. H. and B. Nielsen (2019). Asymptotic theory for cointegration analysis when the cointegration rank is deficient. *Econometrics 7*, 6.

Bykhovskaya, A. and V. Gorin (2022). Cointegration in large VARs. *Annals of Statistics*. forthcoming.

Caner, M. (1998). Tests for cointegration with infinite variance errors. *Journal of Econometrics 86*, 155–175.

Cavaliere, G., I. Georgiev, and A. R. Taylor (2018). Unit root inference for non-stationary linear processes driven by infinite variance innovations. *Econometric Theory 34*, 302–348.

Cavaliere, G., A. Rahbek, and A. R. Taylor (2012). Bootstrap determination of the co-integration rank in vector autoregressive models. *Econometrica 80*, 1721–1740.

Cavaliere, G. and A. R. Taylor (2009). Heteroskedastic time series with a unit root. *Econometric Theory*, 1228–1276.

Chan, N. H. and R. Zhang (2012). Non-stationary autoregressive processes with infinite variance. *Journal of Time Series Analysis 33*, 916–934.

Chen, L., J. J. Dolado, and J. Gonzalo (2021). Quantile factor models. *Econometrica 89*(2), 875–910.

Clauset, A., C. R. Shalizi, and M. E. Newman (2009). Power-law distributions in empirical data. *SIAM review 51*, 661–703.

Davis, R. and S. Resnick (1985). Limit theory for moving averages of random variables with regularly varying tail probabilities. *The Annals of Probability*, 179–195.

Davis, R. and S. Resnick (1986). Limit theory for the sample covariance and correlation functions of moving averages. *The Annals of Statistics*, 533–558.

Davis, R. A. (2010). Heavy tails. *Encyclopedia of Quantitative Finance*.

Davis, R. A., O. Pfaffel, and R. Stelzer (2014). Limit theory for the largest eigenvalues of sample covariance matrices with heavy-tails. *Stochastic Processes and their Applications 124*, 18–50.

Degiannakis, S., G. Filis, G. Siourounis, and L. Trapani (2021). Superkurtosis. *Journal of Money, Credit and Banking*. forthcoming.

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

Embrechts, P., C. Klüppelberg, and T. Mikosch (2013). *Modelling extremal events: for insurance and finance*, Volume 33. Springer Science & Business Media.

Engle, R. and C. W. Granger (1987). Co-integration and error correction: Representation, estimation, and testing. *Econometrica 55*, 251–276.

Falk, B. and C.-H. Wang (2003). Testing long-run PPP with infinite-variance returns. *Journal of Applied Econometrics 18*, 471–484.

Fan, J., H. Liu, and W. Wang (2018). Large covariance estimation through elliptical factor models. *Annals of Statistics 46*, 1383.

Fasen, V. (2013). Time series regression on integrated continuous-time processes with heavy and light tails. *Econometric Theory*, 28–67.

Gabaix, X. (1999). Zipf's law for cities: an explanation. *The Quarterly Journal of Economics 114*, 739–767.

Gonzalo, J. and C. Granger (1995). Estimation of common long-memory components in cointegrated systems. *Journal of Business & Economic Statistics 13*, 27–35.

Hall, P., L. Peng, and Q. Yao (2002). Prediction and nonparametric estimation for time series with heavy tails. *Journal of Time Series Analysis 23*(3), 313–331.

Hallin, M., R. Van den Akker, and B. J. Werker (2011). A class of simple semiparametrically efficient rank-based unit root tests. *Journal of Econometrics 163*, 200–214.

Hallin, M., R. van den Akker, and B. J. Werker (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.

He, Y., X. Kong, L. Yu, and X. Zhang (2022). Large-dimensional factor analysis without moment constraints. *Journal of Business & Economic Statistics 40*, 302–312.

Horváth, L. and L. Trapani (2019). Testing for randomness in a random coefficient autoregression model. *Journal of Econometrics 209*, 338–352.

Ibragimov, M. and R. Ibragimov (2018). Heavy tails and upper-tail inequality: The case of Russia. *Empirical Economics 54*, 823–837.

Jach, A. and P. Kokoszka (2004). Subsampling unit root tests for heavy-tailed observations. *Methodology and Computing in Applied Probability 6*, 73–97.

Johansen, S. (1991). Estimation and hypothesis testing of cointegration vectors in Gaussian vector autoregressive models. *Econometrica*, 1551–1580.

Kasa, K. (1992). Common stochastic trends in international stock markets. *Journal of Monetary Economics 29*, 95–124.

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

Liang, C. and M. Schienle (2019). Determination of vector error correction models in high dimensions. *Journal of Econometrics 208*, 418–441.

McElroy, T. and D. N. Politis (2003). Limit theorems for heavy-tailed random fields with subsampling applications. *Mathematical Methods of Statistics 12*, 305–328.

Nielsen, M. Ø. (2010). Nonparametric cointegration analysis of fractional systems with unknown integration orders. *Journal of Econometrics 155*, 170–187.

Nyblom, J. and A. Harvey (2000). Tests of common stochastic trends. *Econometric theory 16*, 176–199.

Onatski, A. and C. Wang (2018). Alternative asymptotics for cointegration tests in large VARs. *Econometrica 86*, 1465–

1478.

Onatski, A. and C. Wang (2021). Spurious factor analysis. *Econometrica 89*, 591–614.

Patilea, V. and H. Raïssi (2014). Testing second-order dynamics for autoregressive processes in presence of time-varying variance. *Journal of the American Statistical Association 109*, 1099–1111.

Paulauskas, V. and S. T. Rachev (1998). Cointegrated processes with infinite variance innovations. *Annals of Applied Probability*, 775–792.

Peña, D. and P. Poncela (2006). Nonstationary dynamic factor analysis. *Journal of Statistical Planning and Inference 136*, 1237–1257.

Petrov, V. (1974). Moments of distributions attracted to stable laws. *Theory of Probability & Its Applications 18*, 569–571.

Pham, T. D. and L. T. Tran (1985). Some mixing properties of time series models. *Stochastic Processes and their Applications 19*, 297–303.

Qu, Z. and P. Perron (2007). A modified information criterion for cointegration tests based on a VAR approximation. *Econometric Theory*, 638–685.

Samorodnitsky, G. and M. S. Taqqu (1994). *Stable non-Gaussian random processes*. Chapman & Hall.

Seneta, E. (2006). *Regularly varying functions*, Volume 508. Springer.

Shao, X. (2015). Self-normalization for time series: a review of recent developments. *Journal of the American Statistical Association 110*, 1797–1817.

She, R. and S. Ling (2020). Inference in heavy-tailed vector error correction models. *Journal of Econometrics 214*, 433–450.

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

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.

Trapani, L. (2016). Testing for (in)finite moments. *Journal of Econometrics 191*, 57–68.

Tu, Y., Q. Yao, and R. Zhang (2020). Error-correction factor models for high-dimensional cointegrated time series. *Statistica Sinica 30*, 1463–1484.

Watson, M. W. (1994). Vector autoregressions and cointegration. *Handbook of Econometrics 4*, 2843–2915.

Yap, S. F. and G. C. Reinsel (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.

Yu, L., Y. He, and X. Zhang (2019). Robust factor number specification for large-dimensional elliptical factor model. *Journal of Multivariate Analysis 174*, 104543.

Zhang, R., P. Robinson, and Q. Yao (2019). Identifying cointegration by eigenanalysis. *Journal of the American Statistical Association 114*, 916–927.