



UvA-DARE (Digital Academic Repository)

Rank based cointegration testing for dynamic panels with fixed T

Juodis, A.

DOI

[10.1007/s00181-017-1304-8](https://doi.org/10.1007/s00181-017-1304-8)

Publication date

2018

Document Version

Final published version

Published in

Empirical Economics

License

CC BY

[Link to publication](#)

Citation for published version (APA):

Juodis, A. (2018). Rank based cointegration testing for dynamic panels with fixed T. *Empirical Economics*, 55(2), 349-389. <https://doi.org/10.1007/s00181-017-1304-8>

General rights

It is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), other than for strictly personal, individual use, unless the work is under an open content license (like Creative Commons).

Disclaimer/Complaints regulations

If you believe that digital publication of certain material infringes any of your rights or (privacy) interests, please let the Library know, stating your reasons. In case of a legitimate complaint, the Library will make the material inaccessible and/or remove it from the website. Please Ask the Library: <https://uba.uva.nl/en/contact>, or a letter to: Library of the University of Amsterdam, Secretariat, Singel 425, 1012 WP Amsterdam, The Netherlands. You will be contacted as soon as possible.

Rank based cointegration testing for dynamic panels with fixed T

Artūras Juodis¹

Received: 10 November 2015 / Accepted: 11 April 2017 / Published online: 7 July 2017
© The Author(s) 2017. This article is an open access publication

Abstract In this paper, we show that the cointegration testing procedure of Binder et al. (Econ Theory 21:795–837, 2005) for Panel Vector Autoregressive model of order 1, PVAR(1) is not valid due to the singularity of the hessian matrix. As an alternative we propose a method of moments based procedure using the rank test of Kleibergen and Paap (J Econom 133:97–126, 2006) for a fixed number of time series observations. The test is shown to be applicable in situations with time-series heteroscedasticity and unbalanced data. The novelty of our approach is that in the construction of the test we exploit the “weakness” of the Anderson and Hsiao (J Econom 18:47–82, 1982) moment conditions. The finite-sample performance of the proposed test statistic is investigated using simulated data. The results indicate that for most scenarios the method has good statistical properties. The proposed test provides little statistical evidence of cointegration in the employment data of Alonso-Borrego and Arellano (J Bus Econ Stat 17:36–49, 1999).

Keywords Dynamic panel data · Panel VAR · Cointegration · Fixed T consistency

1 Introduction

In this paper, we consider the cointegration testing problem for Panel VAR model of order 1 with a fixed time dimension. Up to date the only testing approach in this

Financial support from the NWO MaGW grant “Likelihood-based inference in dynamic panel data models with endogenous covariates” is gratefully acknowledged.

✉ Artūras Juodis
a.juodis@rug.nl

¹ Department of Economics, Econometrics and Finance, Faculty of Economics and Business, University of Groningen, Nettelbosje 2, 9747 AE Groningen, The Netherlands

case is the likelihood ratio test based on the transformed maximum likelihood (TML) estimator of Binder et al. (2005) [hereafter BHP]. However, in the univariate setup it is known that for data with autoregressive parameter close to unity, the likelihood approach does not have a gaussian asymptotic limit, see e.g. Krueger (2013). We extend that result to multivariate setting and argue that the cointegration testing procedure of Binder et al. (2005) is not valid due to the singularity of the corresponding expected hessian matrix.

To the best of our knowledge, in the fixed T dynamic panel data (DPD) literature no feasible method of moments (or least-squares) alternative to likelihood based cointegration testing procedures is available. The main reason for the absence of method of moments based alternatives is that jacobian matrix of the Anderson and Hsiao (1982) moment conditions is of reduced rank, when process is cointegrated. It is natural to use this information and consider a rank based cointegration test, based on the rank of the jacobian matrix. In this paper, we propose such a test and show that it is applicable in situations with time-series heteroscedasticity and, unlike the likelihood based tests, the new test does not require any numerical optimization. At the same time, this procedure cannot provide inference that is uniform over the parameter space, as the asymptotic distribution of the test depends on the properties of the initial condition.

In the Monte Carlo section of this paper, we investigate the finite sample properties of the proposed procedure. We find that the new testing procedure provides a good size control as well as high power in most of the designs considered. However, in some setups this test lacks power if the data generating process for the initial condition substantially deviates from stationarity.

The paper is structured as follows. In Sect. 2, we briefly present the model, the testing problem at hand and the results for the testing procedure of Binder et al. (2005). Rank-based cointegration testing procedure is formally introduced in Sect. 3. In Sect. 4, we continue with the finite sample performance by means of a Monte Carlo analysis. In Sect. 5, we illustrate the testing procedure using the data of Alonso-Borrego and Arellano (1999). Section 6 concludes.

Here we briefly discuss notation. **Bold** upper-case letters are used to denote the original parameters, i.e. $\{\Phi, \Sigma, \Psi\}$, while the lower-case letters $\{\phi, \sigma, \psi\}$ denote $\text{vec}(\cdot)$ ($\text{vech}(\cdot)$ for symmetric matrices) of corresponding parameters, in the univariate setup corresponding parameters are denoted by $\{\phi, \sigma^2, \psi^2\}$. We use $\rho(A)$ to denote the spectral radius¹ of a matrix $A \in \mathbb{R}^{n \times n}$. We define $\bar{y}_{i-} \equiv (1/T) \sum_{t=1}^T y_{i,t-1}$ and similarly $\bar{y}_i \equiv (1/T) \sum_{t=1}^T y_{i,t}$. We use \tilde{x} to indicate variables after Within Group transformation (for example $\tilde{y}_{i,t} \equiv y_{i,t} - \bar{y}_i$), while \check{x} are used for variables after a “quasi-averaging” transformation.² For further details, see Abadir and Magnus (2002). Where necessary, we use the 0 subscript to denote the true value of the parameters, e.g. Φ_0 .

¹ $\rho(A) \equiv \max_k (|\lambda_k|)$, where λ_k are (possibly complex) eigenvalues of a matrix A .

² $\check{y}_i \equiv \bar{y}_i - y_{i,0}$ and $\check{y}_{i-} \equiv \bar{y}_{i-} - y_{i,0}$.

2 Cointegration testing for fixed T panels

2.1 The model

In this paper, we consider the following PVAR(1) specification

$$y_{i,t} = \eta_i + \Phi y_{i,t-1} + \varepsilon_{i,t}, \quad i = 1, \dots, N, \quad t = 1, \dots, T, \quad (1)$$

where $y_{i,t}$ is an $[m \times 1]$ vector, Φ is an $[m \times m]$ matrix of parameters that are of main interest, η_i is an $[m \times 1]$ vector of unobserved individual specific covariates, and $\varepsilon_{i,t}$ is an $[m \times 1]$ vector of innovations independent across i , with zero mean and some finite covariance matrix. If one sets $m = 1$, the model reduces to the univariate linear dynamic panel data model with autoregressive dynamics.

Throughout this paper we assume that η_i satisfy the so-called “common dynamics” assumption

$$\eta_i = (I_m - \Phi)\mu_i,$$

where $\Pi \equiv \Phi - I_m$. If at least one eigenvalue of Φ is equal to unity this assumption ensures that there is no discontinuity in the data generating process (DGP), for further discussion, see e.g. BHP.

Assuming common dynamics we can rewrite the model in (1) as

$$\Delta y_{i,t} = \Pi u_{i,t-1} + \varepsilon_{i,t}, \quad i = 1, \dots, N, \quad t = 1, \dots, T. \quad (2)$$

Here we define $u_{i,t-1} \equiv y_{i,t-1} - \mu_i$. We say that series $y_{i,t}$ are *cointegrated* if the Π matrix is of reduced rank $0 < r < m$.³ In particular, there exist full column rank (of rank r) matrices α_r and β_r such that:⁴

$$\Phi = I_m + \alpha_r \beta_r', \quad (3)$$

where r is the rank of Π . Matrices α_r and β_r are not unique, as for any $[r \times r]$ invertible matrix U

$$\alpha_r \beta_r' = \alpha_r U U^{-1} \beta_r' = \alpha_r^* \beta_r'^*.$$

This is the so-called “rotation problem”. As a result, it is a usual practice in the literature to introduce identifying restrictions on α_r or β_r . To construct the test statistic that we formally introduce in Sect. 3.1, we follow common practise and assume that $\beta_r' = (I_r, \Delta)$, where Δ is an $[r \times m - r]$ matrix.

³ Unlike time series models, we do not define cointegration as a property of time series, as in our setup we keep T fixed.

⁴ We slightly abuse the notation in this case, so that it remains consistent with the general practice of the time series cointegration literature.

As a natural starting point one can consider the fixed effects (FE) estimator for Φ parameter

$$\hat{\Phi}' = \left(\sum_{i=1}^N \sum_{t=1}^T \tilde{y}_{i,t-1} \tilde{y}'_{i,t-1} \right)^{-1} \sum_{i=1}^N \sum_{t=1}^T \tilde{y}_{i,t-1} \tilde{y}'_{i,t}. \tag{4}$$

It is well known that the estimator of this type is inconsistent for fixed T , see e.g. Nickell (1981) and Hahn and Kuersteiner (2002). For the unit root case, i.e. $\Phi_0 = I_m$ (correspondingly $r = 0$), one can show that

$$\text{plim}_{N \rightarrow \infty} \hat{\Phi} = \left(1 - \frac{3}{(T + 1)} \right) I_m, \tag{5}$$

suggesting that for $T > 2$ we can obtain a consistent estimate of $\Phi_0 = I_m$, by considering the bias corrected estimator of the form

$$\hat{\Phi}_{BC} \equiv \frac{T + 1}{T - 2} \hat{\Phi}. \tag{6}$$

The bias corrected estimator $\hat{\Phi}_{BC}$ can be used for unit root testing, see e.g. Harris and Tzavalis (1999) and Kruiniger and Tzavalis (2002) for the univariate case. However, no comparable result is available when series are cointegrated, i.e. $0 < r < m$. Alternatively, one can rely on the bias-corrected approaches of Bun and Carree (2005), Ramalho (2005) or Dhaene and Jochmans (2016) to obtain a consistent estimator of Φ . However, these procedures generally fail to guarantee correct asymptotic inference for reasons similar to those discussed in the next section.⁵

2.2 The likelihood ratio test of BHP

In this section we introduce likelihood based testing and estimation approach advocated by BHP. First, we list the assumptions (that we abbreviate as, TML) used to derive asymptotic distribution of the transformed maximum likelihood estimator:

- (TML 1) The error terms $\varepsilon_{i,t}$ are i.i.d. across i and uncorrelated over time $E[\varepsilon_{i,t} \varepsilon'_{i,s}] = \mathbf{0}_m$ for $s \neq t$, and $E[\varepsilon_{i,t} \varepsilon'_{i,t}] = \Sigma$ for $t > 0$. $E[\|\varepsilon_{i,t}\|^4] < \infty$ holds $\forall t$.
- (TML 2) The initial deviations $\mathbf{u}_{i,0} \equiv \mathbf{y}_{i,0} - \boldsymbol{\mu}_i$ are i.i.d. across i , with $E[\mathbf{u}_{i,0}] = \mathbf{0}_m$ and positive definite $E[\mathbf{u}_{i,0} \mathbf{u}'_{i,0}] = \Sigma_{u_0}$. $E[\|\mathbf{u}_{i,0}\|^4] < \infty$ holds.
- (TML 3) The following moment restrictions are satisfied: $E[\Pi \mathbf{u}_{i,0} \varepsilon'_{i,t}] = \mathbf{0}_m$ for all i and $t = 1, \dots, T$.
- (TML 4) $N \rightarrow \infty$, T is fixed.
- (TML 5) Denote by $\boldsymbol{\kappa}$ a $[k \times 1]$ vector of unknown coefficients. $\boldsymbol{\kappa} \in \Gamma$, where Γ is a compact subset of \mathbb{R}^k and $\boldsymbol{\kappa}_0 \in \text{interior}(\Gamma)$, while $\rho(\Phi_0) \leq 1$.

Assumption (TML 1) is the no serial correlation assumption. Note how Assumption (TML 2) places moment restrictions only on one linear combination of $\mathbf{y}_{i,0}$ and $\boldsymbol{\mu}_i$,

⁵ Dhaene and Jochmans (2016) prove singularity of the hessian matrix for $m = 1$.

rather than separately on $y_{i,0}$ and μ_i . Assumption (TML 3) imposes zero covariance restriction on initial deviation and the error terms, which is a standard assumption in the literature. For $\Pi = \mathbf{O}$, $E[u_{i,0}e'_{i,t}]$ is unrestricted, as $\Delta y_{i,1}$ is not a function of $u_{i,0}$. Assumption (TML 5) is the main exception as compared to the setup in e.g. Juodis (2016), as in this paper we allow the maximum eigenvalue of the autoregressive matrix Φ_0 to be 1. The exact components and dimension k of the κ vector are related to a particular parametrization of the parameter space used for estimation. Assumptions TML are almost identical to the corresponding assumption FE in Binder et al. (2005), the only difference is that some of their assumptions are “high-level” (e.g. bounds on covariances), while we consider “low-level” assumptions by imposing restrictions directly on $e_{i,t}$ and $u_{i,0}$.

The quasi log-likelihood function for $\Delta Y_i = \text{vec}(\Delta y_{i,1}, \dots, \Delta y_{i,T})$ is then defined as follows (up to a constant):

$$\ell(\kappa) \equiv -\frac{N}{2} \log |\Sigma_{\Delta\tau}| - \frac{N}{2} \text{tr} \left(\left(\mathbf{R}' \Sigma_{\Delta\tau}^{-1} \mathbf{R} \right) \frac{1}{N} \sum_{i=1}^N \Delta Y_i \Delta Y_i' \right), \tag{7}$$

where $\kappa = (\phi', \sigma', \psi')'$, hence $k = m^2 + m(m + 1)$ and Ψ is the variance-covariance matrix of the initial observation $\Delta y_{i,1}$. The $\Sigma_{\Delta\tau}$ matrix has a block tri-diagonal structure, with $-\Sigma$ on first lower and upper off-diagonal blocks, and 2Σ on all but the first (1, 1) diagonal blocks. The first (1,1) block is set to Ψ which takes into account the fact that we do not restrict $\Delta y_{i,1}$ to be covariance stationary.⁶ The $[mT \times mT]$ \mathbf{R} matrix has \mathbf{I}_m matrices on the diagonal blocks, and $-\Phi$ on the first lower off-diagonal blocks.

In Juodis (2016) it is shown that the log-likelihood function of BHP can be substantially simplified to

$$\begin{aligned} \ell(\kappa) = & -\frac{N}{2} \left((T - 1) \log |\Sigma| + \text{tr} \left(\Sigma^{-1} \frac{1}{N} \sum_{i=1}^N \sum_{t=1}^T (\tilde{y}_{i,t} - \Phi \tilde{y}_{i,t-1})(\tilde{y}_{i,t} - \Phi \tilde{y}_{i,t-1})' \right) \right) \\ & - \frac{N}{2} \left(\log |\Theta| + \text{tr} \left(\Theta^{-1} \frac{T}{N} \sum_{i=1}^N (\dot{y}_i - \Phi \dot{y}_{i-})(\dot{y}_i - \Phi \dot{y}_{i-})' \right) \right), \end{aligned} \tag{8}$$

where $\kappa = (\phi', \sigma', \theta')'$ and $\Theta \equiv T(\Psi - \Sigma) + \Sigma$.

If the matrix $\Phi - \mathbf{I}_m = \alpha_r \beta_r'$ is of a reduced rank r (cointegration),⁷ this information can be taken into account in estimation and used for testing. To avoid rotational indeterminacy, one can use the same parametrization as BHP and set $\beta_r' = \delta_r' \mathbf{H}_r' + \mathbf{b}_r'$, where \mathbf{H}_r and \mathbf{b}_r are known and δ_r is an $[m - r \times r]$ (given $0 < r < m$) matrix of parameters. The parameter set in this case is defined as $\kappa_r = ((\text{vec } \alpha_r)', (\text{vec } \delta_r)', \sigma', \theta')'$. Binder et al. (2005) suggest to use the likelihood function (8) in constructing the likelihood

⁶ However, in this setup we still, for simplicity, assume that the initial observation has a zero mean, i.e. $E[\Delta y_{i,1}] = \mathbf{0}_m$.

⁷ Here subscript r is introduced to highlight that these matrices are of rank r .

ratio test statistic to consistently estimate rank r . In particular, the likelihood ratio test statistics for the null hypothesis $H_0 : r_0 = r$ against the alternative $H_A : r_A = r + 1$ is of the form

$$LR(r, r + 1) = -2(\ell(\kappa_r) - \ell(\kappa_{r+1})). \quad (9)$$

Similarly to the classical maximum eigenvalue test of [Johansen \(1991\)](#) for $N = 1$, the overall testing procedure can be performed sequentially, i.e. by first considering $r_0 = 0$ ($\Pi_0 = \mathbf{O}_m$), and if rejected, proceeding with $r_0 = 1$, and so on. BHP argue that this procedure under $H_0 : r_0 = r$ has a $\chi^2(\cdot)$ asymptotic limit. In particular, in Remark 4.1. they note that: “Unlike in time-series models, first differencing in panels with T fixed still allows identification and estimation of the long-run (level) relations that are of economic interest, irrespective of the unit root and cointegrating properties of the $y_{i,t}$ process.” As we show next, this conclusion is not completely correct.

One of the standard regularity conditions for extremum estimators, is that the asymptotic (or expected) hessian matrix, $\mathcal{H}_\ell \equiv E[\mathcal{H}^N(\kappa_0)]$ is positive definite. In [Bond et al. \(2005\)](#), authors showed that for the TML estimator of [Hsiao et al. \(2002\)](#) (which is a special case of [Binder et al. 2005](#) for $m = 1$) this regularity condition is violated. In the next theorem we show that the same conclusion extends to a more general case with $m \geq 1$.

Theorem 1 (Singularity) *Let Assumptions TML be satisfied. Then at $\Phi_0 = \mathbf{I}_m$ the \mathcal{H}_ℓ matrix is singular, i.e.*

$$|\mathcal{H}_\ell| = 0. \quad (10)$$

Proof In the Appendix. □

As the TML estimator can be seen as a non-linear MM estimator with the score vector defining the moment conditions, singularity of the \mathcal{H}_ℓ matrix can be seen as a “weak instrument” problem (using the GMM notation). The singularity result in Theorem 1 is of special interest when the inference regarding the rank of $\mathbf{I}_m - \Phi_0$ is concerned.

It is important to note that despite singularity of \mathcal{H}_ℓ , the TML estimator $\hat{\kappa}_{TML\ell}$ remains consistent, hence the identification part of Remark 4.1. in BHP is correct. However, as a result of singularity the limiting distribution for this estimator is non-standard. Using the approach of [Roznitky et al. \(2000\)](#), [Ahn and Thomas \(2006\)](#) showed that in the univariate model (i.e. $m = 1$), the TML estimator of ϕ converges at the $N^{1/4}$ rate to a non-standard distribution.⁸ Additionally, they show that LR test statistic for $H_0 : \phi_0 = 1$ has a mixture distribution, of a $\chi^2(1)$ random variable and a degenerate random variable that takes value 0 with probability 1, with equal mixing weights of 0.5. In this paper, we do not attempt to study the distributional consequences of the singularity for the LR test and leave it for future research.⁹ Based on results in [Dovonon and Renault \(2009\)](#) (for GMM), it is known that for general rank deficiencies the maximal rate of convergence is $N^{1/4}$. However, no results regarding the behavior of the estimator (see discussion in [Dovonon and Hall 2016](#)) and the LR ratio test in

⁸ [Kruiniger \(2013\)](#) extended their results by allowing cross-sectional heteroscedasticity in the error terms.

⁹ Our numerical simulations suggest that the rank of \mathcal{H}_ℓ has rank deficiency larger than one (for $m = 2$ the rank of \mathcal{H}_ℓ is equal to 7, while full rank is 10), hence results of [Roznitky et al. \(2000\)](#) need to be generalized taking into account this possibility.

cases like ours are available. As a result, it is not obvious that using the critical values from the χ^2 distribution with $(m - r)^2$ degrees of freedom results in a conservative test.

Although the unit root model is not of prime importance for the main topic of this paper, Theorem 1 provides a natural starting point for intuition of the next result. For the unit root case (i.e. $\Phi_0 = I_m$) the expression for \mathcal{H}_ℓ simplifies dramatically as $\Sigma_0 = \Theta_0$. That allowed us in Theorem 1 to show that $|\mathcal{H}_\ell| = 0$ for any value of Σ_0 and T . Unfortunately, no result of this type is available when Π is of reduced rank $r > 0$. However, some special results can be derived for $T = 2$.

Proposition 1 *Let Φ_0 be such that $\text{rk } \Pi_0 = r$ and $T = 2$ then*

$$\text{rk } \mathcal{H}_\ell \leq 0.5m(m - 1) + r^2. \quad (11)$$

Proof In the Appendix. □

This quantity is smaller than m^2 for all $m \leq 4$ (note that the bivariate PVAR model is analyzed in most empirical studies with limited number of time-series observations). It follows that for cases of most empirical value the expected hessian matrix is singular and the corresponding estimator does not have a normal limiting distribution. Although in this paper we do not prove more general results for $T > 2$, we performed numerous numerical evaluations of \mathcal{H}_ℓ for larger values of T and different combinations of population matrices in the bivariate setup.¹⁰ For all setups we found that the expected hessian matrix is singular for $r < m$ and of full rank otherwise. Given these results the unit root and cointegration testing procedure of BHP that is based on asymptotic $\chi^2(\cdot)$ critical values is not asymptotically valid.

Remark 1 Alternatively, instead of considering likelihood function for observations in first differences one can consider a correlated random effects likelihood function (conditional on $y_{i,0}$) as in [Arellano \(2016\)](#) and [Kruiniger \(2013\)](#). Although we do not formally consider a possible singularity of the hessian matrix for that estimator, we conjecture that the main conclusions of this paper are also applicable to that approach ([Ahn and Thomas 2006](#); [Kruiniger 2013](#) proved this for $m = 1$).

Remark 2 Note that the results of this section are derived under assumption that Ψ is estimated without any restrictions, i.e. as suggested by [Binder et al. \(2005\)](#). If one instead imposes some restrictions on this parameter matrix, e.g. covariance stationary, it is possible that the expected hessian matrix has full rank. For example, [Kruiniger \(2008\)](#) considers univariate results, where he shows that for $\phi_0 = 1$, the TML estimator retains standard asymptotic properties if the stationarity assumption is used in estimation.

¹⁰ In particular, setups of BHP were considered.

3 Jacobian based testing

3.1 Regularity conditions

In this section we propose an alternative approach to cointegration testing. To explain the intuition of our approach, consider the following Anderson and Hsiao (1982) moment conditions for Panel VAR(1) model (see e.g. BHP)

$$\text{vec} \left(E[(\Delta y_{i,t} - \Phi \Delta y_{i,t-1}) y'_{i,t-2}] \right) = \mathbf{0}_{m^2}, \quad t = 2, \dots, T, \tag{12}$$

where only the most recent lag, $y_{i,t-2}$ is used as an instrument (other choices are discussed later in the paper). The (minus) jacobian of these moment conditions is given by

$$\left(E[\Delta y_{i,t-1} y'_{i,t-2}] \right)' \otimes \mathbf{I}_m, \quad t = 2, \dots, T. \tag{13}$$

From the properties of the Kronecker product it follows that the rank of this matrix is determined by the rank of the matrix inside the brackets.¹¹ That term can be expanded as follows (upon redefining $t \rightarrow t + 1$, as the previous expression is well defined for $t - 1 = T$)

$$E[\Delta y_{i,t} y'_{i,t-1}] = \Pi E[u_{i,t-1} y'_{i,t-1}] + E[\epsilon_{i,t} y'_{i,t-1}]. \tag{14}$$

Under the no serial correlation assumption, e.g. (TML 1), $E[\epsilon_{i,t} y'_{i,t-1}] = \mathbf{0}_m$, while the first term is the product of rank r and rank $r_{uy} \leq m$ matrices. As a result $\text{rk} (E[\Delta y_{i,t} y'_{i,t-1}]) = \min (r, r_{uy}) < m$ and this leads to a violation of the “relevance” condition for the Instrumental Variable (IV) estimator, thus the Anderson and Hsiao (1982) moment conditions cannot be used to consistently estimate Φ .¹² However, we can use the jacobian matrix directly to test for cointegration, avoiding the estimation step.

Next, we list assumptions that are sufficient to derive the asymptotic properties of the testing approach that we introduce in this section. For the purpose of this section we deviate from (TML) assumptions and restrict moments of μ_i and $y_{i,0}$ separately, rather than their linear combination.

- (A.1) The error terms $\epsilon_{i,t}$ are i.i.d. across i and uncorrelated over time, $E[\epsilon_{i,t} \epsilon'_{i,s}] = \mathbf{0}_m$ for $s \neq t$, and $E[\epsilon_{i,t} \epsilon'_{i,t}] = \Sigma_t$ for $t > 0$. $E[\|\epsilon_{i,t}\|^4] < \infty$ holds $\forall t$.
- (A.2) The μ_i are i.i.d. across i , with $E[\mu_i] = \mathbf{0}_m$ and $E[\mu_i \mu'_i] = \Sigma_\mu$. Furthermore, for all i and $t \geq 0$, $E[\mu_i \epsilon'_{i,t}] = \mathbf{0}_m$. $E[\|\mu_i\|^4] < \infty$ holds.

Note that we allow $\epsilon_{i,t}$ to be heteroscedastic over time. However, cross-sectional heteroscedasticity is in general not allowed, as in this case $E[\Delta y_{i,t} y'_{i,t-1}]$ is individual specific, and we cannot consistently estimate both the mean and the variance of $\Delta y_{i,t} y'_{i,t-1}$. In particular, for this reason we assume that μ_i are iid. As we consider

¹¹ See e.g. Magnus and Neudecker (2007).

¹² Note that for large T consistent estimation is possible, but with non-standard distribution theory, see Phillips (2015).

fixed T panels, it is important to explicitly specify DGP for the initial conditions. For this, we assume that $y_{i,0}$ are of the following form

$$y_{i,0} = \Upsilon \mu_i + \varepsilon_{i,0}. \tag{15}$$

Here Υ is an $[m \times m]$ matrix that controls the degree of non-stationarity in the initial condition, i.e. if $\Upsilon \neq I_m$, initial condition is effect non-stationarity.¹³ Note that if $\Upsilon = I_m$ the assumption (A.2) can be somewhat relaxed allowing heterogenous μ_i . What is left is to specify assumptions on $\varepsilon_{i,0}$. Below we list a few DGPs for $\varepsilon_{i,0}$ that can be used for our purpose.

(DGP.1) $\varepsilon_{i,0} \sim (\mathbf{0}_m, \Sigma_0)$ with Σ_0 positive (semi-)definite matrix.

(DGP.2) $\varepsilon_{i,0} = \sum_{l=0}^M \Phi^l \varepsilon_{i,-l}^*$. Here M is assumed to be finite.

(DGP.3) $\varepsilon_{i,0} = \sum_{l=0}^\infty (\Phi^l - C) \varepsilon_{i,-l}^* + C \xi_i$. Here ξ_i is an $[m \times 1]$ vector of the (independent) individual-specific initialization effects.¹⁴

In what follows, we assume that all random variables in (DGP.1)–(DGP.3) satisfy assumptions (A.1)–(A.2). Furthermore, in (DGP.3) for simplicity all $\varepsilon_{i,-l}^*$ are homoscedastic over the time-series dimension. This restriction can be relaxed by assuming that all Σ_{-l}^* are appropriately summable as $l \rightarrow \infty$. (DGP.3) initialization was used in the Monte Carlo studies of BHP and is motivated by the Granger Representation Theorem, see e.g. Theorem 4.2 in Johansen (1995). The (DGP.2), was e.g. used by Hayakawa (2016). It is important to emphasize that all three DGP are well defined for all values of r .¹⁵ $E[\varepsilon_{i,t} y'_{i,t-1}] = \mathbf{0}_m$ is a direct implication of Assumptions (A.1)–(A.2) and (DGP.1)–(DGP.3).

3.2 Rank test

In this paper, we use the generalized rank test of Kleibergen and Paap (2006) as a basis for our testing procedure. Here we briefly introduce their testing procedure and later apply it to our problem. In construction of the rank test Kleibergen and Paap (2006) use the property that any $[k \times f]$ matrix D can be decomposed as:

$$D = A_q B_q + A_{q,\perp} A_q B_{q,\perp},$$

where A_q is a $[(k - q) \times (f - q)]$ matrix and all \perp matrices are defined in the usual way. For $A_q = \mathbf{0}$ the rank of D is determined by the rank of $A_q B_q$. The procedure in Kleibergen and Paap (2006) is based on testing if A_q is equal to $\mathbf{0}_{(k-q) \times (f-q)}$, with matrices A_q, B_q, A_q obtained using the singular value decomposition (SVD). In our case, we consider singular value decomposition of jacobian matrix, i.e. $D = E[\Delta y_{i,t} y'_{i,t-1}]$.

¹³ Also referred as “mean non-stationarity”, see e.g. Bun and Sarafidis (2015).

¹⁴ Here $C \equiv \beta_\perp (\alpha'_\perp \beta_\perp)^{-1} \alpha'_\perp$ is an $m - r$ rank matrix, while $\alpha_\perp, \beta_\perp$ are the orthogonal complements of α, β .

¹⁵ In (DGP.3) for $\rho(\Phi) < 1$ we have $C = \mathbf{0}_m$, resulting in stationary initialization. On the other hand, $\Phi = I_m$ implies $C = I_m$ (by definition) so that (DGP.3) and (DGP.2) coincide (by redefining M to $M + 1$).

Next, we define the sample analogue of D :

$$\overline{\Delta \mathbf{y}_{i,t} \mathbf{y}'_{i,t-1}} \equiv \frac{1}{N} \sum_{i=1}^N \Delta \mathbf{y}_{i,t} \mathbf{y}'_{i,t-1}.$$

Applying the standard Lindeberg–Levý CLT, it follows that:

$$\sqrt{N} \text{vec} \left(\overline{\Delta \mathbf{y}_{i,t} \mathbf{y}'_{i,t-1}} - E[\Delta \mathbf{y}_{i,t} \mathbf{y}'_{i,t-1}] \right) \xrightarrow{d} N(\mathbf{0}_{m^2}, \mathbf{V}), \quad t = 2, \dots, T.$$

Here the full rank matrix \mathbf{V} can be consistently estimated using its finite sample counterpart:

$$\mathbf{V}_N = \frac{1}{N} \sum_{i=1}^N \text{vec}(\Delta \mathbf{y}_{i,t} \mathbf{y}'_{i,t-1}) \text{vec}(\Delta \mathbf{y}_{i,t} \mathbf{y}'_{i,t-1})' - \overline{\text{vec} \Delta \mathbf{y}_{i,t} \mathbf{y}'_{i,t-1}} \overline{\text{vec} \Delta \mathbf{y}_{i,t} \mathbf{y}'_{i,t-1}}'.$$

Consequently, the estimator $\overline{\Delta \mathbf{y}_{i,t} \mathbf{y}'_{i,t-1}}$ satisfies sufficient conditions in Kleibergen and Paap (2006).¹⁶ As a result one can apply Theorem 1 of Kleibergen and Paap (2006) to the problem at hand:

Theorem 2 *Let Assumptions (A.1)–(A.2) be satisfied with $\boldsymbol{\varepsilon}_{i,0}$ generated by one of (DGP.1)–(DGP.3), then:*

$$\sqrt{N} \hat{\boldsymbol{\lambda}}_r \xrightarrow{d} N(\mathbf{0}_{(m-r)^2}, \boldsymbol{\Omega}_r),$$

where

$$\begin{aligned} \hat{\boldsymbol{\lambda}}_r &= \text{vec} \hat{\mathbf{A}}_r, \quad \hat{\mathbf{A}}_r = \hat{\mathbf{A}}'_{r,\perp} \overline{\Delta \mathbf{y}_{i,t} \mathbf{y}'_{i,t-1}} \hat{\mathbf{B}}'_{r,\perp}, \\ \boldsymbol{\Omega}_r &= (\mathbf{B}_{r,\perp} \otimes \mathbf{A}'_{r,\perp}) \mathbf{V} (\mathbf{B}_{r,\perp} \otimes \mathbf{A}'_{r,\perp})' \end{aligned}$$

Furthermore, under $H_0 : \text{rk} E[\Delta \mathbf{y}_{i,t} \mathbf{y}'_{i,t-1}] = r$, the test statistic:

$$rk(r) = N \hat{\boldsymbol{\lambda}}_r' \boldsymbol{\Omega}_r^{-1} \hat{\boldsymbol{\lambda}}_r$$

converges in distribution to a $\chi^2(\cdot)$ random variable with $(m - r)^2$ degrees of freedom.

Matrices \mathbf{A} and \mathbf{B} in Theorem 2 are obtained from the SVD of $\overline{\Delta \mathbf{y}_{i,t} \mathbf{y}'_{i,t-1}}$. An operational version of the $rk(r)$ test statistic is obtained by replacing the (unknown) matrix $\boldsymbol{\Omega}_r$ with some consistent estimator. An obvious choice for $\hat{\boldsymbol{\Omega}}_r$ is given by:

$$\hat{\boldsymbol{\Omega}}_r = (\hat{\mathbf{B}}_{r,\perp} \otimes \hat{\mathbf{A}}'_{r,\perp}) \mathbf{V}_N (\hat{\mathbf{B}}_{r,\perp} \otimes \hat{\mathbf{A}}'_{r,\perp})'.$$

¹⁶ $\overline{\Delta \mathbf{y}_{i,t} \mathbf{y}'_{i,t-1}}$ satisfies **Assumption 1** (asymptotic normal distribution), while \mathbf{V} satisfies **Assumption 2** (full rank). Note that if \mathbf{V} is of reduced rank the result in Theorem 2 can be modified as in Andrews (1987).

Note that this test statistic uses the unrestricted estimate of $E[\Delta y_{i,t} y'_{i,t-1}]$, hence we do not explicitly specify the alternative hypothesis (as it is done for LR test of e.g. Johansen 1991 or Binder et al. 2005). However, as we discuss next, this testing approach has power only towards alternative with $r_0 > r$, i.e. test rejects if the true rank is larger than the hypothesized one.

The result in Theorem 2 suggests that one can use a sequential testing procedure in order to determine the rank of $E[\Delta y_{i,t} y'_{i,t-1}]$. In particular, one can begin with $H_0 : \text{rk } E[\Delta y_{i,t} y'_{i,t-1}] = 0$, and if rejected, consider $H_0 : \text{rk } E[\Delta y_{i,t} y'_{i,t-1}] = 1$, and so on, until the first non-rejection. The construction of such sequential procedure does not differ from the one suggested in Johansen (1991) or Binder et al. (2005). However, $E[\Delta y_{i,t} y'_{i,t-1}]$ is not of the prime interest for us, as we are interested in testing the rank of Π . This begs the question:

Under which conditions one can interpret rejection/non-rejection of the $rk(r)$ test as an evidence regarding the rank of Π ?

If one rejects the null hypothesis $H_0 : \text{rk } E[\Delta y_{i,t} y'_{i,t-1}] = r$, one can also reject $H_0 : \text{rk } \Pi_0 = r$, as the rank of $\text{rk } E[\Delta y_{i,t} y'_{i,t-1}] \leq \text{rk } (\Pi)$. However, our assumptions do not ensure that $\Pi E[u_{i,t-1} y'_{i,t-1}]$ has a reduced rank if and only if $y_{i,t}$ are cointegrated (the “if” part was established above). Hence, it still remains to be investigated under which conditions the $E[\Delta y_{i,t} y'_{i,t-1}]$ term is of reduced rank if and only if Π is of reduced rank. Let us investigate this issue more closely by expanding the $E[u_{i,t-1} y'_{i,t-1}]$ term (for $t \geq 2$):

$$\begin{aligned}
 E[u_{i,t-1} y'_{i,t-1}] &= E \left[\left(\Phi^{t-1} u_{i,0} + \sum_{s=0}^{t-2} \Phi^s \varepsilon_{i,t-s-1} \right) \right. \\
 &\quad \left. \left(\mu_i + \Phi^{t-1} u_{i,0} + \sum_{s=0}^{t-2} \Phi^s \varepsilon_{i,t-s-1} \right)' \right] \\
 &= \underbrace{\Phi^{t-1} (\mathbf{Y} - \mathbf{I}_m) \Sigma_{\mu} (\Phi^{t-1} (\mathbf{Y} - \mathbf{I}_m))' + \Phi^{t-1} (\mathbf{Y} - \mathbf{I}_m) \Sigma_{\mu}}_{p.s.d} \\
 &\quad + \underbrace{\sum_{s=0}^{t-2} \Phi^s \Sigma_{t-1-s} \Phi^{s'}}_{p.d.} + E[\Phi^{t-1} \varepsilon_{i,0} \varepsilon'_{i,0} (\Phi^{t-1})']
 \end{aligned}$$

In the effect-stationary setup ($\mathbf{Y} = \mathbf{I}_m$) all terms involving \mathbf{Y} are equal to \mathbf{O}_m . Furthermore, the third term is a p.d. matrix as all Σ_s matrices are positive definite. Moreover, (DGP.1)–(DGP.3) assumptions are sufficient to conclude that $E[\Phi^{t-1} \varepsilon_{i,0} \varepsilon'_{i,0} (\Phi^{t-1})']$ is also at least positive semi-definite (p.s.d.) matrix. Thus we conclude that the “only if” part is also valid for $\mathbf{Y} = \mathbf{I}_m$.

Unfortunately, there is a lot of evidence in the DPD literature suggesting that in general this assumption can be too restrictive, see e.g. Arellano (2003) and Roodman (2009). If $\mathbf{Y} \neq \mathbf{I}_m$, the first, third and fourth terms are p.s.d. matrices, while it

is not immediately clear what happens with the second term. For the procedure to consistently estimate the rank of Π (and not only underestimate it), one has to place additional restriction

(IDN) Matrix $\Phi^{t-1}(\mathbf{Y} - \mathbf{I}_m)\Sigma_\mu$ is such that $E[\mathbf{u}_{i,t-1}\mathbf{y}'_{i,t-1}]$ has a full rank m .

Note that restriction is “high-level” as it imposes restrictions not directly on parameters, but instead on their non-linear function. Intuitively, assumption (IDN) is satisfied if $\Phi^{t-1}(\mathbf{Y} - \mathbf{I}_m)\Sigma_\mu$ is “large”, with eigenvalues sufficiently bounded away from zero. However, it is not a trivial task to identify the parameter space of $\{\Phi, \mathbf{Y}, \Sigma_\mu\}$ for the aforementioned condition to be satisfied. One special case is obtained for $\mathbf{Y} = \mathbf{I}_m$ (effect stationarity) with other matrices being unrestricted (at least finite). If we can ensure that $\Phi^{t-1}(\mathbf{Y} - \mathbf{I}_m)\Sigma_\mu$ is such that $E[\mathbf{u}_{i,t-1}\mathbf{y}'_{i,t-1}]$ has full rank m , then $E[\Pi\mathbf{u}_{i,t-1}\mathbf{y}'_{i,t-1}]$ has reduced rank r if and only if $\mathbf{y}_{i,t-1}$ are cointegrated.¹⁷ In the Monte Carlo section of this paper, we check the adequacy of the proposed procedure by considering different values of \mathbf{Y} that are mentioned in the literature.

The test statistic in Theorem 2 is based only on one time series observation (in a sense that if $T > 2$, then we can construct a test statistic for every value of t , but $t = 1$). However, it is not the most efficient way of using the time series information provided. Instead, all time series observations can be pooled into one test statistic to test the rank of:¹⁸

$$\overline{\Delta\mathbf{y}_{i,t}\mathbf{y}'_{i,t-1T}} = \frac{1}{N} \sum_{i=1}^N \frac{1}{T-1} \sum_{t=2}^T \Delta\mathbf{y}_{i,t}\mathbf{y}'_{i,t-1}. \tag{16}$$

For any fixed value of T , the $\overline{\Delta\mathbf{y}_{i,t}\mathbf{y}'_{i,t-1T}}$ term satisfies the sufficient conditions for the CLT, so that the results of Theorem 2 can be extended, with V_N for this case given by:

$$V_N = \frac{1}{N} \sum_{i=1}^N \text{vec} \left(\frac{1}{T-1} \sum_{t=2}^T \Delta\mathbf{y}_{i,t}\mathbf{y}'_{i,t-1} \right) \text{vec} \left(\frac{1}{T-1} \sum_{t=2}^T \Delta\mathbf{y}_{i,t}\mathbf{y}'_{i,t-1} \right)' - \text{vec} \overline{\Delta\mathbf{y}_{i,t}\mathbf{y}'_{i,t-1T}} \text{vec} \overline{\Delta\mathbf{y}_{i,t}\mathbf{y}'_{i,t-1T}}'. \tag{17}$$

In the next section we use “rk-J” to denote the jacobian based cointegration test for $\overline{\Delta\mathbf{y}_{i,t}\mathbf{y}'_{i,t-1T}}$.

Until now we considered only the jacobian of Anderson and Hsiao (1982) moment conditions, however, for $T > 2$ further lags $\mathbf{y}_{i,t-j}$, can be used. Nevertheless, it is not clear that the use of lags j larger than $j > 1$ still ensures that, even in the effect stationary case, $E[\Delta\mathbf{y}_{i,t}\mathbf{y}'_{i,t-jT}]$ has reduced rank r if and only if $\text{rk } \Pi = r$. Moreover, the power of the test might be substantially affected by the choice of lags, as with any

¹⁷ Note that positive definiteness of $E[\mathbf{u}_{i,t-1}\mathbf{y}'_{i,t-1}]$ is a sufficient, but not a necessary condition. The term can be negative definite or even indefinite, as long as it has full rank.

¹⁸ In principle, other pooling schemes with weighted averages are possible, but for ease of exposition in this paper we consider simple time average.

alternative close to the unit circle we encounter the weak instruments problem for any distanced lags. On the other hand, we can expect a better test power to the alternatives with substantially lower $\rho(\Phi)$.

Remark 3 If the model contains time effects λ_t , the test statistic needs to be modified using variables in deviations from their cross-sectional averages $\check{y}_{i,t} \equiv y_{i,t} - (1/N) \sum_{i=1}^N y_{i,t}$ rather than levels.

Remark 4 One important advantage of the proposed test statistic is the additional flexibility while dealing with unbalanced panels. As long as for every individual i at least one $\Delta y_{i,t} y'_{i,t-1}$ ($t > 1$) term is available, the test statistic can be computed. The only difference as compared to the balanced case is that an individual contribution to $\Delta y_{i,t} y'_{i,t-1}$ is no longer a simple averages with $T - 1$ terms, but has an individual specific number of observations $T_i - 1$.

Remark 5 The testing procedure remains valid if, as suggested by [Kleibergen and Paap \(2006\)](#), instead of $\Delta y_{i,t} y'_{i,t-1}$ we investigate the rank of $D = G_N \Delta y_{i,t} y'_{i,t-1} F_N$ (for any full rank matrices $\text{plim}_{N \rightarrow \infty} G_N = G$ and $\text{plim}_{N \rightarrow \infty} F_N = F$). One interesting special case is obtained when we set $G_N = I_m$ and $F_N^{-1} = \frac{1}{N} \sum_{i=1}^N \frac{1}{T-1} \sum_{t=2}^T y_{i,t-1} y'_{i,t-1}$, as in this case we are testing the rank of the pooled OLS estimator $\hat{\Pi}$. Even though the estimator itself is inconsistent (due to the presence of the unobserved heterogeneity), as we show in this paper, it can be used for estimation of $\text{rk } \Pi_0$.

3.3 Discussions

In this section we summarize some of the underlying assumptions, and related problems, for the rk-J test.

Effect non-stationarity Recall that results in [Theorem 2](#) are written in terms of the rank of $\Delta y_{i,t} y'_{i,t-1}$ rather than Π . This suggests that if one uses this rank test to perform the sequential procedure in testing the rank of Π the procedure is “conservative”, i.e. as for some values of \mathcal{Y} , the jacobian can be of a reduced rank, even if Π is of full rank. However, the rank of $\Delta y_{i,t} y'_{i,t-1}$ can never be larger than the rank of Π . In such situations, the rk-J procedure controls the size of the test for Π , i.e. it rejects in at most $\alpha\%$ cases, and it never gets larger than the nominal level, thus this test controls the size uniformly over (Σ, \mathcal{Y}) . However, for some combinations of the nuisance parameters (Σ, \mathcal{Y}) , the power of this test does not converge to 1 as $N \rightarrow \infty$ even if $\text{rk}(\Pi) > r_0$, and as a result such testing procedure lacks power. Thus, it is difficult to draw general conclusions about the properties of the rk-J test when one does not reject the null hypothesis and it is likely that $y_{i,t}$ process is not effect stationary.

Common dynamics assumption Throughout the paper we maintain the common dynamics assumption for η_i . In the univariate case, it is known that if this assumption is satisfied the moment conditions are not relevant at unity, see a more detailed discussion in [Bun and Kleibergen \(2016\)](#). On the other hand, if the common dynamics assumption

is violated, it is possible to have a full rank jacobian even for $\Phi_0 = I_m$, see the aforementioned paper and the discussion in Hayakawa and Nagata (2016). Hence, even if Π matrix is of reduced rank $r < m$ the rank of the jacobian matrix can be of full rank m , when the common dynamics assumption is violated. In this case the rejection of the null hypothesis of the rk-J test, is not informative about the underlying rank of the Π .¹⁹

Initialization For initialization we assumed that $\text{var } \epsilon_{i,0}$ is well defined irrespective of the time-series properties of the data. In the univariate setting, it is known that if e.g. $E[\lim_{\phi \rightarrow 1} (1 - \phi)\epsilon_{i,0}^2] > 0$ then the Anderson and Hsiao (1982) moment conditions have a full rank jacobian matrix. Nevertheless, as discussed in Bun and Kleibergen (2016) this does not imply that ϕ parameter is identified.²⁰ Note that the initializations of this type would imply that the cross-sectional average of $y_{i,t}$ is not well defined for any $t \geq 0$ which is a rather unrealistic assumption to make.

These issues cannot be underestimated in empirical work. However, at the same time we acknowledge that in order to obtain testing procedures that controls size uniformly²¹ one would have to rely on procedures that are numerically challenging, i.e. subset inference using the continuously updated GMM estimator. We should also emphasize that most of the testing procedures for dynamic panel data (especially for persistent data) fail to guarantee uniform inference over the parameter space of autoregressive parameter and/or initialization of the initial condition.²²

4 Monte Carlo simulations

To the best of our knowledge only the BHP study provides results on cointegration analysis for panels with fixed T .²³ Hence, for the main building blocks of the finite-sample studies performed in this paper we take the setups from BHP, but we provide an extended set of scenarios. Only bivariate panels are considered, thus the only null hypothesis we are testing is:

$$H_0: \text{rk } \Pi = 1 \tag{18}$$

For simplicity we use (DGP.2) for initialization:

$$y_{i,0} = \Upsilon \mu_i + \epsilon_{i,0}, \quad \epsilon_{i,0} \sim N \left(\mathbf{0}_2, \sum_{j=0}^M \Phi^j \Sigma \Phi^{j'} \right). \tag{19}$$

¹⁹ However, in order to accommodate η_i that does not satisfy the common dynamics assumption the data generating process of the initial condition needs to be modified.

²⁰ For example if $\text{var } \epsilon_{i,0} = \sigma^2 / (1 - \phi^2)$.

²¹ As e.g. in Andrews and Cheng (2012).

²² For example in the panel unit root testing literature this issue is prominent, see discussion in Moon et al. (2007) and also Westerlund (2016).

²³ Other studies, like Mutl (2009) adapted setups of BHP.

$\varepsilon_{i,t}$ for all i are generated as follows

$$\varepsilon_{i,t} \sim N(\mathbf{0}_2, \Sigma), \quad t > 0.$$

We assume that the error terms are normally distributed i.i.d. both across individuals and time with zero mean and variance-covariance matrix Σ (to be specified later). We set $M = 50^{24}$ and the number of replications $B = 10000$.

We generate the individual heterogeneity (μ_i) using the exactly same procedure as in BHP:

$$\mu_i = \tau \left(\frac{q_i - 1}{\sqrt{2}} \right) \check{\eta}_i, \quad q_i \sim \chi^2(1), \quad \check{\eta}_i \sim N(\mathbf{0}_2, \Sigma), \quad (20)$$

where we set $\text{vech } \Sigma = (.05, .03, .05)'$. Following BHP, the variance-covariance matrix of $\check{\eta}_i$ coincides with the corresponding variance-covariance matrix used in generating $\varepsilon_{i,t}$.

Before summarizing the design parameters for this Monte Carlo study, recall that Π can be rewritten as (for $m = 2$):

$$\Pi = \alpha\beta' + \lambda\alpha_{\perp}\beta'_{\perp}$$

We set $\lambda = 0$ to study the size of the test, while non-zero values of λ are used to investigate power. In particular, the following values of λ are considered:

$$\lambda = \{-0.7; -0.3; -0.1; -0.05; -0.01; -0.005; 0.0\}. \quad (21)$$

In order to reduce the dimensionality of the parameter space we assume that vectors α and β are of the following structure:

$$\alpha = \alpha\iota_2, \quad \beta' = (1, -0.2),$$

and $\alpha = \{-0.1; -0.5\}$. Below we summarize main design parameters of this Monte Carlo study.

$$N = \{50; 250; 500\}, \quad T = \{3; 5; 7\}, \quad \tau = \{1; 5\}.$$

As we discussed in Sect. 3.2 in the effect non-stationary case the particular choice of $\{\Upsilon, \Sigma\}$ and τ might substantially influence the performance of the test statistic. To address these concerns the following five choices of Υ are considered:²⁵

$$\Upsilon = \left\{ 0.5I_2; I_2; 1.5I_2; I_2 - \Phi^{10}; \begin{pmatrix} .85 & .15 \\ .00 & .85 \end{pmatrix} \right\}.$$

²⁴ Results for $M = 5$ are qualitatively and quantitatively similar to the ones presented in this paper.

²⁵ Later we use notation $\Upsilon^{(q)}$ with q indicating the particular element of this set.

The choice of $\mathbf{Y}^{(4)}$ is motivated by the finite start-up assumption, so that the individual specific effects are accumulated only over 9 periods. The particular choice of $S = 10$ was rather arbitrary and is not empirically or theoretically motivated.²⁶

Comparing our setup to BHP, we can see that design 3 of BHP is achieved when $\alpha = -0.5$ and $\lambda = 0.0$ (as they consider size). In order to match our designs with the empirical application, we also considered $N = 750$, however the results are qualitatively and quantitatively similar to $N = 500$, thus omitted. Other design parameters are also chosen to match some of the properties of the empirical application, as $\mathbf{Y}^{(5)}$ is based on the estimates in Arellano (2016) obtained from the bivariate panel of Spanish firm data.

In terms of the test power, we suspect that it should be decreasing with $|\lambda|$, with almost no power against alternatives with $\lambda \approx 0$. However, it is very likely that for general \mathbf{Y} matrices the power curve might not be monotonic because λ not only controls the rank of $\mathbf{\Pi}$ but as well (indirectly) the eigenvalues of the $E[\mathbf{u}_{i,t-1}\mathbf{y}'_{i,t-1}]$ matrix. Hence, for some specific choices of \mathbf{Y} we can observe the weak instruments problem of Anderson and Hsiao (1982) moment conditions that is not caused by the reduced rank of $\mathbf{\Pi}$ matrix.

4.1 Results

The results for all designs are summarized at the top part of Tables 6, 7, and 8 ($\theta = 0$). All rejection frequencies are rounded up to two digits. Empty entries indicate maximal power of 1, 00.

General patterns First of all, we can observe that rejection frequencies are monotonically decreasing in $|\lambda|$ for the vast majority of designs without spatial dependence. As we discussed in Sect. 3.2 this property should not be taken as granted for the rk-J test (as dependence on Φ is non-linear). For lower values of N the test tends to be undersized for $T = 3$ and oversized for $T = 7$.²⁷ In the effect stationary case τ does not play substantial role and only affects the \mathbf{V} matrix, but we can still observe that higher value of τ is associated with slightly lower power. For $N = 500$, the rk-J test has notable power even when λ is very close to 0. For instance, all rejection frequencies in the effect stationary designs at $\lambda = 0.005$ are above 30% and 25% for $\alpha = -0.5$ and $\alpha = -0.1$ respectively. In the vast majority of cases with size distortions being of similar magnitude, the test power for $\alpha = -0.5$ tends to be higher than for $\alpha = -0.1$.

Effect non-stationarity and non-monotonic power curves First, we consider rejection frequencies for $\mathbf{Y} = 0.5 \times \mathbf{I}_m$ as this case is most exceptional in terms of observed patterns. In this case we observe power curves that are not monotonic for $\alpha = -0.1$ (especially for $N = 250$) and sharply decreasing for $\alpha = -0.5$ if $\tau = 5$ and $T = 3$. It can be intuitively explained as in this case the effect non-stationarity term in $E[\Delta\mathbf{y}_{i,t}\mathbf{y}'_{i,t-1}]$ is negative, driving the whole expression towards the zero matrix (recall the analysis in Hayakawa 2009 for the univariate case). Thus, we have

²⁶ Setting $S = 50$ would be another option, but it is of similar arbitrariness.

²⁷ As in this case orders of magnitude for N and T are not substantially different we suspect that critical values obtained as $N, T \rightarrow \infty$ (jointly) might be more appropriate.

a weak instrument problem under the alternative hypothesis that is not induced by cointegration.²⁸ By varying λ parameter we directly vary the relative contributions of time invariant and time varying parts of the variance components in $\text{var } y_{i,t}$. For larger values of $|\lambda|$ the time invariant part is more pronounced, resulting in substantial effects of the “negative” effect stationarity. On the other hand, for $|\lambda| \approx 0$ the idiosyncratic part is dominant and there is no substantial effects of the “negative” effect non-stationary initialization.

Remark 6 This non-monotonicity is further illustrated in Tables 6, 7, where we show how the minimum eigenvalue of the jacobian matrix changes for different nuisance parameters (for very larger N). Those patterns resemble power curves of the rk-J test as presented in Fig. 1.

As it can be expected, the results for $\Upsilon = 1.5 \times \mathbf{I}_m$ are more straightforward. In this case the power curves are monotonic, and rejection frequencies are uniformly dominating the ones from effect stationary case irrespective of other design parameters. Results for $\Upsilon^{(4)}$ seem to combine the properties of both $\Upsilon^{(3)}$ and $\Upsilon^{(1)}$.²⁹ Finally, the results of $\Upsilon^{(5)}$ are somewhat in between those of $\Upsilon^{(1)}$ and $\Upsilon^{(2)}$, but are slightly closer to $\Upsilon^{(2)}$. It serves as an indication that the off-diagonal element in $\Upsilon^{(5)}$ is not of any great importance (given the choice of other design parameters).

Remark 7 In this paper, we do not provide extensive results for the TML estimator of Binder et al. (2005). The main reason for this (besides theoretical problems discussed in Sect. 2.2) is possibly bimodal log-likelihood function (see e.g. Calzolari and Magazzini 2012; Bun et al. 2016; Juodis 2016). For model with stable dynamics, Juodis (2016) presents several alternatives how one can choose the maximizer of the log-likelihood function from the set of local minimizers. Unfortunately, no results are available for non-stationary dynamics analyzed in this paper. Thus, in order to avoid the situation in which unintentionally test procedure based on the TML estimator performs sub-optimally, we present only some limited results, see Table 5. Results suggest that for alternatives close to the null hypothesis LR test has low power, as the critical value from the $\chi^2(1)$ distribution is too large. On the other hand, for some very distant alternative (where the rk-J test struggles to reject the null hypothesis), LR test has sizeable power.

Remark 8 As a robustness check in 6, we also consider model with spatial dependence in the error terms. Evidence of the uniform upward shift in the size can be observed when designs with spatial dependence are considered.

²⁸ Some preliminary MC results, not presented in this paper suggest that effect of τ in this setup is non-monotonic. In the sense that higher values of τ lead to increase of power rather than further decrease. At least for this particular design it seems that $\tau = 5$ represents the close to worst possible scenario as minimum is reached for $\tau \approx 6.2$.

²⁹ From $\Upsilon^{(1)}$ some non-monotonicities are inherited. Apart from that, the superior test power properties (as compared to the effect stationary case) of $\Upsilon^{(3)}$ are dominant. This combined behavior is due to the fact that $\Upsilon^{(4)}$ is changing with λ . In designs with λ substantially lower than 0 we have $\Upsilon^{(4)} \approx \mathbf{I}_m$, consecutively the weak instrument problem under alternative is less pronounced.

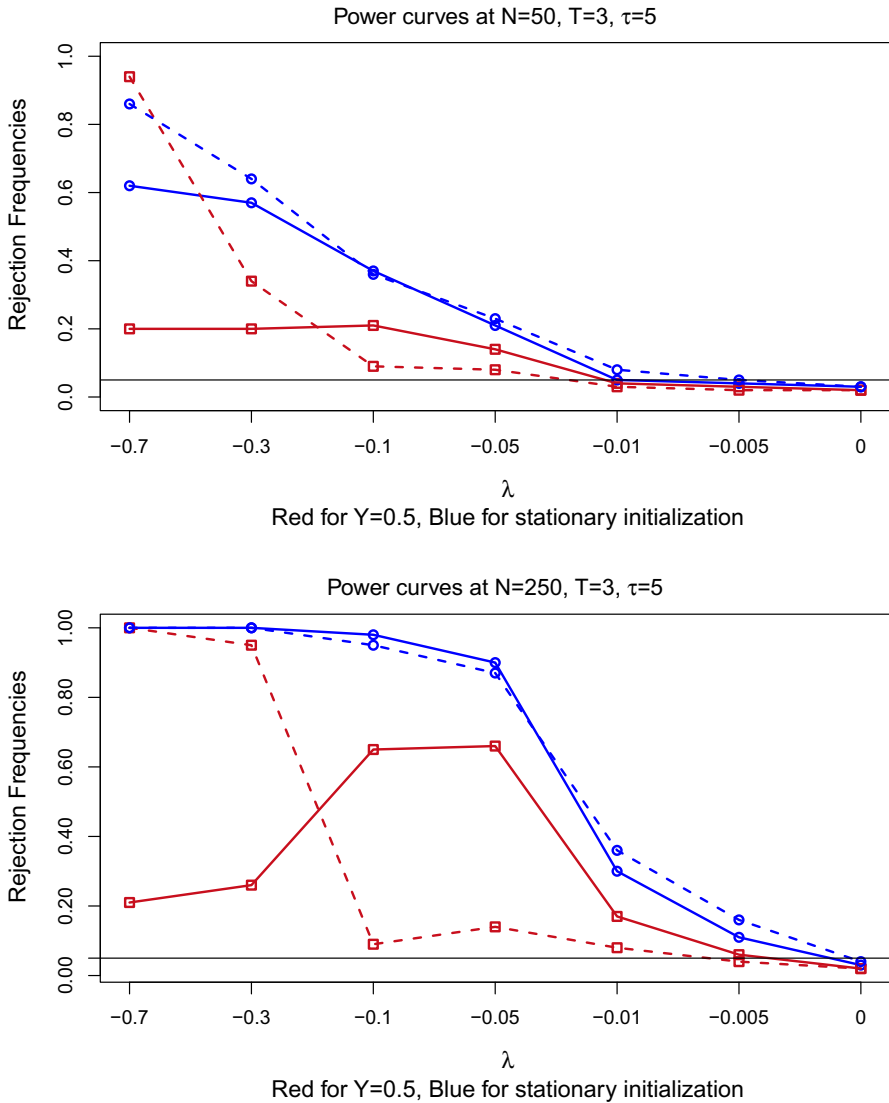


Fig. 1 Red (squares) $\Upsilon = 0.5I_2$, Blue (circles) $\Upsilon = I_2$. Straight line $\alpha = -0.1$. Dashed line $\alpha = -0.5$. (Color figure online)

5 Empirical illustration

5.1 Data

In this section, we analyze the Spanish firm panel dataset covering 1983–1990 of 738 manufacturing companies from [Alonso-Borrego and Arellano \(1999\)](#). This datasets constitutes a balanced panel of manufacturing companies recorded in the database of

Table 1 Descriptive statistics of the dependant variables

Year	log (employment)				log (wages)			
	Mean	Median	Min	Max	Mean	Median	Min	Max
1983	4.83	4.82	2.30	9.31	0.45	0.47	-0.89	1.32
1984	4.83	4.79	2.30	9.29	0.43	0.45	-1.13	1.29
1985	4.83	4.78	2.30	9.21	0.45	0.47	-0.91	1.33
1986	4.84	4.79	2.40	9.05	0.52	0.53	-0.92	1.55
1987	4.86	4.84	2.40	8.97	0.58	0.59	-0.65	1.67
1988	4.88	4.83	2.30	8.91	0.61	0.62	-0.77	1.74
1989	4.90	4.86	2.40	8.85	0.67	0.67	-0.62	1.83
1990	4.90	4.87	2.30	8.80	0.74	0.75	-0.75	1.90

the Bank of Spain's Central Balance Sheet Office from 1983 to 1990. As it contains data only for firms that were observed for the full time span and in all years satisfied specific coherency requirements, it cannot be considered as being a random sample from the population of all firms. For example, this dataset only contains firms that have majority private shareholding, thus state-owned companies are not represented. Thus all results need to be interpreted as conditional on the underlying characteristics used for sample selection.

We construct a bivariate PVAR(1) model with logarithms of employment and wages as dependent variables. Table 1 contains year specific descriptive statistics for these two variables.³⁰ Given that cross-sectional means for both variables differ substantially (especially that of wages) between the beginning and the end points, we follow other papers that considered this dataset (e.g. Arellano 2016) and include the time effects in the model, i.e. we consider variables in their deviations from the corresponding year specific cross-sectional averages. The sensitivity to the cross-sectional demeaning is discussed later in this section.

5.2 Results

In contrast to the previous studies that used this data, we investigate the time-series properties of this dataset in a greater detail. In particular, previous studies assumed that GMM and ML estimators are well-behaved, i.e. unit roots and cointegration were excluded a priori. However, estimation results in Arellano (2016) indicate that some estimated parameter values can be close to unit circle, thus non-stationary behaviour cannot be excluded beforehand. In order to elaborate on those observations, we consider a slightly larger set of estimators to obtain point estimates for Φ that are valid under different sets of assumption. The results in Table 2 are in line with those in Arellano (2016), with many close to unity point estimates of ϕ_{11} . The estimates for ϕ_{22} , on the other hand, are further away from unity, suggesting that both variables can be potentially cointegrated. In order to investigate for possible cointegration in this

³⁰ For a more detailed description of the data, please refer to Alonso-Borrego and Arellano (1999).

Table 2 Estimation results based on full sample

Estimator	ϕ_{11}	ϕ_{21}	ϕ_{12}	ϕ_{22}
AB (2)	0.86	-0.02	0.14	0.36
AB (1)	0.86	-0.03	0.12	0.28
Sys (2)	1.00	0.06	0.07	0.81
Sys (1)	0.99	0.05	0.07	0.81
FE	0.71	0.06	0.08	0.44
FEBC (HK)	0.98	0.02	0.14	0.62
FEBC (K, Sys(2))	1.02	0.02	0.08	0.77
FEBC (SPJ)	1.01	0.02	0.05	0.78
FEBC (BC)	1.05	-0.02	0.04	0.74
TMLE ($r = 1$)	1.00	0.00	0.07	0.68
TMLE ($r = 2$)	1.01	0.01	0.08	0.68

Here “AB(·)” and “Sys(·)” are the estimators of [Arellano and Bond \(1991\)](#) and [Blundell and Bond \(1998\)](#), respectively. The numbers in brackets indicate, whether these are “two-step” or “one-step” estimates. “FE” denotes the fixed effects estimator. “FEBC” (from top to the bottom) are the bias-correcting fixed effects estimators of [Hahn and Kuersteiner \(2002\)](#), [Kiviet \(1995\)](#) (using “Sys(2)” as the plug-in estimator), [Dhaene and Jochmans \(2015\)](#), [Bun and Carree \(2005\)](#), [Juodis \(2013\)](#). “TMLE” are the Transformed Maximum likelihood estimators with and without rank restrictions imposed on Φ

dataset, we make use of the rk-J procedure that was introduced earlier in this paper.³¹ Specifically, we test if there is a single cointegrating relationship between firms level employment and wages, i.e. $H_0 : r_0 = 1$.³²

First, we apply the rk test of [Kleibergen and Paap \(2006\)](#) directly to GMM estimates $\hat{\Pi}$. We restrict the set of GMM estimators to two step estimators that are also presented in BHP: “AB-GMM” stands for the estimator of [Arellano and Bond \(1991\)](#), while “Sys-GMM” is the estimator of [Blundell and Bond \(1998\)](#) that incorporates moment conditions in levels. Second, the LR tests based on the transformed maximum likelihood function of BHP (LR-TMLE) and maximum likelihood function of [Arellano \(2016\)](#) (as mentioned in Remark 1), (LR-RMLE), are considered. Finally, the “rk-J” test of Sect. 3.2 is considered. Under $H_0 : rk \Pi_0 = 1$, if no singularities in the corresponding asymptotic distributions are present, all tests have a $\chi^2(1)$ limit. Note that we present results for “AB-GMM” for informal comparison only, as under $H_0 : r_0 = 1$ this estimator is not consistent. Results are summarized in Table 3.

From Table 3 we can see that only the rk-J test based on the [Anderson and Hsiao \(1982\)](#) moment conditions rejects H_0 . Results for system GMM estimator are mixed, as based on [Windmeijer \(2005\)](#) corrected standard errors the null hypothesis is not rejected, while it is rejected when using the conventional two-step standard errors. Numerous reasons might account for differences in conclusions. First of all, we suspect

³¹ Other testing procedures are described below.

³² We focus only on testing $r = 1$ vs. $r = 2$, as e.g. using LR test based on the TML estimator the $H_0 : r_0 = 0$ is rejected against the alternative $H_A : r_0 = 1$ with the value of the test statistic equal to 117.561. This value is substantial even if the $\chi^2(1)$ does not provide a correct asymptotic approximation.

Table 3 Cointegration testing based on full sample

Name	Test statistic
AB-GMM	14.46 (7.20)
Sys-GMM	4.88** (1.31)
LR-TMLE	0.59
LR-RMLE	0.55
rk-J	13.35***

For GMM estimators test statistics based on Windmeijer (2005) corrected 2-step standard errors, are presented in parenthesis. To define significance we use the critical value from the $\chi^2(1)$ distribution. The 5% critical value is 3.84

* $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$

that the initialization moment conditions of the System estimator are not valid and it does not come as a surprise that this estimator fails to reject H_0 . Hayakawa and Nagata (2016) provide some evidence based on an incremental Sargan test in support of the latter statement.³³ Another explanation of results in Table 3 might be the low power of cointegration test used directly on the estimate of Π .

Now we turn our attention to the likelihood ratio tests. Based on analytical results in this paper for $T = 2$ we can suspect that the likelihood procedures under H_0 of cointegration lack power for close alternatives (recall limited MC results in Table 5), as $\chi^2(1)$ is a poor approximation of the finite sample distribution. Furthermore, we know that both likelihood methods are robust to violations of mean stationarity, but are not so to time-series heteroscedasticity. Thus, we can not rule out the possibility that it can be one of the reasons for divergence in conclusions.³⁴

5.3 Sub-sample analysis

In the previous section we investigated the relationships between the firm level employment and wages in the model estimated using the full length of the dataset. Using the rk-J procedure, no significant statistical evidence was found favouring cointegration between the two variables. In this section, we investigate the sensitivity of this conclusion to smaller values of T by means of the analysis over sub-samples of the original data. We also check the sensitivity towards inclusion/non-inclusion of the time effects in the model. The results of this section are summarized in Table 4, where in total 6 different sub-samples are considered. First of all, we observe that the non-inclusion of time effects leads to an increase of the test statistic in all sub-samples. The difference is especially pronounced for all subsamples with 1990 as the final year. Overall, irrespective of the time span considered the $rk - J$ statistic rejects the null hypothesis if the time-effects are not included, i.e. data is not cross-sectionally demeaned before estimation.

³³ However, this testing procedure cannot be used if series are cointegrated.

³⁴ Arellano (2003) presents some evidence of time-series heteroscedasticity in this dataset.

Table 4 Sub-sample $rk - J$ test

Years	T	Time effects	
		Yes	No
1983–1990	7	13.35***	29.01***
1983–1989	6	16.09***	35.04***
1983–1988	5	15.60***	28.35***
1983–1987	4	18.74***	20.14***
1984–1990	6	4.59**	27.19***
1985–1990	5	2.57	21.54***
1986–1990	4	0.79	15.94***

* $p < 0.1$; ** $p < 0.05$;

*** $p < 0.01$

The same cannot be generally said when data is cross-sectionally demeaned. Note how the value of test statistic increases as T increases for sub-sample ending in 1990. In particular, for $T = \{4; 5\}$ the null hypothesis is not rejected at any conventional significance level. This behavior emphasizes the value of additional time-series observations and possible lack of power for small values of T . As it can be seen from Table 4 the observations for 1983 are especially informative about the properties of the bivariate system, as for all sub-samples starting in 1983 the test statistic always rejects the null hypothesis.

Overall, omission of time-effects from the model does not affect the conclusions from Sect. 5.2. However, a moderate amount of time variation in the magnitude of test statistics suggests that this conclusion is sensitive to different estimation horizons.

6 Conclusions

In this paper, we study the properties of the standard Anderson and Hsiao (1982) moment conditions in a PVAR(1) for cointegrated processes. Under the assumptions similar to Binder et al. (2005) we show that these moment conditions are of reduced rank if the process is cointegrated. Based on this observation we propose a rank based test for the null hypothesis of cointegration. We prove that testing procedure in Binder et al. (2005) is invalid due to the singularity of the hessian matrix for persistent data. Monte Carlo results suggest that for most designs, the new test is reasonably sized and has good power properties but might exhibit non-monotonic power curves for models with substantial effect non-stationarity. We apply our testing procedure to the Spanish manufacturing data of Alonso-Borrego and Arellano (1999) and, unlike the test of BHP, we find no evidence of cointegration.

Acknowledgements This paper greatly benefited from comments made by two anonymous referees. Previous versions of this paper, were presented at the Tinbergen Institute, Netherlands Econometrics Study Group 2013 (Amsterdam) and “Conference on Cross-sectional Dependence in Panel Data” in Cambridge 2013. I would like to thank Ramon van den Akker, Peter Boswijk, Maurice Bun and Vasilis Sarafidis for their comments and suggestions.

Open Access This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution,

and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

Appendix

Proofs: Transformed maximum likelihood

Notation

The commutation matrix $\mathbf{K}_{a,b}$ is defined such that for any $[a \times b]$ matrix \mathbf{A} , $\text{vec}(\mathbf{A}') = \mathbf{K}_{a,b} \text{vec}(\mathbf{A})$. The duplication matrix \mathbf{D}_m is defined such that for symmetric $[a \times a]$ matrix $\text{vec } \mathbf{A} = \mathbf{D}_m \text{vech } \mathbf{A}$.

First, we define a set of new auxiliary variables, which are handy during the derivations of differentials:

$$\begin{aligned} \mathbf{Z}_N(\boldsymbol{\kappa}) &\equiv \frac{1}{N} \sum_{i=1}^N \sum_{t=1}^T (\tilde{\mathbf{y}}_{i,t} - \boldsymbol{\Phi} \tilde{\mathbf{y}}_{i,t-1})(\tilde{\mathbf{y}}_{i,t} - \boldsymbol{\Phi} \tilde{\mathbf{y}}_{i,t-1})', \\ \mathbf{Q}_N(\boldsymbol{\kappa}) &\equiv \frac{1}{N} \sum_{i=1}^N \sum_{t=1}^T \tilde{\mathbf{y}}_{i,t-1}(\tilde{\mathbf{y}}_{i,t} - \boldsymbol{\Phi} \tilde{\mathbf{y}}_{i,t-1})', \\ \mathbf{M}_N(\boldsymbol{\kappa}) &\equiv \frac{T}{N} \sum_{i=1}^N (\ddot{\mathbf{y}}_i - \boldsymbol{\Phi} \ddot{\mathbf{y}}_{i-})(\ddot{\mathbf{y}}_i - \boldsymbol{\Phi} \ddot{\mathbf{y}}_{i-})', \quad \mathbf{N}_N(\boldsymbol{\kappa}) \equiv \frac{T}{N} \sum_{i=1}^N \ddot{\mathbf{y}}_{i-}(\ddot{\mathbf{y}}_i - \boldsymbol{\Phi} \ddot{\mathbf{y}}_{i-})', \\ \mathbf{R}_N &\equiv \frac{1}{N} \sum_{i=1}^N \sum_{t=1}^T \tilde{\mathbf{y}}_{i,t-1} \tilde{\mathbf{y}}'_{i,t-1}, \quad \mathbf{P}_N \equiv \frac{T}{N} \sum_{i=1}^N \ddot{\mathbf{y}}_{i-} \ddot{\mathbf{y}}'_{i-}, \quad \boldsymbol{\Xi} \equiv \sum_{l=0}^{T-2} (T-1-l) \boldsymbol{\Phi}_0^l. \end{aligned}$$

Results from [Binder et al. \(2005\)](#) and [Juodis \(2016\)](#)

Define

$$\mathbf{W}_N(\boldsymbol{\kappa}) \equiv \boldsymbol{\Sigma}^{-1} \sum_{i=1}^N \sum_{t=1}^T (\tilde{\mathbf{y}}_{i,t} - \boldsymbol{\Phi} \tilde{\mathbf{y}}_{i,t-1}) \tilde{\mathbf{y}}'_{i,t-1} + T \boldsymbol{\Theta}^{-1} \sum_{i=1}^N (\ddot{\mathbf{y}}_i - \boldsymbol{\Phi} \ddot{\mathbf{y}}_{i-}) \ddot{\mathbf{y}}'_{i-},$$

then the score vector associated with the log-likelihood function (8) is given by:

$$\nabla(\boldsymbol{\kappa}) = \begin{pmatrix} \text{vec}(\mathbf{W}_N(\boldsymbol{\kappa})) \\ \mathbf{D}'_m \text{vec} \left(-\frac{N}{2} (\boldsymbol{\Sigma}^{-1} ((T-1)\boldsymbol{\Sigma} - \mathbf{Z}_N(\boldsymbol{\kappa}))\boldsymbol{\Sigma}^{-1}) \right) \\ \mathbf{D}'_m \text{vec} \left(-\frac{N}{2} (\boldsymbol{\Theta}^{-1} (\boldsymbol{\Theta} - \mathbf{M}_N(\boldsymbol{\kappa}))\boldsymbol{\Theta}^{-1}) \right) \end{pmatrix}. \quad (22)$$

For $0 < r < m$ we have the following result.

Corollary 1 *Let Assumptions TML be satisfied. Then the restricted score vector associated with the log-likelihood function (8) under cointegrating restrictions is given by:*

$$\nabla_r(\kappa_r) = \begin{pmatrix} \text{vec}(\mathbf{W}_N(\kappa_r)\boldsymbol{\beta}_r) \\ \text{vec}(\boldsymbol{\alpha}'_r\mathbf{W}_N(\kappa_r)\mathbf{H}_r) \\ \mathbf{D}'_m \text{vec}\left(-\frac{N}{2}(\boldsymbol{\Sigma}^{-1}((T-1)\boldsymbol{\Sigma} - \mathbf{Z}_N(\kappa_r))\boldsymbol{\Sigma}^{-1})\right) \\ \mathbf{D}'_m \text{vec}\left(-\frac{N}{2}(\boldsymbol{\Theta}^{-1}(\boldsymbol{\Theta} - \mathbf{M}_N(\kappa_r))\boldsymbol{\Theta}^{-1})\right) \end{pmatrix}. \tag{23}$$

Proof of this corollary is a trivial extension of the corresponding result in Juodis (2016) for the unrestricted model. In the special case where $m = 2$ and $r = 1$, $\boldsymbol{\delta}_1$ is a scalar while $\boldsymbol{\alpha}$ is a $[2 \times 1]$ vector implying that the corresponding entries of the $\nabla_r(\kappa_r)$ vector do not have a $\text{vec}(\cdot)$ operator in them.

Identification with unit roots

Proof of Theorem 1 In order to evaluate the expected hessian matrix, we need to separately calculate the expected value for every term presented in the formula for the second differential at the DGP values κ_0 . First of all, we note that $E[(T-1)\boldsymbol{\Sigma}_0 - \mathbf{Z}_N(\kappa_0)] = \mathbf{O}_m$ as well as $E[\boldsymbol{\Theta}_0 - \mathbf{M}_N(\kappa_0)] = \mathbf{O}_m$, hence the contribution to the hessian matrix of the first four terms in the expression for $d^2\ell(\kappa)$ is zero. Following, e.g. Juodis (2013) it can be shown:

$$E[\boldsymbol{Q}_N(\kappa_0)] = -\frac{1}{T} \left(\sum_{l=0}^{T-2} (T-1-l)\boldsymbol{\Phi}_0^l \right) \boldsymbol{\Sigma}_0. \tag{24}$$

Due to the fact that exact expressions for other terms are rather messy in the general case, they are derived only for the particular case where $\boldsymbol{\Phi}_0 = \mathbf{I}_m$. Under this assumption and assumptions of Binder et al. (2005), it then follows that $\boldsymbol{\Theta}_0 = \boldsymbol{\Sigma}_0$. The DGP in this case under restriction of common dynamics, simplifies to:

$$y_{i,t} = y_{i,t-1} + \boldsymbol{\varepsilon}_{i,t} = y_{i,0} + \sum_{k=1}^t \boldsymbol{\varepsilon}_{i,k}, \quad t \geq 1.$$

At first we consider expectations of the $N_N(\kappa_0)$ term which we can evaluate based on the general result derived in Lemma A.2 Juodis (2016):

$$\begin{aligned} E[N_N(\kappa_0)'] &= (\mathbf{I}_m - \boldsymbol{\Phi}_0) E[\mathbf{u}_{i,0}\mathbf{u}'_{i,0}] (\mathbf{I}_m - \boldsymbol{\Phi}_0)' \boldsymbol{\Xi}' + \frac{1}{T} \boldsymbol{\Sigma}_0 \boldsymbol{\Xi}' \\ &= \frac{1}{T} \boldsymbol{\Sigma}_0 \boldsymbol{\Xi}' = \frac{(T-1)T}{2T} \boldsymbol{\Sigma}_0 = \frac{T-1}{2} \boldsymbol{\Sigma}_0. \end{aligned}$$

It then similarly follows from the general result in Lemma A.1 Juodis (2016) that:

$$E[\mathbf{P}_N] = E \left[\frac{T}{N} \sum_{i=1}^N \ddot{y}_i - \dot{y}'_{i-} \right]$$

$$\begin{aligned}
 &= \frac{1}{T} \boldsymbol{\Sigma} (\mathbf{I}_m - \boldsymbol{\Phi}_0) \text{E} [\mathbf{u}_{i,0} \mathbf{u}'_{i,0}] (\mathbf{I}_m - \boldsymbol{\Phi}_0)' \boldsymbol{\Sigma}' \\
 &\quad + \frac{1}{T} \sum_{t=0}^{T-2} \left(\sum_{j=0}^t \boldsymbol{\Phi}_0^j \right) \boldsymbol{\Sigma}_0 \left(\sum_{j=0}^t \boldsymbol{\Phi}_0^j \right)' \\
 &= \frac{1}{T} \sum_{t=0}^{T-2} (t+1)^2 \boldsymbol{\Sigma}_0 = \frac{(T-1)(2T-1)}{6} \boldsymbol{\Sigma}_0.
 \end{aligned}$$

What is left is to evaluate the term involving \mathbf{R}_N :

$$\begin{aligned}
 \text{E}[\mathbf{R}_N] &= \text{E} \left[\sum_{t=0}^{T-1} \mathbf{y}_{i,t} \mathbf{y}'_{i,t} - T \bar{\mathbf{y}}_i \bar{\mathbf{y}}'_i \right] \\
 &= \text{E} \left[\sum_{t=1}^{T-1} \left(\sum_{k=1}^t \boldsymbol{\varepsilon}_{i,k} \right) \left(\sum_{k=1}^t \boldsymbol{\varepsilon}_{i,k} \right)' \right. \\
 &\quad \left. - \frac{1}{T} \left(\sum_{k=0}^{T-2} (T-1-k) \boldsymbol{\varepsilon}_{i,1+k} \right) \left(\sum_{k=0}^{T-2} (T-1-k) \boldsymbol{\varepsilon}_{i,1+k} \right)' \right] \\
 &= \text{E} \left[\sum_{t=1}^{T-1} \left(\sum_{k=1}^t \boldsymbol{\varepsilon}_{i,k} \boldsymbol{\varepsilon}'_{i,k} \right) - \frac{1}{T} \left(\sum_{k=0}^{T-2} (T-1-k)^2 \boldsymbol{\varepsilon}_{i,1+k} \boldsymbol{\varepsilon}'_{i,1+k} \right) \right] \\
 &= \left(\sum_{t=1}^{T-1} \left(\sum_{k=1}^t 1 \right) - \frac{1}{T} \left(\sum_{k=0}^{T-2} (T-1-k)^2 \right) \right) \boldsymbol{\Sigma}_0 \\
 &= \left(\sum_{t=1}^{T-1} \left(t - \frac{t^2}{T} \right) \right) \boldsymbol{\Sigma}_0 = \left(\frac{T(T-1)}{2} - \frac{(2T-1)(T-1)}{6} \right) \boldsymbol{\Sigma}_0 \\
 &= \frac{(T+1)(T-1)}{6} \boldsymbol{\Sigma}_0.
 \end{aligned}$$

In [Juodis \(2016\)](#) it is shown that the hessian matrix is of the following form (if we neglect terms that evaluated at the true values $\boldsymbol{\kappa}_0$ have expectation \mathbf{O}):

$$\begin{aligned}
 &\mathcal{H}^N(\boldsymbol{\kappa}) \\
 &= -N \begin{pmatrix} \mathbf{R}_N \otimes \boldsymbol{\Sigma}^{-1} + \mathbf{P}_N \otimes \boldsymbol{\Theta}^{-1} & (\mathbf{Q}_N \boldsymbol{\Sigma}^{-1} \otimes \boldsymbol{\Sigma}^{-1}) \mathbf{D}_m & (N_N \boldsymbol{\Theta}^{-1} \otimes \boldsymbol{\Theta}^{-1}) \mathbf{D}_m \\ \mathbf{D}'_m (\boldsymbol{\Sigma}^{-1} \mathbf{Q}'_N \otimes \boldsymbol{\Sigma}^{-1}) & \frac{(T-1)}{2} \mathbf{D}'_m (\boldsymbol{\Sigma}^{-1} \otimes \boldsymbol{\Sigma}^{-1}) \mathbf{D}_m & \mathbf{O}_p \\ \mathbf{D}'_m (\boldsymbol{\Theta}^{-1} N'_N \otimes \boldsymbol{\Theta}^{-1}) & \mathbf{O}_p & \frac{1}{2} \mathbf{D}'_m (\boldsymbol{\Theta}^{-1} \otimes \boldsymbol{\Theta}^{-1}) \mathbf{D}_m \end{pmatrix}.
 \end{aligned} \tag{25}$$

Plugging in these expressions into the formula for the hessian and evaluating it at the true value κ_0 :

$$\mathcal{H}_\ell = \frac{(T-1)}{2} \begin{pmatrix} T(\Sigma_0 \otimes \Sigma_0^{-1}) & -(I_m \otimes \Sigma_0^{-1}) \mathbf{D}_m & (I_m \otimes \Sigma_0^{-1}) \mathbf{D}_m \\ -\mathbf{D}'_m(I_m \otimes \Sigma_0^{-1}) & \mathbf{D}'_m(\Sigma_0^{-1} \otimes \Sigma_0^{-1}) \mathbf{D}_m & \mathbf{O}_p \\ \mathbf{D}'_m(I_m \otimes \Sigma_0^{-1}) & \mathbf{O}_p & \frac{1}{T-1} \mathbf{D}'_m(\Sigma_0^{-1} \otimes \Sigma_0^{-1}) \mathbf{D}_m \end{pmatrix}. \tag{26}$$

Using the formula for the determinant of partitioned matrix, that:

$$|\mathcal{H}_\ell| \propto \left(|H| \times \left| \frac{1}{T-1} \mathbf{D}'_m(\Sigma_0^{-1} \otimes \Sigma_0^{-1}) \mathbf{D}_m \right| \times \left| \mathbf{D}'_m(\Sigma_0^{-1} \otimes \Sigma_0^{-1}) \mathbf{D}_m \right| \right), \tag{27}$$

where \propto is the proportionality sign. For future reference, define

$$\begin{aligned} H &= (E[\mathbf{R}_N] \otimes \Sigma_0^{-1} + E[\mathbf{P}_N] \otimes \Theta_0^{-1}) \\ &\quad - \left[\frac{1}{T-1} \left((E[\mathbf{Q}_N(\kappa_0)] \Sigma_0^{-1}) \otimes (E[\mathbf{Q}_N(\kappa_0)] \Sigma_0^{-1})' \right) \right. \\ &\quad \left. + \left((E[N_N(\kappa_0)] \Theta_0^{-1}) \otimes (E[N_N(\kappa_0)] \Theta_0^{-1})' \right) \right] \mathbf{K}_m \\ &\quad - \left[\frac{1}{T-1} \left((E[\mathbf{Q}_N(\kappa_0)] \Sigma_0^{-1} E[\mathbf{Q}_N(\kappa_0)]') \otimes \Sigma_0^{-1} \right) \right. \\ &\quad \left. + \left((E[N_N(\kappa_0)] \Theta_0^{-1} E[N_N(\kappa_0)]') \otimes \Theta_0^{-1} \right) \right]. \end{aligned}$$

Thus, H in this case given by:

$$\begin{aligned} H &= T \left((\Sigma_0 \otimes \Sigma_0^{-1}) - (I_m \otimes \Sigma_0^{-1}) \mathbf{D}_m \mathbf{D}_m^+ (\Sigma_0 \otimes \Sigma_0) (\mathbf{D}_m \mathbf{D}_m^+)' (I_m \otimes \Sigma_0^{-1}) \right) \\ &= T \left((\Sigma_0 \otimes \Sigma_0^{-1}) - (I_m \otimes \Sigma_0^{-1}) \mathbf{D}_m \mathbf{D}_m^+ (\Sigma_0 \otimes \Sigma_0) (I_m \otimes \Sigma_0^{-1}) \right) \\ &= T \left((\Sigma_0 \otimes \Sigma_0^{-1}) - \frac{1}{2} (I_m \otimes \Sigma_0^{-1}) (I_m + \mathbf{K}_m) (\Sigma_0 \otimes \Sigma_0) (I_m \otimes \Sigma_0^{-1}) \right) \\ &= \frac{T}{2} \left((\Sigma_0 \otimes \Sigma_0^{-1}) - (I_m \otimes \Sigma_0^{-1}) \mathbf{K}_m (\Sigma_0 \otimes \Sigma_0) (I_m \otimes \Sigma_0^{-1}) \right) \\ &= \frac{T}{2} \left((\Sigma_0 \otimes \Sigma_0^{-1}) - \mathbf{K}_m \right). \end{aligned}$$

Then by means of the matrix determinant Lemma:

$$\begin{aligned} |H| &\propto |I_{m^2} - (I_m \otimes \Sigma_0^{-1}) \mathbf{K}_m (\Sigma_0 \otimes I_m)| |\mathbf{K}_m| \\ &\propto |I_{m^2} - (I_m \otimes \Sigma_0^{-1}) \mathbf{K}_m (\Sigma_0 \otimes I_m)| \\ &\propto |I_{m^2} - \mathbf{K}_m| = 0, \end{aligned}$$

where in the first line we used the fact that $\mathbf{K}_m = \mathbf{K}_m^{-1}$ and the second line follows from the fact that $|\mathbf{K}_m| = (-1)^{0.5m(m-1)}$, while $|I_{m^2} - \mathbf{K}_m| = 0$ follows trivially

from the fact that $\text{rk}(\mathbf{I}_{m^2} - \mathbf{K}_m) = 0.5m(m - 1)$. Hence we have proved that the \mathcal{H}_ℓ matrix is not invertible.

Cointegration and $T = 2$

Proof of Proposition 1 Using derivations of previous section it follows that:

$$\begin{aligned} E[\mathbf{P}_N] &= E[\mathbf{R}_N] = \frac{1}{2}\boldsymbol{\Psi}_0 \\ E[\mathbf{N}_N(\boldsymbol{\kappa}_0)] &= \frac{1}{2}\boldsymbol{\Theta}_0 \\ E[\mathbf{Q}_N(\boldsymbol{\kappa}_0)] &= -\frac{1}{2}\boldsymbol{\Sigma}_0 \end{aligned}$$

Corresponding \mathbf{H} matrix:

$$\begin{aligned} \mathbf{H} &= \frac{1}{2}\boldsymbol{\Psi}_0 \otimes (\boldsymbol{\Theta}_0^{-1} + \boldsymbol{\Sigma}_0^{-1}) \\ &\quad - \frac{1}{2}(\mathbf{I}_m \otimes \boldsymbol{\Sigma}_0^{-1}) \mathbf{D}_m \mathbf{D}_m^+ (\boldsymbol{\Sigma}_0 \otimes \boldsymbol{\Sigma}_0) (\mathbf{D}_m \mathbf{D}_m^+)' (\mathbf{I}_m \otimes \boldsymbol{\Sigma}_0^{-1}) \\ &\quad - \frac{1}{2}(\mathbf{I}_m \otimes \boldsymbol{\Theta}_0^{-1}) \mathbf{D}_m \mathbf{D}_m^+ (\boldsymbol{\Theta}_0 \otimes \boldsymbol{\Theta}_0) (\mathbf{D}_m \mathbf{D}_m^+)' (\mathbf{I}_m \otimes \boldsymbol{\Theta}_0^{-1}) \\ &= \frac{1}{2}\boldsymbol{\Psi}_0 \otimes (\boldsymbol{\Theta}_0^{-1} + \boldsymbol{\Sigma}_0^{-1}) - \frac{1}{2}(\boldsymbol{\Sigma}_0 \otimes \mathbf{I}_m) (\mathbf{D}_m \mathbf{D}_m^+)' (\mathbf{I}_m \otimes \boldsymbol{\Sigma}_0^{-1}) \\ &\quad - \frac{1}{2}(\boldsymbol{\Theta}_0 \otimes \mathbf{I}_m) (\mathbf{D}_m \mathbf{D}_m^+)' (\mathbf{I}_m \otimes \boldsymbol{\Theta}_0^{-1}) \\ &\propto \boldsymbol{\Psi}_0 \otimes (\boldsymbol{\Theta}_0^{-1} + \boldsymbol{\Sigma}_0^{-1}) - \mathbf{K}_m - \frac{1}{2}(\boldsymbol{\Theta}_0 \otimes \boldsymbol{\Theta}_0^{-1}) - \frac{1}{2}(\boldsymbol{\Sigma}_0 \otimes \boldsymbol{\Sigma}_0^{-1}) \\ &\propto (\boldsymbol{\Sigma}_0 \otimes \boldsymbol{\Theta}_0^{-1} - \mathbf{K}_m) + (\boldsymbol{\Theta}_0 \otimes \boldsymbol{\Sigma}_0^{-1} - \mathbf{K}_m). \end{aligned}$$

We can express $\boldsymbol{\Theta}_0 = \boldsymbol{\Sigma}_0 + \mathbf{a}\mathbf{a}'$ with \mathbf{a} being of rank r . Then:

$$\begin{aligned} \mathbf{H} &\propto (\mathbf{I}_m \otimes \boldsymbol{\Theta}_0^{-1}) (\boldsymbol{\Sigma}_0 \otimes \boldsymbol{\Sigma}_0 + \boldsymbol{\Theta}_0 \otimes \boldsymbol{\Theta}_0 - 2(\boldsymbol{\Sigma}_0 \otimes \boldsymbol{\Theta}_0) \mathbf{K}_m) (\mathbf{I}_m \otimes \boldsymbol{\Sigma}_0^{-1}) \\ &= (\mathbf{I}_m \otimes \boldsymbol{\Theta}_0^{-1}) (2(\boldsymbol{\Sigma}_0 \otimes \boldsymbol{\Sigma}_0)(\mathbf{I}_{m^2} - \mathbf{K}_m) \\ &\quad + (\boldsymbol{\Sigma}_0 \otimes \mathbf{a}\mathbf{a}')(\mathbf{I}_{m^2} - \mathbf{K}_m) + \mathbf{a}\mathbf{a}' \otimes \mathbf{a}\mathbf{a}') (\mathbf{I}_m \otimes \boldsymbol{\Sigma}_0^{-1}) \\ &= (\mathbf{I}_m \otimes \boldsymbol{\Theta}_0^{-1}) ((2\boldsymbol{\Sigma}_0 \otimes \boldsymbol{\Sigma}_0 + \boldsymbol{\Sigma}_0 \otimes \mathbf{a}\mathbf{a}')(\mathbf{I}_{m^2} - \mathbf{K}_m) + \mathbf{a}\mathbf{a}' \otimes \mathbf{a}\mathbf{a}') (\mathbf{I}_m \otimes \boldsymbol{\Sigma}_0^{-1}). \end{aligned}$$

Thus using basic inequalities for the rank of matrix sum we can deduce that $\text{rk } \mathbf{H}$ is no greater than $0.5m(m - 1) + r^2$, due to the full rank of $(2\boldsymbol{\Sigma}_0 \otimes \boldsymbol{\Sigma}_0 + \boldsymbol{\Sigma}_0 \otimes \mathbf{a}\mathbf{a}')$ matrix.

Monte Carlo results

Robustness check: spatial dependence

To allow for possible spatial dependence, the $\boldsymbol{\varepsilon}_{i,t}$ for all t are generated with Spatial MA process:

$$\begin{aligned}\boldsymbol{\varepsilon}_{i,t} &= \theta \sum_{j=1}^N \omega_{i,j} \boldsymbol{\zeta}_{j,t} + \boldsymbol{\zeta}_{i,t}, \quad t > 0 \\ \boldsymbol{\zeta}_{i,t} &\sim N(\mathbf{0}_2, \boldsymbol{\Sigma}), \\ \boldsymbol{\zeta}_{i,0} &\sim N\left(\mathbf{0}_2, \sum_{j=0}^M \boldsymbol{\Phi}^j \boldsymbol{\Sigma} \boldsymbol{\Phi}^{j'}\right).\end{aligned}$$

The θ parameter controls the degree of cross-sectional dependence between units. For $\theta = 0$ we have i.i.d. dataset, while for $\theta \neq 0$ the cross-sectional units are *weakly* correlated. To illustrate the effect of spatial dependence we consider one value for θ , namely $\theta = 0.5$. Spatial correlation matrix \mathbf{W}_N with a typical (i, j) element given by $\omega_{i,j}$ is assumed to be **1 ahead - 1 behind** circular, so that every individual i is directly linked only with individuals $i - 1$ and $i + 1$.³⁵ The particular choice of the spatial matrix \mathbf{W}_N is motivated by the study in Baltagi et al. (2007), where in the context of the panel unit root testing it is shown that the tests are mostly distorted for this choice of spatial matrix.³⁶ The results are presented at the bottom of Tables 6, 7, and 8.

Evidence of the uniform upward shift in the size can be observed when designs with spatial dependence ($\theta = 0.5$) are considered. This upward movement does not come as a surprise because similar patterns have been documented in the panel unit root testing literature. However, the same conclusion can not be reached regarding the test power, as for most scenarios it changes marginally and does not show any clear patterns in terms of magnitude and direction. More importantly, major size distortions *do* not disappear for $N = 500$, thus as can be expected the fact that we use the variance-covariance matrix that ignores the presence of spatial dependence has a pronounced result. On the other hand, the fact that by ignoring the spatial dependence the rank based test statistic is only mildly oversized, suggests that the procedure developed in this paper is relatively robust to deviations from the i.i.d. assumption.

Tables

See Tables 5, 6, 7, 8, 9, 10, and 11.

³⁵ The circle is closed by connecting $i = 1$ with $i = N$.

³⁶ For a graphical illustration see Figure 2 of the aforementioned paper.

Table 5 MC results for LR test based on TML

Test	$\{T, \tau\} \setminus \lambda$	-0.70	-0.30	-0.10	-0.05	-0.01	-0.005	.00
rk-J	{3; 1}					0.90	0.49	0.04
	{3; 5}	0.27	0.35	0.94	0.97	0.70	0.27	0.03
	{5; 1}						0.77	0.03
	{5; 5}		0.91	0.98		0.93	0.59	0.05
	{7; 1}						0.93	0.05
	{7; 5}			0.99	0.99	0.98	0.78	0.04
	{3; 1}	0.96	0.65	0.04	0.01	0.01	0.01	0.01
TML	{3; 5}		0.88	0.06	0.02	0.01	0.01	0.01
	{5; 1}		0.95	0.13	0.03	0.03	0.02	0.02
	{5; 5}			0.14	0.05	0.03	0.03	0.03
	{7; 1}			0.46	0.08	0.01	0.01	0.01
	{7; 5}			0.51	0.09	0.01	0.02	0.02

Results are based on 500 replications. Here $\alpha = -0.1$, while $\mathcal{Y}^{(1)}$ is used for initialization and $N = 750$. Empty entries correspond to 100% empirical rejection frequencies. As starting values for TML estimators we use the bias corrected Fixed Effects estimators as define in Table 2

Table 6 Eigenvalues

α	$\mathcal{R}^{(q)}$	$\{T, \tau\} \setminus \lambda$	-0.70	-0.30	-0.10	-0.05	-0.01	-0.005	0.00
-0.1	1	{3; 1}	0.07	0.06	0.04	0.03	0.02	0.02	0.00
		{3; 5}	0.03	0.03	0.03	0.02	0.02	0.01	0.00
		{5; 1}	0.08	0.06	0.04	0.03	0.02	0.02	0.00
		{5; 5}	0.07	0.04	0.03	0.03	0.02	0.01	0.00
		{7; 1}	0.08	0.06	0.04	0.03	0.02	0.02	0.00
		{7; 5}	0.09	0.05	0.03	0.03	0.02	0.01	0.00
	2	{3; 1}	0.08	0.06	0.04	0.04	0.02	0.02	0.00
		{3; 5}	0.08	0.06	0.04	0.04	0.02	0.02	0.00
		{5; 1}	0.08	0.06	0.04	0.04	0.03	0.02	0.00
		{5; 5}	0.08	0.06	0.04	0.04	0.02	0.02	0.00
		{7; 1}	0.08	0.06	0.04	0.04	0.03	0.02	0.00
		{7; 5}	0.08	0.06	0.04	0.04	0.02	0.02	0.00
	3	{3; 1}	0.10	0.08	0.05	0.04	0.03	0.02	0.00
		{3; 5}	0.17	0.13	0.08	0.06	0.02	0.01	0.00
		{5; 1}	0.09	0.07	0.05	0.04	0.03	0.02	0.00
		{5; 5}	0.14	0.12	0.08	0.06	0.02	0.01	0.00
		{7; 1}	0.09	0.07	0.05	0.04	0.03	0.02	0.00
		{7; 5}	0.12	0.11	0.08	0.06	0.02	0.01	0.00
	1	{3; 1}	0.13	0.05	0.03	0.03	0.02	0.02	0.00
		{3; 5}	0.18	0.06	0.00	0.00	0.01	0.00	0.00
		{5; 1}	0.14	0.06	0.03	0.03	0.02	0.02	0.00
		{5; 5}	0.21	0.07	0.02	0.01	0.00	0.00	0.00
		{7; 1}	0.14	0.06	0.03	0.03	0.02	0.01	0.00
		{7; 5}	0.19	0.08	0.03	0.02	0.01	0.01	0.00
-0.5	2	{3; 1}	0.12	0.06	0.04	0.02	0.02	0.01	0.00
		{3; 5}	0.12	0.06	0.04	0.02	0.02	0.01	0.00
		{5; 1}	0.12	0.06	0.04	0.02	0.02	0.01	0.00
		{5; 5}	0.12	0.06	0.04	0.02	0.02	0.01	0.00
		{7; 1}	0.12	0.06	0.04	0.02	0.02	0.01	0.00
		{7; 5}	0.12	0.06	0.04	0.02	0.02	0.01	0.00
	3	{3; 1}	0.15	0.08	0.03	0.02	0.02	0.01	0.00
		{3; 5}	0.27	0.12	0.06	0.04	0.02	0.01	0.00
		{5; 1}	0.13	0.07	0.03	0.02	0.02	0.01	0.00
		{5; 5}	0.17	0.09	0.05	0.04	0.02	0.01	0.00
		{7; 1}	0.13	0.06	0.03	0.02	0.02	0.01	0.00
		{7; 5}	0.14	0.07	0.04	0.03	0.02	0.01	0.00

The minima of modulus' of eigenvalues of $\Delta y_{i,t} y'_{i,t-1T}$ for $N = 200,000$

Table 7 Eigenvalues

α	$\mathcal{R}^{(q)}$	$\{T, \tau\} \setminus \lambda$	-0.70	-0.30	-0.10	-0.05	-0.01	-0.005	0.00
-0.1	1	{3; 1}	0.18	0.31	0.51	0.51	0.46	0.46	0.45
		{3; 5}	0.04	0.09	0.43	0.40	0.29	0.28	0.27
		{5; 1}	0.42	0.66	1.03	1.01	0.93	0.92	0.91
		{5; 5}	0.41	0.43	0.99	0.81	0.59	0.57	0.54
		{7; 1}	0.68	1.05	1.57	1.52	1.39	1.38	1.37
		{7; 5}	0.92	1.09	1.58	1.25	0.91	0.87	0.84
	2	{3; 1}	0.22	0.36	0.54	0.53	0.50	0.50	0.50
		{3; 5}	0.22	0.36	0.54	0.53	0.50	0.50	0.50
		{5; 1}	0.43	0.72	1.08	1.07	1.01	1.00	1.00
		{5; 5}	0.44	0.72	1.08	1.07	1.01	1.01	1.01
		{7; 1}	0.65	1.09	1.61	1.60	1.51	1.51	1.51
		{7; 5}	0.65	1.09	1.62	1.60	1.51	1.51	1.51
	3	{3; 1}	0.28	0.46	0.61	0.61	0.61	0.61	0.62
		{3; 5}	0.46	0.74	0.88	0.90	1.01	1.04	1.07
		{5; 1}	0.49	0.85	1.20	1.21	1.20	1.21	1.22
		{5; 5}	0.70	1.29	1.72	1.78	1.95	2.00	2.05
		{7; 1}	0.68	1.22	1.79	1.81	1.78	1.79	1.80
		{7; 5}	0.85	1.73	2.52	2.64	2.83	2.89	2.95
	1	{3; 1}	0.36	0.15	0.09	0.09	0.09	0.09	0.09
		{3; 5}	0.50	0.25	0.03	0.02	0.04	0.04	0.05
		{5; 1}	0.78	0.31	0.20	0.19	0.19	0.19	0.19
		{5; 5}	1.18	0.64	0.15	0.05	0.01	0.01	0.01
		{7; 1}	1.02	0.48	0.31	0.30	0.30	0.31	0.31
		{7; 5}	1.61	0.88	0.27	0.15	0.10	0.10	0.10
-0.5	2	{3; 1}	0.30	0.16	0.12	0.12	0.12	0.12	0.12
		{3; 5}	0.30	0.16	0.12	0.12	0.12	0.12	0.12
		{5; 1}	0.60	0.31	0.24	0.24	0.24	0.24	0.24
		{5; 5}	0.60	0.31	0.24	0.24	0.24	0.24	0.24
		{7; 1}	0.90	0.47	0.36	0.35	0.36	0.36	0.36
		{7; 5}	0.90	0.47	0.36	0.36	0.36	0.36	0.36
	3	{3; 1}	0.36	0.22	0.18	0.18	0.18	0.18	0.18
		{3; 5}	0.75	0.60	0.44	0.42	0.41	0.41	0.41
		{5; 1}	0.65	0.41	0.33	0.32	0.32	0.32	0.32
		{5; 5}	0.96	1.04	0.74	0.68	0.63	0.63	0.62
		{7; 1}	0.93	0.55	0.46	0.45	0.44	0.44	0.44
		{7; 5}	1.18	1.20	0.95	0.85	0.79	0.78	0.78

The minima of modulus' of eigenvalues of $\overline{\mathbf{u}_{i,t-1} \mathbf{y}'_{i,t-1T}}$ for $N = 200,000$

Table 8 Monte Carlo results for $\alpha = -0.1, \Upsilon^{(1)-(3)}$

θ	$\{N, T, \tau\} \backslash \lambda$	$\Upsilon^{(1)}$					$\Upsilon^{(2)}$					$\Upsilon^{(3)}$											
		-.70	-.30	-.10	-.05	-.01	-.0005	.00	.70	.30	.10	.05	.01	.0005	.00	.70	.30	.10	.05	.01	.0005	.00	
0.0	50, 3, 1	.83	.81	.63	.36	.07	.04	.03	.90	.86	.64	.37	.07	.04	.03	.94	.90	.71	.42	.08	.05	.03	.03
	50, 3, 5	.20	.20	.21	.14	.04	.03	.02	.62	.57	.37	.21	.05	.04	.03	.94	.88	.69	.43	.09	.05	.03	.03
	50, 5, 1	.98	.95	.91	.75	.16	.08	.04	.99	.93	.93	.76	.17	.08	.04	.99	.96	.81	.19	.09	.04	.04	.04
	50, 5, 5	.68	.40	.40	.33	.08	.04	.03	.88	.83	.69	.51	.12	.07	.03	.99	.93	.79	.20	.10	.04	.04	.04
	50, 7, 1	.97	.97	.97	.92	.29	.12	.05	.99	.99	.93	.30	.13	.05	.05	.95	.31	.14	.05	.05	.05	.05	.05
	50, 7, 5	.97	.63	.52	.48	.15	.07	.04	.95	.92	.84	.71	.22	.11	.05	.98	.93	.31	.14	.05	.05	.05	.05
	250, 3, 1	.99			.41	.16	.04				.99	.41	.16	.04		.44	.17	.04					
	250, 3, 5	.21	.26	.65	.66	.17	.06	.02	.98	.90	.30	.11	.03	.11	.03	.99	.44	.17	.04				
	250, 5, 1	.97	.65	.80	.85	.46	.17	.03			.77	.34	.05	.34	.05	.78	.35	.05					
	250, 5, 5	.97	.65	.80	.85	.46	.17	.03			.65	.28	.04	.28	.04	.76	.34	.05					
	250, 7, 1				.92	.51	.05	.05			.92	.51	.05	.51	.05	.93	.52	.05					
	250, 7, 5		.94	.88	.90	.66	.29	.04			.84	.43	.05	.43	.05	.91	.49	.05					
	500, 3, 1				.75	.32	.04				.75	.33	.04	.33	.04	.77	.34	.05					
	500, 3, 5	.24	.31	.85	.91	.45	.16	.03			.62	.26	.04	.26	.04	.75	.33	.05					
	500, 5, 1				.97	.62	.05	.05			.97	.61	.05	.61	.05	.97	.62	.05					
	500, 5, 5		.80	.93	.96	.79	.39	.03			.93	.53	.05	.53	.05	.97	.60	.05					
	500, 7, 1				.80	.05		.05			.80	.05		.80	.05	.81	.05						
	500, 7, 5		.99	.97	.98	.90	.58	.05			.99	.72	.05	.72	.05	.78	.34	.05					

Table 8 continued

θ	$\{N, T, \tau\} \setminus \lambda$	$\mathcal{Y}^{(1)}$					$\mathcal{Y}^{(2)}$					$\mathcal{Y}^{(3)}$											
		-.70	-.30	-.10	-.05	-.01	-.0005	.00	-.70	-.30	-.10	-.05	-.01	-.0005	.00	-.70	-.30	-.10	-.05	-.01	-.0005	.00	
0.5	50, 3, 1	.84	.82	.63	.37	.09	.07	.05	.89	.85	.64	.37	.09	.06	.05	.93	.88	.69	.41	.10	.07	.05	.00
	50, 3, 5	.27	.27	.27	.18	.06	.05	.04	.67	.61	.40	.24	.07	.05	.04	.93	.86	.67	.41	.10	.07	.05	.00
	50, 5, 1	.99	.97	.91	.74	.20	.11	.07	.99	.99	.93	.75	.20	.11	.07	.99	.95	.79	.21	.12	.07	.00	.00
	50, 5, 5	.75	.50	.49	.39	.11	.07	.05	.91	.86	.73	.53	.15	.09	.06	.99	.93	.77	.21	.12	.07	.00	.00
	50, 7, 1	.98	.72	.63	.56	.19	.11	.08	.97	.94	.87	.75	.25	.14	.08	.99	.94	.33	.17	.08	.00	.00	.00
	50, 7, 5	.40	.47	.80	.78	.24	.11	.04	.97	.98	.92	.75	.25	.14	.07	.99	.92	.32	.17	.08	.00	.00	.00
	250, 3, 1				.99	.42	.19	.06			.99	.42	.19	.07									
	250, 3, 5	.40	.47	.80	.78	.24	.11	.04	.97	.98	.92	.75	.25	.14	.07	.99	.92	.32	.17	.08	.00	.00	.00
	250, 5, 1				.75	.36	.08				.74	.36	.08										
	250, 5, 5	.98	.81	.91	.93	.55	.23	.05	.90	.90	.92	.33	.15	.05	.08	.99	.94	.20	.07	.00	.00	.00	.00
	250, 7, 1				.90	.52	.09				.90	.51	.09										
	250, 7, 5		.97	.95	.96	.74	.38	.07			.84	.45	.08										
	500, 3, 1				.72	.34	.08				.72	.34	.08										
	500, 3, 5	.50	.61	.95	.97	.54	.22	.06	.95	.95	.63	.28	.07	.08	.08	.99	.94	.33	.17	.08	.00	.00	.00
	500, 5, 1				.95	.61	.08				.95	.60	.08										
	500, 5, 5		.92	.98	.99	.86	.47	.07	.92	.92	.54	.08	.08	.08	.08	.99	.94	.20	.07	.00	.00	.00	.00
	500, 7, 1				.77	.08					.77	.08											
	500, 7, 5		.99	.99	.99	.96	.65	.07	.99	.99	.99	.71	.08	.08	.08	.99	.94	.20	.07	.00	.00	.00	.00

$\theta = 0$ corresponds to the model without spatial dependence (Sect. 4, while $\theta = 0.5$ the model with spatial dependence of circular form from 6. Empty entries correspond to 100% empirical rejection frequencies

Table 9 Monte Carlo results for $\alpha = -0.1$. $\Upsilon^{(4),(6)}$

θ	$\{N, T, \tau\} \backslash \lambda$	$\Upsilon^{(5)}$													
		-0.70	-0.30	-0.10	-0.05	-0.01	-0.005	.00	-0.70	-0.30	-0.10	-0.05	-0.01	-0.005	.00
0.0	50, 3, 1	.92	.89	.66	.38	.07	.04	.03	.88	.84	.64	.37	.07	.04	.03
	50, 3, 5	.79	.81	.52	.21	.04	.03	.03	.36	.39	.30	.19	.05	.03	.02
50, 5, 1	50, 5, 1	.99	.99	.94	.76	.17	.08	.04	.99	.97	.92	.76	.17	.08	.04
	50, 5, 5	.97	.98	.85	.50	.09	.05	.03	.66	.59	.56	.45	.11	.06	.03
50, 7, 1	50, 7, 1	.99	.99	.99	.93	.29	.12	.05	.99	.99	.98	.92	.30	.13	.05
	50, 7, 5	.99	.95	.95	.71	.17	.09	.04	.84	.73	.70	.63	.21	.10	.04
250, 3, 1	250, 3, 1					.43	.16	.04				.99	.41	.16	.04
	250, 3, 5				.94	.22	.08	.03	.74	.85	.92	.85	.29	.11	.03
250, 5, 1	250, 5, 1					.79	.35	.05				.77	.34	.05	.05
	250, 5, 5				.99	.52	.19	.04	.95	.93	.97	.97	.63	.27	.04
250, 7, 1	250, 7, 1					.93	.52	.05				.92	.51	.05	.05
	250, 7, 5					.71	.32	.04	.99	.97	.99	.99	.83	.42	.05
500, 3, 1	500, 3, 1					.77	.33	.04				.75	.32	.04	.04
	500, 3, 5					.50	.18	.03	.89	.96	.99	.99	.62	.25	.04
500, 5, 1	500, 5, 1					.98	.63	.05				.97	.61	.05	.05
	500, 5, 5					.81	.40	.04	.99	.99		.92	.52	.04	.04
500, 7, 1	500, 7, 1					.81	.81	.05				.80	.80	.05	.05
	500, 7, 5					.92	.60	.05				.99	.72	.05	.05

Table 9 continued

θ	$\{N, T, \tau\} \setminus \lambda$	$\mathbf{Y}^{(4)}$					$\mathbf{Y}^{(5)}$								
		-.70	-.30	-.10	-.05	-.01	-.0005	.00	-.70	-.30	-.10	-.05	-.01	-.0005	.00
0.5	50, 3, 1	.87	.84	.63	.37	.10	.06	.05	.87	.84	.63	.37	.10	.06	.05
	50, 3, 5	.44	.45	.35	.22	.07	.05	.04	.44	.45	.35	.22	.07	.05	.04
	50, 5, 1	.99	.98	.92	.74	.20	.11	.07	.99	.98	.92	.74	.20	.11	.07
	50, 5, 5	.75	.68	.63	.49	.15	.09	.06	.75	.68	.63	.49	.15	.09	.06
	50, 7, 1			.98	.92	.32	.16	.08			.98	.92	.32	.16	.08
	50, 7, 5	.89	.81	.77	.68	.24	.13	.07	.89	.81	.77	.68	.24	.13	.07
	250, 3, 1				.99	.42	.19	.07				.99	.42	.19	.07
	250, 3, 5	.87	.93	.95	.88	.32	.15	.05	.87	.93	.95	.88	.32	.15	.05
	250, 5, 1				.74	.74	.36	.08				.74	.74	.36	.08
	250, 5, 5	.98	.97	.99	.99	.64	.30	.07	.98	.97	.99	.99	.64	.30	.07
	250, 7, 1				.90	.90	.51	.09				.90	.90	.51	.09
	250, 7, 5		.99	.99	.83	.83	.45	.08		.99	.99	.83	.83	.45	.08
	500, 3, 1				.72	.72	.34	.08				.72	.72	.34	.08
	500, 3, 5	.97	.99		.99	.62	.28	.06		.99		.99	.62	.28	.06
	500, 5, 1				.95	.95	.60	.08				.95	.95	.60	.08
	500, 5, 5				.91	.91	.54	.08				.91	.91	.54	.08
	500, 7, 1				.77	.77	.77	.08				.77	.77	.77	.08
	500, 7, 5				.98	.98	.71	.08				.98	.98	.71	.08

See Table 8

Table 10 Monte Carlo results for $\alpha = -0.5$. $\Upsilon^{(1)-(3)}$

θ	$(N, T, \tau) \setminus \lambda$	$\Upsilon^{(1)}$					$\Upsilon^{(2)}$					$\Upsilon^{(3)}$									
		-.70	-.30	-.10	-.05	-.01	-.005	.00	-.70	-.30	-.10	-.05	-.01	-.005	.00	-.70	-.30	-.10	-.05	-.01	-.005
0.0	50, 3, 1	.99	.88	.59	.36	.08	.05	.04	.99	.97	.72	.45	.10	.06	.04	.84	.54	.12	.07	.03	
	50, 3, 5	.94	.34	.09	.08	.03	.02	.02	.86	.64	.36	.23	.08	.05	.03	.97	.83	.54	.13	.08	.04
	50, 5, 1	.99	.91	.79	.25	.13	.06	.97	.85	.29	.15	.07	.99	.89	.30	.16	.06				
	50, 5, 5	.88	.26	.21	.08	.05	.03	.97	.89	.74	.59	.22	.13	.06	.97	.99	.85	.27	.14	.06	
	50, 7, 1	.98	.94	.43	.23	.10	.97	.45	.24	.10	.98	.45	.23	.09	.98	.45	.23	.09			
	50, 7, 5	.99	.54	.40	.18	.10	.06	.97	.90	.80	.37	.21	.09	.95	.99	.98	.95	.40	.21	.07	
	250, 3, 1	.99	.98	.42	.16	.03	.99	.47	.20	.04	.49	.22	.05								
	250, 3, 5	.95	.09	.14	.08	.04	.02	.95	.87	.36	.16	.04	.47	.21	.05						
	250, 5, 1	.82	.42	.06	.82	.42	.06	.82	.42	.06	.82	.42	.06	.82	.42	.05					
	250, 5, 5	.56	.24	.13	.06	.02	.99	.72	.37	.06	.99	.72	.37	.06	.78	.39	.05				
	250, 7, 1	.95	.60	.07	.94	.60	.07	.94	.60	.07	.95	.60	.06	.95	.60	.06					
	250, 7, 5	.92	.66	.40	.22	.04	.89	.54	.07	.93	.55	.06									
	500, 3, 1	.76	.35	.04	.78	.37	.05	.79	.38	.05											
	500, 3, 5	.09	.20	.20	.08	.02	.99	.67	.31	.04	.76	.36	.05								
	500, 5, 1	.98	.68	.06	.98	.68	.05	.98	.68	.05	.98	.68	.05								
	500, 5, 5	.79	.23	.14	.07	.03	.94	.61	.05	.97	.64	.05									
	500, 7, 1	.85	.06	.85	.06	.85	.06	.85	.06	.85	.06										
	500, 7, 5	.99	.79	.48	.30	.04	.99	.79	.48	.30	.04	.99	.79	.48	.30	.04	.99	.79	.48	.30	.04

Table 10 continued

θ	$\{N, T, \tau\} \backslash \lambda$	$\Upsilon^{(1)}$					$\Upsilon^{(2)}$					$\Upsilon^{(3)}$								
		-.70	-.30	-.10	-.05	-.01	-.0005	.00	-.70	-.30	-.10	-.05	-.01	-.0005	.00	-.70	-.30	-.10	-.05	-.01
0.5	50, 3, 1	.99	.92	.64	.40	.12	.08	.06	.97	.73	.47	.14	.09	.06	.83	.54	.15	.09	.06	
	50, 3, 5	.94	.39	.12	.10	.04	.04	.03	.71	.41	.27	.10	.07	.05	.98	.81	.53	.15	.10	
	50, 5, 1		.99	.93	.80	.29	.17	.10	.98	.84	.31	.19	.11		.99	.87	.32	.19	.11	
	50, 5, 5		.89	.35	.28	.11	.07	.05	.93	.78	.63	.26	.16	.10	.99	.84	.30	.17	.09	
	50, 7, 1		.99	.95	.45	.27	.14		.96	.46	.27	.14			.97	.45	.27	.13		
	50, 7, 5		.99	.64	.51	.23	.14	.09	.98	.93	.83	.40	.25	.13	.99	.95	.42	.24	.10	
	250, 3, 1			.98	.45	.21	.06		.99	.48	.23	.07			.49	.24	.08			
	250, 3, 5		.93	.17	.14	.06	.03	.02	.97	.89		.38	.19	.06	.99	.46	.23	.08		
	250, 5, 1			.79	.43	.09			.79	.44	.09				.79	.44	.09			
	250, 5, 5		.68	.45	.23	.11	.03		.72	.39	.09				.76	.40	.08			
	250, 7, 1			.92	.59	.11			.92	.59	.11				.93	.59	.10			
	250, 7, 5		.95	.83	.59	.34	.07		.88	.55	.10				.91	.55	.09			
	500, 3, 1			.74	.38	.07			.75	.38	.08				.76	.39	.08			
	500, 3, 5		.20	.14	.10	.05	.02		.67	.34	.07				.74	.37	.07			
	500, 5, 1			.96	.66	.09			.96	.66	.09				.96	.66	.09			
	500, 5, 5		.86	.53	.31	.16	.03		.94	.61	.08				.96	.63	.08			
	500, 7, 1			.83	.10				.82	.10					.83	.09				
	500, 7, 5		.99	.93	.77	.53	.07		.99	.78	.09				.80	.08				

See Table 8

Table 11 Monte Carlo results for $\alpha = -0.5$. $\Upsilon^{(4),(6)}$

θ	$\{N, T, \tau\} \setminus \lambda$	$\Upsilon^{(4)}$										$\Upsilon^{(5)}$																	
		-.70	-.30	-.10	-.05	-.01	-.005	.00	.04	.06	.11	.16	.21	.26	.31	.36	.41	.46	.51	.56	.61	.66	.71	.76	.81	.86	.91	.96	.99
0.0	50, 3, 1		.95	.74	.46	.11	.06	.04	.98	.92	.68	.43	.10	.06	.04	.98	.92	.68	.43	.10	.06	.04	.98	.92	.68	.43	.10	.06	.04
	50, 3, 5		.50	.45	.26	.09	.05	.04	.58	.30	.22	.17	.07	.04	.03	.58	.30	.22	.17	.07	.04	.03	.58	.30	.22	.17	.07	.04	.03
50, 5, 1	50, 5, 1		.99	.98	.86	.30	.15	.07	.98	.98	.95	.83	.28	.15	.07	.98	.98	.95	.83	.28	.15	.07	.98	.98	.95	.83	.28	.15	.07
	50, 5, 5		.68	.82	.63	.27	.14	.07	.92	.57	.44	.43	.19	.11	.05	.92	.57	.44	.43	.19	.11	.05	.92	.57	.44	.43	.19	.11	.05
50, 7, 1	50, 7, 1				.97	.47	.25	.10			.99	.95	.45	.24	.10			.99	.95	.45	.24	.10			.99	.95	.45	.24	.10
	50, 7, 5		.84	.94	.81	.46	.25	.10	.82	.82	.62	.62	.34	.20	.08	.82	.82	.62	.62	.34	.20	.08	.82	.82	.62	.62	.34	.20	.08
250, 3, 1	250, 3, 1				.99	.50	.21	.04			.99	.99	.47	.20	.04			.99	.99	.47	.20	.04			.99	.99	.47	.20	.04
	250, 3, 5		.95	.98	.89	.47	.20	.04	.91	.59	.58	.60	.28	.13	.03	.91	.59	.58	.60	.28	.13	.03	.91	.59	.58	.60	.28	.13	.03
250, 5, 1	250, 5, 1				.84	.44	.06				.84	.44	.06					.84	.44	.06					.84	.44	.06		
	250, 5, 5		.97		.99	.47	.06	.06	.89	.89	.81	.88	.67	.35	.05	.89	.89	.81	.88	.67	.35	.05	.89	.89	.81	.88	.67	.35	.05
250, 7, 1	250, 7, 1				.96	.63	.07				.96	.63	.07					.96	.63	.07					.96	.63	.07		
	250, 7, 5		.99		.95	.66	.07	.07	.99	.99	.92	.96	.87	.54	.06	.99	.99	.92	.96	.87	.54	.06	.99	.99	.92	.96	.87	.54	.06
500, 3, 1	500, 3, 1				.81	.39	.05				.81	.39	.05					.81	.39	.05					.81	.39	.05		
	500, 3, 5		.99		.99	.39	.04	.04	.98	.78	.81	.85	.57	.27	.03	.98	.78	.81	.85	.57	.27	.03	.98	.78	.81	.85	.57	.27	.03
500, 5, 1	500, 5, 1				.98	.71	.05				.98	.71	.05					.98	.71	.05					.98	.71	.05		
	500, 5, 5				.99	.73	.05	.05	.99	.97	.94	.98	.93	.60	.05	.99	.97	.94	.98	.93	.60	.05	.99	.97	.94	.98	.93	.60	.05
500, 7, 1	500, 7, 1				.87	.87	.06				.87	.87	.06					.87	.87	.06					.87	.87	.06		
	500, 7, 5				.88	.88	.06	.06	.88	.98	.98	.80	.80	.80	.06	.88	.98	.98	.80	.80	.80	.06	.88	.98	.98	.80	.80	.80	.06

Table 11 continued

θ	$\{N, T, \tau\} \backslash \lambda$	$\mathbf{Y}^{(4)}$										$\mathbf{Y}^{(5)}$																	
		-.70	-.30	-.10	-.05	-.01	-.0005	.00	.07	.10	.14	.19	.23	.28	.33	.39	.46	.50	.55	.60	.66	.71	.77	.83	.89	.95	.99		
0.5	50, 3, 1		.96	.74	.49	.14	.10	.07								.94	.70	.46	.14									.09	.06
	50, 3, 5		.58	.48	.30	.11	.08	.05								.39	.28	.21	.09									.06	.04
	50, 5, 1			.98	.85	.33	.19	.11								.99	.96	.83	.31									.19	.11
	50, 5, 5		.76	.85	.66	.30	.18	.10								.67	.54	.50	.23									.15	.08
	50, 7, 1				.96	.47	.28	.14								.99	.99	.95	.46									.27	.14
	50, 7, 5		.90	.96	.84	.47	.28	.14								.88	.71	.69	.37									.23	.12
	250, 3, 1				.99	.50	.24	.07										.99	.48									.23	.07
	250, 3, 5		.97	.99	.91	.47	.23	.06								.78	.73	.71	.33									.17	.05
	250, 5, 1				.99	.81	.45	.10										.96	.79									.44	.09
	250, 5, 5		.99		.99	.81	.46	.09								.95	.92	.95	.70									.38	.09
	250, 7, 1				.93	.93	.60	.10										.92	.92									.59	.11
	250, 7, 5				.94	.94	.63	.10									.97	.99	.99	.87								.55	.10
	500, 3, 1				.77	.77	.41	.08											.75									.39	.08
	500, 3, 5				.77	.77	.40	.07								.93	.92	.93	.62									.31	.06
	500, 5, 1				.97	.97	.68	.09										.99	.96									.66	.09
	500, 5, 5				.97	.97	.70	.08								.99	.98	.99	.93									.61	.08
	500, 7, 1				.84	.84	.84	.10										.10	.83									.83	.10
	500, 7, 5				.85	.85	.85	.09										.09	.99									.79	.09

See Table 8

References

- Abadir KM, Magnus JR (2002) Notation in econometrics: a proposal for a standard. *Econom J* 5:76–90
- Ahn SC, Thomas GM (2006) Likelihood based inference for dynamic panel data models. Mimeo, New York City
- Alonso-Borrego C, Arellano M (1999) Symmetrically normalized instrumental-variable estimation using panel data. *J Bus Econ Stat* 17:36–49
- Anderson TW, Hsiao C (1982) Formulation and estimation of dynamic models using panel data. *J Econom* 18:47–82
- Andrews DWK (1987) Asymptotic results for generalized wald tests. *Econom Theory* 3(3):348–358
- Andrews DWK, Cheng X (2012) Estimation and inference with weak, semi-strong, and strong identification. *Econometrica* 80(5):2153–2211
- Arellano M (2003) Panel data econometrics. Advanced texts in econometrics. Oxford University Press, Oxford
- Arellano M (2016) Modeling optimal instrumental variables for dynamic panel data models. *Res Econ* 70(2):238–261
- Arellano M, Bond S (1991) Some tests of specification for panel data: Monte Carlo evidence and an application to employment equations. *Rev Econ Stud* 58:277–297
- Baltagi BH, Bresson G, Pirotte A (2007) Panel unit root tests and spatial dependence. *J Appl Econom* 22:339–360
- Binder M, Hsiao C, Pesaran MH (2005) Estimation and inference in short panel vector autoregressions with unit root and cointegration. *Econom Theory* 21:795–837
- Blundell RW, Bond S (1998) Initial conditions and moment restrictions in dynamic panel data models. *J Econom* 87:115–143
- Bond S, Nauges C, Windmeijer F (2005) Unit roots: identification and testing in micro panels. Mimeo, New York City
- Bun MJG, Carree MA (2005) Bias-corrected estimation in dynamic panel data models. *J Bus Econ Stat* 23(2):200–210
- Bun MJG, Kleibergen FR (2016) Identification and inference in moments based analysis of linear dynamic panel data models. uvA-Econometrics Working Paper Series
- Bun MJG, Sarafidis V (2015) Dynamic panel data models. In: Baltagi BH (ed) The oxford handbook of panel data, Chap 3. Oxford University Press, Oxford
- Bun MJG, Carree MA, Juodis A (2016) On maximum likelihood estimation of dynamic panel data models. *Oxford Bull Econ Stat* (forthcoming)
- Calzolari G, Magazzini L (2012) Autocorrelation and masked heterogeneity in panel data models estimated by maximum likelihood. *Empir Econ* 43(1):145–152
- Dhaene G, Jochmans K (2015) Split-panel jackknife estimation of fixed-effect models. *Rev Econ Stud* 82(3):991–1030
- Dhaene G, Jochmans K (2016) Likelihood inference in an autoregression with fixed effects. *Econom Theory* 32(5):1178–1215
- Dovonon P, Hall A (2016) The asymptotic properties of gmm and indirect inference under second-order identification. Mimeo, New York City
- Dovonon P, Renault E (2009) Gmm overidentification test with first order underidentification. Mimeo, New York City
- Hahn J, Kuersteiner G (2002) Asymptotically unbiased inference for a dynamic panel model with fixed effects when both n and t are large. *Econometrica* 70(4):1639–1657
- Harris RD, Tzavalis E (1999) Inference for unit roots in dynamic panels where the time dimension is fixed. *J Econom* 91(2):201–226
- Hayakawa K (2009) On the effect of mean-nonstationarity in dynamic panel data models. *J Econom* 153:133–135
- Hayakawa K (2016) An improved gmm estimation of panel var models. *Comput Stat Data Anal* 100:240–264
- Hayakawa K, Nagata S (2016) On the behavior of the gmm estimator in persistent dynamic panel data models with unrestricted initial conditions. *Comput Stat Data Anal* 100:265–303
- Hsiao C, Pesaran MH, Tahmiscioglu AK (2002) Maximum likelihood estimation of fixed effects dynamic panel data models covering short time periods. *J Econom* 109:107–150
- Johansen S (1991) Estimation and hypothesis testing of cointegration vectors in gaussian vector autoregressive models. *Econometrica* 59(6):1551–1580

- Johansen S (1995) Likelihood-based inference in cointegrated vector autoregressive models. Advanced texts in econometrics. Oxford University Press, Oxford
- Juodis A (2013) A note on bias-corrected estimation in dynamic panel data models. *Econ Lett* 118:435–438
- Juodis A (2016) First difference transformation in panel var models: robustness, estimation and inference. *Econom Rev* (forthcoming)
- Kiviet JF (1995) On bias, inconsistency, and efficiency of various estimators in dynamic panel data models. *J Econom* 68:53–78
- Kleibergen FR, Paap R (2006) Generalized reduced rank tests using the singular value decomposition. *J Econom* 133:97–126
- Kruiniger H (2008) Maximum likelihood estimation and inference methods for the covariance stationary panel ar(1)/unit root model. *J Econom* 144:447–464
- Kruiniger H (2013) Quasi ml estimation of the panel ar(1) model with arbitrary initial conditions. *J Econom* 173:175–188
- Kruiniger H, Tzavalis E (2002) Testing for unit roots in short dynamic panels with serially correlated and heteroscedastic disturbance terms, working paper 459, Queen Mary, University of London
- Magnus JR, Neudecker H (2007) Matrix differential calculus with applications in statistics and econometrics. Wiley, Hoboken
- Moon HR, Perron B, Phillips PCB (2007) Incidental trends and the power of panel unit root tests. *J Econom* 141(2):416–459
- Mutl J (2009) Panel var models with spatial dependence. Mimeo, New York City
- Nickell S (1981) Biases in dynamic models with fixed effects. *Econometrica* 49:1417–1426
- Phillips PCB (2015) Dynamic panel Anderson–Hsiao estimation with roots near unity. *Econom Theory* (forthcoming)
- Ramalho JJS (2005) Feasible bias-corrected ols, within-groups, and first-differences estimators for typical micro and macro ar(1) panel data models. *Empir Econ* 30:735–748
- Roodman D (2009) A note on the theme of too many instruments. *Oxford Bull Econ Stat* 71:135–158
- Roznitzky A, Cox DR, Bottai M, Robins J (2000) Likelihood-based inference with singular information matrix. *Bernoulli* 6(2):243–284
- Westerlund J (2016) Pooled panel unit root tests and the effect of past initialization. *Econom Rev* 35(3):396–427. doi:[10.1080/07474938.2013.833829](https://doi.org/10.1080/07474938.2013.833829)
- Windmeijer F (2005) A finite sample correction for the variance of linear efficient two-step gmm estimators. *J Econom* 126(1):25–51