UvA ECONOMETRICS

Discussion Paper: 2014/08

Cointegration Testing in Panel VAR Models Under Partial Identification and Spatial Dependence

Artūras Juodis

www.ase.uva.nl/uva-econometrics

Amsterdam School of Economics
Department of Economics & Econometrics
Valckenierstraat 65-67
1018 XE AMSTERDAM
The Netherlands

UvA UNIVERSITEIT VAN AMSTERDAM

Cointegration Testing in Panel VAR Models Under Partial Identification and Spatial Dependence☆

Artūras Juodis[a,b]

[a] *Amsterdam School of Economics, University of Amsterdam, Valckenierstraat 65-67, 1018 XE, Amsterdam, The Netherlands*
[b] *Tinbergen Institute, The Netherlands*

Abstract

This paper considers the Panel Vector Autoregressive Models of order 1 (PVAR(1)) with possibly spatially dependent error terms. We prove that the cointegration testing procedure of Binder, Hsiao, and Pesaran (2005) is not valid due to the singularity of the corresponding Hessian matrices under pure unit roots or cointegrated processes. As an alternative we propose a simple Method of Moments based cointegration test using the rank test of Kleibergen and Paap (2006) for fixed number of time observations. The test is shown to be robust to time series heteroscedasticity as well as unbalanced panels. The novelty of our approach is that we exploit the "weakness" of the Anderson and Hsiao (1982) moment conditions in the construction of the new test. The finite-sample performance of the proposed test statistic is investigated using the simulated data. The results show that for most scenarios the method performs well in terms of both size and power. The proposed test is applied to employment and wage equations using Spanish firm data of Alonso-Borrego and Arellano (1999) and the results show little evidence for cointegration.

Keywords: Dynamic Panel Data, Panel VAR, Cointegration, Heteroscedasticity, Spatial Dependence, Fixed T Consistency.

JEL: C13, C33.

☆I would like to thank Peter Boswijk, Maurice Bun and Vasilis Sarafidis, as well as seminar participants at Tinbergen Institute, Netherlands Econometrics Study Group 2013 and "Conference on Cross-sectional Dependence in Panel Data" in Cambridge 2013, for their comments and suggestions. Financial support from the NWO MaGW grant "Likelihood-based inference in dynamic panel data models with endogenous covariates" is gratefully acknowledged.

Email address: `a.juodis@uva.nl` (Artūras Juodis)

1. Introduction

The standard textbook treatment of econometrics assumes that estimation and hypothesis testing are the two sides of the same coin, with the latter being impossible to implement without the former. More importantly, for most of standard estimation methods regularity conditions necessary for hypothesis testing are equivalent to those of estimation. As a prototypical example consider the full rank assumption in the (Generalized) Method of Moments estimation that is needed both for estimation and hypothesis testing using Wald test. Unfortunately, for some econometric models this assumption can be too strong, resulting in a partial identification only, see e.g. Phillips (1989).

In their pioneering work Anderson and Rubin (1949) advocated the idea that it is possible to perform hypothesis testing for a simultaneous equations model under regularity conditions weaker than estimation conditions.[1] Thus, in some situations it is possible to avoid the estimation step and to perform the hypothesis testing directly. In this paper we apply a similar principle by turning a disadvantageous situation from estimating point of view into an advantageous one for hypothesis testing.

We consider the cointegration testing problem for the Panel VAR model of order 1 with fixed time dimension. Up to date the only method proposed for cointegration testing is the Transformed Maximum Likelihood testing procedure of Binder et al. (2005)[henceforth BHP]. However in the univariate setup it is known that for persistent data this likelihood approach does not have a Gaussian asymptotic limit. In this paper, we similarly prove that cointegration testing procedure of Binder et al. (2005) is not valid due to the singularity of the corresponding Hessian matrices under pure unit roots or cointegrated processes.

To the best of our knowledge currently in the Dynamic Panel Data (DPD) literature with fixed number of time periods no feasible Method of Moments (MM) alternative to likelihood based cointegration testing procedures was proposed. The main reason for the absence of MM alternatives is the partial identification issue of the standard Anderson and Hsiao (1982)[AH] moment conditions when the process is cointegrated, as the Jacobian of these moment conditions are of reduced rank. Therefore we propose a rank based

[1] For more recent papers please refer to Stock and Wright (2000) and Kleibergen (2005) *inter alia*.

cointegration test for the Jacobian of the aforementioned moment conditions. We show that the proposed test is robust to time series heteroscedasticity and, unlike the likelihood based tests, (e.g. the aforementioned TML estimator requires estimation of all variances parameters) our test does not require any numerical optimization algorithms.

In the Monte Carlo section of this paper we investigate the finite sample properties of the proposed procedure. We find that new cointegration testing procedure provides good size control as well as high power in most designs consider. Furthermore, in situations where error terms are spatially correlated, but this information is not taken into account, we found that only minor size distortions are visible.

The paper is structured as follows. In Section 2 we briefly present the model and the testing problem at hand. In Section 3 we present results for the likelihood based testing procedure of Binder et al. (2005). Rank-based cointegration testing procedure is formally introduced in Section 4. In Section 5 we continue with the finite sample performance by means of a Monte Carlo analysis. In Section 6 we illustrate the testing procedure using the Spanish manufacturing data of Alonso-Borrego and Arellano (1999). Finally, we conclude in Section 7.

Here we briefly discuss notation. **Bold** upper-case Greek letters are used to denote the original parameters, i.e. $\{\boldsymbol{\Phi}, \boldsymbol{\Sigma}, \boldsymbol{\Psi}\}$, while the lower-case Greek letters $\{\boldsymbol{\phi}, \boldsymbol{\sigma}, \boldsymbol{\psi}\}$ will denote $\operatorname{vec}(\cdot)$ ($\operatorname{vech}(\cdot)$ for symmetric matrices) of corresponding parameters, in the univariate setup corresponding parameters will be denoted by $\{\phi, \sigma^2, \psi^2\}$. We use $\rho(\boldsymbol{A})$ to denote the spectral radius[2] of a matrix $\boldsymbol{A} \in \mathbb{R}^{n\times n}$. The commutation matrix $\mathbf{K}_{a,b}$ is defined such that for any $[a \times b]$ matrix $\boldsymbol{A}$, $\operatorname{vec}(\boldsymbol{A}') = \mathbf{K}_{a,b} \operatorname{vec}(\boldsymbol{A})$. The duplication matrix $\mathbf{D}_m$ is defined such that for symmetric $[a \times a]$ matrix $\operatorname{vec} \boldsymbol{A} = \mathbf{D}_m \operatorname{vech} \boldsymbol{A}$. We define $\bar{\boldsymbol{y}}_{i-} \equiv (1/T)\sum_{t=1}^{T} \boldsymbol{y}_{i,t-1}$ and similarly $\bar{\boldsymbol{y}}_i \equiv (1/T)\sum_{t=1}^{T} \boldsymbol{y}_{i,t}$. We will use $\tilde{x}$ to indicate variables after Within Group transformation (for example $\tilde{\boldsymbol{y}}_{i,t} \equiv \boldsymbol{y}_{i,t} - \bar{\boldsymbol{y}}_i$), while $\ddot{\boldsymbol{x}}$ will be used for variables after a "quasi-averaging" transformation.[3]. For further details regarding the notation used in this paper, see Abadir and Magnus (2002). Where necessary to avoid confusion, we will use the subscript 0 to denote the true value of the parameters, e.g. $\boldsymbol{\Phi}_0$

[2] $\rho(\boldsymbol{A}) \equiv \max_i(|\lambda_i|)$, where λ_i's are (possibly complex) eigenvalues of a matrix $\boldsymbol{A}$.

[3] $\ddot{\boldsymbol{y}}_i \equiv \bar{\boldsymbol{y}}_i - \boldsymbol{y}_{i,0}$ and $\ddot{\boldsymbol{y}}_{i-} \equiv \bar{\boldsymbol{y}}_{i-} - \boldsymbol{y}_{i,0}$.

or $\boldsymbol{\Sigma}_0$.

2. Model

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

$$\boldsymbol{y}_{i,t} = \boldsymbol{\eta}_i + \boldsymbol{\Phi}\boldsymbol{y}_{i,t-1} + \boldsymbol{\varepsilon}_{i,t}, \quad i = 1, \ldots, N, \quad t = 1, \ldots, T, \tag{1}$$

where $\boldsymbol{y}_{i,t}$ is an $[m \times 1]$ vector, $\boldsymbol{\Phi}$ is an $[m \times m]$ matrix of parameters to be estimated, $\boldsymbol{\eta}_i$ is an $[m \times 1]$ vector of fixed effects and $\boldsymbol{\varepsilon}_{i,t}$ is an $[m \times 1]$ vector of innovations independent across i, with zero mean and covariance matrix $\boldsymbol{\Sigma}_t$. If we set $m = 1$ the model reduces to the linear DPD model with AR(1) dynamics. Throughout this paper we maintain the assumption that data in levels $\{\boldsymbol{y}_{i,t}\}$ is available for all $t = \{0, \ldots, T\}$ and $i = \{1, \ldots, N\}$.

We assume that $\boldsymbol{\eta}_i$ satisfy the so-called "common dynamics"("common factor") assumption:

$$\boldsymbol{\eta}_i = (\boldsymbol{I}_m - \boldsymbol{\Phi})\boldsymbol{\mu}_i.$$

If at least one eigenvalue is equal to unity this assumption ensures that there is no discontinuity in DGP, for further discussion, see e.g. BHP. Assuming common dynamics we can rewrite the model in (1) as:

$$\Delta\boldsymbol{y}_{i,t} = \boldsymbol{\Pi}\boldsymbol{u}_{i,t-1} + \boldsymbol{\varepsilon}_{i,t}, \quad i = 1, \ldots, N, \quad t = 1, \ldots, T.$$

Here we define $\boldsymbol{\Pi} = \boldsymbol{\Phi} - \boldsymbol{I}_m$ and $\boldsymbol{u}_{i,t-1} \equiv \boldsymbol{y}_{i,t-1} - \boldsymbol{\mu}_i$. We say that series $\boldsymbol{y}_{i,t}$ are *cointegrated* if the $\boldsymbol{\Pi}$ matrix is of reduced rank.[4] In particular, there exist $[m \times r]$ matrices $\boldsymbol{\alpha}_r$ and $\boldsymbol{\beta}_r$[5] of full column rank such that:

$$\boldsymbol{\Phi} = \boldsymbol{I}_m + \boldsymbol{\alpha}_r\boldsymbol{\beta}_r',$$

where r is the rank of $\boldsymbol{\Pi}$. In general matrices $\boldsymbol{\alpha}_r$ and $\boldsymbol{\beta}_r$ are not unique as for any $[r \times r]$ invertible matrix $\boldsymbol{U}$:

$$\boldsymbol{\alpha}_r\boldsymbol{\beta}_r' = \boldsymbol{\alpha}_r\boldsymbol{U}\boldsymbol{U}^{-1}\boldsymbol{\beta}_r' = \boldsymbol{\alpha}_r^*\boldsymbol{\beta}_r^{*'}.$$

[4] Note that unlike in pure time series models, we do not define cointegration as a property of time series because we keep T fixed.

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

This is the so-called *rotation problem*. As a result, it is a usual practice in the cointegration literature to impose identifying restrictions on $\boldsymbol{\alpha}_r$ or $\boldsymbol{\beta}_r$. The exact normalization of $\boldsymbol{\beta}_r$ matrix is important for existing procedures that are available in the literature, but not for the testing procedure that we formally introduce in Section 4.

3. Existing procedures. Transformed Maximum Likelihood of Binder et al. (2005)

We begin this section by recalling the list of Standard Assumptions (**SA**) used to derive asymptotic distribution of the estimator in BHP:[6]

(SA.1) The disturbances $\boldsymbol{\varepsilon}_{i,t}, t \leq T$, are i.i.d. for all i, with $\mathrm{E}[\boldsymbol{\varepsilon}_{i,t}] = \mathbf{0}_m$ and $\mathrm{E}[\boldsymbol{\varepsilon}_{i,t}\boldsymbol{\varepsilon}_{i,s}'] = 1_{(s=t)}\boldsymbol{\Sigma}_0$, $\boldsymbol{\Sigma}_0$ being a p.d. matrix.

(SA.2) The initial deviation $\boldsymbol{u}_{i,0} \equiv \boldsymbol{y}_{i,0} - \boldsymbol{\mu}_i$ are i.i.d. across cross-sectional units, with $\mathrm{E}[\boldsymbol{u}_{i,0}] = \mathbf{0}_m$ and finite (constant) variance $\boldsymbol{\Psi}_{u,0}$, where $\boldsymbol{\eta}_i = (\boldsymbol{I}_m - \boldsymbol{\Phi}_0)\boldsymbol{\mu}_i$.

(SA.3) The following moment restrictions are satisfied: $\mathrm{E}[\boldsymbol{u}_{i,0}\boldsymbol{\varepsilon}_{i,t}'] = \mathbf{O}_m$ for all i and $t = \{1, \ldots, T\}$.

(SA.4) $N \to \infty$, but T is fixed.

(SA.5) Denote by $\boldsymbol{\kappa}$ a $[m^2 + 2p \times 1]$ vector of unknown coefficients. $\boldsymbol{\kappa} \in \boldsymbol{\Gamma}$, where $\boldsymbol{\Gamma}$ is a compact subset of $\mathbb{R}^{m^2+2p}$ and $\boldsymbol{\kappa}_0 \in \mathrm{interior}(\boldsymbol{\Gamma})$, while $\rho(\boldsymbol{\Phi}_0) \leq 1$.

A detailed discussion of these assumptions can be found in Juodis (2014). The main exception is **(SA.5)** where we allow maximum eigenvalue of the autoregressive matrix $\boldsymbol{\Phi}_0$ to be 1. The exact components of the $\boldsymbol{\kappa}$ vector are related to a particular parametrization of the parameter space used for estimation.

The quasi log-likelihood function can then be defined for $\Delta\boldsymbol{Y}_i = \mathrm{vec}\left(\Delta\boldsymbol{y}_{i,1}, \ldots, \Delta\boldsymbol{y}_{i,T}\right)$ as follows:

$$\ell(\boldsymbol{\kappa}) = c - \frac{N}{2}\log|\boldsymbol{\Sigma}_{\Delta\boldsymbol{\tau}}| - \frac{N}{2}\mathrm{tr}\left(\left(\boldsymbol{R}'\boldsymbol{\Sigma}_{\Delta\boldsymbol{\tau}}^{-1}\boldsymbol{R}\right)\frac{1}{N}\sum_{i=1}^{N}\Delta\boldsymbol{Y}_i\Delta\boldsymbol{Y}_i'\right), \tag{2}$$

[6] Here $p = 0.5(m(m+1))$.

where $\boldsymbol{\kappa} = (\boldsymbol{\phi}', \boldsymbol{\sigma}', \boldsymbol{\psi}')'$ and $\boldsymbol{\Psi}$ is the variance-covariance matrix of the initial observation $\Delta \boldsymbol{y}_{i,1}$. The $\boldsymbol{\Sigma}_{\Delta \boldsymbol{\tau}}$ matrix has a block tridiagonal structure, with $-\boldsymbol{\Sigma}$ on first lower and upper off-diagonal blocks, and $2\boldsymbol{\Sigma}$ on all but first (1,1) diagonal blocks. The first (1,1) block is set to $\boldsymbol{\Psi}$ which takes into account the fact that we do not restrict $\Delta \boldsymbol{y}_{i,1}$ to be covariance stationary.[7] The $[mT \times mT]$ $\boldsymbol{R}$ matrix has $\boldsymbol{I}_m$ elements on the diagonal blocks, and $-\boldsymbol{\Phi}$ on the first lower off-diagonal blocks.

In Juodis (2014) it is shown that the log-likelihood function of BHP can be expressed in the following way:

$$\begin{aligned}\ell(\boldsymbol{\kappa}) = c - \frac{N}{2}\left((T-1)\log|\boldsymbol{\Sigma}| + \operatorname{tr}\left(\boldsymbol{\Sigma}^{-1}\frac{1}{N}\sum_{i=1}^{N}\sum_{t=1}^{T}(\tilde{\boldsymbol{y}}_{i,t} - \boldsymbol{\Phi}\tilde{\boldsymbol{y}}_{i,t-1})(\tilde{\boldsymbol{y}}_{i,t} - \boldsymbol{\Phi}\tilde{\boldsymbol{y}}_{i,t-1})'\right)\right) \\ - \frac{N}{2}\left(\log|\boldsymbol{\Theta}| + \operatorname{tr}\left(\boldsymbol{\Theta}^{-1}\frac{T}{N}\sum_{i=1}^{N}(\ddot{\boldsymbol{y}}_i - \boldsymbol{\Phi}\ddot{\boldsymbol{y}}_{i-})(\ddot{\boldsymbol{y}}_i - \boldsymbol{\Phi}\ddot{\boldsymbol{y}}_{i-})'\right)\right),\end{aligned} \tag{3}$$

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

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

then the score vector associated with the log-likelihood function (3) is given by:

$$\nabla(\boldsymbol{\kappa}) = \begin{pmatrix} \operatorname{vec}(\boldsymbol{W}_N(\boldsymbol{\kappa})) \\ \mathbf{D}_m' \operatorname{vec}(-\frac{N}{2}(\boldsymbol{\Sigma}^{-1}((T-1)\boldsymbol{\Sigma} - \boldsymbol{Z}_N(\boldsymbol{\kappa}))\boldsymbol{\Sigma}^{-1})) \\ \mathbf{D}_m' \operatorname{vec}(-\frac{N}{2}(\boldsymbol{\Theta}^{-1}(\boldsymbol{\Theta} - \boldsymbol{M}_N(\boldsymbol{\kappa}))\boldsymbol{\Theta}^{-1})) \end{pmatrix}. \tag{4}$$

When the process $\{\boldsymbol{y}_{i,t}\}_{t=0}^{T}$ is cointegrated and consequently the matrix $\boldsymbol{\Phi} - \boldsymbol{I}_m = \boldsymbol{\alpha}\boldsymbol{\beta}'$ is of reduced rank r, then this information can be taken into account while deriving the score and the Hessian of the log-likelihood function. For this purpose we use the same parametrization as BHP for $\boldsymbol{\beta}'$ to avoid rotational indeterminacy. Namely we set $\boldsymbol{\beta}' = \boldsymbol{\delta}'\boldsymbol{H}' + \boldsymbol{b}'$, where both $\boldsymbol{H}$ and $\boldsymbol{b}$ are given and $\boldsymbol{\delta}$ is an $[m - r \times r]$ (assuming $0 < r < m$) matrix. The parameter set in this case is defined as $\boldsymbol{\kappa} = ((\operatorname{vec}\boldsymbol{\alpha})', (\operatorname{vec}\boldsymbol{\delta}')', \boldsymbol{\sigma}', \boldsymbol{\theta}')'$.

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

Corollary 1. *Let Assumptions **SA** be satisfied. Then the restricted score vector associated with the log-likelihood function* (3) *under cointegrating restrictions is given by:*

$$\nabla_r(\boldsymbol{\kappa}) = \begin{pmatrix} \mathrm{vec}\,(\boldsymbol{W}_N(\boldsymbol{\kappa})\boldsymbol{\beta}) \\ \mathrm{vec}\,(\boldsymbol{\alpha}'\boldsymbol{W}_N(\boldsymbol{\kappa})\boldsymbol{H}) \\ \mathbf{D}_m'\,\mathrm{vec}\,(-\frac{N}{2}(\boldsymbol{\Sigma}^{-1}((T-1)\boldsymbol{\Sigma} - \boldsymbol{Z}_N(\boldsymbol{\kappa}))\boldsymbol{\Sigma}^{-1})) \\ \mathbf{D}_m'\,\mathrm{vec}\,(-\frac{N}{2}(\boldsymbol{\Theta}^{-1}(\boldsymbol{\Theta} - \boldsymbol{M}_N(\boldsymbol{\kappa}))\boldsymbol{\Theta}^{-1})) \end{pmatrix}. \tag{5}$$

In the special case where $m = 2$ and $r = 1$, $\boldsymbol{\delta}$ is a scalar while $\boldsymbol{\alpha}$ is a $[2 \times 1]$ vector implying that the corresponding entries of the $\nabla_r(\boldsymbol{\kappa})$ vector do not have a $\mathrm{vec}\,(\cdot)$ operator in them.

Using fairly standard consistency and asymptotic normality results for M-estimators Binder et al. (2005) conclude:

Proposition 1 (Asymptotic normality in Binder et al. (2005))**.** *Under Assumptions **SA**, the TMLE defined by maximizing* (3) *is consistent. Furthermore, under these assumptions:*

$$\sqrt{N}\,(\hat{\boldsymbol{\kappa}} - \boldsymbol{\kappa}_0) \xrightarrow{d} \mathrm{N}(\mathbf{0}, \mathfrak{B}_{QML}),$$

where:

$$\mathfrak{B}_{QML} = \boldsymbol{\mathcal{H}}_\ell^{-1}\boldsymbol{\mathcal{I}}_\ell\boldsymbol{\mathcal{H}}_\ell^{-1},$$
$$\boldsymbol{\mathcal{H}}_\ell = \lim_{N\to\infty} \mathrm{E}\left[-\frac{1}{N}\boldsymbol{\mathcal{H}}^N(\boldsymbol{\kappa}_0)\right], \textit{and} \quad \boldsymbol{\mathcal{I}}_\ell = \lim_{N\to\infty} \mathrm{E}\left[\frac{1}{N}\sum_{i=1}^N \nabla^{(i)}(\boldsymbol{\kappa}_0)\nabla^{(i)}(\boldsymbol{\kappa}_0)'\right].$$

It can be clearly seen that a necessary condition for the limiting distribution to be well defined is that the $\boldsymbol{\mathcal{H}}_\ell$ matrix is p.d.. When both the mean and the variance of $\Delta\boldsymbol{y}_{i,1}$ are unrestricted and treated as free parameters, Bond et al. (2005) showed that the TML estimator of Hsiao et al. (2002) is locally non-identified at the unit root. Here by local non-identification we mean singularity of $\boldsymbol{\mathcal{H}}_\ell$. In the next theorem we will show that the same conclusion holds for more general case with $m \geq 1$.

Theorem 1 (Singularity)**.** *Let Assumptions **SA** be satisfied. Then at $\boldsymbol{\Phi}_0 = \boldsymbol{I}_m$ the $\boldsymbol{\mathcal{H}}_\ell$*

matrix is equal to:

$$\boldsymbol{\mathcal{H}}_\ell = \frac{(T-1)}{2}\begin{pmatrix} T(\boldsymbol{\Sigma}_0 \otimes \boldsymbol{\Sigma}_0^{-1}) & -(\boldsymbol{I}_m \otimes \boldsymbol{\Sigma}_0^{-1})\,\mathbf{D}_m & (\boldsymbol{I}_m \otimes \boldsymbol{\Sigma}_0^{-1})\,\mathbf{D}_m \\ -\,\mathbf{D}_m'(\boldsymbol{I}_m \otimes \boldsymbol{\Sigma}_0^{-1}) & \mathbf{D}_m'(\boldsymbol{\Sigma}_0^{-1} \otimes \boldsymbol{\Sigma}_0^{-1})\,\mathbf{D}_m & \mathbf{O}_p \\ \mathbf{D}_m'(\boldsymbol{I}_m \otimes \boldsymbol{\Sigma}_0^{-1}) & \mathbf{O}_p & \frac{1}{T-1}\,\mathbf{D}_m'(\boldsymbol{\Sigma}_0^{-1} \otimes \boldsymbol{\Sigma}_0^{-1})\,\mathbf{D}_m \end{pmatrix} \tag{6}$$

furthermore, this matrix is singular.

Proof. In Appendix A.1. □

The singularity result in Theorem 1 is of special interest when the inference regarding the rank of $\boldsymbol{I}_m - \boldsymbol{\Phi}_0$ is concerned. As this TML estimator can be seen as a non-linear MM estimator with the score vector defining the moment conditions, singularity of $\boldsymbol{\mathcal{H}}_\ell$ matrix can be translated to the GMM language as a weak instrument problem.

Remark 1. The conclusion of Theorem 1 is in sharp contradiction with Remark 4.1. of BHP, where they note that:[8] "*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* $\boldsymbol{y}_{i,t}$ *process.*".

It is important to note that despite singularity of $\boldsymbol{\mathcal{H}}_\ell$, the TML estimator $\hat{\boldsymbol{\kappa}}_{TMLE}$ remains consistent, but does not have a standard limiting distribution.

Ahn and Thomas (2006) in their study used the approach of Roznitzky, Cox, Bottai, and Robins (2000) to show that $\hat{\phi}$ converges at rate $N^{1/4}$ to a non-standard distribution. Later on their results are extended to i.h.d. setting in Kruiniger (2013) . Furthermore, it is proved that the limiting distribution of the LR test statistic for $H_0 : \phi_0 = 1$ is a mixture of $\chi^2(1)$ distributed random variable and zero. In this paper we will not attempt to study the distributional consequences of the singularity for the TMLE and LR test and leave it for future research, as preliminary numerical simulations suggest that the rank of $\boldsymbol{\mathcal{H}}_\ell$ has rank deficiency larger than one (for $m = 2$ the rank of $\boldsymbol{\mathcal{H}}_\ell$ is equal to 7, while full rank is 10), hence results of Roznitzky et al. (2000) need to be generalized taking into

[8]Notation is appropriately adjusted to the notation used in this paper

account this possibility. Based on recent work of Dovonon and Renault (2009) in the GMM literature it is know that for general rank deficiencies the maximal rate of convergence is $N^{1/4}$. However no results regarding the behavior of the LR ratio test in cases like ours are available.

Although the case with pure unit roots 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.

Using the block structure of the Hessian matrix it can be easily seen that $|\boldsymbol{\mathcal{H}}_\ell| \propto |\boldsymbol{B}|$. Here $\boldsymbol{B}$ is given by:[9]

$$\begin{aligned}\boldsymbol{B} &= (\mathrm{E}[\boldsymbol{R}_N] \otimes \boldsymbol{\Sigma}_0^{-1} + \mathrm{E}[\boldsymbol{P}_N] \otimes \boldsymbol{\Theta}_0^{-1}) \\ &- \left[\frac{1}{T-1}\Big(\big(\mathrm{E}[\boldsymbol{Q}_N(\boldsymbol{\kappa}_0)]\boldsymbol{\Sigma}_0^{-1}\big) \otimes \big(\mathrm{E}[\boldsymbol{Q}_N(\boldsymbol{\kappa}_0)]\boldsymbol{\Sigma}_0^{-1}\big)'\Big) + \Big(\big(\mathrm{E}[\boldsymbol{N}_N(\boldsymbol{\kappa}_0)]\boldsymbol{\Theta}_0^{-1}\big) \otimes \big(\mathrm{E}[\boldsymbol{N}_N(\boldsymbol{\kappa}_0)]\boldsymbol{\Theta}_0^{-1}\big)'\Big)\right]\mathbf{K}_m \\ &- \left[\frac{1}{T-1}\big(\big(\mathrm{E}[\boldsymbol{Q}_N(\boldsymbol{\kappa}_0)]\boldsymbol{\Sigma}_0^{-1}\,\mathrm{E}[\boldsymbol{Q}_N(\boldsymbol{\kappa}_0)]'\big) \otimes \boldsymbol{\Sigma}_0^{-1}\big) + \big(\big(\mathrm{E}[\boldsymbol{N}_N(\boldsymbol{\kappa}_0)]\boldsymbol{\Theta}_0^{-1}\,\mathrm{E}[\boldsymbol{N}_N(\boldsymbol{\kappa}_0)]'\big) \otimes \boldsymbol{\Theta}_0^{-1}\big)\right]\end{aligned}$$

For the unit root case (i.e. $\boldsymbol{\Phi}_0 = \boldsymbol{I}_m$) this expression simplifies dramatically as $\boldsymbol{\Sigma}_0 = \boldsymbol{\Theta}_0$. That allowed us in Theorem 1 to show that $|\boldsymbol{B}| = 0$ for any value of $\boldsymbol{\Sigma}_0$ and T. Unfortunately, no such strong result is available when $\boldsymbol{\Pi}$ is of reduced rank $0 < r < m$. However, some weaker results can be derived, that provide intuition for singularity of the Hessian matrix. The next Proposition summarizes results available for $T = 2$.

Proposition 2. *Let $\boldsymbol{\Phi}_0$ be such that $\mathrm{rk}\,(\boldsymbol{\Phi}_0 - mI_m) = r$ and $T = 2$ then:*

$$\mathrm{rk}\,\boldsymbol{B} \leq 0.5m(m-1) + r^2. \tag{7}$$

Proof. In Appendix A.2. □

This quantity is smaller than m^2 for all $m \leq 4$. Given that in most cases bivariate PVAR is considered it follows that expected hessian matrix is singular and the corresponding estimator does not have a normal limiting distribution. Although this paper does not contain more general results for $T > 2$, we performed numerous numerical evaluations of $\boldsymbol{B}$ for larger values of T and different combinations of population matrices in the bivariate

[9] For exact definitions of $\boldsymbol{R}_N, \boldsymbol{P}_N, \boldsymbol{Q}_N(\boldsymbol{\kappa}), \boldsymbol{N}_N(\boldsymbol{\kappa})$ please refer to Appendix.

setup.[10] For all setups it was found that the expected hessian matrix is singular for $r < m$ and non-singular 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.

4. Jacobian based cointegration testing

To explain the intuition of our approach lets consider the following (standard) AH moment conditions for Panel VAR(1) model:

$$\operatorname{vec} \mathrm{E}[(\Delta \boldsymbol{y}_{i,t} - \boldsymbol{\Phi} \Delta \boldsymbol{y}_{i,t-1}) \boldsymbol{y}'_{i,t-2}] = \mathbf{0}_{m^2}, \quad t = 2, \ldots, T.$$

The (minus) Jacobian of these moment conditions is given by:

$$\left(\mathrm{E}[\Delta \boldsymbol{y}_{i,t-1} \boldsymbol{y}'_{i,t-2}]\right)' \otimes \boldsymbol{I}_m, \quad t = 2, \ldots, T. \tag{8}$$

It follows from the properties of the Kronecker product that the rank of this matrix is determined by the rank of the matrix in the brackets.[11] The expected value of this term is given by (upon redefining $t \to t+1$, as the previous expression is well defined for $t = T+1$):

$$\mathrm{E}[\Delta \boldsymbol{y}_{i,t} \boldsymbol{y}'_{i,t-1}] = \boldsymbol{\Pi}\, \mathrm{E}[\boldsymbol{u}_{i,t-1} \boldsymbol{y}'_{i,t-1}] + \mathrm{E}[\boldsymbol{\varepsilon}_{i,t} \boldsymbol{y}'_{i,t-1}]. \tag{9}$$

Under usual regularity conditions of the DPD literature[12] the second term is equal to $\mathbf{O}_m$, while the first term is the product of rank r and rank m matrices. As a result $\operatorname{rk}\left(\mathrm{E}[\Delta \boldsymbol{y}_{i,t} \boldsymbol{y}'_{i,t-1}]\right)$ is equal to r and leads to violation of the "relevance" condition for the Instrumental Variable (IV) estimator. In such situation we can not estimate $\boldsymbol{\Phi}$ consistently from the AH moment conditions. However, we can use the Jacobian matrix directly avoiding the estimation step to test for cointegration.

4.1. Regularity conditions

In the previous section we have presented the intuition of the proposed method, but it still remains to be investigated under which conditions the $\mathrm{E}[\Delta \boldsymbol{y}_{i,t} \boldsymbol{y}'_{i,t-1}]$ term is of reduced

[10] In particular, setups of BHP were considered.
[11] See e.g. Magnus and Neudecker (2007).
[12] That we will state formally in the next section.

rank if and only if $\boldsymbol{\Pi}$ is of reduced rank. We assume that the initial conditions $\boldsymbol{y}_{i,0}$ are of the following form:

$$\boldsymbol{y}_{i,0} = \boldsymbol{\Upsilon}\boldsymbol{\mu}_i + \boldsymbol{\varepsilon}_{i,0}.$$

Here $\boldsymbol{\Upsilon}$ is an $[m \times m]$ matrix that allows for a possible effect non-stationarity[13] of the initial condition if $\boldsymbol{\Upsilon} \neq \boldsymbol{I}_m$. For the purpose of this section instead of considering assumptions **SA**, we consider more primitive assumptions.

(A.1) The error terms $\boldsymbol{\varepsilon}_{i,t}$ are i.i.d. across cross sectional units and uncorrelated over time $\mathrm{E}[\boldsymbol{\varepsilon}_{i,t}\boldsymbol{\varepsilon}'_{i,s}] = \mathbf{O}_m$ for $s \neq t$. Variance of the error terms is a constant p.d. matrix $\operatorname{var}\boldsymbol{\varepsilon}_{i,t} = \boldsymbol{\Sigma}_t$ for $t > 0$. Furthermore, the higher order moment condition $\mathrm{E}[\|\boldsymbol{\varepsilon}_{i,t}\|^4] < \infty$ holds $\forall t$.

(A.2) The fixed effects $\boldsymbol{\mu}_i$ are i.i.d. across cross sectional units and have zero mean with a p.d. variance-covariance matrix $\boldsymbol{\Sigma}_{\boldsymbol{\mu}}$. Furthermore, for all i and $t \geq 0$ $\mathrm{E}[\boldsymbol{\mu}_i\boldsymbol{\varepsilon}'_{i,t}] = \mathbf{O}_m$. The higher order moment condition $\mathrm{E}[\|\boldsymbol{\mu}_i\|^4] < \infty$ holds.

Particularly we allow $\boldsymbol{\varepsilon}_{i,t}$ to be heteroscedastic over time.

$\mathrm{E}[\boldsymbol{\varepsilon}_{i,t}\boldsymbol{y}'_{i,t-1}] = \mathbf{O}_m$ is a direct implication of Assumptions **(A.1)-(A.2)**. However, they do not ensure that $\boldsymbol{\Pi}\,\mathrm{E}[\boldsymbol{u}_{i,t-1}\boldsymbol{y}'_{i,t-1}]$ has a reduced rank if and only if $\boldsymbol{y}_{i,t}$ are cointegrated. Lets investigate this issue more closely by expanding the $\mathrm{E}[\boldsymbol{u}_{i,t-1}\boldsymbol{y}'_{i,t-1}]$ term (for $t \geq 2$):

$$\begin{aligned}
\mathrm{E}[\boldsymbol{u}_{i,t-1}\boldsymbol{y}'_{i,t-1}] &= \mathrm{E}\left[\left(\boldsymbol{\Phi}^{t-1}\boldsymbol{u}_{i,0} + \sum_{s=0}^{t-2}\boldsymbol{\Phi}^s\boldsymbol{\varepsilon}_{i,t-s-1}\right)\left(\boldsymbol{\mu}_i + \boldsymbol{\Phi}^{t-1}\boldsymbol{u}_{i,0} + \sum_{s=0}^{t-2}\boldsymbol{\Phi}^s\boldsymbol{\varepsilon}_{i,t-s-1}\right)'\right] \\
&= \mathrm{E}\left[\left(\boldsymbol{\Phi}^{t-1}(\boldsymbol{\Upsilon} - \boldsymbol{I}_m)\boldsymbol{\mu}_i + \sum_{s=0}^{t-1}\boldsymbol{\Phi}^s\boldsymbol{\varepsilon}_{i,t-s-1}\right)\left((\boldsymbol{I}_m + \boldsymbol{\Phi}^{t-1}(\boldsymbol{\Upsilon} - \boldsymbol{I}_m))\boldsymbol{\mu}_i + \sum_{s=0}^{t-1}\boldsymbol{\Phi}^s\boldsymbol{\varepsilon}_{i,t-s-1}\right)'\right] \\
&= \underbrace{\boldsymbol{\Phi}^{t-1}(\boldsymbol{\Upsilon} - \boldsymbol{I}_m)\boldsymbol{\Sigma}_{\boldsymbol{\mu}}(\boldsymbol{\Phi}^{t-1}(\boldsymbol{\Upsilon} - \boldsymbol{I}_m))'}_{p.s.d} + \boldsymbol{\Phi}^{t-1}(\boldsymbol{\Upsilon} - \boldsymbol{I}_m)\boldsymbol{\Sigma}_{\boldsymbol{\mu}} \\
&+ \underbrace{\sum_{s=0}^{t-2}\boldsymbol{\Phi}^s\boldsymbol{\Sigma}_{t-1-s}\boldsymbol{\Phi}^{s'}}_{p.d.} + \mathrm{E}[\boldsymbol{\Phi}^{t-1}\boldsymbol{\varepsilon}_{i,0}\boldsymbol{\varepsilon}'_{i,0}(\boldsymbol{\Phi}^{t-1})']
\end{aligned}$$

In the effect-stationary case ($\boldsymbol{\Upsilon} = \boldsymbol{I}_m$) all terms involving $\boldsymbol{\Upsilon}$ are equal to $\mathbf{O}_m$. However, if that is not the case we have that the first term is a p.s.d. matrix, while it is not

[13]It is often called “mean non-stationarity” by other scholars.

immediately clear what happens with the second term. The third term is a p.d. matrix as all $\boldsymbol{\Sigma}_s$'s matrices are positive definite. The analysis of the last term is more subtle as it requires explicit assumptions regarding the DGP for $\boldsymbol{\varepsilon}_{i,0}$. In general, we are looking at $\boldsymbol{\varepsilon}_{i,0}$ such that at least the product matrix $\mathrm{E}[\boldsymbol{\Phi}^{t-1}\boldsymbol{\varepsilon}_{i,0}\boldsymbol{\varepsilon}_{i,0}'(\boldsymbol{\Phi}^{t-1})']$ is well defined and non-negative definite. Below we summarize a few DGP's for $\boldsymbol{\varepsilon}_{i,0}$ currently used in the literature that satisfy this condition.

(DGP.1) $\boldsymbol{\varepsilon}_{i,0} \sim IID(\mathbf{0}, \boldsymbol{\Sigma}_0)$ with $\boldsymbol{\Sigma}_0$ constant p.s.d. matrix (independent of other DGP parameters).

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

(DGP.3) $\boldsymbol{\varepsilon}_{i,0} = \sum_{l=0}^{\infty} \left(\boldsymbol{\Phi}^l - \boldsymbol{C}\right) \boldsymbol{\varepsilon}_{i,-l} + \boldsymbol{C}\boldsymbol{\xi}_i$. Here $\boldsymbol{\xi}_i$ is an $[m \times 1]$ vector of the (independent) individual-specific initialization effects, while $\boldsymbol{C} = \boldsymbol{\beta}_\perp \left(\boldsymbol{\alpha}_\perp' \boldsymbol{\beta}_\perp\right)^{-1} \boldsymbol{\alpha}_\perp'$ is an $m - r$ rank matrix.

For simplicity in what follows, we assume that all $\boldsymbol{\varepsilon}_{i,-l}$ are homoscedastic over time. To simplify matters we assume that all random variables in **(DGP.1)**-**(DGP.3)** satisfy assumptions **(A.1)**-**(A.2)**. The **(DGP.3)** initialization was used in the Monte Carlo studies of BHP and is motivated by the *Granger Representation Theorem*, see e.g. Johansen (1995)[Theorem 4.2]. The **(DGP.2)**, among others, was used in Hayakawa (2013).

It is important to emphasize that *all* three DGP's are well defined for all rank values of r. For $\rho(\boldsymbol{\Phi}) < 1$ we have $\boldsymbol{C} = \mathbf{O}_m$ resulting in stationary initialization. On the other hand, $\boldsymbol{\Phi} = \boldsymbol{I}_m$ implies $\boldsymbol{C} = \boldsymbol{I}_m$ (by definition) so that **(DGP.3)** and **(DGP.2)** coincide (by redefining M to $M + 1$).

For **(DGP.3)** by construction of the $\boldsymbol{C}$ matrix we have that $\boldsymbol{\Pi}\boldsymbol{C} = \mathbf{O}_m$, and thus $\boldsymbol{\Phi}^{t-1}\boldsymbol{C} = \boldsymbol{C}$. Combining these results

$$\mathrm{E}[\boldsymbol{\Pi}\boldsymbol{\Phi}^{t-1}\boldsymbol{\varepsilon}_{i,0}\boldsymbol{\varepsilon}_{i,0}'(\boldsymbol{\Phi}^{t-1})'] = \boldsymbol{\Pi}\boldsymbol{\Phi}^{t-1}\left(\sum_{l=0}^{\infty}\left(\boldsymbol{\Phi}^l - \boldsymbol{C}\right)\boldsymbol{\Sigma}\left(\boldsymbol{\Phi}^l - \boldsymbol{C}\right)'\right)(\boldsymbol{\Phi}^{t-1})',$$

where existence of $\sum_{l=0}^{\infty}\left(\boldsymbol{\Phi}^l - \boldsymbol{C}\right)\boldsymbol{\Sigma}\left(\boldsymbol{\Phi}^l - \boldsymbol{C}\right)'$ is implied by the absolute summability of $\{\boldsymbol{\Phi}^l - \boldsymbol{C}\}_{l=0}^{\infty}$, see e.g. Lütkepohl (2006). Furthermore, it is obvious that $\sum_{l=0}^{\infty}\left(\boldsymbol{\Phi}^l - \boldsymbol{C}\right)\boldsymbol{\Sigma}\left(\boldsymbol{\Phi}^l - \boldsymbol{C}\right)'$ is a p.s.d. matrix and consecutively that $\left(\mathrm{E}[\boldsymbol{\Phi}^{t-1}\boldsymbol{\varepsilon}_{i,0}\boldsymbol{\varepsilon}_{i,0}'(\boldsymbol{\Phi}^{t-1})']\right)$ is a p.s.d. matrix.

If we can ensure that $\boldsymbol{\Phi}^{t-1}(\boldsymbol{\Upsilon}-\boldsymbol{I}_m)\boldsymbol{\Sigma}_{\boldsymbol{\mu}}$ is such that $\mathrm{E}[\boldsymbol{u}_{i,t-1}\boldsymbol{y}'_{i,t-1}]$ has full rank m, then $\mathrm{E}[\boldsymbol{\Pi}\boldsymbol{u}_{i,t-1}\boldsymbol{y}'_{i,t-1}]$ has reduced rank r if and only if $\boldsymbol{y}_{i,t-1}$ are cointegrated.[14] However, it is not a trivial task to identify the parameter space of $\{\boldsymbol{\Phi}, \boldsymbol{\Upsilon}, \boldsymbol{\Sigma}_{\boldsymbol{\mu}}\}$ for the aforementioned condition to be satisfied. One special case is obtained for $\boldsymbol{\Upsilon} = \boldsymbol{I}_m$ (effect stationarity) with other matrices being unrestricted (at least finite). 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 (2003b) and Roodman (2009). In the Monte Carlo simulations we will check the adequacy of the proposed procedure by considering different values of $\boldsymbol{\Upsilon}$ that are mentioned in the literature.

4.2. Rank Test

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

$$\boldsymbol{D} = \boldsymbol{A}_q\boldsymbol{B}_q + \boldsymbol{A}_{q,\perp}\boldsymbol{\Lambda}_q\boldsymbol{B}_{q,\perp},$$

where all $\perp$ matrices are defined in the usual way and $\boldsymbol{\Lambda}_q$ is an $[(k-q)\times(m-q)]$ matrix. For $\boldsymbol{\Lambda}_q = \mathbf{O}$ the rank of $\boldsymbol{D}$ is determined by the rank of $\boldsymbol{A}_q\boldsymbol{B}_q$. The procedure in KP is based on testing if $\boldsymbol{\Lambda}_q$ is equal to $\mathbf{O}_m$, with matrices $\boldsymbol{A}_q, \boldsymbol{B}_q, \boldsymbol{\Lambda}_q$ obtained using the singular value decomposition (SVD). In our case matrix $\boldsymbol{D}$ is the $[m \times m]$ matrix $\mathrm{E}[\Delta\boldsymbol{y}_{i,t}\boldsymbol{y}'_{i,t-1}]$.

We define the following cross-sectional average:

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

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

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

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

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

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

Consecutively the estimator $\overline{\Delta \boldsymbol{y}_{i,t}\boldsymbol{y}'_{i,t-1}}$ satisfies sufficient conditions in KP.[15] As a result one can use the Theorem **1** of KP 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:*

$$\sqrt{N}\hat{\boldsymbol{\lambda}}_r \xrightarrow{d} \mathbf{N}(\mathbf{0}_r, \boldsymbol{\Omega}_r),$$

where:

$$\hat{\boldsymbol{\lambda}}_r = \operatorname{vec}\hat{\boldsymbol{\Lambda}}_r, \quad \hat{\boldsymbol{\Lambda}}_r = \hat{\boldsymbol{A}}'_{r,\perp}\overline{\Delta \boldsymbol{y}_{i,t}\boldsymbol{y}'_{i,t-1}}\hat{\boldsymbol{B}}'_{r,\perp},$$
$$\boldsymbol{\Omega}_r = (\boldsymbol{B}_{r,\perp} \otimes \boldsymbol{A}'_{r,\perp})\boldsymbol{V}(\boldsymbol{B}_{r,\perp} \otimes \boldsymbol{A}'_{r,\perp})'$$

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

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

converges in distribution to a $\chi^2((m-r)^2)$ distributed random variable.

All $\boldsymbol{A}$ and $\boldsymbol{B}$ variables in Theorem 2 are obtained from the SVD of the $\overline{\Delta \boldsymbol{y}_{i,t}\boldsymbol{y}'_{i,t-1}}$ matrix. 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:

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

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 test statistic for every value of t, but $t = 1$). Of

[15] $\overline{\Delta \boldsymbol{y}_{i,t}\boldsymbol{y}'_{i,t-1}}$ satisfies **Assumption 1** (asymptotic normal distribution), while $\boldsymbol{V}$ satisfies **Assumption 2**(full rank).

course, it is not the most efficient way of how the information can be used. Instead, all time series observations can be pooled into one test statistic for testing rank of:[16]

$$\overline{\Delta \boldsymbol{y}_{i,t}\boldsymbol{y}'_{i,t-1}}_T = \frac{1}{N}\sum_{i=1}^{N}\frac{1}{T-1}\sum_{t=2}^{T}\Delta \boldsymbol{y}_{i,t}\boldsymbol{y}'_{i,t-1}. \tag{10}$$

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

$$\begin{aligned}\boldsymbol{V}_N = \frac{1}{N}\sum_{i=1}^{N}&\operatorname{vec}\left(\frac{1}{T-1}\sum_{t=2}^{T}\Delta \boldsymbol{y}_{i,t}\boldsymbol{y}'_{i,t-1}\right)\operatorname{vec}\left(\frac{1}{T-1}\sum_{t=2}^{T}\Delta \boldsymbol{y}_{i,t}\boldsymbol{y}'_{i,t-1}\right)' \\ &- \operatorname{vec}\overline{\Delta \boldsymbol{y}_{i,t}\boldsymbol{y}'_{i,t-1}}_T \operatorname{vec}\overline{\Delta \boldsymbol{y}_{i,t}\boldsymbol{y}'_{i,t-1}}_T{}'.\end{aligned} \tag{11}$$

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

Up to this stage we considered only the Jacobian of Anderson and Hsiao (1982) moment conditions, however for $T > 2$ further lags can be used. The particular choice of lags used is subject to the same “arbitrariness” as the choice of moment conditions for the Arellano and Bond (1991)[AB] estimator. More importantly, it is not clear that the use of lags larger than 1 still ensures that $\mathrm{E}[\overline{\Delta \boldsymbol{y}_{i,t}\boldsymbol{y}'_{i,t-j}}_T]$ has reduced rank r if and only if $\operatorname{rk}\boldsymbol{\Pi} = r$ (even in the effect stationary case). Moreover, the power of the test might be substantially affected by the choice of lags, as with any alternative close to the unit circle we encounter the weak instruments problem for any distanced lags. On the other hand, we can expect better test power to alternatives with substantially lower $\rho(\boldsymbol{\Phi})$.

Remark 2. If the model contains time effects λ_t, the test statistic is based on variables in deviations from the cross-sectional averages $\check{\boldsymbol{y}}_{i,t} \equiv \boldsymbol{y}_{i,t} - (1/N)\sum_{i=1}^{N}\boldsymbol{y}_{i,t}$ rather than levels (similarly to the standard GMM treatment).

Remark 3. 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

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

least one $\Delta \boldsymbol{y}_{i,t}\boldsymbol{y}'_{i,t-1}$ $(t > 1)$ term is available, the test statistic can be computed. The only difference as compared to the unbalanced case is that individual contributions to $\overline{\Delta \boldsymbol{y}_{i,t}\boldsymbol{y}'_{i,t-1}}_T$ are no longer simple averages with $T-1$ terms involved, but have individual specific number of observations $T(i)-1$.

Remark 4. The testing procedure remains valid if instead of $\overline{\Delta \boldsymbol{y}_{i,t}\boldsymbol{y}'_{i,t-1}}_T$ we can investigate the rank of $\boldsymbol{G}_N\overline{\Delta \boldsymbol{y}_{i,t}\boldsymbol{y}'_{i,t-1}}_T\boldsymbol{F}_N$ as suggested by KP, for any full rank matrices $\text{plim}_{N\to\infty}\boldsymbol{G}_N = \boldsymbol{G}$ and $\text{plim}_{N\to\infty}\boldsymbol{F}_N = \boldsymbol{F}$. One interesting case is obtained when we set $\boldsymbol{G}_N = \boldsymbol{I}_m$ and $\boldsymbol{F}_N^{-1} = \frac{1}{N}\sum_{i=1}^N \frac{1}{T-1}\sum_{t=2}^T \boldsymbol{y}_{i,t-1}\boldsymbol{y}'_{i,t-1}$, as in this case we are testing the rank of pooled OLS estimator $\hat{\boldsymbol{\Pi}}$. Even though the estimator itself is inconsistent (due to the presence of fixed effects in the model), it can be used for consistent estimation of $\text{rk}\,\boldsymbol{\Pi}_0$.

5. Monte Carlo

To the best of our knowledge only the BHP study provides results on cointegration analysis for panels with fixed T.[17] Hence, for the main building blocks of the finite-sample studies performed in this paper we take the setups from BHP, but we provide extended range of scenarios. Only *bivariate* panels are considered, thus the only null hypothesis we are testing is:

$$\mathbf{H}_0 : \text{rk}\,\boldsymbol{\Pi} = 1 \tag{12}$$

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

$$\boldsymbol{y}_{i,0} = \boldsymbol{\Upsilon}_i\boldsymbol{\mu}_i + \boldsymbol{\varepsilon}_{i,0}. \tag{13}$$

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

$$\boldsymbol{\varepsilon}_{i,t} = \theta\sum_{j=1}^N \omega_{i,j}\boldsymbol{\zeta}_{j,t} + \boldsymbol{\zeta}_{i,t}, \quad i = 1,\dots,N; \quad t = 0,\dots,T.$$

$$\boldsymbol{\zeta}_{i,t} \sim IID(\mathbf{0}_2, \boldsymbol{\Sigma}), \quad i = 1,\dots,N; \quad t > 0,\dots,T.$$

$$\boldsymbol{\zeta}_{i,0} \sim IID\left(\mathbf{0}_2, \sum_{j=0}^M \boldsymbol{\Phi}^j\boldsymbol{\Sigma}\boldsymbol{\Phi}^{j'}\right), \quad i = 1,\dots,N.$$

[17] Other studies, like Mutl (2009) adapted setups of BHP.

In what follows we will alow for cross-sectional heterogeneity in $\boldsymbol{\Upsilon}_i$ but not in $\boldsymbol{\Sigma}$. We set $M = 50$[18] and the number of Monte Carlo replications $B = 10000$.

We generate the individual heterogeneity $\boldsymbol{\mu}_i$ using exactly the same procedure as in BHP:

$$\boldsymbol{\mu}_i = \tau \left(\frac{q_i - 1}{\sqrt{2}} \right) \check{\boldsymbol{\eta}}_i, \quad q_i \sim \chi^2(1), \quad \check{\boldsymbol{\eta}}_i \sim \mathrm{N}(\mathbf{0}_2, \boldsymbol{\Sigma}). \tag{14}$$

We assume that the error terms are normally distributed i.i.d. both across individuals and time with zero mean and variance-covariance matrix $\boldsymbol{\Sigma}$ (to be specified later).

Before summarizing the Design parameters for this Monte Carlo study recall that $\boldsymbol{\Pi}$ can be rewritten as (for $m = 2$):

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

We set $\lambda = 0$ to study the size of the test, while non-zero values of λ are used to investigate power. In order to reduce the dimensionality of the parameter space we assume that vectors $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ are of the following structure:

$$\boldsymbol{\alpha} = \alpha \boldsymbol{\imath}_2, \quad \boldsymbol{\beta}' = (1, -0.2).$$

All Design parameters are summarized in Table 1.[19]

Comparing our Designs to those present in the literature, we can see that Design **3** of BHP is achieved when $\alpha = -0.5$.

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. Spatial correlation matrix $\boldsymbol{W}_N$ is assumed to be **1 ahead - 1 behind** circular, so that every individual i is directly linked only with individuals $i - 1$ and $i + 1$.[20] The particular choice of the spatial matrix $\boldsymbol{W}_N$ 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.[21] Thus, we suspect that by choosing this particular matrix

---

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

[19] Later we will use notation $\boldsymbol{\Upsilon}^{(i)}$ with i indicating the particular row of Table 1.

[20] The circle is closed by connecting $i = 1$ with $i = N$.

[21] For a graphical illustration see Figure 2 of the aforementioned paper.

Table 1: Design parameters.

N	T	τ	α	θ	λ	vech $\boldsymbol{\Sigma}$	$\boldsymbol{\Upsilon}$
50	3	1	-.1	.0	-.700	$(.05, .03, .05)'$	$0.5\boldsymbol{I}_2$
250	5	5	-.5	.5	-.300		$\boldsymbol{I}_2$
500	7				-.100		$1.5\boldsymbol{I}_2$
					-.050		$\boldsymbol{I}_2 - \boldsymbol{\Phi}^{10}$
					-.010		$\begin{pmatrix} .85 & .15 \\ .00 & .85 \end{pmatrix}$
					-.005		
					.000		

we will put the proposed cointegration test under close to the *least favorable* conditions in terms of size distortions.

As we have discussed in Section 2 in the effect non-stationary case the particular choice of $\{\boldsymbol{\Upsilon}, \boldsymbol{\Sigma}\}$ and τ might substantially influence the performance of the test statistic. For this reason we consider two different choices of $\boldsymbol{\Sigma}$ matrix.

The choice of $\boldsymbol{\Upsilon}^{(4)}$ 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.[22] $\boldsymbol{\Upsilon}^{(5)}$ is based on the estimates in Arellano (2003a) 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 $\boldsymbol{\Upsilon}$ matrices the power curve might not be monotonic because λ not only controls the rank of $\boldsymbol{\Pi}$ but as well (indirectly) the eigenvalues of the $\mathrm{E}[\boldsymbol{u}_{i,t-1}\boldsymbol{y}'_{i,t-1}]$ matrix. Hence, for some specific choices of $\boldsymbol{\Upsilon}$ we can observe the weak instruments problem of the AB moment conditions that is not caused by the reduced rank of $\boldsymbol{\Pi}$ matrix.

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

5.1. Results

The Results for all Designs are summarized in Tables B.3-B.6, henceforth Tables. All rejection frequencies are rounded two digits. Empty spots indicate maximal power of 1.00. Numbers in **bold** indicate that the actual size is equal to the nominal one.

General Trends. First of all, we can observe that rejection frequencies are monotonically declining in $|\lambda|$ for the vast majority of Designs without spatial dependence. As we have discussed in Section 4.2 this property should not be taken as granted for the rk-J test (as dependence on $\boldsymbol{\Phi}$ is non-linear). For lower values of N the test tends to be undersized for $T = 3$ and oversized for $T = 7$.[23] In the effect stationary case τ does not play substantial role and only affects the $\boldsymbol{V}$ 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.

As we have mentioned in the previous section, the comparison between different values of α is not totally *fair* and should be interpreted with caution (in the effect stationary case R^2_Δ for $\alpha = -0.5$ is roughly 5 times higher than the one for $\alpha = -0.1$.). 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 $\alpha = -0.1$.

Spatial Dependence. 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

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

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 i.i.d. assumption.

Effect Non-stationarity and Non-monotonic power curves. Firstly, we consider rejection frequencies for $\boldsymbol{\Upsilon} = 0.5\boldsymbol{I}_m$ as this case is most exceptional in terms of observed patters. 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 $\mathrm{E}[\Delta \boldsymbol{y}_{i,t}\boldsymbol{y}'_{i,t-1}]$ is *negative*, driving the whole expression towards zero matrix (recall the analysis in Hayakawa (2009) for the univariate case). Thus, we have a weak instrument problem under the alternative hypothesis that is not induced by cointegration. These patterns are present irrespective of whether spatial dependence is present or not.[24] By varying λ parameter we directly vary the relative contributions of fixed effects and idiosyncratic parts of the variance components in $\operatorname{var} \boldsymbol{y}_{i,t}$. For larger values of $|\lambda|$ the fixed effects 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.

As it can be expected, the results for $\boldsymbol{\Upsilon} = 1.5\boldsymbol{I}_m$ 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 $\boldsymbol{\Upsilon}^{(4)}$ seem to combine the properties of both $\boldsymbol{\Upsilon}^{(3)}$ and $\boldsymbol{\Upsilon}^{(1)}$.[25] Finally, the results of $\boldsymbol{\Upsilon}^{(5)}$ are

[24]Some preliminary MC results, not presented in this paper suggest that effect of τ in this setup is not-monotonic. In the sense that higher values of τ will 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 $\tau \approx 6.2$.

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

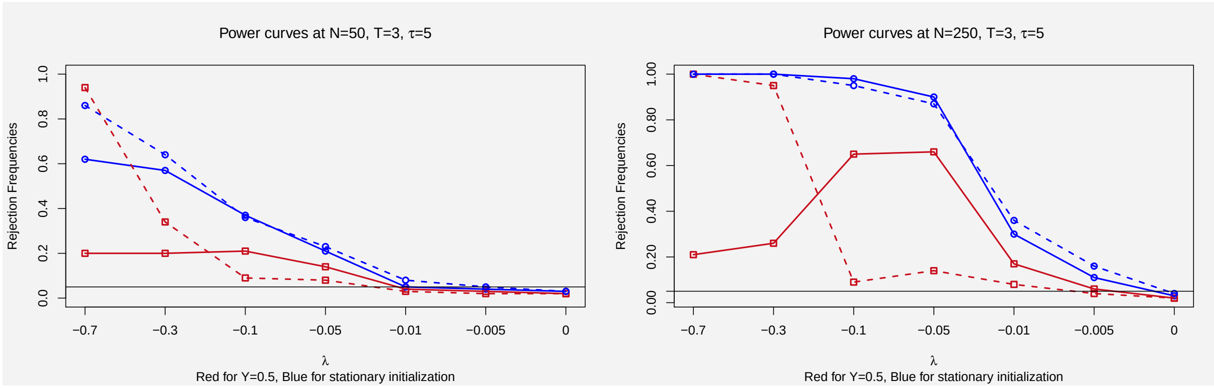

Figure 1: Red (squares) $\boldsymbol{\Upsilon} = 0.5\boldsymbol{I}_2$, Blue (circles) $\boldsymbol{\Upsilon} = \boldsymbol{I}_2$. No Spatial Dependence $\theta = 0$. Straight line $\alpha = -0.1$. Dashed line $\alpha = -0.5$.

somewhat in between those of $\boldsymbol{\Upsilon}^{(1)}$ and $\boldsymbol{\Upsilon}^{(2)}$, but are slightly closer to $\boldsymbol{\Upsilon}^{(2)}$. It serves as an indication that the off-diagonal element in $\boldsymbol{\Upsilon}^{(5)}$ is not of any great importance (given the particular choice of other design parameters).

6. Empirical Illustration

In this section, we use the rk-J procedure to test for cointegration in Spanish firm panel dataset covering **1983-1990** of **738** manufacturing companies as in Alonso-Borrego and Arellano (1999). We construct a bivariate PVAR(1) model where dependent variables are log's of employment and wages. It is reasonable to assume that time effects are present in the model so we explicitly consider variables in their deviations from the cross-sectional averages. Several alternative approaches for cointegration testing are considered.

Firstly, we apply the rk test of KP directly to GMM estimates $\hat{\boldsymbol{\Pi}}$. We restrict 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 "System" estimator of Blundell and Bond (1998) which incorporates moment conditions based on the initial condition. Secondly, the LR tests based on the Transformed Maximum Likelihood function of BHP (LR-TMLE) and Conditional Maximum Likelihood function of Arellano (2003a) (LR-CMLE) are considered. Finally, the **rk-J** test of Section 4.2 is considered. Under $\mathbf{H}_0 : \operatorname{rk} \boldsymbol{\Pi} = 1$ all tests have limiting χ^2 distribution with one degree of freedom.

consecutively the *weak instrument problem under alternative* is less pronounced.

Note that we present results for **AB-GMM** for informal comparison, as under $\mathbf{H}_0$ this estimator is not consistent. Results are summarized in Table 2:

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

Table 2: Cointegration testing. The 5% critical value is 3.84. In the parenthesis for GMM estimators test statistics based on Windmeijer (2005) corrected 2-step S.E.

From Table 2 we can see that only the rk-J test based on the AH moment conditions clearly rejects $\mathbf{H}_0$. Results for Sys-GMM estimator are mixed, as based on Windmeijer (2005) corrected S.E. the null hypothesis is not rejected, while using conventional two-step S.E. hypothesis is rejected (with $p-$ value of 0.0272). Numerous reasons might account for differences in conclusions. First of all, we suspect 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 $\mathbf{H}_0$. Hayakawa and Nagata (2012) provide some evidence based on an incremental Sargan test in support of the latter statement.[26] Another explanation of results in Table 2 might be the low power of cointegration test used directly on the estimate of $\boldsymbol{\Pi}$. Some preliminary MC results suggest that for the System estimator it might be the case, because the (size adjusted) power of the rk-J test dominates the power of rk-SysGMM in most setups of Table 1.

Now we turn our attention to Likelihood Ratio tests. Based on analytical results in this paper for $T = 2$ we can suspect that likelihood procedures under $\mathbf{H}_0$ of cointegration do not control size, 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

[26]However, this testing procedure can not be used if series are cointegrated.

the possibility that it can be one of the reasons for divergence in conclusions.[27]

7. Conclusions

In this paper we have studied the properties of the standard Anderson and Hsiao (1982) moment conditions in a PVAR(1) when the process is cointegrated. Under the assumptions similar to Binder et al. (2005) we have shown that these moment conditions are of reduced rank if and only if the process is cointegrated. Based on this observation we have proposed a rank based test to test the null hypothesis of cointegration, that is robust to time-series heteroscedasticity and can accommodate unbalanced sampling. Furthermore, we have proved that cointegration testing procedure in Binder et al. (2005) is not valid due to the singularity of the Hessian matrix both under unit roots and cointegration. In the Monte Carlo study we found evidence that the new test is reasonably sized and has good power properties in most cases but might exhibit non-monotonic power curves for models with substantial effect non-stationarity. We have applied our testing procedure to the Spanish manufacturing data of Alonso-Borrego and Arellano (1999) and unlike the test of BHP we have rejected the null hypothesis of cointegration.

References

Abadir, K. M. and J. R. Magnus (2002): "Notation in Econometrics: A Proposal for a Standard," *Econometrics Journal*, 5, 76–90.

Ahn, S. C. and G. M. Thomas (2006): "Likelihood Based Inference for Dynamic Panel Data Models," Unpublished Manuscript.

Alonso-Borrego, C. and M. Arellano (1999): "Symmetrically Normalized Instrumental-Variable Estimation using Panel Data," *Journal of Business & Economic Statistics*, 17, 36–49.

Anderson, T. W. and C. Hsiao (1982): "Formulation and Estimation of Dynamic Models Using Panel Data," *Journal of Econometrics*, 18, 47–82.

[27] Arellano (2003b) presents some evidence of time-series heteroscedasticity in this dataset.

Anderson, T. W. and H. Rubin (1949): "Estimation of the Parameters of a Single Equation in a Complete System of Stochastic Equations," *Annals of Mathematical Statistics*, 20, 46–63.

Arellano, M. (2003a): "Modeling Optimal Instrumental Variables for Dynamic Panel Data Models," Unpublished manuscript.

——— (2003b): *Panel Data Econometrics*, Advanced Texts in Econometrics, Oxford University Press.

Arellano, M. and S. Bond (1991): "Some Tests of Specification for Panel Data: Monte Carlo Evidence and an Application to Employment Equations," *Review of Economic Studies*, 58, 277–297.

Baltagi, B. H., G. Bresson, and A. Pirotte (2007): "Panel Unit Root Tests and Spatial Dependence," *Journal of Applied Econometrics*, 22, 339–360.

Binder, M., C. Hsiao, and M. H. Pesaran (2005): "Estimation and Inference in Short Panel Vector Autoregressions with Unit Root and Cointegration," *Econometric Theory*, 21, 795–837.

Blundell, R. W. and S. Bond (1998): "Initial Conditions and Moment Restrictions in Dynamic Panel Data Models," *Journal of Econometrics*, 87, 115–143.

Bond, S., C. Nauges, and F. Windmeijer (2005): "Unit Roots: Identification and Testing in Micro Panels," Working paper.

Dovonon, P. and E. Renault (2009): "GMM Overidentification Test with First Order Underidentification," Working Paper.

Hayakawa, K. (2009): "On the Effect of Mean-Nonstationarity in Dynamic Panel Data Models," *Journal of Econometrics*, 153, 133–135.

——— (2013): "An Improved GMM Estimation of Panel VAR Models," Working paper.

Hayakawa, K. and S. Nagata (2012): "On the Behavior of the GMM Estimator in Persistent Dynamic Panel Data Models with Unrestricted Initial Conditions," Working Paper.

Hsiao, C., M. H. Pesaran, and A. K. Tahmiscioglu (2002): "Maximum Likelihood Estimation of Fixed Effects Dynamic Panel Data Models Covering Short Time Periods," *Journal of Econometrics*, 109, 107–150.

Johansen, S. (1995): *Likelihood-Based Inference in Cointegrated Vector Autoregressive Models*, Advanced Texts in Econometrics, Oxford University Press.

Juodis, A. (2013): "A note on bias-corrected estimation in dynamic panel data models," *Economics Letters*, 118, 435–438.

——— (2014): "First Difference Transformation in Panel VAR models: Robustness, Estimation and Inference," UvA-Econometrics working paper 2013/06.

Kleibergen, F. R. (2005): "Testing Parameters in GMM without Assuming that They are Identified," *Econometrica*, 73, 1103–1123.

Kleibergen, F. R. and R. Paap (2006): "Generalized Reduced Rank Tests Using the Singular Value Decomposition," *Journal of Econometrics*, 133, 97–126.

Kruiniger, H. (2013): "Quasi ML Estimation of the Panel AR(1) Model with Arbitrary Initial Conditions," *Journal of Econometrics*, 173, 175–188.

Lütkepohl, H. (2006): *New Introduction to Multiple Time Series Analysis*, Springer, 2nd ed.

Magnus, J. R. and H. Neudecker (2007): *Matrix Differential Calculus with Applications in Statistics and Econometrics*, John Wiley & Sons.

Mutl, J. (2009): "Panel VAR Models with Spatial Dependence," Working Paper.

Phillips, P. C. B. (1989): "Partially Identified Econometric Models," *Econometric Theory*, 5, 181–240.

Roodman, D. (2009): "A Note on the Theme of Too Many Instruments," *Oxford Bulletin of Economics and Statistics*, 71, 135–158.

Roznitzky, A., D. R. Cox, M. Bottai, and J. Robins (2000): "Likelihoood-Based Inference with Singular Information Matrix," *Bernoulli*, 6, 243–284.

Stock, J. and J. Wright (2000): "GMM with Weak Identification," *Econometrica*, 68, 1055–1096.

Windmeijer, F. (2005): "A Finite Sample Correction for the Variance of Linear Efficient Two-Step GMM Estimators," *Journal of Econometrics*, 126, 25–51.

Appendices

Appendix A. Transformed Maximum Likelihood

Firstly, we define a set of new auxiliary variables, which will be handy during the derivations of differentials:

$$\boldsymbol{Z}_N(\boldsymbol{\kappa}) \equiv \frac{1}{N}\sum_{i=1}^{N}\sum_{t=1}^{T}(\tilde{\boldsymbol{y}}_{i,t} - \boldsymbol{\Phi}\tilde{\boldsymbol{y}}_{i,t-1})(\tilde{\boldsymbol{y}}_{i,t} - \boldsymbol{\Phi}\tilde{\boldsymbol{y}}_{i,t-1})', \quad \boldsymbol{Q}_N(\boldsymbol{\kappa}) \equiv \frac{1}{N}\sum_{i=1}^{N}\sum_{t=1}^{T}\tilde{\boldsymbol{y}}_{i,t-1}(\tilde{\boldsymbol{y}}_{i,t} - \boldsymbol{\Phi}\tilde{\boldsymbol{y}}_{i,t-1})',$$

$$\boldsymbol{M}_N(\boldsymbol{\kappa}) \equiv \frac{T}{N}\sum_{i=1}^{N}(\ddot{\boldsymbol{y}}_{i} - \boldsymbol{\Phi}\ddot{\boldsymbol{y}}_{i-})(\ddot{\boldsymbol{y}}_{i} - \boldsymbol{\Phi}\ddot{\boldsymbol{y}}_{i-})', \quad \boldsymbol{N}_N(\boldsymbol{\kappa}) \equiv \frac{T}{N}\sum_{i=1}^{N}\ddot{\boldsymbol{y}}_{i-}(\ddot{\boldsymbol{y}}_{i} - \boldsymbol{\Phi}\ddot{\boldsymbol{y}}_{i-})',$$

$$\boldsymbol{R}_N \equiv \frac{1}{N}\sum_{i=1}^{N}\sum_{t=1}^{T}\tilde{\boldsymbol{y}}_{i,t-1}\tilde{\boldsymbol{y}}_{i,t-1}', \quad \boldsymbol{P}_N \equiv \frac{T}{N}\sum_{i=1}^{N}\ddot{\boldsymbol{y}}_{i-}\ddot{\boldsymbol{y}}_{i-}', \quad \boldsymbol{\Xi} \equiv \sum_{l=0}^{T-2}(T-1-l)\boldsymbol{\Phi}_0^l.$$

Appendix A.1. Identification with unit roots

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 $\boldsymbol{\kappa}_0$. First of all, we note that $\mathrm{E}[(T-1)\boldsymbol{\Sigma}_0 - \boldsymbol{Z}_N(\boldsymbol{\kappa}_0)] = \mathbf{O}_m$ as well as $\mathrm{E}[\boldsymbol{\Theta}_0 - \boldsymbol{M}_N(\boldsymbol{\kappa}_0)] = \mathbf{O}_m$, hence the contribution to the Hessian matrix of the first four terms in the expression for $\mathrm{d}^2\ell(\boldsymbol{\kappa})$ is zero. Following, e.g. Juodis (2013) it can be shown:

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

Due to the fact that exact expressions for other terms are rather messy in the general case, they will be derived only for the particular case where $\boldsymbol{\Phi}_0 = \boldsymbol{I}_m$. 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:

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

At first we consider expectations of the $\boldsymbol{N}_N(\boldsymbol{\kappa}_0)$ term which we can evaluate based on the general result derived in Lemma A.2 Juodis (2014):

$$\begin{aligned}
\mathrm{E}[\boldsymbol{N}_N(\boldsymbol{\kappa}_0)'] &= (\boldsymbol{I}_m - \boldsymbol{\Phi}_0)\,\mathrm{E}\left[\boldsymbol{u}_{i,0}\boldsymbol{u}_{i,0}'\right](\boldsymbol{I}_m - \boldsymbol{\Phi}_0)'\boldsymbol{\Xi}' + \frac{1}{T}\boldsymbol{\Sigma}_0\boldsymbol{\Xi}' \\
&= \frac{1}{T}\boldsymbol{\Sigma}_0\boldsymbol{\Xi}' = \frac{(T-1)T}{2T}\boldsymbol{\Sigma}_0 = \frac{T-1}{2}\boldsymbol{\Sigma}_0.
\end{aligned}$$

It then similarly follows from the general result in Lemma A.1 Juodis (2014) that:

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

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

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

In Juodis (2014) 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 $\mathbf{O}$):

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

Plugging in these expressions into the formula for the Hessian and evaluating it at the true value $\boldsymbol{\kappa}_0$:

$$\boldsymbol{\mathcal{H}}_\ell = \frac{(T-1)}{2} \begin{pmatrix} T(\boldsymbol{\Sigma}_0 \otimes \boldsymbol{\Sigma}_0^{-1}) & -(\boldsymbol{I}_m \otimes \boldsymbol{\Sigma}_0^{-1})\,\mathbf{D}_m & (\boldsymbol{I}_m \otimes \boldsymbol{\Sigma}_0^{-1})\,\mathbf{D}_m \\ -\mathbf{D}_m'(\boldsymbol{I}_m \otimes \boldsymbol{\Sigma}_0^{-1}) & \mathbf{D}_m'(\boldsymbol{\Sigma}_0^{-1} \otimes \boldsymbol{\Sigma}_0^{-1})\,\mathbf{D}_m & \mathbf{O}_p \\ \mathbf{D}_m'(\boldsymbol{I}_m \otimes \boldsymbol{\Sigma}_0^{-1}) & \mathbf{O}_p & \frac{1}{T-1}\,\mathbf{D}_m'(\boldsymbol{\Sigma}_0^{-1} \otimes \boldsymbol{\Sigma}_0^{-1})\,\mathbf{D}_m \end{pmatrix}. \tag{A.3}$$

Using the formula for the determinant of Partitioned Matrix, that:

$$|\boldsymbol{\mathcal{H}}_\ell| \propto \left(|\boldsymbol{B}| \times \left|\frac{1}{T-1}\,\mathbf{D}_m'(\boldsymbol{\Sigma}_0^{-1} \otimes \boldsymbol{\Sigma}_0^{-1})\,\mathbf{D}_m\right| \times |\mathbf{D}_m'(\boldsymbol{\Sigma}_0^{-1} \otimes \boldsymbol{\Sigma}_0^{-1})\,\mathbf{D}_m|\right), \tag{A.4}$$

with $\boldsymbol{B}$ in this case given by:

$$\begin{aligned} \boldsymbol{B} &= T\left((\boldsymbol{\Sigma}_0 \otimes \boldsymbol{\Sigma}_0^{-1}) - (\boldsymbol{I}_m \otimes \boldsymbol{\Sigma}_0^{-1})\,\mathbf{D}_m\,\mathbf{D}_m^{+}(\boldsymbol{\Sigma}_0 \otimes \boldsymbol{\Sigma}_0)(\mathbf{D}_m\,\mathbf{D}_m^{+})'(\boldsymbol{I}_m \otimes \boldsymbol{\Sigma}_0^{-1})\right) \\ &= T\left((\boldsymbol{\Sigma}_0 \otimes \boldsymbol{\Sigma}_0^{-1}) - (\boldsymbol{I}_m \otimes \boldsymbol{\Sigma}_0^{-1})\,\mathbf{D}_m\,\mathbf{D}_m^{+}(\boldsymbol{\Sigma}_0 \otimes \boldsymbol{\Sigma}_0)(\boldsymbol{I}_m \otimes \boldsymbol{\Sigma}_0^{-1})\right) \\ &= T\left((\boldsymbol{\Sigma}_0 \otimes \boldsymbol{\Sigma}_0^{-1}) - \frac{1}{2}(\boldsymbol{I}_m \otimes \boldsymbol{\Sigma}_0^{-1})(\boldsymbol{I}_m + \mathbf{K}_m)(\boldsymbol{\Sigma}_0 \otimes \boldsymbol{\Sigma}_0)(\boldsymbol{I}_m \otimes \boldsymbol{\Sigma}_0^{-1})\right) \\ &= \frac{T}{2}\left((\boldsymbol{\Sigma}_0 \otimes \boldsymbol{\Sigma}_0^{-1}) - (\boldsymbol{I}_m \otimes \boldsymbol{\Sigma}_0^{-1})\,\mathbf{K}_m(\boldsymbol{\Sigma}_0 \otimes \boldsymbol{\Sigma}_0)(\boldsymbol{I}_m \otimes \boldsymbol{\Sigma}_0^{-1})\right) \\ &= \frac{T}{2}\left((\boldsymbol{\Sigma}_0 \otimes \boldsymbol{\Sigma}_0^{-1}) - \mathbf{K}_m\right). \end{aligned}$$

Then by means of the Matrix determinant Lemma:

$$\begin{aligned} |\boldsymbol{B}| &\propto |\boldsymbol{I}_{m^2} - (\boldsymbol{I}_m \otimes \boldsymbol{\Sigma}_0^{-1})\,\mathbf{K}_m(\boldsymbol{\Sigma}_0 \otimes \boldsymbol{I}_m)||\mathbf{K}_m| \\ &\propto |\boldsymbol{I}_{m^2} - (\boldsymbol{I}_m \otimes \boldsymbol{\Sigma}_0^{-1})\,\mathbf{K}_m(\boldsymbol{\Sigma}_0 \otimes \boldsymbol{I}_m)| \\ &\propto |\boldsymbol{I}_{m^2} - \mathbf{K}_m| = 0, \end{aligned}$$

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 $|\boldsymbol{I}_{m^2} - \mathbf{K}_m| = 0$ follows trivially from the fact that $\operatorname{rk}(\boldsymbol{I}_{m^2} - \mathbf{K}_m) = 0.5m(m-1)$. Hence we have proved that the $\boldsymbol{\mathcal{H}}_\ell$ matrix is not invertible. $\square$

Appendix A.2. Cointegration and $T = 2$

Proposition 2. Using derivations of previous section it follows that:

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

Corresponding $\boldsymbol{B}$ matrix:

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

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

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

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

Appendix B. Tables and Graphs

Table B.3: $\alpha = -0.1$. $\boldsymbol{\Upsilon}^{(1)-(3)}$. Empty entries correspond to 100% empirical rejection frequencies.

		$\boldsymbol{\Upsilon}^{(1)}$							$\boldsymbol{\Upsilon}^{(2)}$							$\boldsymbol{\Upsilon}^{(3)}$						
θ	$\{N,T,\tau\} \setminus \lambda$	-.70	-.30	-.10	-.05	-.01	-.005	.00	-.70	-.30	-.10	-.05	-.01	-.005	.00	-.70	-.30	-.10	-.05	-.01	-.005	.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
	50,3,5	.20	.20	.21	.14	.04	.03	.02	.62	.57	.37	.21	.05	.04	.03	.94	.88	.69	.43	.09	.05	.03
	50,5,1	.98	.95	.91	.75	.16	.08	.04		.99	.93	.76	.17	.08	.04		.99	.96	.81	.19	.09	.04
	50,5,5	.68	.40	.40	.33	.08	.04	.03	.88	.83	.69	.51	.12	.07	.03		.99	.93	.79	.20	.10	.04
	50,7,1		.99	.97	.92	.29	.12	**.05**			.99	.93	.30	.13	**.05**				.95	.31	.14	**.05**
	50,7,5	.97	.63	.52	.48	.15	.07	.04	.95	.92	.84	.71	.22	.11	**.05**			.98	.93	.31	.14	**.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				.99	.44	.17	.04
	250,5,1					.77	.34	.04					.77	.34	**.05**					.78	.35	**.05**
	250,5,5	.97	.65	.80	.85	.46	.17	.03					.65	.28	.04					.76	.34	**.05**
	250,7,1					.92	.51	**.05**					.92	.51	**.05**					.93	.52	**.05**
	250,7,5		.94	.88	.90	.66	.29	.04					.84	.43	**.05**					.91	.49	**.05**
	500,3,1					.75	.32	.04					.75	.33	.04					.77	.34	**.05**
	500,3,5	.24	.31	.85	.91	.45	.16	.03					.62	.26	.04					.75	.33	**.05**
	500,5,1					.97	.62	**.05**					.97	.61	**.05**					.97	.62	**.05**
	500,5,5		.80	.93	.96	.79	.39	.03					.93	.53	**.05**					.97	.60	**.05**
	500,7,1						.80	**.05**						.80	**.05**						.81	**.05**
	500,7,5		.99	.97	.98	.90	.58	**.05**					.99	.72	**.05**						.78	**.05**
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**
	50,3,5	.27	.27	.27	.18	.06	.05	.04	.67	.61	.40	.24	.07	.05	.04	.93	.86	.67	.41	.10	.07	**.05**
	50,5,1	.99	.97	.91	.74	.20	.11	.07	.99	.99	.93	.75	.20	.11	.07		.99	.95	.79	.21	.12	.07
	50,5,5	.75	.50	.49	.39	.11	.07	**.05**	.91	.86	.73	.53	.15	.09	.06		.99	.93	.77	.21	.12	.07
	50,7,1		.99	.98	.91	.31	.16	.08			.99	.92	.32	.16	.08			.99	.94	.33	.17	.08
	50,7,5	.98	.72	.63	.56	.19	.11	.07	.97	.94	.87	.75	.25	.14	.07			.98	.92	.32	.17	.08
	250,3,1				.99	.42	.19	.06				.99	.42	.19	.07				.99	.44	.20	.07
	250,3,5	.40	.47	.80	.78	.24	.11	.04			.98	.92	.33	.15	**.05**				.99	.43	.20	.07
	250,5,1					.75	.36	.08					.74	.36	.08					.75	.37	.08
	250,5,5	.98	.81	.91	.93	.55	.23	**.05**					.65	.31	.07					.74	.36	.08
	250,7,1					.90	.52	.09					.90	.51	.09					.91	.52	.08
	250,7,5		.97	.95	.96	.74	.38	.07					.84	.45	.08					.89	.50	.08
	500,3,1					.72	.34	.08					.72	.34	.08					.73	.35	.08
	500,3,5	.50	.61	.95	.97	.54	.22	.06					.63	.28	.07					.72	.34	.07
	500,5,1					.95	.61	.08					.95	.60	.08					.96	.61	.08
	500,5,5		.92	.98	.99	.86	.47	.07					.92	.54	.08					.95	.60	.08
	500,7,1						.77	.08						.77	.08						.78	.08
	500,7,5			.99	.99	.96	.65	.07					.99	.71	.08					.99	.76	.08

Table B.4: $\alpha = -0.1$. $\boldsymbol{\Upsilon}^{(4),(6)}$. Empty entries correspond to 100% empirical rejection frequencies.

		$\boldsymbol{\Upsilon}^{(4)}$							$\boldsymbol{\Upsilon}^{(5)}$						
θ	$\{N,T,\tau\} \setminus \lambda$	-.70	-.30	-.10	-.05	-.01	-.005	.00	-.70	-.30	-.10	-.05	-.01	-.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		.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			.99	.93	.29	.12	**.05**		.99	.98	.92	.30	.13	**.05**
	50,7,5	.99		.95	.71	.17	.09	.04	.84	.73	.70	.63	.21	.10	.04
	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					.79	.35	**.05**					.77	.34	**.05**
	250,5,5				.99	.52	.19	.04	.95	.93	.97	.97	.63	.27	.04
	250,7,1					.93	.52	**.05**					.92	.51	**.05**
	250,7,5					.71	.32	.04	.99	.97	.99	.99	.83	.42	**.05**
	500,3,1					.77	.33	.04					.75	.32	.04
	500,3,5					.50	.18	.03	.89	.96	.99	.99	.62	.25	.04
	500,5,1					.98	.63	**.05**					.97	.61	**.05**
	500,5,5					.81	.40	.04	.99	.99			.92	.52	.04
	500,7,1						.81	**.05**						.80	**.05**
	500,7,5					.92	.60	**.05**					.99	.72	**.05**
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	.36	.08					.74	.36	.08
	250,5,5	.98	.97	.99	.99	.64	.30	.07	.98	.97	.99	.99	.64	.30	.07
	250,7,1					.90	.51	.09					.90	.51	.09
	250,7,5		.99	.99		.83	.45	.08		.99	.99		.83	.45	.08
	500,3,1					.72	.34	.08					.72	.34	.08
	500,3,5	.97	.99		.99	.62	.28	.06	.97	.99		.99	.62	.28	.06
	500,5,1					.95	.60	.08					.95	.60	.08
	500,5,5					.91	.54	.08					.91	.54	.08
	500,7,1						.77	.08						.77	.08
	500,7,5					.98	.71	.08					.98	.71	.08

Table B.5: $\alpha = -0.5$. $\boldsymbol{\Upsilon}^{(1)-(3)}$. Empty entries correspond to 100% empirical rejection frequencies.

		$\boldsymbol{\Upsilon}^{(1)}$							$\boldsymbol{\Upsilon}^{(2)}$							$\boldsymbol{\Upsilon}^{(3)}$						
θ	$\{N,T,\tau\} \setminus \lambda$	-.70	-.30	-.10	-.05	-.01	-.005	.00	-.70	-.30	-.10	-.05	-.01	-.005	.00	-.70	-.30	-.10	-.05	-.01	-.005	.00
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	.99	.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	.96	.85	.27	.14	.06
	50,7,1			.98	.94	.43	.23	.10				.97	.45	.24	.10				.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	**.05**
	250,5,5			.56	.24	.13	.06	.02				.99	.72	.37	.06					.78	.39	**.05**
	250,7,1					.95	.60	.07					.94	.60	.07					.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**
	500,5,5			.79	.23	.14	.07	.03					.94	.61	**.05**					.97	.64	**.05**
	500,7,1						.85	.06						.85	.06						.85	.06
	500,7,5			.99	.79	.48	.30	.04					.99	.79	.06						.82	**.05**
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	.90	.71	.41	.27	.10	.07	**.05**		.98	.81	.53	.15	.10	.06
	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**	.98	.93	.78	.63	.26	.16	.10	.98	.99	.97	.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	.97		.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

Table B.6: $\alpha = -0.5$. $\boldsymbol{\Upsilon}^{(4),(6)}$. Empty entries correspond to 100% empirical rejection frequencies.

		$\boldsymbol{\Upsilon}^{(4)}$							$\boldsymbol{\Upsilon}^{(5)}$						
θ	$\{N,T,\tau\} \backslash \lambda$	-.70	-.30	-.10	-.05	-.01	-.005	.00	-.70	-.30	-.10	-.05	-.01	-.005	.00
0.0	50,3,1		.95	.74	.46	.11	.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
	50,5,1		.99	.98	.86	.30	.15	.07		.98	.95	.83	.28	.15	.07
	50,5,5		.68	.82	.63	.27	.14	.07	.92	.57	.44	.43	.19	.11	**.05**
	50,7,1				.97	.47	.25	.10			.99	.95	.45	.24	.10
	50,7,5		.84	.94	.81	.46	.25	.10		.82	.62	.62	.34	.20	.08
	250,3,1				.99	.50	.21	.04				.99	.47	.20	.04
	250,3,5		.95	.98	.89	.47	.20	.04	.91	.59	.58	.60	.28	.13	.03
	250,5,1					.84	.44	.06					.82	.42	**.05**
	250,5,5		.97		.99	.84	.47	.06		.89	.81	.88	.67	.35	**.05**
	250,7,1					.96	.63	.07					.94	.60	.07
	250,7,5		.99			.95	.66	.07		.99	.92	.96	.87	.54	.06
	500,3,1					.81	.39	**.05**					.78	.37	**.05**
	500,3,5		.99		.99	.80	.39	.04	.98	.78	.81	.85	.57	.27	.03
	500,5,1					.98	.71	**.05**					.98	.68	**.05**
	500,5,5					.99	.73	**.05**		.97	.94	.98	.93	.60	**.05**
	500,7,1						.87	.06						.85	.06
	500,7,5						.88	.06			.98		.99	.80	.06
0.5	50,3,1		.96	.74	.49	.14	.10	.07	.99	.94	.70	.46	.14	.09	.06
	50,3,5		.58	.48	.30	.11	.08	**.05**	.68	.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	.95	.67	.54	.50	.23	.15	.08
	50,7,1				.96	.47	.28	.14			.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	.96	.78	.73	.71	.33	.17	**.05**
	250,5,1					.81	.45	.10					.79	.44	.09
	250,5,5		.99		.99	.81	.46	.09		.95	.92	.95	.70	.38	.09
	250,7,1					.93	.60	.10					.92	.59	.11
	250,7,5					.94	.63	.10			.97	.99	.87	.55	.10
	500,3,1					.77	.41	.08					.75	.39	.08
	500,3,5					.77	.40	.07	.99	.93	.92	.93	.62	.31	.06
	500,5,1					.97	.68	.09					.96	.66	.09
	500,5,5					.97	.70	.08		.99	.98	.99	.93	.61	.08
	500,7,1						.84	.10						.83	.10
	500,7,5						.85	.09					.99	.79	.09