UPPSALA UNIVERSITET

Working Paper 2012:1

Department of Statistics

An LM type test for idiosyncratic unit roots in a dynamic-factor model with integrated factors

Xingwu Zhou and Martin Solberger

Working Paper 2012:1
June 2012
Department of Statistics
Uppsala University
Box 513
SE-751 20 UPPSALA
SWEDEN

Working papers can be downloaded from www.statistics.uu.se

Title: An LM type test for idiosyncratic unit roots in a dynamic-factor model with integrated factors
Authors: Xingwu Zhou and Martin Solberger
E-mail: xingwu.zhou@statistics.uu.se, martin.solberger@statistics.uu.se

An LM-type test for idiosyncratic unit roots in a dynamic-factor model with integrated factors

Xingwu Zhou* and Martin Solberger†
Department of Statistics, Uppsala University

Abstract

We consider an exact factor model with integrated factors and propose an LM-type test for unit roots in the idiosyncratic component. We show that, when the number of time points (T) tends to infinity the limiting distribution of the LM-statistic is a weighted sum of independent χ_1^2-variables, and when T tends to infinity followed by the number of panel individuals (N) tending to infinity, the limiting distribution is standard normal. In a simulation study, the proposed test shows better local power than the pooled Fisher-type test of Bai and Ng (2004) when the factors are integrated.

JEL: C12, C23
Keywords: Panel unit root, Dynamic factors, Lagrange multiplier

1 Introduction

First generation panel unit root tests, such as in Levin, Lin and Chu (2002) and Im, Pesaran and Shin (2003), are successful in increasing power compared to the simple time series case. However, the assumption of cross-sectional independence is often too restrictive and may result in severe size distortion. This has been demonstrated both when the dependence across units arise through temporal correlations (see e.g. O'Connel, 1998), and through common stochastic trends, i.e. when the panel individuals are cointegrated (see e.g. Urbain and Westerlund, 2006). A second generation of tests has focused on overcoming this issue, whereby a popular way is to assume a factor structure. This is done by e.g. Choi (2002), Phillips and Sul (2003), Moon and Perron (2004), Bai and Ng (2004) and Pesaran (2006, 2007). Conveniently these models coincide with a framework that has been used for a long

*Department of Statistics, Uppsala University, Sweden, Box SE-751-20.
E-mail: xingwu.zhou@statistics.uu.se
†Department of Statistics, Uppsala University, Sweden, Box SE-751-20.
E-mail: martin.solberger@statistics.uu.se

time within different fields of economics, such as business cycle analysis, monetary policy analysis and risk diversification (see e.g. Breitung and Eickmeier, 2006, for an overview).

The main idea is that a low number of factors are able to capture, in large, the comovement of several units, thereby reducing the complexity of the analysis. The panel model may then be decomposed as

$$x_{i,t} = \mu_i + \xi_{i,t} + u_{i,t}; \quad i = 1, 2, ..., N; \ \ t = 1, 2, .., T \tag{1}$$

where μ_i is an individual-specific mean, $u_{i,t}$ is an idiosyncratic component and $\xi_{i,t}$ is an unobservable common component driven by common factors such that $\xi_{i,t} = \boldsymbol{\lambda}_i' \mathbf{f}_t$, where $\mathbf{f}_t = (f_{1t}, f_{2t}, ..., f_{rt})'$ is a $r \times 1$ vector of dynamic factors and $\boldsymbol{\lambda}_i' = (\lambda_{i1}, \lambda_{i2}, ..., \lambda_{ir})$ is a $r \times 1$ vector of factor loadings.[1]

If there is a unit root in $x_{i,t}$, it could be due to non-stationary factors and/or non-stationary idiosyncratic components. However, the literature in general assumes that the factors are stationary. The common component is thus primarily seen as a way to control for cross-sectional dependence rather than a source for cumulative shocks. For this reason, the PANIC approach by Bai and Ng (2004) has gained a wide popularity, as they separately test for unit roots in the factors and the idiosyncratic components.

In this paper we assume that the factors are all $I(1)$. Conditional on this, we derive a homogenous LM-type test for unit roots in the idiosyncratic components, which is demonstrated to have better power than the pooled test in Bai and Ng (2004) when the factors are integrated. The assumption of integrated factors is restrictive, and we propose that it may be relaxed using a similar framework as considered here. Also, including unobservable common stochastic trends has natural implications for a cointegrating context, as considered by e.g. Gengenbach, Palm and Urbain (2006) and Bai, Kao and Ng (2009), to which our framework may be extended.

The rest of the paper is organized as follows: Section 2 presents the framework and Section 3 derives the LM-statistic and its asymptotic properties. In Section 4 we demonstrate, through Monte Carlos simulations, the size of our test-statistic and compare size-adjusted power with the pooled test of Bai and Ng (2004) when the factors are integrated. Section 5 concludes. All mathematical derivations are placed in appendices. Appendix A derives the score and information matrix and proofs of theorems are in Appendix B.

Notation: $\varphi_1(\mathbf{A}) \geq \varphi_2(\mathbf{A}) \geq ... \geq \varphi_n(\mathbf{A})$ denotes the eigenvalues, and $||\mathbf{A}|| = [tr(\mathbf{A}\mathbf{A}')]^{1/2}$ denotes the Frobenius norm of a matrix $\mathbf{A} \in \mathbb{R}^{n \times n}$. $\mathbf{I}_n$ is the $n \times n$ identity matrix and $diag\,(a_1, a_2, ..., a_n)$ is a $n \times n$ diagonal matrix with entries $a_1, a_2, ..., a_n$. For limits, $T \to$ denotes limit taken over T with N fixed and

[1] The model is often combined with a set of regressors, such that the factors and idiosyncratic component make up the error part, i.e. say $y_{i,t} = \mu_i + \boldsymbol{\beta}_i' \boldsymbol{x}_{i,t} + \varepsilon_{i,t}$, where $\varepsilon_{i,t} = \boldsymbol{\lambda}_i' \mathbf{f}_t + u_{i,t}$.

$T, N \to$ denotes sequential limit with limit taken over T followed by limit taken over N. For convergence, $\xrightarrow{p}$ denotes convergence in probability and $\xrightarrow{d}$ denotes convergence in distribution.

2 The framework

We may also write (1) as the vector process,

$$\mathbf{x}_t = (x_{1t}, x_{2t}, ..., x_{Nt})' = \boldsymbol{\mu} + \boldsymbol{\Lambda}\mathbf{f}_t + \mathbf{u}_t, \quad (2)$$

where $\mathbf{u}_t = (u_{1t}, u_{2t}, ..., u_{Nt})'$, $\boldsymbol{\mu} = (\mu_1, \mu_2, ..., \mu_N)'$ and $\boldsymbol{\Lambda}$ is the $N \times r$ matrix of factor loadings. This model is static, which is a special case of the more general dynamic factor model of Forni, Hallin, Lippi and Reichlin (2000) where the factor loadings are lag polynomials. Assuming that the lags are finite, these models have a static representation as in (2), which permits the use of estimation through principal components as in e.g. Stock and Watson (2002) and Bai and Ng (2004). Here we use a likelihood based framework and derive an LM-type statistic for testing idiosyncratic unit roots in the model (2).

If $\mathbf{x}_t$ is non-stationary, it could be due to non-stationary factors and/or non-stationary idiosyncratic components. We assume that $\mathbf{x}_t \sim I(1)$ due to at least $\mathbf{f}_t \sim I(1)$, where $\mathbf{f}_t$ is fully integrated, i.e. $f_{jt} \sim I(1)$ for all j. We then propose a homogenous LM-type-test for a complete unit root in $\mathbf{u}_t$ versus the alternative of $u_{i,t} \sim I(0)$ for all i.[2] A similar framework was recently studied by Breitung and Das (2008). They consider three null hypothesis: (i) $\mathbf{f}_t \sim I(1), \mathbf{u}_t \sim I(1)$ (ii) $\mathbf{f}_t \sim I(0)$, $\mathbf{u}_t \sim I(1)$ and (iii) $\mathbf{f}_t \sim I(1)$, $\mathbf{u}_t \sim I(0)$. The test we propose is thus, in a non-strict sense, a test for Case (i) v.s. Case (iii), where under the latter we have cross-unit cointegration. One of the key results in Breitung and Das (2008), is that none of the statistics they consider handle this case well. We impose the following restrictions:[3]

Assumption 1. The idiosyncratic components are AR(1), $u_{i,t} = \rho u_{i,t-1} + \varepsilon_{i,t}$, where $\rho \in (-1, 1]$ and where $\varepsilon_{i,t} \sim \mathcal{N}\left(0, \sigma_\varepsilon^2\right)$ are i.i.d. with $\sigma_\varepsilon^2 < \infty$.
Assumption 2. The factors are random walks $f_{j,t} = f_{j,t-1} + e_{j,t}$, where $e_{j,t} \sim \mathcal{N}(0, 1)$ are i.i.d.
Assumption 3. The factor loadings are non-random with $||\boldsymbol{\lambda}_i|| < \infty$.
Assumption 4. The processes $e_{j,t}$ and $\varepsilon_{i,t}$ are independently distributed.

[2] This is restrictive, but not unrealistic. Eickmeier (2009), for instance, find that five out of five factors are integrated, and then proceed to test for non-stationarity in the idiosyncratic components. Feasible procedures for finding the number of factors under the assumptions in this paper are Bai and Ng (2002). See also general discussion in Bai (2004).

[3] We maintain these assumptions through out the paper, but propose that all of them may be relaxed at the cost of more complicated mathematical derivations.

Under Assumption 1, the model (2) is also *exact* (or strict) as opposed to *approximate* where cross-sectional correlations in the errors $\varepsilon_{i,t}$ are allowed (c.f. Chamberlain and Rothschild, 1983). The unit variance in Assumption 2 is standard and can be made without loss of generality. Assumptions 3-4 enable us to identify the factor structure, and are also standard in factor analysis. The assumption of normality enables us to formulate the likelihood. Finally, the framework is homogenous, since ρ is the same across individuals.

Our interest here is to test for a full unit root in $\mathbf{u}_t$ versus a homogenous alternative,

$$\begin{aligned} H_0 &: \rho = 1 | f_{jt} \sim I(1), \forall j \in \{1, ..., r\} \\ H_1 &: \rho < 1 | f_{jt} \sim I(1), \forall j \in \{1, ..., r\}. \end{aligned}$$

Because the panel data is non-stationary, we take differences. Let $T^* = T - 1$ and let $\mathbf{D}$ be the $T^* \times T$ first-difference-matrix,

$$\mathbf{D} = \begin{pmatrix} -1 & 1 & 0 & \cdots & 0 \\ 0 & -1 & 1 & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & -1 & 1 \end{pmatrix}.$$

Further, let $\mathbf{f}_j = (f_{j1}, f_{j2}, ..., f_{jT})'$ be a vector that tracks the jth factor over time and in levels. Then, from Assumption 2, $\mathbf{Df}_j \sim \mathcal{N}_{T^*}(\mathbf{0}, \mathbf{I})$. Likewise, let $\mathbf{u}_i = (u_{i1}, u_{i2}, ..., u_{iT})'$. If there is a unit root in $\mathbf{u}_i$, then $\mathbf{Du}_i \sim \mathcal{N}_{T^*}\left(\mathbf{0}, \sigma_\varepsilon^2 \mathbf{I}\right)$. Conversely, for the stationary case $\mathbf{u}_i \sim \mathcal{N}_T\left(\mathbf{0}, \sigma_\varepsilon^2 \mathbf{\Pi}(\rho)\right)$, where (see e.g. Van der Leeuw, 1994)

$$\mathbf{\Pi}_{k,s}(\rho) = \begin{cases} \frac{1}{1-\rho^2} & \text{for } k = s \\ \frac{\rho^{|k-s|}}{1-\rho^2} & \text{for } k \neq s \end{cases}, \tag{3}$$

where $\mathbf{\Pi}_{k,s}$ denotes the element corresponding to the kth row and sth column. It is then straightforward to show that $\mathbf{Du}_i \sim \mathcal{N}_{T^*}\left(\mathbf{0}, \sigma_\varepsilon^2 \mathbf{\Psi}(\rho)\right)$, where,

$$\mathbf{\Psi}_{k,s}(\rho) = \begin{cases} \frac{2}{1+\rho} & \text{for } k = s \\ -\frac{\rho^{||k-s|-1|}(1-\rho)}{1+\rho} & \text{for } k \neq s \end{cases}. \tag{4}$$

Our analysis rests on the observation, that while the non-stationary case is not defined by (3), it is defined by (4), where $\mathbf{\Psi}(1) = \mathbf{I}_{T^*}$ Also, the matrix $\mathbf{\Psi}$ is twice differentiable and continuous around the non-stationary point $\rho = 1$, and is thus well-defined for constructing the LM-statistic.

Let $\mathbf{Y} = \mathbf{XD}'$, where $\mathbf{X} = (\mathbf{x}_1, \mathbf{x}_2, ..., \mathbf{x}_T)$ is the observed panel data. The differenced and stacked panel data, $\mathbf{Y}_v = \text{vec}(\mathbf{Y}) = (\mathbf{y}_2', \mathbf{y}_3', ..., \mathbf{y}_T')'$, has covariance matrix,

$$\mathbf{\Sigma} = E\left(\mathbf{Y}_v \mathbf{Y}_v'\right) = \left(\mathbf{I}_{T^*} \otimes \mathbf{\Lambda\Lambda}'\right) + \sigma_\varepsilon^2 \left[\mathbf{\Psi}(\rho) \otimes \mathbf{I}_N\right]$$

where $\otimes$ denotes Kronecker product. Under the null hypothesis $\rho = 1$, we have that

$$\mathbf{\Sigma} = \left(\mathbf{I}_{T^*} \otimes \mathbf{\Lambda\Lambda}'\right) + \sigma_\varepsilon^2 \mathbf{I}_{NT^*} = \left[\mathbf{I}_{T^*} \otimes \left(\mathbf{\Lambda\Lambda}' + \sigma_\varepsilon^2 \mathbf{I}_N\right)\right],$$

such that, for $\mathbf{y}_t = \Delta \mathbf{x}_t = (\Delta x_{1t}, \Delta x_{2t}, ..., \Delta x_{Nt})'$,

$$E\left(\mathbf{y}_t \mathbf{y}_t'\right) = \mathbf{\Lambda\Lambda}' + \sigma_\varepsilon^2 \mathbf{I}_N = \mathbf{\Omega}.$$

Hence, under the null hypothesis, the maximum likelihood estimators are the same as in the conventional factor model, where, because $\mathbf{\Lambda\Lambda}'$ has at most rank r and is positive semidefinite, we may use a reduced singular value decomposition of $\mathbf{\Lambda}$ and factorize as

$$\mathbf{\Lambda\Lambda}' = \mathbf{AHA}',$$

where $\mathbf{H} = diag\left[\varphi_1(\mathbf{\Lambda}'\mathbf{\Lambda}), \varphi_2(\mathbf{\Lambda}'\mathbf{\Lambda}), ..., \varphi_r(\mathbf{\Lambda}'\mathbf{\Lambda})\right]$ and $\mathbf{A}$ $(N \times r)$ is semi-orthogonal such that $\mathbf{A}'\mathbf{A} = \mathbf{I}_r$. It is well-known (see e.g. Stoica and Jansson, 2009) that the maximum likelihood estimator of $\mathbf{A}$ is the set of eigenvectors of $\mathbf{S} = \frac{1}{T^*}\mathbf{YY}'$ associated with the r largest eigenvalues $\widehat{\varphi}_1(\mathbf{S}) \geq \widehat{\varphi}_2(\mathbf{S}) \geq ... \geq \widehat{\varphi}_r(\mathbf{S})$, and that the MLE:s of $\mathbf{H}$ and σ_ε^2 are

$$\widehat{\sigma}_\varepsilon^2 = \frac{1}{(N-r)} \sum_{i=r+1}^{N} \widehat{\varphi}_i\left(\mathbf{S}\right) \tag{5}$$

and

$$\widehat{\mathbf{H}} = \widehat{\mathbf{\Phi}} - \widehat{\sigma}_\varepsilon^2 \mathbf{I}_r, \tag{6}$$

where $\widehat{\mathbf{\Phi}} = diag\left[\widehat{\varphi}_1(\mathbf{S}), \widehat{\varphi}_2(\mathbf{S}), ..., \widehat{\varphi}_r(\mathbf{S})\right]$. From this we define the MLE of $\mathbf{\Omega}$,

$$\mathbf{S}_{01} = \widehat{\mathbf{\Omega}} = \widehat{\mathbf{A}}\widehat{\mathbf{H}}\widehat{\mathbf{A}}' + \widehat{\sigma}_\varepsilon^2 \mathbf{I}_N. \tag{7}$$

The corresponding sample covariances are

$$\mathbf{Y}_v \mathbf{Y}_v' = \begin{pmatrix} \mathbf{S}_{22} & \mathbf{S}_{23} & \cdots & \mathbf{S}_{2T} \\ \mathbf{S}_{32} & \mathbf{S}_{33} & \cdots & \mathbf{S}_{3T} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{S}_{T2} & \mathbf{S}_{T3} & \cdots & \mathbf{S}_{TT} \end{pmatrix}$$

where $\mathbf{S}_{t,s} = \mathbf{y}_t \mathbf{y}_s'$ for $t, s = 2, 3, ..., T$. For future reference we define

$$\mathbf{S}_0 = \sum_{j=2}^{T} \mathbf{S}_{j,j} = \sum_{t=2}^{T} \mathbf{y}_t \mathbf{y}_t' = \mathbf{YY}', \tag{8}$$

$$\mathbf{S}_{00} = \sum_{j=2}^{T} \sum_{k=2}^{T} \mathbf{S}_{j,k} = \sum_{j=2}^{T} \sum_{k=2}^{T} \mathbf{y}_j \mathbf{y}_k' = \left(\sum_{t=2}^{T} \mathbf{y}_t\right)\left(\sum_{t=2}^{T} \mathbf{y}_t\right)'. \tag{9}$$

3 The LM-statistic

Let $\boldsymbol{\theta} = \left(\text{vec}(\boldsymbol{\Lambda})', \sigma_\varepsilon^2, \rho\right)$ be the parameter set. Assuming normality, the log-likelihood with respect to the differenced and stacked data is

$$l(\boldsymbol{\theta}) = -\frac{NT^*}{2}\log 2\pi - \frac{1}{2}\log|\boldsymbol{\Sigma}| - \frac{1}{2}\mathbf{Y}_v'\boldsymbol{\Sigma}^{-1}\mathbf{Y}_v \ . \tag{10}$$

Based on (10), we derive the LM-type statistic

$$\vartheta(N, T^*) = \left[\frac{\partial l}{\partial \rho}\sqrt{I^{\rho\rho}}\right]_{\rho=1}, \tag{11}$$

where, from Appendix A1,

$$\left.\frac{\partial l}{\partial \rho}\right|_{\rho=1} = \frac{\sigma_\varepsilon^2}{2}\left[\frac{1}{2}T^* tr(\boldsymbol{\Omega}^{-1}) - tr\left(\boldsymbol{\Omega}^{-1}\mathbf{S}_0\boldsymbol{\Omega}^{-1}\right) + \frac{1}{2}tr\left(\boldsymbol{\Omega}^{-1}\mathbf{S}_{00}\boldsymbol{\Omega}^{-1}\right)\right],$$

and where $I^{\rho\rho}$ is the corresponding lower right scalar of the inverse of the information matrix

$$I(\boldsymbol{\theta})^{-1} = \begin{bmatrix} I_{\boldsymbol{\Lambda}\boldsymbol{\Lambda}'} & I_{\sigma_\varepsilon^2\boldsymbol{\Lambda}'} & I_{\rho\boldsymbol{\Lambda}'} \\ I_{\boldsymbol{\Lambda}\sigma_\varepsilon^2} & I_{\sigma_\varepsilon^2\sigma_\varepsilon^2} & I_{\rho\sigma_\varepsilon^2} \\ I_{\boldsymbol{\Lambda}\rho} & I_{\sigma_\varepsilon^2\rho} & I_{\rho\rho} \end{bmatrix}^{-1} = \begin{bmatrix} I_{11} & I_{12} \\ I_{21} & I_{\rho\rho} \end{bmatrix}^{-1} = \begin{bmatrix} I^{11} & I^{12} \\ I^{21} & I^{\rho\rho} \end{bmatrix}, \tag{12}$$

where

$$\underset{(Nr+1)\times(Nr+1)}{I_{11}} = \begin{bmatrix} I_{\boldsymbol{\Lambda}\boldsymbol{\Lambda}'} & I_{\sigma_\varepsilon^2\boldsymbol{\Lambda}'} \\ I_{\boldsymbol{\Lambda}\sigma_\varepsilon^2} & I_{\sigma_\varepsilon^2\sigma_\varepsilon^2} \end{bmatrix}, \ \underset{(Nr+1)\times 1}{I_{12}} = \begin{bmatrix} I_{\rho\boldsymbol{\Lambda}'} \\ I_{\rho\sigma_\varepsilon^2} \end{bmatrix}, \ \underset{1\times(Nr+1)}{I_{21}} = \begin{bmatrix} I_{\boldsymbol{\Lambda}\rho} & I_{\sigma_\varepsilon^2\rho} \end{bmatrix}.$$

This setup is well defined when $r = 1$, but as we show in Appendix A2, the information matrix has reduced rank when $r > 1$, due to $\frac{1}{2}r(r-1)$ linear combinations in the block $\left[I_{\boldsymbol{\Lambda}\boldsymbol{\Lambda}'} \ \ I_{\sigma_\varepsilon^2\boldsymbol{\Lambda}'} \ \ I_{\rho\boldsymbol{\Lambda}'}\right]$. This is however balanced by another linear combination such that $I_{21}I_{11}^{-1}I_{12}/I_{\rho\rho}$ is constant whenever I_{11} has full rank. Without loss of generality we may therefore assume that the linearly dependent, and hence redundant, rows and columns that reduce the rank of $I(\boldsymbol{\theta})$ have been removed. We then have the following result.

Theorem 1 *For any T, the test-statistic (11) is*

$$\vartheta = \frac{(T-1)\,tr(\mathbf{S}_{01}^{-1}) - 2tr(\mathbf{S}_{01}^{-1}\mathbf{S}_0\mathbf{S}_{01}^{-1}) + tr(\mathbf{S}_{01}^{-1}\mathbf{S}_{00}\mathbf{S}_{01}^{-1})}{\sqrt{2\,(T-1)\,(T-2)\,tr(\mathbf{S}_{01}^{-1}\mathbf{S}_{01}^{-1})}}, \tag{13}$$

where $\mathbf{S_{01}}$, $\mathbf{S_0}$ and $\mathbf{S_{00}}$ are given by (7), (8) and (9) respectively.

The limiting distribution of ϑ over T is a weighted sum of independent χ_1^2 (chi-square with one degrees of freedom) variables as stated in Theorem 2.

Theorem 2 *For $1 \leq r < N$ and $\eta_i = \varphi_i\left(\mathbf{\Lambda}'\mathbf{\Lambda}\right)$ where $\eta_1 \geq \eta_2 \geq ... \geq \eta_r \geq 0$, the asymptotic distribution of the LM-type statistic (11) as $T \to \infty$ is*

$$\vartheta \xrightarrow{d} \frac{1}{\sqrt{2\sum_{i=1}^{N} w_i^2}} \left[\sum_{i=1}^{N} w_i \chi_{1,i}^2 - \sum_{i=1}^{N} w_i \right], \tag{14}$$

where $\chi_{1,i}^2$ are independent over i and where $w_i = \begin{cases} \frac{\sigma_\varepsilon^2}{\eta_i + \sigma_\varepsilon^2} & \text{for } 1 \leq i \leq r \\ 1 & \text{for } r < i \leq N \end{cases}$.

Finding a closed form expression of (14) is non-trivial, even if the weights w_i were known.[4] However, let $\delta_i = \frac{\eta_i}{\sigma_\varepsilon^2}$ where $0 \leq \delta_r \leq \delta_{r-1} \leq ... \leq \delta_1 < \infty$, and hence, for $0 < i \leq r$, $w_i = \frac{1}{1+\delta_i}$ lie between 0 and 1. Thus, (14) has two extreme cases, namely (i) $\lim\limits_{\delta_r \to \infty} w_i = 0$, $\forall i \leq r$, in which case the limiting distribution is the standardized expression

$$\frac{1}{\sqrt{2u}} \left(\chi_u^2 - u\right) \tag{15}$$

with degrees of freedom $u = N - r$, and (ii) $\lim\limits_{\delta_1 \to 0} w_i = 1$, $\forall i \leq r$, in which case the limiting distribution is (15) with degrees of freedom $u = N$. For all other cases we propose that, by fitting first and second moments, (14) can be approximated with the following result.[5]

Proposition 1 *Let ϑ have the weighted χ^2-distribution in (14). Matching first and second moments yields*

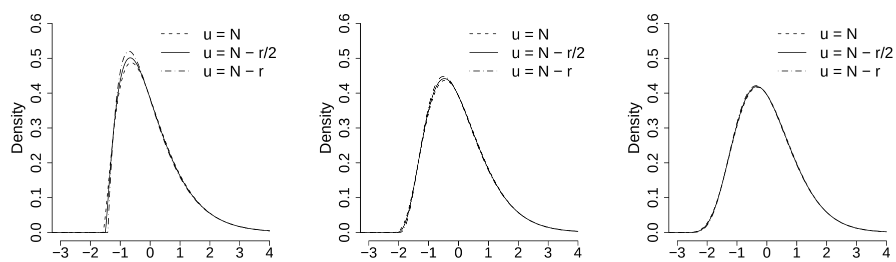
$$\vartheta \overset{app.}{\sim} \frac{1}{\sqrt{2u}} \left(\chi_u^2 - u\right), \tag{16}$$

where $u = \dfrac{\left(\sum_{i=1}^{N} w_i\right)^2}{\sum_{i=1}^{N} w_i^2} = \frac{\left[tr\left(\mathbf{\Omega}^{-1}\right)\right]^2}{tr\left(\mathbf{\Omega}^{-1}\mathbf{\Omega}^{-1}\right)}$.

[4] Except when all weights are either 0 or 1.

[5] The exact distrubution of a weighted sum of χ^2-variables is generally some function of the gamma distribution. For more precise approximations see e.g. Liu, Tang and Zhang (2009) and references therein.

Because $N - r \leq \sum_{i=1}^{N} w_i^2 \leq \sum_{i=1}^{N} w_i \leq N$, in the extreme cases (14) and (16) coincide. The degrees of freedom is thus suitably approximated with $u \approx N - \frac{r}{2}$. Naturally, for large enough N, and assuming $N >> r$, the differences in distribution between the extreme cases and the approximation (16) with degrees of freedom $u = N - r/2$ are very small. Figure 1 illustrates the convergence when keeping the ratio of N to r fixed as 5 to 1 for some different N and r.



(a) $N = 5, r = 1$ (b) $N = 10, r = 2$ (c) $N = 20, r = 4$

Figure 1. Functional forms of (16) with degrees of freedom $N - r$, $N - r/2$ and N with $\{N, r\}$ as $\{5, 1\}$ (a), $\{10, 2\}$ (b) and $\{20, 4\}$ (c).

When the limit is taken over T followed by limit taken over N, the limiting distribution is standard normal, as stated in Theorem 3.

Theorem 3 *As $T, N \to \infty$, the asymptotic distribution of the LM-statistic (11) is standard normal,*

$$\vartheta \xrightarrow{d} \mathcal{N}(0, 1).$$

Note that, even though we use sequential convergence here, in order to get a standard distribution we do not need N to go to infinity, since for any fixed N the asymptotic distribution for T tending to infinity is approximately the standardized χ^2 distribution (16). The number of cross-sectional individuals simply decides the functional form, for any fixed r.

4 Size and power

As the asymptotic distribution of the test-statistic (13) is the sum of independent χ_1^2-variables in (14), one way to find the critical values is to estimate the corresponding eigenvalues and variance and find the ratios $\eta_i/\sigma_\varepsilon^2$. Here we are rather interested in how well the approximation (16) works also for finite T. The test considered here is left-tailed. Thus the asymptotic critical values are found from the transformation of the left, say 5%, percentile of χ^2_u

$$C_\vartheta = \frac{\chi^2_{u,0.05} - u}{\sqrt{2u}}, \tag{17}$$

where, from Proposition 1,

$$u = \frac{(N - r + \sum_{i=1}^{r} \frac{1}{1+\eta_i/\sigma^2_\varepsilon})^2}{N - r + \sum_{i=1}^{r} \frac{1}{(1+\eta_i/\sigma^2_\varepsilon)^2}}.$$

The results in Section 3 show that as long as N and T are large enough, the magnitude of the factor loadings and the error variance will not matter. Here we also illustrate that choosing $u = N - \frac{r}{2}$ will in general provide a good fit. We generate data from Equation (1) with

$$f_{j,t} = f_{j,t-1} + e_{j,t}, \quad e_{j,t} \sim \mathcal{N}(0,1), \; f_{j,0} = 0, \forall j, \tag{18}$$

$$u_{i,t} = \rho_i u_{i,t-1} + \varepsilon_{i,t}, \quad \varepsilon_{i,t} \sim \mathcal{N}(0,\sigma^2_\varepsilon), \; u_{i,0} = 0, \forall i, \tag{19}$$

$$\boldsymbol{\lambda}_i \sim \mathcal{N}_r(\mathbf{0}, \mathbf{I}\sigma^2_\lambda). \tag{20}$$

Figure 2 displays histograms of 10,000 simulated values on the statistic (13) for some finite T and N for $r = 1$ with $\rho_i = 1$ for all i and $\sigma^2_\lambda = \sigma^2_\varepsilon = 1$. The dashed line is the approximate asymptotic distribution (16) with $u = N - \frac{r}{2}$. Note

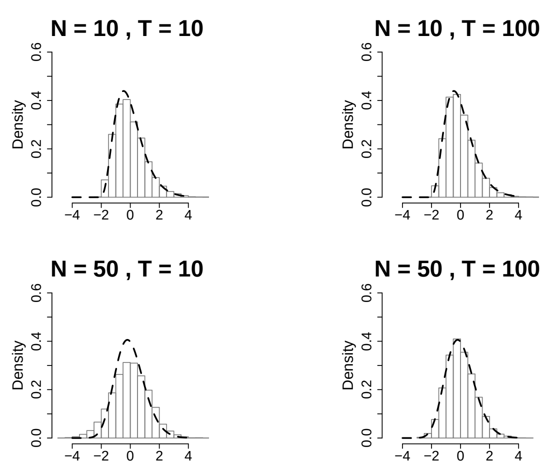


Figure 2. Simulated values on the LM-statistic (13) and approximate asymptotic distribution with degrees of freedom $u = N - r/2$ (dashed line) for $r = 1$.

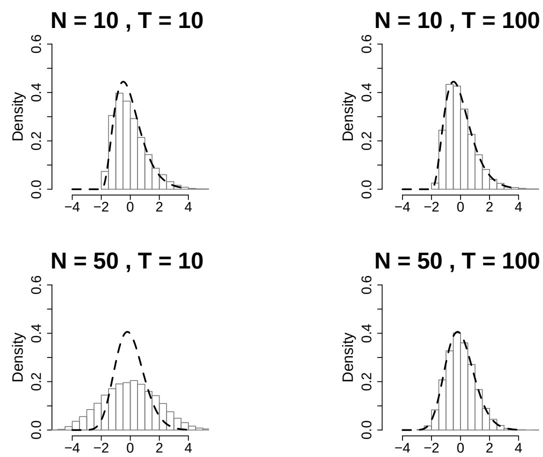


Figure 3. Simulated values on the LM-statistic (13) and approximate asymptotic distribution with degrees of freedom $u = N - r/2$ (dashed line) for $r = 3$.

that even for small N and T the approximation seems to provide a quite good fit. The exception is when N is large and T is small, which is not surprising as the asymptotic distribution of the LM-statistic is derived as $T, N \to \infty$. Figure 3 displays the result for three factors. Again, the approximate asymptotic distribution seems to provide a reasonable fit, except when N is large and T is small, in which case the statistic will be oversized.

For Table 1 we change the variances σ_ε^2 and σ_λ^2 in (19) and (20), where the latter will alter the eigenvalues of $\mathbf{\Lambda}'\mathbf{\Lambda}$. Changing these parameters will thus push the r first weights in (14) towards the extreme cases 0 and 1. We then tabulate the rejection rates using the critical values from the approximate asymptotic distribution. Also, in Table 2 we have tabulated the simulated 5% critical values in $1,000,000$ replications from (18-20) with $\rho_i = 1$ for all i and $\sigma_\lambda^2 = \sigma_\varepsilon^2 = 1$. In Table 1, the rejection rates tend to be too high for small T, but with marginal differences between the different settings. This indicates that the r weights in (14) are likely to have a small impact on the distribution of the test-statistic (13) even for small N and T, and that the critical values in Table 2 should be quite precise also for finite N and T.

Next we compare size-adjusted power of the LM test with the pooled Fisher-type test in Bai and Ng (2004) (BN hereafter). The PANIC approach in BN is to first take first-differences on the panel (2) and then extract principal components, whereby the differenced idiosyncratic components are left as residuals. Augmented Dickey Fuller (ADF) regressions without constants are then applied to the re-

Table 1. Empirical size (%) with respect to the approximate asymptotic distribution

N	T	$\sigma_\lambda^2 = 1$ $\sigma_\varepsilon^2 = 1$		$\sigma_\lambda^2 = 100$ $\sigma_\varepsilon^2 = 1$		$\sigma_\lambda^2 = 1$ $\sigma_\varepsilon^2 = 100$	
		$r = 1$	$r = 3$	$r = 1$	$r = 3$	$r = 1$	$r = 3$
10	10	6.6	7.7	6.9	8.8	5.4	5.8
	25	5.1	4.2	5.1	4.5	4.7	3.9
	50	4.8	4.5	4.8	4.5	4.8	4.7
	100	4.8	4.5	4.9	4.5	4.9	5.0
25	10	8.7	15.9	8.9	16.3	7.8	15.0
	25	5.5	6.3	5.5	6.4	4.9	5.3
	50	5.0	5.3	5.0	5.5	4.8	4.4
	100	5.1	5.1	5.1	5.1	5.0	4.8
50	10	11.2	21.6	11.3	21.8	10.4	21.1
	25	6.0	8.0	6.0	7.9	5.5	6.7
	50	5.2	5.4	5.2	5.3	5.0	4.8
	100	5.1	4.9	5.1	5.0	4.9	4.6
100	10	14.1	26.8	14.2	26.7	13.6	26.6
	25	6.9	10.0	6.9	10.1	6.5	9.7
	50	5.5	6.4	5.5	6.4	5.3	5.9
	100	5.1	4.9	5.1	4.9	5.1	4.9

Note: The data is generated as $x_{i,t} = \boldsymbol{\lambda}_i'\mathbf{f}_t + e_{i,t}$, where $\Delta\mathbf{f}_t \sim \mathcal{N}_r(\mathbf{0}, \mathbf{I})$, $\boldsymbol{\lambda}_i \sim \mathcal{N}_r(\mathbf{0}, \sigma_\lambda^2\mathbf{I})$ (generated once and then kept fixed), $u_{i,t} = u_{i,t-1} + \varepsilon_{i,t}$, $\varepsilon_{i,t} \sim \mathcal{N}(0, \sigma_\varepsilon^2)$. The replication number is 10,000. The asymptotic critical values from (17) with $u = N - \frac{r}{2}$ are used.

accumulated idiosyncratics. The Fisher-type statistic in BN is

$$P_{\hat{e}}^c = \frac{-2\sum_{i=1}^N log p_{\hat{e}}^c(i) - 2N}{\sqrt{4N}}, \tag{21}$$

where $p_{\hat{e}}^c(i)$ are the p-values of the individual ADF tests. It is demonstrated in BN, that when the factors are indeed random walks the test-statistic (21) has much higher power than taking the mean of the ADF t-statistics in the sense of Im, Pesaran and Shin (2003). We therefore only compare our test with $P_{\hat{e}}^c$. To find the size-adjusted power of $P_{\hat{e}}^c$, a two-step simulation procedure is used. First, we locate the p-values. For the case of a random walk without constant, i.e. $u_t = \rho u_{t-1} + \varepsilon_t$, the OLS estimator of ρ is $\hat{\rho} = \frac{\sum_{t=1}^T u_{t-1}u_t}{\sum_{t=1}^T u_{t-1}^2}$, with the corresponding t-statistic

$$t_T = \frac{T^{-1}\sum_{t=1}^T u_{t-1}\varepsilon_t}{\{T^{-2}\sum_{t=1}^T u_{t-1}^2\}^{1/2}\{s_T^2\}^{1/2}}, \tag{22}$$

where $s_T^2 = \sum_{t=1}^T (u_t - \hat{\rho}u_{t-1})^2/(T-1)$. In 10,000 replications, we simulate the

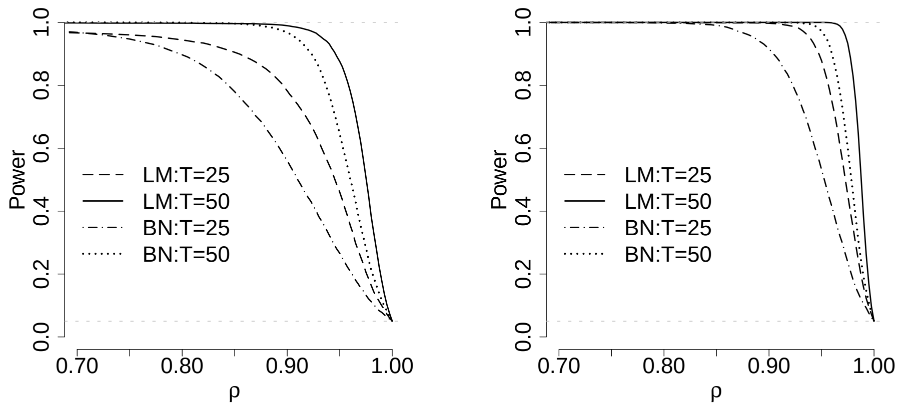
Table 2. Critical values for $r = 1$ to $r = 3$

	N \ T	10	15	20	25	50	75	100	∞
r = 1	10	-1.399	-1.361	-1.350	-1.348	-1.343	-1-342	-1.342	-1.35
	15	-1.530	-1.448	-1.427	-1.418	-1.408	-1.407	-1.406	-1.41
	20	-1.638	-1.515	-1.481	-1.465	-1.446	-1.446	-1.442	-1.44
	25	-1.726	-1.563	-1.513	-1.495	-1.472	-1.469	-1.469	-1.47
	50	-2.087	-1.751	-1.642	-1.596	-1.540	-1.530	-1.534	-1.52
	75	-2.401	-1.895	-1.736	-1.663	-1.574	-1.556	-1.554	-1.55
	100	-2.695	-2.033	-1.813	-1.716	-1.598	-1.575	-1.571	-1.56
	∞								-1.64
r = 2	10	-1.414	-1.343	-1.328	-1.326	-1.321	-1.321	-1.320	-1.34
	15	-1.617	-1.465	-1.426	-1.414	-1.398	-1.396	-1.395	-1.40
	20	-1.798	-1.567	-1.499	-1.474	-1.446	-1.438	-1.436	-1.44
	25	-1.952	-1.645	-1.554	-1.514	-1.472	-1.466	-1.462	-1.47
	50	-2.586	-1.958	-1.750	-1.659	-1.554	-1.534	-1.532	-1.52
	75	-3.087	-2.224	-1.914	-1.772	-1.591	-1.566	-1.558	-1.55
	100	-3.525	-2.461	-2.051	-1.868	-1.630	-1.588	-1.576	-1.56
	∞								-1.64
r = 3	10	-1.421	-1.327	-1.306	-1.301	-1.297	-1.298	-1.297	-1.33
	15	-1.705	-1.483	-1.429	-1.409	-1.386	-1.385	-1.384	-1.40
	20	-1.949	-1.613	-1.518	-1.480	-1.439	-1.432	-1.429	-1.44
	25	-2.157	-1.722	-1.588	-1.532	-1.471	-1.462	-1.459	-1.46
	50	-3.005	-2.161	-1.858	-1.722	-1.562	-1.536	-1.530	-1.52
	75	-3.659	-2.524	-2.078	-1.873	-1.616	-1.574	-1.560	-1.55
	100	-4.207	-2.830	-2.282	-2.011	-1.663	-1.603	-1.584	-1.56
	∞								-1.64

Note: The values correspond to the left 5%-tail. For finite N and T, the critical values are found from $1,000,000$ observations on (13) where the data is generated from (18-20) with $\boldsymbol{\lambda}_i \sim \mathcal{N}_r(\mathbf{0}, \mathbf{I})$ (generated once and then kept fixed) and $\sigma_\varepsilon^2 = 1$. The asymptotic critical values (∞) are found from (17) with $u = N - r/2$.

distribution of (22) with $\rho = 1$, $\varepsilon_t \sim \mathcal{N}(0,1)$ and starting value $u_0 = 0$ for $T = 10, 25, 50, 75, 100, 200$. As in BN, a table is then constructed to map 300 points with corresponding p-values. Second, we do $5,000$ replications generating data from (18-20) with $\rho_i = \rho \in [0,1]$ and $\sigma_\lambda^2 = \sigma_\varepsilon^2 = 1$ to compute the size-adjusted power. The size-adjusted power for the statistics ϑ and $P_{\hat{e}}^c$ for some finite N and T when $r = 1$ is shown in Figure 4.[6]

[6]In BN, the suggested number of lags for the ADF-regression are $k = 4[min(N,T)/100]^{1/4}$. However, since we consider a simple random walk, this choice will systematically overestimate the number of lags, which for small N and T will have a great impact on the size-adjusted power. We therefore choose $k = 0$, and aim to maximize the power of the $P_{\hat{e}}^c$ statistic.



(d) $N = 10$ (e) $N = 25$

Figure 4. Size-adjusted power of ϑ (LM) and $P^c_{\hat{e}}$ (BN) when $r = 1$ for $T = \{25, 50\}$ for $N = 10$ (a) and $N = 25$ (b)

Note that the LM test in general has higher power than $P^c_{\hat{e}}$, except for when T is small and ρ is far from the unit root. Especially, and perhaps most interestingly, the LM-test seems to have higher power close to the unit root.[7] Consequently, we study the local power using a near unit root function in the sense of Phillips (1987), but adjusted for panels (see e.g Breitung and Das, 2005). We let

$$\rho = 1 - \frac{c}{T\sqrt{N}}, \tag{23}$$

for $c = 1$ and $c = 5$. The results for $r = 1$ and $r = 3$ are tabulated in Table 3. Note that as T increases the LM-statistic has substantially higher local power than $P^c_{\hat{e}}$, both for $r = 1$ and $r = 3$. The LM test proposed here is both homogenous and cross-sectionally homoscedastic. We therefore test for robustness against these assumptions. Table 4 displays the result from 5,000 replications where we let (i) $\rho_i \sim U(0.98, 1)$ together with $\sigma_{\varepsilon,i} = 1$ (left side), and (ii) $\sigma_{\varepsilon,i} \sim U(0, 5)$ together with $\rho_i = \rho = 0.99$ for power and $\rho_i = \rho = 1$ for size (right side). The LM-statistic seems to be quite robust against, at least some, heterogeneity in ρ, in terms of size-adjusted power. When considering cross-sectional heteroscedasticity, the size-adjusted power of the LM-statistic remains relatively high compared with BN, but it then suffers from being oversized. This distortion in size is enhanced when N is much larger than T, which is in line with the earlier findings from the size simulations reported in Table 1.

[7]The pattern is the same for $r > 1$, but is left out for ease of exposition.

Table 3. Size-adjusted local power (%) of the LM-statistic and $P_{\hat{e}}^c$

N	T	LM ϑ, r = 1		BN $P_{\hat{e}}^c$, r = 1		LM ϑ, r = 3		BN $P_{\hat{e}}^c$, r = 3	
		$c=1$	$c=5$	$c=1$	$c=5$	$c=1$	$c=5$	$c=1$	$c=5$
10	10	8.5	32.6	6.9	18.9	5.6	10.9	5.8	9.5
	25	9.7	53.0	8.2	32.0	9.0	33.3	7.8	22.1
	50	11.3	63.4	8.6	37.2	9.5	43.0	8.1	29.3
	100	12.6	68.7	9.4	41.6	11.1	53.2	8.4	35.8
	200	12.5	69.7	9.1	41.0	10.9	56.8	8.1	34.5
25	10	8.8	41.1	7.7	22.7	5.2	7.5	5.4	8.1
	25	11.1	71.2	8.5	37.1	11.2	57.7	7.9	28.5
	50	12.6	80.2	9.2	44.7	11.7	71.0	8.0	37.9
	100	13.0	83.7	9.1	47.9	12.9	77.3	8.9	43.9
	200	15.3	85.9	8.6	46.0	12.8	80.8	8.9	42.3
50	10	6.6	30.3	6.8	21.5	5.0	5.8	5.1	7.6
	25	12.8	80.2	8.7	40.3	9.7	59.4	7.5	31.6
	50	13.8	86.2	8.8	46.0	13.8	83.1	9.0	43.8
	100	14.7	90.1	9.2	50.1	13.9	86.2	9.4	46.8
	200	15.5	91.5	9.1	51.1	15.1	89.9	9.2	45.2
100	10	6.1	19.6	6.4	17.9	5.1	5.3	5.3	6.5
	25	12.3	82.9	8.3	41.4	8.7	55.7	8.1	32.9
	50	14.4	89.5	9.4	49.4	13.7	85.5	9.8	46.9
	100	14.7	92.4	9.6	53.6	15.5	91.0	9.2	51.2
	200	16.6	93.3	9.3	51.7	15.7	92.2	9.3	50.6
200	10	5.7	10.6	5.9	12.9	4.9	5.0	5.1	5.9
	25	11.1	79.9	8.5	41.5	7.6	43.1	7.1	28.1
	50	15.2	92.4	9.3	49.7	12.3	82.9	8.6	44.3
	100	16.5	93.5	9.3	53.5	14.7	92.1	9.4	52.0
	200	17.0	94.9	8.7	50.9	15.4	93.5	9.3	53.2

Note: The data is generated as $x_{i,t} = \boldsymbol{\lambda}_i'\mathbf{f}_t + u_{i,t}$, where $\Delta\mathbf{f}_t \sim \mathcal{N}_r(\mathbf{0},\mathbf{I})$, $\boldsymbol{\lambda}_i \sim \mathcal{N}_r(\mathbf{0},\mathbf{I})$ (generated once and then kept fixed), $u_{i,t} = \rho u_{i,t-1} + \varepsilon_{i,t}$, $\varepsilon_{i,t} \sim \mathcal{N}(0,1)$, $\rho = 1 - \frac{c}{T\sqrt{N}}$ and the replication number is 5,000.

5 Conclusions

In this paper we derive a homogenous LM test for idiosyncratic unit roots in the exact static factor model conditional on that all the factors are Gaussian random walks. We show that the asymptotic distribution, as T tends to infinity for any fixed N, is a weighted sum of independent χ_1^2-variables, and standard normal as T tends to infinity followed by N tending to infinity. The exact distribution for any fixed N over T is non-trivial but, fitting first and second moments, it can be appropriately approximated by a χ^2-distribution with $N-r/2$ degrees of freedom, where r is the number of non-stationary factors.

We carry out monte-carlo simulations to investigate the size and power properties of the LM test. The approximate asymptotic distribution is demonstrated

Table 4. Robustness to heterogeneity (left side) and cross-sectional heteroscedasticity (right side)

N	T	$\sigma_{\varepsilon,i}=1,\forall i,\rho_i\sim U(0.98,1)$				$\sigma_{\varepsilon,i}\sim U(0,5)$, power: $\rho_i=0.99,\forall i$, size: $\rho_i=1,\forall i$							
		$r=1$		$r=3$		$r=1$				$r=3$			
		LM^p	BN^p	LM^p	BN^p	LM^s	LM^p	BN^s	BN^p	LM^s	LM^p	BN^s	BN^p
10	10	6.1	5.8	5.3	5.4	8.7	6.5	5.9	4.5	7.7	4.8	9.6	5.1
	25	9.2	8.0	9.0	7.7	7.8	11.0	6.1	6.9	5.2	8.3	7.2	6.5
	50	19.9	13.6	16.1	12.3	7.7	19.1	5.9	9.9	6.0	15.7	7.7	8.9
	100	48.7	29.6	38.8	24.9	7.0	41.5	5.3	19.9	6.0	31.3	7.8	16.0
	200	86.1	66.7	75.0	55.7	7.2	71.3	5.1	42.0	6.4	59.0	8.5	30.0
25	10	6.8	6.1	5.2	5.3	12.3	8.4	9.3	6.4	18.0	4.6	13.2	4.8
	25	13.6	9.7	13.4	9.0	10.5	18.4	7.3	8.9	9.4	14.5	8.6	7.8
	50	36.1	19.2	30.2	17.1	9.7	38.5	6.6	17.5	9.6	30.7	7.2	14.0
	100	78.2	48.2	73.1	43.3	9.6	75.9	5.7	42.5	9.7	65.3	6.5	34.1
	200	99.5	92.5	99.0	88.5	10.5	98.6	5.8	85.4	10.0	95.9	6.8	75.1
50	10	6.1	6.1	5.0	5.0	15.3	7.1	12.7	7.2	22.8	4.4	19.7	4.9
	25	19.6	11.7	13.8	9.7	11.4	28.0	8.3	13.4	11.7	17.2	8.8	10.1
	50	51.5	25.4	47.8	25.2	11.5	60.8	6.6	29.4	10.4	54.6	7.9	25.6
	100	93.6	66.6	89.5	61.6	10.8	96.9	5.6	72.1	10.2	93.5	6.1	64.0
	200	99.9	99.3	99.9	98.6	9.9	100	5.9	99.4	10.1	100	6.0	98.1
100	10	6.0	6.5	5.1	5.2	17.4	5.4	16.2	6.2	27.7	3.7	29.6	4.2
	25	30.2	15.9	17.7	13.9	12.5	37.3	8.5	17.1	14.3	21.3	11.0	14.3
	50	81.8	45.8	76.6	43.9	11.8	83.1	6.6	48.5	11.5	77.6	8.0	43.2
	100	99.9	94.3	99.7	92.5	10.6	99.9	6.1	94.2	10.4	99.8	6.6	91.1
	200	100	100	100	100	10.8	100	6.9	100	10.6	100	6.9	100
200	10	6.0	6.5	4.9	5.2	19.9	4.7	20.1	5.3	31.2	3.5	44.1	3.9
	25	46.8	25.4	22.0	17.4	12.8	51.3	8.5	25.2	17.0	25.0	12.9	16.6
	50	98.4	73.6	95.0	67.0	11.4	98.1	6.4	74.0	12.4	94.5	9.4	64.5
	100	100	99.9	100	99.8	11.3	100	5.8	99.9	10.8	100	6.6	99.6
	200	100	100.0	100	100	10.3	100	6.4	100	10.3	100	7.0	100

Note: For superscripts: p denotes size-adjusted power and s denotes size. The data is generated as $x_{i,t}=\boldsymbol{\lambda}_i'\mathbf{f}_t+u_{i,t}$, where $\Delta\mathbf{f}_t\sim\mathcal{N}_r(0,\mathbf{I})$, $\boldsymbol{\lambda}_i\sim\mathcal{N}(\mathbf{0},\mathbf{I}_r)$ (generated once and then kept fixed) and $u_{i,t}=\rho_i u_{i,t-1}+\varepsilon_{i,t}$ with $\varepsilon_{i,t}\sim\mathcal{N}(0,\sigma^2_{\varepsilon,i})$. For the left side: $\rho_i\sim U(0.98,1)$ (generated once and then kept fixed) and $\sigma_{\varepsilon,i}=1,\forall i$. For the right side: $\sigma_{\varepsilon,i}\sim U(0,5)$ (generated once and then kept fixed), $\rho_i=0.99,\forall i$, for power and $\rho_i=1,\forall i$, for size. The replication number is 5,000.

to fit very well even for finite samples, as long as the time dimension is not much smaller than the cross-sectional dimension, in which case the statistic is oversized. Further, the proposed statistic is demonstrated to have substantially higher local power than the pooled Fisher-type test of Bai and Ng (2004) when all the factors are integrated, and also to be quite robust against some heterogeneity and cross-sectional heteroscedasticity.

Acknowledgements

The authors wish to thank Rolf Larsson and Johan Lyhagen for the valuable comments. Financial support from the Jan Wallander and Tom Hedelius Foundation through the research grants P2005-0117:1 and P2009-0189:1 is gratefully acknowledged.

Appendix A Score and information

Let l be the log-likelihood (10). The score vector is,

$$s(\boldsymbol{\theta}) = \left(\frac{\partial l}{\partial \theta_1}, ..., \frac{\partial l}{\partial \theta_{Nr+2}} \right)' = \left[\left(\frac{\partial l}{\partial \text{vec}\boldsymbol{\Lambda}} \right)', \frac{\partial l}{\partial \sigma_\varepsilon^2}, \frac{\partial l}{\partial \rho} \right]'$$

where, for $\boldsymbol{\Sigma} = E(\mathbf{S})$ (see e.g. Hartley and Rao, 1967),

$$\begin{aligned} \frac{\partial l}{\partial \theta_i} &= -\frac{1}{2} tr \left(\frac{\partial \boldsymbol{\Sigma}}{\partial \theta_i} \boldsymbol{\Sigma}^{-1} \right) + \frac{1}{2} tr \left(\frac{\partial \boldsymbol{\Sigma}}{\partial \theta_i} \boldsymbol{\Sigma}^{-1} \mathbf{S} \boldsymbol{\Sigma}^{-1} \right) \\ &= -\frac{1}{2} tr \left[\frac{\partial \boldsymbol{\Sigma}}{\partial \theta_i} \left(\boldsymbol{\Sigma}^{-1} - \boldsymbol{\Sigma}^{-1} \mathbf{S} \boldsymbol{\Sigma}^{-1} \right) \right]. \end{aligned} \tag{A1}$$

The information matrix, element by element, is

$$I(\boldsymbol{\theta}) = \begin{pmatrix} I_{\theta_1\theta_1} & I_{\theta_1\theta_2} & \cdots & I_{\theta_1\theta_{Nr+2}} \\ I_{\theta_2\theta_1} & I_{\theta_2\theta_2} & \cdots & I_{\theta_2\theta_{Nr+2}} \\ \vdots & \vdots & \ddots & \vdots \\ I_{\theta_{Nr+2}\theta_1} & I_{\theta_{Nr+2}\theta_2} & \cdots & I_{\theta_{Nr+2}\theta_{Nr+2}} \end{pmatrix},$$

where (see e.g. Harville, 1977),

$$I_{\theta_i\theta_j} = -E\left(\frac{\partial^2 l}{\partial \theta_i \partial \theta_j} \right) = \frac{1}{2} tr \left(\frac{\partial \boldsymbol{\Sigma}}{\partial \theta_i} \boldsymbol{\Sigma}^{-1} \frac{\partial \boldsymbol{\Sigma}}{\partial \theta_j} \boldsymbol{\Sigma}^{-1} \right). \tag{A2}$$

We also make use of the following standard results (A3-A6), which can be found in e.g. Magnus and Neudecker (2001, pp. 28-31). For any matrix $\mathbf{A} \in \mathbb{R}^{m \times n}$ with $\mathbf{a}_i$ as its ith column, the vec operator is such that $\text{vec}(\mathbf{A}) = (\mathbf{a}_1', \mathbf{a}_2', ..., \mathbf{a}_n')'$. For any additional matrices $\mathbf{B} \in \mathbb{R}^{n \times q}$, $\mathbf{C} \in \mathbb{R}^{q \times k}$, $\mathbf{D} \in \mathbb{R}^{k \times r}$, we have that,

$$\text{vec}(\mathbf{ABC}) = (\mathbf{C}' \otimes \mathbf{A}) \text{vec}(\mathbf{B}), \tag{A3}$$

$$tr(\mathbf{ABCD}) = \text{vec}(\mathbf{D}')' (\mathbf{C}' \otimes \mathbf{A}) \text{vec}(\mathbf{B}), \tag{A4}$$

where $\otimes$ denotes Kronecker product. Also, for square matrices $\mathbf{A} \in \mathbb{R}^{n \times n}$ and $\mathbf{B} \in \mathbb{R}^{m \times m}$,

$$tr(\mathbf{A} \otimes \mathbf{B}) = tr(\mathbf{A}) \, tr(\mathbf{B}), \tag{A5}$$

and for non-singular matrices $\mathbf{C} \in \mathbb{R}^{q \times q}$ and $\mathbf{D} \in \mathbb{R}^{k \times k}$,

$$(\mathbf{C} \otimes \mathbf{D})^{-1} = \left(\mathbf{C}^{-1} \otimes \mathbf{D}^{-1} \right). \tag{A6}$$

A1 The case $r = 1$

For the one-factor-case, $\boldsymbol{\Lambda}$ is the $N \times 1$ vector $(\lambda_1, \lambda_2, ..., \lambda_N)'$. We then have that

$$\boldsymbol{\Lambda\Lambda}' = \begin{pmatrix} \lambda_1^2 & \lambda_1\lambda_2 & \cdots & \lambda_1\lambda_N \\ \lambda_2\lambda_1 & \lambda_2^2 & \cdots & \lambda_2\lambda_N \\ \vdots & \vdots & \ddots & \vdots \\ \lambda_N\lambda_1 & \lambda_N\lambda_2 & \cdots & \lambda_N^2 \end{pmatrix},$$

where

$$\frac{\partial \mathbf{\Lambda}\mathbf{\Lambda}'}{\partial \lambda_i} = (\mathbf{e}_i \otimes \mathbf{\Lambda}') + (\mathbf{e}_i' \otimes \mathbf{\Lambda}) = \mathbf{\Gamma}_i,$$

where $\mathbf{e}_i$ is a $N \times 1$ vector for which the ith element equals 1 and all other elements equal 0. Thus

$$\frac{\partial \mathbf{\Sigma}}{\partial \lambda_i} = \left(\mathbf{I}_{T^*} \otimes \frac{d\mathbf{\Lambda}\mathbf{\Lambda}'}{d\lambda_i}\right) = \mathbf{I}_{T^*} \otimes \mathbf{\Gamma}_i, \tag{A7}$$

$$\frac{\partial \mathbf{\Sigma}}{\partial \sigma_\varepsilon^2} = \mathbf{\Psi} \otimes \mathbf{I}_N, \tag{A8}$$

$$\frac{\partial \mathbf{\Sigma}}{\partial \rho} = \sigma_\varepsilon^2 \Big(\frac{\partial \mathbf{\Psi}}{\partial \rho} \otimes \mathbf{I}_N\Big). \tag{A9}$$

For the last derivative, let $\mathbf{\Psi}^1 = \frac{\partial \mathbf{\Psi}}{\partial \rho}$, where

$$\mathbf{\Psi}^1(\rho) = \frac{\partial \mathbf{\Psi}}{\partial \rho} = \frac{1}{(1+\rho)^2} \begin{pmatrix} -2 & 2 & \rho^2+2\rho-1 & \cdots & d^* \\ 2 & -2 & 2 & & \\ \rho^2+2\rho-1 & 2 & -2 & & \\ \vdots & & & \ddots & \\ d^* & & & & -2 \end{pmatrix},$$

with $d^* = \left[(T-2)\rho^{T-3} - (T-3)\rho^{T-4}\right](\rho+1) - \rho^{T-3}(\rho-1)$. Then

$$\left.\frac{\partial \mathbf{\Sigma}}{\partial \rho}\right|_{\rho=1} = \sigma_\varepsilon^2 \left[\mathbf{\Psi}^1(1) \otimes \mathbf{I}_N\right], \tag{A10}$$

where,

$$\mathbf{\Psi}^1(1) = \begin{pmatrix} -\frac{1}{2} & \frac{1}{2} & \cdots & \frac{1}{2} \\ \frac{1}{2} & -\frac{1}{2} & \ddots & \vdots \\ \vdots & \ddots & \ddots & \frac{1}{2} \\ \frac{1}{2} & \cdots & \frac{1}{2} & -\frac{1}{2} \end{pmatrix}. \tag{A11}$$

Further, because the stacked data provides us with a one-sample covariance matrix, we have that $\mathbf{\Sigma} = E(\mathbf{Y}_v \mathbf{Y}_v')$, where, using (A6),

$$\mathbf{\Sigma}^{-1} = \left(\mathbf{I}_{T^*} \otimes \mathbf{\Omega}^{-1}\right) \tag{A12}$$

such that, using (A5),

$$tr\left(\mathbf{\Sigma}^{-1}\right) = tr\left(\mathbf{I}_{T^*} \otimes \mathbf{\Omega}^{-1}\right) = T^* tr\left(\mathbf{\Omega}^{-1}\right). \tag{A13}$$

Also, for future reference, note that

$$\mathbf{\Sigma}^{-1}\mathbf{Y}_v\mathbf{Y}_v'\mathbf{\Sigma}^{-1} = \begin{pmatrix} \mathbf{\Omega}^{-1}\mathbf{S}_{22}\mathbf{\Omega}^{-1} & \mathbf{\Omega}^{-1}\mathbf{S}_{23}\mathbf{\Omega}^{-1} & \cdots & \mathbf{\Omega}^{-1}\mathbf{S}_{2T}\mathbf{\Omega}^{-1} \\ \mathbf{\Omega}^{-1}\mathbf{S}_{32}\mathbf{\Omega}^{-1} & \mathbf{\Omega}^{-1}\mathbf{S}_{33}\mathbf{\Omega}^{-1} & \cdots & \mathbf{\Omega}^{-1}\mathbf{S}_{3T}\mathbf{\Omega}^{-1} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{\Omega}^{-1}\mathbf{S}_{T2}\mathbf{\Omega}^{-1} & \mathbf{\Omega}^{-1}\mathbf{S}_{T3}\mathbf{\Omega}^{-1} & \cdots & \mathbf{\Omega}^{-1}\mathbf{S}_{TT}\mathbf{\Omega}^{-1} \end{pmatrix}, \tag{A14}$$

so that,

$$tr\left(\mathbf{\Sigma}^{-1}\mathbf{Y}_v\mathbf{Y}_v'\mathbf{\Sigma}^{-1}\right) = \sum_{j=2}^{T} tr\left(\mathbf{\Omega}^{-1}\mathbf{S}_{jj}\mathbf{\Omega}^{-1}\right) = tr\Big(\mathbf{\Omega}^{-1}\sum_{j=2}^{T}\mathbf{S}_{jj}\mathbf{\Omega}^{-1}\Big). \tag{A15}$$

Because the MLE:s are such that $\frac{\partial l}{\partial \lambda_i}|_{\mathbf{\Lambda}=\hat{\mathbf{\Lambda}}} = 0$ and $\frac{\partial l}{\partial \sigma_\varepsilon^2}|_{\sigma_\varepsilon^2=\hat{\sigma}_\varepsilon^2} = 0$, the score vector is

$$s(\boldsymbol{\theta}) = \left(\mathbf{0}, \left.\frac{\partial l}{\partial \rho}\right|_{\rho=1}\right)',$$

with the non-zero part, from (A1), (A9) and (A12),

$$\left.\frac{\partial l}{\partial \rho}\right|_{\rho=1} = -\frac{1}{2}tr\Big[(\mathbf{\Psi}^1(1)\otimes\mathbf{\Omega}^{-1}) - \left(\mathbf{\Psi}^1(1)\otimes\mathbf{\Omega}^{-1}\right)\mathbf{Y}_v\mathbf{Y}_v'\left(\mathbf{I}_{T^*}\otimes\mathbf{\Omega}^{-1}\right)\Big]\sigma_\varepsilon^2,$$

where, using the results (A3-A6) and (A13-A15),

$$\begin{aligned} tr\Big[\left(\mathbf{\Psi}^1(1)\otimes\mathbf{\Omega}^{-1}\right)\mathbf{Y}_v\mathbf{Y}_v'\left(\mathbf{I}_{T^*}\otimes\mathbf{\Omega}^{-1}\right)\Big] &= \frac{1}{2}tr\Big(\mathbf{\Omega}^{-1}\sum_{j=2}^{T}\sum_{k\neq j}^{T}\mathbf{S}_{j,k}\mathbf{\Omega}^{-1} - \mathbf{\Omega}^{-1}\sum_{j=2}^{T}\mathbf{S}_{j,j}\mathbf{\Omega}^{-1}\Big) \\ &= \frac{1}{2}tr\Big(\mathbf{\Omega}^{-1}\sum_{j=2}^{T}\sum_{k=2}^{T}\mathbf{S}_{j,k}\mathbf{\Omega}^{-1} - 2\mathbf{\Omega}^{-1}\sum_{j=2}^{T}\mathbf{S}_{j,j}\mathbf{\Omega}^{-1}\Big). \end{aligned}$$

Hence

$$\begin{aligned} \left.\frac{\partial l}{\partial \rho}\right|_{\rho=1} = \ & \frac{\sigma_\varepsilon^2}{2}\Big[\frac{1}{2}T^*tr(\mathbf{\Omega}^{-1}) - tr\Big(\mathbf{\Omega}^{-1}\sum_{j=2}^{T}\mathbf{S}_{j,j}\mathbf{\Omega}^{-1}\Big) \\ & + \frac{1}{2}tr\Big(\mathbf{\Omega}^{-1}\sum_{j=2}^{T}\sum_{k=2}^{T}\mathbf{S}_{j,k}\mathbf{\Omega}^{-1}\Big)\Big]. \end{aligned} \tag{A16}$$

For the information matrix we have, using (A2) and (A3-A6),

$$I_{\lambda_i\lambda_j}\big|_{\rho=1} = \frac{1}{2}tr\left(\mathbf{I}_{T^*}\otimes\mathbf{\Gamma}_i\mathbf{\Omega}^{-1}\mathbf{\Gamma}_j\mathbf{\Omega}^{-1}\right) = \frac{1}{2}T^*tr(\mathbf{\Gamma}_i\mathbf{\Omega}^{-1}\mathbf{\Gamma}_j\mathbf{\Omega}^{-1}), \tag{A17}$$

$$I_{\lambda_i\sigma_\varepsilon^2}\big|_{\rho=1} = \frac{1}{2}tr\left[(\mathbf{\Psi}(1)\otimes\mathbf{\Gamma}_i\mathbf{\Omega}^{-1}\mathbf{\Omega}^{-1})\right] = \frac{1}{2}T^*tr(\mathbf{\Gamma}_i\mathbf{\Omega}^{-1}\mathbf{\Omega}^{-1}), \tag{A18}$$

$$I_{\lambda_i\rho}\big|_{\rho=1} = \frac{1}{2}\sigma_\varepsilon^2 tr\left[(\mathbf{\Psi}^1(1)\otimes\mathbf{\Gamma}_i\mathbf{\Omega}^{-1}\mathbf{\Omega}^{-1})\right] = -\frac{1}{4}\sigma_\varepsilon^2T^*tr(\mathbf{\Gamma}_i\mathbf{\Omega}^{-1}\mathbf{\Omega}^{-1}), \tag{A19}$$

$$I_{\sigma_\varepsilon^2\sigma_\varepsilon^2}\big|_{\rho=1} = \frac{1}{2}tr\left[\mathbf{\Psi}(1)\,\mathbf{\Psi}(1)\otimes\mathbf{\Omega}^{-1}\mathbf{\Omega}^{-1}\right] = \frac{1}{2}T^*tr(\mathbf{\Omega}^{-1}\mathbf{\Omega}^{-1}), \tag{A20}$$

$$I_{\sigma_\varepsilon^2\rho}\big|_{\rho=1} = \frac{1}{2}tr\left[\mathbf{\Psi}(1)\,\mathbf{\Psi}^1(1)\otimes\mathbf{\Omega}^{-1}\mathbf{\Omega}^{-1}\right] = -\frac{\sigma_\varepsilon^2}{4}T^*tr(\mathbf{\Omega}^{-1}\mathbf{\Omega}^{-1}), \tag{A21}$$

$$I_{\rho\rho}\big|_{\rho=1} = \frac{1}{2}\sigma_\varepsilon^4 tr\left[\mathbf{\Psi}^1(1)\,\mathbf{\Psi}^1(1)\otimes\mathbf{\Omega}^{-1}\mathbf{\Omega}^{-1}\right] = \frac{\sigma_\varepsilon^4}{8}T^{*^2}tr(\mathbf{\Omega}^{-1}\mathbf{\Omega}^{-1}). \tag{A22}$$

A2 The general case, $r > 0$

For the general case the derivative of $\mathbf{\Sigma}$ by $\lambda_{i,k}$ is

$$\left.\frac{\partial \mathbf{\Sigma}}{\partial \lambda_{i,k}}\right|_{\rho=1} = \mathbf{I}_{T^*} \otimes \mathbf{\Gamma}_{i,k},$$

where $\mathbf{\Gamma}_{i,k} = (\mathbf{e}_i \otimes \boldsymbol{\lambda}_k') + (\mathbf{e}_i' \otimes \boldsymbol{\lambda}_k)$ and where $\boldsymbol{\lambda}_k$ is the kth column vector of $\mathbf{\Lambda}$.

The score part w.r.t. ρ is the same as (A16) while the information becomes a $(Nr+2) \times (Nr+2)$ matrix, where

$$\left. I_{\lambda_{i,k},\lambda_{j,s}}\right|_{\rho=1} = \frac{1}{2} T^* tr(\mathbf{\Gamma}_{i,k} \mathbf{\Omega}^{-1} \mathbf{\Gamma}_{j,s} \mathbf{\Omega}^{-1}), \tag{A23}$$

$$\left. I_{\lambda_{i,k},\sigma_\varepsilon^2}\right|_{\rho=1} = \frac{1}{2} T^* tr(\mathbf{\Gamma}_{i,k} \mathbf{\Omega}^{-1} \mathbf{I}_N \mathbf{\Omega}^{-1}), \tag{A24}$$

$$\left. I_{\lambda_{i,k},\rho}\right|_{\rho=1} = -\frac{1}{4} \sigma_\varepsilon^2 T^* tr(\mathbf{\Gamma}_{i,k} \mathbf{\Omega}^{-1} \mathbf{I}_N \mathbf{\Omega}^{-1}), \tag{A25}$$

and where the information parts w.r.t. σ_ε^2 and ρ are the same as (A20) to (A22).

By construction, the information matrix I has reduced rank when there is more than one factor. To be more precise $Rank\,(I) = (Nr+2) - r(r-1)/2$. However, due to a second linear relationship, this does not affect our inference. First, let

$$\begin{bmatrix} I_{\mathbf{\Lambda}\mathbf{\Lambda}'} \\ I_{\mathbf{\Lambda}\sigma_\varepsilon^2} \\ I_{\mathbf{\Lambda}\rho} \end{bmatrix} = (\mathbf{A}_1, \mathbf{A}_2, ..., \mathbf{A}_r),$$

where $\mathbf{A}_k$ are submatrices, each of size $(Nr+2) \times N$, for $k = 1, ..., r$. Then from the results (A23-A25), we have that the i row of $\mathbf{A}_k$ is

$$\mathbf{a}_{i,k} = \left[\frac{1}{2} T^* tr(\mathbf{\Omega}^{-1} \widetilde{\mathbf{\Gamma}} \mathbf{\Omega}^{-1} \mathbf{\Gamma}_{1,k}), \frac{1}{2} T^* tr(\mathbf{\Omega}^{-1} \widetilde{\mathbf{\Gamma}} \mathbf{\Omega}^{-1} \mathbf{\Gamma}_{2,k}), ..., \frac{1}{2} T^* tr(\mathbf{\Omega}^{-1} \widetilde{\mathbf{\Gamma}} \mathbf{\Omega}^{-1} \mathbf{\Gamma}_{N,k}) \right],$$

where

$$\widetilde{\mathbf{\Gamma}} = \begin{cases} \mathbf{\Gamma}_{j,s} & \text{if } \ i \leq Nr \\ \mathbf{I}_N & \text{if } i = Nr+1 \\ -\frac{\sigma_\varepsilon^2}{2} \mathbf{I}_N & \text{if } i = Nr+2 \end{cases}, \text{ for } i = (s-1)N + j, \ j = 1, ..., N, \ s = 1, ..., r.$$

Also, we have that

$$\sum_{j=1}^N \mathbf{\Gamma}_{j,k} \lambda_{j,l} = \sum_{j=1}^N \left[(\mathbf{e}_j \otimes \boldsymbol{\lambda}_k') + (\boldsymbol{\lambda}_k \otimes \mathbf{e}_j') \right] \lambda_{j,l} = \left(\sum_{j=1}^N \mathbf{e}_j \lambda_{j,l} \otimes \boldsymbol{\lambda}_k' \right) + \left(\boldsymbol{\lambda}_k \otimes \sum_{j=1}^N \mathbf{e}_j' \lambda_{j,l} \right)$$

$$= (\boldsymbol{\lambda}_l \otimes \boldsymbol{\lambda}_k') + (\boldsymbol{\lambda}_k \otimes \boldsymbol{\lambda}_l') = \sum_{j=1}^N \mathbf{\Gamma}_{j,l} \lambda_{j,k},$$

such that

$$\mathbf{a}_{i,k}\boldsymbol{\lambda}_l = \frac{1}{2}T^*\sum_{j=1}^{N} tr(\boldsymbol{\Omega}^{-1}\widetilde{\boldsymbol{\Gamma}}\boldsymbol{\Omega}^{-1}\boldsymbol{\Gamma}_{j,k})\lambda_{j,l} = \frac{1}{2}T^*\sum_{j=1}^{N} tr(\boldsymbol{\Omega}^{-1}\widetilde{\boldsymbol{\Gamma}}\boldsymbol{\Omega}^{-1}\boldsymbol{\Gamma}_{j,k}\lambda_{j,l}) \tag{A26}$$

$$= \frac{1}{2}T^* tr(\boldsymbol{\Omega}^{-1}\widetilde{\boldsymbol{\Gamma}}\boldsymbol{\Omega}^{-1}\sum_{j=1}^{N}\boldsymbol{\Gamma}_{j,k}\lambda_{j,l}) = \frac{1}{2}T^* tr(\boldsymbol{\Omega}^{-1}\widetilde{\boldsymbol{\Gamma}}\boldsymbol{\Omega}^{-1}\sum_{j=1}^{N}\boldsymbol{\Gamma}_{j,l}\lambda_{j,k})$$

$$= \mathbf{a}_{i,l}\boldsymbol{\lambda}_k,$$

where the number of such linear combinations are $\sum_{k=1}^{r}(r-k) = \frac{1}{2}r(r-1)$. Thus the information matrix has reduced rank and is non-invertible. However, another linear relationship will show useful. Let,

$$I_{11} = \begin{pmatrix} I_{\boldsymbol{\Lambda}\boldsymbol{\Lambda}'} & I_{\sigma_\varepsilon^2\boldsymbol{\Lambda}'} \\ I_{\boldsymbol{\Lambda}\sigma_\varepsilon^2} & I_{\sigma_\varepsilon^2\sigma_\varepsilon^2} \end{pmatrix} = (\mathbf{b}_1, \mathbf{b}_2, ..., \mathbf{b}_{Nr+1}),$$

where $\mathbf{b}_l$ are vectors, each of size $(Nr+1)\times 1$, for $l = 1, ..., Nr+1$. Then

$$\mathbf{b}_{Nr+1} = \begin{pmatrix} I_{\sigma_\varepsilon^2\lambda_{1,1}} \\ \vdots \\ I_{\sigma_\varepsilon^2\lambda_{N,1}} \\ \vdots \\ I_{\sigma_\varepsilon^2\lambda_{N,r}} \\ I_{\sigma_\varepsilon^2\sigma_\varepsilon^2} \end{pmatrix} = \begin{pmatrix} \frac{T^*}{2}tr(\boldsymbol{\Gamma}_{1,1}\boldsymbol{\Omega}^{-2}) \\ \vdots \\ \frac{T^*}{2}tr(\boldsymbol{\Gamma}_{N,1}\boldsymbol{\Omega}^{-2}) \\ \vdots \\ \frac{T^*}{2}tr(\boldsymbol{\Gamma}_{N,r}\boldsymbol{\Omega}^{-2}) \\ \frac{T^*}{2}tr(\boldsymbol{\Omega}^{-2}) \end{pmatrix}.$$

Further, we have that

$$I_{12} = \begin{bmatrix} I_{\rho\boldsymbol{\Lambda}'} \\ I_{\rho\sigma_\varepsilon^2} \end{bmatrix} = \begin{pmatrix} I_{\rho\lambda_{1,1}} \\ \vdots \\ I_{\rho\lambda_{N,1}} \\ \vdots \\ I_{\rho\lambda_{N,r}} \\ I_{\rho\sigma_\varepsilon^2} \end{pmatrix} = \begin{pmatrix} -\frac{\sigma_\varepsilon^2}{4}T^*tr(\boldsymbol{\Gamma}_{1,1}\boldsymbol{\Omega}^{-2}) \\ \vdots \\ -\frac{\sigma_\varepsilon^2}{4}T^*tr(\boldsymbol{\Gamma}_{N,1}\boldsymbol{\Omega}^{-2}) \\ \vdots \\ -\frac{\sigma_\varepsilon^2}{4}T^*tr(\boldsymbol{\Gamma}_{N,r}\boldsymbol{\Omega}^{-2}) \\ -\frac{\sigma_\varepsilon^2}{4}T^*tr(\boldsymbol{\Omega}^{-2}) \end{pmatrix},$$

which is also $(Nr+1)\times 1$. Thus

$$I_{12} = -\frac{\sigma_\varepsilon^2}{2}\mathbf{b}_{Nr+1}. \tag{A27}$$

Appendix B Proofs

The maximum likelihood estimators are consistent, implying for (7) that $\widehat{\mathbf{A}}\widehat{\mathbf{H}}\widehat{\mathbf{A}}' \xrightarrow{p} \boldsymbol{\Lambda}\boldsymbol{\Lambda}'$ and $\widehat{\sigma}_\varepsilon^2 \xrightarrow{p} \sigma_\varepsilon^2 \Longrightarrow \widehat{\boldsymbol{\Omega}} \xrightarrow{p} \boldsymbol{\Omega}$. Further, since under the null hypothesis $E(\mathbf{S}_{ii}) = \boldsymbol{\Omega}$ for all i, $\frac{1}{T^*}\sum_{j=2}^{T}\mathbf{S}_{j,j} \xrightarrow{p} \boldsymbol{\Omega}$ by the weak law of large numbers.

To prove Theorem 1, we need the following lemma.

Lemma 1 *If* I_{11} *is non-singular, then*

$$\frac{I_{21}I_{11}^{-1}I_{12}}{I_{\rho\rho}} = \frac{1}{T^*}$$

Proof of Lemma 1. Define $I_{11} = (\mathbf{b}_1, ..., \mathbf{b}_{Nr+1})$ where $\mathbf{b}_i$ is the ith column vector of I_{11} for $i = 1, 2, ..., Nr+1$ and where, from result (A27), $I_{12} = -\frac{\sigma_\varepsilon^2}{2}\mathbf{b}_{Nr+1}$. Also, define $I_{11}^{-1} = (\mathbf{c}_1, ..., \mathbf{c}_{Nr+1})'$, where $\mathbf{c}_i'$ $(1 \times (Nr+1))$ is the ith row vector of the inverse of I_{11} for $i = 1, 2, ..., Nr+1$. By the definition of an inverse, $\mathbf{c}_k'\mathbf{b}_l = 1$ if $k = l$ and 0 otherwise. Hence

$$I_{11}^{-1}I_{12} = -\frac{\sigma_\varepsilon^2}{2}\begin{pmatrix} \mathbf{c}_1' \\ \vdots \\ \mathbf{c}_{Nr+1}' \end{pmatrix}\mathbf{b}_{Nr+1} = \begin{pmatrix} 0 \\ \vdots \\ 0 \\ -\frac{\sigma_\varepsilon^2}{2} \end{pmatrix},$$

and

$$I_{21}I_{11}^{-1}I_{12} = \begin{bmatrix} I_{\mathbf{\Lambda}\rho} & I_{\sigma_\varepsilon^2\rho} \end{bmatrix}\begin{pmatrix} 0 \\ \vdots \\ 0 \\ -\frac{\sigma_\varepsilon^2}{2} \end{pmatrix} = I_{\rho\sigma_\varepsilon^2}(-\frac{\sigma_\varepsilon^2}{2}) = \frac{\sigma_\varepsilon^4}{8}T^* tr(\mathbf{\Omega}^{-1}\mathbf{\Omega}^{-1}) = \frac{I_{\rho\rho}}{T^*}.$$

■

Proof of Theorem 1. Assume that $I(\boldsymbol{\theta})$ has full rank such that (12) exists. Using a standard result, and assuming that I_{11} has full rank, the lower right scalar of the inverse of the information matrix is

$$\begin{aligned} I^{\rho\rho} &= (I_{\rho\rho} - I_{21}I_{11}^{-1}I_{12})^{-1} \\ &= I_{\rho\rho}^{-1}\left(1 - \frac{I_{21}I_{11}^{-1}I_{12}}{I_{\rho\rho}}\right)^{-1}. \end{aligned}$$

Thus, using Lemma 1 and the result (A22),

$$I^{\rho\rho} = I_{\rho\rho}^{-1}\left(1 - \frac{1}{T^*}\right)^{-1} = \left[\frac{\sigma_\varepsilon^4}{8}T^{*^2} tr(\mathbf{\Omega}^{-1}\mathbf{\Omega}^{-1})\right]^{-1}\left(\frac{T^*}{T^*-1}\right). \tag{B1}$$

We know from result (A26) that when $r > 1$, there are $r(r-1)/2$ linear combinations in $I(\boldsymbol{\theta})$ whereby I_{11} is singular. However, we may without loss of generality assume that these redundant columns and rows have been removed, such that

$$I^*(\boldsymbol{\theta}) = \begin{bmatrix} I_{11}^* & I_{12}^* \\ I_{21}^* & I_{\rho\rho} \end{bmatrix},$$

has full rank and where I_{11}^* is invertible such that Lemma 1 holds with $\frac{I_{21}^* I_{11}^{*-1} I_{12}^*}{I_{\rho\rho}} = \frac{1}{T^*}$, leaving the result unchanged. Taking the square root of the estimated (B1) and multiplying it with the estimated (A16) then gives the test-statistic (13). ■

Proof of Theorem 2. Let $\psi_i = \varphi_i(\mathbf{\Omega})$ denote the eigenvalues of $\mathbf{\Omega}$ and let $\eta_i = \varphi_i\left(\mathbf{\Lambda\Lambda}'\right)$ denote the eigenvalues of $\mathbf{\Lambda\Lambda}'$. Then the eigenvalues of $\mathbf{\Omega}$ are $\psi_1 = \eta_1 + \sigma_\varepsilon^2, ..., \psi_r = \eta_r + \sigma_\varepsilon^2, \psi_{r+1} = ... = \psi_N = \sigma_\varepsilon^2$, so that

$$tr(\mathbf{\Omega}^{-1}) = \sum_{i=1}^{N} \frac{1}{\psi_i} = \sum_{i=1}^{r} \frac{1}{\eta_i + \sigma_\varepsilon^2} + \frac{N-r}{\sigma_\varepsilon^2}, \tag{B2}$$

$$tr(\mathbf{\Omega}^{-1}\mathbf{\Omega}^{-1}) = \sum_{i=1}^{N} \left(\frac{1}{\psi_i}\right)^2 = \sum_{i=1}^{r} \left(\frac{1}{\eta_i + \sigma_\varepsilon^2}\right)^2 + \frac{N-r}{\sigma_\varepsilon^4}. \tag{B3}$$

The first term of (13) is, because $\lim\limits_{T\to\infty} \frac{T-1}{\sqrt{(T-1)(T-2)}} = 1$ and $\mathbf{S}_{01} \xrightarrow{p} \mathbf{\Omega}$, and using (B2-B3),

$$\frac{T^* tr(\mathbf{S}_{01}^{-1})}{\sqrt{2T^*\left(T^*-1\right) tr(\mathbf{S}_{01}^{-1}\mathbf{S}_{01}^{-1})}} \xrightarrow{p} \frac{1}{\sqrt{2}} \frac{tr(\mathbf{\Omega}^{-1})}{\sqrt{tr(\mathbf{\Omega}^{-1}\mathbf{\Omega}^{-1})}}$$

$$= \frac{1}{\sqrt{2}} \frac{\sum\limits_{i=1}^{r} \frac{\sigma_\varepsilon^2}{\eta_i+\sigma_\varepsilon^2} + N - r}{\sqrt{\sum\limits_{i=1}^{r}\left(\frac{\sigma_\varepsilon^2}{\eta_i+\sigma_\varepsilon^2}\right)^2 + N - r}} = \frac{1}{\sqrt{2}} \frac{\nu}{\sqrt{\nu^*}} \ ,$$

where $\nu = \sum\limits_{i=1}^{r} \frac{1}{\frac{\eta_i}{\sigma_\varepsilon^2}+1} + N - r = \sum\limits_{i}^{N} w_i$ and $\nu^* = \sum\limits_{i=1}^{r} \frac{1}{\left(\frac{\eta_i}{\sigma_\varepsilon^2}+1\right)^2} + N - r = \sum\limits_{i}^{N} w_i^2$.

The second term of (13) is, because $\lim\limits_{T\to\infty} \frac{\sqrt{(T-1)(T-2)}}{T-1} = 1$, $\frac{1}{T^*}\mathbf{S}_0 \xrightarrow{p} \mathbf{\Omega}$ and $\mathbf{S}_{01} \xrightarrow{p} \mathbf{\Omega}$, and using (B2-B3),

$$\frac{-2tr(\mathbf{S}_{01}^{-1}\frac{1}{T^*}\mathbf{S}_0\mathbf{S}_{01}^{-1})}{\frac{1}{T^*}\sqrt{2T^*\left(T^*-1\right) tr(\mathbf{S}_{01}^{-1}\mathbf{S}_{01}^{-1})}} \xrightarrow{p} -\sqrt{2}\frac{tr(\mathbf{\Omega}^{-1})}{\sqrt{tr(\mathbf{\Omega}^{-1}\mathbf{\Omega}^{-1})}} = -\sqrt{2}\frac{\nu}{\sqrt{\nu^*}} \ .$$

For the third term, note that by the central limit theorem,

$$\frac{\mathbf{\Omega}^{-\frac{1}{2}}\left(\sum_{t=2}^{T}\mathbf{y}_t - \mathbf{0}\right)}{\sqrt{T^*}} \xrightarrow{d} \mathbf{Z} \sim \mathcal{N}_N(\mathbf{0}, \mathbf{I}), \quad \text{as } T \to \infty.$$

Hence

$$\frac{\mathbf{\Omega}^{-\frac{1}{2}}(\sum_{t=2}^{T}\mathbf{y}_t)(\sum_{t=2}^{T}\mathbf{y}_t)'\mathbf{\Omega}^{-\frac{1}{2}}}{T^*} \xrightarrow{d} \mathbf{ZZ}', \quad \text{as } T \to \infty.$$

Also, let $\mathbf{\Omega}$ have spectral decomposition $\mathbf{\Omega} = \mathbf{UVU}'$, implying $\mathbf{\Omega}^{-1} = \mathbf{UV}^{-1}\mathbf{U}'$ where $\mathbf{V} = diag\left[\varphi_1(\mathbf{\Omega}), \varphi_2(\mathbf{\Omega}), ..., \varphi_N(\mathbf{\Omega})\right]$. Then $\mathbf{Z}_2 = \mathbf{U}'\mathbf{Z}_1$ is an orthogonal transformation such that $\mathbf{Z}_2 \sim \mathbf{Z}_1$. Using these results and that $\lim\limits_{T\to\infty} \frac{\sqrt{(T-1)(T-2)}}{T-1} = 1$ and $tr(\mathbf{S}_{01}^{-1}\mathbf{S}_{01}^{-1}) \xrightarrow{p}$

$\frac{1}{\sigma_\varepsilon^4} v^*$, the third term of (13) is

$$\frac{tr(\mathbf{S}_{01}^{-1} \frac{1}{T^*} \mathbf{S}_{00} \mathbf{S}_{01}^{-1})}{\frac{1}{T^*}\sqrt{2T^*\,(T^*-1)\,tr(\mathbf{S}_{01}^{-1}\mathbf{S}_{01}^{-1})}} \xrightarrow{d} \frac{\sigma_\varepsilon^2}{\sqrt{2}\sqrt{\nu^*}} tr(\mathbf{\Omega}^{-1/2}\mathbf{Z}\mathbf{Z}'\mathbf{\Omega}^{-1/2})$$
$$= \frac{\sigma_\varepsilon^2}{\sqrt{2}\sqrt{\nu^*}} \mathbf{Z}'\mathbf{U}\mathbf{V}^{-1}\mathbf{U}'\mathbf{Z} \sim \frac{\sigma_\varepsilon^2}{\sqrt{2}\sqrt{\nu^*}} \mathbf{Z}'\mathbf{V}^{-1}\mathbf{Z} = \frac{\sigma_\varepsilon^2}{\sqrt{2}\sqrt{\nu^*}} \sum_{i=1}^{N} Z_i^2 \frac{1}{\psi_i}$$
$$= \frac{1}{\sqrt{2}\sqrt{\nu^*}} \left(\frac{\sigma_\varepsilon^2}{\eta_1+\sigma_\varepsilon^2} Z_1^2 + ... + \frac{\sigma_\varepsilon^2}{\eta_r+\sigma_\varepsilon^2} Z_r^2 + Z_{r+1}^2 + ... + Z_N^2 \right)$$
$$= \frac{1}{\sqrt{2}\sqrt{\nu^*}} \left[\left(\sum\nolimits_{i=1}^{r} \frac{\sigma_\varepsilon^2}{\eta_i+\sigma_\varepsilon^2} \chi_{1,i}^2 \right) + \chi_{1,r+1}^2 + ... + \chi_{1,N}^2 \right].$$

Put together, we have that

$$\vartheta \xrightarrow{d} \frac{1}{\sqrt{2}} \frac{\nu}{\sqrt{\nu^*}} - \sqrt{2} \frac{\nu}{\sqrt{\nu^*}} + \frac{1}{\sqrt{2}\sqrt{\nu^*}} \Big[\Big(\sum_{i=1}^{r} \frac{\sigma_\varepsilon^2}{\eta_i+\sigma_\varepsilon^2} \chi_{1,i}^2 \Big) + \chi_{1,r+1}^2 + ... + \chi_{1,N}^2 \Big]$$
$$= \frac{1}{\sqrt{2}\sqrt{\nu^*}} \Big[\Big(\sum_{i=1}^{r} \frac{\sigma_\varepsilon^2}{\eta_i+\sigma_\varepsilon^2} \chi_{1,i}^2 + \chi_{1,r+1}^2 + ... + \chi_{1,N}^2 \Big) - \Big(\sum_{i=1}^{r} \frac{\sigma_\varepsilon^2}{\eta_i+\sigma_\varepsilon^2} + N - r \Big) \Big].$$

■

Proof of Proposition 1. Let $\delta_i = \frac{\eta_i}{\sigma_\varepsilon^2}$, and let as before $\nu = \sum\limits_{i=1}^{N} w_i$ and $\nu^* = \sum\limits_{i=1}^{N} w_i^2$. The weighted sum of $\chi_{(1)}^2$ in (14) can be written as

$$\frac{1}{\delta_1+1} Z_1^2 + ... + \frac{1}{\delta_r+1} Z_r^2 + Z_{r+1}^2 + ... + Z_N^2 = \nu \sum_{i=1}^{N} a_i Z_i^2,$$

where $a_i = \frac{1}{(\delta_i+1)\nu}$ for $i = 1, ..., r$ and $a_i = \frac{1}{\nu}$ for $i = r+1, ..., N$, such that $\sum\limits_{i=1}^{N} a_i = 1$. Then by fitting first and second moments (see e.g. Satterthwaite, 1946)

$$\sum_{i=1}^{N} a_i Z_i^2 \overset{app.}{\sim} \frac{\chi_{(u)}^2}{u},$$

where

$$u = \frac{2\Big[E(\sum_{i=1}^{N} a_i Z_i^2)\Big]^2}{Var(\sum_{i=1}^{N} a_i Z_i^2)} = \frac{2(\sum_{i=1}^{N} a_i)^2}{2\sum_{i=1}^{N} a_i^2} = \frac{1}{\frac{1}{\nu^2}\Big(\sum_{i=1}^{r} \frac{1}{(\delta_i+1)^2} + N - r\Big)} = \frac{\nu^2}{\nu^*}\ .$$

So, in analogous derivations as in the proof of Theorem 2, the three terms of (13) are

$$\frac{1}{\sqrt{2}} \frac{\nu}{\sqrt{\nu^*}} = \frac{1}{\sqrt{2}} \sqrt{u}\ , \quad -\sqrt{2} \frac{\nu}{\sqrt{\nu^*}} = -\sqrt{2}\sqrt{u}\ , \quad \frac{1}{\sqrt{2}\sqrt{\nu^*}} \frac{\nu \chi_{(u)}^2}{u} = \frac{1}{\sqrt{2}} \frac{\chi_{(u)}^2}{\sqrt{u}}\ ,$$

such that

$$\vartheta \overset{app.}{\sim} \frac{1}{\sqrt{2u}} \left(\chi^2_{(u)} - u \right).$$

■

Proof of Theorem 3. We check Lyapounov's condition on the fourth moment. Let

$$S_N = \frac{1}{\delta_1 + 1} Z_1^2 + ... + \frac{1}{\delta_r + 1} Z_r^2 + Z_{r+1}^2 + ... + Z_N^2 = \sum_{i=1}^{N} w_i Z_i^2 = \sum_{i=1}^{N} X_i.$$

The fourth moment for X_i clearly exists, since it is proportional to the eighth moment of the standard normal. Then, using the binomial theorem $(a+b)^4 = a^4 + 4a^3b + 6a^2b^2 + 4ab^3 + b^4$, and some moments of the standard normal distribution; $E\left(Z_i^2\right) = 1$, $E\left(Z_i^4\right) = 3$, $E\left(Z_i^6\right) = 15$ and $E\left(Z_i^8\right) = 105$, we have that

$$\begin{aligned} \sum_{i=1}^{N} E\left|X_i - E\left(X_i\right)\right|^4 &= \sum_{i=1}^{N} E\{X_i^4 + [E(X_i)]^4 + 6X_i^2[E(X_i)]^2 \\ &\quad - 4X_i^3 E(X_i) - 4X_i[E(X_i)]^3\} \\ &= 60 \sum_{i=1}^{N} w_i^4 \leq 60N, \end{aligned}$$

since $w_i \in \left\{ \begin{array}{ll} (0,1), & 1 \leq i \leq r \\ 1, & r < i \leq N \end{array} \right.$. Also, let

$$s_n^2 = \sum_{i=1}^{N} Var\left(X_i\right) = \sum_{i=1}^{N} w_i^2 \left\{ E(Z_i^4) - [E(Z_i^2)]^2 \right\} = 2 \sum_{i=1}^{N} w_i^2,$$

where

$$s_n^4 = 4 \Big(\sum_{i=1}^{N} w_i^2 \Big)^2 \geq 4\left(N - r\right)^2.$$

Then the Lyapounov condition is satisfied

$$\frac{\sum_{i=1}^{N} E\left|X_i - E\left(X_i\right)\right|^4}{s_n^4} \leq \frac{15N}{\left(N-r\right)^2} \to 0, \;\; \textit{as } N \to \infty,$$

and the central limit theorem follows

$$\frac{\sum_{i=1}^{N} \left[X_i - E\left(X_i\right)\right]}{s_n} \xrightarrow{d} \mathcal{N}\left(0, 1\right), \text{ as } N \to \infty.$$

■

References

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

Bai, J., Kao, C., Ng, S. (2009), "Panel cointegration with global stochastic trends", *Journal of Econometrics*, 149, 82-99.

Bai, J., Ng, S. (2002), "Determining the number of factors in approximate factor models", *Econometrica*, 70, 191-221.

Bai, J., Ng, S. (2004), "A panic attack on unit roots and cointegration", *Econometrica*, 72, 1127-1177.

Breitung, J., Das, S. (2005), "Panel unit root tests under cross-sectional dependence", *Statistica Neerlandica*, 59, 414-433.

Breitung, J., Das, S. (2008), "Testing for unit roots in panels with a factor structure", *Econometric Theory*, 24, 88-108.

Breitung, J., Eickmeier, S. (2006), "Dynamic factor models", *Allgemeines statistiches Archiv*, 90, 27-42.

Chamberlain, G., Rothschild, M. (1983), "Arbitrage, factor structure, and mean-variance analysis on large asset markets", *Econometrica*, 51, 1281-1304.

Choi, I. (2002), "Combination unit root tests for cross-sectionally correlated panels", *Mimeo*, Hong Kong University of Science and Technology.

Eickmeier, S. (2009), "Comovements and heterogeneity in the euro area analyzed in a non-stationary dynamic factor model", *Journal of Applied Econometrics*, 24, 933-959.

Forni, M., Hallin, M., Lippi, M., Reichlin, L. (2000), The generalized dynamic-factor model: identification and estimation", *The review of Economics and Statistics*, 82, 540-554.

Gengenbach, C., Palm, F.C., Urbain, J-P. (2006), "Cointegration testing in panels with common factors", *Oxford Bulletin of Economics and Statistics*, 68, Supplement, 683-719.

Hartley, H.O., Rao, J.N.K. (1967), "Maximum-likelihood estimation for the mixed analysis of variance model", *Biometrika*, 54, 93-108.

Harville, D.A. (1977), "Maximum likelihood approaches to variance component estimation and to related problems", *Journal of the American Statistical Association*, 72, 320-338.

Im, K.S., Pesaran, M.H., Shin, Y. (2003), "Testing for unit roots in heterogenous panels", *Journal of Econometrics*, 115, 53-74.

van der Leeuw, J. (1994), "The covariance matrix of ARMA errors in closed form", *Journal of Econometrics*, 63, 397-405.

Levin, A., Lin, C.F., Chu, C.S.J. (2002), "Unit root tests in panel data: asymptotic and finite-sample properties", *Journal of Econometrics*, 108, 1-24.

Liu, H., Tang, Y., Zhang, H.H. (2009), "A new chi-square approximation to the distribution of non-negative definite quadratic forms in non-central normal variables", *Computational Statistics and Data Analysis*, 53, 853-856.

Magnus, J.R., Neudecker, H. (2001), *Matrix Differential Calculus with Applications in Statistics and Econometrics*, Wiley, Chichester.

Moon, H.R., Perron, B. (2004), "Testing for a unit root in panels with dynamic factors", *Journal of Econometrics*, 122, 81-126.

O'Connel, P.G.J. (1998), "The overvaluation of purchasing power parity", *Journal of International Economics*, 44, 1-19.

Pesaran, M.H. (2006), "Estimation and inference in large heterogenous panels with a multifactor error structure, *Econometrica*, 74, 967-1012.

Pesaran, M.H. (2007), "A simple unit root test in the presence of cross-section dependence", *Journal of Applied Econometrics*, 22, 265-312.

Phillips, P.C.B. (1987), "Towards a unified asymptotic theory for autoregression", *Biometrika*, 74, 535-547.

Phillips, P.C.B., Sul, D. (2003), "Dynamic panel estimation and homogeneity testing under cross section dependence", *Econometrics Journal*, 6, 217-259.

Satterthwaite, F.E. (1946), "An approximate distribution of estimates of variance components", *Biometrics Bulletin*, 2, 110-114.

Stock, J.H., Watson, M. W. (2002), "Macroeconomic forecasting using diffusion indexes", *Journal of Business and Economic Statistics*, 20, 147-162.

Stoica, P., Jansson, M. (2009), "On maximum likelihood estimation in factor analysis - an algebraic derivation", *Signal Processing*, 89, 1260-1262.

Urbain, J-P., Westerlund, J. (2006), "Spurious regression in nonstationary panels with cross-unit cointegration", *METEOR Research Memoranda No. 057*.