

A Lagrange multiplier-type test for idiosyncratic unit roots in the exact factor model under misspecification

Citation for published version (APA):

Zhou, X., & Solberger, M. (2013). *A Lagrange multiplier-type test for idiosyncratic unit roots in the exact factor model under misspecification*. Maastricht University, Graduate School of Business and Economics. GSBE Research Memoranda No. 058 <https://doi.org/10.26481/umagsb.2013058>

Document status and date:

Published: 01/01/2013

DOI:

[10.26481/umagsb.2013058](https://doi.org/10.26481/umagsb.2013058)

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](#)

General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal.

If the publication is distributed under the terms of Article 25fa of the Dutch Copyright Act, indicated by the "Taverne" license above, please follow below link for the End User Agreement:

www.umlib.nl/taverne-license

Take down policy

If you believe that this document breaches copyright please contact us at:

repository@maastrichtuniversity.nl

providing details and we will investigate your claim.

Xingwu Zhou, Martin Solberger

**A Lagrange Multiplier-type
test for idiosyncratic unit
roots in the exact factor
model under misspecification**

RM/13/058

GSBE

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

P.O. Box 616
NL- 6200 MD Maastricht
The Netherlands

A LAGRANGE MULTIPLIER-TYPE TEST FOR IDIOSYNCRATIC UNIT ROOTS IN THE EXACT FACTOR MODEL UNDER MISSPECIFICATION

Xingwu Zhou and Martin Solberger
Uppsala University [¶]

Abstract

We consider an exact factor model and derive a Lagrange multiplier-type test for unit roots in the idiosyncratic components. The asymptotic distribution of the statistic is derived under the misspecification that the differenced factors are white noise. We prove that the asymptotic distribution is independent of the distribution of the factors, and that the factors are allowed to be integrated, cointegrate, or be stationary. In a simulation study, size and power is compared with some popular second generation panel unit root tests. The simulations suggest that our statistic is well-behaved in terms of size and that it is powerful and robust in comparison with existing tests.

JEL: C12, C23

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

1 Introduction

This paper adapts to the growing literature on panel unit-root tests that let the cross-sectional dependence be modeled by a factor representation which is static in the factor loadings, but where the factors themselves are dynamic.¹ Let $x_{i,t}$ be the i th variable that is observed in time period t , where $i = 1, 2, \dots, N$ and $t = 1, 2, \dots, T$, and suppose that it admits a static factor representation (omitting any deterministic terms),

$$x_{i,t} = \lambda_i' \mathbf{f}_t + u_{i,t}, \tag{1}$$

where $u_{i,t}$ is an unobservable dynamic idiosyncratic component and $\lambda_i' \mathbf{f}_t$ is the common component, where $\mathbf{f}_t = (f_{1,t}, f_{2,t}, \dots, f_{r,t})'$ is an $r \times 1$ vector of unobservable dynamic factors and $\lambda_i = (\lambda_{i,1}, \lambda_{i,2}, \dots, \lambda_{i,r})'$ is an $r \times 1$ vector of factor loadings. Thus, only the left side of Equation (1) is observed while all elements on the right side are unobservable. Assuming independence between the factors and the idiosyncratic components, the contemporaneous covariance of the N -variate process $\mathbf{x}_t = (x_{1,t}, x_{2,t}, \dots, x_{N,t})'$ is $\Sigma_{xx} = \text{Var}(\mathbf{x}_t) = \Lambda \Sigma_{ff} \Lambda' + \Sigma_{uu}$,

[¶]Department of Statistics, Uppsala University, Sweden, Box SE-751-20.

E-mail addresses: xingwu.zhou@statistics.uu.se; martin.solberger@statistics.uu.se

¹As opposed to the generalized dynamic factor model of Forni, Hallin, Lippi, and Reichlin (2000) where the factor loadings are lag-polynomials and the common shocks (factors) are white noise.

where $\Sigma_{ff} = \text{Var}(\mathbf{f}_t)$ and $\Sigma_{uu} = \text{Var}(\mathbf{u}_t)$, where \mathbf{u}_t is defined analogously to \mathbf{x}_t . There is a distinction between the *approximate* factor model, as defined by Chamberlain and Rothschild (1983), where the eigenvalues of $\Lambda \Sigma_{ff} \Lambda'$ are $\mathcal{O}(N)$ and the largest eigenvalue of Σ_{uu} is bounded, and the *exact* factor model where Σ_{uu} is restricted to be diagonal. As such, the exact factor model is nested within the approximate factor model.

The usual workhorse for the static factor model is the method of principal components (e.g. Stock and Watson, 2002; Bai and Ng, 2004). Bai (2003) show that, under limited time series correlation in $x_{i,t}$ and limited cross-unit correlation in $u_{i,t}$ admitting an approximate factor model, the common component $\lambda'_i \mathbf{f}_t$ is consistently estimated with principal components at rate $\min(\sqrt{N}, \sqrt{T})$ as N and T tend to infinity jointly, without any additional restrictions on the relationship between N and T . This is used by Bai and Ng (2004) in the popular PANIC (Panel Analysis of Nonstationarity in Idiosyncratic and Common components) procedure to perform unit root tests on the idiosyncratic component, because the idiosyncratic component is consistently estimated as a residual to the panel (1). However, the principal components estimators are suitable for large factor models, and some of the focus has now moved to question efficiency and the size of panels (e.g. Boivin and Ng, 2006). To achieve efficiency the natural choice is to consider likelihood based methods, or likelihood-type methods, as in Choi (2011), Breitung and Tenhofen (2011), Doz, Giannone, and Reichlin (2012), and Bai and Li (2012b,a).

Bai and Li (2012b) consider the exact factor model, and propose to optimize a misspecified likelihood function. The parameter estimators are then quasi-likelihood estimators, which are shown by Bai and Li (2012a) to be consistent also when the exact factor model is misspecified contra the more general approximate factor model. Let $\Phi_t = E(\mathbf{u}_t \mathbf{u}_t')$, which allows for time heteroscedasticity. Also, suppose the factors are fixed constants and let $\mathbf{M}_{ff} = \frac{1}{T-1} \sum_{t=1}^T (\mathbf{f}_t - \bar{\mathbf{f}}_t)(\mathbf{f}_t - \bar{\mathbf{f}}_t)'$, where $\bar{\mathbf{f}}_t = \frac{1}{T} \sum_{t=1}^T \mathbf{f}_t$, and suppose that $\lim_{T \rightarrow \infty} \mathbf{M}_{ff}$ exists and is positive definite. Bai and Li (2012b,a) consider maximizing the objective function

$$l = -\frac{1}{2N} \log |\bar{\Sigma}_{xx}| - \frac{1}{2N} \text{tr}(\bar{\Sigma}_{xx}^{-1} \mathbf{M}_{zz}), \quad (2)$$

where $\bar{\Sigma}_{xx} = \Lambda \mathbf{M}_{ff} \Lambda' + \Phi_0$, and where $\Phi_0 = \text{diag}(\phi_{0,1}^2, \phi_{0,2}^2, \dots, \phi_{0,N}^2)$ is a diagonal matrix holding the diagonal elements of $T^{-1} \sum_{t=1}^T \Phi_t$. Bai and Li (2012a) show that, under quite general forms of idiosyncratic autocorrelation, time heteroscedasticity and cross-sectional correlations admitting an approximate factor model, the quasi-MLEs of Λ , \mathbf{M}_{ff} and Φ_0 are consistent (under certain identifying restrictions) with average convergence rates

$$\begin{aligned} \frac{1}{N} \sum_{i=1}^N \frac{1}{\hat{\phi}_{0,i}^2} \|\hat{\lambda}_i - \lambda_i\|^2 &= O_p(T^{-1}) + O_p(N^{-2}), \\ \frac{1}{N} \sum_{i=1}^N (\hat{\phi}_{0,i}^2 - \phi_{0,i}^2)^2 &= O_p(T^{-1}) + O_p(N^{-2}), \\ \|\widehat{\mathbf{M}}_{ff} - \mathbf{M}_{ff}\|^2 &= O_p(T^{-1}) + O_p(N^{-2}). \end{aligned}$$

If the exact factor model is the true model, then the convergence rates are $O_p(T^{-1})$, i.e. the $O_p(N^{-2})$ term then falls out. An important property of the quasi-MLEs is that the results hold also if the factors have arbitrary dynamic properties with $E\|\mathbf{f}_t\|^4 < \infty$.

In this paper we combine the quasi-likelihood based methods of Bai and Li (2012b,a) with the full maximum likelihood approach for the covariance stationary/unit root AR(1) panel model in Krueger (2008). But, we do not employ the usual procedure of taking differences, then estimating, then re-accumulating, and then performing unit root tests, as in Bai and Ng (2004). Instead, we base our test directly on the likelihood function of the differenced data, and construct a homogenous LM-type test for idiosyncratic unit roots. We restrict ourselves to concern the exact factor model with AR(1) dynamics in the idiosyncratic component. However, we show that the factors may have arbitrary dynamics as long as they are difference stationary, and propose that a more general test that allows also for higher order idiosyncratic dynamics may be derived from the same principles as in this paper.

The rest of the paper is organized as follows: Section 2 goes through the framework and the model assumptions. Section 3 derives the LM-type statistic and its asymptotic distribution. In Section 4 we evaluate, through Monte Carlo simulation, size and power of the LM-type statistic and compare the performance with some existing second generation panel unit root tests. Section 5 concludes. All mathematical derivations are placed in an appendix.

Notation: For a matrix $\mathbf{A} \in \mathbb{R}^{n \times n}$, $\varphi_1(\mathbf{A}) \geq \varphi_2(\mathbf{A}) \geq \dots \geq \varphi_n(\mathbf{A})$ denote the eigenvalues and $\text{dg}(\mathbf{A})$ is the operator which stacks the diagonal elements in an $n \times 1$ vector. 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 $\text{vec}(\mathbf{A})$ 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 i th row and j th column, $\text{diag}(a_1, a_2, \dots, a_n)$ is an $n \times n$ diagonal matrix with entries a_1, a_2, \dots, a_n , and \mathbf{I}_n is the $n \times n$ identity matrix. For limits, $T \rightarrow$ denotes limit taken over T ; $(T, N)_s \rightarrow$ denotes sequential limit with limit taken over T followed by limit taken over N ; and $N, T \rightarrow$ denotes joint limit, where N and T go to infinity simultaneously. \xrightarrow{p} (\xrightarrow{d}) denotes convergence in probability (distribution) and $\overset{s}{\sim}$ denotes similar matrices (see Appendix A2).

2 The framework

Let $i = 1, 2, \dots, N$ denote the cross-sectional individuals and $t = 1, 2, \dots, T$ denote the time points. We assume the panel data has the following representation:

$$\mathbf{x}_t = (x_{1,t}, x_{2,t}, \dots, x_{N,t})' = \boldsymbol{\mu} + \boldsymbol{\Lambda}\mathbf{f}_t + \mathbf{u}_t, \quad (3)$$

where $\boldsymbol{\mu} = (\mu_1, \mu_2, \dots, \mu_N)'$ are constants, $\mathbf{f}_t = (f_{1,t}, f_{2,t}, \dots, f_{r,t})'$ are unobservable dynamic factors, $\mathbf{u}_t = (u_{1,t}, u_{2,t}, \dots, u_{N,t})'$ are unobservable dynamic idiosyncratic components and $\boldsymbol{\Lambda} = [\lambda_{i,j}]_{N \times r}$ is the matrix of factor loadings. Through out the paper we will assume that \mathbf{x}_t is integrated of at most order 1, where $\mathbf{x}_t \sim I(1)$ means $x_{i,t} \sim I(1)$ for at least one i . Further, if $x_{i,t} \sim I(1)$, then the nonstationarity could come from one of the factors or from the idiosyncratic component:

$$\begin{aligned} x_{i,t} \sim I(1) &\Leftrightarrow u_{i,t} \sim I(1) \text{ and/or } \exists j : f_{j,t} \sim I(1), \\ x_{i,t} \sim I(0) &\Leftrightarrow u_{i,t} \sim I(0) \text{ and } \forall j : f_{j,t} \sim I(0). \end{aligned}$$

The interest here is to test for unit roots in the idiosyncratic components \mathbf{u}_t using maximum likelihood, and where we want this test to be asymptotically independent of the distribution of the factors. Similar to the procedures in Bai and Ng (2004, 2010) we proceed by taking first-differences. Let $\mathbf{y}_t = \Delta \mathbf{x}_t = (\Delta x_{1,t}, \Delta x_{2,t}, \dots, \Delta x_{N,t})'$ and let similarly $\mathbf{g}_t = \Delta \mathbf{f}_t$ and $\mathbf{z}_t = \Delta \mathbf{u}_t$, which by assumption are all stationary processes. The contemporaneous covariances are now $\Sigma_{yy} = \text{Var}(\mathbf{y}_t) = \Lambda \Sigma_{gg} \Lambda' + \Sigma_{zz}$, where $\Sigma_{gg} = \text{Var}(\Delta \mathbf{f}_t)$ and $\Sigma_{zz} = \text{Var}(\Delta \mathbf{u}_t)$. Based on the likelihood function of the differenced and stacked panel data we derive a Lagrange multiplier test for idiosyncratic unit roots. We make the following general assumptions, drawing partly on the assumptions in Bai (2003) and Bai and Ng (2002, 2004):

Assumption 2.1 (factors) The factors admit the representation $(1 - L)\mathbf{f}_t = \mathbf{C}(L)\mathbf{v}_t$, where $\mathbf{C}(L) = \sum_{m=0}^{\infty} \mathbf{C}_m L^m$ is an $r \times r$ lag polynomial matrix, and where;

- (i) $\mathbf{v}_t \sim \mathcal{N}(\mathbf{0}, \Sigma_{vv})$ is iid, where $E\|\mathbf{v}_t\|^4 < \infty$;
- (ii) $\Sigma_{gg} = \sum_{m=0}^{\infty} \mathbf{C}_m \Sigma_{vv} \mathbf{C}_m' > 0$;
- (iii) $\sum_{m=0}^{\infty} m \|\mathbf{C}_m\| < \infty$;
- (iv) $\text{rank}(\mathbf{C}(1)) = r_1$, where $0 \leq r_1 \leq r$.

Assumption 2.2 (idiosyncratic components) The idiosyncratic components are AR(1) processes; $(1 - \rho_i L)u_{i,t} = \varepsilon_{i,t}$, where $\rho_i \in (-1, 1]$, $\varepsilon_{i,t} \sim \mathcal{N}(0, \sigma_{\varepsilon,i}^2)$ with $\sigma_{\varepsilon,i}^2 < \infty$, and where $E(\varepsilon_{i,t} \varepsilon_{l,s}) = 0$ for all $i \neq l$ ($i, l = 1, 2, \dots, N$) and all $t, s = 1, 2, \dots, T$.

Assumption 2.3 (independence) The errors $v_{j,t}$ and $\varepsilon_{i,t}$ are mutually independent at all leads and lags, such that $E(v_{j,t} \varepsilon_{i,s}) = 0$ for all $j = 1, 2, \dots, r$, $i = 1, 2, \dots, N$ and $t, s = 1, 2, \dots, T$.

Assumption 2.4 (starting values) For stationary processes $f_{j,t}$ and $u_{i,t}$, the starting values $f_{j,0}$ and $u_{i,0}$ come from the stationary distributions, while for nonstationary processes they are $\mathcal{O}_p(1)$.

Assumption 2.1 is the same dynamic assumption on the factors as made in Bai and Ng (2004), except we assume normally distributed shocks. It allows for r_1 stochastic trends among the factors. Under Assumption 2.2, the idiosyncratic components are allowed to be cross-sectionally heteroscedastic but time heteroscedasticity is ruled out. Also, because $\Sigma_{zz} = E(\mathbf{z}_t \mathbf{z}_t')$ is a diagonal matrix, where $\mathbf{z}_t = (\Delta u_{1,t}, \Delta u_{2,t}, \dots, \Delta u_{N,t})'$, we have the exact factor model. Assumption 2.3 is a standard assumption in factor analysis to identify the covariance structure (see e.g. Bai, 2003), and Assumption 2.4 ensures us that the stochastic processes are well-behaved with respect to their initial values.

Our likelihood based analysis rests on the same observation as is made by Kruiniger (2008). For notational convenience, let $T^* = T - 1$, and let $\mathbf{D} = [D_{s,t}]_{T^* \times T}$ be the first-difference matrix, i.e. $D_{s,t} = -1$ if $t = s$, $D_{s,t} = 1$ if $t = s + 1$, and zero otherwise. Also, let $\mathbf{u}_i = (u_{i,1}, u_{i,2}, \dots, u_{i,T})'$. If \mathbf{u}_i has a unit root, then $\mathbf{D}\mathbf{u}_i \sim \mathcal{N}_T(\mathbf{0}, \sigma_{\varepsilon,i}^2 \mathbf{I})$. Conversely, if \mathbf{u}_i is stationary and the starting value $u_{i,0}$ is chosen according to Assumption 2.4, then it is

straightforward to show that $\mathbf{D}\mathbf{u}_i \sim \mathcal{N}_{\mathcal{T}}(\mathbf{0}, \sigma_{\varepsilon_i}^2 \mathbf{\Psi})$, where $\mathbf{\Psi}(\rho_i) = [\Psi_{k,m}^i]_{T^* \times T^*}$ with

$$\Psi_{k,m}^i = \begin{cases} \frac{2}{1+\rho_i} & \text{for } k = m \\ -\frac{\rho_i^{|k-s|-1}(1-\rho_i)}{1+\rho_i} & \text{for } k \neq m \end{cases}. \quad (4)$$

The covariance matrix $\mathbf{\Psi}$ has two convenient properties; (a) $\mathbf{\Psi}(1) = \mathbf{I}_{T^*}$, i.e. it is correctly defined in the nonstationary point; and (b) it is infinitely many times continuously differentiable at and in the neighborhood of the unit root $\rho_i = 1$ (see Kruiniger, 2008, Lemma 7).²

Assumption 2.5 (factor loadings) It holds that;

- (i) $\|\lambda_i\| < \infty$ for all $i = 1, 2, \dots, N$;
- (ii) $N^{-1} \mathbf{\Lambda}' \mathbf{\Sigma}_{zz}^{-1} \mathbf{\Lambda} \rightarrow \mathbf{\Sigma}_1 > 0$, as $N \rightarrow \infty$.

Assumption 2.6 (idiosyncratic homogeneity) The idiosyncratic components are homogenous, $\forall i: \rho_i = \rho$.

Assumption 2.5 is the same assumption as is made in Bai and Li (2012b,a), and is needed to put the necessary identifying restrictions on the factor loadings that allow for quasi maximum likelihood estimation. It can also be seen as the analog of the usual assumption of identifiable pervasive factors (see e.g. Bai and Ng, 2004), where $N^{-1} \mathbf{\Lambda}' \mathbf{\Lambda}$ converges to some positive definite matrix as $N \rightarrow \infty$. Here the convergence concerns the loadings weighted by the idiosyncratic variances, and implies that the eigenvalues of $\mathbf{\Lambda}' \mathbf{\Sigma}_{zz}^{-1} \mathbf{\Lambda}$ are $\mathcal{O}(N)$. Assumption 2.6, which may be relaxed, simplifies the analysis. This assumption only affects power, and we generally expect homogenous tests to have power also under a heterogenous alternative. Imposing this assumption the hypotheses of interest are

$$\mathcal{H}_0 : \rho = 1, \text{ vs } \mathcal{H}_1 : \rho < 1.$$

Misspecification 1 The differenced factors are multivariate white noise, drawn from the normal distribution, $\Delta \mathbf{f}_t \sim \mathcal{N}_r(\mathbf{0}, \mathbf{I})$.

Misspecification 1 implies that the factors in levels are independent random walks. By imposing this misspecification we have an approximating model in the same sense as e.g. Doz, Giannone, and Reichlin (2012). Proceeding with the likelihood based inference will in this sense result in a quasi-likelihood statistic. We prove later that our statistic to be proposed is asymptotically robust against this misspecification, and that the factors are allowed to have the representation in Assumption 2.1 as long as both N and T tend to infinity.

Consider the observed panel data $\mathbf{X} = (\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_T)$ and let $\mathbf{Y} = \mathbf{X}\mathbf{D}'$ denote the $N \times T^*$ first-differenced data. Imposing Misspecification 1, the differenced and stacked data of size $NT^* \times 1$, $\mathbf{Y}_v = \text{vec}(\mathbf{Y})$, has covariance matrix

$$\mathbf{\Sigma} = \text{Var}(\mathbf{Y}_v) = (\mathbf{I}_{T^*} \otimes \mathbf{\Lambda}\mathbf{\Lambda}') + (\mathbf{\Psi}(\rho) \otimes \mathbf{\Sigma}_{\varepsilon\varepsilon}),$$

²The autocovariance matrix of a stationary AR(1) process, $\mathbf{\Pi}$ say, may be found from e.g. van der Leeuw (1994) and Karanasos (1998). Straightforward algebra will show that $\mathbf{D}\mathbf{\Pi}\mathbf{D}' = \mathbf{\Psi}$, where $\mathbf{\Psi}$ is a matrix with elements (4).

where \otimes denotes the Kronecker product, $\Psi(\rho)$ is the idiosyncratic autocovariance matrix defined before, and $\Sigma_{\varepsilon\varepsilon} = E(\varepsilon_t \varepsilon_t') = \text{diag}(\sigma_{\varepsilon,1}^2, \sigma_{\varepsilon,2}^2, \dots, \sigma_{\varepsilon,N}^2)$ is the idiosyncratic error variance, where $\varepsilon_t = (\varepsilon_{1,t}, \varepsilon_{2,t}, \dots, \varepsilon_{N,t})'$ are the idiosyncratic error terms. Under the null hypothesis $\rho = 1$ we have that $\Sigma = (\mathbf{I}_{T^*} \otimes \Lambda \Lambda') + (\mathbf{I}_{T^*} \otimes \Sigma_{\varepsilon\varepsilon}) = (\mathbf{I}_{T^*} \otimes \Omega)$, where $\Omega = (\Lambda \Lambda' + \Sigma_{\varepsilon\varepsilon})$, such that for $\mathbf{y}_t = \Delta \mathbf{x}_t = (\Delta x_{1t}, \Delta x_{2t}, \dots, \Delta x_{Nt})'$,

$$\text{Cov}(\mathbf{y}_t, \mathbf{y}_s) = \begin{cases} \Omega & \text{if } s = t \\ \mathbf{0} & \text{if } s \neq t \end{cases}. \quad (5)$$

The corresponding sample covariances, $\mathbf{Y}_v \mathbf{Y}_v'$, consists of the 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

$$\mathbf{S}_0 = \sum_{t=2}^T \mathbf{S}_{t,t} = \sum_{t=2}^T \mathbf{y}_t \mathbf{y}_t' = \mathbf{Y} \mathbf{Y}', \quad (6)$$

and

$$\mathbf{S}_{00} = \sum_{t=2}^T \sum_{s=2}^T \mathbf{S}_{t,s} = \sum_{t=2}^T \sum_{s=2}^T \mathbf{y}_t \mathbf{y}_s' = \left(\sum_{t=2}^T \mathbf{y}_t \right) \left(\sum_{t=2}^T \mathbf{y}_t \right)'. \quad (7)$$

Note that if Misspecification 1 is true, then we have, under the null hypothesis, that $E(\mathbf{S}_{t,t}) = \Omega$ for all t , such that by the weak law of large numbers

$$\frac{1}{T^*} \sum_{t=2}^T \mathbf{S}_{t,t} = \frac{1}{T^*} \mathbf{S}_0 \xrightarrow{p} \Omega, \quad (8)$$

and by the central limit theorem

$$\frac{1}{\sqrt{T^*}} \Omega^{-\frac{1}{2}} \left(\sum_{t=2}^T \mathbf{y}_t - \mathbf{0} \right) \xrightarrow{d} \mathbf{Z} \sim \mathcal{N}_N(\mathbf{0}, \mathbf{I}). \quad (9)$$

3 The LM-type statistic

Let $\Lambda_v = \text{vec}(\Lambda)$ and $\sigma = \text{dg}(\Sigma_{\varepsilon\varepsilon})$, and define the parameter vector $\theta = (\Lambda_v', \sigma', \rho)'$ with $K = Nr + N + 1$ parameters. Assuming normality, the log-likelihood with respect to the differenced and stacked data is

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

To be clear in notation we make the following definition:

Definition 3.1 (Score and information) *For any parameter subsets $\omega, \nu \subseteq \theta$ with K_ω and K_ν parameters respectively, where ω_k denotes the k th parameter in ω and ν_q denotes the q th parameter in ν , the score w.r.t. ω is $\mathbf{V}_\omega \equiv \frac{\partial l(\theta)}{\partial \omega} \equiv [V_{\omega_k}]_{K_\omega \times 1}$, where $V_{\omega_k} = \frac{\partial l(\theta)}{\partial \omega_k}$, and the information w.r.t. ω and ν is $\mathbf{J}_{\omega\nu'} \equiv -E \left(\frac{\partial^2 l(\theta)}{\partial \omega \partial \nu'} \right) \equiv [J_{\omega_k \nu_q}]_{K_\omega \times K_\nu}$, where $J_{\omega_k \nu_q} = -E \left(\frac{\partial^2 l(\theta)}{\partial \omega_k \partial \nu_q} \right)$.*

The LM-statistic is defined as

$$LM = \mathbf{V}'_{\theta} \mathbf{J}_{\theta\theta'}^{-1} \mathbf{V}_{\theta} \Big|_{\theta=\tilde{\theta}}, \quad (11)$$

where $\tilde{\theta} = (\tilde{\Lambda}'_v, \tilde{\sigma}', 1)'$ is the restricted maximum likelihood estimator under \mathcal{H}_0 . The score and information under Assumptions 2.1-2.6, but imposing Misspecification 1, are derived in Appendix A1. Let $\theta = (\theta'_1, \theta'_2)'$ with θ_2 holding the parameters that are subject to restrictions under the null hypothesis. Under restricted maximum likelihood estimation we have that $\mathbf{V}_{\theta}|_{\theta=\tilde{\theta}} = (\mathbf{0}, \mathbf{V}'_{\theta_2}|_{\theta=\tilde{\theta}})'$. The information matrix and its inverse may then be partitioned into the relevant blocks;

$$\mathbf{J} = \begin{pmatrix} \mathbf{J}_{11} & \mathbf{J}_{12} \\ \mathbf{J}_{21} & \mathbf{J}_{22} \end{pmatrix}; \quad \mathbf{J}^{-1} = \begin{pmatrix} \mathbf{J}^{11} & \mathbf{J}^{12} \\ \mathbf{J}^{21} & \mathbf{J}^{22} \end{pmatrix},$$

where $\mathbf{J}_{11} = \mathbf{J}_{\theta_1\theta'_1}$, $\mathbf{J}_{12} = \mathbf{J}'_{21} = \mathbf{J}_{\theta_1\theta'_2}$, and $\mathbf{J}_{22} = \mathbf{J}_{\theta_2\theta'_2}$, such that the LM-statistic (11) becomes

$$LM = \mathbf{V}'_{\theta_2} \mathbf{J}^{22} \mathbf{V}_{\theta_2} \Big|_{\theta=\tilde{\theta}}. \quad (12)$$

Condition 3.1 (asymptotic block-diagonality) *There exists a sequence of matrices of constants $\mathbf{C}_{i,NT} \rightarrow \infty$ for $i = 1, 2$ such that:*

$$\begin{aligned} \mathbf{C}_{1,NT}^{-1} \mathbf{J}_{11} \mathbf{C}'_{1,NT} &\rightarrow F_{11}; & \mathbf{C}_{1,NT}^{-1} \mathbf{J}_{12} \mathbf{C}'_{2,NT} &\rightarrow \mathbf{0}; \\ \mathbf{C}_{2,NT}^{-1} \mathbf{J}_{21} \mathbf{C}'_{1,NT} &\rightarrow \mathbf{0}; & \mathbf{C}_{2,NT}^{-1} \mathbf{J}_{22} \mathbf{C}'_{2,NT} &\rightarrow F_{22}, \end{aligned}$$

where $\mathbf{C}^{-'} = (\mathbf{C}^{-1})'$, and where F_{11} and F_{22} are invertible matrices.

Condition 3.1, which is inspired by Solo (1984), implies that

$$\begin{aligned} \mathbf{J}^{22} &= \left(\mathbf{J}_{22} - \mathbf{J}_{21} \mathbf{J}_{11}^{-1} \mathbf{J}_{12} \right)^{-1} = \mathbf{C}'_{2,NT} \left[\mathbf{C}_{2,NT}^{-1} \mathbf{J}_{22} \mathbf{C}'_{2,NT} - \right. \\ &\quad \left. \left(\mathbf{C}_{2,NT}^{-1} \mathbf{J}_{21} \mathbf{C}'_{1,NT} \right) \left(\mathbf{C}_{1,NT}^{-1} \mathbf{J}_{11} \mathbf{C}'_{1,NT} \right)^{-1} \left(\mathbf{C}_{1,NT}^{-1} \mathbf{J}_{12} \mathbf{C}'_{2,NT} \right) \right]^{-1} \mathbf{C}_{2,NT}^{-1} \rightarrow \mathbf{J}_{22}^{-1}, \end{aligned}$$

and likewise that $\mathbf{J}^{11} \rightarrow \mathbf{J}_{11}^{-1}$. We may analogously use probability limits, meaning that Condition 3.1 holds in probability, in which case, letting $\tilde{\mathbf{J}}$ denote $\mathbf{J}|_{\theta=\tilde{\theta}}$, we have that $\tilde{\mathbf{J}}^{22} \xrightarrow{p} \tilde{\mathbf{J}}_{22}^{-1}$ and $\tilde{\mathbf{J}}^{11} \xrightarrow{p} \tilde{\mathbf{J}}_{11}^{-1}$. Let similarly $\tilde{\mathbf{V}}$ denote $\mathbf{V}|_{\theta=\tilde{\theta}}$. Whenever Condition 3.1 holds in probability, the statistic (12) can be written as

$$LM = \tilde{\mathbf{V}}'_{\theta_2} \tilde{\mathbf{J}}^{22} \tilde{\mathbf{V}}_{\theta_2} \simeq LM^* = \tilde{\mathbf{V}}'_{\theta_2} \tilde{\mathbf{J}}_{22}^{-1} \tilde{\mathbf{V}}_{\theta_2}, \quad (13)$$

in probability as $\mathbf{C}_{i,NT} \rightarrow \infty$, in the sense that $|LM - LM^*| = o_p(1)$. For the case considered in this paper $\theta_2 = \rho$, so the information matrix may be decomposed as;

$$\mathbf{J}_{11} = \begin{pmatrix} \mathbf{J}_{\Lambda_v \Lambda'_v} & \mathbf{J}_{\Lambda_v \sigma'} \\ \mathbf{J}_{\sigma \Lambda'_v} & \mathbf{J}_{\sigma \sigma'} \end{pmatrix}; \quad \mathbf{J}_{21} = \mathbf{J}'_{12} = \begin{pmatrix} \mathbf{J}_{\rho \Lambda'_v} & \mathbf{J}_{\rho \sigma'} \end{pmatrix}; \quad \mathbf{J}_{22} = \mathbf{J}_{\rho \rho}. \quad (14)$$

Here we show that the information matrix is asymptotically block-diagonal in probability as $T \rightarrow \infty$, and consider the LM-type statistic

$$\vartheta = \tilde{V}_\rho \sqrt{\tilde{J}_{\rho\rho}^{-1}}, \quad (15)$$

for which, if Condition 3.1 holds, we have that $\vartheta^2 \simeq LM$ in probability. To show that Condition 3.1 holds for fixed N , the following lemma is sufficient.

Lemma 3.1 *Under Assumptions 2.1-2.6 but imposing Misspecification 1, the following terms are $\mathcal{O}_p(1)$ for all $i, l = 1, 2, \dots, N$ and all $j, q = 1, 2, \dots, r$:*

$$(i) \frac{\tilde{J}_{\lambda_{ij}\lambda_{lq}}}{N^2T}; \quad (ii) \frac{\tilde{J}_{\lambda_{ij}\sigma_{\epsilon_{il}}^2}}{T}; \quad (iii) \frac{\tilde{J}_{\lambda_{ij}\rho}}{N^{3/2}T}; \quad (iv) \frac{\tilde{J}_{\sigma_{\epsilon_{il}}^2, \sigma_{\epsilon_{il}}^2}}{T}; \quad (v) \frac{\tilde{J}_{\rho\sigma_{\epsilon_{il}}^2}}{T}; \quad (vi) \frac{\tilde{J}_{\rho\rho}}{NT^2}.$$

If we let $\mathbf{C}_{1,NT} = \text{diag}(N\sqrt{T}\mathbf{I}_{Nr}, \sqrt{T}\mathbf{I}_N)$ and $\mathbf{C}_{2,NT} = T\sqrt{N}$, then we have that, using Lemma 3.1,

$$\begin{aligned} \mathbf{C}_{1,NT}^{-1} \tilde{\mathbf{J}}_{11} \mathbf{C}_{1,NT}^{-1} &= \begin{pmatrix} \mathbf{O}_p(1)_{Nr \times Nr} & \mathbf{O}_p(N^{-1})_{Nr \times N} \\ \mathbf{O}_p(N^{-1})_{N \times Nr} & \mathbf{O}_p(1)_{N \times N} \end{pmatrix} ((Nr + N) \times (Nr + N)), \\ \mathbf{C}_{2,NT}^{-1} \tilde{\mathbf{J}}_{21} \mathbf{C}_{1,NT}^{-1} &= \begin{pmatrix} \mathbf{O}_p(T^{-1/2})_{1 \times Nr} & \mathbf{O}_p((NT)^{-1/2})_{1 \times N} \end{pmatrix} (1 \times (Nr + N)), \\ \mathbf{C}_{2,NT}^{-1} \tilde{\mathbf{J}}_{22} \mathbf{C}_{2,NT}^{-1} &= \mathbf{C}_{2,NT}^{-1} \tilde{J}_{\rho\rho} \mathbf{C}_{2,NT}^{-1} = \mathcal{O}_p(1), \end{aligned}$$

where normal \mathbf{O} 's denote scalars and boldface \mathbf{O} 's denote matrices. Hence, the information matrix is asymptotically block diagonal for any fixed N as T tends to infinity, because then, $\mathbf{C}_{1,NT}^{-1} \tilde{\mathbf{J}}_{12} \mathbf{C}_{2,NT}^{-1} \xrightarrow{p} \mathbf{0}$, $\mathbf{C}_{2,NT}^{-1} \tilde{\mathbf{J}}_{21} \mathbf{C}_{1,NT}^{-1} \xrightarrow{p} \mathbf{0}$, $\mathbf{C}_{1,NT}^{-1} \tilde{\mathbf{J}}_{11} \mathbf{C}_{1,NT}^{-1} \xrightarrow{p} F_{11}$ for some matrix F_{11} , and $\mathbf{C}_{2,NT}^{-1} \tilde{\mathbf{J}}_{22} \mathbf{C}_{2,NT}^{-1} \xrightarrow{p} F_{22}$ for some scalar F_{22} implying $\tilde{\mathbf{J}}^{22} \xrightarrow{p} \tilde{J}_{22}^{-1} = \tilde{J}_{\rho\rho}^{-1}$. Note that we do not need to check if F_{11} is invertible, because the corresponding part of the information matrix does not enter the LM-statistic. Without loss of generality we may assume that F_{11} is either non-singular or that any linearly dependent, and hence redundant, rows and columns have been removed. Also, from the proof of Lemma 3.1 (vi) in Appendix A2 we have that $F_{22} = \frac{1}{8N} \sum_{i=1}^N (\eta_i + 1)^{-2} \rightarrow \frac{1}{8}$, as $N \rightarrow \infty$, where $\eta_i = \varphi_i (\boldsymbol{\Lambda}' \boldsymbol{\Sigma}_{\epsilon\epsilon}^{-1} \boldsymbol{\Lambda})$.

It follows for the LM-statistic (13) that $LM^* = \tilde{V}_\rho \tilde{J}_{\rho\rho}^{-1} \tilde{V}_\rho$, and using the results (A2) and (A3) from Appendix A1, the LM-type statistic (15) is

$$\vartheta = \tilde{V}_\rho \sqrt{\tilde{J}_{\rho\rho}^{-1}} = \frac{T^* \text{tr}(\tilde{\boldsymbol{\Sigma}}_{\epsilon\epsilon} \mathbf{S}_{01}^{-1}) - 2 \text{tr}(\tilde{\boldsymbol{\Sigma}}_{\epsilon\epsilon} \mathbf{S}_{01}^{-1} \mathbf{S}_0 \mathbf{S}_{01}^{-1}) + \text{tr}(\tilde{\boldsymbol{\Sigma}}_{\epsilon\epsilon} \mathbf{S}_{01}^{-1} \mathbf{S}_{00} \mathbf{S}_{01}^{-1})}{\sqrt{2T^* \text{tr}(\tilde{\boldsymbol{\Sigma}}_{\epsilon\epsilon} \mathbf{S}_{01}^{-1} \tilde{\boldsymbol{\Sigma}}_{\epsilon\epsilon} \mathbf{S}_{01}^{-1})}}, \quad (16)$$

where $\mathbf{S}_{01} = \boldsymbol{\Omega}(\tilde{\boldsymbol{\theta}}) = \tilde{\boldsymbol{\Lambda}} \tilde{\boldsymbol{\Lambda}}' + \tilde{\boldsymbol{\Sigma}}_{\epsilon\epsilon}$, and \mathbf{S}_0 and \mathbf{S}_{00} are given by (6) and (7) respectively. Here $\boldsymbol{\Sigma}_{\epsilon\epsilon}$ can be consistently estimated with the EM algorithm for the different schemes IC1 - IC5 in Bai and Li (2012b,a). In this paper we use IC3, which identifies an orthogonal rotation of $\boldsymbol{\Lambda}$, say $\boldsymbol{\Lambda}^*$, up to a column sign change. This is done by imposing the identifying restrictions that $\boldsymbol{\Lambda}^* \boldsymbol{\Sigma}_{\epsilon\epsilon} \boldsymbol{\Lambda}^* / N$ is a diagonal matrix with distinct values (in descending order) and that

$\mathbf{M}_{gg} = \mathbf{I}_r$, where \mathbf{M}_{gg} is the differenced analog to \mathbf{M}_{ff} defined for the objective function (2). This combined restriction is the most general of the five proposed by Bai and Li (2012b,a) in the sense that estimates under the remaining four schemes can be found explicitly from IC3. To see how this restriction works, suppose there exists an orthogonal matrix \mathbf{Q} with the property that $\mathbf{Q}'\boldsymbol{\Lambda}'\boldsymbol{\Sigma}_{\varepsilon\varepsilon}\boldsymbol{\Lambda}\mathbf{Q}$ is a diagonal matrix. Then $\boldsymbol{\Lambda}^* = \boldsymbol{\Lambda}\mathbf{Q}$ fulfills the restriction, and the covariances are equally identified because $\boldsymbol{\Lambda}^*\boldsymbol{\Lambda}^{*\prime} = \boldsymbol{\Lambda}\boldsymbol{\Lambda}'$. By assumption there exists asymptotically at least one solution with this property, namely the left eigenvectors associated with the (positive) eigenvalues of $\boldsymbol{\Sigma}_1$ defined in Assumption 2.5 (ii). The EM algorithm will iteratively find an orthogonal rotation with the property just described, and assuming $\tilde{\boldsymbol{\Lambda}}^*$ has the same column signs as those of $\boldsymbol{\Lambda}^*$, we have that $\tilde{\boldsymbol{\Lambda}}^* \xrightarrow{p} \boldsymbol{\Lambda}^*$ as $N, T \rightarrow \infty$. Further, through the EM algorithm we may find $\tilde{\boldsymbol{\Phi}}_0$, associated with the objective function (2). After taking differences we have that $\tilde{\boldsymbol{\Phi}}_0 \xrightarrow{p} \boldsymbol{\Sigma}_{\varepsilon\varepsilon}$ as $N, T \rightarrow \infty$. For notational simplicity, from here on we let $\tilde{\boldsymbol{\Sigma}}_{\varepsilon\varepsilon} = \tilde{\boldsymbol{\Phi}}_0$.

Let as before $\mathbf{y}_t = \Delta\mathbf{x}_t$ and $\mathbf{g}_t = \Delta\mathbf{f}_t$, and define the second moment of the sample

$$\mathbf{M}_{yy} = \frac{1}{T^*} \sum_{t=2}^T (\mathbf{y}_t - \bar{\mathbf{y}}_t)(\mathbf{y}_t - \bar{\mathbf{y}}_t)',$$

where $\bar{\mathbf{y}}_t = \frac{1}{T^*} \sum_{t=2}^T \mathbf{y}_t$. Let $\boldsymbol{\theta}_1^{(k)} = (\boldsymbol{\Lambda}^{(k)}, \boldsymbol{\Sigma}_{\varepsilon\varepsilon}^{(k)})$ denote the estimator at the k th iteration from the EM algorithm, which is updated by (see Bai and Li, 2012b,a, and references therein):

$$\begin{aligned} \boldsymbol{\Lambda}^{(k+1)} &= \left[\frac{1}{T^*} \sum_{t=2}^T E \left(\mathbf{y}_t \mathbf{g}_t' | \mathbf{Y}, \boldsymbol{\theta}_1^{(k)} \right) \right] \left[\frac{1}{T^*} \sum_{t=2}^T E \left(\mathbf{g}_t \mathbf{g}_t' | \mathbf{Y}, \boldsymbol{\theta}_1^{(k)} \right) \right]^{-1}, \\ \boldsymbol{\Sigma}_{\varepsilon\varepsilon}^{(k+1)} &= \text{diag}^* \left(\mathbf{M}_{yy} - \boldsymbol{\Lambda}^{(k+1)} \boldsymbol{\Lambda}^{(k)\prime} \left(\boldsymbol{\Sigma}_{yy}^{(k)} \right)^{-1} \mathbf{M}_{yy} \right), \end{aligned}$$

where diag^* sets the off-diagonal elements to zero, and where

$$\begin{aligned} \frac{1}{T^*} \sum_{t=2}^T E \left(\mathbf{y}_t \mathbf{g}_t' | \mathbf{Y}, \boldsymbol{\theta}_1^{(k)} \right) &= \mathbf{M}_{yy} \left(\boldsymbol{\Sigma}_{yy}^{(k)} \right)^{-1} \boldsymbol{\Lambda}^{(k)}, \\ \frac{1}{T^*} \sum_{t=2}^T E \left(\mathbf{g}_t \mathbf{g}_t' | \mathbf{Y}, \boldsymbol{\theta}_1^{(k)} \right) &= \boldsymbol{\Lambda}^{(k)\prime} \left(\boldsymbol{\Sigma}_{yy}^{(k)} \right)^{-1} \mathbf{M}_{yy} \left(\boldsymbol{\Sigma}_{yy}^{(k)} \right)^{-1} \boldsymbol{\Lambda}^{(k)} + \mathbf{I}_r - \boldsymbol{\Lambda}^{(k)\prime} \left(\boldsymbol{\Sigma}_{yy}^{(k)} \right)^{-1} \boldsymbol{\Lambda}^{(k)}, \end{aligned}$$

with $\boldsymbol{\Sigma}_{yy}^{(k)} = \boldsymbol{\Lambda}^{(k)} \boldsymbol{\Lambda}^{(k)\prime} + \boldsymbol{\Sigma}_{\varepsilon\varepsilon}^{(k)}$. The algorithm is typically initiated with the principal components estimator, where the iteration is repeated until $\|\boldsymbol{\theta}_1^{(k+1)} - \boldsymbol{\theta}_1^{(k)}\|$ reaches some sufficiently small tolerance level. This yields, under the null hypothesis, that $\mathbf{S}_{01} = \boldsymbol{\Sigma}_{yy}^{(k)} = \tilde{\boldsymbol{\Lambda}}^* \tilde{\boldsymbol{\Lambda}}^{*\prime} + \tilde{\boldsymbol{\Sigma}}_{\varepsilon\varepsilon} \xrightarrow{p} \boldsymbol{\Omega}$, as $N, T \rightarrow \infty$, where $\boldsymbol{\Omega}$ is the contemporaneous covariance matrix defined in (5). Note then that, asymptotically, the restrictions under IC3 and any potential errors in column signs will not pose a problem for the statistic (16), because $\tilde{\boldsymbol{\Lambda}}^* \tilde{\boldsymbol{\Lambda}}^{*\prime} \xrightarrow{p} \boldsymbol{\Lambda}\boldsymbol{\Lambda}'$.

The case when $\boldsymbol{\Sigma}_{\varepsilon\varepsilon} = \sigma^2 \mathbf{I}_N$ (the exact factor model with spherical noise) was considered by Zhou and Solberger (2012). We then have explicit maximum likelihood estimators of $\boldsymbol{\Lambda}$ and σ^2 (see e.g. Doz, Giannone, and Reichlin, 2012; Stoica and Jansson, 2009), which reduce the computational burden. However, neglecting cross-sectional heteroscedasticity will cause the LM-statistic to be oversized.

We now consider sequential limits, by first studying the behavior of (16) when N is fixed as $T \rightarrow \infty$, and then the behavior as $(T, N)_s \rightarrow \infty$.

Theorem 3.1 Let $\eta_i = \varphi_i(\mathbf{\Lambda}'\mathbf{\Sigma}_{\varepsilon\varepsilon}^{-1}\mathbf{\Lambda})$ where $\eta_1 \geq \eta_2 \geq \dots \geq \eta_r \geq 0$ for $1 \leq r < N$, where r is the number of factors. Under Assumptions 2.1-2.6, but imposing Misspecification 1, the asymptotic distribution of the LM-type statistic (16) as $T \rightarrow \infty$ is the following weighted χ_1^2 (chi-square with one degrees of freedom) distribution:

$$\vartheta \xrightarrow{d} \frac{1}{\sqrt{2 \sum_{i=1}^N w_i^2}} \left(\sum_{i=1}^N w_i \chi_{1,i}^2 - \sum_{i=1}^N w_i \right), \quad (17)$$

where $\chi_{1,i}^2$ are independent over $i = 1, 2, \dots, N$ and where $w_i = \begin{cases} (1 + \eta_i)^{-1} & \text{if } 1 \leq i \leq r \\ 1 & \text{if } r < i \leq N \end{cases}$.

The estimated weights \tilde{w}_i can be found with the eigenvalues of $\tilde{\mathbf{\Lambda}}'\tilde{\mathbf{\Sigma}}_{\varepsilon\varepsilon}^{-1}\tilde{\mathbf{\Lambda}}$.³ But we also consider an approximation due to Satterthwaite (1946), which was illustrated by Zhou and Solberger (2012) to fit very well for the statistic (15) under cross-sectional idiosyncratic homoscedasticity. Rewrite the weighted sum of χ_1^2 in (17) as $\sum_{i=1}^N w_i \chi_{1,i}^2 = \pi \sum_{i=1}^N a_i \chi_{1,i}^2$, where $\pi = \sum_{i=1}^N w_i$ and $a_i = w_i/\pi$ such that $\sum_{i=1}^N a_i = 1$. We can then find a positive integer d such that $\sum_{i=1}^N a_i \chi_{1,i}^2$ and χ_d^2/d have equal first and second moments. The solution for this can be rearranged as

$$\vartheta \stackrel{app.}{\sim} \frac{1}{\sqrt{2d}} (\chi_d^2 - d), \quad (18)$$

with $d = \frac{[\text{tr}(\mathbf{\Sigma}_{\varepsilon\varepsilon}\mathbf{\Omega}^{-1})]^2}{\text{tr}(\mathbf{\Sigma}_{\varepsilon\varepsilon}\mathbf{\Omega}^{-1}\mathbf{\Sigma}_{\varepsilon\varepsilon}\mathbf{\Omega}^{-1})}$.⁴

Because $N - r \leq \sum_{i=1}^N w_i^2 \leq \sum_{i=1}^N w_i \leq N$, it follows that also $N - r \leq d \leq N$, and that a rough approximation of the degrees of freedom is the interpolation $d = N - \frac{r}{2}$, which gets more precise as N grows (with r fixed). Note also that Assumption 2.5 (ii) and the restriction under IC3 imply that the eigenvalues of $\tilde{\mathbf{\Lambda}}'\tilde{\mathbf{\Sigma}}_{\varepsilon\varepsilon}^{-1}\tilde{\mathbf{\Lambda}}$ are $\mathcal{O}_p(N)$ such that for all $i \leq r$, $\tilde{w}_i \xrightarrow{p} 0$ as $N \rightarrow \infty$, while for the remaining $N - r$ (non-stochastic) weights we have that $\tilde{w}_i = 1$ for $i = r + 1, r + 2, \dots, N$. Thus, the fixed- N -asymptotic distribution (17) may be decomposed as

$$\vartheta \sim \frac{\chi_{(N-r)}^2 - (N-r)}{\sqrt{2(N-r) + \mathcal{O}_p(N^{-2})}} + \mathcal{O}_p(N^{-3/2}),$$

which will tend to the standard normal distribution as N grows by the central limit theorem. The following theorem states that this holds true also if Misspecification 1 is violated.

³Note that the eigenvalues of $\mathbf{\Lambda}'\mathbf{\Sigma}_{\varepsilon\varepsilon}^{-1}\mathbf{\Lambda}$ correspond to the r largest eigenvalues of $\mathbf{\Sigma}_{\varepsilon\varepsilon}^{-1/2}\mathbf{\Lambda}\mathbf{\Lambda}'\mathbf{\Sigma}_{\varepsilon\varepsilon}^{-1/2}$, which is also similar to $\mathbf{\Sigma}_{\varepsilon\varepsilon}^{-1}\mathbf{\Lambda}\mathbf{\Lambda}'$. Similar matrices have the same characteristic polynomial, and they therefore have the same eigenvalues (see Appendix A2)

⁴The proof of this approximation is analogous to the proof of Proposition 1 in Zhou and Solberger (2012), and is left out.

Theorem 3.2 Under Assumptions 2.1-2.6, the asymptotic distribution of the LM-type statistic (16) as $(T, N)_s \rightarrow \infty$ is standard normal, $\vartheta \xrightarrow{d} \mathcal{N}(0, 1)$.

A direct consequence of Theorem 3.2 is that the LM-type test is asymptotically independent of the distribution of the factors as long as they fall under Assumption 2.2, which is quite general. The factors may be independent $I(1)$ processes, cointegrate, or be $I(0)$ processes.

Remarks 1 The assumptions in this paper could be considered strict, and though there is room for generalizations, they are outside the scope of this paper. However, some of the assumptions deserve to be commented; We consider AR(1) processes for the idiosyncratic components. For processes of higher order a similar procedure could be used. Consider for example the AR(2) case, which, assuming Gaussian errors, can be decomposed as

$$\begin{aligned} u_{i,t} &= \rho_i u_{i,t-1} + \varepsilon_{i,t}, \\ \varepsilon_{i,t} &= \gamma_i \varepsilon_{i,t-1} + v_{i,t}, \end{aligned}$$

where $|\gamma_i| < 1$ and $v_{i,t}$ is iid $\mathcal{N}(0, \sigma_{v,i}^2)$, and where the unit root hypothesis corresponds to $\rho_i = 1, \forall i$. If we let $\Psi(\rho_i, \gamma_i)$ denote the autocovariance matrix for a differenced AR(2) process, then $\Psi(1, \gamma_i) = \Pi(\gamma_i)$, where Π is the autocovariance matrix for a stationary AR(1) process in levels. Here the N nuisance parameters γ_i will have to be concentrated out. For the AR(p) case there will be $N(p-1)$ nuisance parameters. The quasi-likelihood procedures of Bai and Li (2012b,a) offer a potential to estimate these parameters. We could potentially also adjust the LM-statistic non-parametrically in the spirit of Phillips and Perron (1988). A second generalization of our test would perhaps be to include a trend in the model (3). A linear trend will have to be treated as a constant in the differenced data and be included in the likelihood function. Demeaning the differenced panel data as in Bai and Ng (2004), i.e. $\Delta x_{i,t} - \overline{\Delta x}_i$, where $\overline{\Delta x}_i = 1/(T-1) \sum_{t=2}^T \Delta x_{i,t}$, and then proceed with the test will not work for the statistic (16), because in the limit as $T \rightarrow \infty$, the stochastic part comes from \mathbf{S}_{00} , which will then sum to zero, and so the statistic converges to a point.

4 Monte Carlo simulations

In this section we compare size and power of our LM-type test with the Fisher-type test proposed by Bai and Ng (2004), the two statistics proposed by Moon and Perron (2004), and the two analogous statistics proposed by Bai and Ng (2010). We first review these tests. If we restrict ourselves to AR(1) idiosyncratic components and a static factor model with intercept but no trend, then the model considered by Bai and Ng (2004, 2010) is

$$\begin{aligned} x_{i,t} &= \mu_i + \lambda_i' \mathbf{f}_t + u_{i,t}, \\ (1-L)\mathbf{f}_t &= \mathbf{C}(L)\mathbf{v}_t, \\ u_{i,t} &= \rho_i u_{i,t-1} + \varepsilon_{i,t}, \end{aligned} \tag{19}$$

where $\mathbf{C}(L)$ has the same properties as in Assumption 2.1, and where the largest eigenvalue of $\boldsymbol{\Sigma}_{\varepsilon\varepsilon} = \text{Var}(\boldsymbol{\varepsilon}_t)$ is bounded, where $\boldsymbol{\varepsilon}_t = (\varepsilon_{1,t}, \varepsilon_{2,t}, \dots, \varepsilon_{N,t})'$. The Fisher-type test in Bai and Ng (2004) is

$$P_{\hat{u}} = \frac{-2 \sum_{i=1}^N \log p_{\hat{u}}^c(i) - 2N}{\sqrt{4N}}, \quad (20)$$

where $p_{\hat{u}}^c(i)$ are p-values of idiosyncratic augmented Dickey-Fuller (ADF) tests. The idiosyncratic components are found by taking first-differences on the panel (19) to find the representation $y_{i,t} = \Delta x_{i,t} = \boldsymbol{\lambda}'_i \mathbf{g}_t + z_{i,t}$, where $\mathbf{g}_t = \Delta \mathbf{f}_t$ and $z_{i,t} = \Delta u_{i,t}$, and then applying the method of principal components to estimate $\mathbf{G} = (\mathbf{g}_1, \mathbf{g}_2, \dots, \mathbf{g}_T)'$ (under the normalization $(T-1)^{-1} \mathbf{G}' \mathbf{G} = \mathbf{I}_r$) as $\sqrt{T-1}$ times the eigenvectors associated with the r largest eigenvalues of the $(T-1) \times (T-1)$ matrix $\mathbf{Y}' \mathbf{Y}$, where $\mathbf{Y} = (\mathbf{y}_1, \mathbf{y}_2, \dots, \mathbf{y}_T)$ was defined in Section 2. This gives the principal component estimator of the factor loadings as $\widehat{\boldsymbol{\Lambda}}_{PC} = (\widehat{\boldsymbol{\lambda}}_1, \widehat{\boldsymbol{\lambda}}_2, \dots, \widehat{\boldsymbol{\lambda}}_N)' = (T-1)^{-1} \mathbf{Y} \widehat{\mathbf{G}}$. The idiosyncratic components, in first differences, are then left as residuals, $\widehat{z}_{i,t} = y_{i,t} - \widehat{\boldsymbol{\lambda}}'_i \widehat{\mathbf{g}}_t$, and the estimated idiosyncratic components in levels are found by re-accumulating, $\widehat{u}_{i,t} = \sum_{s=2}^t \widehat{z}_{i,s}$. Under the null hypothesis $\rho_i = 1$ for all i , the statistic (20) tends to the standard normal distribution as $N, T \rightarrow \infty$.

Moon and Perron (2004) consider the model:

$$\begin{aligned} x_{i,t} &= \mu_i + x_{i,t}^0, \\ x_{i,t}^0 &= \rho_i^0 x_{i,t-1}^0 + \zeta_{i,t}, \\ \zeta_{i,t} &= \boldsymbol{\lambda}'_i \mathbf{f}_t^0 + u_{i,t}^0, \end{aligned} \quad (21)$$

with initial condition $x_{i,0}^0 = 0$. The models (19) and (21) are related. If $x_{i,0} = x_{i,0}^0 = 0$ for all i , and if we impose the restriction $\rho_i = \rho_i^0 = \rho$ for all i , then the models are parametrically equivalent with $(1 - \rho L) \mathbf{f}_t = \mathbf{f}_t^0$ and $(1 - \rho L) u_{i,t} = u_{i,t}^0$. That is, the factors and the idiosyncratic components are restricted to have the same order of integration. If we impose heterogeneity, then the models are only approximately equivalent in the parameters. Also, the respective assumptions made on the processes are somewhat different. In principle, Moon and Perron (2004) make the same assumption for \mathbf{f}_t^0 as Bai and Ng (2004), and we, do for $(1 - L) \mathbf{f}_t$.

Let $\overline{\mathbf{X}} = (\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_N)$ and $\overline{\mathbf{X}}_{-1} = (\mathbf{x}_{-1,1}, \mathbf{x}_{-1,2}, \dots, \mathbf{x}_{-1,N})$ be the $T \times N$ matrices with columns $\mathbf{x}_i = (x_{i,1}, x_{i,2}, \dots, x_{i,T})'$ and $\mathbf{x}_{-1,i} = (x_{i,0}, x_{i,1}, \dots, x_{i,T-1})'$ respectively. Define analogously the $T \times N$ matrix of errors in Equation (21) as $\boldsymbol{\Xi} = (\boldsymbol{\zeta}_1, \boldsymbol{\zeta}_2, \dots, \boldsymbol{\zeta}_N)$, where $\boldsymbol{\zeta}_i = (\zeta_{i,1}, \zeta_{i,2}, \dots, \zeta_{i,T})'$, and consider the estimator $\widehat{\boldsymbol{\Xi}} = \overline{\mathbf{X}} - \widehat{\rho}_{\text{pool}}^0 \overline{\mathbf{X}}_{-1}$, where $\widehat{\rho}_{\text{pool}}^0$ is the pooled autoregressive estimator,

$$\widehat{\rho}_{\text{pool}}^0 = \frac{\text{tr}(\overline{\mathbf{X}}_{-1}' \overline{\mathbf{X}})}{\text{tr}(\overline{\mathbf{X}}_{-1}' \overline{\mathbf{X}}_{-1})}.$$

The errors follow a factor model. As such, apply the method of principal components to estimate $\widehat{\mathbf{F}}^0 = (\mathbf{f}_1^0, \mathbf{f}_2^0, \dots, \mathbf{f}_T^0)'$ (under the normalization $T^{-1} \mathbf{F}^0 \mathbf{F}^0 = \mathbf{I}_r$) as \sqrt{T} times the eigenvectors associated with the r largest eigenvalues of the $T \times T$ matrix $\widehat{\boldsymbol{\Xi}} \widehat{\boldsymbol{\Xi}}'$. This gives the principal component estimator of the factor loadings as $\widehat{\boldsymbol{\Lambda}}_{PC}^0 = (\widehat{\boldsymbol{\lambda}}_1^0, \widehat{\boldsymbol{\lambda}}_2^0, \dots, \widehat{\boldsymbol{\lambda}}_N^0)' = T^{-1} \widehat{\boldsymbol{\Xi}}' \widehat{\mathbf{F}}^0$, from which we define the re-scaled estimated loadings $\widehat{\boldsymbol{\Lambda}}^0 = N^{-1/2} \widehat{\boldsymbol{\Lambda}}_{PC}^0 (\widehat{\boldsymbol{\Lambda}}_{PC}^0)' \widehat{\boldsymbol{\Lambda}}_{PC}^0)^{1/2}$. Next,

let $\mathbf{Q}_{\hat{\lambda}} = \mathbf{I}_N - \mathbf{P}_{\hat{\lambda}}$, where $\mathbf{P}_{\hat{\lambda}} = \hat{\Lambda}^0 (\hat{\Lambda}^{0'} \hat{\Lambda}^0)^{-1} \hat{\Lambda}^{0'}$ is the matrix that projects onto the space orthogonal to $\hat{\Lambda}^0$, and let $\hat{u}_{i,t}^0 = \hat{\Xi} \mathbf{Q}_{\hat{\lambda}}$. Suppose $u_{i,t}^0$ has variance $\sigma_{u^0,i}^2$, long-run variance $\omega_{u^0,i}^2$, and one-sided variance $\zeta_{u^0,i} = (\omega_{u^0,i}^2 - \sigma_{u^0,i}^2)/2$, and define the average long-run variances

$$\omega_{u^0}^2 = \frac{1}{N} \sum_{i=1}^N \omega_{u^0,i}^2, \quad (22)$$

$$\kappa_{u^0}^4 = \frac{1}{N} \sum_{i=1}^N \omega_{u^0,i}^4, \quad (23)$$

$$\zeta_{u^0} = \frac{1}{N} \sum_{i=1}^N \zeta_{u^0,i}. \quad (24)$$

The individual long-run variances are consistently estimated with

$$\hat{\omega}_{u^0,i}^2 = \sum_{j=-T+1}^{T-1} K\left(\frac{j}{h}\right) \hat{\Gamma}_i(j), \quad (25)$$

$$\hat{\zeta}_{u^0,i} = \sum_{j=1}^{T-1} K\left(\frac{j}{h}\right) \hat{\Gamma}_i(j), \quad (26)$$

where $\hat{\Gamma}_i(j) = T^{-1} \sum_t \hat{u}_{i,t}^0 \hat{u}_{i,t+j}^0$ with summation \sum_t defined over $1 \leq t, t+j \leq T$, and $K(\cdot)$ is some kernel function admitting Assumptions 10-13 in Moon and Perron (2004) with bandwidth parameter h . Also, let $\hat{\omega}_{u^0,i}^4 = (\hat{\omega}_{u^0,i}^2)^2$.

Finally, define the pooled (de-factored and autocorrelation-adjusted) autoregressive estimator

$$\hat{\rho}_{\text{pool}}^* = \frac{\text{tr}(\bar{\mathbf{X}}_{-1} \mathbf{Q}_{\hat{\lambda}} \bar{\mathbf{X}}') - NT \zeta_{u^0}}{\text{tr}(\bar{\mathbf{X}}_{-1} \mathbf{Q}_{\hat{\lambda}} \bar{\mathbf{X}}_{-1}')}. \quad (27)$$

Moon and Perron (2004) propose the following two statistics:

$$t_A = \frac{T\sqrt{N}(\hat{\rho}_{\text{pool}}^* - 1)}{\sqrt{2\hat{\kappa}_{u^0}^4 / \hat{\omega}_{u^0}^4}}, \quad (27)$$

$$t_B = T\sqrt{N}(\hat{\rho}_{\text{pool}}^* - 1) \left(\frac{\hat{\omega}_{u^0}}{\hat{\kappa}_{u^0}^2} \right) \sqrt{\frac{1}{T^2 N} \text{tr}(\bar{\mathbf{X}}_{-1} \mathbf{Q}_{\hat{\lambda}} \bar{\mathbf{X}}_{-1}')}, \quad (28)$$

where $\hat{\omega}_{u^0} = \sqrt{\hat{\omega}_{u^0}^2}$, $\hat{\omega}_{u^0}^4 = (\hat{\omega}_{u^0}^2)^2$, and $\hat{\kappa}_{u^0}^2 = \sqrt{\hat{\kappa}_{u^0}^4}$. Under the null hypothesis $\rho_i^0 = 1$ for all i , the statistics (27) and (28) tend to the standard normal distribution as $N, T \rightarrow \infty$ with $N/T \rightarrow 0$.

Based on the procedures of Moon and Perron (2004), Bai and Ng (2010) propose two pooled tests applied to the idiosyncratic terms in the panel (19) which are first estimated by PANIC, $\hat{u}_{i,t} = \sum_{s=2}^T \hat{z}_{i,s}$. Let $\hat{\mathbf{U}}_{-1}$ and $\hat{\mathbf{U}}$ be the $(T-2) \times N$ matrices holding the estimated idiosyncratic components, where $\hat{\mathbf{U}}_{-1}$ and $\hat{\mathbf{U}}$ are defined analogously to $\bar{\mathbf{X}}_{-1}$ and $\bar{\mathbf{X}}$ above (or if the index in levels starts at $t=0$, then $\hat{\mathbf{U}}_{-1}$ and $\hat{\mathbf{U}}$ are $(T-1) \times N$ matrices). Then let $\hat{\rho}^+$ be the bias-adjusted pooled OLS estimator of ρ in $\hat{u}_{i,t} = \rho \hat{u}_{i,t-1} + \varepsilon_{i,t}$,

$$\hat{\rho}^+ = \frac{\text{tr}(\hat{\mathbf{U}}_{-1}' \hat{\mathbf{U}}) - NT \zeta_{\varepsilon}}{\text{tr}(\hat{\mathbf{U}}_{-1}' \hat{\mathbf{U}}_{-1})},$$

where $\widehat{\zeta}_\varepsilon = N^{-1} \sum_{i=1}^N \widehat{\zeta}_{\varepsilon,i}$ is the analog to (24) based on the residuals $\widehat{\varepsilon}_{i,t} = \widehat{u}_{i,t} - \widehat{\rho} \widehat{u}_{i,t-1}$, where $\widehat{\rho} = \text{tr}(\widehat{\mathbf{U}}'_{-1} \widehat{\mathbf{U}}) / \text{tr}(\widehat{\mathbf{U}}'_{-1} \widehat{\mathbf{U}}_{-1})$ is the standard pooled OLS estimator. For the same residuals, define analogously to (22) and (23) the estimators $\widehat{\omega}_\varepsilon^2 = \frac{1}{N} \sum_{i=1}^N \widehat{\omega}_{\varepsilon,i}^2$ and $\widehat{\kappa}_\varepsilon^4 = \frac{1}{N} \sum_{i=1}^N \widehat{\omega}_{\varepsilon,i}^4$. Bai and Ng (2010) propose the following two statistics:

$$P_A = \frac{T\sqrt{N}(\widehat{\rho}^+ - 1)}{\sqrt{2\widehat{\kappa}_\varepsilon^4 / \widehat{\omega}_\varepsilon^4}}, \quad (29)$$

$$P_B = T\sqrt{N}(\widehat{\rho}^+ - 1) \left(\frac{\widehat{\omega}_\varepsilon}{\widehat{\kappa}_\varepsilon^2} \right) \sqrt{\frac{1}{T^2 N} \text{tr}(\widehat{\mathbf{U}}'_{-1} \widehat{\mathbf{U}}_{-1})}, \quad (30)$$

where as before $\widehat{\omega}_\varepsilon = \sqrt{\widehat{\omega}_\varepsilon^2}$, $\widehat{\omega}_\varepsilon^4 = (\widehat{\omega}_\varepsilon^2)^2$, and $\widehat{\kappa}_\varepsilon^2 = \sqrt{\widehat{\kappa}_\varepsilon^4}$. Under the null hypothesis $\rho_i = 1$ for all i , the statistics (29) and (30) tend to the standard normal distribution as $N, T \rightarrow \infty$ with $N/T \rightarrow 0$. Bai and Ng (2010) also propose to pool the univariate tests by Sargan and Bhargava (1983). In the simulation study in the same paper, this pooling is clearly outperformed by the statistics (27),(28),(29) and (30). We therefore do not consider this pooling.

We now proceed with the simulation study. The factor model (3) may be written as

$$x_{i,t} = \mu_i + \sum_{j=1}^r \lambda_{i,j} f_{j,t} + u_{i,t}.$$

Here we set the constants to zero, $\mu_i = 0$, and consider AR(1) dynamics;

$$\begin{aligned} f_{j,t} &= \alpha_j f_{j,t-1} + v_{j,t}, \\ u_{i,t} &= \rho_i u_{i,t-1} + \varepsilon_{i,t}, \end{aligned}$$

where $v_{j,t} \sim \mathcal{N}(0, 1)$, $\varepsilon_{i,t} \sim \mathcal{N}(0, \sigma_{\varepsilon,i}^2)$, $\lambda_{i,j} \sim \mathcal{N}(1, \sigma_\lambda^2)$, and the initial values are set to zero $f_{j,0} = u_{i,0} = 0$. That is, we set the initial value of the panel to zero, $x_{i,0} = 0$, which is included in the sample so that we have $T + 1$ observations on every panel individual $i = 1, 2, \dots, N$. We do this so that the statistics (27) and (28) are well-defined. We apply the following parameter settings, where for each setting we consider one factor ($r = 1$) and three factors ($r = 3$) respectively:

Setting 1: (local power, nonstationary factors),

Rho's: $\rho_i = \rho = 1 - \frac{c}{T\sqrt{N}}$. For size $c = 0$. For power $c = 5$.

Alpha's: $\alpha_j = \alpha = 1$.

Standard deviations: $\sigma_\lambda = 1$, $\sigma_{\varepsilon,i} \sim U(1, 3)$.

Setting 2: (local power, stationary factors),

Rho's: $\rho_i = \rho = 1 - \frac{c}{T\sqrt{N}}$. For size $c = 0$. For power $c = 5$.

Alpha's: For $r = 1$, $\alpha = 0.8$. For $r = 3$, $\alpha_1 = 0.8$, $\alpha_2 = 0.4$, $\alpha_3 = 0.2$.

Standard deviations: $\sigma_\lambda = 1$, $\sigma_{\varepsilon,i} \sim U(1, 3)$.

Setting 3: (local power, locally stationary factors, estimated number of factors)

Rho's: $\rho_i = \rho = 1 - \frac{c}{T\sqrt{N}}$. For size $c = 0$. For power $c = 5$.

Alpha's: $\alpha_j = \alpha = 1 - \frac{c}{T\sqrt{N}}$.

Standard deviations: $\sigma_\lambda = 1$, $\sigma_{\varepsilon,i} \sim U(1, 3)$.

Number of factors (r) is estimated.

Setting 4: Same as Setting 3, except with standard deviations $\sigma_\lambda = 4$, $\sigma_{\varepsilon,i} \sim U(1,3)$.

Setting 5: Same as Setting 3, except under the approximate factor model allowing for idiosyncratic cross-sectional correlation.

The dimensions we consider are $N = 10, 25, 50, 100$ and $T = 20, 40, 80, 160, 320$, where for all settings we look at local power. However, under the alternative hypothesis we set the drift term at $c = 5$, which should be considered a quite rough neighborhood of the unit root, because it starts at $\rho = 0.921$ corresponding to the dimensions $N = 10$ and $T = 20$, and it ends at $\rho = 0.998$ corresponding to the dimensions $N = 100$ and $T = 320$. For the LM-statistic (16) we use the approximate asymptotic distribution (18) with the proposed degrees of freedom $d = \text{tr}(\tilde{\Sigma}_{\varepsilon\varepsilon}\mathbf{S}_{01}^{-1})^2 / \text{tr}(\tilde{\Sigma}_{\varepsilon\varepsilon}\mathbf{S}_{01}^{-1}\tilde{\Sigma}_{\varepsilon\varepsilon}\mathbf{S}_{01}^{-1})$. For small N and large T this distribution is only correct when Misspecification 1 is true. However, we expect it to approximate the true distribution also when the factors are stationary. For large both N and T the asymptotic distribution is standard normal regardless of whether the factors are nonstationary, stationary, or cointegrate, as stated in Theorem 3.2. As suggested by Bai and Ng (2004), for the Fisher-type statistic (20) we size-adjust the individual ADF-tests by first running a simulation in 10,000 replications to find the p-values for finite T . For the estimation of the long-run variances in (25) and (26), which are needed for the statistics (27), (28) (29), and (30), we use the Bartlett kernel. For the Fisher-type statistic (20) we use the right tail of the asymptotic distribution for rejection of the null hypothesis, and for all other statistics we use the left tail of the respective asymptotic distributions. For each statistic this is in the direction of the alternative hypothesis. The nominal size is set to 5%.

In Setting 1 we consider only nonstationary factors. Hence, under this setting Misspecification 1 is true. As a consequence, the Moon and Perron model (21) is not correctly defined under the alternative of stationary idiosyncratic components, because then $\alpha \neq \rho$. We may view this as a misspecification on the statistics (27) and (28). However, the model (21) is correctly defined under the null hypothesis, because then $\alpha = \rho$. In Setting 2 we consider only stationary factors, and thereby break Misspecification 1. In this case the Moon and Perron model (21) is ill-defined both under the null hypothesis and under the alternative hypothesis. For Settings 3, 4 and 5, we estimate the number of factors. As shown by e.g. Moon and Perron (2004) and Gengenbach, Palm, and Urbain (2010), misspecifying the number of factors can have a distinct impact on size and power in small samples. For the LM-statistic (16), and the statistics (20), (29) and (30), we use the IC_{p1} criterion from Bai and Ng (2002). Applied to the differenced data, this criterion will consistently estimate the number of factors r , as $N, T \rightarrow \infty$, by finding, for some $k \in \{1, 2, \dots, r_{\max}\}$,

$$\hat{r} = \arg \min_{1 \leq k \leq r_{\max}} = \ln \left(V(k, \hat{\mathbf{g}}_t^k) \right) + k \left(\frac{N+T}{NT} \right) \ln \left(\frac{NT}{N+T} \right),$$

where $\mathbf{g}_t = \Delta \mathbf{f}_t$, and where $V(r, \hat{\mathbf{g}}^k) = \min_{\Lambda} \frac{1}{NT} \sum_{i=1}^N \sum_{t=1}^T \left(y_{i,t} - \lambda_i^{k'} \hat{\mathbf{g}}_t^k \right)^2$ is the scaled sum of squared residuals with $y_{i,t} = \Delta x_{i,t}$. Moon and Perron (2004) define the analog to this criterion based on the de-factored data by letting $V(k, \hat{\mathbf{g}}_t^k) = \text{tr}(\hat{\mathbf{\Xi}}' \mathbf{Q}_{\hat{\lambda}} \hat{\mathbf{\Xi}}) / NT$, which will consistently estimate the number of factors for the model (21) as $N, T \rightarrow \infty$. This adjusted criterion is

used for the statistics (27) and (28). Note that in Settings 3,4 and 5 all models are correctly defined under both the null hypothesis and the alternative hypothesis, because at all times $\alpha = \rho$. Thus, the dynamic properties of the factors change with the dynamic properties of the idiosyncratic components. Under the null hypothesis the factors are nonstationary and under the alternative hypothesis they are locally stationary. In Setting 4 we have the same settings as in Setting 3 except we increase the variance in the factor loadings. An increase in the importance of the factor loadings was reported by Moon and Perron (2004) to cause size distortions for the statistics (27) and (28). Lastly, in Setting 5 we consider generating data under the approximate factor model by imposing idiosyncratic cross-sectional correlations. Strictly, only the Fisher-type test allows for an approximate factor model, but we expect the other statistics to be somewhat robust against mild cross-sectional correlations in the idiosyncratic component. Let the correlation between the idiosyncratic errors be denoted $\tau_{i,l} = E(\varepsilon_{i,t}\varepsilon_{l,t})$ for $i, l = 1, 2, \dots, N$. Then the largest eigenvalue of $\Sigma_{zz} = \text{Var}(\Delta \mathbf{u}_t) = \Sigma_{\varepsilon\varepsilon}$ is bounded by $\max_i \sum_{l=1}^N |\tau_{i,l}|$ (see e.g. Bai, 2003; Bai and Ng, 2004). So, if for all i , $\sum_{l=1}^N |\tau_{i,l}| < \infty$, then we have an approximate factor model. To impose idiosyncratic cross-sectional correlations we let $\Sigma_{\varepsilon\varepsilon} = \mathbf{A}\mathbf{B}\mathbf{A}$, where $\mathbf{A} = \text{diag}(\sigma_1, \sigma_2, \dots, \sigma_N)$, where as before $\sigma_i \sim U(1,3)$, and where $\mathbf{B} = [b_{i,l}]_{N \times N}$ is a symmetric Toeplitz matrix with elements $b_{i,l} = \delta^{|i-l|}$, $\delta \in (-1,1)$. We then have that $\Sigma_{\varepsilon\varepsilon} = [\tau_{i,l}]_{N \times N}$ with elements $\tau_{i,l} = \delta^{|i-l|}\sigma_i\sigma_l$. This admits the approximate factor model because we have, letting $\sigma_{\max} = \max[\text{dg}(\mathbf{A})]$, for all i that

$$\sum_{l=1}^N |\tau_{i,l}| \leq \sigma_{\max}^2 \sum_{l=1}^N |\delta^{|i-l|}| \ll 2\sigma_{\max}^2 \sum_{l=0}^{N-1} |\delta|^l = 2\sigma_{\max}^2 \frac{1-|\delta|^N}{1-|\delta|} < \infty. \quad (-1 < \delta < 1)$$

Here we set $\delta = 0.2$, i.e. we allow for moderate idiosyncratic cross-sectional correlations, and generate the idiosyncratic errors as $\varepsilon_t \sim \mathcal{N}(\mathbf{0}, \Sigma_{\varepsilon\varepsilon})$.

The results under Setting 1 are presented in Table 1. Here and for all other tables $\bar{\vartheta}$ denotes the infeasible solution to the LM-statistic ϑ where we treat Λ and $\Sigma_{\varepsilon\varepsilon}$ as known. The difference between $\bar{\vartheta}$ and ϑ then points to inefficiencies in finite samples and possible biases resulting from imposing the restrictions under IC3. It seems the proposed LM-statistic is relatively well-behaved in terms of size compared with the other statistics. The statistics P_A , P_B , t_A and t_B all have significant size-distortions when N is small, and the distortions are quite severe when the number of factors is increased to 3. For P_A and t_A , this size-distortion is persistent even when N and T increase. All statistics have size-distortions when N is large and T is small, e.g. when $N = 100$ and $T = 20$. This is somewhat unexpected for the Fisher-type test (which requires that $N, T \rightarrow \infty$ without any restriction on the relation between N and T), but expected for the other statistics, because the LM-statistic is derived sequentially as $(T, N)_s \rightarrow \infty$, and P_A , P_B , t_A and t_B require $N/T \rightarrow 0$. In terms of power the Fisher-type test $P_{\hat{u}}$ is clearly outperformed by all other statistics, who perform similarly for large N and T . For $N \leq 50$, P_A and t_A have the highest local power, but they are also the most oversized. It is then natural to compare size-adjusted power, where for $N \geq 25$ the LM-statistic generally has the highest size-adjusted local power. For large N and T , the LM-statistic also has the highest local power without adjusting for size; for $N = 100$ and $T \geq 80$ when $r = 1$, and for $N = 100$ and $T \geq 40$ when $r = 3$. This is in line with theory, because the LM-statistic is predicted to belong to the most powerful tests in a shrinking neighborhood of the null

Table 2. (Setting 2) Empirical size and local power (%) under stationary factors

N	T	Size							Local power							Size-adj. local power							
		$\bar{\vartheta}$	ϑ	$P_{\bar{u}}$	P_A	P_B	t_A	t_B	$\bar{\vartheta}$	ϑ	$P_{\bar{u}}$	P_A	P_B	t_A	t_B	$\bar{\vartheta}$	ϑ	$P_{\bar{u}}$	P_A	P_B	t_A	t_B	
r = 1	10	20	8.5	5.2	7.7	11.9	6.8	14.5	9.1	65.0	50.8	39.7	73.7	60.4	88.7	79.3	52.3	49.7	30.5	51.2	52.9	66.7	66.3
		40	6.9	5.1	7.6	12.7	7.1	14.4	8.6	69.7	65.2	46.5	85.1	73.4	91.1	81.9	62.7	65.0	38.1	62.9	64.6	69.6	69.6
		80	6.6	5.4	6.3	13.7	7.9	14.7	7.9	73.4	71.4	47.3	89.5	79.8	91.9	83.1	66.0	69.7	41.2	68.1	68.4	73.9	74.3
		160	5.9	5.4	7.0	13.9	7.9	14.3	7.9	76.0	74.5	49.6	91.8	82.1	92.8	84.3	71.5	72.5	43.6	72.8	73.3	73.5	74.9
	320	6.2	5.6	6.3	14.8	7.7	14.6	7.9	77.1	75.8	50.1	92.7	83.5	93.2	84.4	72.4	73.4	45.2	75.1	74.7	76.0	75.4	
	25	20	5.7	3.3	6.2	6.6	4.0	9.9	6.1	77.3	66.6	45.4	76.5	67.5	88.7	82.8	74.4	74.4	41.7	71.1	71.5	79.3	80.0
		40	5.7	4.2	5.5	8.2	5.2	10.0	6.2	82.2	79.1	49.7	85.3	77.8	90.3	84.0	80.0	82.1	47.5	77.5	76.8	80.7	80.9
		80	5.2	4.2	5.2	8.9	5.6	9.7	6.4	85.7	83.7	48.0	89.5	83.3	91.5	85.9	84.9	86.2	47.4	80.6	80.9	82.2	82.6
		160	4.5	4.3	5.8	8.7	5.3	9.5	5.7	86.0	85.3	52.4	90.5	84.6	91.2	86.0	87.5	87.3	48.9	84.5	83.9	84.0	84.1
	320	5.5	5.2	5.0	10.0	6.6	10.4	6.8	87.5	87.4	51.3	91.6	86.4	92.3	86.8	86.2	86.8	50.8	82.5	81.9	82.7	82.9	
	50	20	5.9	2.6	5.5	4.9	3.4	10.2	7.2	85.0	72.4	45.2	75.2	68.7	89.8	85.4	82.9	83.2	43.3	75.6	75.6	80.5	81.2
		40	5.6	3.6	5.2	6.2	4.1	9.4	6.3	87.9	83.9	49.0	84.1	79.0	90.9	87.3	86.3	87.0	48.4	81.2	81.6	84.6	85.1
80		5.3	4.3	4.1	7.5	5.1	8.6	6.1	90.3	88.9	47.4	89.1	84.4	92.2	87.9	89.8	90.6	52.3	84.0	83.7	85.1	85.1	
160		4.9	4.3	5.3	6.9	4.9	7.8	5.5	91.0	90.5	53.0	90.3	86.2	91.5	87.8	91.2	91.9	51.2	86.9	86.5	86.3	87.1	
320	5.2	5.1	4.6	8.4	5.9	8.7	6.3	91.9	91.6	50.4	91.4	88.0	92.2	88.7	91.6	91.5	53.0	85.8	86.1	86.2	86.5		
100	20	5.8	2.3	4.7	3.9	2.9	10.5	8.4	87.7	73.3	45.8	74.4	70.1	91.3	88.6	85.7	84.9	47.5	77.9	78.2	82.4	82.2	
	40	4.9	3.1	4.8	5.2	4.1	8.4	6.8	91.1	86.5	50.4	84.5	81.0	91.5	89.7	91.2	91.7	51.1	84.0	84.3	86.4	86.2	
	80	5.1	4.1	3.6	6.2	4.7	7.8	5.8	92.6	90.4	46.8	88.4	85.2	92.0	89.1	92.3	92.5	55.9	85.8	85.9	87.5	87.9	
	160	4.4	4.0	5.0	6.4	4.6	7.1	5.5	93.8	93.3	51.1	90.2	87.8	92.0	89.6	94.8	94.7	51.4	88.1	88.3	88.5	88.6	
320	5.3	4.7	3.5	7.7	5.9	8.3	6.4	94.0	93.8	49.4	91.6	88.8	92.5	89.9	93.9	94.3	55.9	87.0	86.8	87.4	87.1		
r = 3	10	20	12.5	4.1	6.8	14.9	9.6	59.8	51.1	68.2	32.8	29.5	64.7	51.3	98.7	96.9	44.8	36.2	23.8	33.9	34.6	40.4	36.4
		40	11.6	5.3	8.1	17.6	11.6	46.6	36.6	73.8	52.4	42.0	80.8	70.0	98.1	95.8	50.2	50.3	32.3	46.7	47.5	51.6	48.6
		80	11.0	6.3	8.1	18.5	12.5	35.0	25.3	77.2	63.9	46.9	88.2	78.1	97.5	93.6	59.2	57.3	34.2	53.2	53.5	56.3	55.7
		160	10.3	6.9	9.5	19.0	11.7	27.0	18.5	79.0	67.2	52.3	88.9	79.8	95.2	89.3	57.1	59.1	35.5	56.1	56.0	57.7	58.2
	320	10.4	7.0	8.9	18.9	11.8	22.7	15.0	80.5	70.4	51.2	90.4	81.4	93.9	86.8	61.9	59.5	36.8	58.4	59.5	59.3	58.6	
	25	20	8.5	2.1	3.4	6.6	4.3	55.3	48.5	77.2	44.7	44.7	65.1	56.7	99.1	97.7	66.1	63.2	34.8	59.4	59.9	30.1	24.0
		40	7.0	3.1	4.6	9.1	6.2	31.1	22.5	83.2	70.6	70.6	82.8	75.4	98.2	95.9	76.6	78.9	44.0	71.5	72.0	68.2	68.6
		80	6.1	4.2	5.1	9.8	6.2	20.5	14.0	84.7	79.0	79.0	88.9	82.5	96.6	93.4	81.7	82.7	44.5	79.3	79.4	79.2	80.0
		160	6.2	4.5	5.7	10.6	7.0	15.5	11.0	85.5	82.7	82.7	90.2	83.9	94.1	90.3	82.2	83.9	49.0	79.5	79.1	80.0	79.9
	320	6.0	5.1	5.6	11.2	7.6	13.8	9.3	86.0	83.6	50.1	91.3	85.6	93.5	88.8	83.4	83.2	46.8	78.5	79.1	78.9	79.3	
	50	20	7.1	1.0	2.3	3.6	2.6	68.4	62.6	83.5	45.3	45.3	62.9	55.9	99.5	98.7	77.6	73.3	39.2	69.1	69.5	16.4	13.5
		40	6.2	2.5	2.6	5.8	3.8	37.4	29.5	88.1	75.8	75.8	81.7	76.3	99.1	97.6	85.3	86.4	50.7	80.0	79.4	71.6	72.6
80		5.2	3.4	3.0	6.7	4.7	21.4	15.5	90.0	85.2	85.2	88.4	83.8	97.5	95.7	89.4	89.5	54.6	84.9	84.7	84.1	84.3	
160		5.6	4.4	4.6	8.0	5.9	14.2	10.6	90.3	88.2	88.2	89.4	85.2	95.4	92.9	88.9	89.5	51.8	88.3	83.3	83.2	83.4	
320	5.3	4.5	4.3	8.7	6.4	12.0	8.4	90.5	89.4	49.0	89.9	85.8	93.4	89.8	89.6	90.5	51.7	83.4	83.0	83.7	83.8		
100	20	6.2	0.3	1.1	2.1	1.6	81.5	77.3	87.4	38.7	38.7	56.0	53.6	99.8	99.5	84.4	82.2	41.4	73.5	73.4	14.0	11.7	
	40	5.6	1.3	2.1	4.0	3.1	48.8	41.8	91.2	75.5	75.5	80.8	77.0	99.5	99.2	90.2	89.8	51.0	84.4	84.1	67.9	69.5	
	80	5.2	2.9	2.3	5.5	4.1	25.9	20.6	92.9	87.3	87.3	88.2	85.0	98.5	97.4	92.5	92.9	52.4	87.2	87.4	85.2	86.5	
	160	5.0	3.8	4.1	6.8	5.2	16.0	12.6	93.7	91.3	91.3	89.7	87.0	96.8	95.3	93.7	93.2	51.7	86.7	86.5	86.0	85.8	
320	5.5	5.0	3.8	7.4	5.7	11.8	9.1	93.1	91.8	48.5	91.0	88.2	94.6	92.8	92.2	92.0	53.1	87.1	87.3	86.7	86.7		

Note: The data is generated as $u_{i,t} = \rho_i u_{i,t-1} + \varepsilon_{i,t}$, $\rho_i = \rho = 1 - \frac{5}{T\sqrt{N}}$, $\varepsilon_{i,t} \sim \mathcal{N}(0, \sigma_{\varepsilon_i}^2)$, $\sigma_{\varepsilon_i} \sim U(1, 3)$, $\lambda_i \sim \mathcal{N}_r(\mathbf{1}, \mathbf{I})$ (standard deviations and loadings are generated once and then kept fixed), $f_{j,t} = \alpha_j f_{j,t-1} + v_{j,t}$, $v_t \sim \mathcal{N}(0, 1)$, where; for $r = 1$, $\alpha = 0.8$; and for $r = 3$, $\alpha_1 = 0.8, \alpha_2 = 0.6, \alpha_3 = 0.4$. The replication number is 5,000.

infeasible statistic and the feasible statistic are over-sized when N is small. However, except for $P_{\bar{u}}$, the size-distortion for ϑ is not worse than for any of the other statistics. As N and T grow the size-distortions for $\bar{\vartheta}$ and ϑ vanish, supporting Theorem 3.2, stating that the LM-statistic is asymptotically robust against stationarity in the factors. This time, when N is large, say $N = 100$, and T is small, say $T = 20$, all statistics are under-sized, except t_A and t_B . Again, P_A is not well-behaved in terms of size even for large dimensions. If we look at size-adjusted power, then the LM-statistic generally has the highest local power for $N \geq 25$.

Table 4. (Setting 4) Empirical size and local power (%) under locally stationary factors, number of factors estimated, and increased variance of factor loadings

N	T	Size							Local power							Size-adj. local power							
		$\bar{\vartheta}$	ϑ	$P_{\bar{u}}$	P_A	P_B	t_A	t_B	$\bar{\vartheta}$	ϑ	$P_{\bar{u}}$	P_A	P_B	t_A	t_B	$\bar{\vartheta}$	ϑ	$P_{\bar{u}}$	P_A	P_B	t_A	t_B	
r = 1	10	20	7.5	3.5	10.7	19.1	12.9	44.8	40.2	61.2	14.2	19.4	37.9	28.4	84.2	78.8	51.4	18.2	9.6	13.3	13.1	23.7	22.3
		40	5.9	3.6	11.3	19.4	12.8	39.8	32.1	64.1	31.1	32.1	61.4	48.4	87.6	79.5	60.4	37.8	16.8	24.9	26.8	27.3	26.1
		80	5.3	3.7	8.8	19.3	12.2	30.5	21.8	66.9	43.6	37.4	74.5	60.3	88.8	79.7	66.1	48.9	24.6	34.1	35.3	36.5	37.9
		160	4.7	4.4	9.5	17.6	10.8	23.2	14.9	69.0	50.6	44.3	80.9	67.0	88.9	77.5	70.4	52.8	29.5	43.2	45.3	45.7	47.6
	320	4.6	4.3	9.1	16.4	9.3	18.8	10.5	71.4	55.3	46.1	85.2	72.1	88.8	77.8	74.0	58.5	30.3	56.0	57.2	57.9	57.3	
	25	20	5.6	4.8	8.0	10.0	6.7	45.1	43.0	76.9	49.6	36.9	62.0	52.6	88.0	84.0	74.9	50.6	27.5	46.8	47.0	17.1	16.0
		40	5.6	4.8	5.9	8.7	5.7	26.7	22.7	81.8	77.6	50.1	84.5	76.9	93.1	88.0	79.5	77.9	46.2	74.9	74.4	11.0	10.4
		80	5.1	4.3	5.5	9.0	5.9	13.6	9.3	85.6	83.4	48.3	89.2	82.5	92.7	87.4	85.1	85.0	45.5	78.9	79.5	67.9	69.3
		160	4.7	4.6	6.0	8.7	5.5	10.2	6.5	85.8	85.2	52.3	90.5	84.2	92.6	86.8	86.9	86.3	48.2	83.3	83.2	84.3	84.3
	320	5.2	5.3	5.1	9.9	6.4	10.2	6.7	87.2	87.1	50.6	91.6	86.0	92.9	87.1	86.8	86.7	50.3	82.0	81.3	82.8	81.7	
	50	20	5.8	5.6	7.5	8.4	6.2	39.5	37.5	84.9	73.0	47.8	75.6	68.9	88.7	85.1	83.0	70.7	38.7	63.9	63.4	14.8	12.9
		40	5.6	4.7	6.2	7.2	5.3	27.0	24.5	87.6	84.5	51.5	84.8	79.1	92.3	88.9	86.2	85.4	47.1	78.4	78.0	11.7	10.6
		80	5.4	4.6	4.8	7.7	5.4	14.9	11.8	90.6	88.8	49.3	89.0	84.3	93.3	89.8	89.9	89.9	50.3	83.0	82.7	49.1	52.4
		160	5.0	4.5	5.6	7.4	4.9	8.9	6.2	91.1	90.6	53.6	90.3	86.0	92.7	89.2	91.2	91.7	51.0	86.4	86.3	86.5	86.3
	320	5.3	5.3	4.4	8.4	5.9	9.2	6.4	91.6	91.6	51.0	91.5	87.6	92.7	89.5	91.2	90.8	53.7	85.8	85.0	87.1	87.0	
	100	20	6.0	7.2	8.5	8.8	7.5	39.2	37.5	87.7	76.5	50.8	76.1	72.2	86.5	84.5	85.5	68.0	38.8	63.0	62.9	14.4	11.3
		40	4.8	4.9	6.1	6.6	5.3	28.3	26.5	91.1	87.7	54.0	85.4	81.7	92.4	89.8	91.5	87.8	49.3	80.2	80.6	11.8	10.2
		80	5.0	4.6	4.2	7.0	5.4	17.0	14.7	92.6	90.8	48.6	88.4	85.2	92.8	90.6	92.5	91.3	52.4	84.3	84.3	6.7	6.6
		160	4.4	4.5	4.9	6.6	5.0	9.7	7.5	93.9	93.3	52.5	90.5	87.8	93.2	91.0	94.8	94.0	53.1	87.8	87.8	85.1	85.7
	320	5.1	5.0	3.9	7.8	6.0	8.7	6.7	94.0	93.8	49.9	91.8	89.0	93.1	90.9	94.0	93.8	54.8	87.0	87.1	87.8	87.8	
r = 3	10	20	7.1	3.8	11.3	20.4	14.2	36.3	30.5	52.5	15.1	23.2	44.9	34.2	86.2	79.3	44.1	18.5	11.4	14.2	14.1	25.0	24.0
		40	6.6	4.0	9.8	21.0	14.6	32.5	24.9	54.1	31.6	33.8	67.5	55.9	88.8	80.8	45.7	36.0	21.3	27.7	27.4	30.3	29.1
		80	4.8	4.1	9.3	21.9	14.9	28.6	20.2	54.9	40.5	39.7	78.4	66.5	89.5	81.1	56.2	43.7	24.9	36.1	36.9	37.4	37.8
		160	5.0	4.8	10.6	21.9	15.5	25.3	18.0	53.9	45.2	43.3	82.1	71.4	88.1	79.2	53.8	46.4	27.8	39.6	41.0	45.3	45.2
	320	4.8	4.9	10.1	21.6	14.9	23.0	16.0	55.0	46.6	44.8	85.0	73.9	87.8	78.6	56.5	47.4	28.1	41.7	43.4	44.2	45.5	
	25	20	7.0	6.2	8.8	13.3	9.8	35.7	33.1	74.7	28.7	25.5	45.7	37.1	84.6	80.7	66.4	23.2	16.2	23.9	24.0	12.6	11.8
		40	6.2	5.4	7.5	11.9	7.7	23.7	19.5	80.2	66.9	44.4	78.9	70.5	91.6	86.5	76.4	65.2	35.1	60.8	60.3	8.3	7.7
		80	5.3	5.0	5.7	11.3	7.1	14.7	10.0	81.8	77.2	47.0	87.1	80.4	92.5	87.2	80.8	77.0	44.0	75.3	75.1	70.0	70.1
		160	5.0	4.9	6.0	10.8	7.3	11.7	8.0	82.6	80.9	42.5	89.3	82.9	91.9	86.2	82.6	81.1	48.6	76.6	76.6	80.5	80.0
	320	5.2	5.0	5.8	11.3	7.4	11.8	7.7	83.5	82.4	49.9	90.6	84.3	92.2	86.5	83.4	82.3	46.9	77.8	77.0	80.4	80.3	
	50	20	6.7	9.1	9.8	12.3	9.7	31.3	29.2	82.8	50.1	35.1	58.6	51.5	84.6	81.8	78.0	33.0	21.9	36.8	36.0	10.2	9.5
		40	5.9	5.3	6.8	9.6	6.9	20.0	16.9	87.2	76.5	47.0	80.1	74.1	90.7	86.9	84.8	75.4	39.2	67.7	68.8	6.9	6.5
		80	5.2	5.2	5.0	8.5	6.2	13.1	10.4	89.2	85.4	48.3	87.3	82.6	92.6	89.2	88.9	84.9	48.6	79.9	79.4	69.1	69.5
		160	5.5	5.1	5.7	8.9	6.3	10.0	7.2	89.6	87.7	52.1	89.0	84.7	92.1	88.3	88.7	87.5	49.0	81.1	81.4	82.7	82.9
	320	5.0	4.9	4.8	8.8	6.4	9.1	6.6	89.9	89.2	50.4	89.4	85.5	91.4	87.4	89.9	89.5	51.8	82.6	81.6	83.7	83.4	
	100	20	6.3	12.5	11.1	13.3	11.8	32.0	30.7	87.2	56.3	38.3	59.7	55.2	81.9	80.0	84.8	24.1	21.4	30.6	30.3	10.0	9.2
		40	5.8	6.6	7.0	9.0	7.0	21.0	18.8	91.0	80.7	47.4	80.1	76.2	89.5	87.2	89.8	75.2	39.8	69.2	69.4	7.2	6.9
		80	5.0	5.4	4.4	7.7	6.1	13.9	12.2	92.9	88.2	46.6	87.5	83.9	92.3	90.2	92.9	86.7	48.5	81.2	81.1	44.1	44.4
		160	5.0	5.0	5.8	7.6	6.0	10.4	8.2	93.5	91.8	51.9	89.9	86.6	92.8	89.9	93.7	91.9	49.8	83.9	83.8	82.8	83.0
	320	5.5	5.3	4.7	7.7	6.0	8.6	6.5	92.9	92.1	50.8	90.8	87.9	92.3	89.7	92.2	91.6	51.5	85.4	85.5	86.6	86.4	

Note: The data is generated as $u_{i,t} = \rho_i u_{i,t-1} + \varepsilon_{i,t}$, $\rho_i = \rho = 1 - \frac{5}{T\sqrt{N}}$, $\varepsilon_{i,t} \sim \mathcal{N}(0, \sigma_{\varepsilon_i}^2)$, $\sigma_{\varepsilon_i} \sim U(1, 3)$, $\lambda_i \sim \mathcal{N}_r(1, \sigma_{\lambda_i}^2 \mathbf{I})$, $\sigma_{\lambda_i} = 4$ (standard deviations and loadings are generated once and then kept fixed), $f_{j,t} = \alpha_j f_{j,t-1} + v_{j,t}$, $v_{j,t} \sim \mathcal{N}(0, 1)$, $\alpha_j = \alpha = \rho$. The replication number is 5,000.

the difference between the size reported in Table 3 and the size reported in Table 4, is that the variance in the factor loadings is increased. The increased size-distortion for t_A and t_B reported in Table 4 is remarkable for $N \leq 25$, suggesting that these statistics are not well-behaved when the variance coming from the common component dominates the variance coming from the idiosyncratic component. For the other statistics the increase in variance in the factor loadings does not have the same impact. When adjusted for size, the LM-statistic again has the highest local power as N and T grow.

perform not specifically worse than under the exact factor model. Again, the LM-statistic has the highest size-adjusted local power for large enough N and T . Though not reported here, we also tried $\delta = 0.5$. The size was then increased for all statistics with size-distortions that were more persistent as the panel dimensions grow. This suggests that further research is needed to investigate what dimensions of the panel that are needed in practice to allow for an approximate factor model.

5 Conclusions

In this paper we derive a homogenous LM-type test for idiosyncratic unit roots in the exact factor model based on the likelihood function of the differenced data. The test allows for idiosyncratic cross-sectional heteroscedasticity, which is consistently estimated with quasi maximum likelihood. By using sequential limits we first derive an intermediate asymptotic distribution for fixed N as T tends to infinity, which is a weighted chi-square distribution, and then show that as N is subsequently passed to infinity the asymptotic distribution is standard normal. The statistic is derived under the misspecification of nonstationary factors, which is shown to be irrelevant asymptotically.

In a simulation study we compare size and local power of the LM-statistic with the Fisher-type test proposed by Bai and Ng (2004), the two statistics in Moon and Perron (2004), and the analogous statistics proposed by Bai and Ng (2010). The results show that the Fisher-type test has by far the lowest local power while the other statistics have comparable local power as N and T grow. However, the LM-statistic is more stable in terms of size. The two statistics of Moon and Perron (2004) are sensitive to misspecifications, and tend to be over-sized when the number of factors is estimated and when the variance coming from the common component dominates the variance coming from the idiosyncratic components. The analogous statistics of Bai and Ng (2010) also tend to be over-sized, but to a less degree. Adjusting for size, the LM-statistic has the highest local power as the panel dimensions increase. The simulation results also suggest that the statistics considered here allow for some mild idiosyncratic cross-section correlation without seriously influencing size, although strictly, only the Fisher-type test is designed for the approximate factor model.

Acknowledgements

The authors wish to thank Rolf Larsson, Johan Lyhagen, and the participants at the seminar series at the Department of Quantitative Economics, Maastricht University, for the valuable discussions. Financial support from the Jan Wallander and Tom Hedelius Foundation through the research grants P2005-0117:1 and P2009-0189:1 is gratefully acknowledged, as well as a travel grant for the second author from Thun's Foundation.

Mathematical appendix

A1 Score and information

Let l be the log-likelihood (10) with $\Sigma = E(\mathbf{Y}_v \mathbf{Y}_v')$, and let $\tilde{\theta} = (\tilde{\Lambda}'_v, \tilde{\sigma}', 1)'$ be the restricted MLE where $\tilde{\Lambda}$ and $\tilde{\sigma}$ are MLEs and $\rho = 1$ is the restriction under the null hypothesis. Also, in accordance with the notation in Section 3, let $\tilde{\Sigma}$ and $\tilde{\Omega}$ denote $\Sigma(\tilde{\theta})$ and $\Omega(\tilde{\theta})$, where $\tilde{\Sigma}^{-1} = (\mathbf{I}_{T^*} \otimes \tilde{\Omega}^{-1})$. By definition, the full score vector evaluated at $\theta = \tilde{\theta}$ is

$$\tilde{\mathbf{V}}_{\theta} = \left[\left(\frac{\partial l}{\partial \text{vec} \Lambda} \right)', \left(\frac{\partial l}{\partial \sigma'} \right)', \frac{\partial l}{\partial \rho} \right]'_{\theta=\tilde{\theta}} = [\mathbf{0}, \tilde{V}_{\rho}]'$$

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

$$\tilde{V}_{\rho} = \frac{\partial \Sigma}{\partial \rho} \Big|_{\theta=\tilde{\theta}} = -\frac{1}{2} \text{tr} \left(\frac{\partial \Sigma}{\partial \rho} \Big|_{\theta=\tilde{\theta}} \tilde{\Sigma}^{-1} \right) + \frac{1}{2} \text{tr} \left(\frac{\partial \Sigma}{\partial \rho} \Big|_{\theta=\tilde{\theta}} \tilde{\Sigma}^{-1} \mathbf{Y}_v \mathbf{Y}_v' \tilde{\Sigma}^{-1} \right), \quad (\text{A1})$$

where $\frac{\partial \Sigma}{\partial \rho} \Big|_{\theta=\tilde{\theta}} = \left(\frac{\partial \Psi}{\partial \rho} \Big|_{\theta=\tilde{\theta}} \otimes \tilde{\Sigma}_{\varepsilon\varepsilon} \right)$, and where it is straightforward to show from (4) that $\frac{\partial \Psi}{\partial \rho} \Big|_{\theta=\tilde{\theta}} = \frac{1}{2} \mathbf{1}\mathbf{1}' - \mathbf{I}_{T^*}$, where $\mathbf{1}$ is a $T^* \times 1$ vector of ones. Thus, using some common results for the trace and the inverse of Kronecker products (see e.g. Magnus and Neudecker, 2001, pp. 28-31), the first part of (A1) is

$$-\frac{1}{2} \text{tr} \left(\frac{\partial \Sigma}{\partial \rho} \Big|_{\theta=\tilde{\theta}} \tilde{\Sigma}^{-1} \right) = -\frac{1}{2} \text{tr} \left[\left(\frac{1}{2} \mathbf{1}\mathbf{1}' - \mathbf{I}_{T^*} \right) \otimes \tilde{\Sigma}_{\varepsilon\varepsilon} \tilde{\Omega}^{-1} \right] = \frac{T^*}{4} \text{tr} \left(\tilde{\Sigma}_{\varepsilon\varepsilon} \tilde{\Omega}^{-1} \right).$$

Similarly, the second part of (A1) is

$$\begin{aligned} \frac{1}{2} \text{tr} \left(\frac{\partial \Sigma}{\partial \rho} \Big|_{\theta=\tilde{\theta}} \tilde{\Sigma}^{-1} \mathbf{Y}_v \mathbf{Y}_v' \tilde{\Sigma}^{-1} \right) &= \frac{1}{2} \text{tr} \left\{ \left[\left(\frac{1}{2} \mathbf{1}\mathbf{1}' - \mathbf{I}_{T^*} \right) \otimes \tilde{\Sigma}_{\varepsilon\varepsilon} \tilde{\Omega}^{-1} \right] \mathbf{Y}_v \mathbf{Y}_v' \left(\mathbf{I}_{T^*} \otimes \tilde{\Omega}^{-1} \right) \right\} \\ &= \frac{1}{4} \text{tr} \left[\left(\mathbf{1}\mathbf{1}' \otimes \mathbf{I}_N \right) \left(\mathbf{I}_{T^*} \otimes \tilde{\Sigma}_{\varepsilon\varepsilon} \tilde{\Omega}^{-1} \right) \mathbf{Y}_v \mathbf{Y}_v' \left(\mathbf{I}_{T^*} \otimes \tilde{\Omega}^{-1} \right) \right] \\ &\quad - \frac{1}{2} \text{tr} \left[\left(\mathbf{I}_{T^*} \otimes \tilde{\Sigma}_{\varepsilon\varepsilon} \tilde{\Omega}^{-1} \right) \mathbf{Y}_v \mathbf{Y}_v' \left(\mathbf{I}_{T^*} \otimes \tilde{\Omega}^{-1} \right) \right], \end{aligned}$$

where, using the definition of the sample covariances in Section 2,

$$\left(\mathbf{I}_{T^*} \otimes \tilde{\Sigma}_{\varepsilon\varepsilon} \tilde{\Omega}^{-1} \right) \mathbf{Y}_v \mathbf{Y}_v' \left(\mathbf{I}_{T^*} \otimes \tilde{\Omega}^{-1} \right) = \begin{pmatrix} \tilde{\Sigma}_{\varepsilon\varepsilon} \tilde{\Omega}^{-1} \mathbf{s}_{22} \tilde{\Omega}^{-1} & \cdots & \tilde{\Sigma}_{\varepsilon\varepsilon} \tilde{\Omega}^{-1} \mathbf{s}_{2T} \tilde{\Omega}^{-1} \\ \vdots & \ddots & \vdots \\ \tilde{\Sigma}_{\varepsilon\varepsilon} \tilde{\Omega}^{-1} \mathbf{s}_{T2} \tilde{\Omega}^{-1} & \cdots & \tilde{\Sigma}_{\varepsilon\varepsilon} \tilde{\Omega}^{-1} \mathbf{s}_{TT} \tilde{\Omega}^{-1} \end{pmatrix},$$

from which it follows that

$$\frac{1}{2} \text{tr} \left(\frac{\partial \Sigma}{\partial \rho} \Big|_{\theta=\tilde{\theta}} \tilde{\Sigma}^{-1} \mathbf{Y}_v \mathbf{Y}_v' \tilde{\Sigma}^{-1} \right) = \frac{1}{4} \text{tr} \left(\tilde{\Sigma}_{\varepsilon\varepsilon} \tilde{\Omega}^{-1} \sum_{t=2}^T \sum_{s=2}^T \mathbf{s}_{t,s} \tilde{\Omega}^{-1} \right) - \frac{1}{2} \text{tr} \left(\tilde{\Sigma}_{\varepsilon\varepsilon} \tilde{\Omega}^{-1} \sum_{t=2}^T \mathbf{s}_{t,t} \tilde{\Omega}^{-1} \right).$$

Put together, noting that $\tilde{\Omega} = \mathbf{S}_{01}$, we have that

$$\tilde{V}_{\rho} = \frac{1}{4} \text{tr} \left(\tilde{\Sigma}_{\varepsilon\varepsilon} \mathbf{S}_{01}^{-1} \mathbf{S}_{00} \mathbf{S}_{01}^{-1} \right) - \frac{1}{2} \text{tr} \left(\tilde{\Sigma}_{\varepsilon\varepsilon} \mathbf{S}_{01}^{-1} \mathbf{S}_0 \mathbf{S}_{01}^{-1} \right) + \frac{T^*}{4} \text{tr} \left(\tilde{\Sigma}_{\varepsilon\varepsilon} \mathbf{S}_{01}^{-1} \right). \quad (\text{A2})$$

Further, let the information matrix be partitioned as in (14). Using another standard result (see e.g Harville, 1977) we have that

$$\begin{aligned}\tilde{J}_{\rho\rho} &= -E \left(\frac{\partial^2 l}{\partial \rho \partial \rho} \Big|_{\theta=\tilde{\theta}} \right) = \frac{1}{2} \text{tr} \left(\frac{\partial \Sigma}{\partial \rho} \Big|_{\theta=\tilde{\theta}} \tilde{\Sigma}^{-1} \frac{\partial \Sigma}{\partial \rho} \Big|_{\theta=\tilde{\theta}} \tilde{\Sigma}^{-1} \right) \\ &= \frac{1}{2} \text{tr} \left[\left(\frac{1}{2} \mathbf{1}\mathbf{1}' - \mathbf{I}_{T^*} \right)^2 \otimes \left(\tilde{\Sigma}_{\varepsilon\varepsilon} \tilde{\Omega}^{-1} \right)^2 \right] \\ &= \frac{1}{8} T^{*2} \text{tr} \left(\tilde{\Sigma}_{\varepsilon\varepsilon} \mathbf{S}_{01}^{-1} \tilde{\Sigma}_{\varepsilon\varepsilon} \mathbf{S}_{01}^{-1} \right).\end{aligned}\tag{A3}$$

For the remaining parts of the information matrix, let \mathbf{e}_i be an $N \times 1$ vector for which the i th element equals 1 and all other elements equal 0, and define $\Gamma_{i,j} = \frac{\partial \Lambda \Lambda'}{\partial \lambda_{i,j}} = \lambda_j \mathbf{e}'_i + \mathbf{e}_i \lambda'_j$, where $\lambda_j = (\lambda_{1,j}, \lambda_{2,j}, \dots, \lambda_{N,j})'$ is the j th column vector of $\Lambda = (\lambda_1, \lambda_2, \dots, \lambda_r) = [\lambda_{i,j}]_{N \times r}$. We then have that

$$\frac{\partial \Sigma}{\partial \sigma_{\varepsilon,i}^2} \Big|_{\theta=\tilde{\theta}} = (\Psi(1) \otimes \mathbf{e}_i \mathbf{e}'_i) = (\mathbf{I}_{T^*} \otimes \mathbf{e}_i \mathbf{e}'_i),$$

and

$$\frac{\partial \Sigma}{\partial \lambda_{i,j}} \Big|_{\theta=\tilde{\theta}} = (\Psi(1) \otimes \tilde{\Gamma}_{i,j}) = (\mathbf{I}_{T^*} \otimes \tilde{\Gamma}_{i,j}),$$

from which it follows, using analogously the relation (A3), that for $i, l = 1, 2, \dots, N$ and $j, q = 1, 2, \dots, r$,

$$\tilde{J}_{\sigma_{\varepsilon,i}^2 \sigma_{\varepsilon,l}^2} = \tilde{J}_{\sigma_{\varepsilon,l}^2 \sigma_{\varepsilon,i}^2} = \frac{T^*}{2} \text{tr} \left(\mathbf{e}_i \mathbf{e}'_i \mathbf{S}_{01}^{-1} \mathbf{e}_l \mathbf{e}'_l \mathbf{S}_{01}^{-1} \right),\tag{A4}$$

$$\tilde{J}_{\rho \sigma_{\varepsilon,i}^2} = \tilde{J}_{\sigma_{\varepsilon,i}^2 \rho} = -\frac{T^*}{4} \text{tr} \left(\tilde{\Sigma}_{\varepsilon\varepsilon} \mathbf{S}_{01}^{-1} \mathbf{e}_i \mathbf{e}'_i \mathbf{S}_{01}^{-1} \right),\tag{A5}$$

$$\tilde{J}_{\lambda_{i,j} \lambda_{l,q}} = \tilde{J}_{\lambda_{l,q} \lambda_{i,j}} = \frac{T^*}{2} \text{tr} \left(\tilde{\Gamma}_{i,j} \mathbf{S}_{01}^{-1} \tilde{\Gamma}_{l,q} \mathbf{S}_{01}^{-1} \right),\tag{A6}$$

$$\tilde{J}_{\lambda_{i,j} \sigma_{\varepsilon,l}^2} = \tilde{J}_{\sigma_{\varepsilon,l}^2 \lambda_{i,j}} = \frac{T^*}{2} \text{tr} \left(\tilde{\Gamma}_{i,j} \mathbf{S}_{01}^{-1} \mathbf{e}_l \mathbf{e}'_l \mathbf{S}_{01}^{-1} \right),\tag{A7}$$

$$\tilde{J}_{\lambda_{i,j} \rho} = \tilde{J}_{\rho \lambda_{i,j}} = -\frac{T^*}{4} \text{tr} \left(\tilde{\Gamma}_{i,j} \mathbf{S}_{01}^{-1} \tilde{\Sigma}_{\varepsilon\varepsilon} \mathbf{S}_{01}^{-1} \right).\tag{A8}$$

A2 Proofs of theorems

We will make repeated use of the Cauchy-Schwarz trace inequality, which can be found in e.g. Magnus and Neudecker (2001, p. 201, in quadratic form); For $\mathbf{A}, \mathbf{B} \in \mathbb{R}^{n \times n}$, $\text{tr}(\mathbf{A}\mathbf{B}) \leq \|\mathbf{A}\| \cdot \|\mathbf{B}\|$. If \mathbf{A} is symmetric and positive semi-definite, then $\|\mathbf{A}\| \leq \text{tr}(\mathbf{A})$. Also, we will make use of the property that similar matrices share eigenvalues, where two matrices $\mathbf{A}, \mathbf{B} \in \mathbb{R}^{n \times n}$ are similar if, for some invertible matrix $\mathbf{C} \in \mathbb{R}^{n \times n}$, $\mathbf{B} = \mathbf{C}^{-1} \mathbf{A} \mathbf{C}$. The result follows immediately from that \mathbf{A} and \mathbf{B} share characteristic polynomial; $|\mathbf{B} - \varphi \mathbf{I}_n| = |\mathbf{C}^{-1}(\mathbf{A} - \varphi \mathbf{I}_n) \mathbf{C}| = |\mathbf{C}^{-1}| \cdot |\mathbf{A} - \varphi \mathbf{I}_n| \cdot |\mathbf{C}| = |\mathbf{A} - \varphi \mathbf{I}_n|$.

For the proofs we need the following lemma:

Lemma A1 For the $N \times r$ matrices $\mathbf{H} = \Sigma_{\varepsilon\varepsilon}^{-1/2} \Lambda$ and $\mathbf{H}_0 = \Sigma_{\varepsilon\varepsilon}^{-1/2} \Lambda \Sigma_{gg}^{1/2}$, where r is the number of factors, it holds that, for $i = 1, 2, \dots, r$, $\varphi_i(\mathbf{H}'\mathbf{H}) = \mathcal{O}(N)$ and $\varphi_i(\mathbf{H}'_0 \mathbf{H}_0) = \mathcal{O}(N)$, corresponding to the non-zero eigenvalues of $\mathbf{H}\mathbf{H}'$ and $\mathbf{H}_0 \mathbf{H}'_0$ respectively.

Proof of Lemma A1. Because $\Sigma_{zz} = \Sigma_{\varepsilon\varepsilon}$, it follows directly from Assumption 2.5 (ii) that all the eigenvalues of $\mathbf{H}'\mathbf{H}$ are $\mathcal{O}(N)$. To see that this corresponds to the non-zero eigenvalues of $\mathbf{H}\mathbf{H}'$, let \mathbf{H} have singular value decomposition (SVD) $\mathbf{U}\mathbf{K}\mathbf{V}'$ where \mathbf{U} is $N \times N$, \mathbf{K} is $N \times r$ and \mathbf{V} is $r \times r$.

Because $N > r$, \mathbf{K} has r singular values and $N - r$ zero-rows, i.e. $\mathbf{K} = [\mathbf{K}_+, \mathbf{0}]'$. 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 $\mathbf{U}_+ \mathbf{K}_+ \mathbf{V}'$ is the reduced SVD such that $\mathbf{H}\mathbf{H}' = \mathbf{U}_+ \mathbf{K}_+^2 \mathbf{U}'_+$, where \mathbf{K}_+^2 has the eigenvalues of $\mathbf{H}'\mathbf{H} = \mathbf{V}\mathbf{K}_+^2 \mathbf{V}'$.

From Assumption 2.1 (ii) we have that $\Sigma_{gg} > 0$, implying that for any non-zero $r \times 1$ vector \mathbf{a} , $\Sigma_{gg} \mathbf{a} \neq \mathbf{0}$. It then immediately follows that, using Assumption 2.5 (ii),

$$N^{-1} \mathbf{H}'_0 \mathbf{H}_0 = N^{-1} \Sigma_{gg}^{1/2} \Lambda' \Sigma_{\varepsilon\varepsilon}^{-1} \Lambda \Sigma_{gg}^{1/2} = N^{-1} \Sigma_{gg}^{1/2} \mathbf{H}' \mathbf{H} \Sigma_{gg}^{1/2} \xrightarrow{p} \Sigma_{gg}^{1/2} \Sigma_1 \Sigma_{gg}^{1/2},$$

where $\Sigma_{gg}^{1/2} \Sigma_1 \Sigma_{gg}^{1/2} > 0$, because $\mathbf{a}' \Sigma_{gg}^{1/2} \Sigma_1 \Sigma_{gg}^{1/2} \mathbf{a} = (\Sigma_{gg}^{1/2} \mathbf{a})' \Sigma_1 (\Sigma_{gg}^{1/2} \mathbf{a}) > 0$. Hence all the eigenvalues of $\mathbf{H}'_0 \mathbf{H}_0$, and likewise the non-zero eigenvalues of $\mathbf{H}_0 \mathbf{H}'_0$, are $\mathcal{O}(N)$. ■

Proof of Lemma 3.1.

(i) Because $\Lambda \Lambda'$ and $\Sigma_{\varepsilon\varepsilon}$ are both symmetric positive definite, $\Omega^k = (\Lambda \Lambda' + \Sigma_{\varepsilon\varepsilon})^k$ is also symmetric positive definite for $k \in \mathbb{R}$, such that $\varphi_i(\Omega^k) = \varphi_i^k(\Omega)$ for $i = 1, 2, \dots, N$, implying $\text{tr}(\Omega^k) = \sum_{i=1}^N \varphi_i^k(\Omega) = \mathcal{O}(N)$. Also, $\Gamma_{ij} = \lambda_j \mathbf{e}'_i + \mathbf{e}_i \lambda'_j$ is symmetric, where $\text{tr}(\Gamma_{ij}^2) = 2\lambda_{ij}^2 + 2\sum_{i=1}^N \lambda_{ij}^2 = \mathcal{O}(N)$. Using these results, the result (A6), and the Cauchy-Schwarz trace-inequality, we have that, as $T \rightarrow \infty$, for $i, l = 1, 2, \dots, N$ and $j, q = 1, 2, \dots, r$,

$$\begin{aligned} \frac{\tilde{J}_{\lambda_{ij}\lambda_{lq}}}{N^2 T} &= \frac{T^*}{2N^2 T} \text{tr} \left(\tilde{\Gamma}_{ij} \mathbf{S}_{00}^{-1} \tilde{\Gamma}_{lq} \mathbf{S}_{00}^{-1} \right) \xrightarrow{p} \frac{1}{2N^2} \text{tr} \left(\Gamma_{ij} \Omega^{-1} \Gamma_{lq} \Omega^{-1} \right) \\ &\leq \frac{1}{2N^2} \|\Gamma_{ij} \Omega^{-1}\| \cdot \|\Gamma_{lq} \Omega^{-1}\| \\ &\leq \frac{1}{2N^2} \|\Gamma_{ij}\| \cdot \|\Gamma_{lq}\| \cdot \|\Omega^{-1}\| \cdot \|\Omega^{-1}\| \\ &= \frac{1}{2N^2} \text{tr} \left(\Gamma_{lq}^2 \right) \text{tr} \left(\Omega^{-2} \right) = \mathcal{O}(1). \end{aligned}$$

(ii) If we let $\Omega^{-1} \Gamma_{ij} \Omega^{-1} = [a_{i,l}]_{N \times N}$, then, using the result (A7), we have that, for $i, l = 1, 2, \dots, N$ and $j = 1, 2, \dots, r$, as $T \rightarrow \infty$,

$$\frac{\tilde{J}_{\lambda_{ij}\sigma_{\varepsilon_l}^2}}{T} = \frac{T^*}{2T} \text{tr} \left(\tilde{\Gamma}_{ij} \mathbf{S}_{01}^{-1} \mathbf{e}_l \mathbf{e}'_l \mathbf{S}_{01}^{-1} \right) \xrightarrow{p} \frac{1}{2} \mathbf{e}'_l \Omega^{-1} \Gamma_{ij} \Omega^{-1} \mathbf{e}_l = \frac{1}{2} a_{l,l} = \mathcal{O}(1).$$

(iii) Let $\sigma_{max} = \max[\text{dg}(\Sigma_{\varepsilon\varepsilon})] = \mathcal{O}(1)$. Because $\Sigma_{\varepsilon\varepsilon}$ is diagonal and, from (i), $\text{tr}(\Omega^k)$ and $\text{tr}(\Gamma_{ij}^2)$ are both $\mathcal{O}(N)$, we have that, using the result (A8), for $i = 1, 2, \dots, N$ and $j = 1, 2, \dots, r$, as $T \rightarrow \infty$,

$$\begin{aligned} \left| \frac{\tilde{J}_{\lambda_{ij}\rho}}{N^{3/2} T} \right| &= \frac{T^*}{4N^{3/2} T} \text{tr} \left(\tilde{\Gamma}_{ij} \mathbf{S}_{01}^{-1} \tilde{\Sigma}_{\varepsilon\varepsilon} \mathbf{S}_{01}^{-1} \right) \xrightarrow{p} \frac{1}{4N^{3/2}} \text{tr} \left(\Gamma_{ij} \Omega^{-1} \Sigma_{\varepsilon\varepsilon} \Omega^{-1} \right), \\ &\leq \frac{1}{4N^{3/2}} \|\Gamma_{ij}\| \cdot \|\Omega^{-1} \Sigma_{\varepsilon\varepsilon}^{1/2}\| \cdot \|\Sigma_{\varepsilon\varepsilon}^{1/2} \Omega^{-1}\| \\ &= \frac{1}{4N^{3/2}} \sqrt{\text{tr} \left(\Gamma_{ij}^2 \right)} \text{tr} \left(\Omega^{-1} \Sigma_{\varepsilon\varepsilon} \Omega^{-1} \right) \\ &\leq \frac{\sigma_{max}}{4N^{3/2}} \sqrt{\text{tr} \left(\Gamma_{ij}^2 \right)} \text{tr} \left(\Omega^{-2} \right) = \mathcal{O}(1). \end{aligned}$$

(iv) If we let $\Omega^{-1} = [b_{i,l}]_{N \times N}$, then, using the result (A4), we have that, for $i = 1, 2, \dots, N$, as $T \rightarrow \infty$,

$$\frac{\tilde{J}_{\sigma_{\varepsilon_i}^2 \sigma_{\varepsilon_l}^2}}{T} = \frac{T^*}{2T} \text{tr} \left(\mathbf{e}'_i \mathbf{S}_{01}^{-1} \mathbf{e}_i \mathbf{e}'_l \mathbf{S}_{01}^{-1} \mathbf{e}_l \right) \xrightarrow{p} \frac{1}{2} \mathbf{e}'_i \Omega^{-1} \mathbf{e}_i \mathbf{e}'_l \Omega^{-1} \mathbf{e}_l = \frac{1}{2} b_{i,i} \times b_{l,l} = \mathcal{O}(1).$$

(v) If we let $\mathbf{\Omega}^{-1}\mathbf{\Sigma}_{\varepsilon\varepsilon}\mathbf{\Omega}^{-1} = [c_{i,l}]_{N \times N}$, then, using the result (A5), we have that, for $i = 1, 2, \dots, N$, as $T \rightarrow \infty$,

$$\frac{\tilde{J}_{\rho\varepsilon i}^2}{T} = -\frac{T^*}{4T} \text{tr} \left(\tilde{\mathbf{\Sigma}}_{\varepsilon\varepsilon} \mathbf{S}_{01}^{-1} \mathbf{e}_i \mathbf{e}_i' \mathbf{S}_{01}^{-1} \right) \xrightarrow{p} -\frac{1}{4} \mathbf{e}_i' \mathbf{\Omega}^{-1} \mathbf{\Sigma}_{\varepsilon\varepsilon} \mathbf{\Omega}^{-1} \mathbf{e}_i = -\frac{1}{4} c_{i,i} = \mathcal{O}(1).$$

(vi) Let $\mathbf{H} = \mathbf{\Sigma}_{\varepsilon\varepsilon}^{-1/2} \mathbf{\Lambda}$, and let $\mathbf{H}\mathbf{H}'$ have eigenvalues $\eta_i = \varphi_i(\mathbf{H}\mathbf{H}')$ for $i = 1, 2, \dots, N$, where from Lemma A1, $\eta_i = \mathcal{O}(N)$ for $i = 1, 2, \dots, r$, and $\eta_i = 0$ for $i = r+1, \dots, N$. We then have that $\mathbf{\Sigma}_{\varepsilon\varepsilon} \mathbf{\Omega}^{-1} \stackrel{s}{\sim} \mathbf{\Sigma}_{\varepsilon\varepsilon}^{-1/2} \mathbf{\Sigma}_{\varepsilon\varepsilon} \mathbf{\Omega}^{-1} \mathbf{\Sigma}_{\varepsilon\varepsilon}^{1/2} = \mathbf{\Sigma}_{\varepsilon\varepsilon}^{1/2} \mathbf{\Omega}^{-1} \mathbf{\Sigma}_{\varepsilon\varepsilon}^{1/2} = (\mathbf{H}\mathbf{H}' + \mathbf{I}_N)^{-1}$, which is symmetric positive definite. Thus $\mathbf{\Sigma}_{\varepsilon\varepsilon} \mathbf{\Omega}^{-1}$ is positive definite, but not symmetric. It follows that $\varphi_i(\mathbf{\Sigma}_{\varepsilon\varepsilon} \mathbf{\Omega}^{-1}) = (\eta_i + 1)^{-1}$, and that for all $k \in \mathbb{R}$,

$$\text{tr} \left[\left(\mathbf{\Sigma}_{\varepsilon\varepsilon} \mathbf{\Omega}^{-1} \right)^k \right] = \sum_{i=1}^N \varphi_i^k \left(\mathbf{\Sigma}_{\varepsilon\varepsilon} \mathbf{\Omega}^{-1} \right) = \sum_{i=1}^N (\eta_i + 1)^{-k} = \mathcal{O}(N).$$

Hence, using (A3),

$$\frac{\tilde{J}_{\rho\rho}}{NT^2} = \frac{T^{*2}}{8NT^2} \text{tr} \left(\tilde{\mathbf{\Sigma}}_{\varepsilon\varepsilon} \mathbf{S}_{01}^{-1} \tilde{\mathbf{\Sigma}}_{\varepsilon\varepsilon} \mathbf{S}_{01}^{-1} \right) \xrightarrow{p} \frac{1}{8N} \text{tr} \left[\left(\mathbf{\Sigma}_{\varepsilon\varepsilon} \mathbf{\Omega}^{-1} \right)^2 \right] = \mathcal{O}(1), \text{ as } T \rightarrow \infty,$$

which completes the proof of Lemma 3.1 ■

Proof of Theorem 3.1. Rewrite the statistic (16) as

$$\vartheta = \frac{\text{tr} \left(\tilde{\mathbf{\Sigma}}_{\varepsilon\varepsilon} \mathbf{S}_{01}^{-1} \right) - 2 \text{tr} \left(\tilde{\mathbf{\Sigma}}_{\varepsilon\varepsilon} \mathbf{S}_{01}^{-1} \frac{1}{T^*} \mathbf{S}_0 \mathbf{S}_{01}^{-1} \right) + \text{tr} \left(\tilde{\mathbf{\Sigma}}_{\varepsilon\varepsilon} \mathbf{S}_{01}^{-1} \frac{1}{T^*} \mathbf{S}_{00} \mathbf{S}_{01}^{-1} \right)}{\sqrt{2 \text{tr} \left(\tilde{\mathbf{\Sigma}}_{\varepsilon\varepsilon} \mathbf{S}_{01}^{-1} \tilde{\mathbf{\Sigma}}_{\varepsilon\varepsilon} \mathbf{S}_{01}^{-1} \right)}}. \quad (\text{A9})$$

From the proof of Lemma 3.1 (vi) we have that $\varphi_i(\mathbf{\Sigma}_{\varepsilon\varepsilon} \mathbf{\Omega}^{-1}) = (\eta_i + 1)^{-1}$, where $\eta_i = \varphi_i(\mathbf{\Sigma}_{\varepsilon\varepsilon}^{-1} \mathbf{\Lambda} \mathbf{\Lambda}')$. Thus, because $\tilde{\mathbf{\Sigma}}_{\varepsilon\varepsilon} \xrightarrow{p} \mathbf{\Sigma}_{\varepsilon\varepsilon}$, $\mathbf{S}_{01}^{-1} \xrightarrow{p} \mathbf{\Omega}^{-1}$, and from (8), $\frac{1}{T^*} \mathbf{S}_0 \xrightarrow{p} \mathbf{\Omega}$, for the first two parts in the numerator of (A9) we have that $\text{tr}(\tilde{\mathbf{\Sigma}}_{\varepsilon\varepsilon} \mathbf{S}_{01}^{-1}) \xrightarrow{p} \text{tr}(\mathbf{\Sigma}_{\varepsilon\varepsilon} \mathbf{\Omega}^{-1}) = \sum_{i=1}^N \frac{1}{\eta_i + 1}$ and $2 \text{tr}(\tilde{\mathbf{\Sigma}}_{\varepsilon\varepsilon} \mathbf{S}_{01}^{-1} \frac{1}{T^*} \mathbf{S}_0 \mathbf{S}_{01}^{-1}) \xrightarrow{p} 2 \text{tr}(\mathbf{\Sigma}_{\varepsilon\varepsilon} \mathbf{\Omega}^{-1}) = 2 \sum_{i=1}^N \frac{1}{\eta_i + 1}$. For the third part, note that the eigenvalues of $\mathbf{\Sigma}_{\varepsilon\varepsilon} \mathbf{\Omega}^{-1}$ are also the eigenvalues of the similar matrix $\mathbf{\Omega}^{-1/2} \mathbf{\Sigma}_{\varepsilon\varepsilon} \mathbf{\Omega}^{-1/2} = \mathbf{\Omega}^{-1/2} \mathbf{\Sigma}_{\varepsilon\varepsilon} \mathbf{\Omega}^{-1/2}$. Let $\mathbf{\Omega}^{-1/2} \mathbf{\Sigma}_{\varepsilon\varepsilon} \mathbf{\Omega}^{-1/2}$ have spectral decomposition $\mathbf{B}\mathbf{A}\mathbf{B}'$, where $\mathbf{A} = \text{diag}((\eta_1 + 1)^{-1}, (\eta_2 + 1)^{-1}, \dots, (\eta_N + 1)^{-1})$. Using this and the CLT from (9) in Section 2 we have that

$$\begin{aligned} \text{tr} \left[\tilde{\mathbf{\Sigma}}_{\varepsilon\varepsilon} \mathbf{S}_{01}^{-1/2} \mathbf{S}_{01}^{-1/2} \frac{1}{\sqrt{T^*}} \left(\sum_{t=2}^T \mathbf{y}_t \right) \frac{1}{\sqrt{T^*}} \left(\sum_{t=2}^T \mathbf{y}_t \right)' \mathbf{S}_{01}^{-1/2} \mathbf{S}_{01}^{-1/2} \right] &\xrightarrow{d} \text{tr} \left(\mathbf{\Sigma}_{\varepsilon\varepsilon} \mathbf{\Omega}^{-1/2} \mathbf{Z}\mathbf{Z}' \mathbf{\Omega}^{-1/2} \right) \\ &= \mathbf{Z}' \mathbf{B}\mathbf{A}\mathbf{B}' \mathbf{Z} \\ &\sim \mathbf{Z}' \mathbf{A}\mathbf{Z} = \sum_{i=1}^N \frac{1}{\eta_i + 1} Z_i^2, \end{aligned}$$

where we have used that $\mathbf{W} = \mathbf{B}'\mathbf{Z}$ is an orthogonal transformation, implying $\mathbf{W} \sim \mathbf{Z}$.

Finally, for the denominator of (A9) we have that, using the continuous mapping theorem,

$$\sqrt{2 \text{tr} \left(\tilde{\mathbf{\Sigma}}_{\varepsilon\varepsilon} \mathbf{S}_{01}^{-1} \tilde{\mathbf{\Sigma}}_{\varepsilon\varepsilon} \mathbf{S}_{01}^{-1} \right)} \xrightarrow{p} \sqrt{2 \text{tr} \left(\mathbf{\Sigma}_{\varepsilon\varepsilon} \mathbf{\Omega}^{-1} \mathbf{\Sigma}_{\varepsilon\varepsilon} \mathbf{\Omega}^{-1} \right)} = \sqrt{2 \sum_{i=1}^N \left(\frac{1}{\eta_i + 1} \right)^2}.$$

Put together, using Slutsky's theorem, (17) follows. Note finally that $\eta_i = \varphi_i(\mathbf{\Sigma}_{\varepsilon\varepsilon}^{-1} \mathbf{\Lambda} \mathbf{\Lambda}')$, where $\mathbf{\Sigma}_{\varepsilon\varepsilon}^{-1} \mathbf{\Lambda} \mathbf{\Lambda}' \stackrel{s}{\sim} \mathbf{\Sigma}_{\varepsilon\varepsilon}^{-1/2} \mathbf{\Lambda} \mathbf{\Lambda}' \mathbf{\Sigma}_{\varepsilon\varepsilon}^{-1/2}$, such that from Lemma A1, $\eta_i = \varphi_i(\mathbf{\Lambda}' \mathbf{\Sigma}_{\varepsilon\varepsilon}^{-1} \mathbf{\Lambda})$ for $i = 1, 2, \dots, r$, and $\eta_i = 0$ for $i = r+1, r+2, \dots, N$. This completes the proof of Theorem 3.1. ■

Proof of Theorem 3.2. Relaxing Misspecification 1, but retaining Assumptions 2.1-2.6, the contemporaneous covariances are, under the null hypothesis,

$$\boldsymbol{\Sigma}_{yy} = \text{Var}(\mathbf{y}_t) = \boldsymbol{\Lambda}_0 \boldsymbol{\Lambda}'_0 + \boldsymbol{\Sigma}_{\varepsilon\varepsilon} \equiv \boldsymbol{\Omega}_0,$$

where $\boldsymbol{\Lambda}_0 = \boldsymbol{\Lambda} \boldsymbol{\Sigma}_{gg}^{1/2}$ with $\boldsymbol{\Sigma}_{gg} > 0$ from Assumption 2.1 (ii). As shown by Bai and Li (2012b), the MLEs in the exact factor model are consistent for $\boldsymbol{\Lambda}_0$ and $\boldsymbol{\Sigma}_{\varepsilon\varepsilon}$, such that $\mathbf{S}_{01} = \tilde{\boldsymbol{\Lambda}}_0 \tilde{\boldsymbol{\Lambda}}'_0 + \tilde{\boldsymbol{\Sigma}}_{\varepsilon\varepsilon} \xrightarrow{p} \boldsymbol{\Omega}_0$ as $T \rightarrow \infty$. Also, because $\mathbf{y}_t \sim I(0)$, \mathbf{y}_t has an MA representation by the Wold decomposition theorem, from which it follows that $\frac{1}{T^*} \mathbf{S}_0 = \frac{1}{T^*} \sum_{t=2}^T \mathbf{y}_t \mathbf{y}'_t \xrightarrow{p} E(\mathbf{y}_t \mathbf{y}'_t) = \boldsymbol{\Omega}_0$.

Consider again the decomposition (A9). For the first two parts in the numerator we have that, as $T \rightarrow \infty$, $\text{tr}(\tilde{\boldsymbol{\Sigma}}_{\varepsilon\varepsilon} \mathbf{S}_{01}^{-1}) \xrightarrow{p} \text{tr}(\boldsymbol{\Sigma}_{\varepsilon\varepsilon} \boldsymbol{\Omega}_0^{-1})$ and $\text{tr}(\tilde{\boldsymbol{\Sigma}}_{\varepsilon\varepsilon} \mathbf{S}_{01}^{-1} \frac{1}{T^*} \mathbf{S}_0 \mathbf{S}_{01}^{-1}) \xrightarrow{p} \text{tr}(\boldsymbol{\Sigma}_{\varepsilon\varepsilon} \boldsymbol{\Omega}_0^{-1})$ where $\boldsymbol{\Sigma}_{\varepsilon\varepsilon} \boldsymbol{\Omega}_0^{-1} \stackrel{s}{\sim} \boldsymbol{\Sigma}_{\varepsilon\varepsilon}^{1/2} \boldsymbol{\Omega}_0^{-1} \boldsymbol{\Sigma}_{\varepsilon\varepsilon}^{1/2} = (\mathbf{H}_0 \mathbf{H}'_0 + \mathbf{I}_N)^{-1}$, with $\mathbf{H}_0 = \boldsymbol{\Sigma}_{\varepsilon\varepsilon}^{-1/2} \boldsymbol{\Lambda}_0$. Let $\eta_{0,i} = \varphi_i(\mathbf{H}_0 \mathbf{H}'_0)$, where from Lemma A1, $\eta_{0,i} = \mathcal{O}(N)$ for $i = 1, 2, \dots, r$, and $\eta_{0,i} = 0$ for $i = r + 1, \dots, N$. Thus, for all $k \in \mathbb{R}$,

$$w_{0,i}^k = \begin{cases} (1 + \eta_{0,i})^{-1} = \mathcal{O}(N^{-k}) & \text{for } i = 1, 2, \dots, r \\ 1 & \text{for } i = r + 1, \dots, N \end{cases} \quad (\text{A10})$$

such that

$$\text{tr}(\boldsymbol{\Sigma}_{\varepsilon\varepsilon} \boldsymbol{\Omega}_0^{-1})^k = \sum_{i=1}^N w_{0,i}^k = \mathcal{O}(N). \quad (\text{A11})$$

Further, we have that $\sum_{t=2}^T \mathbf{y}_t = \sum_{t=2}^T \Delta \mathbf{x}_t = \mathbf{x}_T - \mathbf{x}_1 = \boldsymbol{\Lambda}(\mathbf{f}_T - \mathbf{f}_1) + (\mathbf{u}_T - \mathbf{u}_1)$. Or if we index from $t = 1$ (in differences), then we have that $\sum_{t=1}^T \mathbf{y}_t = \mathbf{x}_T - \mathbf{x}_0 = \boldsymbol{\Lambda}(\mathbf{f}_T - \mathbf{f}_0) + (\mathbf{u}_T - \mathbf{u}_0)$, where by Assumption 2.5, \mathbf{f}_0 and \mathbf{u}_0 come from some stationary distributions, which only matters under the alternative, and under the null they could be $\mathbf{O}_p(1)$. Regardless of where the index starts, for the third part in the numerator of (A9) we have, under the null hypothesis as $T \rightarrow \infty$, that

$$\begin{aligned} \frac{1}{T^*} \mathbf{S}_{00} &= \frac{\left(\sum_{t=2}^T \mathbf{y}_t\right) \left(\sum_{t=2}^T \mathbf{y}_t\right)'}{\sqrt{T^*} \sqrt{T^*}} \\ &= \frac{\boldsymbol{\Lambda} \mathbf{f}_T \mathbf{f}'_T \boldsymbol{\Lambda}' + \mathbf{u}_T \mathbf{u}'_T}{T^*} + \frac{\boldsymbol{\Lambda} \mathbf{f}_T \mathbf{u}'_T + \mathbf{u}_T \mathbf{f}'_T \boldsymbol{\Lambda}'}{T^*} + \mathbf{o}_p(1) \\ &\xrightarrow{d} \boldsymbol{\Lambda} \boldsymbol{\Gamma}_{gg}^{1/2} \mathbf{Z}_1 \mathbf{Z}'_1 \boldsymbol{\Gamma}_{gg}^{1/2} \boldsymbol{\Lambda}' + \boldsymbol{\Sigma}_{\varepsilon\varepsilon}^{1/2} \mathbf{Z}_2 \mathbf{Z}'_2 \boldsymbol{\Sigma}_{\varepsilon\varepsilon}^{1/2} + \boldsymbol{\Lambda} \boldsymbol{\Gamma}_{gg}^{1/2} \mathbf{Z}_1 \mathbf{Z}'_2 \boldsymbol{\Sigma}_{\varepsilon\varepsilon}^{1/2} + \boldsymbol{\Sigma}_{\varepsilon\varepsilon}^{1/2} \mathbf{Z}_2 \mathbf{Z}'_1 \boldsymbol{\Lambda} \boldsymbol{\Gamma}_{gg}^{1/2} \\ &= \mathcal{I}_1 + \mathcal{I}_2 + \mathcal{I}_3 + \mathcal{I}_4, \end{aligned}$$

where $\mathbf{Z}_1 \sim \mathcal{N}(\mathbf{0}, \mathbf{I}_r)$ and $\mathbf{Z}_2 \sim \mathcal{N}(\mathbf{0}, \mathbf{I}_N)$ are independent, $\boldsymbol{\Gamma}_{gg}$ ($r \times r$) is the long run covariance matrix of $\mathbf{g}_t = \Delta \mathbf{f}_t$, $\mathcal{I}_1, \mathcal{I}_2, \mathcal{I}_3, \mathcal{I}_4$ have obvious definitions and the $\mathbf{o}_p(1)$ term is due to the initial values.⁵ We then have that, as $T \rightarrow \infty$,

$$\text{tr} \left(\tilde{\boldsymbol{\Sigma}}_{\varepsilon\varepsilon} \mathbf{S}_{01}^{-1} \frac{\mathbf{S}_{00}}{T^*} \mathbf{S}_{01}^{-1} \right) \xrightarrow{d} \text{tr} \left[\boldsymbol{\Sigma}_{\varepsilon\varepsilon} \boldsymbol{\Omega}_0^{-1} (\mathcal{I}_1 + \mathcal{I}_2 + \mathcal{I}_3 + \mathcal{I}_4) \boldsymbol{\Omega}_0^{-1} \right],$$

where

- (i) $\text{tr}(\boldsymbol{\Sigma}_{\varepsilon\varepsilon} \boldsymbol{\Omega}_0^{-1} \mathcal{I}_1 \boldsymbol{\Omega}_0^{-1}) = \mathcal{O}_p(N^{-1})$,
- (ii) $\text{tr}(\boldsymbol{\Sigma}_{\varepsilon\varepsilon} \boldsymbol{\Omega}_0^{-1} \mathcal{I}_2 \boldsymbol{\Omega}_0^{-1}) = \mathcal{O}_p(N)$,
- (iii) $\text{tr}(\boldsymbol{\Sigma}_{\varepsilon\varepsilon} \boldsymbol{\Omega}_0^{-1} \mathcal{I}_3 \boldsymbol{\Omega}_0^{-1}) = \text{tr}(\boldsymbol{\Sigma}_{\varepsilon\varepsilon} \boldsymbol{\Omega}_0^{-1} \mathcal{I}_4 \boldsymbol{\Omega}_0^{-1}) = \mathcal{O}_p(1)$.

⁵Using the central limit theorem for stationary dependent processes (see e.g. Lütkepohl, 2007, p. 691, Proposition C.13(3)) we have that $\frac{1}{\sqrt{T^*}} \sum \Delta \mathbf{f}_t = \frac{1}{\sqrt{T^*}} \mathbf{f}_T + \mathbf{o}_p(1) \xrightarrow{d} \boldsymbol{\Gamma}_{gg}^{1/2} \mathbf{Z}_1$. Also, because $\mathbf{u}_T = \sum_{t=2}^T \boldsymbol{\varepsilon}_t$, where $\boldsymbol{\varepsilon}_t = (\varepsilon_{1,t}, \varepsilon_{2,t}, \dots, \varepsilon_{N,t})'$ is iid, we have by the central limit theorem that $\frac{1}{\sqrt{T^*}} \sum \Delta \mathbf{u}_t = \frac{1}{\sqrt{T^*}} \mathbf{u}_T + \mathbf{o}_p(1) \xrightarrow{d} \boldsymbol{\Sigma}_{\varepsilon\varepsilon}^{1/2} \mathbf{Z}_2$.

Proof of (i): Let $\text{tr}(\boldsymbol{\Sigma}_{\varepsilon\varepsilon}\boldsymbol{\Omega}_0^{-1}\mathcal{I}_1\boldsymbol{\Omega}_0^{-1}) = \mathbf{Z}'_1\mathbf{M}_1\mathbf{Z}_1$, where $\mathbf{M}_1 = \boldsymbol{\Gamma}_{gg}^{1/2}\boldsymbol{\Lambda}'\boldsymbol{\Omega}_0^{-1}\boldsymbol{\Sigma}_{\varepsilon\varepsilon}\boldsymbol{\Omega}_0^{-1}\boldsymbol{\Lambda}\boldsymbol{\Gamma}_{gg}^{1/2}$, which is positive semi-definite, because it is quadratic around $\boldsymbol{\Sigma}_{\varepsilon\varepsilon} > 0$. Similarly, $\boldsymbol{\Lambda}'\boldsymbol{\Omega}_0^{-1}\boldsymbol{\Sigma}_{\varepsilon\varepsilon}\boldsymbol{\Omega}_0^{-1}\boldsymbol{\Lambda} \geq 0$. We now show that the non-negative eigenvalues of \mathbf{M}_1 are bounded by showing that the trace of \mathbf{M}_1 ($r \times r$) is bounded:

$$\begin{aligned}\text{tr}(\mathbf{M}_1) &= \text{tr}(\boldsymbol{\Lambda}'\boldsymbol{\Omega}_0^{-1}\boldsymbol{\Sigma}_{\varepsilon\varepsilon}\boldsymbol{\Omega}_0^{-1}\boldsymbol{\Lambda}\boldsymbol{\Gamma}_{gg}) \leq \|\boldsymbol{\Lambda}\boldsymbol{\Lambda}'\boldsymbol{\Omega}_0^{-1}\boldsymbol{\Sigma}_{\varepsilon\varepsilon}\boldsymbol{\Omega}_0^{-1}\| \cdot \|\boldsymbol{\Gamma}_{gg}\| \\ &\leq \text{tr}(\boldsymbol{\Lambda}\boldsymbol{\Lambda}'\boldsymbol{\Omega}_0^{-1}\boldsymbol{\Sigma}_{\varepsilon\varepsilon}\boldsymbol{\Omega}_0^{-1})\|\boldsymbol{\Gamma}_{gg}\| \\ &= \text{tr}\left\{[(\boldsymbol{\Lambda}\boldsymbol{\Lambda}' + \boldsymbol{\Sigma}_{\varepsilon\varepsilon}) - \boldsymbol{\Sigma}_{\varepsilon\varepsilon}]\boldsymbol{\Omega}_0^{-1}\boldsymbol{\Sigma}_{\varepsilon\varepsilon}\boldsymbol{\Omega}_0^{-1}\right\}\|\boldsymbol{\Gamma}_{gg}\| \\ &= \text{tr}\left[\left(\boldsymbol{\Sigma}_{\varepsilon\varepsilon}\boldsymbol{\Omega}_0^{-1}\right) - \text{tr}\left(\boldsymbol{\Sigma}_{\varepsilon\varepsilon}\boldsymbol{\Omega}_0^{-1}\boldsymbol{\Sigma}_{\varepsilon\varepsilon}\boldsymbol{\Omega}_0^{-1}\right)\right]\|\boldsymbol{\Gamma}_{gg}\|,\end{aligned}$$

where $\|\boldsymbol{\Gamma}_{gg}\| < \infty$ (r is fixed). Using (A10-A11) we have that $\text{tr}(\boldsymbol{\Sigma}_{\varepsilon\varepsilon}\boldsymbol{\Omega}_0^{-1}) - \text{tr}(\boldsymbol{\Sigma}_{\varepsilon\varepsilon}\boldsymbol{\Omega}_0^{-1}\boldsymbol{\Sigma}_{\varepsilon\varepsilon}\boldsymbol{\Omega}_0^{-1}) = \sum_{i=1}^N w_{0,i} - \sum_{i=1}^N w_{0,i}^2 = N - r + \sum_{i=1}^r w_{0,i} - (N - r + \sum_{i=1}^r w_{0,i}^2) = \mathcal{O}_p(N^{-1})$, because $\sum_{i=1}^r w_{0,i} = \mathcal{O}_p(N^{-1})$ and $\sum_{i=1}^r w_{0,i}^2 = \mathcal{O}_p(N^{-2})$, implying $\max_j \varphi_j(\mathbf{M}_1) = \mathcal{O}_p(N^{-1})$ for $j = 1, 2, \dots, r$. Thus, because $\chi_{1,j}^2 = \mathcal{O}_p(1)$ for any j ,

$$\mathbf{Z}'_1\mathbf{M}_1\mathbf{Z}_1 = \sum_{j=1}^r \varphi_j(\mathbf{M}_1)\chi_{1,j}^2 = \mathcal{O}_p(N^{-1}).$$

Proof of (ii): Let $\text{tr}(\boldsymbol{\Sigma}_{\varepsilon\varepsilon}\boldsymbol{\Omega}_0^{-1}\mathcal{I}_2\boldsymbol{\Omega}_0^{-1}) = \mathbf{Z}'_2\mathbf{M}_2\mathbf{Z}_2$, where $\mathbf{M}_2 = \boldsymbol{\Sigma}_{\varepsilon\varepsilon}^{1/2}\boldsymbol{\Omega}_0^{-1}\boldsymbol{\Sigma}_{\varepsilon\varepsilon}\boldsymbol{\Omega}_0^{-1}\boldsymbol{\Sigma}_{\varepsilon\varepsilon}^{1/2}$ is $N \times N$. Because $\mathbf{M}_2 \stackrel{s}{\sim} \boldsymbol{\Omega}_0^{-1}\boldsymbol{\Sigma}_{\varepsilon\varepsilon}\boldsymbol{\Omega}_0^{-1}\boldsymbol{\Sigma}_{\varepsilon\varepsilon}$, we have that, using (A10),

$$\mathbf{Z}'_2\mathbf{M}_2\mathbf{Z}_2 = \sum_{i=1}^N w_{0,i}^2\chi_{i,1}^2 = \sum_{i=r+1}^N w_{0,i}^2\chi_{i,1}^2 + \mathcal{O}_p(N^{-2}) = \mathcal{O}_p(N),$$

since for $i = r + 1, r + 2, \dots, N$, $w_{0,i} = 1$, and for any i , $\chi_{1,i}^2 = \mathcal{O}_p(1)$.

Proof of (iii): It is readily verified that $\text{tr}(\boldsymbol{\Sigma}_{\varepsilon\varepsilon}\boldsymbol{\Omega}_0^{-1}\mathcal{I}_3\boldsymbol{\Omega}_0^{-1}) = \text{tr}\left(\boldsymbol{\Sigma}_{\varepsilon\varepsilon}\boldsymbol{\Omega}_0^{-1}\mathcal{I}_4\boldsymbol{\Omega}_0^{-1}\right)$, where

$$\begin{aligned}\text{tr}\left(\boldsymbol{\Sigma}_{\varepsilon\varepsilon}\boldsymbol{\Omega}_0^{-1}\mathcal{I}_3\boldsymbol{\Omega}_0^{-1}\right) &= \text{tr}\left(\boldsymbol{\Sigma}_{\varepsilon\varepsilon}^{1/2}\boldsymbol{\Omega}_0^{-1}\boldsymbol{\Lambda}\boldsymbol{\Gamma}_{gg}^{1/2}\mathbf{Z}_1\mathbf{Z}'_2\boldsymbol{\Sigma}_{\varepsilon\varepsilon}^{1/2}\boldsymbol{\Omega}_0^{-1}\boldsymbol{\Sigma}_{\varepsilon\varepsilon}^{1/2}\right) \\ &\leq \|\boldsymbol{\Sigma}_{\varepsilon\varepsilon}^{1/2}\boldsymbol{\Omega}_0^{-1}\boldsymbol{\Lambda}\boldsymbol{\Gamma}_{gg}^{1/2}\mathbf{Z}_1\| \cdot \|\mathbf{Z}'_2\boldsymbol{\Sigma}_{\varepsilon\varepsilon}^{1/2}\boldsymbol{\Omega}_0^{-1}\boldsymbol{\Sigma}_{\varepsilon\varepsilon}^{1/2}\| \\ &= \sqrt{\text{tr}\left(\boldsymbol{\Sigma}_{\varepsilon\varepsilon}^{1/2}\boldsymbol{\Omega}_0^{-1}\boldsymbol{\Lambda}\boldsymbol{\Gamma}_{gg}^{1/2}\mathbf{Z}_1\mathbf{Z}'_1\boldsymbol{\Gamma}_{gg}^{1/2}\boldsymbol{\Lambda}'\boldsymbol{\Omega}_0^{-1}\boldsymbol{\Sigma}_{\varepsilon\varepsilon}^{1/2}\right)} \\ &\quad \times \sqrt{\text{tr}\left(\boldsymbol{\Sigma}_{\varepsilon\varepsilon}^{1/2}\boldsymbol{\Omega}_0^{-1}\boldsymbol{\Sigma}_{\varepsilon\varepsilon}^{1/2}\mathbf{Z}_2\mathbf{Z}'_2\boldsymbol{\Sigma}_{\varepsilon\varepsilon}^{1/2}\boldsymbol{\Omega}_0^{-1}\boldsymbol{\Sigma}_{\varepsilon\varepsilon}^{1/2}\right)} \\ &= \sqrt{\mathbf{Z}'_1\mathbf{M}_1\mathbf{Z}_1} \cdot \sqrt{\mathbf{Z}'_2\mathbf{M}_2\mathbf{Z}_2} = \mathcal{O}_p(N^{-1/2}) \cdot \mathcal{O}_p(N^{1/2}) = \mathcal{O}_p(1).\end{aligned}$$

Lastly, for the denominator of (A9) we have that, using (A11),

$$\sqrt{2 \text{tr}\left(\tilde{\boldsymbol{\Sigma}}_{\varepsilon\varepsilon}\mathbf{S}_{01}^{-1}\tilde{\boldsymbol{\Sigma}}_{\varepsilon\varepsilon}\mathbf{S}_{01}^{-1}\right)} \xrightarrow{p} \sqrt{2 \text{tr}\left(\boldsymbol{\Sigma}_{\varepsilon\varepsilon}\boldsymbol{\Omega}_0^{-1}\boldsymbol{\Sigma}_{\varepsilon\varepsilon}\boldsymbol{\Omega}_0^{-1}\right)} = \mathcal{O}_p(N^{1/2}).$$

Put together, we have that, as $T \rightarrow \infty$,

$$\vartheta \xrightarrow{d} \frac{\text{tr}\left(\boldsymbol{\Sigma}_{\varepsilon\varepsilon}\boldsymbol{\Omega}_0^{-1}\mathcal{I}_2\boldsymbol{\Omega}_0^{-1}\right) - \text{tr}\left(\boldsymbol{\Sigma}_{\varepsilon\varepsilon}\boldsymbol{\Omega}_0^{-1}\right)}{\sqrt{2 \text{tr}\left(\boldsymbol{\Sigma}_{\varepsilon\varepsilon}\boldsymbol{\Omega}_0^{-1}\boldsymbol{\Sigma}_{\varepsilon\varepsilon}\boldsymbol{\Omega}_0^{-1}\right)}} + \mathcal{O}_p(N^{-1/2}),$$

where, using (A10),

$$\frac{\text{tr} \left(\boldsymbol{\Sigma}_{\varepsilon\varepsilon} \boldsymbol{\Omega}_0^{-1} \mathcal{I}_2 \boldsymbol{\Omega}_0^{-1} \right) - \text{tr} \left(\boldsymbol{\Sigma}_{\varepsilon\varepsilon} \boldsymbol{\Omega}_0^{-1} \right)}{\sqrt{2 \text{tr} \left(\boldsymbol{\Sigma}_{\varepsilon\varepsilon} \boldsymbol{\Omega}_0^{-1} \boldsymbol{\Sigma}_{\varepsilon\varepsilon} \boldsymbol{\Omega}_0^{-1} \right)}} = \frac{\sum_{i=1}^N w_{0,i}^2 Z_i^2 - \sum_{i=1}^N w_{0,i}}{\sqrt{2 \sum_{i=1}^N w_{0,i}^2}} \sim \frac{\mathcal{A}_1}{\mathcal{A}_2} + \mathcal{A}_3,$$

with

$$\begin{aligned} \mathcal{A}_1 &= \sum_{i=r+1}^N \chi_{1,i}^2 - (N-r) = \chi_{(N-r)}^2 - (N-r), \\ \mathcal{A}_2 &= \sqrt{2 \sum_{i=1}^N w_{0,i}^2} = \sqrt{2 \sum_{i=r+1}^N w_{0,i}^2} + \mathcal{O}_p(N^{-2}) = \sqrt{2(N-r)} + \mathcal{O}_p(N^{-2}), \\ \mathcal{A}_3 &= \frac{\sum_{i=1}^r w_{0,i}^2 \chi_{1,i}^2 - \sum_{i=1}^r w_{0,i}}{\sqrt{2 \sum_{i=1}^N w_{0,i}^2}} = \mathcal{O}_p(N^{-3/2}). \end{aligned}$$

Because $\mathcal{A}_1/\mathcal{A}_2 \xrightarrow{d} \mathcal{N}(0,1)$ as $N \rightarrow \infty$ by the central limit theorem, Theorem 3.2 follows. ■

References

- BAI, J. (2003): “Inferential theory for factor models of large dimensions,” *Econometrica*, 71, 135–171.
- 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.
- (2010): “Panel unit root tests with cross-section dependence: a further investigation,” *Econometric Theory*, 26, 1088–1114.
- BOIVIN, J., AND S. NG (2006): “Are more data always better for factor analysis,” *Journal of Econometrics*, 132, 169–194.
- BREITUNG, J., AND J. TENHOFEN (2011): “GLS estimation of dynamic factor models,” *Journal of the American Statistical Association*, 106, 1150–1166.
- CHAMBERLAIN, G., AND M. ROTHSCILD (1983): “Arbitrage, factor structure, and mean-variance analysis on large asset markets,” *Econometrica*, 51, 1281–1304.
- CHOI, I. (2011): “Efficient estimation of nonstationary factor models,” Working Papers 1101, Research Institute for Market Economy, Sogang University.

- 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.
- FORNI, M., M. HALLIN, M. LIPPI, AND L. REICHLIN (2000): "The generalized dynamic-factor model: identification and estimation," *The Review of Economics and Statistics*, 82, 540–554.
- GENGENBACH, C., F. C. PALM, AND J.-P. URBAIN (2010): "Panel unit root tests in the presence of cross-sectional dependencies: comparison and implications for modelling," *Econometric Reviews*, 29, 111–145.
- 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.
- KARANASOS, M. (1998): "A new method for obtaining the autocovariance of an ARMA model: an exact form solution," *Econometric Theory*, 14, 622–640.
- KRUINIGER, H. (2008): "Maximum likelihood estimation and inference methods for the covariance stationary panel AR(1)/unit root model," *Journal of Econometrics*, 144, 447–464.
- LÜTKEPOHL, H. (2007): *New Introduction to Multiple Time Series Analysis*. Springer, Berlin.
- 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.
- PHILLIPS, P. C. B., AND P. PERRON (1988): "Testing for a unit root in time series," *Biometrika*, 75, 335–346.
- SARGAN, J. D., AND A. BHARGAVA (1983): "Testing residuals from least squares regression for being generated by the Gaussian random walk," *Econometrica*, 51, 153–174.
- 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 (2002): "Macroeconomic forecasting using diffusion indexes," *Journal of Business and Economic Statistics*, 20, 147–162.
- 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.