LM-type tests for idiosyncratic and common unit roots in the exact factor model with AR(1) dynamics

Citation for published version (APA):

Document status and date:
Published: 01/01/2013

Document Version:
Publisher's PDF, also known as Version of record

Please check the document version of this publication:

• A submitted manuscript is the version of the article upon submission and before peer-review. There can be important differences between the submitted version and the official published version of record. People interested in the research are advised to contact the author for the final version of the publication, or visit the DOI to the publisher's website.
• The final author version and the galley proof are versions of the publication after peer review.
• The final published version features the final layout of the paper including the volume, issue and page numbers.

Link to publication

Martin Solberger, Xingwu Zhou

LM-type tests for idiosyncratic and common unit roots in the exact factor model with AR(1) dynamics

GSBE

Maastricht University School of Business and Economics
Graduate School of Business and Economics

LM-TYPE TESTS FOR IDIOSYNCRATIC AND COMMON UNIT ROOTS IN THE EXACT FACTOR MODEL WITH AR(1) DYNAMICS

Martin Solberger and Xingwu Zhou

Uppsala University[¶]

Abstract

Recent developments within the panel unit-root literature have illustrated how the exact factor model serves as a parsimonious framework and allows for consistent maximum likelihood inference even when it is misspecified contra the more general approximate factor model. In this paper we consider an exact factor model with AR(1) dynamics and propose LM-type tests for idiosyncratic and common unit roots. We derive the asymptotic distributions and carry out simulations to investigate size and power of the tests in finite samples, as well as compare the performance with some existing tests.

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

1 Introduction

For panel unit root tests there has lately been a great deal of focus on factor models. The reason for this is that they provide at least two important properties; (i) they have an economic rationale, in the sense that economic theory may predict that shocks to a system will affect its individuals (e.g. countries) proportionally and that a relatively small set of shocks may drive a large system (see e.g. Breitung and Eickmeier, 2006, for an overview); and (ii) they implicitly control for a cross-sectional dependence. It is fair to say that the latter has been the main motivation for the so called second generation panel unit root tests, such as Phillips and Sul (2003), Bai and Ng (2004), Moon and Perron (2004) and Pesaran (2007).

Suppose we have the following panel model with N individuals and T time periods:

$$x_{i,t} = \mu_i + \lambda_i' \mathbf{f}_t + u_{i,t}, \qquad i = 1, 2, \ldots, N; t = 1, 2, \ldots, T, \qquad (1)$$

where μ_i is an individual-specific constant, $\mathbf{f}_t = (f_{1,t}, f_{2,t}, \ldots, f_{r,t})'$ is an $r \times 1$ vector of unobservable dynamic factors, $\lambda_i = (\lambda_{i,1}, \lambda_{i,2}, \ldots, \lambda_{i,r})'$ is an $r \times 1$ vector of factor loadings and

[¶]Department of Statistics, Uppsala University, Sweden, Box SE-751-20.
E-mail addresses: martin.solberger@statistics.uu.se; xingwu.zhou@statistics.uu.se

$u_{i,t}$ is a dynamic idiosyncratic component. The exact factor model, which assumes cross-sectionally independent idiosyncratic terms, was given some attention in the early modeling with dynamic factor models in small samples (e.g. Sargent and Sims, 1977; Engle and Watson, 1981; Stock and Watson, 1989). This was later considered to be too strict for most economic time series, and the approximate factor model of Chamberlain and Rothschild (1983), allowing for moderate idiosyncratic cross-sectional dependencies, has become popular (e.g. Stock and Watson, 2002; Bai and Ng, 2004). The standard tool for these models are principal components, while maximum likelihood has generally been regarded as computationally infeasible due to the large number of parameters. Recently, however, Bai and Li (2012a) (see also Doz, Giannone, and Reichlin, 2012) have considered quasi-maximum likelihood estimation in a misspecified exact factor model when the true model is an approximate factor model, and shown that the quasi-MLEs are still consistent. This suggests that the exact factor model is attractive for large panels as a parsimonious representation allowing for consistent likelihood-based inference also under misspecification.

In this paper we consider the exact factor model under AR(1) dynamics, but assume that it is correctly specified, and propose three LM-type tests for unit roots in the idiosyncratic and common components. The null hypothesis is specified as "nonstationary factors and nonstationary idiosyncratic components", which is tested versus "stationary idiosyncratic components", "stationary factors", and "stationary factors and/or stationary idiosyncratic components". We perform a simulation study to investigate the size and power of the proposed statistics, and compare their performance with the PANIC (Panel Analysis of Nonstationarity in Idiosyncratic and Common components) framework of Bai and Ng (2004). The simulation results suggest that the likelihood approach is more powerful than PANIC. Also, when testing for the number of stochastic trends among the factors, PANIC has virtually no power when the factors are locally stationary. Though it will be necessary to extend our model to allow for more general dynamics, the present framework constitutes an important special case and the results should contribute to the development of likelihood based unit root tests in dynamic factor models.

The rest of the paper is organized as follows: Section 2 describes the general framework. Section 3 derives the LM-type statistics and their limiting distributions. Section 4 evaluates the size and power of the tests in finite samples through Monte Carlo simulations and Section 5 concludes. All lengthy mathematical derivations are placed in appendices; Appendix A derives the score and information matrix and proofs are in Appendix B.

Notation: For a matrix $\mathbf{A} \in \mathbb{R}^{n \times n}$, $\varphi_1(\mathbf{A}) \geq \varphi_2(\mathbf{A}) \geq \cdots \geq \varphi_n(\mathbf{A})$ denote the eigenvalues. For a matrix $\mathbf{A} \in \mathbb{R}^{n \times m}$, $||\mathbf{A}|| = [\text{tr}(\mathbf{A}\mathbf{A}')]^{1/2}$ denotes the Frobenius norm and $\mathbf{A}_v = \text{vec}(\mathbf{A})$ is the vectorization operator which stacks the columns in an $nm \times 1$ vector. Unless specified differently, $[a_{i,j}]_{n \times m}$ is an $n \times m$ matrix with element $a_{i,j}$ corresponding to the ith row and jth column, $\text{diag}(a_1, a_2, \ldots, a_n)$ is an $n \times n$ diagonal matrix with entries $a_1, a_2, \ldots, a_n$ and $\mathbf{I}_n$ is the $n \times n$ identity matrix. For limits, $T \rightarrow$ denotes limit taken over T with N fixed and $(T, N)_s \rightarrow$ denotes sequential limit with limit taken over T followed by limit taken over N. $\overset{p}{\rightarrow} (\overset{d}{\rightarrow})$ denotes convergence in probability (distribution).

2 The framework

We consider the panel model (1), which for convenience is restated here. In vector notation we have

$$\mathbf{x}_t = (x_{1,t}, x_{2,t}, \ldots, x_{N,t})' = \boldsymbol{\mu} + \boldsymbol{\Lambda}\mathbf{f}_t + \mathbf{u}_t, \tag{2}$$

where $\boldsymbol{\mu} = (\mu_1, \mu_2, \ldots, \mu_N)'$ are the constants, $\mathbf{f}_t = (f_{1,t}, f_{2,t}, \ldots, f_{r,t})'$ are the unobservable dynamic factors, $\mathbf{u}_t = (u_{1,t}, u_{2,t}, \ldots, u_{N,t})'$ are the dynamic idiosyncratic components and $\boldsymbol{\Lambda} = [\lambda_{i,j}]_{N \times r}$ is the matrix of factor loadings. The contemporaneous covariances are then $\boldsymbol{\Sigma}_x = \boldsymbol{\Lambda}\boldsymbol{\Sigma}_f\boldsymbol{\Lambda}' + \boldsymbol{\Sigma}_u$, where $\boldsymbol{\Sigma}_f = Var(\mathbf{f}_t)$ and $\boldsymbol{\Sigma}_u = Var(\mathbf{u}_t)$. If the largest eigenvalue of $\boldsymbol{\Sigma}_u$ is bounded and the eigenvalues of $\boldsymbol{\Lambda}\boldsymbol{\Sigma}_f\boldsymbol{\Lambda}'$ are not bounded, then we have the approximate factor model by Chamberlain and Rothschild (1983). Further, if $\boldsymbol{\Sigma}_u$ is restricted to be diagonal, then we have the exact factor model, which is the direct dynamic generalization of the classical factor model. Note then that the exact factor model is nested within the approximate factor model. We impose the following restrictions, admitting the exact factor model:

Assumption 1 The idiosyncratic components are AR(1) processes, $u_{i,t} = \rho_i u_{i,t-1} + \varepsilon_{i,t}$, where $\rho_i \in (-1, 1]$, $\varepsilon_{i,t} \sim \mathcal{N}(0, \sigma_{\varepsilon,i}^2)$ are independent and identically distributed (iid) over time with $\sigma_{\varepsilon,i}^2 < \infty$, and $E(\varepsilon_{i,t}\varepsilon_{l,s}) = 0$ for all $i \neq l$ $(i, l = 1, 2, \ldots, N)$ and all $t, s = 1, 2, \ldots, T$.
Assumption 2 The factors are AR(1) processes, $f_{j,t} = \alpha_j f_{j,t-1} + v_{j,t}$, where $\alpha_j \in (-1, 1]$, $v_{j,t}$ are iid $\mathcal{N}(0, 1)$, and $E(v_{j,t}v_{q,s}) = 0$ for all $j \neq q$ $(j, q = 1, 2, \ldots, r)$ and all $t, s = 1, 2, \ldots, T$.
Assumption 3 The factor loadings are non-random with $||\lambda_i|| < \infty$, and $\frac{1}{N}\sum_{i=1}^{N} \lambda_i \lambda_i'$ converges to some positive definite matrix $\boldsymbol{\Sigma}_\Lambda$.
Assumption 4 The processes $v_{j,t}$ and $\varepsilon_{i,t}$ are independently distributed.
Assumption 5 For stationary processes the starting values $f_{j,0}$ and $u_{i,0}$ come from the stationary distributions, and for nonstationary processes they are $\mathcal{O}_p(1)$.

Assumptions 1 and 2 imply that the idiosyncratic components are cross-sectionally independent, so that we have the exact factor model, and likewise that the factors are mutually independent. The assumption of normally distributed errors will allow us to obtain the likelihood of the factor model, where the assumption of unit variance in the factor errors can be made without loss of generality, because, unless we impose further restrictions, the factors and the factor loadings are not separately identified. Assumption 3 implies that the eigenvalues of $\boldsymbol{\Lambda}\boldsymbol{\Lambda}'$ are $\mathcal{O}(N)$, which together with Assumption 4 allows us to identify the factor structure. These assumptions are standard in factor models (see e.g. Bai and Ng, 2002; Bai, 2003; Bai and Ng, 2004, and references therein). Assumption 5 ensures us that the stochastic processes are well-behaved with respect to their initial values. Together, we may summarize these assumptions as the exact factor model with Gaussian AR(1) dynamics.

Because the panel data is allowed to be nonstationary, we will proceed by taking first differences. For notational convenience, let $T^* = T - 1$, and let also $\mathbf{D} = [D_{\tau,t}]_{T^* \times T}$ be a first-difference matrix, i.e. $D_{\tau,t} = -1$ if $\tau = t$, $D_{\tau,t} = 1$ if $\tau = t - 1$, and zero otherwise. Consider the ith idiosyncratic component and let $\mathbf{u}_i = (u_{i,1}, u_{i,2}, \ldots, u_{i,T})'$. If $\mathbf{u}_i$ is stationary,

then $\mathbf{u}_i \sim \mathcal{N}_T(\mathbf{0}, \sigma_{\varepsilon,i}^2 \mathbf{\Pi}(\rho_i))$, where $\mathbf{\Pi}(\rho_i) = [\Pi_{k,m}^i]_{T \times T}$ is an autocovariance matrix, with

$$\Pi_{k,m}^i = \begin{cases} \frac{1}{1-\rho_i^2} & \text{for } k = m \\ \frac{\rho_i^{|k-s|}}{1-\rho_i^2} & \text{for } k \neq m \end{cases},$$

which can be found from e.g. van der Leeuw (1994). The autocovariance matrix has the useful property that the nonstationary point exists after taking first-differences. It is straightforward to show that $\mathbf{Du}_i \sim \mathcal{N}_{T^*}(\mathbf{0}, \sigma_{\varepsilon,i}^2 \mathbf{\Psi}(\rho_i))$, where $\mathbf{\Psi}(\rho_i) = [\Psi_{k,m}^i]_{T^* \times T^*}$, with

$$\Psi_{k,m}^i = \begin{cases} \frac{2}{1+\rho_i} & \text{for } k = m \\ -\frac{\rho_i^{||k-s|-1|}(1-\rho_i)}{1+\rho_i} & \text{for } k \neq m \end{cases}. \tag{3}$$

Here $\mathbf{\Psi}$ is twice continuously differentiable around the nonstationary point $\mathbf{\Psi}(1) = \mathbf{I}_{T^*}$. Also, the nonstationary point is correctly defined, because if there is a unit root in $\mathbf{u}_i$, then $\mathbf{Du}_i \sim \mathcal{N}_{T^*}(\mathbf{0}, \sigma_{\varepsilon,i}^2 \mathbf{I})$. The matrix $\mathbf{\Psi}$ is therefore well-defined for constructing likelihood-based statistics for a unit root hypothesis. Similarly, for the jth factor, let $\mathbf{f}_j = (f_{j,1}, f_{j,2}, \ldots, f_{j,T})'$. If there is a unit root in $f_{j,t}$, then, from Assumption 2, $\mathbf{Df}_j \sim \mathcal{N}_{T^*}(\mathbf{0}, \mathbf{I})$, while for the stationary case $\mathbf{Df}_j \sim \mathcal{N}_{T^*}(\mathbf{0}, \mathbf{\Psi}(\alpha_j))$, where $\mathbf{\Psi}(\alpha_j)$ is a matrix with elements analogous to (3).

Assumption 6 The idiosyncratic components and the factors are separate homogenous groups. That is, $\rho_i = \rho$ for all i, and $\alpha_j = \alpha$ for all j.

Assumption 7 The idiosyncratic components are cross-sectionally homoscedastic. That is, for all i, $\sigma_{\varepsilon,i}^2 = \sigma_\varepsilon^2$.

Assumptions 6 and 7, which may be relaxed, simplify the analysis, and will allow us to use explicit maximum likelihood estimators. Assumption 6 is not particularly restrictive, at least not from a statistical point of view, because we expect homogenous tests to have power also against a heterogenous alternative. Assumption 7 is restrictive and was documented by Zhou and Solberger (2012) to give size distortions if violated. Relaxing the latter assumption will prompt us to use numerical maximum likelihood, say using the EM-algorithm. Some consistency results were recently derived by Bai and Li (2012b), but this is not pursued here. Instead we maintain these assumptions through out the paper to be able to rely on explicit MLEs. The results are nevertheless important, because relaxing Assumption 7 is likely to result in test-statistics with similar asymptotic distributions. The null hypothesis of interest here is

$$\mathcal{H}_0 : \rho = 1, \alpha = 1,$$

which, under Assumptions 1-7, may be equivalently stated as

$$\mathcal{H}_0 : \mathbf{f}_t \sim I(1), \mathbf{u}_t \sim I(1). \tag{4}$$

Let $\mathbf{Y} = \mathbf{XD}'$, where $\mathbf{X} = (\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_T)$ is the $N \times T$ matrix of observed panel data and let $\mathbf{M} = \mathbf{\Lambda \Lambda}'$. The differenced and stacked panel data, $\mathbf{Y}_v = \text{vec}(\mathbf{Y}) = (\mathbf{y}_2', \mathbf{y}_3', \ldots, \mathbf{y}_T')'$, has covariance matrix $\mathbf{\Sigma} = E(\mathbf{Y}_v \mathbf{Y}_v') = [\mathbf{\Psi}(\alpha) \otimes \mathbf{M}] + \sigma_\varepsilon^2 [\mathbf{\Psi}(\rho) \otimes \mathbf{I}_N]$ where $\otimes$ denotes the

4

Kronecker product. Under the null hypothesis $\rho = \alpha = 1$ we have that

$$\boldsymbol{\Sigma} = (\mathbf{I}_{T^*} \otimes \mathbf{M}) + \sigma_\varepsilon^2 \mathbf{I}_{NT^*} = \left[\mathbf{I}_{T^*} \otimes \left(\mathbf{M} + \sigma_\varepsilon^2 \mathbf{I}_N\right)\right] = (\mathbf{I}_{T^*} \otimes \boldsymbol{\Omega}), \tag{5}$$

where $\boldsymbol{\Omega} = \left(\mathbf{M} + \sigma_\varepsilon^2 \mathbf{I}_N\right)$, implying for $\mathbf{y}_t = \Delta\mathbf{x}_t = (\Delta x_{1t}, \Delta x_{2t}, \dots, \Delta x_{Nt})'$ that

$$E\left(\mathbf{y}_t \mathbf{y}_s'\right) = \begin{cases} \boldsymbol{\Omega} & \text{if } s = t \\ \mathbf{0} & \text{if } s \neq t \end{cases}. \tag{6}$$

In this case we have an explicit MLE of $\boldsymbol{\Omega}$. Use a reduced singular value decomposition of $\boldsymbol{\Lambda}$ and factorize as $\boldsymbol{\Lambda}\boldsymbol{\Lambda}' = \mathbf{A}\mathbf{H}\mathbf{A}'$, where $\mathbf{H} = \text{diag}\left[\varphi_1(\boldsymbol{\Lambda}'\boldsymbol{\Lambda}), \varphi_2(\boldsymbol{\Lambda}'\boldsymbol{\Lambda}), \dots, \varphi_r(\boldsymbol{\Lambda}'\boldsymbol{\Lambda})\right]$ has the non-zero eigenvalues of $\boldsymbol{\Lambda}\boldsymbol{\Lambda}'$ and $\mathbf{A}$ is semi-orthogonal such that $\mathbf{A}'\mathbf{A} = \mathbf{I}_r$.[1] 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 $\mathbf{S} = \frac{1}{T^*}\mathbf{Y}\mathbf{Y}'$ associated with the r largest eigenvalues $\widehat{\varphi}_1(\mathbf{S}) \geq \widehat{\varphi}_2(\mathbf{S}) \geq \cdots \geq \widehat{\varphi}_r(\mathbf{S})$, and that the MLEs of $\mathbf{H}$ and σ_ε^2 are $\widehat{\sigma}_\varepsilon^2 = \frac{1}{(N-r)}\sum_{i=r+1}^N \widehat{\varphi}_i(\mathbf{S})$ and $\widehat{\mathbf{H}} = \widehat{\boldsymbol{\Phi}} - \widehat{\sigma}_\varepsilon^2 \mathbf{I}_r$, where $\widehat{\boldsymbol{\Phi}} = \text{diag}\left[\widehat{\varphi}_1(\mathbf{S}), \widehat{\varphi}_2(\mathbf{S}), \dots, \widehat{\varphi}_r(\mathbf{S})\right]$. From this we define the MLE of $\boldsymbol{\Omega}$,

$$\mathbf{S}_{01} = \widehat{\boldsymbol{\Omega}} = \widehat{\mathbf{M}} + \widehat{\sigma}_\varepsilon^2 \mathbf{I}_N, \tag{7}$$

where $\widehat{\mathbf{M}} = \widehat{\mathbf{A}}\widehat{\mathbf{H}}\widehat{\mathbf{A}}'$.

The sample covariances, $\mathbf{Y}_v \mathbf{Y}_v'$, consist of T^{*2} blocks, $\mathbf{S}_{t,s} = \mathbf{y}_t \mathbf{y}_s'$ for $t, s = 2, 3, \dots, T$, each block of size $N \times N$. For these blocks we define, for future reference,

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

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

3 The LM-type statistics

Let $\boldsymbol{\theta} = (\boldsymbol{\Lambda}_v', \sigma_\varepsilon^2, \alpha, \rho)'$ be the parameter vector holding $K = Nr + 3$ parameters. 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 three LM-type tests with three different alternative hypotheses. Let $\widetilde{\boldsymbol{\theta}}$ be the restricted maximum likelihood estimator under $\mathcal{H}_0$. The LM-statistic is defined as

$$LM = \left.\mathbf{V}(\boldsymbol{\theta})'\mathbf{J}(\boldsymbol{\theta})^{-1}\mathbf{V}(\boldsymbol{\theta})\right|_{\boldsymbol{\theta}=\widetilde{\boldsymbol{\theta}}}, \tag{11}$$

where $\mathbf{V}(\boldsymbol{\theta}) \equiv \partial l(\boldsymbol{\theta})/\partial\boldsymbol{\theta}$ is the score vector and $\mathbf{J}(\boldsymbol{\theta}) \equiv [J_{\theta_k\theta_q}]_{K\times K}$ is the information matrix, where $J_{\theta_k\theta_q} = -E\left(\partial^2 l(\boldsymbol{\theta})/\partial\theta_k\partial\theta_q\right)$ for $k, q = 1, 2, \dots, K$. Also, define for any subsets $\boldsymbol{\omega}, \boldsymbol{v} \subseteq \boldsymbol{\theta}$ with K_ω and K_v parameters respectively, $\mathbf{V}(\boldsymbol{\omega}) \equiv \partial l(\boldsymbol{\theta})/\partial\boldsymbol{\omega}$ and $\mathbf{J}_{\boldsymbol{\omega}\boldsymbol{v}'} \equiv [J_{\omega_k v_q}]_{K_\omega \times K_v}$ for

---

[1]To clarify: Let $\boldsymbol{\Lambda}$ $(N \times r)$ have SVD $\mathbf{U}\boldsymbol{\Sigma}\mathbf{V}'$ where $\mathbf{U}$ is $N \times N$, $\boldsymbol{\Sigma}$ is $N \times r$ and $\mathbf{V}$ is $r \times r$. Because $N > r$, $\boldsymbol{\Sigma}$ has r singular values and $N - r$ zero-rows, i.e. $\boldsymbol{\Sigma} = [\boldsymbol{\Sigma}_+, \mathbf{0}]'$. Let correspondingly $\mathbf{U} = [\mathbf{U}_+, \mathbf{U}_0]$, where $\mathbf{U}_+'\mathbf{U}_+ = \mathbf{I}_r$, but $\mathbf{U}_+\mathbf{U}_+' \neq \mathbf{I}_N$. Then $\mathbf{U}_+\boldsymbol{\Sigma}_+\mathbf{V}'$ is the reduced SVD such that $\boldsymbol{\Lambda}\boldsymbol{\Lambda}' = \mathbf{U}_+\boldsymbol{\Sigma}_+^2\mathbf{U}_+'$, where $\boldsymbol{\Sigma}_+^2$ has the eigenvalues of $\boldsymbol{\Lambda}'\boldsymbol{\Lambda} = \mathbf{V}\boldsymbol{\Sigma}_+^2\mathbf{V}'$, corresponding to the non-zero eigenvalues of $\boldsymbol{\Lambda}\boldsymbol{\Lambda}'$.

integers $k = 1, 2, \ldots, K_\omega$ and $q = 1, 2, \ldots, K_\nu$ (see Appendix A for specific derivations of the score and the information). Now, let the parameter vector be partitioned as $\boldsymbol{\theta} = (\boldsymbol{\theta}_1', \boldsymbol{\theta}_2')'$ where $\boldsymbol{\theta}_2 = (\alpha, \rho)'$ is the vector subject to restrictions. Because under restricted maximum likelihood estimation we have that $\mathbf{V}(\boldsymbol{\theta}) = \left[\mathbf{0}, (\partial l\, (\boldsymbol{\theta})\, / \partial \boldsymbol{\theta}_2)' \right]'$, the information matrix may be partitioned into the relevant blocks,

$$\mathbf{J} = \left(\begin{array}{cc} \mathbf{J}_{11} & \mathbf{J}_{12} \\ \mathbf{J}_{21} & \mathbf{J}_{22} \end{array} \right),$$

where $\mathbf{J}_{22} = -E \left(\frac{\partial^2 l(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}_2 \partial \boldsymbol{\theta}_2'} \right)$. If we define the inverse as

$$\mathbf{J}^{-1} = \left(\begin{array}{cc} \mathbf{J}^{11} & \mathbf{J}^{12} \\ \mathbf{J}^{21} & \mathbf{J}^{22} \end{array} \right),$$

then the LM-statistic (11) becomes

$$LM = \left(\frac{\partial l\, (\boldsymbol{\theta})}{\partial \boldsymbol{\theta}_2} \right)' \mathbf{J}^{22} \left(\frac{\partial l\, (\boldsymbol{\theta})}{\partial \boldsymbol{\theta}_2} \right) \Bigg|_{\boldsymbol{\theta} = \widetilde{\boldsymbol{\theta}}}. \tag{12}$$

If the information matrix is block-diagonal in the specific way that $\mathbf{J}_{21} = \mathbf{J}_{12}' = \mathbf{0}$, then we simply have that $\mathbf{J}^{22} = \mathbf{J}_{22}^{-1}$. When the information matrix is not block-diagonal in a suitable way, we may still rely on some asymptotic results (see Solo, 1984). For notational simplicity, let $\widetilde{\mathbf{J}} = \mathbf{J}|_{\boldsymbol{\theta} = \widetilde{\boldsymbol{\theta}}}$. Suppose there exist a sequence of matrices of constants $\mathbf{C}_{i,NT} \to \infty$ for $i = 1, 2$ such that the following block-diagonal condition holds:

$$\mathbf{C}_{1,NT}^{-1} \widetilde{\mathbf{J}}_{11} \mathbf{C}_{1,NT}^{-'} \xrightarrow{p} \boldsymbol{\pi}_{11}; \quad \mathbf{C}_{1,NT}^{-1} \widetilde{\mathbf{J}}_{12} \mathbf{C}_{2,NT}^{-'} \xrightarrow{p} \mathbf{0};$$
$$\mathbf{C}_{2,NT}^{-1} \widetilde{\mathbf{J}}_{21} \mathbf{C}_{1,NT}^{-'} \xrightarrow{p} \mathbf{0}; \quad \mathbf{C}_{2,NT}^{-1} \widetilde{\mathbf{J}}_{22} \mathbf{C}_{2,NT}^{-'} \xrightarrow{p} \boldsymbol{\pi}_{22},$$

where $\mathbf{C}^{-'} = (\mathbf{C}^{-1})'$, and $\boldsymbol{\pi}_{11}$ and $\boldsymbol{\pi}_{22}$ are full rank matrices. Then we have that

$$\widetilde{\mathbf{J}}^{22} = (\widetilde{\mathbf{J}}_{22} - \widetilde{\mathbf{J}}_{21} \widetilde{\mathbf{J}}_{11}^{-1} \widetilde{\mathbf{J}}_{12})^{-1} = \mathbf{C}_{2,NT}^{-'} \Big[\mathbf{C}_{2,NT}^{-1} \widetilde{\mathbf{J}}_{22} \mathbf{C}_{2,NT}^{-'} -$$

$$\left(\mathbf{C}_{2,NT}^{-1} \widetilde{\mathbf{J}}_{21} \mathbf{C}_{1,NT}^{-'} \right) \left(\mathbf{C}_{1,NT}^{-1} \widetilde{\mathbf{J}}_{11} \mathbf{C}_{1,NT}^{-'} \right)^{-1} \left(\mathbf{C}_{1,NT}^{-1} \widetilde{\mathbf{J}}_{12} \mathbf{C}_{2,NT}^{-'} \right) \Big]^{-1} \mathbf{C}_{2,NT}^{-1}$$

$$\xrightarrow{p} \widetilde{\mathbf{J}}_{22}^{-1}.$$

Thus, letting $\widetilde{\mathbf{V}} = \mathbf{V}|_{\boldsymbol{\theta} = \widetilde{\boldsymbol{\theta}}}$, Equation (12) can be written as

$$LM = \widetilde{\mathbf{V}}(\boldsymbol{\theta}_2)' \widetilde{\mathbf{J}}^{22} \widetilde{\mathbf{V}}(\boldsymbol{\theta}_2) \simeq LM^* = \widetilde{\mathbf{V}}(\boldsymbol{\theta}_2)' \widetilde{\mathbf{J}}_{22}^{-1} \widetilde{\mathbf{V}}(\boldsymbol{\theta}_2), \tag{13}$$

in probability as $\mathbf{C}_{i,NT} \to \infty$, in the sense that $|LM - LM^*| = o_p(1)$. In the following sections we will make use of this property.

For the null hypothesis (4) we postulate three alternative hypotheses:

$$\mathcal{H}_{1a} : \mathbf{f}_t \sim I(1), \mathbf{u}_t \sim I(0),$$
$$\mathcal{H}_{1b} : \mathbf{f}_t \sim I(0), \mathbf{u}_t \sim I(1),$$
$$\mathcal{H}_{1c} : \neg(\mathbf{f}_t \sim I(1), \mathbf{u}_t \sim I(1)),$$

where ¬ denotes the logical complement "not". The first two hypotheses are also considered by Breitung and Das (2008), but as null hypotheses. They conclude that treating $\mathcal{H}_{1a}$ as a null hypothesis is especially difficult, because OLS and GLS based statistics as considered by Phillips and Sul (2003), Moon and Perron (2004) and Breitung and Das (2005) are no longer valid. Here the tests are reversed, in the sense that we consider these cases as alternative hypotheses. Our simulation results in Section 4 indicate that the size and power of the test-statistic for $\mathcal{H}_0$ vs $\mathcal{H}_{1a}$ is asymptotically independent of the conditioning on $I(1)$-factors, and, similarly, that the size and power of the test-statistic for $\mathcal{H}_0$ vs $\mathcal{H}_{1b}$ is asymptotically independent of the conditioning of $I(1)$-idiosyncratic components. Although we do not provide a formal proof, we conjecture this to be true. For the third alternative hypothesis, rejection should be regarded as "at least one of $\mathbf{f}_t$ and $\mathbf{u}_t$ are stationary". To our knowledge, no one has proposed a similar test before.

Before we go in to the specific cases, we introduce the following important lemma.

Lemma 1 *Under Assumptions 1-7, for $i, l = 1, 2, \ldots, N$ and $j, q = 1, 2, \ldots, r$, the following terms are $\mathcal{O}_p(1)$:*

$$
\begin{array}{lllll}
(i) \ \dfrac{\widetilde{J}_{\lambda_{i,j}\lambda_{l,q}}}{N^2 T^*}; & (ii) \ \dfrac{\widetilde{J}_{\lambda_{i,j}\sigma_\varepsilon^2}}{NT^*}; & (iii) \ \dfrac{\widetilde{J}_{\lambda_{i,j}\alpha}}{NT^*}; & (iv) \ \dfrac{\widetilde{J}_{\lambda_{i,j}\rho}}{NT^*}; & (v) \ \dfrac{\widetilde{J}_{\sigma_\varepsilon^2\sigma_\varepsilon^2}}{NT^*}; \\[3mm]
(vi) \ \dfrac{\widetilde{J}_{\alpha\sigma_\varepsilon^2}}{T^*}; & (vii) \ \dfrac{\widetilde{J}_{\rho\sigma_\varepsilon^2}}{NT^*}; & (viii) \ \dfrac{\widetilde{J}_{\alpha\alpha}}{T^{*2}}; & (ix) \ \dfrac{\widetilde{J}_{\alpha\rho}}{T^{*2}}; & (x) \ \dfrac{\widetilde{J}_{\rho\rho}}{NT^{*2}}.
\end{array}
$$

3.1 Case A, alternative hypothesis with $\mathbf{u}_t \sim I(0)$

For case A we condition on integrated factors by keeping $\alpha = 1$ fixed. Thus, the parameter vector may be effectively partitioned as $\theta_{1,a} = (\mathbf{\Lambda}_v', \sigma_\varepsilon^2)'$ and $\theta_{2,a} = \rho$. We then consider the "square-root" LM-statistic

$$
\vartheta_a = \left. \frac{\partial l}{\partial \rho} \sqrt{J^{22}} \right|_{\theta = \widetilde{\theta}}, \tag{14}
$$

to test the hypotheses

$$
\mathcal{H}_0 : \rho = 1 | \alpha = 1,
$$
$$
\mathcal{H}_1 : \rho < 1 | \alpha = 1.
$$

Zhou and Solberger (2012) show that for any fixed $N > r$ we have the relation $J^{22} = J_{22}^{-1} \left(1 - \frac{1}{T^*}\right)$, and that the LM-type statistic (14) is

$$
\vartheta_a = \frac{T^* \operatorname{tr} \left(\mathbf{S}_{01}^{-1} \right) - 2 \operatorname{tr} \left(\mathbf{S}_{01}^{-1} \mathbf{S}_0 \mathbf{S}_{01}^{-1} \right) + \operatorname{tr} \left(\mathbf{S}_{01}^{-1} \mathbf{S}_{00} \mathbf{S}_{01}^{-1} \right)}{\sqrt{2T^* (T^* - 1) \operatorname{tr} \left(\mathbf{S}_{01}^{-1} \mathbf{S}_{01}^{-1} \right)}}, \tag{15}
$$

where $\mathbf{S}_{01}$, $\mathbf{S}_0$ and $\mathbf{S}_{00}$ are given by (7), (8) and (9) respectively. Let $\eta_i = \phi_i(\mathbf{\Lambda}'\mathbf{\Lambda})$ denote the ordered eigenvalues of $\mathbf{\Lambda}'\mathbf{\Lambda}$, $\eta_1 \geq \eta_2 \geq \cdots \geq \eta_r > 0$, corresponding to the r non-zero eigenvalues of $\mathbf{\Lambda}\mathbf{\Lambda}'$. For any fixed N, the statistic (15) converges as $T \to \infty$ to a weighted

χ^2-distribution,

$$\vartheta_a \xrightarrow{d} \frac{1}{\sqrt{2\sum_{i=1}^{N} w_i^2}} \left(\sum_{i=1}^{N} w_i \chi_{1,i}^2 - \sum_{i=1}^{N} w_i \right), \tag{16}$$

where $\chi_{1,i}^2$ (Chi-square with one degree of freedom) are independent over $i = 1, 2, \ldots, N$ with weights $w_i = \sigma_\varepsilon^2 / (\eta_i + \sigma_\varepsilon^2)$ for $1 \le i \le r$ and $w_i = 1$ for $r < i \le N$.[2] This distribution is non-standard, but there exist many relatively precise approximations (see e.g. Liu, Tang, and Zhang, 2009). Especially, it can be suitably approximated by (see Proposition 1 in Section 3.2 or Zhou and Solberger, 2012)

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

with $u = \left[\text{tr}(\Omega^{-1}) \right]^2 / \text{tr}(\Omega^{-1}\Omega^{-1})$, where Ω is the contemporaneous covariance matrix in (6). In a simulation study, this approximation is demonstrated by Zhou and Solberger (2012) to fit very well as long as T is large enough, say $T > 25$, but also for smaller T as long as N is equally small. For large both N and T, the approximation is not necessary as the limiting distribution is standard normal,

$$\vartheta_a \xrightarrow{d} \mathcal{N}(0,1), \text{ as } (T, N)_s \to \infty.$$

We may also consider another approach where we impose the property of a block-diagonal information matrix. Here the information matrix has the decomposition

$$\mathbf{J}_{11,a} = \begin{pmatrix} \mathbf{J}_{\Lambda_v \Lambda_v'} & \mathbf{J}_{\Lambda_v \sigma_\varepsilon^2} \\ \mathbf{J}_{\sigma_\varepsilon^2 \Lambda_v'} & \mathbf{J}_{\sigma_\varepsilon^2 \sigma_\varepsilon^2} \end{pmatrix}; \quad \mathbf{J}_{21,a} = \mathbf{J}_{12,a}' = \begin{pmatrix} \mathbf{J}_{\rho \Lambda_v'} & \mathbf{J}_{\rho \sigma_\varepsilon^2} \end{pmatrix}; \quad \mathbf{J}_{22,a} = J_{\rho\rho}. \tag{17}$$

So if we let $\mathbf{C}_{1,NT^*,a} = \text{diag}(N\sqrt{T^*}\mathbf{I}_{Nr}, \sqrt{NT^*})$ and $\mathbf{C}_{2,NT^*,a} = T^*\sqrt{N}$, then by applying Lemma 1 it is easily established that the block-diagonal condition holds for any fixed N as $T \to \infty$. Let $\mathbf{C} = \text{diag}(\mathbf{C}_{1,NT^*,a}, \mathbf{C}_{2,NT^*,a})$. Then

$$\mathbf{C}^{-1}\widetilde{\mathbf{J}}(\theta)\mathbf{C}^{-\prime} = \begin{pmatrix} \mathbf{O}_p(1) & \mathbf{O}_p\left(\frac{1}{\sqrt{N}}\right) & \mathbf{O}_p\left(\frac{1}{\sqrt{NT^*}}\right) \\ \mathbf{O}_p\left(\frac{1}{\sqrt{N}}\right) & \mathcal{O}_p(1) & \mathcal{O}_p\left(\frac{1}{\sqrt{T^*}}\right) \\ \mathbf{O}_p\left(\frac{1}{\sqrt{NT^*}}\right) & \mathcal{O}_p\left(\frac{1}{\sqrt{T^*}}\right) & \mathcal{O}_p(1) \end{pmatrix},$$

where $\mathbf{O}$ denotes a matrix and $\mathcal{O}$ denotes a scalar. That is; corresponding to the lower right scalar, $\mathbf{C}_{2,NT,a}^{-1}\widetilde{\mathbf{J}}_{22,a}\mathbf{C}_{2,NT,a}^{-\prime} = \mathcal{O}_p(1)$; corresponding to the upper left $(Nr + 1) \times (Nr + 1)$ block, $\mathbf{C}_{1,NT,a}^{-1}\widetilde{\mathbf{J}}_{11,a}\mathbf{C}_{1,NT,a}^{-\prime} = \left\{ \left[\mathbf{O}_p(1), \mathbf{O}_p\left(\frac{1}{\sqrt{N}}\right)\right]', \left[\mathbf{O}_p\left(\frac{1}{\sqrt{N}}\right), \mathcal{O}_p(1)\right]' \right\}$; and corresponding to the lower left $1 \times (Nr + 1)$ block (with the transpose corresponding to the upper right $(Nr + 1) \times 1$ block), $\mathbf{C}_{2,NT,a}^{-1}\widetilde{\mathbf{J}}_{21,a}\mathbf{C}_{1,NT,a}^{-\prime} = \left[\mathbf{O}_p\left(\frac{1}{\sqrt{NT^*}}\right), \mathcal{O}_p\left(\frac{1}{\sqrt{T^*}}\right)\right]'$. Hence, $\widetilde{\mathbf{J}}(\theta)$ is

---

[2]We assume here, in convention with most of the literature, that the number of factors is known. Typically we can use the criteria in Bai and Ng (2002) to consistently estimate the number of factors.

block-diagonal in probability for any fixed N as T tends to infinity, because as $T \to \infty$, $\mathbf{C}_{2,NT,a}^{-1} \widetilde{\mathbf{J}}_{22,a} \mathbf{C}_{2,NT,a}^{-\prime} \xrightarrow{p} \pi_{22,a}$ for some scalar $\pi_{22,a}$, $\mathbf{C}_{1,NT,a}^{-1} \widetilde{\mathbf{J}}_{11,a} \mathbf{C}_{1,NT,a}^{-\prime} \xrightarrow{p} \pi_{11,a}$ for some matrix $\pi_{11,a}$, $\mathbf{C}_{1,NT,a}^{-1} \widetilde{\mathbf{J}}_{12,a} \mathbf{C}_{2,NT,a}^{-\prime} \xrightarrow{p} \mathbf{0}$, and $\mathbf{C}_{2,NT,a}^{-1} \widetilde{\mathbf{J}}_{21,a} \mathbf{C}_{1,NT,a}^{-\prime} \xrightarrow{p} \mathbf{0}$. Here we do not need find the exact form of $\pi_{11,a}$, because asymptotically this part will not enter the test-statistic. Without loss of generality we may assume that $\pi_{11,a}$ is some full rank matrix.[3] Also, from the proof of Lemma 1 (x) (see Appendix B), it follows that $\pi_{22,a} = \frac{1}{8N} \sum_{i=1}^{N} w_i^2 \to \frac{1}{8}$ as $(T,N)_s \to \infty$, where w_i are the weights in (16).

Imposing a block-diagonal information matrix, and using the results from Appendix A, it is straightforward to construct the block-diagonal analogue of (14) as

$$\vartheta_a \simeq \vartheta_a^* = \widetilde{\mathbf{V}}(\rho) \sqrt{\widetilde{J}_{\rho\rho}^{-1}} = \frac{T^* \operatorname{tr}\left(\mathbf{S}_{01}^{-1}\right) - 2 \operatorname{tr}\left(\mathbf{S}_{01}^{-1} \mathbf{S}_0 \mathbf{S}_{01}^{-1}\right) + \operatorname{tr}\left(\mathbf{S}_{01}^{-1} \mathbf{S}_{00} \mathbf{S}_{01}^{-1}\right)}{\sqrt{2T^{*2} \operatorname{tr}\left(\mathbf{S}_{01}^{-1} \mathbf{S}_{01}^{-1}\right)}},$$

where $\widetilde{\mathbf{V}}(\rho)$ and $\widetilde{J}_{\rho\rho}$ denote $\mathbf{V}(\rho)|_{\theta=\widetilde{\theta}}$ and $J_{\rho\rho}|_{\theta=\widetilde{\theta}}$ respectively. Clearly, ϑ_a^* and ϑ_a in (15) have the same asymptotic distribution as $T \to \infty$ and as $(T,N)_s \to \infty$.

3.2 Case B, alternative hypothesis with $\mathbf{f}_t \sim I(0)$

For Case B we condition on integrated idiosyncratic components by keeping $\rho = 1$ fixed, and consider the hypotheses

$$\mathcal{H}_0 : \alpha = 1 | \rho = 1,$$
$$\mathcal{H}_1 : \alpha < 1 | \rho = 1.$$

We can then partition the parameter vector such that $\theta_{1,b} = (\Lambda_v', \sigma_\varepsilon^2)'$ and $\theta_{2,b} = \alpha$, and where the information matrix has the decomposition

$$\mathbf{J}_{11,b} = \begin{pmatrix} \mathbf{J}_{\Lambda_v \Lambda_v'} & \mathbf{J}_{\Lambda_v \sigma_\varepsilon^2} \\ \mathbf{J}_{\sigma_\varepsilon^2 \Lambda_v'} & J_{\sigma_\varepsilon^2 \sigma_\varepsilon^2} \end{pmatrix}; \quad \mathbf{J}_{21,b} = \mathbf{J}_{12,b}' = \begin{pmatrix} \mathbf{J}_{\alpha \Lambda_v'} & J_{\alpha \sigma_\varepsilon^2} \end{pmatrix}; \quad \mathbf{J}_{22,b} = J_{\alpha\alpha}. \quad (18)$$

Again, we consider the "square-root" LM-statistic,

$$\vartheta_b = \left. \frac{\partial l}{\partial \alpha} \sqrt{J^{22}} \right|_{\theta=\widetilde{\theta}}.$$

Let $\mathbf{C} = \operatorname{diag}(\mathbf{C}_{1,NT,b}, \mathbf{C}_{2,NT,b})$, where $\mathbf{C}_{1,NT,b} = \operatorname{diag}(N\sqrt{T^*}\mathbf{I}_{Nr}, \sqrt{NT^*})$ and $\mathbf{C}_{2,NT,b} = T^*$. From Lemma 1 it is straightforward to show that

$$\mathbf{C}^{-1} \widetilde{\mathbf{J}}(\theta) \mathbf{C}^{-\prime} = \begin{pmatrix} \mathbf{O}_p(1) & \mathbf{O}_p\left(\frac{1}{\sqrt{N}}\right) & \mathbf{O}_p\left(\frac{1}{\sqrt{T^*}}\right) \\ \mathbf{O}_p\left(\frac{1}{\sqrt{N}}\right) & \mathcal{O}_p(1) & \mathcal{O}_p\left(\frac{1}{\sqrt{NT^*}}\right) \\ \mathbf{O}_p\left(\frac{1}{\sqrt{T^*}}\right) & \mathcal{O}_p\left(\frac{1}{\sqrt{NT^*}}\right) & \mathcal{O}_p(1) \end{pmatrix},$$

---

[3] Even if $\pi_{11,a}$ does not have full rank, we may simply assume that any linearly dependent, and therefore redundant, rows or columns have been removed such that $\pi_{11,a}$ is invertible.

such that

$$\vartheta_b \simeq \vartheta_b^* = \widetilde{\mathbf{V}}(\alpha)\sqrt{\widetilde{J}_{\alpha\alpha}^{-1}} = \frac{T^* \operatorname{tr}\left(\widehat{\mathbf{M}}\mathbf{S}_{01}^{-1}\right) - 2\operatorname{tr}\left(\widehat{\mathbf{M}}\mathbf{S}_{01}^{-1}\mathbf{S}_0\mathbf{S}_{01}^{-1}\right) + \operatorname{tr}\left(\widehat{\mathbf{M}}\mathbf{S}_{01}^{-1}\mathbf{S}_{00}\mathbf{S}_{01}^{-1}\right)}{\sqrt{2T^{*2}\operatorname{tr}\left(\widehat{\mathbf{M}}\mathbf{S}_{01}^{-1}\widehat{\mathbf{M}}\mathbf{S}_{01}^{-1}\right)}}, \qquad (19)$$

where $\widetilde{\mathbf{V}}(\alpha) = \mathbf{V}(\alpha)|_{\theta=\widetilde{\theta}}$ and $\widetilde{J}_{\alpha\alpha} = J_{\alpha\alpha}|_{\theta=\widetilde{\theta}}$ are found in Appendix A, and $\widehat{\mathbf{M}} = \widehat{\mathbf{A}}\widehat{\mathbf{H}}\widehat{\mathbf{A}}'$ was defined in Section 2.[4] Similar as with Case A, the asymptotic distribution as $T \to \infty$ is a weighted sum of χ_1^2-variables as stated in Theorem 1.

Theorem 1 *For $1 \leq r < N$, and $\eta_j = \varphi_j(\mathbf{\Lambda}'\mathbf{\Lambda})$, where $\eta_1 \geq \eta_2 \geq \cdots \geq \eta_r > 0$, the asymptotic distribution of the LM-type statistic (19) as $T \to \infty$ is*

$$\vartheta_b^* \xrightarrow{d} \frac{1}{\sqrt{2\sum_{j=1}^r \overline{w}_j^2}}\left(\sum_{j=1}^r \overline{w}_j \chi_{1,j}^2 - \sum_{j=1}^r \overline{w}_j\right), \qquad (20)$$

where $\chi_{1,j}^2$ are independent over $j = 1, 2, \ldots, r$ and where $\overline{w}_j = \frac{\eta_j}{\sigma_\varepsilon^2 + \eta_j}$, and the asymptotic distribution as $(T, N)_s \to \infty$ is

$$\vartheta_b^* \xrightarrow{d} \frac{1}{\sqrt{2r}}\left(\chi_r^2 - r\right). \qquad (21)$$

Note that for the weights $\overline{w}_j$ in (20) and the corresponding weights w_i in (16), we have the relation $\overline{w}_j + w_i = 1$ for $j, i = 1, 2, \ldots, r$. Because $\eta_i = \mathcal{O}(N)$, implying $\overline{w}_j \to 1$ and $w_i \to 0$ for $j, i = 1, 2, \ldots, r$ as $N \to \infty$, the limiting distribution (21) follows as $(T, N)_s \to \infty$, and likewise the standard normal distribution for (16). For small N we follow Zhou and Solberger (2012) and consider an approximation due to Satterthwaite (1946). Let $\nu_1 = \sum_{j=1}^r \overline{w}_j$ such that $\sum_{j=1}^r \overline{w}_j \chi_{1,j}^2 = \nu_1 \sum_{j=1}^r a_j \chi_{1,j}^2$, where $a_j = \overline{w}_j/\nu_1$ and $\sum_{j=1}^r a_j = 1$. Then, for some u, we have approximately that

$$\sum_{j=1}^r a_j \chi_{1,j}^2 \overset{app.}{\sim} \tfrac{1}{u}\chi_u^2, \qquad (22)$$

in the sense of fitting first and second moments of the left and right side. The solution for u with respect to the statistic (19) is given by the following proposition:

Proposition 1 *Let ϑ_b^* have the weighted χ^2-distribution (20). Matching first and second moments of (22) yields*

$$\vartheta_b^* \overset{app.}{\sim} \frac{1}{\sqrt{2u_b}}\left(\chi_{u_b}^2 - u_b\right),$$

where $u_b = \frac{[\operatorname{tr}(\mathbf{M}\mathbf{\Omega}^{-1})]^2}{\operatorname{tr}(\mathbf{M}\mathbf{\Omega}^{-1}\mathbf{M}\mathbf{\Omega}^{-1})}$.

---

[4]From the proof of Lemma 1 (viii) in Appendix B it follows that, as $T \to \infty$, $\mathbf{C}_{2,NT,b}^{-1}\widetilde{\mathbf{J}}_{22,b}\mathbf{C}_{2,NT,b}^{-\prime} \xrightarrow{p} \pi_{22,b} = \frac{1}{8}\sum_{j=1}^r \overline{w}_j^2$, where $\overline{w}_j = 1 - w_j$ for $j = 1, 2, \ldots, r$, and where $= \pi_{22,b} \to \frac{r}{8}$ as $(T, N)_s \to \infty$.

3.3 Case C, alternative hypothesis with $f_t \sim I(0)$ and/or $u_t \sim I(0)$

For Case C we only keep restrictions under the null hypothesis, and consider the hypotheses

$$\mathcal{H}_0 : \alpha = 1, \rho = 1,$$
$$\mathcal{H}_1 : \alpha < 1 \text{ and/or } \rho < 1.$$

We may then partition the parameter vector such that $\boldsymbol{\theta}_{1,c} = (\boldsymbol{\Lambda}_v', \sigma_\varepsilon^2)'$ and $\boldsymbol{\theta}_{2,c} = (\alpha, \rho)'$, and decompose the information matrix as

$$\mathbf{J}_{11,c} = \begin{pmatrix} J_{\Lambda_v \Lambda_v'} & J_{\Lambda_v \sigma_\varepsilon^2} \\ J_{\sigma_\varepsilon^2 \Lambda_v'} & J_{\sigma_\varepsilon^2 \sigma_\varepsilon^2} \end{pmatrix}; \quad \mathbf{J}_{21,c} = \mathbf{J}_{12,c}' = \begin{pmatrix} J_{\alpha \Lambda_v'} & J_{\alpha \sigma_\varepsilon^2} \\ J_{\rho \Lambda_v'} & J_{\rho \sigma_\varepsilon^2} \end{pmatrix}; \quad \mathbf{J}_{22,c} = \begin{pmatrix} J_{\alpha\alpha} & J_{\alpha\rho} \\ J_{\rho\alpha} & J_{\rho\rho} \end{pmatrix}. \tag{23}$$

If we define $\mathbf{C} = \mathrm{diag}(\mathbf{C}_{1,NT^*,c}, \mathbf{C}_{2,NT^*,c})$, where $\mathbf{C}_{1,NT,c} = \mathrm{diag}(N\sqrt{T^*}\mathbf{I}_{Nr}, \sqrt{NT^*})$ and $\mathbf{C}_{2,NT,c} = \mathrm{diag}(T^*, T^*\sqrt{N})$, then, using Lemma 1, we have that

$$\mathbf{C}^{-1}\widetilde{\mathbf{J}}(\boldsymbol{\theta})\mathbf{C}^{-\prime} = \begin{pmatrix} \mathbf{O}_p(1) & \mathbf{O}_p\left(\frac{1}{\sqrt{N}}\right) & \mathbf{O}_p\left(\frac{1}{\sqrt{T^*}}\right) & \mathbf{O}_p\left(\frac{1}{\sqrt{NT^*}}\right) \\ \mathbf{O}_p\left(\frac{1}{\sqrt{N}}\right) & \mathcal{O}_p(1) & \mathcal{O}_p\left(\frac{1}{\sqrt{NT^*}}\right) & \mathcal{O}_p\left(\frac{1}{\sqrt{T^*}}\right) \\ \mathbf{O}_p\left(\frac{1}{\sqrt{T^*}}\right) & \mathcal{O}_p\left(\frac{1}{\sqrt{NT^*}}\right) & \mathcal{O}_p(1) & \mathcal{O}_p\left(\frac{1}{\sqrt{N}}\right) \\ \mathbf{O}_p\left(\frac{1}{\sqrt{NT^*}}\right) & \mathcal{O}_p\left(\frac{1}{\sqrt{T^*}}\right) & \mathcal{O}_p\left(\frac{1}{\sqrt{N}}\right) & \mathcal{O}_p(1) \end{pmatrix}. \tag{24}$$

Hence, as $T \to \infty$, $\widetilde{\mathbf{J}}(\boldsymbol{\theta})$ is block-diagonal in probability. Also, corresponding to the lower right 2×2 block of (24), we have that, as $(T,N)_s \to \infty$,

$$\mathbf{C}_{2,NT,c}^{-1}\widetilde{\mathbf{J}}_{22,c}\mathbf{C}_{2,NT,c}^{-\prime} = \left[\left(\mathcal{O}_p(1), \mathcal{O}_p\left(\frac{1}{\sqrt{N}}\right)\right)', \left(\mathcal{O}_p\left(\frac{1}{\sqrt{N}}\right), \mathcal{O}_p(1)\right)'\right] \xrightarrow{p} \mathrm{diag}\left(\frac{r}{8}, \frac{1}{8}\right),$$

where we have used the results in proof of Lemma 1 (viii) and (x) in Appendix B.

Here we consider an LM-type test where the right score vector of (13) is replaced with $\mathbf{V}^* = \frac{1}{2\sqrt{2}}(\sqrt{2}T^*, T^*\sqrt{N})'$;

$$\vartheta_c^* = \mathbf{V}(\boldsymbol{\theta}_2)'\mathbf{J}_{22}^{-1}\mathbf{V}^*\Big|_{\boldsymbol{\theta}=\widetilde{\boldsymbol{\theta}}} = \frac{T^*\left\{\widetilde{\mathbf{V}}(\alpha)\left(\widetilde{J}_{\rho\rho} - \sqrt{\frac{N}{2}}\widetilde{J}_{\alpha\rho}\right) + \widetilde{\mathbf{V}}(\rho)\left(\sqrt{\frac{N}{2}}\widetilde{J}_{\alpha\alpha} - \widetilde{J}_{\alpha\rho}\right)\right\}}{2\left(\widetilde{J}_{\alpha\alpha}\widetilde{J}_{\rho\rho} - \widetilde{J}_{\alpha\rho}^2\right)}, \tag{25}$$

where, from Appendix A,

$$\widetilde{\mathbf{V}}(\alpha) = \frac{1}{4}\mathrm{tr}\left(\widehat{\mathbf{M}}\mathbf{S}_{01}^{-1}\mathbf{S}_{00}\mathbf{S}_{01}^{-1}\right) - \frac{1}{2}\mathrm{tr}\left(\widehat{\mathbf{M}}\mathbf{S}_{01}^{-1}\mathbf{S}_0\mathbf{S}_{01}^{-1}\right) + \frac{T^*}{4}\mathrm{tr}\left(\widehat{\mathbf{M}}\mathbf{S}_{01}^{-1}\right),$$

$$\widetilde{\mathbf{V}}(\rho) = \widehat{\sigma}_\varepsilon^2\left[\frac{1}{4}\mathrm{tr}\left(\mathbf{S}_{01}^{-1}\mathbf{S}_{00}\mathbf{S}_{01}^{-1}\right) - \frac{1}{2}\mathrm{tr}\left(\mathbf{S}_{01}^{-1}\mathbf{S}_0\mathbf{S}_{01}^{-1}\right) + \frac{T^*}{4}\mathrm{tr}\left(\mathbf{S}_{01}^{-1}\right)\right],$$

$$\widetilde{J}_{\alpha\alpha} = \frac{T^{*2}}{8}\mathrm{tr}\left(\widehat{\mathbf{M}}\mathbf{S}_{01}^{-1}\widehat{\mathbf{M}}\mathbf{S}_{01}^{-1}\right),$$

$$\widetilde{J}_{\rho\rho} = \frac{T^{*2}\widehat{\sigma}_\varepsilon^4}{8}\mathrm{tr}\left(\mathbf{S}_{01}^{-1}\mathbf{S}_{01}^{-1}\right),$$

$$\widetilde{J}_{\alpha\rho} = \frac{T^{*2}\widehat{\sigma}_\varepsilon^2}{8}\mathrm{tr}\left(\widehat{\mathbf{M}}\mathbf{S}_{01}^{-1}\mathbf{S}_{01}^{-1}\right),$$

where $\mathbf{S}_{01}, \mathbf{S}_0, \mathbf{S}_{00}, \widehat{\mathbf{M}}$ and $\widehat{\sigma}_\varepsilon^2$ were all defined before. The limiting distribution of the statistic

Table 1. Critical values for Case C

sig.	No. of factors (r) for ϑ_c^*							No. of factors (r) for ϑ_c^{**}						
	1	2	3	4	5	6	7	1	2	3	4	5	6	7
1%	-2.96	-2.82	-2.73	-2.67	-2.62	-2.59	-2.56	-2.73	-2.82	-2.87	-2.91	-2.93	-2.96	-2.98
5%	-2.22	-2.07	-1.98	-1.92	-1.88	-1.85	-1.83	-2.00	-2.07	-2.11	-2.13	-2.15	-2.16	-2.17
10%	-1.82	-1.66	-1.57	-1.52	-1.48	-1.46	-1.44	-1.61	-1.66	-1.69	-1.70	-1.72	-1.73	-1.73

(25) is given by Theorem 2.

Theorem 2 *As $(T, N)_s \to \infty$, the asymptotic distribution of the LM-type statistic (25) is*

$$\vartheta_c^* \overset{d}{\to} \left(\tfrac{1}{r}\mathcal{V}_1 - 1\right) + \mathcal{V}_2,$$

where $\mathcal{V}_1 \sim \chi^2(r)$ and $\mathcal{V}_2 \sim \mathcal{N}(0,1)$ are independent.

Another interesting case is found by letting $\mathbf{V}^* = \frac{1}{2\sqrt{2}}(T^*\sqrt{r}, T^*\sqrt{N})'$, such that

$$\vartheta_c^{**} = \mathbf{V}(\boldsymbol{\theta}_2)'\mathbf{J}_{22}^{-1}\mathbf{V}^*\Big|_{\boldsymbol{\theta}=\tilde{\boldsymbol{\theta}}} = \frac{T^*\left\{\widetilde{\mathbf{V}}(\alpha)\left(\widetilde{J}_{\rho\rho} - \sqrt{\frac{N}{r}}\widetilde{J}_{\alpha\rho}\right) + \widetilde{\mathbf{V}}(\rho)\left(\sqrt{\frac{N}{r}}\widetilde{J}_{\alpha\alpha} - \widetilde{J}_{\alpha\rho}\right)\right\}}{\sqrt{\frac{8}{r}\left(\widetilde{J}_{\alpha\alpha}\widetilde{J}_{\rho\rho} - \widetilde{J}_{\alpha\rho}^2\right)}}. \tag{26}$$

The limiting distribution for this statistic is given by Theorem 3. In essence, the statistics ϑ_c^* and ϑ_c^{**} are combinations of Case A and Case B. When $r = 2$ they are the same statistic, but when $r > 2$, ϑ_c^{**} will put more weight than ϑ_c^* on the information coming from the factors, while vice versa if $r = 1$. Thus, if $r > 2$ we expect ϑ_c^{**} to be more sensitive to stationarity in the factors, and ϑ_c^* to be more sensitive to stationarity in the idiosyncratic components, and vice versa if $r = 1$. The critical values for ϑ_c^* and ϑ_c^{**} are found in Table 1.

Theorem 3 *As $(T, N)_s \to \infty$, the asymptotic distribution of the LM-type statistic (26) is*

$$\vartheta_c^{**} \overset{d}{\to} \mathcal{Z}_1 + \mathcal{Z}_2,$$

where $\mathcal{Z}_1 \sim \frac{1}{\sqrt{2r}}\left(\chi_r^2 - r\right)$ and $\mathcal{Z}_2 \sim \mathcal{N}(0,1)$ are independent.

4   Monte Carlo simulations

In this section we evaluate size and power of the proposed LM-type statistics, and compare the performance with the PANIC framework of Bai and Ng (2004).

Equation (2) may be written in panel notation as

$$x_{i,t} = \mu_i + \sum_{j=1}^{r} \lambda_{i,j} f_{j,t} + u_{i,t}, \qquad i = 1, 2, \ldots, N; t = 1, 2, \ldots, T.$$

Here we generate data with

$$\mu_i = 0, \tag{27}$$

$$f_{j,t} = \alpha_j f_{j,t-1} + v_{j,t}, \quad v_{j,t} \sim \mathcal{N}(0,1), \ f_{j,-49} = 0, \tag{28}$$

$$u_{i,t} = \rho_i u_{i,t-1} + \varepsilon_{i,t}, \quad \varepsilon_{i,t} \sim \mathcal{N}(0,\sigma_\varepsilon^2), \ u_{i,-49} = 0, \tag{29}$$

$$\lambda_{i,j} \sim \mathcal{N}(1,\sigma_\lambda^2), \tag{30}$$

where the first 50 observations are discarded to ridden the impact of the initial values when the processes $f_{j,t}$ and $u_{i,t}$ are stationary. We consider the following DGPs:

> *DGP A* : Data is generated from $(27 - 30)$ with $\alpha_j = \alpha = 1$ and $\rho_i \leq 1$,
>
> *DGP B* : Data is generated from $(27 - 30)$ with $\alpha_j \leq 1$ and $\rho_i = \rho = 1$,
>
> *DGP C* : Data is generated from $(27 - 30)$ with $\alpha_j \leq 1$ and $\rho_i \leq 1$.

That is, DGP A corresponds to the null and alternative for Case A, DGP B corresponds to the null and alternative for Case B, and DGP C corresponds to the null and alternative for Case C. For all DGPs we set $\sigma_\lambda^2 = \sigma_\varepsilon^2 = 1$, and choose nominal significance level 5%.

Table 2 shows the result for DGPs A-C under homogeneity, i.e. $\alpha_j = \alpha$ for all j and $\rho_i = \rho$ for all i. The panel dimensions are $N = 25$ and $T = 100$, and the number of factors are $r = 1, 3, 5$. Table 3 shows the results under the same settings, but for the larger sample sizes $N = 50$ and $T = 200$. Consider first the statistic ϑ_a^*. As expected this statistic has the highest power of the proposed statistics under DGP A, because ϑ_a^* is derived conditional on $I(1)$-factors. Note, however, that under DGP B the size-adjusted power of ϑ_a^* is close to nominal size, indicating that the test-statistic is robust to the conditioning on integrated factors. This result is strengthened when we consider the larger sample sizes reported in Table 3. For the statistic ϑ_b^* the pattern is reversed. That is, under DGP B ϑ_b^* has the highest power, which is expected because the statistic is derived conditional on $I(1)$-idiosyncratic components. In general the power is quite low, but increases when the number of factors increases. However, as we will illustrate later, ϑ_b^* still has higher local power than PANIC. Note that, analogously, under DGP A the size-adjusted power of ϑ_b^* is close to nominal size for all choices of ρ, which indicates that the test-statistic is robust to the conditioning on integrated idiosyncratic components. Again, this result is strengthened when we consider the larger sample sizes in Table 3.

The benefit of using the joint statistics ϑ_c^* and ϑ_c^{**} can be seen under DGP C. For instance, in Table 2, when $r = 5$ and $\alpha = \rho = 0.99$, the powers of ϑ_c^* and ϑ_c^{**} are about 54%, while the powers of ϑ_a^* and ϑ_b^* are about 43% and 15% respectively. Table 2 and Table 3 also show that, as expected, ϑ_c^{**} is more sensitive than ϑ_c^* to stationary idiosyncratic components if $r = 1$, and vice versa if $r > 2$, which is seen under DGP A. Conversely, ϑ_c^* is more sensitive than ϑ_c^{**} to stationary factors if $r = 1$, and vice versa if $r > 2$, which is seen under DGP B.

Because, at least to our knowledge, the statistics of the type ϑ_c^* and ϑ_c^{**} have not been considered before in the literature, in a strict sense we have no other tests to compare with. Here we compare size and power with PANIC, where we look at DGP A and DGP B sep-

Table 2. Empirical size-adjusted power of test-statistics (%) under homogeneity for $N = 25, T = 100$

	α	ρ	ϑ_a^*	ϑ_b^*	ϑ_c^*	ϑ_c^{**}	ϑ_a^*	ϑ_b^*	ϑ_c^*	ϑ_c^{**}	ϑ_a^*	ϑ_b^*	ϑ_c^*	ϑ_c^{**}
				$r=1$				$r=3$				$r=5$		
DGP A	1	0.99	54.3	4.6	41.7	45.0	50.1	5.0	37.0	35.2	43.7	5.7	35.4	30.6
	1	0.95	100	4.5	80.1	86.9	100	5.3	84.0	79.3	99.7	5.9	87.1	74.8
	1	0.9	100	4.5	84.7	90.8	100	5.4	89.8	85.3	100	6.2	93.0	82.3
	1	0.8	100	4.8	86.4	92.1	100	5.6	91.9	87.7	100	6.3	94.6	85.4
	1	0.5	100	4.7	87.4	93.0	100	5.9	92.9	89.1	100	6.5	95.3	87.1
	1	0	100	4.7	88.0	93.4	100	6.0	93.2	89.8	100	7.1	95.5	87.9
DGP B	0.99	1	4.9	6.9	6.2	6.3	4.6	10.7	6.7	7.3	4.5	13.4	7.8	9.6
	0.95	1	4.7	10.0	8.6	8.3	4.4	33.8	11.9	13.9	4.2	54.3	14.0	19.4
	0.9	1	4.6	13.2	9.9	9.3	4.4	57.4	14.5	16.8	4.1	82.4	16.3	24.4
	0.8	1	4.6	17.0	10.6	9.9	4.3	77.9	15.6	18.6	3.9	95.7	17.7	26.7
	0.5	1	4.6	22.8	11.6	10.4	4.3	93.1	16.6	19.8	3.9	99.7	18.5	27.8
	0	1	4.6	28.2	11.8	10.8	4.4	99.0	16.7	20.0	3.8	100	18.5	28.4
DGP C	1	1	5.0	5.0	5.0	5.0	5.0	5.0	5.0	5.0	5.0	5.0	5.0	5.0
	0.99	0.99	54.5	6.4	50.0	53.5	50.0	11.0	52.2	51.9	43.4	14.9	53.7	53.8
	0.95	0.95	100	10.5	99.3	99.8	100	40.1	100	100	100	65.3	100	100
	0.8	0.5	100	21.7	100	100	100	94.3	100	100	100	100	100	100
	0.5	0.8	100	31.0	100	100	100	99.9	100	100	100	100	100	100
	0	0.9	100	41.5	100	100	100	100	100	100	100	100	100	100
	0.9	0	100	15.6	100	100	100	74.4	100	100	100	97.4	100	100

Note: The data is generated as $f_{j,t} = \alpha f_{j,t-1} + v_{j,t}$, $v_{j,t} \sim \mathcal{N}(0,1)$, $u_{i,t} = \rho u_{i,t-1} + \varepsilon_{i,t}$, $\varepsilon_{i,t} \sim \mathcal{N}(0, \sigma_\varepsilon^2)$, $\lambda_i \sim \mathcal{N}_r(\mathbf{1}, \sigma_\lambda^2 \mathbf{I})$. The loadings are generated once and then kept fixed. The replication number is 5,000.

arately. Evaluating ϑ_c^* and ϑ_c^{**} under these DGPs will then show, in comparison to PANIC, how they respond to stationarity in the idiosyncratic components and stationarity in the factors separately. Under DGP A we compare size and local power with the Fisher-type test in Bai and Ng (2004),

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

where $p_{\hat{u}}^c(i)$ are p-values of idiosyncratic augmented Dickey-Fuller (ADF) tests. The idiosyncratic components are found by first taking first-differences on the panel (2), and then applying principal components to estimate the factors in first differences, $\widehat{\Delta \mathbf{f}}_t$, and the associate factor loading, $\widehat{\mathbf{\Lambda}}$, such that the idiosyncratic components, in first differences, are left as residuals, $\widehat{\Delta u}_{i,t} = \Delta x_{i,t} - \widehat{\lambda}_i' \widehat{\Delta \mathbf{f}}_t$. The estimated idiosyncratic components in levels are then found by re-accumulating, $\widehat{u}_{i,t} = \sum_{s=2}^{t} \widehat{\Delta u}_{i,s}$. As $N \to \infty$ we have that $P_{\hat{u}}^c \xrightarrow{d} \mathcal{N}(0,1)$ by the central limit theorem. As suggested by Bai and Ng (2004, p. 1148), before pooling we first locate the p-values of the individual Dickey-Fuller tests for finite T in a simulation of 10,000 replications.[5]

---

[5]Because we consider AR(1) processes, to be fair we add no lags in the Dickey-Fuller tests. The criteria

Table 3. Empirical size-adjusted power of test-statistics (%) under homogeneity for $N = 50, T = 200$

	α	ρ	ϑ_a^*	ϑ_b^*	ϑ_c^*	ϑ_c^{**}	ϑ_a^*	ϑ_b^*	ϑ_c^*	ϑ_c^{**}	ϑ_a^*	ϑ_b^*	ϑ_c^*	ϑ_c^{**}
				$r=1$				$r=3$				$r=5$		
DGP A	1	0.99	100	5.2	83.7	88.9	100	5.3	91.0	86.0	99.9	5.3	94.2	84.1
	1	0.95	100	5.1	93.1	96.7	100	5.5	98.5	96.9	100	5.4	99.6	96.8
	1	0.9	100	4.9	94.0	97.3	100	5.6	98.8	97.5	100	5.5	99.7	97.7
	1	0.8	100	4.8	94.3	97.6	100	5.6	99.1	97.9	100	5.7	99.7	98.0
	1	0.5	100	4.9	94.6	97.7	100	5.7	99.2	98.1	100	5.9	99.7	98.2
	1	0	100	5.0	94.7	97.8	100	5.8	99.2	98.1	100	6.1	99.7	98.4
DGP B	0.99	1	4.8	7.3	6.8	6.5	4.6	18.8	10.2	10.9	4.6	28.7	11.5	14.2
	0.95	1	4.9	13.9	9.6	8.9	4.5	62.8	14.9	16.7	4.2	90.2	16.6	23.3
	0.9	1	4.9	18.9	10.5	9.5	4.4	86.3	15.9	17.8	4.2	99.1	17.7	26.0
	0.8	1	4.9	23.1	10.7	9.7	4.5	96.5	16.5	18.8	4.1	100	18.4	27.4
	0.5	1	5.0	32.7	11.0	9.8	4.6	99.8	16.8	19.3	4.1	100	18.8	28.0
	0	1	4.9	42.2	11.2	9.9	4.5	100	16.9	19.5	4.1	100	18.9	28.3
DGP C	1	1	5.0	5.0	5.0	5.0	5.0	5.0	5.0	5.0	5.0	5.0	5.0	5.0
	0.99	0.99	100	7.7	97.5	98.8	100	19.5	99.8	99.6	99.9	30.4	100	99.7
	0.95	0.95	100	15.0	100	100	100	70.0	100	100	100	95.2	100	100
	0.8	0.5	100	28.8	100	100	100	99.9	100	100	100	100	100	100
	0.5	0.8	100	45.7	100	100	100	100	100	100	100	100	100	100
	0	0.9	100	59.6	100	100	100	100	100	100	100	100	100	100
	0.9	0	100	20.8	100	100	100	94.4	100	100	100	100	100	100

Note: The data is generated as $f_{j,t} = \alpha f_{j,t-1} + v_{j,t}$, $v_{j,t} \sim \mathcal{N}(0,1)$, $u_{i,t} = \rho u_{i,t-1} + \varepsilon_{i,t}$, $\varepsilon_{i,t} \sim \mathcal{N}(0,\sigma_\varepsilon^2)$, $\lambda_i \sim \mathcal{N}_r(\mathbf{1}, \sigma_\lambda^2 \mathbf{I})$. The loadings are generated once and then kept fixed. The replication number is 5,000.

Table 4 shows the result under DGP A for $r = 1, 2, 3$, where we have also imposed local heterogeneity by letting

$$\rho_i = 1 - \frac{c_i}{T\sqrt{N}}, \quad c_i \sim U(1, 10).$$

The statistic ϑ_a^* has satisfactory size properties and quite substantially higher local power than $P_{\hat{u}}^c$ (denoted BN in Table 4) for all dimensions considered here, which was already found for the similar statistic ϑ_a in Zhou and Solberger (2012). The statistics ϑ_c^* and ϑ_c^{**} tend to be undersized for finite N and T, but in general they still have higher local power than $P_{\hat{u}}^c$. For example, when $N = 60$ and $T = 100$, and for one factor, the local power of ϑ_c^*, ϑ_c^{**} and $P_{\hat{u}}^c$ are 51%, 54% and 37% respectively. The differences in power are enlarged when considering size-adjusted power. The results also show that even though the tests derived in this paper are homogenous they still have power against a local heterogenous alternative.

To test for non-stationarity in the common component the literature offers only a handful

---

suggested by Bai and Ng, $k = 4[min(N,T)/100]^{1/4}$, would indeed tend to overestimate the number of lags, and have a negative impact on power.

Table 4. Empirical size and local power (%) under DGP A with heterogenous idiosyncratic components

	N	T	Size				Local power				Size-adj. local power			
			ϑ_a^*	ϑ_c^*	ϑ_c^{**}	BN	ϑ_a^*	ϑ_c^*	ϑ_c^{**}	BN	ϑ_a^*	ϑ_c^*	ϑ_c^{**}	BN
$r=1$	30	50	4.4	3.2	3.0	7.2	47.5	32.6	33.5	33.6	49.8	38.7	40.9	26.5
		100	5.0	4.1	3.9	5.5	58.2	40.9	43.1	32.6	58.2	44.8	47.5	30.5
		200	5.3	4.3	3.9	6.3	70.3	47.8	50.8	41.6	69.4	50.6	55.1	36.9
	60	100	4.4	3.8	3.6	4.8	70.7	51.2	54.1	36.7	73.3	55.7	59.6	37.2
		200	5.5	3.8	3.9	6.0	80.4	58.0	62.7	46.0	79.0	61.3	66.3	42.1
$r=2$	40	50	4.8	3.3	3.3	7.8	40.2	33.3	33.3	33.5	49.3	42.3	42.3	25.4
		100	4.5	3.5	3.5	5.3	62.8	44.8	44.8	34.3	64.9	50.8	50.8	33.2
		200	4.8	4.1	4.1	5.9	75.5	52.7	52.7	42.7	76.2	56.5	56.5	39.2
	80	100	4.8	3.6	3.6	5.2	69.8	52.0	52.0	34.4	70.4	57.2	57.2	33.4
		200	5.0	4.4	4.4	6.0	79.5	59.2	59.2	44.3	79.5	61.6	61.6	38.8
$r=3$	50	50	4.9	2.9	2.9	8.1	47.9	34.5	32.0	33.2	48.0	42.8	40.3	23.2
		100	4.4	3.8	3.8	5.5	64.7	47.7	44.6	34.0	66.6	53.9	50.3	32.5
		200	5.1	4.2	4.3	6.3	76.7	55.8	52.4	44.6	76.2	59.1	55.0	39.9
	100	100	5.0	3.5	3.3	4.7	70.1	53.6	49.6	33.8	70.2	59.3	55.9	34.8
		200	5.6	4.0	4.1	6.2	79.6	61.3	56.7	43.3	77.9	65.0	60.2	39.2

Note: The data is generated as $f_{j,t} = f_{j,t-1} + v_{j,t}$, $v_{j,t} \sim \mathcal{N}(0,1)$, $u_{i,t} = \rho_i u_{i,t-1} + \varepsilon_{i,t}$, $\varepsilon_{i,t} \sim \mathcal{N}(0,\sigma_\varepsilon^2)$, $\rho_i = 1 - \frac{c_i}{T\sqrt{N}}$, $c_i \sim U(1,10)$, $\lambda_i \sim \mathcal{N}_r(\mathbf{1}, \sigma_\lambda^2 \mathbf{I})$. The loadings and the drift terms c_i are generated once and then kept fixed. The replication number is 5,000.

of methods. Bai (2004) and Peña and Poncela (2006) derive tests for the number of nonstationary factors when the idiosyncratic components are $I(0)$. Hence these tests are not applicable here. In case of several factors, the PANIC procedure of Bai and Ng (2004) offers sequential tests for the number of stochastic trends among the factors based on Stock and Watson (1988). In case of a single factor they propose an ADF test on the estimated factor. For $r > 1$ we consider the statistic MQ_c^c (Bai and Ng, 2004, p. 1133). This statistic is calculated as follows: Start with the hypothesis $\mathcal{H}_0 : r_1 = m = r$, where r_1 is the number of stochastic trends and r is the number of factors. Demean the estimated factors as $\widehat{\mathbf{f}}_t^c = \widehat{\mathbf{f}}_t - \bar{\widehat{\mathbf{f}}}_t$, where $\bar{\widehat{\mathbf{f}}}_t = (T-1)^{-1} \sum_t^T \widehat{\mathbf{f}}_t$, and let $\widehat{\mathbf{B}}_t^c = \widehat{\beta}_\perp' \widehat{\mathbf{f}}_t^c$, where $\widehat{\beta}_\perp$ are the m eigenvectors associated with the m largest eigenvalues of $T^{-2} \sum_{t=2}^T \widehat{\mathbf{f}}_t^c \widehat{\mathbf{f}}_t^{c\prime}$. Next, let $\widehat{\xi}_t^c$ be the residuals from estimating a VAR(1) model for $\widehat{\mathbf{B}}_t^c$, and calculate

$$\widehat{\mathbf{\Sigma}}_1^c = \sum_{j=1}^J K(j) \left(T^{-1} \sum_{t=2}^T \widehat{\xi}_{t-j}^c \widehat{\xi}_t^{c\prime} \right),$$

where, for $j = 0, 1, \ldots, J$, $K(j) = 1 - j/(J+1)$ are truncated Bartlett Kernel weights. Here we set $J = 4 \, \text{ceil}[\min(N,T)/100]^{1/4}$ as suggested by Bai and Ng (2004, p. 1150). Finally, let

Table 5. Empirical size and local power (%) under DGP B with homogeneous factors

	N	T	Size				Local power				Size-adj. local power			
			ϑ_b^*	ϑ_c^*	ϑ_c^{**}	BN	ϑ_b^*	ϑ_c^*	ϑ_c^{**}	BN	ϑ_b^*	ϑ_c^*	ϑ_c^{**}	BN
$r = 1$	10	50	5.5	3.3	3.0	6.4	7.7	5.0	4.2	9.4	7.1	7.0	7.0	7.8
		100	5.3	3.0	2.6	5.7	8.8	4.7	4.1	8.8	8.1	7.2	7.2	7.6
		200	5.4	3.7	3.3	5.1	9.4	5.1	4.5	8.0	8.6	7.6	7.4	7.8
	50	100	5.0	4.2	4.0	5.6	7.2	5.3	4.9	6.6	7.2	6.2	6.3	6.1
		200	5.1	4.4	4.1	5.6	8.1	5.8	5.6	6.0	7.9	6.7	6.4	5.4
	100	200	5.0	4.2	4.2	5.6	7.5	5.8	5.4	5.6	7.5	6.9	6.4	5.0
$r = 3$	10	50	4.9	4.5	4.7	1.5	14.8	10.0	10.6	1.9	15.2	10.7	11.4	6.6
		100	4.2	4.7	5.0	3.1	15.6	10.4	11.5	4.4	18.8	11.0	11.5	6.6
		200	5.2	4.3	4.6	4.1	18.7	11.2	12.4	5.9	18.4	12.8	14.0	6.8
	50	100	4.5	4.2	4.2	3.3	12.5	7.4	7.7	3.8	13.2	8.7	8.8	5.6
		200	5.1	4.1	4.2	3.8	14.3	7.9	8.4	4.2	14.1	9.6	9.7	5.5
	100	200	5.0	4.3	4.2	3.7	11.3	7.0	7.3	3.9	11.3	8.5	8.7	5.6
$r = 5$	10	50	4.3	5.8	6.0	0.3	22.2	14.9	17.8	0.4	25.9	12.8	15.2	5.3
		100	4.4	6.6	6.9	1.8	24.8	16.4	20.3	2.7	28.1	13.1	15.8	6.4
		200	4.7	5.8	6.6	3.7	28.6	18.0	22.5	4.6	30.0	15.5	19.1	5.9
	50	100	4.6	4.2	4.3	1.7	18.5	8.0	9.7	2.4	20.1	9.3	11.0	5.4
		200	5.1	4.1	4.3	3.6	21.4	9.1	11.0	4.2	21.0	10.7	12.8	5.8
	100	200	4.9	4.1	4.0	4.0	16.8	7.2	8.5	3.9	16.9	8.9	10.6	5.1

Note: The data is generated as $u_{i,t} = u_{i,t-1} + \varepsilon_{i,t}$, $\varepsilon_{i,t} \sim \mathcal{N}(0,\sigma_\varepsilon^2)$, $f_{j,t} = \alpha f_{j,t-1} + v_{j,t}$, $v_{j,t} \sim \mathcal{N}(0,1)$, $\alpha = 1 - \frac{5}{T\sqrt{N}}$, $\lambda_i \sim \mathcal{N}_r(\mathbf{1}, \sigma_\lambda^2 \mathbf{I})$. The loadings are generated once and then kept fixed. The replication number is 5,000.

$v_c^c(m)$ be the smallest eigenvalue of

$$\widehat{\mathbf{\Phi}}_c^c(m) = \frac{1}{2} \left[\sum_{t=2}^T \left(\widehat{\mathbf{B}}_t^c \widehat{\mathbf{B}}_{t-1}^{c\prime} + \widehat{\mathbf{B}}_{t-1}^c \widehat{\mathbf{B}}_t^{c\prime} \right) - T \left(\widehat{\mathbf{\Sigma}}_1^c + \widehat{\mathbf{\Sigma}}_1^{c\prime} \right) \right] \left(\sum_{t=2}^T \widehat{\mathbf{B}}_{t-1}^c \widehat{\mathbf{B}}_{t-1}^{c\prime} \right)^{-1},$$

and calculate the test-statistic $MQ_c^c(m) = T\left[v_c^c(m) - 1\right]$. The critical values are found from Bai and Ng (2004, Table 1, p. 1136). If $\mathcal{H}_0 : r_1 = m$ is rejected, then repeat for $\mathcal{H}_0 : r_1 = m - 1$ until the number of stochastic trends is decided. Here we are not concerned with the specific number of stochastic trends that the sequential procedure will decide. Instead, we consider DGP B with only stationary factors and compare rejection rates for the statistics ϑ_b^*, ϑ_c^* and ϑ_c^{**} with the number of times MQ_c^c rejects the null $\mathcal{H}_0 : r_1 = r$, i.e. concluding correctly that "the number of stochastic trends among the factors is less than the number of factors". As before, the number of factors is treated as known.

Table 5 shows the result under DGP B for $r = 1, 3, 5$, where we have imposed local homogeneity by letting

$$\alpha = 1 - \frac{c}{T\sqrt{N}}, \quad c = 5.$$

Here BN denotes the ADF test when $r = 1$ and MQ_c^c when $r > 1$. The simulations reveal some interesting results. First, for $r = 1$ all tests have very low local power, but they are more or less equivalent in terms of size-adjusted local power. Also, here ϑ_c^* and ϑ_c^{**} are undersized in such a manner that the local power is sometimes under nominal size. When we increase the number of factors, the local power and the size-adjusted local power increase for the proposed statistics ϑ_b^*, ϑ_c^* and ϑ_c^{**}. However, for MQ_c^c, as we increase the number of factors, the size-adjusted local power remains roughly constant (or even decreases), and it is very low. This indicates that the statistic has a hard time rejecting the null of "the number of stochastic trends equals the number of factors" when the factors are close to having unit roots. Also, MQ_c^c is at times severely undersized with the result that it too has local power below nominal size. This seems to be especially true when N is not large enough relative the number of factors.

5 Conclusions

The exact factor model has recently been proposed as a parsimonious representation of the static factor model allowing for consistent likelihood based inference also when the true model is the approximate factor model. In this paper we extend the results of Zhou and Solberger (2012) who derived the limiting distribution of an LM-type test for idiosyncratic unit roots in the exact static factor model conditional on $I(1)$ factors. We derive the limiting distribution of an LM-type test for unit roots in the factors conditional on $I(1)$ idiosyncratic components, and two additional LM-type tests, which test jointly for unit roots in the factors and the idiosyncratic components. In a simulation study we investigate the size and power properties of the proposed statistics, and compare the performance with the PANIC procedure of Bai and Ng (2004). The simulations suggest that; (i) the proposed test-statistics are robust to the conditioning on, respectively, $I(1)$ factors and $I(1)$ idiosyncratic components if the panel dimensions are large enough, which may then be seen as a misspecification leading to correct inference for large panels; (ii) the likelihood based statistics have higher local power than the PANIC procedure which is based on principal components; and (iii) PANIC has virtually no power against locally stationary factors. The results in this paper are derived under the important AR(1) case, which should contribute to the development of likelihood based unit root tests in dynamic factor models with more general dynamic properties than considered here.

Acknowledgements

Appendix A Score and information

Let l be the log-likelihood (10) with $\boldsymbol{\Sigma} = E\left(\mathbf{Y}_v\mathbf{Y}_v'\right)$, and let $\widetilde{\boldsymbol{\theta}} = (\widehat{\boldsymbol{\Lambda}}_v', \widehat{\sigma}^2, 1, 1)'$ be the restricted MLE where $\widehat{\boldsymbol{\Lambda}}$ and $\widehat{\sigma}^2$ are MLEs and $\alpha = \rho = 1$ is the restriction under the null hypothesis. Also, for the matrices $\boldsymbol{\Sigma}(\boldsymbol{\theta})$ and $\boldsymbol{\Omega}(\boldsymbol{\theta})$, let $\widetilde{\boldsymbol{\Sigma}}$ and $\widehat{\boldsymbol{\Omega}}$ denote $\boldsymbol{\Sigma}(\widetilde{\boldsymbol{\theta}})$ and $\boldsymbol{\Omega}(\widetilde{\boldsymbol{\theta}})$, where $\widetilde{\boldsymbol{\Sigma}}^{-1} = (\mathbf{I}_{T^*} \otimes \widetilde{\boldsymbol{\Omega}}^{-1})$. Similarly, for any subsets $\boldsymbol{\omega}, \boldsymbol{\nu} \subseteq \boldsymbol{\theta}$, let $\widetilde{\mathbf{V}}(\boldsymbol{\omega})$ and $\widetilde{\mathbf{J}}_{\boldsymbol{\omega}\boldsymbol{\nu}'}$ denote $\mathbf{V}(\boldsymbol{\omega})|_{\boldsymbol{\theta}=\widetilde{\boldsymbol{\theta}}}$ and $\mathbf{J}_{\boldsymbol{\omega}\boldsymbol{\nu}'}|_{\boldsymbol{\theta}=\widetilde{\boldsymbol{\theta}}}$ respectively. By definition, the full score vector evaluated at $\boldsymbol{\theta} = \widetilde{\boldsymbol{\theta}}$ is for the three cases respectively;

$$\widetilde{\mathbf{V}}(\boldsymbol{\theta}) = \left[\left(\frac{\partial l}{\partial \mathrm{vec}\boldsymbol{\Lambda}}\right)', \frac{\partial l}{\partial \sigma_\varepsilon^2}, \frac{\partial l}{\partial \alpha}, \frac{\partial l}{\partial \rho}\right]'_{\boldsymbol{\theta}=\widetilde{\boldsymbol{\theta}}} = \left[\mathbf{0}, \widetilde{\mathbf{V}}(\alpha), \widetilde{\mathbf{V}}(\rho)\right]', \text{ for Case C;}$$

$$\widetilde{\mathbf{V}}(\boldsymbol{\theta}) = \left[\left(\frac{\partial l}{\partial \mathrm{vec}\boldsymbol{\Lambda}}\right)', \frac{\partial l}{\partial \sigma_\varepsilon^2}, \frac{\partial l}{\partial \alpha}\right]'_{\boldsymbol{\theta}=\widetilde{\boldsymbol{\theta}}} = \left[\mathbf{0}, \widetilde{\mathbf{V}}(\alpha)\right]', \text{ for Case B;}$$

$$\widetilde{\mathbf{V}}(\boldsymbol{\theta}) = \left[\left(\frac{\partial l}{\partial \mathrm{vec}\boldsymbol{\Lambda}}\right)', \frac{\partial l}{\partial \sigma_\varepsilon^2}, \frac{\partial l}{\partial \rho}\right]'_{\boldsymbol{\theta}=\widetilde{\boldsymbol{\theta}}} = \left[\mathbf{0}, \widetilde{\mathbf{V}}(\rho)\right]', \text{ for Case A.}$$

Using a standard result (see e.g. Hartley and Rao, 1967) we have that

$$\widetilde{\mathbf{V}}(\alpha) = \left.\frac{\partial l}{\partial \alpha}\right|_{\boldsymbol{\theta}=\widetilde{\boldsymbol{\theta}}} = -\frac{1}{2}\,\mathrm{tr}\left(\left.\frac{\partial \boldsymbol{\Sigma}}{\partial \alpha}\right|_{\boldsymbol{\theta}=\widetilde{\boldsymbol{\theta}}} \widetilde{\boldsymbol{\Sigma}}^{-1}\right) + \frac{1}{2}\,\mathrm{tr}\left(\left.\frac{\partial \boldsymbol{\Sigma}}{\partial \alpha}\right|_{\boldsymbol{\theta}=\widetilde{\boldsymbol{\theta}}} \widetilde{\boldsymbol{\Sigma}}^{-1}\mathbf{Y}_v\mathbf{Y}_v'\widetilde{\boldsymbol{\Sigma}}^{-1}\right), \tag{A1}$$

where $\left.\frac{\partial \boldsymbol{\Sigma}}{\partial \alpha}\right|_{\boldsymbol{\theta}=\widetilde{\boldsymbol{\theta}}} = \left(\left.\frac{\partial \boldsymbol{\Psi}}{\partial \alpha}\right|_{\boldsymbol{\theta}=\widetilde{\boldsymbol{\theta}}} \otimes \widehat{\mathbf{M}}\right)$, and where straightforward derivations will show that

$$\left.\frac{\partial \boldsymbol{\Psi}}{\partial \alpha}\right|_{\boldsymbol{\theta}=\widetilde{\boldsymbol{\theta}}} = \begin{pmatrix} -\frac{1}{2} & \frac{1}{2} & \cdots & \frac{1}{2} \\ \frac{1}{2} & -\frac{1}{2} & \ddots & \vdots \\ \vdots & \ddots & \ddots & \frac{1}{2} \\ \frac{1}{2} & \cdots & \frac{1}{2} & -\frac{1}{2} \end{pmatrix} = \frac{1}{2}\mathbf{1}\mathbf{1}' - \mathbf{I}_{T^*}.$$

The first part of (A1) is, using well-known results for the trace and the inverse of Kronecker products (see e.g. Magnus and Neudecker, 2001, pp. 28-31),

$$-\frac{1}{2}\,\mathrm{tr}\left(\left.\frac{\partial \boldsymbol{\Sigma}}{\partial \alpha}\right|_{\boldsymbol{\theta}=\widetilde{\boldsymbol{\theta}}} \widetilde{\boldsymbol{\Sigma}}^{-1}\right) = -\frac{1}{2}\,\mathrm{tr}\left[\left(\tfrac{1}{2}\mathbf{1}\mathbf{1}' - \mathbf{I}_{T^*}\right) \otimes \widehat{\mathbf{M}}\widetilde{\boldsymbol{\Omega}}^{-1}\right] = \frac{T^*}{4}\,\mathrm{tr}\left(\widehat{\mathbf{M}}\widetilde{\boldsymbol{\Omega}}^{-1}\right).$$

Similarly, the second part of (A1) is

$$\begin{aligned}
\frac{1}{2}\,\mathrm{tr}\left(\left.\frac{\partial \boldsymbol{\Sigma}}{\partial \alpha}\right|_{\boldsymbol{\theta}=\widetilde{\boldsymbol{\theta}}} \widetilde{\boldsymbol{\Sigma}}^{-1}\mathbf{Y}_v\mathbf{Y}_v'\widetilde{\boldsymbol{\Sigma}}^{-1}\right) =& \frac{1}{2}\,\mathrm{tr}\left\{\left[\left(\tfrac{1}{2}\mathbf{1}\mathbf{1}' - \mathbf{I}_{T^*}\right) \otimes \widehat{\mathbf{M}}\widetilde{\boldsymbol{\Omega}}^{-1}\right]\mathbf{Y}_v\mathbf{Y}_v'\left(\mathbf{I}_{T^*} \otimes \widetilde{\boldsymbol{\Omega}}^{-1}\right)\right\} \\
=& \frac{1}{4}\,\mathrm{tr}\left[\left(\mathbf{1}\mathbf{1}' \otimes \mathbf{I}_N\right)\left(\mathbf{I}_{T^*} \otimes \widehat{\mathbf{M}}\widetilde{\boldsymbol{\Omega}}^{-1}\right)\mathbf{Y}_v\mathbf{Y}_v'\left(\mathbf{I}_{T^*} \otimes \widetilde{\boldsymbol{\Omega}}^{-1}\right)\right] \\
& -\frac{1}{2}\,\mathrm{tr}\left[\left(\mathbf{I}_{T^*} \otimes \widehat{\mathbf{M}}\widetilde{\boldsymbol{\Omega}}^{-1}\right)\mathbf{Y}_v\mathbf{Y}_v'\left(\mathbf{I}_{T^*} \otimes \widetilde{\boldsymbol{\Omega}}^{-1}\right)\right],
\end{aligned}$$

where

$$\left(\mathbf{I}_{T^*} \otimes \widehat{\mathbf{M}}\widetilde{\boldsymbol{\Omega}}^{-1}\right)\mathbf{Y}_v\mathbf{Y}_v'\left(\mathbf{I}_{T^*} \otimes \widetilde{\boldsymbol{\Omega}}^{-1}\right) = \begin{pmatrix} \widehat{\mathbf{M}}\widetilde{\boldsymbol{\Omega}}^{-1}\mathbf{S}_{22}\widetilde{\boldsymbol{\Omega}}^{-1} & \cdots & \widehat{\mathbf{M}}\widetilde{\boldsymbol{\Omega}}^{-1}\mathbf{S}_{2T}\widetilde{\boldsymbol{\Omega}}^{-1} \\ \vdots & \ddots & \vdots \\ \widehat{\mathbf{M}}\widetilde{\boldsymbol{\Omega}}^{-1}\mathbf{S}_{T2}\widetilde{\boldsymbol{\Omega}}^{-1} & \cdots & \widehat{\mathbf{M}}\widetilde{\boldsymbol{\Omega}}^{-1}\mathbf{S}_{TT}\widetilde{\boldsymbol{\Omega}}^{-1} \end{pmatrix}.$$

Thus it follows that

$$\frac{1}{2}\,\mathrm{tr}\left(\left.\frac{\partial \boldsymbol{\Sigma}}{\partial \alpha}\right|_{\boldsymbol{\theta}=\widetilde{\boldsymbol{\theta}}} \widetilde{\boldsymbol{\Sigma}}^{-1}\mathbf{Y}_v\mathbf{Y}_v'\widetilde{\boldsymbol{\Sigma}}^{-1}\right) = \frac{1}{4}\,\mathrm{tr}\left(\widehat{\mathbf{M}}\widetilde{\boldsymbol{\Omega}}^{-1}\sum_{t=2}^{T}\sum_{s=2}^{T}\mathbf{S}_{t,s}\widetilde{\boldsymbol{\Omega}}^{-1}\right) - \frac{1}{2}\,\mathrm{tr}\left(\widehat{\mathbf{M}}\widetilde{\boldsymbol{\Omega}}^{-1}\sum_{t=2}^{T}\mathbf{S}_{t,t}\widetilde{\boldsymbol{\Omega}}^{-1}\right),$$

and, put together,

$$\widetilde{\mathbf{V}}(\alpha) = \frac{1}{4}\operatorname{tr}\left(\widehat{\mathbf{M}}\widetilde{\boldsymbol{\Omega}}^{-1}\mathbf{S}_{00}\widetilde{\boldsymbol{\Omega}}^{-1}\right) - \frac{1}{2}\operatorname{tr}\left(\widehat{\mathbf{M}}\widetilde{\boldsymbol{\Omega}}^{-1}\mathbf{S}_{0}\widetilde{\boldsymbol{\Omega}}^{-1}\right) + \frac{T^*}{4}\operatorname{tr}\left(\widehat{\mathbf{M}}\widetilde{\boldsymbol{\Omega}}^{-1}\right). \tag{A2}$$

Because $\left.\frac{\partial \boldsymbol{\Sigma}}{\partial \rho}\right|_{\boldsymbol{\theta}=\widetilde{\boldsymbol{\theta}}} = \widehat{\sigma}_\varepsilon^2\left(\left.\frac{\partial \boldsymbol{\Psi}}{\partial \rho}\right|_{\boldsymbol{\theta}=\widetilde{\boldsymbol{\theta}}} \otimes \mathbf{I}_N\right)$, where $\left.\frac{\partial \boldsymbol{\Psi}(\rho)}{\partial \rho}\right|_{\boldsymbol{\theta}=\widetilde{\boldsymbol{\theta}}} = \frac{1}{2}\mathbf{1}\mathbf{1}' - \mathbf{I}_{T^*}$, similar derivations will show that (see also Zhou and Solberger, 2012)

$$\widetilde{\mathbf{V}}(\rho) = \left.\frac{\partial l}{\partial \rho}\right|_{\boldsymbol{\theta}=\widetilde{\boldsymbol{\theta}}} = \widehat{\sigma}_\varepsilon^2\left[\frac{1}{4}\operatorname{tr}\left(\widetilde{\boldsymbol{\Omega}}^{-1}\mathbf{S}_{00}\widetilde{\boldsymbol{\Omega}}^{-1}\right) - \frac{1}{2}\operatorname{tr}\left(\widetilde{\boldsymbol{\Omega}}^{-1}\mathbf{S}_{0}\widetilde{\boldsymbol{\Omega}}^{-1}\right) + \frac{T^*}{4}\operatorname{tr}\left(\widetilde{\boldsymbol{\Omega}}^{-1}\right)\right]. \tag{A3}$$

Let the information matrices for Case A, Case B and Case C be partitioned as in (17), (18) and (23) respectively. Using another standard result (see e.g. Harville, 1977) we have that

$$\begin{aligned}
\widetilde{J}_{\alpha\alpha} &= -E\left(\left.\frac{\partial^2 l}{\partial\alpha\partial\alpha}\right|_{\boldsymbol{\theta}=\widetilde{\boldsymbol{\theta}}}\right) = \frac{1}{2}\operatorname{tr}\left(\left.\frac{\partial \boldsymbol{\Sigma}}{\partial \alpha}\right|_{\boldsymbol{\theta}=\widetilde{\boldsymbol{\theta}}}\widetilde{\boldsymbol{\Sigma}}^{-1}\left.\frac{\partial \boldsymbol{\Sigma}}{\partial \alpha}\right|_{\boldsymbol{\theta}=\widetilde{\boldsymbol{\theta}}}\widetilde{\boldsymbol{\Sigma}}^{-1}\right) \\
&= \frac{1}{2}\operatorname{tr}\left[\left(\frac{1}{2}\mathbf{1}\mathbf{1}' - \mathbf{I}_{T^*}\right)^2 \otimes \left(\widehat{\mathbf{M}}\widetilde{\boldsymbol{\Omega}}^{-1}\right)^2\right] \\
&= \frac{T^{*2}}{8}\operatorname{tr}\left(\widehat{\mathbf{M}}\widetilde{\boldsymbol{\Omega}}^{-1}\widehat{\mathbf{M}}\widetilde{\boldsymbol{\Omega}}^{-1}\right).
\end{aligned} \tag{A4}$$

Analogous derivations using the relation (A4) will show that

$$\widetilde{J}_{\alpha\rho} = \widetilde{J}_{\rho\alpha} = \frac{T^{*2}\widehat{\sigma}_\varepsilon^2}{8}\operatorname{tr}\left(\widehat{\mathbf{M}}\widetilde{\boldsymbol{\Omega}}^{-1}\widetilde{\boldsymbol{\Omega}}^{-1}\right), \tag{A5}$$

$$\widetilde{J}_{\rho\rho} = \frac{T^{*2}\widehat{\sigma}_\varepsilon^4}{8}\operatorname{tr}\left(\widetilde{\boldsymbol{\Omega}}^{-1}\widetilde{\boldsymbol{\Omega}}^{-1}\right), \tag{A6}$$

and, because $\left.\frac{\partial \boldsymbol{\Sigma}}{\partial \sigma^2}\right|_{\boldsymbol{\theta}=\widetilde{\boldsymbol{\theta}}} = \mathbf{I}_{NT^*}$, we have that

$$\widetilde{J}_{\sigma_\varepsilon^2\sigma_\varepsilon^2} = \frac{T^*}{2}\operatorname{tr}\left(\widetilde{\boldsymbol{\Omega}}^{-1}\widetilde{\boldsymbol{\Omega}}^{-1}\right), \tag{A7}$$

$$\widetilde{J}_{\alpha\sigma_\varepsilon^2} = \widetilde{J}_{\sigma_\varepsilon^2\alpha} = -\frac{T^*}{4}\operatorname{tr}\left(\widetilde{\boldsymbol{\Omega}}^{-1}\widehat{\mathbf{M}}\widetilde{\boldsymbol{\Omega}}^{-1}\right), \tag{A8}$$

$$\widetilde{J}_{\rho\sigma_\varepsilon^2} = \widetilde{J}_{\sigma_\varepsilon^2\rho} = -\frac{\widehat{\sigma}_\varepsilon^2 T^*}{4}\operatorname{tr}\left(\widetilde{\boldsymbol{\Omega}}^{-1}\widetilde{\boldsymbol{\Omega}}^{-1}\right). \tag{A9}$$

For the remaining parts of the information matrix, let $\lambda_{i,j}$ ($1 \le i \le N$, $1 \le j \le r$) denote the ith parameter in $\boldsymbol{\lambda}_j = (\lambda_{1,j}, \lambda_{2,j}, \ldots, \lambda_{N,j})'$, the jth column vector of $\boldsymbol{\Lambda} = (\boldsymbol{\lambda}_1, \boldsymbol{\lambda}_2, \ldots, \boldsymbol{\lambda}_r)$, and let $\boldsymbol{\Gamma}_{i,j} = \frac{\partial \boldsymbol{\Lambda}\boldsymbol{\Lambda}'}{\partial \lambda_{i,j}} = \boldsymbol{\lambda}_j \mathbf{e}_i' + \mathbf{e}_i \boldsymbol{\lambda}_j'$, where $\mathbf{e}_i$ is an $N \times 1$ vector for which the ith element equals 1 and all other elements equal 0. Then $\frac{\partial \boldsymbol{\Sigma}}{\partial \lambda_{i,j}} = (\boldsymbol{\Psi} \otimes \boldsymbol{\Gamma}_{i,j})$, where, because $\boldsymbol{\Psi}(1) = \mathbf{I}_{T^*}$, we have that $\left.\frac{\partial \boldsymbol{\Sigma}}{\partial \lambda_{i,j}}\right|_{\boldsymbol{\theta}=\widetilde{\boldsymbol{\theta}}} = (\mathbf{I}_{T^*} \otimes \widetilde{\boldsymbol{\Gamma}}_{i,j})$. It then follows, using analogously the relation (A4), that for $1 \le i,l \le N$ and $1 \le j,q \le r$,

$$\widetilde{J}_{\lambda_{i,j}\lambda_{l,q}} = \widetilde{J}_{\lambda_{l,q}\lambda_{i,j}} = \frac{T^*}{2}\operatorname{tr}\left(\widetilde{\boldsymbol{\Gamma}}_{i,j}\widetilde{\boldsymbol{\Omega}}^{-1}\widetilde{\boldsymbol{\Gamma}}_{l,q}\widetilde{\boldsymbol{\Omega}}^{-1}\right), \tag{A10}$$

$$\widetilde{J}_{\lambda_{i,j}\sigma_\varepsilon^2} = \widetilde{J}_{\sigma_\varepsilon^2\lambda_{i,j}} = \frac{T^*}{2}\operatorname{tr}\left(\widetilde{\boldsymbol{\Gamma}}_{i,j}\widetilde{\boldsymbol{\Omega}}^{-1}\widetilde{\boldsymbol{\Omega}}^{-1}\right), \tag{A11}$$

$$\widetilde{J}_{\lambda_{i,j}\alpha} = \widetilde{J}_{\alpha\lambda_{i,j}} = -\frac{T^*}{4}\operatorname{tr}\left(\widetilde{\boldsymbol{\Gamma}}_{i,j}\widetilde{\boldsymbol{\Omega}}^{-1}\widehat{\mathbf{M}}\widetilde{\boldsymbol{\Omega}}^{-1}\right), \tag{A12}$$

$$\widetilde{J}_{\lambda_{i,j}\rho} = \widetilde{J}_{\rho\lambda_{i,j}} = -\frac{T^*\sigma_\varepsilon^2}{4}\operatorname{tr}\left(\widetilde{\boldsymbol{\Gamma}}_{i,j}\widetilde{\boldsymbol{\Omega}}^{-1}\widetilde{\boldsymbol{\Omega}}^{-1}\right). \tag{A13}$$

Appendix B Proofs

Before we go in to the proofs we present some general results. Under Assumptions 1-7, and under the null hypothesis, the maximum likelihood estimators are consistent for fixed N, implying, as $T \to \infty$,

$$\widehat{\mathbf{A}}\widehat{\mathbf{H}}\widehat{\mathbf{A}}' \xrightarrow{p} \mathbf{\Lambda}\mathbf{\Lambda}', \tag{B1}$$

$$\widehat{\sigma}_\varepsilon^2 \xrightarrow{p} \sigma_\varepsilon^2, \tag{B2}$$

$$\mathbf{S}_{01} \xrightarrow{p} \mathbf{\Omega}. \tag{B3}$$

Also, because under the null hypothesis $E(\mathbf{S}_{tt}) = \mathbf{\Omega}$ for all t, we have that, by the weak law of large numbers,

$$\frac{1}{T^*}\sum_{t=2}^{T} \mathbf{S}_{t,t} = \frac{1}{T^*}\mathbf{S}_0 \xrightarrow{p} \mathbf{\Omega}, \text{ as } T \to \infty, \tag{B4}$$

and by the central limit theorem

$$\frac{1}{\sqrt{T^*}}\mathbf{\Omega}^{-\frac{1}{2}}\left(\sum_{t=2}^{T}\mathbf{y}_t - \mathbf{0}\right) \xrightarrow{d} \mathbf{Z} \sim \mathcal{N}_N(\mathbf{0}, \mathbf{I}), \text{ as } T \to \infty. \tag{B5}$$

Further, let $\psi_i = \varphi_i(\mathbf{\Omega})$ denote the eigenvalues of $\mathbf{\Omega}$, where $\psi_1 \geq \psi_2 \geq \cdots \geq \psi_N \geq 0$, and let $\eta_i = \varphi_i(\mathbf{\Lambda}'\mathbf{\Lambda})$ denote the non-zero eigenvalues of $\mathbf{\Lambda}\mathbf{\Lambda}'$, where $\eta_1 \geq \eta_2 \geq \cdots \geq \eta_r > 0$. Then the eigenvalues of $\mathbf{\Omega}$ are $\psi_1 = \eta_1 + \sigma_\varepsilon^2, \ldots, \psi_r = \eta_r + \sigma_\varepsilon^2, \psi_{r+1} = \cdots = \psi_N = \sigma_\varepsilon^2$, so that

$$\text{tr}\left(\mathbf{\Omega}^{-k}\right) = \sum_{i=1}^{N}\frac{1}{\psi_i^k} = \frac{1}{\sigma_\varepsilon^{2k}}\sum_{i=1}^{N} w_i^k, \tag{B6}$$

where $w_i = \begin{cases} \frac{\sigma_\varepsilon^2}{\eta_i + \sigma_\varepsilon^2} & \text{for } 1 \leq i \leq r \\ 1 & \text{for } r < i \leq N \end{cases}$ was defined in Section 3.1. It then follows that

$$\text{tr}\left(\mathbf{\Lambda}\mathbf{\Lambda}'\mathbf{\Omega}^{-1}\right) = \text{tr}\left[\left(\mathbf{\Omega} - \sigma_\varepsilon^2\mathbf{I}_N\right)\mathbf{\Omega}^{-1}\right] = \text{tr}\left(\mathbf{I}_N - \sigma_\varepsilon^2\mathbf{\Omega}^{-1}\right) = \sum_{i=1}^{N}(1 - w_i) = \sum_{j=1}^{r}\overline{w}_j, \tag{B7}$$

where $\overline{w}_j = \frac{\eta_j}{\eta_j + \sigma_\varepsilon^2}$ was defined in Section 3.2. Likewise we have that

$$\text{tr}\left(\mathbf{\Lambda}\mathbf{\Lambda}'\mathbf{\Omega}^{-1}\mathbf{\Lambda}\mathbf{\Lambda}'\mathbf{\Omega}^{-1}\right) = \text{tr}\left[\left(\mathbf{I}_N - \sigma_\varepsilon^2\mathbf{\Omega}^{-1}\right)\left(\mathbf{I}_N - \sigma_\varepsilon^2\mathbf{\Omega}^{-1}\right)\right]$$
$$= \text{tr}\left[\mathbf{I}_N - 2\sigma_\varepsilon^2\mathbf{\Omega}^{-1} + \sigma_\varepsilon^4\mathbf{\Omega}^{-1}\mathbf{\Omega}^{-1}\right] = \sum_{i=1}^{N}(1 - w_i)^2 = \sum_{j=1}^{r}\overline{w}_j^2. \tag{B8}$$

Also, Assumption 3 implies that the eigenvalues of $\mathbf{\Lambda}'\mathbf{\Lambda}$ (likewise the non-zero eigenvalues of $\mathbf{\Lambda}\mathbf{\Lambda}'$) are $\mathcal{O}(N)$. Hence, as $N \to \infty$ the smallest non-zero eigenvalue of $\mathbf{\Lambda}\mathbf{\Lambda}'$ diverges, $\eta_r = \min(\eta_j) \to \infty$, such that for all $k \in \mathbb{R}$, $\lim_{\eta_r \to \infty}\left(\frac{\eta_i}{\sigma_\varepsilon^2 + \eta_i}\right)^k = 1$, and for all $k > 0$, $\lim_{\eta_r \to \infty}\left(\frac{\sigma_\varepsilon^2}{\sigma_\varepsilon^2 + \eta_i}\right)^k = 0$, implying for $j = 1, 2, \ldots, r$ and $i = 1, 2, \ldots, N$ that

$$\{\forall j | \overline{w}_j \to 1\} \text{ as } N \to \infty, \tag{B9}$$

$$\{\forall i \leq r | w_i \to 0\} \text{ as } N \to \infty. \tag{B10}$$

Because for all $i > r$, $w_i = 1$, it follows that for all $k \geq 0$,

$$\sum_{i=1}^{N} w_i^k = \mathcal{O}(N) \Rightarrow \text{tr}\left(\mathbf{\Omega}^{-k}\right) = \mathcal{O}(N). \tag{B11}$$

21

Proof of Lemma 1. We will make repeated use of the Cauchy-Schwarz trace-inequality; $\text{tr}(\mathbf{AB}) \leq ||\mathbf{A}|| \cdot ||\mathbf{B}||$ for $\mathbf{A}, \mathbf{B} \in \mathbb{R}^{n \times n}$.

(i) Because $\mathbf{\Gamma}_{i,j} = \boldsymbol{\lambda}_j \mathbf{e}_i' + \mathbf{e}_i \boldsymbol{\lambda}_j'$ is symmetric we have that $\text{tr}(\mathbf{\Gamma}_{i,j}^2) = 2\lambda_{i,j}^2 + 2\sum_{i=1}^N \lambda_{i,j}^2 = \mathcal{O}(N)$. Using this result, the Cauchy-Schwarz trace-inequality, and the results (A10) and (B11), we have that, for $i, l = 1, 2, \ldots, N$ and $j, q = 1, 2, \ldots, r$,

$$\frac{J_{\lambda_{i,j}\lambda_{l,q}}}{N^2 T^*} = \frac{1}{2N^2} \text{tr}\left(\mathbf{\Gamma}_{i,j}\mathbf{\Omega}^{-1}\mathbf{\Gamma}_{l,q}\mathbf{\Omega}^{-1}\right) \leq \frac{1}{2N^2} ||\mathbf{\Gamma}_{i,j}\mathbf{\Omega}^{-1}|| \cdot ||\mathbf{\Gamma}_{l,q}\mathbf{\Omega}^{-1}||$$

$$\leq \frac{1}{2N^2} ||\mathbf{\Gamma}_{i,j}|| \cdot ||\mathbf{\Gamma}_{l,q}|| \cdot ||\mathbf{\Omega}^{-1}|| \cdot ||\mathbf{\Omega}^{-1}||$$

$$= \frac{1}{2N^2} \sqrt{\text{tr}\left(\mathbf{\Gamma}_{i,j}^2\right)} \sqrt{\text{tr}\left(\mathbf{\Gamma}_{l,q}^2\right)} \text{tr}\left(\mathbf{\Omega}^{-2}\right) = \mathcal{O}(1).$$

(ii) Similar as in (i); using the result (A11), for $i = 1, 2, \ldots, N$ and $j = 1, 2, \ldots, r$,

$$\frac{J_{\lambda_{i,j}\sigma_\varepsilon^2}}{NT^*} = \frac{1}{2N} \text{tr}\left(\mathbf{\Gamma}_{i,j}\mathbf{\Omega}^{-1}\mathbf{\Omega}^{-1}\right) \leq \frac{1}{2N} ||\mathbf{\Gamma}_{i,j}|| \cdot ||\mathbf{\Omega}^{-2}|| = \frac{1}{2N} \sqrt{\text{tr}(\mathbf{\Gamma}_{i,j}^2)} \sqrt{\text{tr}\left(\mathbf{\Omega}^{-4}\right)} = \mathcal{O}(1).$$

(iii) It follows from (i) and (ii) that $\text{tr}\left(\mathbf{\Gamma}_{i,j}\mathbf{\Omega}^{-1}\right)$ and $\text{tr}\left(\mathbf{\Gamma}_{i,j}\mathbf{\Omega}^{-1}\mathbf{\Omega}^{-1}\right)$ are both $\mathcal{O}(N)$. Thus, using (A12), we have that, for $i = 1, 2, \ldots, N$ and $j = 1, 2, \ldots, r$,

$$\left|\frac{J_{\lambda_{i,j}\alpha}}{NT^*}\right| = \frac{1}{4N} \text{tr}\left(\mathbf{\Lambda}\mathbf{\Lambda}'\mathbf{\Omega}^{-1}\mathbf{\Gamma}_{i,j}\mathbf{\Omega}^{-1}\right) = \frac{1}{4N} \text{tr}\left[\left(\mathbf{\Lambda}\mathbf{\Lambda}' + \sigma_\varepsilon^2 \mathbf{I}_N\right)\mathbf{\Omega}^{-1}\mathbf{\Gamma}_{i,j}\mathbf{\Omega}^{-1} - \sigma_\varepsilon^2 \mathbf{\Omega}^{-1}\mathbf{\Gamma}_{i,j}\mathbf{\Omega}^{-1}\right]$$

$$= \frac{1}{4N} \text{tr}\left(\mathbf{\Gamma}_{i,j}\mathbf{\Omega}^{-1} - \sigma_\varepsilon^2 \mathbf{\Omega}^{-1}\mathbf{\Gamma}_{i,j}\mathbf{\Omega}^{-1}\right) = \mathcal{O}(1).$$

(iv) Similar as in (ii); using the result (A13), for $i = 1, 2, \ldots, N$ and $j = 1, 2, \ldots, r$,

$$\frac{J_{\lambda_{i,j}\rho}}{NT^*} = \frac{\sigma_\varepsilon^2}{4N} \text{tr}\left(\mathbf{\Gamma}_{i,j}\mathbf{\Omega}^{-1}\mathbf{\Omega}^{-1}\right) = \mathcal{O}(1).$$

(v) Using (A7) and (B11), $\frac{J_{\sigma_\varepsilon^2 \sigma_\varepsilon^2}}{NT^*} = \frac{1}{2N} \text{tr}\left(\mathbf{\Omega}^{-1}\mathbf{\Omega}^{-1}\right) = \mathcal{O}(1)$.

(vi) Using (A8) and (B6),

$$\left|\frac{J_{\sigma_\varepsilon^2 \alpha}}{T^*}\right| = \frac{1}{4} \text{tr}(\mathbf{\Omega}^{-1}\mathbf{\Lambda}\mathbf{\Lambda}'\mathbf{\Omega}^{-1}) = \frac{1}{4} \text{tr}\left[\mathbf{\Omega}^{-1}\left(\mathbf{\Lambda}\mathbf{\Lambda}' + \sigma_\varepsilon^2 \mathbf{I}_N\right)\mathbf{\Omega}^{-1} - \sigma_\varepsilon^2 \mathbf{\Omega}^{-1}\mathbf{\Omega}^{-1}\right]$$

$$= \frac{1}{4}\left[\text{tr}\left(\mathbf{\Omega}^{-1}\right) - \sigma_\varepsilon^2 \text{tr}\left(\mathbf{\Omega}^{-1}\mathbf{\Omega}^{-1}\right)\right] = \frac{1}{4\sigma_\varepsilon^2}\left(\sum_{i=1}^N w_i - \sum_{i=1}^N w_i^2\right) = \mathcal{O}(1).$$

(vii) Using (A9) and (B11), $\left|\frac{J_{\sigma_\varepsilon^2 \rho}}{NT^*}\right| = \frac{\sigma_\varepsilon^2}{4N} \text{tr}\left(\mathbf{\Omega}^{-1}\mathbf{\Omega}^{-1}\right) = \mathcal{O}(1)$.

(viii) Using (A4) and (B8), $\frac{J_{\alpha\alpha}}{T^{*2}} = \frac{1}{8}\sum_{j=1}^r \overline{w}_j^2 \leq \frac{r}{8}$, where, using (B9), $\lim_{N\to\infty} \frac{1}{8}\sum_{j=1}^r \overline{w}_j^2 = \frac{r}{8}$.

(ix) Similar as for (vi); using (A5), $\frac{J_{\alpha\rho}}{T^{*2}} = \frac{\sigma_\varepsilon^2}{8} \text{tr}\left(\mathbf{\Omega}^{-1}\mathbf{\Lambda}\mathbf{\Lambda}'\mathbf{\Omega}^{-1}\right) = \mathcal{O}(1)$.

(x) Using (A6) and (B6), $\frac{J_{\rho\rho}}{NT^{*2}} = \frac{1}{8N}\sum_{i=1}^N w_i^2 \leq \frac{1}{8}$, where, using (B10), $\lim_{N\to\infty} \frac{1}{8N}\sum_{i=1}^N w_i^2 = \frac{1}{8}$.

Replacing the elements J with their constrained counterparts $\widetilde{J} = J|_{\boldsymbol{\theta}=\widetilde{\boldsymbol{\theta}}}$ will shift $\mathcal{O}$ into $\mathcal{O}_p$. This completes the proof of Lemma 1. ∎

Proof of Theorem 1. Decompose the statistic (19) as $\vartheta_b^* = \left(\mathcal{I}_1 + \mathcal{I}_2 + \mathcal{I}_3\right) / \sqrt{\mathcal{I}_4}$, where

$$\mathcal{I}_1 = \frac{1}{T^*} \operatorname{tr}\left(\widehat{\mathbf{M}}\mathbf{S}_{01}^{-1}\mathbf{S}_{00}\mathbf{S}_{01}^{-1}\right), \; \mathcal{I}_2 = -\frac{1}{T^*}2\operatorname{tr}\left(\widehat{\mathbf{M}}\mathbf{S}_{01}^{-1}\mathbf{S}_0\mathbf{S}_{01}^{-1}\right),$$

$$\mathcal{I}_3 = \operatorname{tr}\left(\widehat{\mathbf{M}}\mathbf{S}_{01}^{-1}\right), \; \mathcal{I}_4 = 2\operatorname{tr}\left(\widehat{\mathbf{M}}\mathbf{S}_{01}^{-1}\widehat{\mathbf{M}}\mathbf{S}_{01}^{-1}\right).$$

Let $\boldsymbol{\Omega}$ have spectral decomposition $\boldsymbol{\Omega} = \mathbf{Q}\boldsymbol{\Xi}\mathbf{Q}'$, where the eigenvalues $\boldsymbol{\Xi} = \operatorname{diag}\left[\psi_1, \psi_2, \dots, \psi_N\right]$ have the properties leading up to (B6). Then $\mathbf{Z}_2 = \mathbf{Q}'\mathbf{Z}_1$ is an orthogonal transformation such that $\mathbf{Z}_2 \sim \mathbf{Z}_1$. Using this and the results (B1 - B6), we have that, as $T \to \infty$,

$$\mathcal{I}_1 = \operatorname{tr}\left[\widehat{\mathbf{M}}\mathbf{S}_{01}^{-1/2}\mathbf{S}_{01}^{-1/2}\frac{1}{\sqrt{T^*}}\left(\sum_{t=2}^{T}\mathbf{y}_t\right)\frac{1}{\sqrt{T^*}}\left(\sum_{t=2}^{T}\mathbf{y}_t\right)'\mathbf{S}_{01}^{-1/2}\mathbf{S}_{01}^{-1/2}\right]$$

$$\xrightarrow{d} \operatorname{tr}\left(\boldsymbol{\Lambda}\boldsymbol{\Lambda}'\boldsymbol{\Omega}^{-1/2}\mathbf{Z}\mathbf{Z}'\boldsymbol{\Omega}^{-1/2}\right)$$

$$= \operatorname{tr}\left[\left(\boldsymbol{\Lambda}\boldsymbol{\Lambda}' + \sigma_\varepsilon^2\mathbf{I}_N\right)\boldsymbol{\Omega}^{-1/2}\mathbf{Z}\mathbf{Z}'\boldsymbol{\Omega}^{-1/2} - \sigma_\varepsilon^2\boldsymbol{\Omega}^{-1/2}\mathbf{Z}\mathbf{Z}'\boldsymbol{\Omega}^{-1/2}\right]$$

$$\sim \operatorname{tr}\left(\mathbf{Z}\mathbf{Z}'\right) - \sigma_\varepsilon^2\operatorname{tr}\left(\mathbf{Z}'\boldsymbol{\Xi}^{-1}\mathbf{Z}\right)$$

$$= \left(\sum_{i=1}^{N}Z_i^2 - \sum_{i=1}^{N}\frac{\sigma_\varepsilon^2}{\psi_i}Z_i^2\right) = \sum_{j=1}^{r}\left(1 - w_j\right)Z_j^2 = \sum_{j=1}^{r}\overline{w}_jZ_j^2 \equiv \sum_{j=1}^{r}\overline{w}_j\chi_{1,j}^2.$$

Further, as $T \to \infty$ we have that, using (B1 - B4) and (B7),

$$\mathcal{I}_2 + \mathcal{I}_3 = -2\operatorname{tr}\left(\widehat{\mathbf{M}}\mathbf{S}_{01}^{-1}\frac{1}{T^*}\mathbf{S}_0\mathbf{S}_{01}^{-1}\right) + \operatorname{tr}\left(\widehat{\mathbf{M}}\mathbf{S}_{01}^{-1}\right) \xrightarrow{p} -\operatorname{tr}\left(\boldsymbol{\Lambda}\boldsymbol{\Lambda}'\boldsymbol{\Omega}^{-1}\right) = -\sum_{j=1}^{r}\overline{w}_j.$$

Finally, for the denominator we have that, as $T \to \infty$, using (B1 - B4) and (B8),

$$\mathcal{I}_4 = 2\operatorname{tr}\left(\widehat{\mathbf{M}}\mathbf{S}_{01}^{-1}\widehat{\mathbf{M}}\mathbf{S}_{01}^{-1}\right) \xrightarrow{p} 2\operatorname{tr}\left(\boldsymbol{\Lambda}\boldsymbol{\Lambda}'\boldsymbol{\Omega}^{-1}\boldsymbol{\Lambda}\boldsymbol{\Lambda}'\boldsymbol{\Omega}^{-1}\right) = \sum_{j=1}^{r}\overline{w}_j^2,$$

where from the continuous mapping theorem, $\mathcal{I}_4^{-1/2} \xrightarrow{p} \left(\sum_{j=1}^{r}\overline{w}_j^2\right)^{-1/2}$. Put together and using Slutsky's theorem the result (20) immediately follows, and using (B9) the result (21) follows;

$$\vartheta_b^* \xrightarrow{d} \frac{1}{\sqrt{2r}}\left(\sum_{j=1}^{r}\chi_{1,j}^2 - r\right) = \frac{1}{\sqrt{2r}}\left(\chi_r^2 - r\right), \text{ as } (T, N)_s \to \infty.$$

This completes the proof of Theorem 1. ∎

Proof of Proposition 1. Let for $1 \leq j \leq r$, $Z_j \sim \mathcal{N}(0,1)$, $\nu_1 = \sum_{j=1}^{r}\overline{w}_j$ and $\nu_2 = \sum_{j=1}^{r}\overline{w}_j^2$, and rewrite the weighted sum of χ_1^2-variables in (20) as $\sum_{j=1}^{r}\overline{w}_jZ_j^2 = \nu_1\sum_{j=1}^{r}a_jZ_j^2$, where $a_j = \frac{\overline{w}_j}{\nu_1}$ such that $\sum_{j=1}^{r}a_j = 1$. We seek an u such that $\sum_{j=1}^{r}a_jZ_j^2$ and $u^{-1}\chi_u^2$ have equivalent first and second moments. For the first moments we have that $\sum_{j=1}^{r}a_jE(Z_j^2) = u^{-1}E(\chi_u^2) = 1$ for any u. For the second moments we have that $E(\sum_{j=1}^{r}a_jZ_j^2)^2 = 2\sum_{j=1}^{r}a_j^2 + 1$, and $E(\chi_u^2/u)^2 = 2/u + 1$. Hence, for both first and second moments to be equal we need that

$$u = \left(\sum_{j=1}^{r}a_j^2\right)^{-1} = \nu_1^2\left(\sum_{j=1}^{r}\overline{w}_j^2\right)^{-1} = \frac{\nu_1^2}{\nu_2} = \frac{[\operatorname{tr}(\boldsymbol{\Lambda}\boldsymbol{\Lambda}'\boldsymbol{\Omega}^{-1})]^2}{\operatorname{tr}(\boldsymbol{\Lambda}\boldsymbol{\Lambda}'\boldsymbol{\Omega}^{-1}\boldsymbol{\Lambda}\boldsymbol{\Lambda}'\boldsymbol{\Omega}^{-1})},$$

where we have used (B7-B8). Thus, because $\sum_{j=1}^{r} \overline{w}_j \chi_{1,j}^2 = \nu_1 \sum_{j=1}^{r} a_j \chi_{1,j}^2 \overset{app.}{\sim} \nu_1 \chi_u^2 / u$,

$$\vartheta_b^* \overset{app.}{\sim} \frac{\nu_1 \left(\frac{\chi_u^2}{u} - 1 \right)}{\sqrt{2\nu_2}} = \frac{\sqrt{u}}{\sqrt{2}} \left(\frac{\chi_u^2}{u} - 1 \right) = \frac{1}{\sqrt{2u}} \left(\chi_u^2 - u \right),$$

as claimed in Proposition 1. ∎

Proof of Theorem 2. Let $F = \text{diag}\left(T^*, T^* \sqrt{N} \right)$. The statistic ϑ_c^* may be decomposed as

$$\vartheta_c^* = \left(\begin{array}{cc} \widetilde{\mathbf{V}}(\alpha) & \widetilde{\mathbf{V}}(\rho) \end{array} \right) F^{-1} \left[F^{-1} \left(\begin{array}{cc} \widetilde{J}_{\alpha\alpha} & \widetilde{J}_{\alpha\rho} \\ \widetilde{J}_{\alpha\rho} & \widetilde{J}_{\rho\rho} \end{array} \right) F^{-1} \right]^{-1} F^{-1} \mathbf{V}^*$$

$$= \left(\begin{array}{cc} \frac{\widetilde{\mathbf{V}}(\alpha)}{T} & \frac{\widetilde{\mathbf{V}}(\rho)}{N^{1/2}T} \end{array} \right) \left(\begin{array}{cc} \frac{\widetilde{J}_{\alpha\alpha}}{T^2} & \frac{\widetilde{J}_{\alpha\rho}}{N^{1/2}T^2} \\ \frac{\widetilde{J}_{\rho\alpha}}{N^{1/2}T^2} & \frac{\widetilde{J}_{\rho\rho}}{NT^2} \end{array} \right)^{-1} \left(\begin{array}{c} \frac{1}{2} \\ \frac{1}{2\sqrt{2}} \end{array} \right),$$

where it follows from the proof of Lemma 1 (viii), (ix) and (x) respectively that, as $(T,N)_s \to \infty$, $\frac{1}{T^{*2}} \widetilde{J}_{\alpha\alpha} \overset{p}{\to} \frac{r}{8}$, $\frac{1}{\sqrt{N}T^{*2}} \widetilde{J}_{\alpha\rho} \overset{p}{\to} 0$ and $\frac{1}{NT^{*2}} \widetilde{J}_{\rho\rho} \overset{p}{\to} \frac{1}{8}$. Let as before Ω have spectral decomposition $\Omega = \mathbf{Q} \Xi \mathbf{Q}'$, where $\Xi = \text{diag}[\psi_1, \psi_2, \ldots, \psi_N]$. The scores w.r.t. α and ρ are given by (A2) and (A3) respectively. Analogous derivations to the proof of Theorem 1 yields, as $T \to \infty$,

$$\frac{1}{T^*} \widetilde{\mathbf{V}}(\alpha) = \frac{1}{4} \text{tr} \left[\widehat{\mathbf{M}} \mathbf{S}_{01}^{-1} \frac{1}{\sqrt{T^*}} \left(\sum_{t=2}^{T} \mathbf{y}_t \right) \frac{1}{\sqrt{T^*}} \left(\sum_{t=2}^{T} \mathbf{y}_t \right)' \mathbf{S}_{01}^{-1} \right] - \frac{1}{2} \text{tr} \left(\widehat{\mathbf{M}} \mathbf{S}_{01}^{-1} \frac{1}{T^*} \mathbf{S}_0 \mathbf{S}_{01}^{-1} \right) + \frac{1}{4} \text{tr} \left(\widehat{\mathbf{M}} \mathbf{S}_{01}^{-1} \right)$$

$$\overset{d}{\to} \frac{1}{4} \text{tr} \left(\Omega^{-1/2} \Lambda \Lambda' \Omega^{-1/2} \mathbf{Z} \mathbf{Z}' \right) - \frac{1}{4} \text{tr} \left(\Lambda \Lambda' \Omega^{-1} \right)$$

$$\sim \frac{1}{4} \text{tr} \left(\mathbf{Z} \mathbf{Z}' \right) - \frac{\sigma_\varepsilon^2}{4} \text{tr} \left(\mathbf{Z}' \Xi^{-1} \mathbf{Z} \right) - \frac{1}{4} \text{tr}(\Lambda \Lambda' \Omega^{-1}) = \frac{1}{4} \left(\sum_{j=1}^{r} \overline{w}_j Z_j^2 - \sum_{j=1}^{r} \overline{w}_j \right) = \mathcal{W}_1,$$

and

$$\frac{1}{T^* \sqrt{N}} \widetilde{\mathbf{V}}(\rho) = \frac{\widehat{\sigma}_\varepsilon^2}{\sqrt{N}} \left[\frac{1}{4} \text{tr} \left(\widetilde{\Omega}^{-1} \frac{1}{T^*} \mathbf{S}_{00} \widetilde{\Omega}^{-1} \right) - \frac{1}{2} \text{tr} \left(\widetilde{\Omega}^{-1} \frac{1}{T^*} \mathbf{S}_0 \widetilde{\Omega}^{-1} \right) + \frac{1}{4} \text{tr} \left(\widetilde{\Omega}^{-1} \right) \right]$$

$$\overset{d}{\to} \frac{\sigma_\varepsilon^2}{\sqrt{N}} \left[\frac{1}{4} \text{tr} \left(\mathbf{Z}' \Omega^{-1} \mathbf{Z} \right) - \frac{1}{4} \text{tr} \left(\Omega^{-1} \right) \right]$$

$$\sim \frac{1}{4\sqrt{N}} \left[\sigma_\varepsilon^2 \text{tr} \left(\mathbf{Z}' \Xi^{-1} \mathbf{Z} \right) - \sum_{i=1}^{N} w_i \right] = \frac{1}{4\sqrt{N}} \left(\sum_{i=1}^{N} w_i Z_i^2 - \sum_{i=1}^{N} w_i \right) = \mathcal{W}_2.$$

Using (B9) it follows that $\mathcal{W}_1 \overset{d}{\to} \frac{1}{4}(\sum_{j=1}^{r} Z_j^2 - r) \equiv \frac{1}{4}(\chi_r^2 - r)$ as $N \to \infty$, and using (B10) it follows that $\mathcal{W}_2 = \frac{1}{4\sqrt{N}}(\sum_{i=r+1}^{N} w_i Z_i^2 - \sum_{i=r+1}^{N} w_i) + o(1) = \frac{1}{4\sqrt{N}}[\chi_{(N-r)}^2 - (N-r)] + o(1)$. Because $\frac{4\sqrt{N}}{\sqrt{2(N-r)}} \mathcal{W}_2 = \frac{1}{\sqrt{2(N-r)}}[\chi_{(N-r)}^2 - (N-r)] + o(1) \overset{d}{\to} \mathcal{N}(0,1)$ as $N \to \infty$ by the central limit theorem, it follows that $\mathcal{W}_2 \to \frac{1}{2\sqrt{2}} \mathcal{N}(0,1)$ as $N \to \infty$. Since $\mathcal{W}_1$ is strictly a function of $Z_1, Z_2, \ldots, Z_r$, and $\mathcal{W}_2$ is asymptotically only a function of $Z_{r+1}, Z_{r+2}, \ldots, Z_N$, $\frac{1}{T^*} \widetilde{\mathbf{V}}(\alpha)$ and $\frac{1}{T^* \sqrt{N}} \widetilde{\mathbf{V}}(\rho)$ are asymptotically independent as $(T,N)_s \to \infty$.

Put together, we have that

$$\vartheta_c^* \xrightarrow{d} \frac{1}{4}\left(\chi_r^2 - r\right)\frac{8}{r}\frac{1}{2} + \frac{1}{2\sqrt{2}}\mathcal{N}(0,1)\frac{8}{2\sqrt{2}} = \frac{\chi_r^2}{r} - 1 + \mathcal{N}(0,1),$$

as claimed in Theorem 2. ∎

The **Proof of Theorem 3** is analogous to the proof of Theorem 2. ∎

References

BAI, J. (2003): "Inferential theory for factor models of large dimensions," *Econometrica*, 71, 135–171.

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

BAI, J., AND K. LI (2012a): "Maximum likelihood estimation and inference for approximate factor models of high dimension," MPRA Paper NO. 42118.

――― (2012b): "Statistical analysis of factor models of high dimension," *The Annals of Statistics*, 40, 436–465.

BAI, J., AND S. NG (2002): "Determining the number of factors in approximate factor models," *Econometrica*, 70, 191–221.

――― (2004): "A PANIC attack on unit roots and cointegration," *Econometrica*, 72, 1127–1177.

BREITUNG, J., AND S. DAS (2005): "Panel unit root tests under cross-sectional dependence," *Statistica Neerlandica*, 59, 414–433.

――― (2008): "Testing for unit roots in panels with a factor structure," *Econometric Theory*, 24, 88–108.

BREITUNG, J., AND S. EICKMEIER (2006): "Dynamic factor models," *Allgemeines Statistisches Archiv*, 90, 27–42.

CHAMBERLAIN, G., AND M. ROTHSCHILD (1983): "Arbitrage, factor structure, and mean-variance analysis on large asset markets," *Econometrica*, 51, 1281–1304.

DOZ, C., D. GIANNONE, AND L. REICHLIN (2012): "A quasi-maximum likelihood approach for large, approximate dynamic factor models," *The Review of Economics and Statistics*, 94, 1014–1024.

ENGLE, R. F., AND M. W. WATSON (1981): "A one-factor multivariate time series model of metropolitan wage rates," *Journal of the American Statistical Association*, 76, 774–781.

HARTLEY, H. O., AND J. N. K. RAO (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.

LIU, H., Y. TANG, AND H. H. ZHANG (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., AND H. NEUDECKER (2001): *Matrix Differential Calculus with Applications in Statistics and Econometrics*. Wiley, Chichester.

MOON, H. R., AND B. PERRON (2004): "Testing for a unit root in panels with dynamic factors," *Journal of Econometrics*, 122, 81–126.

PEÑA, D., AND P. PONCELA (2006): "Nonstationary dynamic factor analysis," *Journal of Statistical Planning and Inference*, 136, 1237–1257.

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

PHILLIPS, P. C. B., AND D. SUL (2003): "Dynamic panel estimation and homogeneity testing under cross section dependence," *Econometrics Journal*, 6, 217–259.

SARGENT, T. J., AND C. A. SIMS (1977): "Business cycle modeling without pretending to have too much a priori economic theory," in *New Methods in Business Cycle Research: Proceedings from a Conference*, ed. by C. A. Sims, pp. 45–110. Federal Reserve Bank of Minneapolis.

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

SOLO, V. (1984): "The order of differencing in ARIMA models," *Journal of the American Statistical Association*, 79, 916–921.

STOCK, J. H., AND M. W. WATSON (1988): "Testing for common trends," *Journal of the American Statistical Association*, 83, 1097–1107.

——— (1989): "New indexes of coincident and leading economic indicator," *NBER Macroeconomics Journal*, 4, 351–394.

——— (2002): "Forecasting using principal components from a large number of predictors," *Journal of the American Statistical Association*, 97, 1167–1179.

STOICA, P., AND M. JANSSON (2009): "On maximum likelihood estimation in factor analysis - an algebraic derivation," *Signal Processing*, 89, 1260–1262.

VAN DER LEEUW, J. (1994): "The covariance matrix of ARMA errors in closed form," *Journal of Econometrics*, 63, 397–405.

ZHOU, X., AND M. SOLBERGER (2012): "An LM-type test for idiosyncratic unit roots in the exact factor model with nonstationary common shocks," Department of Statistics, Uppsala University. Mimeo.