Econometrics and Statistics xxx (xxxx) xxx

Contents lists available at ScienceDirect

Econometrics and Statistics

journal homepage: www.elsevier.com/locate/ecosta

Estimating the number of common trends in large T and N factor models via canonical correlations analysis

Massimo Franchi [a,*], Iliyan Georgiev [b], Paolo Paruolo [c]

[a] Sapienza University of Rome, P.le Aldo Moro 5, Rome, 00185, Italy
[b] University of Bologna, Piazza Scaravilli 2, Bologna, 40126, Italy
[c] Joint Research Centre (JRC), Via Enrico Fermi 2749, Ispra (VA), 21027, Italy

ARTICLE INFO

Article history:
Received 29 May 2023
Revised 25 September 2023
Accepted 9 October 2023
Available online xxx

JEL classification:
C32

Keywords:
Common trends
Nonstationary large panels
Factor models
Unit root processes
Cointegration
Canonical Correlations Analysis

ABSTRACT

Asymptotic results for canonical correlations are derived when the analysis is performed between levels and cumulated levels of N time series of length T, generated by a factor model with s common stochastic trends. For $T \to \infty$ and fixed N and s, the largest s squared canonical correlations are shown to converge to a non-degenerate limit distribution while the remaining $N - s$ converge in probability to 0. Furthermore, if s grows at most linearly in N, the largest s squared canonical correlations are shown to converge in probability to 1 as $(T, N)_{seq} \to \infty$. This feature allows one to estimate the number of common trends as the integer with largest decrease in adjacent squared canonical correlations. The maximal gap equals 1 in the limit and this criterion is shown to be consistent. A Monte Carlo simulation study illustrates the findings.

1. Introduction

Many economic and financial time series exhibit persistent co-movements, that reflect smooth economic changes over time. These phenomena are well replicated by linear processes with common stochastic trends, i.e. with fewer stochastic trends than variables, see Stock and Watson (1988), or, equivalently, with cointegration, see Engle and Granger (1987). Because of this, the number of common trends plays a central role in the analysis of economic and financial time series.

The number of common stochastic trends, s say, and the number of linearly independent cointegrating relations, r say, are complementary, i.e. $N = r + s$, where N is the number of time series, in case of processes integrated of order 1 or 0, indicated as $I(1)$, $I(0)$. This type of complementarity is proved in Granger-type representation theorems, see e.g. Johansen (1996) and references therein.

Cointegration analysis initially focused on the case of small N and large T, where T is the number of time periods, see e.g. Engle and Granger (1987); Stock and Watson (1988, 1993); Ahn and Reinsel (1990); Phillips (1991), Johansen (1988, 1991, 1996). These contributions were derived in a linear process framework, typically of Vector AutoRegressive (VAR) or Vector AutoRegressive Moving Average (VARMA) type, and results involved large-T asymptotics. Simulations in Abadir et al. (1999) and Gonzalo and Pitarakis (1999) showed that the small-T properties of many of these inferential procedures differ from the large-T asymptotic distributions, and these discrepancies worsen as N increases.

* Corresponding author.
E-mail addresses: massimo.franchi@uniroma1.it (M. Franchi), i.georgiev@unibo.it (I. Georgiev), paolo.paruolo@ec.europa.eu (P. Paruolo).

https://doi.org/10.1016/j.ecosta.2023.10.001

Please cite this article as: M. Franchi, I. Georgiev and P. Paruolo, Estimating the number of common trends in large T and N factor models via canonical correlations analysis, Econometrics and Statistics, https://doi.org/10.1016/j.ecosta.2023.10.001

In parallel to the cointegration literature, Dynamic Factor Models (DFMs) were introduced and discussed by Geweke (1977) and Sargent and Sims (1977). They were further developed by Forni et al. (2000, 2004, 2005); Forni and Lippi (2001); Bai (2003); Bai and Ng (2002, 2007). DFMs were introduced as models for large panels of stationary time series with stationary co-movements, see e.g. the surveys in Stock and Watson (2010); Anderson et al. (2022); Lippi et al. (2023).

Early DFM contributions, with the notable exception of Bai (2004), were based on the assumption of stationarity of variables and factors; the corresponding analysis could be applied to panels of nonstationary time series only after prior variable transformations, such as first differencing or de-trending. In this setup, a few stationary dynamic factors are the driving forces responsible for a large fraction of the variation of the variables.

Criteria and tests for the number of common factors in DFM are largely available in a stationary setting, see e.g. Onatski (2010); Ahn and Horenstein (2013), and Trapani (2018) for recent results; Bai (2004) has long been the only contribution dealing with the non-stationary case, see Barigozzi and Trapani (2022).

More recent strands of literature consider stochastic trends in large dimensional systems. In a VAR context, Onatski and Wang (2018, 2019) show that the empirical distribution function of the squared sample canonical correlations that appear in the likelihood ratio (LR) rank test converges to the Wachter distribution when N and T go to infinity simultaneously and proportionally, and r/N converges to 0. This result explains the observed LR rank tests over-rejection of the null hypothesis of no cointegration.

Using a similar asymptotic setting, Bykhovskaya and Gorin (2022) propose a modification of the LR rank statistics to test the null of no cointegration in a VAR(1) and Liang and Schienle (2019) provide a Lasso-type consistent estimator of the number s of stochastic trends when N is large and can increase with T.

Zhang et al. (2019) propose a consistent estimator of s and of the cointegrating space for nonstationary time series of unknown, possibly fractional, integration orders, both when N is fixed and T increases, and when N diverges slowly with T. In a DFM context, Barigozzi and Trapani (2022) and Barigozzi et al. (2022) propose estimators of s for large N and T, where the latter paper allows also for the possible presence of innovations with heavy tails.

The present paper is closely related to the recent developments in the analysis of stochastic trends in a DFM framework. It proposes and discusses a criterion to determine the number s of common stochastic trends in the N-dimensional DFM $X_t = \Lambda F_t + \xi_t$, where the common factors F_t are $I(1)$ and cointegrated and the idiosyncratic components ξ_t are $I(0)$ and non-cointegrated, but where F_t and ξ_t are both serially and cross-sectionally correlated. The present setup is a particular case of the one in Barigozzi and Trapani (2022), who allow F_t also to contain a drift, and of the one in Barigozzi et al. (2021), see also Barigozzi et al. (2020), where the idiosyncratic component ξ_t can also be $I(1)$ and possibly cointegrated. Moreover, the present setup includes the one in Bai (2004), where F_t is a pure random walk and ξ_t is cross-sectionally uncorrelated.

The proposed criterion to determine the number s of common stochastic trends is based on a Canonical Correlations Analysis (CCA) between levels of the variables X_t and cumulated levels $\mathcal{S}X_t$, where the operator $\mathcal{S}$ performs the cumulation of the time series, see Gregoir (1999). Results are based on large T and N sequential asymptotics, first for $T \to \infty$ and then for $N \to \infty$, indicated as $(T, N)_{seq} \to \infty$, under the assumption that s grows at most linearly in N. It is shown that the s largest squared canonical correlations converge to 1 in probability, while the remaining ones converge to 0 as $(T, N)_{seq} \to \infty$. This feature allows to estimate the number of common trends consistently as the number for which consecutive squared canonical correlations have the largest decrease. The Monte Carlo simulations show that the proposed criterion is accurate for $N - s$ not too small, when the sample size T is at least 3 times larger than the cross-sectional dimension N.

The assumption that s grows at most linearly in N implies that $r = N - s \to \infty$ as N diverges. This is in line with the tradition of DFMs where the number of common trends is assumed fixed for increasing cross sectional dimension N, see e.g. Barigozzi et al. (2022); Barigozzi and Trapani (2022). On the other hand, assuming $s = \lfloor cN \rfloor$ with $0 < c < 1$ is complementary to the large dimensional VAR analysis in Onatski and Wang (2018) and Bykhovskaya and Gorin (2022), who assume $s/N \to 1$, which is excluded in the present paper because $c \neq 1$.

CCA has been used to analyze links among time series in several areas, including in likelihood-based analysis of cointegration in VARs, see Johansen (1988, 1991, 1996) and Ahn and Reinsel (1990). In those settings, CCA is performed between differences ΔX_t and lagged levels X_{t-1} after correcting for $\Delta X_{t-1}, \dots \Delta X_{t-k+1}$ and deterministic terms. In this paper, CCA is applied to levels X_t and cumulated levels $\mathcal{S}X_t$; to the best of the authors' knowledge, Breitung (1998) appears to be the only paper considering such a CCA, although it did not consider the large N analysis of this paper.

On the contrary, the idea of performing a principal component analysis of levels and cumulated levels and exploiting the difference in the stochastic order of the eigenvalues has already been employed in the literature, see e.g. Breitung (2002); Harris (1997); Nyblom and Harvey (2000); Nielsen et al. (2022).

The $(T, N)_{seq} \to \infty$ sequential asymptotics is derived in two stages. The first part (for $T \to \infty$ and fixed N, s) shows that the largest s squared canonical correlations converge to the R^2 of the regression of the limit non-stationary part of X_t (driven by s Brownian motions) on the limit of $\mathcal{S}X_t$, which consists of $N - s$ Brownian motions plus the cumulation of the s Brownian motions present in X_t; the $N - s$ smallest squared canonical correlations converge in probability to zero. In the second part (for $N \to \infty$ and s growing at most linearly in N), it is found that the largest s squared canonical correlations converge to 1, i.e. the limit densities of these squared canonical correlations shrink toward the point mass distribution at 1 as $(T, N)_{seq} \to \infty$.

The asymptotics for large N in the second part uses an extension of the results in Phillips (1998), who showed that the R^2 of an $L^2[0, 1]$-regression of a Brownian motion $W(u)$ on a k-variate independent standard Brownian motion $B^{(k)}(u)$, indicated as $B^{(k)}(u) = BM(I_k)$, approaches one in probability as k diverges. Here and in the following I_k indicates the iden-

tity matrix of order k. This convergence is due to a functional approximation of the sample paths of $W(u)$ and not to an increasing degree of statistical dependence between the stochastic processes $W(u)$ and $B^{(k)}(u)$, as represented by the statistical regression function $E(W(u)|B^{(k)}(u))$; thus, this convergence is often understood as spurious, see Johansen (2012).

More precisely, consider the statistical regression function $E(W(u)|B^{(k)}(u))$, which is 0 if $W(u)$ and $B^{(k)}(u)$ are independent. Let R_k^2 indicate the R^2 of the $L^2[0,1]$-regression of $W(u)$ on $B^{(k)}(u)$ and let $s_k^2 := \int_0^1 (W(u) - E(W(u)|B^{(k)}(u)))^2 du$ indicate the square of the $L^2[0,1]$-norm of the statistical regression error. From a statistical standpoint, spuriousness of the $L^2[0,1]$-regression for $k \to \infty$ can be understood as $R_k^2 \xrightarrow{p} 1$ with s_k^2 stochastically bounded away from 0; this is what Phillips (1998) proved when $W(u)$ and $B^{(k)}(u)$ are independent and $B^{(k)}(u) = BM(I_k)$.

The present paper extends this spurious regression result to the case of dependent $W(u)$ and $B^{(k)}(u)$, with $B^{(k)}(u) = BM(\Sigma_k)$, $\Sigma_k > 0$ (positive definite). The proof in the present paper shows that also in this case $R_k^2 \xrightarrow{p} 1$ with s_k^2 stochastically bounded away from 0, i.e. that the same spurious regression results hold. Note that the present setting includes the case $E(W(u)|B^{(k)}(u)) = 0$ and $\Sigma_k = I_k$ considered in Phillips (1998) as a special case.

The rest of the paper is organized as follows: Section 2 discusses the class of processes considered in the paper, Section 3 derives their Common Trends representation, Section 4 presents the statistical analysis, Section 5 reports the main results, Section 6 illustrates simulation results, and Section 7 concludes. Appendix A contains proofs and Appendix B tables of the Monte Carlo simulations.

2. Setup

Consider the (large-dimensional) Dynamic Factor Model (DFM)

$$X_t = \Lambda F_t + \xi_t, \quad \Lambda \in \mathbb{R}^{N \times N_F}, \quad \operatorname{rank} \Lambda = N_F < N, \tag{1}$$

where $\{X_t, \xi_t\}_{t\in\mathbb{Z}}$ are $N \times 1$ and $\{F_t\}_{t\in\mathbb{Z}}$ is $N_F \times 1$; the differences of the common factors F_t and the idiosyncratic component ξ_t are assumed to be linear processes generated by independent shocks $\{\zeta_t, \varepsilon_t\}_{t\in\mathbb{Z}}$ of dimension $N_\zeta \times 1$ and $N \times 1$ respectively, $N_\zeta \le N_F$, i.e. with

$$\eta_t := \begin{pmatrix} \zeta_t \\ \varepsilon_t \end{pmatrix} \sim i.i.d.(0, \Omega), \quad \Omega = \begin{pmatrix} I & 0 \\ 0 & \Omega_\varepsilon \end{pmatrix} > 0, \quad N_\eta := N_\zeta + N, \tag{2}$$

of dimension $N_\eta \times 1$, and

$$\begin{matrix} \Delta F_t = C(L)\zeta_t \\ \xi_t = B(L)\varepsilon_t \end{matrix}, \quad \begin{matrix} 0 < s := \operatorname{rank} C(1) < N_F \\ B(1) \text{ nonsingular,} \end{matrix} \quad 0 < N_\zeta \le N_F < N < N_\eta, \tag{3}$$

where, for a generic process $\{w_t\}_{t\in\mathbb{Z}}$, $\Delta w_t := w_t - w_{t-1}$ performs the difference of the time series and $C(z) = \sum_{n=0}^{\infty} C_n z^n$, $B(z) = \sum_{n=0}^{\infty} B_n z^n$ converge for all $|z| < 1 + \delta$ for some $\delta > 0$.

Observe that $0 < s < N$, and ΔF_t and ξ_t are mutually uncorrelated; moreover, note that the convergence assumption for $C(z)$ and $B(z)$ is for instance satisfied if $\Delta F_t = C(L)\zeta_t$ and $\xi_t = B(L)\varepsilon_t$ are the MA representations of stationary VARMA processes. Fractional processes are excluded by the assumed convergence on $|z| < 1 + \delta$, see e.g. Johansen (2008, Definition 1).

The present setup is a particular case of the one in Barigozzi and Trapani (2022), who also allow for a random walk with drift in F_t. It is also a particular case of the one in Barigozzi et al. (2021), see also Barigozzi et al. (2020), where the idiosyncratic component ξ_t can be $I(1)$ and possibly cointegrated, and it includes the one in Bai (2004), where $F_t = F_{t-1} + \zeta_t$ is a pure random walk and Ω_ε is diagonal.

The next definition presents the notions of $I(d)$ process in Johansen (1996) and $I_{nc}(d)$ process in Franchi and Paruolo (2019).

Definition 2.1. ($I(d)$ and $I_{nc}(d)$, $d = 0, 1, \dots$) Consider a linear process $u_t = A(L)w_t$, where $w_t \sim i.i.d.(0, \Omega_w)$ and $A(z) = \sum_{n=0}^{\infty} A_n z^n$ converges for all $|z| < 1 + \delta$ for some $\delta > 0$. $\{u_t\}_{t\in\mathbb{Z}}$ is said to be integrated of order 0 – indicated as $I(0)$ – if $A(1) \neq 0$; $\{u_t\}_{t\in\mathbb{Z}}$ is said to be $I(0)$ and non-cointegrated – indicated as $I_{nc}(0)$ – if it is $I(0)$ and, in addition, $A(1)$ has full row rank. If $\Delta^d z_t \sim I(0)$, then $\{z_t\}_{t\in\mathbb{Z}}$ is said to be $I(d)$, while if $\Delta^d z_t \sim I_{nc}(0)$, then $\{z_t\}_{t\in\mathbb{Z}}$ is said to be $I(d)$ and non-cointegrated, indicated as $I_{nc}(d)$.

First observe that a linear process can be stationary but not $I(0)$, such as for instance $u_t = w_t - w_{t-1}$ with u_t, w_t as in Definition 2.1. Next consider the cumulation operator $\mathcal{S}$ in Gregoir (1999), defined as follows: for a generic process $\{w_t\}_{t\in\mathbb{Z}}$, its cumulation $\{\mathcal{S}w_t\}_{t\in\mathbb{Z}}$ is defined as $\mathcal{S}w_t := 1_{(t>1)} \sum_{i=1}^{t} w_i - 1_{(t\le 1)} \sum_{i=t+1}^{0} w_i$. The cumulation operator $\mathcal{S}$ satisfies $\mathcal{S}\Delta w_t = w_t - w_0$, $\mathcal{S}1 = t$ and $\Delta \mathcal{S} w_t = w_t$.

Further note that the cumulation of a stationary process may as well be stationary; for example cumulating $u_t = \Delta w_t$ one finds $\mathcal{S}\Delta w_t = w_t - w_0$, which is stationary up to the choice of initial values w_0. This cannot happen if the process is $I(0)$, i.e. the cumulation of an $I(0)$ process gives an $I(1)$ process by definition. Hence the order of integration of the cumulation of a stationary process is not well defined while it is uniquely determined if the process is $I(0)$.

Remark that the full rank assumption on $B(1)$ in (3) implies that the idiosyncratic components ξ_t are $I_{nc}(0)$ while the reduced rank assumption on $C(1)$ in (3) implies that the common factors F_t are $I(1)$ and cointegrated.

The distinction between $I(0)$ and $I_{nc}(0)$ allows one to establish the order of integration when taking linear combinations in the following sense: while $a'u_t \sim I(0)$ for any nonzero a if $u_t \sim I_{nc}(0)$, $a'u_t$ is stationary but not necessarily $I(0)$ when $u_t \sim I(0)$. Thus the order of integration of the cumulation of $a'u_t$ is not well defined if $u_t \sim I(0)$ while it is equal to 1 if $u_t \sim I_{nc}(0)$. The CCA performed in Section 4 below involves X_t and its cumulation and their integration properties in specific directions, see (9) and (11), hence the need for the distinction between I and I_{nc}.

3. Representation

This section derives the Common Trends (CT) representation of X_t, see (8) below. Consider (3), substitute $C(z) = C(1) + (1-z)C_1(z)$ in $\Delta F_t = C(L)\zeta_t$ and cumulate $\Delta F_t = C(1)\zeta_t + C_1(L)\Delta\zeta_t$; this leads to the Common Trends (CT) representation of the factors

$$F_t = C(1)\mathcal{S}\zeta_t + C_1(L)(\zeta_t - \zeta_0) + F_0, \quad 0 < s = \operatorname{rank} C(1) < N_F. \tag{4}$$

Set w.l.o.g. $F_0 = \zeta_0 = 0$ and substitute (4) and $\xi_t = B(L)\varepsilon_t$ into (1) to find the CT representation of the observables

$$X_t = C_x\mathcal{S}\eta_t + B_x(L)\eta_t, \quad \begin{matrix} C_x = (\Lambda C(1), 0) \\ B_x(z) = (\Lambda C_1(z), B(z)) \end{matrix}, \quad \eta_t = \begin{pmatrix} \zeta_t \\ \varepsilon_t \end{pmatrix}, \tag{5}$$

where C_x and $B_x(z)$ are $N \times N_\eta$ and partitioned conformably with η_t, $\operatorname{rank} C_x = \operatorname{rank} C(1) = s$ and $\operatorname{rank} B_x(1) = \operatorname{rank} B(1) = N$. Next consider a rank factorization of $C(1)$,

$$C(1) = \gamma\omega', \tag{6}$$

where γ is $N_F \times s$ and ω is $N_\zeta \times s$ both of full column rank s and let

$$\psi = \Lambda\gamma, \quad \kappa = \begin{pmatrix} \omega \\ 0 \end{pmatrix}, \tag{7}$$

where ψ is $N \times s$ and κ is $N_\zeta \times s$, both of full column rank s; observe that $\psi\kappa'$ is a rank factorization of C_x. Eq. (5) and (7) give the CT representation of the observables

$$X_t = \psi\kappa'\mathcal{S}\eta_t + B_x(L)\eta_t \sim I(1). \tag{8}$$

Observe that (8) gives a decomposition of X_t into an $I(1)$ component $\psi\kappa'\mathcal{S}\eta_t$ and a $I_{nc}(0)$ component $B_x(L)\eta_t$; that is, $X_t \sim I(1)$ has s common trends $\kappa'\mathcal{S}\eta_t \sim I_{nc}(1)$ loaded by ψ and $N-s$ cointegrating relations $\psi_\perp' X_t \sim I_{nc}(0)$. Here and in the following $a_\perp$ indicates a basis of the orthogonal complement of $\operatorname{col}(a)$, the linear space spanned by the columns of the matrix a.

The loading matrix ψ is tall and of full column rank s, with basis of its orthogonal complement $\psi_\perp$ of the form $\psi_\perp = (\Lambda_\perp, \bar{\Lambda}\gamma_\perp)$, where $\bar{a}$ indicates $\bar{a} := a(a'a)^{-1}$ for any tall matrix a of full column rank. Hence from (8) one has

$$\psi_\perp' X_t = \psi_\perp' B_x(L)\eta_t \sim I_{nc}(0), \quad \bar{\psi}' X_t = \kappa'\mathcal{S}\eta_t + \bar{\psi}' B_x(L)\eta_t \sim I_{nc}(1). \tag{9}$$

Remark that no further decrement in the order of integration is possible in (9) because $\psi_\perp' X_t$ and $\bar{\psi}' X_t$ are both I_{nc}. The aim of this paper is to discuss the estimation of s from data on X_t.

4. Canonical correlation analysis of levels and cumulated levels

Consider a CCA of the levels X_t, see (8), and cumulated levels

$$\mathcal{S}X_t = \psi\kappa'\mathcal{S}^2\eta_t + B_x(L)\mathcal{S}\eta_t \sim I(2), \tag{10}$$

which satisfy, see (9),

$$\psi_\perp'\mathcal{S}X_t \sim I_{nc}(1), \quad \bar{\psi}'\mathcal{S}X_t \sim I_{nc}(2). \tag{11}$$

Remark that no further decrement in the order of integration is possible in (11) because $\psi_\perp'\mathcal{S}X_t$ and $\bar{\psi}'\mathcal{S}X_t$ are both I_{nc}; moreover, unlike X_t, $\mathcal{S}X_t$ is nonstationary in all directions.

The CCA delivers the squared canonical correlation λ_i and the canonical variates $\varphi_i' X_t$ and $\varsigma_i'\mathcal{S}X_t$, $i = 1, \dots, N$, s.t. the i-th squared canonical correlation λ_i is maximal subject to the constraint that $\varphi_i' X_t$ ($\varsigma_i'\mathcal{S}X_t$) is uncorrelated with the previously found canonical variates $\varphi_1' X_t, \dots, \varphi_{i-1}' X_t$ ($\varsigma_1'\mathcal{S}X_t, \dots, \varsigma_{i-1}'\mathcal{S}X_t$). By construction, the squared canonical correlations satisfy $1 \geq \lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_N \geq 0$ and the $N \times N$ matrices $V = (\varphi_1, \dots, \varphi_N)$ and $Q = (\varsigma_1, \dots, \varsigma_N)$ are nonsingular.

The CCA is performed as follows: let

$$z_t = \begin{pmatrix} z_{1t} \\ z_{2t} \end{pmatrix}, \quad z_{1t} = X_t, \quad z_{2t} = \mathcal{S}X_t, \tag{12}$$

see (8) and (10), and define the sample moments

$$T^{-1}\sum_{t=1}^{T} z_t z_t' = \begin{pmatrix} S_{11} & S_{12} \\ S_{21} & S_{22} \end{pmatrix}, \quad S_{ij} = T^{-1}\sum_{t=1}^{T} z_{it} z_{jt}'. \tag{13}$$

The $N \times j$, $j = 1, \ldots, N$, full column rank matrices φ and ς that deliver the canonical variates are those that maximize the function

$$f(\varphi, \varsigma) = \frac{|\varphi' S_{12} \varsigma \varsigma' S_{21} \varphi|}{|\varphi' S_{11} \varphi||\varsigma' S_{22} \varsigma|},$$

see e.g. Johansen (1996) and references therein. The maximal value of $f(\varphi, \varsigma)$ is equal to $\prod_{i=1}^{j} \lambda_i$ and it is attained for

$$\varphi = (\varphi_1, \ldots, \varphi_j), \quad \varsigma = (\varsigma_1, \ldots, \varsigma_j),$$

where $V = (\varphi_1, \ldots, \varphi_N)$ are eigenvectors associated to the eigenvalues $\Lambda = \text{diag}(\lambda_1, \ldots, \lambda_N)$ in the basic eigenvalue problem

$$|\lambda S_{11} - S_{12} S_{22}^{-1} S_{21}| = 0, \tag{14}$$

and $Q = (\varsigma_1, \ldots, \varsigma_N) = S_{22}^{-1} S_{21} V \Lambda^{-1/2}$. By definition, V and Λ satisfy

$$S_{11} V \Lambda = S_{12} S_{22}^{-1} S_{21} V, \quad V' S_{11} V = I, \quad V' S_{12} S_{22}^{-1} S_{21} V = \Lambda. \tag{15}$$

Premultiplying the first equation by V' one has $V' S_{11} V \Lambda = V' S_{12} S_{22}^{-1} S_{21} V$ and because $\Lambda = \text{diag}(\lambda_1, \ldots, \lambda_N)$ and $V = (\varphi_1, \ldots, \varphi_N)$, one finds

$$\lambda_i = \frac{\varphi_i' S_{12} S_{22}^{-1} S_{21} \varphi_i}{\varphi_i' S_{11} \varphi_i}, \quad i = 1, \ldots, N.$$

This shows that the i-th squared canonical correlation λ_i coincides with the R^2 of a regression of the i-th canonical variate $\varphi_i' X_t$ on $\mathcal{S}X_t$.

5. Main results

This section states the main results of the paper in Theorems 5.1, 5.4 and in Corollary 5.5. The first part, see Theorem 5.1, is for $T \to \infty$ and fixed N, s. It shows that the largest s squared canonical correlations $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_s$ converge to the R^2 of the regression of the limit non-stationary part of X_t (driven by s Brownian motions) on the limit of $\mathcal{S}X_t$, which consists of $N - s$ Brownian motions plus the cumulation of the s Brownian motions present in X_t. The $N - s$ smallest squared canonical correlations $\lambda_{s+1} \geq \lambda_{s+2} \geq \cdots \geq \lambda_N$ converge in probability to zero. The second part, see Theorem 5.4, is for $N \to \infty$ in the large-T limits, with s growing at most linearly in N. It finds that the largest s squared canonical correlation converge to 1, i.e. the limit densities of these squared canonical correlation shrink toward the point mass distribution at 1 as $(T, N)_{seq} \to \infty$. Corollary 5.5 combines Theorems 5.1 and 5.4 to show that the gap between consecutive squared canonical correlations $\lambda_j - \lambda_{j+1}$ converges to 0 for $j < s$ and for $j > s$, while it converges to 1 for $j = s$ as $(T, N)_{seq} \to \infty$.

The first result discusses the limit of the squared canonical correlations λ_i, $i = 1, \ldots, N$, as $T \to \infty$. Observe that a functional central limit theorem applies to $\mathcal{S}\eta_t$, namely

$$T^{-\frac{1}{2}} \sum_{t=1}^{\lfloor Tu \rfloor} \eta_t \xrightarrow{w} W(u), \quad \text{var}\, W(1) = \Omega, \tag{16}$$

where $\lfloor \cdot \rfloor$ denotes the integer part, $W(u) = BM(\Omega)$ is a $N_\eta \times 1$ Brownian motion in $u \in [0, 1]$, with covariance Ω, see (2); finally $\xrightarrow{w}$ indicates weak convergence of probability measures in $D[0, 1]$, the space of all right continuous real-valued functions having finite left limits on [0,1] endowed with the sup norm.

Next define the following quantities that depend on $W(u)$ in (16):

$$Y(u) = \kappa' W(u), \quad K(u) = \begin{pmatrix} \int_0^u Y(v) \mathrm{d}v \\ \psi_\perp' B_x(1) W(u) \end{pmatrix}, \quad \ell_{ij} = \int_0^1 i(u) j(u)' \mathrm{d}u, \quad i, j \in \{Y, K\}, \tag{17}$$

where ψ and κ are defined in (7) and $B_x(z)$ is defined in (5); note that $Y(u)$ is of dimension $s \times 1$ and $K(u)$ of dimension $N \times 1$. Observe that the Brownian motions $\kappa' W(u)$ and $\psi_\perp' B_x(1) W(u)$ are in general correlated.

Theorem 5.1. *(Limit of the squared canonical correlations as $T \to \infty$) Let X_t be generated as in (1), (2), (3); then (16) holds, and as $T \to \infty$ one has that:*

(i) *the s largest squared canonical correlations $1 \geq \lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_s$ of (14)* jointly *satisfy*

$$\lambda_i \xrightarrow{w} \mu_i \overset{a.s.}{>} 0, \quad i = 1, \ldots, s, \quad T \to \infty,$$

where μ_i is an eigenvalue of $\ell_{YY}^{-1} \ell_{YK} \ell_{KK}^{-1} \ell_{KY}$ and $1 \geq \mu_1 \geq \mu_2 \geq \cdots \geq \mu_s \overset{a.s.}{>} 0$;

(ii) *the N-s smallest squared canonical correlations $\lambda_{s+1} \geq \lambda_{s+2} \geq \cdots \geq \lambda_N$ of (14) satisfy*

$$\lambda_i \xrightarrow{p} 0, \quad i = s + 1, \ldots, N.$$

Proof. See Appendix A. □

Theorem 5.1 shows that the behavior of the s largest squared canonical correlations differs from that of the remaining $N-s$. In fact the s largest squared canonical correlations $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_s$ converge weakly to the eigenvalues of the $s \times s$ random matrix $\ell_{YY}^{-1}\ell_{YK}\ell_{KK}^{-1}\ell_{KY}$, which is positive definite a.s.. This implies that the i-th squared canonical correlation λ_i, $i = 1, \ldots, s$, is positive a.s. in the limit. On the contrary, the $N-s$ smallest squared canonical correlations $\lambda_{s+1} \geq \lambda_{s+2} \geq \cdots \geq \lambda_N$ converge in probability to zero without any normalization in T.

Observe that the process $\mathcal{S}X_t$ is nonstationary in all directions, see (11), while X_t is nonstationary only in the s directions of ψ, see (9). Then, because $\bar{\psi}'X_t$ is asymptotically correlated with $\mathcal{S}X_t$, the s largest squared canonical correlations $\lambda_1, \ldots, \lambda_s$ tend to positive limits as $T \to \infty$. On the contrary, because $\psi_\perp' X_t$ is asymptotically uncorrelated with $\mathcal{S}X_t$, the $N-s$ smallest squared canonical correlations $\lambda_{s+1}, \ldots, \lambda_N$ converge in probability to 0 as $T \to \infty$.

The second step (for $N \to \infty$) in the sequential asymptotics is derived under the following assumption.

Assumption 5.2. (s grows at most linearly in N) The number s of stochastic trends can be:

(i) a fixed positive integer, or
(ii) $s = \lfloor g(N) \rfloor$ with $1 \leq g(N) = o(N)$ and non-decreasing in N, or
(iii) $s = \lfloor cN \rfloor$ with $0 < c < 1$.

Remark 5.3 Assumption 5.2 implies $N - s \to \infty$ as $N \to \infty$**.** Assumption 5.2 accommodates different behavior for s as $N \to \infty$. In all cases, however, one has that $N - s \to \infty$ as $N \to \infty$. In fact in case (i) of a fixed number s of common trends, the fact that $N - s \to \infty$ follows directly. In case (ii) of $s = \lfloor g(N) \rfloor$, one has that $N - s > N - \lfloor cN \rfloor$ for N large enough and any $0 < c < 1$, because $g(N)/N = o(1)$ by Assumption 5.2.(ii). This implies $N - s \to \infty$. Finally, in case (iii) the number of common trends s is a constant proportion of N, and one has $N - s \approx N(1-c) \to \infty$ because $1 - c > 0$ by Assumption 5.2.(iii).

Under Assumption 5.2, the next result shows that the largest s squared canonical correlations converge to 1 as $(T, N)_{seq} \to \infty$, i.e. the limit densities of these squared canonical correlations shrink toward the point mass distribution at 1.

Theorem 5.4. (*$(T, N)_{seq} \to \infty$ limit of the s largest squared canonical correlations) Under Assumption 5.2, the ordered eigenvalues μ_i, $i = 1, \ldots, s$, of $\ell_{YY}^{-1}\ell_{YK}\ell_{KK}^{-1}\ell_{KY}$, see Theorem 5.1, all converge in probability to 1 as $N \to \infty$.*

Proof. See Appendix A. □

Theorem 5.4 shows that under Assumption 5.2 the s largest squared canonical correlations $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_s$ all converge in probability to 1 as $(T, N)_{seq} \to \infty$. The proof of this result extends the results in Phillips (1998), who showed that any Brownian motion has an L^2 expansion, with random coefficients, in terms of independent standard Brownian motions, and the R^2 of the L^2 regression of an arbitrary Brownian motion on independent standard Brownian motions tends to one as the number of regressors tends to ∞.

Here it is shown that the same result applies to arbitrary Brownian motions with nonsingular variance that are not necessarily independent of the regressand. The proof exploits the fact that μ_i is the R^2 of the L^2 regression of the Brownian motion $c_i'Y(u)$ on the process $K(u)$, where c_i is an eigenvector of $\ell_{YY}^{-1}\ell_{YK}\ell_{KK}^{-1}\ell_{KY}$ associated to μ_i and $K(u)$ contains the $(N-s)$-dimensional Brownian motion $\psi_\perp' B_x(1)W(u)$. Assumption 5.2 ensures that $N - s$ diverges, see Remark 5.3, and this implies that $\mu_i \xrightarrow{p} 1$.

Combining Theorems 5.1 and 5.4, one has that the gap between consecutive squared canonical correlations $\lambda_j - \lambda_{j+1}$ converges to 0 for $j < s$ and for $j > s$, while it converges to 1 for $j = s$ as $(T, N)_{seq} \to \infty$; this leads to the definition of the following estimator of the number of common trends:

$$\widehat{s} := \underset{j \in \{1, \ldots, N-1\}}{\operatorname{argmax}} \, (\lambda_j - \lambda_{j+1}) \tag{18}$$

called the maximal gap estimator in the following. Note that $0 < \widehat{s} < N$ in parallel to the assumption that $0 < s < N$.

Corollary 5.5. (*$(T, N)_{seq} \to \infty$ limit of the maximal gap estimator) Let $\widehat{s}$ be the maximal gap estimator of s defined in (18); then, under Assumption 5.2, $\widehat{s} \xrightarrow{p} s$ when first $T \to \infty$ and then $N \to \infty$.*

6. Simulations

The performance of the maximal gap estimator is investigated by simulation using two different Data Generating Processes, (DGPs): DGP 1, see Section 6.1 below, is inspired by the Monte Carlo study in Bai (2004) and DGP 2, see Section 6.2 below, by the one in Barigozzi and Trapani (2022).

In the simulations, the number of variables is set to $N = 50, 100, 300$, the sample size to $T = (2, 3, 4, 5, 10)N$ and the number of stochastic trends to $s = 1, 2, 3, 4, 5, 10, 20, 30, 40$; the number of Monte Carlo replications is set to 10^4 in every simulation. In the following, a_{ij} indicates the (i, j) element of a matrix A unless stated otherwise.

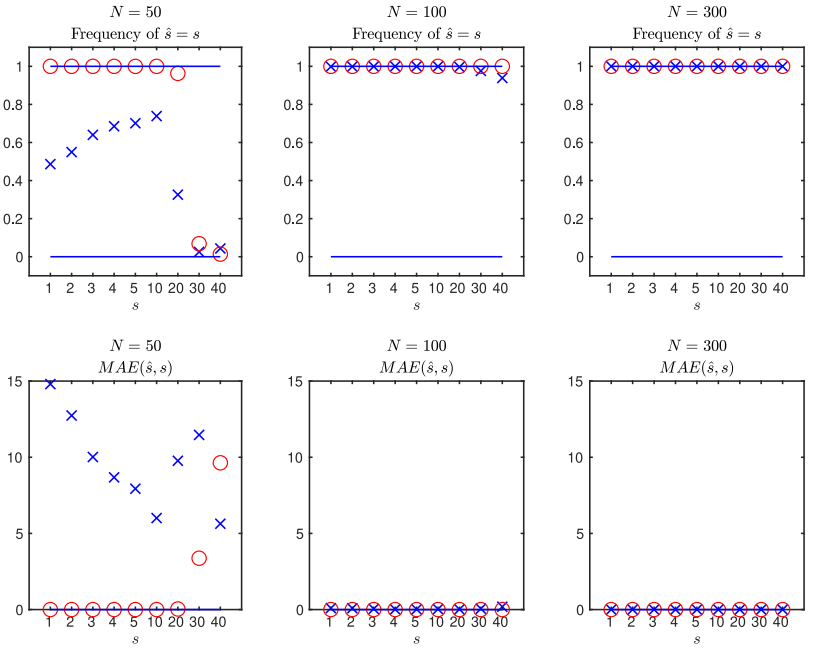

Fig. 1. DGP 1 with $\rho = 0.5$, $T = 3N(\times)$ and $T = 5N(\circ)$. Frequency of $\hat{s} = s$, i.e. $\frac{1}{MC}\sum_{n=1}^{MC} 1_{\hat{s}_n = s}$, and mean absolut error $MAE(\hat{s}, s) = \frac{1}{MC}\sum_{n=1}^{MC} |\hat{s}_n - s|$ over $MC = 10^4$ replications for DGP 1, see Section 6.1, with $\rho = 0.5$.

6.1. DGP 1: Bai (2004)

Following Bai (2004), Section 6, the factor structure of DGP 1 is

$$X_t = \Lambda_1 f_t + \Lambda_2 f_{t-1} + \xi_t,$$

where the $s \times 1$ stochastic trends are generated as $f_t = f_{t-1} + \zeta_t$, $\zeta_t \sim N(0, I_s)$, the $N \times s$ loadings Λ_h, $h = 1, 2$, as $\lambda_{h,ij} \sim N(0, 1)$, independently of ζ_t and the $N \times 1$ idiosyncratic component as

$$\xi_t = A\xi_{t-1} + \varepsilon_t + B\varepsilon_{t-1}, \quad \begin{array}{ll} A = \mathrm{diag}(a_1, \ldots, a_N) & a_i \sim U[-\rho, \rho] \\ B = \mathrm{diag}(b_1, \ldots, b_N), & b_i \sim U[-\rho, \rho] \\ \varepsilon_t \sim N(0, \Omega_\varepsilon) & \Omega_\varepsilon = (0.5^{|i-j|})_{i,j=1,\ldots,N} \end{array}, \quad 0 < \rho < 1,$$

all independent, where $U[a, b]$ indicates the uniform distribution on the interval $[a, b]$. This is an extension of the design considered in Bai (2004). The present setup satisfies (1) - (2): in fact

$$X_t = \Lambda F_t + \xi_t, \quad \Lambda = (\Lambda_1, \Lambda_2), \qquad F_t = \begin{pmatrix} f_t \\ f_{t-1} \end{pmatrix},$$

$B(z) = (I - Az)^{-1}(I + Bz) = \sum_{n=0}^{\infty} B_n z^n$ converges for all $|z| < 1 + \delta$ for $\delta = -1 + \min_{i=1,\ldots,N} 1/|a_i| > 0$ and $B(1)$ is nonsingular.

The complete set of results of the simulations are grouped in Table B.1 and Table B.2 in Appendix B.1 according to the value of ρ: non-autocorrelated idiosyncratic terms ($\rho = 0$) and mildly autocorrelated idiosyncratic terms ($\rho = 0.5$). The overall conclusions are that the percentage of correct detection of the number of stochastic trends, i.e. the frequency of $\hat{s} = s$, is 100% (or very close to it) in most cases when the sample size T is at least 3 times larger than the cross-sectional dimension N and $N - s$ is not too small.

A graphical illustration of the results for the $\rho = 0.5$ case is presented in Figure 1, which reports frequencies of correct estimation of s and the associated mean absolute estimation error.

6.2. DGP 2: Barigozzi and Trapani (2022)

Following Barigozzi and Trapani (2022), Section 5, the factor structure of DGP 2 is

$$X_t = \Lambda F_t + \xi_t, \quad \Lambda = (\Lambda_2, \Lambda_3), \quad F_t := \left(f_t^{(2)'}, f_t^{(3)'}\right)', \quad \xi_t = \sqrt{\theta} u_t, \tag{19}$$

where, following their original notation and with independent draws, one has that:

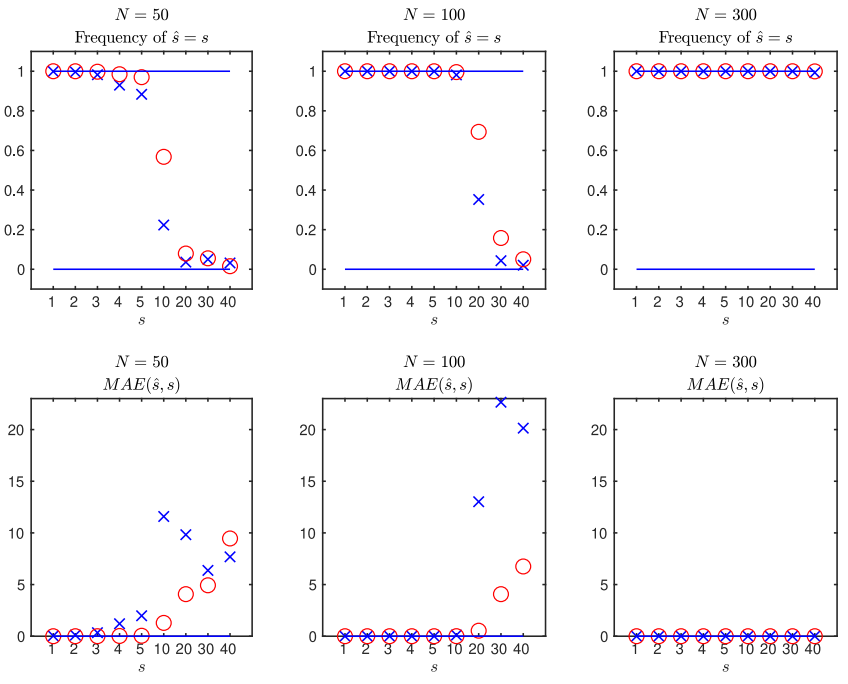

Fig. 2. DGP 2 with $r_3 = 3$, $T = 4N(\times)$ and $T = 5N(\circ)$. Frequency of $\hat{s} = s$, i.e. $\frac{1}{MC}\sum_{n=1}^{MC} 1_{\hat{s}_n = s}$, and mean absolut error $MAE(\hat{s}, s) = \frac{1}{MC}\sum_{n=1}^{MC} |\hat{s}_n - s|$ over $MC = 10^4$ replications for DGP 2, see Section 6.2, with $r_3 = 3$ stationary factors.

(*i*) the $s \times 1$ stochastic trends $f_t^{(2)}$ are generated as

$$f_t^{(2)} = f_{t-1}^{(2)} + e_t^{(2)}, \quad e_t^{(2)} = \mathrm{diag}(\rho_1, ..., \rho_s) e_{t-1}^{(2)} + \varepsilon_t^{(2)}, \quad \begin{matrix} \rho_i \sim U[0.4, 0.8] \\ \varepsilon_t^{(2)} \sim N(0, I_s) \end{matrix}; \tag{20}$$

(*ii*) the $r_3 \times 1$ stationary factors $f_t^{(3)}$ are generated as

$$f_t^{(3)} = \mathrm{diag}(\alpha_1, ..., \alpha_{r_3}) f_{t-1}^{(3)} + \varepsilon_t^{(3)}, \quad \begin{matrix} \alpha_i \sim U[-0.5, 0.5] \\ \varepsilon_t^{(3)} \sim N(0, I_{r_3}) \end{matrix}; \tag{21}$$

(*iii*) the $N \times 1$ innovations u_t in the idiosyncratic component $\xi_t = \sqrt{\theta} u_t$ are generated as

$$u_t = 0.5 u_{t-1} + B\upsilon_t, \quad \upsilon_t \sim N(0, I_N), \tag{22}$$

where B is s.t. $b_{ij} = 0.5^{|j-i|}$ for $|j - i| \le \min(\lfloor N/20 \rfloor, 10)$ and 0 elsewhere.

In the setup of Barigozzi and Trapani (2022), F_t can also include a random walk with drift $f_t^{(1)}$ and B is s.t. $b_{ij} = 0.5$ for $|j - i| \le \min(\lfloor N/20 \rfloor, 10)$, $b_{ii} = 1$ and 0 elsewhere; the present modification of B avoids warnings of numerical singularity of the inverses in the CCA analysis.

X_t in (19) is thus generated as follows: define A as a random matrix of dimension $N \times (s + r_3)$ with elements independently drawn from a $N(0, 1)$ distribution, compute its QR decomposition, $QR = A$ say, and set $\Lambda = \sqrt{N} Q$; this satisfies $\Lambda'\Lambda = NI_N$.

Next generate $f_t^{(j)}$, $j = 2, 3$, as in (20), (21); indicating the resulting realization of the factors as $\widetilde{f}_t^{(j)}$, $j = 2, 3$, one then computes

$$k_2 = \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\left(\lambda_i^{(2)'} \Delta \widetilde{f}_t^{(2)}\right)^2, \quad k_3 = \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\left(\lambda_i^{(3)'} \widetilde{f}_t^{(3)}\right)^2$$

where $\lambda_i^{(j)'}$ is row i of Λ_j, and one rescales the factors as $f_t^{(j)} = \widetilde{f}_t^{(j)} / \sqrt{k_j}$, $j = 2, 3$. Finally one generates $\xi_t = \sqrt{\theta} u_t$, where u_t is as in (22) and sets

$$\theta = 0.5 \cdot \frac{\sum_{i=1}^{N}\sum_{t=1}^{T}\left(\lambda_i^{(2)'} \Delta \widetilde{f}_t^{(2)} + \lambda_i^{(3)'} \Delta \widetilde{f}_t^{(3)}\right)^2}{\sum_{i=1}^{N}\sum_{t=1}^{T}(\Delta u_{it})^2};$$

one then computes X_t from (19).

Table B.3 and Table B.4 in Appendix B.2 respectively report the estimation frequencies in the presence of $r_3 = 0, 3$ stationary factors $f_t^{(3)}$. The overall conclusions are:

(*i*) the results do not change with the number $r_3 = 0, 3$ of stationary factors;
(*ii*) when the cross-sectional dimension is moderate ($N = 50$ and $N = 100$) and the number of stochastic trends is moderate ($s \leq 5$), the percentage of correct detection of the number of stochastic trends s, $\widehat{s} = s$, is 100% (or very close to it) in most cases when the sample size T is at least 4 times larger than the cross-sectional dimension N;
(*iii*) when the cross-sectional dimension is large ($N = 300$) a sample size T is at least 4 times larger than the cross-sectional dimension N generally guarantees a frequency of correct detection of 1 (or very close to it).

Figure 2 reports frequencies of correct estimation of s and the associated mean absolute estimation error, for the $r_3 = 3$ case.

7. Conclusions

This paper discusses the limit behavior for CCA performed on the levels and cumulated levels of a vector time series when the number of common trends s grows at most linearly in N and both the time and the cross sectional dimensions T and N diverge sequentially.

It is found that the largest s squared canonical correlations tend to 1 due to a spurious regression phenomenon while the smallest $N - s$ squared canonical correlations tend to 0; this fact is used to estimate s consistently via the index of the largest gap between squared canonical correlations. Monte Carlo simulations show that the proposed criterion is accurate when the sample size T is at least 3 times larger than the cross-sectional dimension N and for a number of stochastic trends s that guarantees a large $N - s$ dimension.

Acknowledgements

The present version of the paper has benefited from valuable comments by two anonymous referees and the editor. We also thank Marco Lippi for helpful discussions on DFMs and Søren Johansen, Anders Rahbek and seminar participants at the Department of Economics, University of Copenhagen. For the simulations, the authors acknowledge the use of TeraStat2, the general-purpose supercomputing cluster of the Department of Statistical Sciences, Sapienza University of Rome, and the precious support from Prof. Umberto Ferraro Petrillo. Partial financial support by PRIN 2010 is gratefully acknowledged by the first author. The views set out in this paper are those of the authors and do not necessarily reflect the ones of the institutions of affiliation.

Appendix A. Proofs

Proof. *Proof of Theorem 5.1.* Eq. (16) holds by Johansen (1996) Theorem B.12, eq. (B.12), see also Phillips and Solo (1992). Next one can apply results for linear processes related to (16) that enter the CCA, following Johansen (1995) (henceforth J95).

Specifically, define $X_{0,1t} := \psi_\perp' X_t$, $\Delta X_{1,1t} := \Delta\psi_\perp' \mathcal{S} X_t = X_{0,1t}$, $\Delta X_{1,2t} := \Delta\psi' X_t$, $\Delta^2 X_{2,1t} := \Delta^2\psi' \mathcal{S} X_t = X_{1,2t}$, see (9) and (11), and observe that they are linear processes of the type $\Delta^j X_{j,ht} = a_j(L)\eta_t = \sum_{i=0}^{\infty} a_{ji}\eta_{t-i}$, $j = 0, 1, 2$ and $h = 1, 2$, with geometrically decreasing coefficients a_{ji}. Hence, applying (A.4), (A.5) and (A.6) in J95 to $z_{1t} = X_t, z_{2t} = \mathcal{S}X_t$ in (12) and using (8), one finds for $t = 1, \ldots, T$ and $t = \lfloor Tu \rfloor$

$$
\begin{aligned}
T^{-\frac{1}{2}}\bar{\psi}' z_{1t} &= T^{-\frac{1}{2}}\kappa'\mathcal{S}\eta_t + o_p(1) \xrightarrow{w} Y(u) := \kappa' W(u),\\
T^{-\frac{1}{2}}\psi_\perp' z_{2t} &= T^{-\frac{1}{2}}\psi_\perp' B_x(1)\mathcal{S}\eta_t + o_p(1) \xrightarrow{w} K_{\psi_\perp}(u) := \psi_\perp' B_x(1) W(u),\\
T^{-\frac{3}{2}}\bar{\psi}' z_{2t} &= T^{-\frac{3}{2}}\kappa'\mathcal{S}^2\eta_t + o_p(1) \xrightarrow{w} K_{\psi}(u) := \int_0^u Y(v)\mathrm{d}v = \int_0^u \kappa' W(v)\mathrm{d}v.
\end{aligned}
$$

Further note that $\psi_\perp' z_{1t}$ is a stationary process with mean 0 and variance $\Sigma_{\psi_\perp\psi_\perp} := \mathrm{var}(\psi_\perp' z_{1t})$. Observe also that $E(K_{\psi_\perp}(u)Y(u)') = (\psi_\perp' B_x(1)\Omega\kappa)u = ((0, \gamma_\perp)' C_1(1)\omega)u$ which is not necessarily equal to 0; this means that $K_{\psi_\perp}(u)$ and $Y(u)$ are in general correlated.

Next define

$$
K(u) = \begin{pmatrix} K_\psi(u) \\ K_{\psi_\perp}(u) \end{pmatrix}, \qquad A_T = \begin{pmatrix} T^{-\frac{1}{2}}\bar{\psi} & \psi_\perp \end{pmatrix}, \quad B_T = \begin{pmatrix} T^{-\frac{3}{2}}\bar{\psi} & T^{-\frac{1}{2}}\psi_\perp \end{pmatrix}; \tag{A.1}
$$

then

$$
N_{11} = A_T' S_{11} A_T \xrightarrow{w} \begin{pmatrix} \ell_{YY} & 0 \\ 0 & \Sigma_{\psi_\perp\psi_\perp} \end{pmatrix}, \quad N_{22} = B_T' S_{22} B_T \xrightarrow{w} \ell_{KK}, \quad N_{12} = A_T' S_{12} B_T \xrightarrow{w} \begin{pmatrix} \ell_{YK} \\ 0 \end{pmatrix}
$$

by (A.7)-(A.10) in J95.

Normalizing $S(\lambda) = \lambda S_{11} - S_{12}S_{22}^{-1}S_{21}$ in (14) using the matrices A_T and B_T in (A.1), one finds

$$
A_T' S(\lambda) A_T = \lambda N_{11} - N_{12}N_{22}^{-1}N_{21} \xrightarrow{w} \lambda \begin{pmatrix} \ell_{YY} & 0 \\ 0 & \Sigma_{\psi_\perp\psi_\perp} \end{pmatrix} - \begin{pmatrix} \ell_{YK}\ell_{KK}^{-1}\ell_{KY} & 0 \\ 0 & 0 \end{pmatrix}; \tag{A.2}
$$

this shows that the s largest λ eigenvalues in (14) converge weakly to the ones of $\ell_{YY}^{-1}\ell_{YK}\ell_{KK}^{-1}\ell_{KY}$ and the smallest $N-s$ eigenvalues λ converge to 0 in probability. □

Proof. *Proof of Theorem 5.4.* From Theorem 5.1 one has that $\lambda_i \xrightarrow{w} \mu_i$ as $T \to \infty$, $i = 1, \ldots, s$, eigenvalue of $\ell_{YY}^{-1}\ell_{YK}\ell_{KK}^{-1}\ell_{KY}$; letting c_i be a corresponding eigenvector, one has $\ell_{YY}c_i\mu_i = \ell_{YK}\ell_{KK}^{-1}\ell_{KY}c_i$ and hence

$$\mu_i = \frac{c_i'\ell_{YK}\ell_{KK}^{-1}\ell_{KY}c_i}{c_i'\ell_{YY}c_i},$$

i.e. μ_i is the R^2 of a $L^2[0,1]$ regression of the Brownian motion $c_i'Y(u) = c_i'\kappa' W(u)$ on $K(u) = (K_\psi(u)', K_{\psi_\perp}(u)')'$, composed by the s-dimensional integrated Brownian motion $K_\psi(u) = \int_0^u Y(v)\mathrm{d}v$ and the $(N-s)$-dimensional Brownian motion $K_{\psi_\perp}(u) = \psi_\perp' B_x(1)W(u)$. Hence, μ_i is bounded below by the R^2 of the regression of $c_i'Y(u)$ on $K_{\psi_\perp}(u)$ alone. It is shown next that the latter R^2 approaches 1 in probability, as $N-s \to \infty$, and hence, so does μ_i.

The underlying fact is Theorem 4.3 of Phillips (1998), asserting that an arbitrary Brownian motion on [0,1] has an $L^2[0,1]$ expansion, with random coefficients, in terms of independent standard Brownian motions defined on an augmented probability space. Specifically, given a standard Brownian motion $W_0(u)$, there exist independent standard Brownian motions $\widehat{W}_k(u)$, $k = 1, 2, \ldots$, independent of $W_0(u)$ and defined on a product extension of the probability space where $W_0(u)$ is defined, and r.v.'s $\widehat{d}_k$ such that, for every $\varepsilon > 0$,

$$\Pr\left(\int_0^1 \left[W_0(u) - \sum_{k=1}^{N-s} \widehat{d}_k \widehat{W}_k(u)\right]^2 \mathrm{d}u < \varepsilon\right) \to 1$$

as $N-s \to \infty$. The previous probability is fully determined by the joint distribution of $W_0(u), \widehat{W}_k(u), \widehat{d}_k$, $k = 1, \ldots, N-s$, whatever the structure of the supporting probability space. Therefore, also

$$\Pr\left(\int_0^1 \left[W_0(u) - \sum_{k=1}^{N-s} d_k W_k(u)\right]^2 \mathrm{d}u < \varepsilon\right) \to 1 \tag{A.3}$$

as $N-s \to \infty$ for

$$W_0(u) := \sigma_{ii.\psi_\perp}^{-1/2}\{c_i'Y(u) - \Sigma_{i\psi_\perp}\Sigma_{\psi_\perp\psi_\perp}^{-1}K_{\psi_\perp}(u)\}$$
$$\left(W_1(u), \ldots, W_{N-s}(u)\right)' := \Sigma_{\psi_\perp\psi_\perp}^{-1/2}K_{\psi_\perp}(u)$$

with

$$\Sigma_{\psi_\perp\psi_\perp} = \psi_\perp' B_x(1)\Omega B_x(1)'\psi_\perp, \qquad \Sigma_{i\psi_\perp} = c_i'\bar{\psi}'C_x\Omega B_x(1)'\psi_\perp,$$
$$\sigma_{ii} = c_i'\bar{\psi}'C_x\Omega C_x'\bar{\psi}c_i, \qquad \sigma_{ii.\psi_\perp} = \sigma_{ii} - \Sigma_{i\psi_\perp}\Sigma_{\psi_\perp\psi_\perp}^{-1}\Sigma_{\psi_\perp i}$$

and d_k (possibly defined on an extension of the original probability space) such that $W_0(u), W_k(u), d_k$ and $W_0(u), \widehat{W}_k(u), \widehat{d}_k$, $k = 1, \ldots, N-s$, have the same joint distribution. Then (A.3) can be reformulated as

$$P\left(\int_0^1 \left[c_i'Y(u) - a_i'K_{\psi_\perp}(u)\right]^2 \mathrm{d}u < \sigma_{ii.\psi_\perp}\varepsilon\right) \to 1$$

as $N-s \to \infty$, with $a_i' = \sigma_{ii.\psi_\perp}^{1/2} d'\Sigma_{\psi_\perp\psi_\perp}^{-1/2} + \Sigma_{i\psi_\perp}\Sigma_{\psi_\perp\psi_\perp}^{-1}$, $d' = (d_1, \ldots, d_{N-s})$. As the R^2 of the regression of $c_i'Y(u)$ on $K_{\psi_\perp}(u)$ is invariant to the normalization of c_i, without loss of generality, one can choose c_i such that $\sigma_{ii} = 1$; then $\int_0^1 [c_i'Y(u)]^2\mathrm{d}u$ is bounded away from zero in probability, as $N-s \to \infty$, and $\sigma_{ii.\psi_\perp} \le \sigma_{ii} = 1$ for all N. It follows that

$$\frac{\int_0^1 \left[c_i'Y(u) - a_i'K_{\psi_\perp}(u)\right]^2 \mathrm{d}u}{\int_0^1 [c_i'Y(u)]^2\mathrm{d}u} \to 0$$

in probability, as $N-s \to \infty$. One less the previous fraction is a lower bound for the R^2 of the regression of $c_i'Y(u)$ on $K_{\psi_\perp}(u)$, implying that this R^2 approaches one in probability, as $N-s \to \infty$. Hence the convergence of μ_i. □

Appendix B. Tables and figures

B1. DGP 1: Bai (2004)

Table B1
Frequency of $\hat{s} = m$, i.e. $\frac{1}{MC}\sum_{n=1}^{MC} 1_{\hat{s}_n=m}$ over $MC = 10^4$ replications for DGP 1, see Section 6.1, with non-autocorrelated idiosyncratic terms ($\rho = 0$).

$\rho = 0$		$N = 50$					$N = 100$					$N = 300$				
		T					T					T				
s	$\hat{s}$	100	150	200	250	500	200	300	400	500	1000	600	900	1200	1500	3000
1	1	0	0.9945	1	1	1	0	1	1	1	1	0	1	1	1	1
	2	0	0.0021	0	0	0	0	0	0	0	0	0	0	0	0	0
	3	0	0.0001	0	0	0	0	0	0	0	0	0	0	0	0	0
	4	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	5	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
2	2	0	0.9935	1	1	1	0	1	1	1	1	0	1	1	1	1
	3	0	0.0026	0	0	0	0	0	0	0	0	0	0	0	0	0
	4	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	5	0	0.0001	0	0	0	0	0	0	0	0	0	0	0	0	0
	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
3	3	0	0.9949	1	1	1	0	1	1	1	1	0	1	1	1	1
	4	0	0.0024	0	0	0	0	0	0	0	0	0	0	0	0	0
	5	0	0.0003	0	0	0	0	0	0	0	0	0	0	0	0	0
	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
4	4	0	0.9943	1	1	1	0	1	1	1	1	0	1	1	1	1
	5	0	0.0022	0	0	0	0	0	0	0	0	0	0	0	0	0
	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	4	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
5	5	0	0.9935	1	1	1	0	1	1	1	1	0	1	1	1	1
	8	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	9	0	0.0042	0.0001	0	0	0	0	0	0	0	0	0	0	0	0
10	10	0	0.9726	0.9999	1	1	0	1	1	1	1	0	1	1	1	1
	11	0	0.0091	0	0	0	0	0	0	0	0	0	0	0	0	0
	12	0	0.0008	0	0	0	0	0	0	0	0	0	0	0	0	0
	18	0	0.0017	0.0003	0	0	0	0	0	0	0	0	0	0	0	0
	19	0	0.076	0.0139	0.0018	0	0	0.0002	0	0	0	0	0	0	0	0
20	20	0	0.7188	0.9832	0.9982	1	0	0.9998	1	1	1	0	1	1	1	1
	21	0	0.0502	0.0025	0	0	0	0	0	0	0	0	0	0	0	0
	22	0	0.0145	0.0001	0	0	0	0	0	0	0	0	0	0	0	0
	28	0	0.0233	0.11	0.1725	0.1626	0	0	0	0	0	0	0	0	0	0
	29	0.0002	0.0252	0.0819	0.1272	0.2308	0	0.0056	0	0	0	0	0	0	0	0
30	30	0.0004	0.027	0.0622	0.0727	0.2897	0	0.9942	1	1	1	0	1	1	1	1
	31	0.0007	0.0243	0.0413	0.0288	0.0005	0	0.0002	0	0	0	0	0	0	0	0
	32	0.0014	0.0265	0.0286	0.0112	0	0	0	0	0	0	0	0	0	0	0
	38	0.0103	0.0394	0.0359	0.0217	0.0024	0	0	0	0	0	0	0	0	0	0
	39	0.0149	0.0404	0.0336	0.0181	0.0005	0	0.0334	0.0004	0	0	0	0	0	0	0
40	40	0.0205	0.0404	0.0304	0.0144	0.0004	0	0.9621	0.9996	1	1	0	1	1	1	1
	41	0.0274	0.0466	0.0304	0.0119	0	0	0.0035	0	0	0	0	0	0	0	0
	42	0.0357	0.0484	0.0263	0.0106	0	0	0.0001	0	0	0	0	0	0	0	0

Table B2
Frequency of $\hat{s} = m$, i.e. $\frac{1}{MC}\sum_{n=1}^{MC} 1_{\hat{s}_n=m}$ over $MC = 10^4$ replications for DGP 1, see Section 6.1, with midly autocorrelated idiosyncratic terms ($\rho = 0.5$).

$\rho = 0.5$		$N = 50$					$N = 100$					$N = 300$				
		T					T					T				
s	$\widehat{s}$	100	150	200	250	500	200	300	400	500	1000	600	900	1200	1500	3000
1	1	0	0.4863	0.9999	1	1	0	0.9983	1	1	1	0	1	1	1	1
	2	0	0.0003	0	0	0	0	0	0	0	0	0	0	0	0	0
	3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	4	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	5	0	0.0002	0.0001	0	0	0	0	0	0	0	0	0	0	0	0
	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
2	2	0	0.5495	0.9999	1	1	0	0.9986	1	1	1	0	1	1	1	1
	3	0	0.0007	0.0001	0	0	0	0	0	0	0	0	0	0	0	0
	4	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	5	0	0.0002	0	0	0	0	0	0	0	0	0	0	0	0	0
	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
3	3	0	0.6399	0.9997	1	1	0	0.9992	1	1	1	0	1	1	1	1
	4	0	0.0007	0.0002	0	0	0	0	0	0	0	0	0	0	0	0
	5	0	0.0002	0	0	0	0	0	0	0	0	0	0	0	0	0
	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
4	4	0	0.6853	0.9999	1	1	0	0.9997	1	1	1	0	1	1	1	1
	5	0	0.0012	0.0001	0	0	0	0	0	0	0	0	0	0	0	0
	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	4	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
5	5	0	0.7011	0.9997	0.9998	1	0	0.9997	1	1	1	0	1	1	1	1
	8	0	0.0001	0	0	0	0	0	0	0	0	0	0	0	0	0
	9	0	0.0033	0.0013	0.0001	0	0	0	0	0	0	0	0	0	0	0
10	10	0	0.7387	0.997	0.9999	1	0	0.9994	1	1	1	0	1	1	1	1
	11	0	0.0121	0.0016	0	0	0	0.0001	0	0	0	0	0	0	0	0
	12	0	0.0029	0	0	0	0	0	0	0	0	0	0	0	0	0
	18	0	0.0097	0.0069	0.0011	0	0	0	0	0	0	0	0	0	0	0
	19	0	0.0736	0.0851	0.0326	0.0018	0	0.0025	0	0	0	0	0	0	0	0
20	20	0	0.3259	0.8757	0.9628	0.9982	0	0.9965	1	1	1	0	1	1	1	1
	21	0	0.0633	0.0256	0.0033	0	0	0.0008	0	0	0	0	0	0	0	0
	22	0	0.0278	0.0032	0.0002	0	0	0	0	0	0	0	0	0	0	0
	28	0.0001	0.0218	0.1066	0.1673	0.1745	0	0	0	0	0	0	0	0	0	0
	29	0.0002	0.0252	0.0841	0.129	0.2343	0	0.0213	0	0	0	0	0	0	0	0
30	30	0.0003	0.0253	0.0624	0.0685	0.2715	0	0.9758	1	1	1	0	1	1	1	1
	31	0.0009	0.0289	0.042	0.0294	0.0009	0	0.0024	0	0	0	0	0	0	0	0
	32	0.0015	0.0238	0.0332	0.0141	0	0	0	0	0	0	0	0	0	0	0
	38	0.0131	0.0365	0.0359	0.0231	0.0024	0	0.0001	0	0	0	0	0	0	0	0
	39	0.0168	0.0401	0.0338	0.0183	0.0022	0	0.0515	0.001	0	0	0	0	0	0	0
40	40	0.0199	0.0434	0.0335	0.0137	0.0001	0	0.9388	0.999	1	1	0	1	1	1	1
	41	0.0249	0.0459	0.0318	0.0136	0.0003	0	0.0065	0	0	0	0	0	0	0	0
	42	0.0332	0.0482	0.0286	0.0127	0	0	0.0003	0	0	0	0	0	0	0	0

B2. DGP 2: Barigozzi and Trapani (2022)

Table B3
Frequency of $\hat{s} = m$, i.e. $\frac{1}{MC}\sum_{n=1}^{MC} 1_{\hat{s}_n=m}$ over $MC = 10^4$ replications for DGP 2, see Section 6.2, with $r_3 = 0$ stationary factors.

$r_3 = 0$		$N = 50$					$N = 100$					$N = 300$				
		T					T					T				
s	$\hat{s}$	100	150	200	250	500	200	300	400	500	1000	600	900	1200	1500	3000
1	1	0	0.0218	0.9974	1	1	0	0.736	1	1	1	0	1	1	1	1
	2	0	0	0.0001	0	0	0	0	0	0	0	0	0	0	0	0
	3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	4	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	5	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	1	0	0	0.0011	0.0006	0	0	0	0	0	0	0	0	0	0	0
2	2	0	0.0249	0.9919	0.9994	1	0	0.7088	1	1	1	0	1	1	1	1
	3	0	0	0.0004	0	0	0	0	0	0	0	0	0	0	0	0
	4	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	5	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	2	0	0	0.004	0.0024	0.0002	0	0	0	0	0	0	0	0	0	0
3	3	0	0.0178	0.9751	0.9973	0.9998	0	0.7318	1	1	1	0	0.9999	1	1	1
	4	0	0	0.0015	0.0002	0	0	0	0	0	0	0	0	0	0	0
	5	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	2	0	0	0	0.0001	0	0	0	0	0	0	0	0	0	0	0
	3	0	0	0.0098	0.0057	0.0004	0	0	0	0	0	0	0	0	0	0
4	4	0	0.0219	0.958	0.9937	0.9996	0	0.6381	1	1	1	0	1	1	1	1
	5	0	0	0.0016	0.0004	0	0	0	0	0	0	0	0	0	0	0
	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	3	0	0	0	0.0001	0	0	0	0	0	0	0	0	0	0	0
	4	0	0	0.0129	0.0111	0.0028	0	0.0001	0.0004	0	0	0	0	0	0	0
5	5	0	0.0232	0.9345	0.9878	0.9972	0	0.4156	0.9996	1	1	0	0.9999	1	1	1
	8	0	0	0.0146	0.0284	0.0057	0	0	0.0002	0	0	0	0	0	0	0
	9	0	0.0004	0.1221	0.1682	0.0668	0	0.0008	0.0133	0.0037	0.0002	0	0	0	0	0
10	10	0	0.0039	0.4562	0.7398	0.9265	0	0.1247	0.9846	0.9963	0.9998	0	0.9988	1	1	1
	11	0	0.0001	0.0303	0.0238	0.0002	0	0	0.0003	0	0	0	0	0	0	0
	12	0	0.0001	0.0093	0.0056	0.0001	0	0	0	0	0	0	0	0	0	0
	18	0	0.0024	0.0801	0.1557	0.151	0	0.0003	0.0366	0.0201	0.0006	0	0	0	0	0
	19	0	0.0031	0.0935	0.2053	0.2661	0	0.0019	0.1717	0.1184	0.015	0	0.0001	0	0	0
20	20	0	0.0037	0.0806	0.1664	0.3991	0	0.007	0.6277	0.8463	0.9844	0	0.9877	1	1	1
	21	0	0.0043	0.0539	0.0707	0.0172	0	0.0004	0.0265	0.0089	0	0	0	0	0	0
	22	0	0.005	0.0429	0.0439	0.002	0	0	0.0056	0.0006	0	0	0	0	0	0
	28	0	0.014	0.063	0.0882	0.1108	0	0	0.0786	0.1496	0.0514	0	0.0001	0	0	0
	29	0	0.0162	0.0651	0.0762	0.0792	0	0.0002	0.0986	0.2293	0.1494	0	0.0046	0.0001	0	0
30	30	0.0004	0.019	0.0616	0.0655	0.0388	0	0.0003	0.0745	0.2417	0.7599	0	0.89	0.9999	1	1
	31	0.0001	0.0207	0.0602	0.0549	0.0119	0	0.0001	0.0353	0.0537	0.0025	0	0	0	0	0
	32	0.0001	0.0227	0.0541	0.0418	0.004	0	0.0004	0.0178	0.0206	0.0001	0	0	0	0	0
	38	0.0126	0.0404	0.0508	0.0382	0.0146	0	0.0006	0.0409	0.1071	0.1449	0	0.0008	0.0002	0	0
	39	0.015	0.0461	0.0432	0.0287	0.0052	0	0.0009	0.0427	0.0962	0.2176	0	0.0135	0.008	0.0004	0
40	40	0.0197	0.0463	0.0354	0.0194	0.0017	0	0.0008	0.0326	0.0782	0.3049	0	0.221	0.9916	0.9996	1
	41	0.0252	0.0463	0.0305	0.0138	0.0006	0	0.0004	0.0287	0.0491	0.0199	0	0.0004	0.0001	0	0
	42	0.0348	0.0526	0.0258	0.01	0.0001	0	0.0009	0.0218	0.0345	0.0034	0	0	0	0	0

Table B4
Frequency of $\hat{s} = m$, i.e. $\frac{1}{MC}\sum_{n=1}^{MC} 1_{\hat{s}_n=m}$ over $MC = 10^4$ replications for DGP 2, see Section 6.2, with $r_3 = 3$ stationary factors.

$r_3 = 3$		$N = 50$					$N = 100$					$N = 300$				
		T					T					T				
s	$\widehat{s}$	100	150	200	250	500	200	300	400	500	1000	600	900	1200	1500	3000
1	1	0	0.0377	0.9988	1	1	0	0.7911	1	1	1	0	1	1	1	1
	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	4	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	5	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	1	0	0	0.0001	0.0001	0.0001	0	0	0	0	0	0	0	0	0	0
2	2	0	0.0352	0.9964	0.9999	0.9999	0	0.7125	1	1	1	0	1	1	1	1
	3	0	0	0.0001	0	0	0	0	0	0	0	0	0	0	0	0
	4	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	5	0	0	0.0001	0	0	0	0	0	0	0	0	0	0	0	0
	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	2	0	0	0.004	0.0023	0.0003	0	0	0.0001	0	0	0	0	0	0	0
3	3	0	0.0286	0.9824	0.9976	0.9997	0	0.6776	0.9999	1	1	0	1	1	1	1
	4	0	0	0.0011	0.0001	0	0	0	0	0	0	0	0	0	0	0
	5	0	0	0.0001	0	0	0	0	0	0	0	0	0	0	0	0
	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	2	0	0	0	0	0.0001	0	0	0	0	0	0	0	0	0	0
	3	0	0	0.0201	0.014	0.0016	0	0	0	0	0	0	0	0	0	0
4	4	0	0.0182	0.9297	0.9848	0.9983	0	0.6669	1	1	1	0	0.9999	1	1	1
	5	0	0	0.0038	0.0008	0	0	0	0	0	0	0	0	0	0	0
	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	3	0	0	0	0.0006	0	0	0	0	0	0	0	0	0	0	0
	4	0	0.0001	0.0322	0.0256	0.0054	0	0	0.0007	0.0001	0	0	0	0	0	0
5	5	0	0.0142	0.8834	0.97	0.9946	0	0.4638	0.9993	0.9999	1	0	0.9999	1	1	1
	8	0	0.0001	0.0439	0.0702	0.0121	0	0	0.0006	0	0	0	0	0	0	0
	9	0	0.0003	0.1438	0.2418	0.0923	0	0.0015	0.0162	0.0042	0	0	0	0	0	0
10	10	0	0.0011	0.2234	0.5678	0.8932	0	0.1357	0.9815	0.9957	1	0	0.9994	1	1	1
	11	0	0.0001	0.0251	0.0349	0.0006	0	0	0.0002	0.0001	0	0	0	0	0	0
	12	0	0.0006	0.0117	0.0099	0	0	0	0	0	0	0	0	0	0	0
	18	0	0.0016	0.0506	0.1294	0.1839	0	0.0003	0.0784	0.0599	0.0014	0	0	0	0	0
	19	0	0.0019	0.0453	0.1149	0.2518	0	0.0004	0.1912	0.1952	0.0246	0	0.0007	0	0	0
20	20	0	0.0022	0.0362	0.0794	0.2455	0	0.0003	0.3523	0.6939	0.9733	0	0.986	1	1	1
	21	0	0.0035	0.0342	0.0569	0.0177	0	0.0001	0.0326	0.0184	0	0	0	0	0	0
	22	0	0.0043	0.0366	0.0463	0.0041	0	0.0001	0.009	0.0032	0	0	0	0	0	0
	28	0	0.0113	0.0492	0.0625	0.0785	0	0.0002	0.0587	0.1566	0.0615	0	0.0003	0	0	0
	29	0.0002	0.016	0.0478	0.0613	0.0462	0	0.0001	0.0547	0.1721	0.1706	0	0.0126	0.0003	0	0
30	30	0.0004	0.0204	0.0502	0.0553	0.0226	0	0	0.0433	0.1579	0.7138	0	0.7199	0.9997	1	1
	31	0.001	0.0226	0.0508	0.0494	0.0101	0	0	0.0239	0.0508	0.0043	0	0.0001	0	0	0
	32	0.0009	0.0252	0.0501	0.0482	0.0052	0	0	0.0156	0.0229	0.0002	0	0	0	0	0
	38	0.0105	0.042	0.0396	0.0278	0.0057	0	0.0004	0.0257	0.0778	0.15	0	0.0011	0.0002	0	0
	39	0.0169	0.0398	0.0325	0.0194	0.0021	0	0.0003	0.0235	0.0709	0.2013	0	0.0146	0.0066	0.0008	0
40	40	0.0225	0.042	0.0319	0.0154	0.0005	0	0.0001	0.0204	0.0504	0.226	0	0.2152	0.993	0.9992	1
	41	0.0295	0.0436	0.0265	0.0108	0.0001	0	0.0005	0.0209	0.0398	0.0226	0	0.0003	0.0002	0	0
	42	0.034	0.0446	0.0229	0.0068	0	0	0.0004	0.0183	0.0312	0.0046	0	0	0	0	0

References

Abadir, K., Hadri, K., Tzavalis, E., 1999. The Influence of VAR Dimensions on Estimator Biases. Econometrica 67, 163–181.
Ahn, S., Horenstein, A., 2013. Eigenvalue Ratio Test for the Number of Factors. Econometrica 81, 1203–1227.
Ahn, S., Reinsel, C., 1990. Estimation for partially nonstationary multivariate autoregressive models. Journal of the American Statistical Association 85, 813–823.
Anderson, B., Deistler, M., Lippi, M., 2022. Linear System Challenges of Dynamic Factor Models. Econometrics 10 (4), 35.
Bai, J., 2003. Inferential Theory for Factor Models of Large Dimensions. Econometrica 71, 135–171.
Bai, J., 2004. Estimating cross-section common stochastic trends in nonstationary panel data. Journal of Econometrics 122, 137–183.
Bai, J., Ng, S., 2002. Determining the Number of Factors in Approximate Factor Models. Econometrica 70, 191–221.
Bai, J., Ng, S., 2007. Determining the Number of Primitive Shocks in Factor Models. Journal of Business & Economic Statistics 25, 52–60.
Barigozzi, M., Cavaliere, G., Trapani, L., 2022. Inference in Heavy-Tailed Nonstationary Multivariate Time Series. Journal of the American Statistical Association. forthcoming
Barigozzi, M., Lippi, M., Luciani, M., 2020. Cointegration and Error Correction Mechanisms for Singular Stochastic Vectors. Econometrics 8.

Barigozzi, M., Lippi, M., Luciani, M., 2021. Large-dimensional Dynamic Factor Models: Estimation of Impulse-Response Functions with I(1) cointegrated factors. Journal of Econometrics 221, 455–482.
Barigozzi, M., Trapani, L., 2022. Testing for common trends in nonstationary large datasets. Journal of Business & Economic Statistics 40, 1107–1122.
Breitung, J. (1998). Canonical correlation statistics for testing the cointegration rank in a reversed order. SFB 373 Discussion Papers n.105, Humboldt University of Berlin.
Breitung, J., 2002. Nonparametric tests for unit roots and cointegration. Journal of Econometrics 108, 343–363.
Bykhovskaya, A., Gorin, V., 2022. Cointegration in large VARs. The Annals of Statistics 50, 1593–1617.
Engle, R., Granger, C., 1987. Co-integration and Error Correction: Representation, Estimation, and Testing. Econometrica 55, 251–276.
Forni, M., Hallin, M., Lippi, M., Reichlin, L., 2000. The Generalized Dynamic Factor Model: Identification and estimation. Review of Economics and Statistics 82, 540–554.
Forni, M., Hallin, M., Lippi, M., Reichlin, L., 2004. The generalized dynamic factor model consistency and rates. Journal of Econometrics 119, 231–255.
Forni, M., Hallin, M., Lippi, M., Reichlin, L., 2005. The Generalized Dynamic Factor Model: One-sided estimation and forecasting. Journal of the American Statistical Association 100, 830–840.
Forni, M., Lippi, M., 2001. The Generalized Dynamic Factor Model: Representation theory. Econometric Theory 17, 1113–1141.
Franchi, M., Paruolo, P., 2019. A general inversion theorem for cointegration. Econometric Reviews 38, 1176–1201.
Geweke, J., 1977. The Dynamic Factor Analysis of Economic Time Series. In: Aigner, D., Goldberger, A. (Eds.), Latent variables in socio-economic models 1. North-Holland
Gonzalo, J., Pitarakis, J., 1999. Dimensionality Effect in Cointegration Analysis. In: Engle, R., White, H. (Eds.), Cointegration, causality and forecasting: Festschrift in honour of clive granger. University Press, Oxford, pp. 212–229.
Gregoir, S., 1999. Multivariate time series with various hidden unit roots, Part I. Econometric Theory 15, 435–468.
Harris, D., 1997. Principal Components Analysis of Cointegrated Time Series. Econometric Theory 13, 529–557.
Johansen, S., 1988. Statistical Analysis of Cointegration Vectors. Journal of Economic Dynamics and Control 12, 231–254.
Johansen, S., 1991. Estimation and Hypothesis Testing of Cointegration Vectors in Gaussian Vector Autoregressive Models. Econometrica 59, 1551–1580.
Johansen, S., 1995. A stastistical analysis of cointegration for I(2) variables. Econometric Theory 11, 25–59.
Johansen, S., 1996. Likelihood-based Inference in Cointegrated Vector Auto-Regressive Models. Oxford University Press.
Johansen, S., 2008. A representation theory for a class of vector autoregressive models for fractional processes. Econometric Theory 24, 651–676.
Johansen, S., 2012. The Analysis of Nonstationary Time Series Using Regression, Correlation and Cointegration. Contemporary Economics 6 (2), 40–57.
Liang, C., Schienle, M., 2019. Determination of vector error correction models in high dimensions. Journal of Econometrics 208, 418–441.
Lippi, M., Deistler, M., Anderson, B., 2023. High-Dimensional Dynamic Factor Models: A Selective Survey and Lines of Future Research. Econometrics and Statistics 26, 3–16.
Nielsen, M., Seo, W., Seong, D., 2022. Inference on the dimension of the nonstationary subspace in functional time series. Econometric Theory. Forthcoming
Nyblom, J., Harvey, A., 2000. Tests of common stochastic trends. Econometric Theory 16, 176–199.
Onatski, A., 2010. Determining the number of factors from empirical distribution of eigenvalues. The Review of Economics and Statistics 92, 1004–1016.
Onatski, A., Wang, C., 2018. Alternative Asymptotics for Cointegration Tests in Large VARs. Econometrica 86, 1465–1478.
Onatski, A., Wang, C., 2019. Extreme canonical correlations and high-dimensional cointegration analysis. Journal of Econometrics 212, 307–322.
Phillips, P., 1991. Optimal Inference in Cointegrated Systems. Econometrica 59, 283–306.
Phillips, P.C.B., 1998. New Tools for Understanding Spurious Regressions. Econometrica 66, 1299–1326.
Phillips, P.C.B., Solo, V., 1992. Asymptotics for linear processes. The Annals of Statistics 20, 971–1001.
Sargent, T., Sims, C., 1977. Business cycle modeling without pretending to have too much a priori economic theory. In: New methods in business cycle research, pp. 45–109. Federal Reserve Bank of Minneapolis
Stock, J., Watson, M., 1988. Testing for common trends. Journal of the American Statistical Society 83, 1097–1107.
Stock, J., Watson, M., 1993. A simple estimator of cointegrating vectors in higher order integrated systems. Econometrica 61, 783–820.
Stock, J., Watson, M., 2010. Dynamic Factor Models. In: Clements, M., Henry, D. (Eds.), Oxford handbook of economic forecasting. Oxford University Press.
Trapani, L., 2018. A Randomized Sequential Procedure to Determine the Number of Factors. Journal of the American Statistical Association 113, 1341–1349.
Zhang, R., Robinson, P., Yao, Q., 2019. Identifying cointegration by eigenanalysis. Journal of the American Statistical Association 114, 916–927.