UPPSALA UNIVERSITET

Working Paper 2012:2

Department of Statistics

A likelihood ratio test for idiosyncratic unit roots in a dynamic-factor model with integrated factors

Martin Solberger

Working Paper 2012:2
June 2012
Department of Statistics
Uppsala University
Box 513
SE-751 20 UPPSALA
SWEDEN

Title:
A likelihood ratio test for idiosyncratic unit roots in a dynamic-factor model with integrated factors
Author: Martin Solberger
E-mail: martin.solberger@statistics.uu.se

A likelihood ratio test for idiosyncratic unit roots in a dynamic-factor model with integrated factors

Martin Solberger*
Department of Statistics, Uppsala University

Abstract

We consider an exact factor model with the restriction of unobservable common stochastic trends imposed by non-stationary factors as considered by Zhou and Solberger (2012). Conditional on this, we propose a homogenous likelihood ratio test for unit roots in the idiosyncratic components. The likelihood approach has long been overlooked in this framework due to numerical burdens, and though parts of the test are found numerically, the homogeneity restrictions imposed here make the test feasible. In a simulation study, for relatively large T, our test shows better local power than the pooled Fisher-type test of Bai and Ng (2004), while it is roughly equivalent to the LM test of Zhou and Solberger (2012).

JEL: C12, C15, C23, C63
Keywords: Panel unit root, Dynamic factors, Maximum likelihood

1 Introduction

Factor models have become a very popular approach among second generation panel unit root tests to control for a cross-sectional dependence. For only one factor, Phillips and Sul (2003) consider defactorizing the data through feasible GLS and Pesaran (2007) considers demeaning the data. Similarly, for a multifactor setting, Moon and Perron (2004) consider defactorizing and Pesaran (2006) consider demeaning. But as advocated by Bai and Ng (2004), the factors are then primarily seen as a way to control for cross-sectional dependence rather than a source for cumulative shocks, and they consider testing separately both for integrated idiosyncratic components and for integrated common factors.

The standard approach in these papers is to use principal components in combination with Dickey-Fuller based statistics or GLS based statistics. Meanwhile, maximum likelihood has been largely overlooked due to numerical burdens.

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

Also, to our knowledge, any likelihood based methods have only been aimed at estimation and forecasting within stationary dynamic factor models, e.g. Engle and Watson (1981), Stock and Watson (1989), Camba-Mendez, Kapetanios, Smith and Weale (2001) and Jungbacker and Koopman (2008). Using maximum likelihood to test for non-stationarity in panel factor models seems to have been overlooked.

Zhou and Solberger (2012) consider a static factor model, also considered by Bai and Ng (2004), and propose an LM test for idiosyncratic unit roots conditional on that all the factors are nonstationary. Though the assumption of integrated factors is restrictive, it could typically arise in applied work for either a large data set or a specific subset within a dynamic-factor model framework (see e.g. Eickmeier, 2009). By simulation, Zhou and Solberger (2012) find that the LM test has substantially higher local power than the pooled Fisher-type test in Bai and Ng (2004) when there are unobservable common stochastic trends imposed by the factors.

In this paper we follow the framework of Zhou and Solberger (2012) and look at the likelihood ratio test for unit roots in the idiosyncratic components conditional on a certain number of non-stationary common factors. The restrictions we impose allow us to rely on standard optimization procedures, from which we provide response surface regression estimates for critical values and compare power with the pooled test of Bai and Ng (2004) and the LM test of Zhou and Solberger (2012). The rest of the paper is organized as follows. Section 2 outlines the framework and Section 3 specifies the likelihood ratio. Section 4 illustrates, through simulation, the power properties of the likelihood ratio test and Section 5 concludes. All proofs are placed in the Appendix.

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

2 The framework

For a panel, let $i = 1, ..., N$ denote the individuals and let $t = 1, ..., T$ denote the time points. We consider the factor model

$$x_{i,t} = \mu_i + \boldsymbol{\lambda}_i'\mathbf{f}_t + u_{i,t}, \tag{1}$$

where $\mathbf{f}_t = (f_{1t}, f_{2t}, ..., f_{rt})'$ is a $r \times 1$ vector of unobservable dynamic factors, $\boldsymbol{\lambda}_i' = (\lambda_{i1}, \lambda_{i2}, ..., \lambda_{ir})$ is a $r \times 1$ vector of factor loadings, $u_{i,t}$ is an idiosyncratic component and μ_i is an individual-specific intercept. It is convenient to rewrite

Equation (1) as an N-dimensional time series,

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

where $\mathbf{u}_t = (u_{1t}, u_{2t}, ..., u_{Nt})'$, $\boldsymbol{\mu} = (\mu_1, \mu_2, ..., \mu_N)'$ and $\boldsymbol{\Lambda}$ is the $N \times r$ matrix of factor loadings. Here non-stationarity in $\mathbf{x}_t$ can arise as a result of non-stationary factors or/and non-stationary idiosyncratics. We assume that $\mathbf{x}_t \sim I(1)$ due to at least $\mathbf{f}_t \sim I(1)$, where $\mathbf{f}_t$ is fully integrated, i.e. $f_{jt} \sim I(1)$ for all j. We then propose a homogenous likelihood ratio test for a unit root in $\mathbf{u}_t$ versus the alternative of $u_{i,t} \sim I(0)$ for all i.[1] Following Zhou and Solberger (2012) we impose the following restrictions.

Assumption 1. The factors admit the VAR(1) representation $\mathbf{f}_t = \mathbf{I}_r\mathbf{f}_{t-1} + \mathbf{e}_t$, where $\mathbf{e}_t \sim \mathcal{N}_r(0, \mathbf{I})$ are i.i.d. (independently and identically distributed).
Assumption 2. The idiosyncratic components admit the VAR(1) representation $\mathbf{u}_t = \rho\mathbf{I}_N\mathbf{u}_{t-1} + \boldsymbol{\varepsilon}_t$, where $\rho \in (-1, 1]$ and where $\boldsymbol{\varepsilon}_t \sim \mathcal{N}_N\left(0, \sigma_\varepsilon^2\mathbf{I}\right)$ are i.i.d. with $\sigma_\varepsilon^2 < \infty$.
Assumption 3. The factor loadings are non-random with $||\boldsymbol{\lambda}_i|| < \infty$.
Assumption 4. The vector processes $\mathbf{e}_t$ and $\boldsymbol{\varepsilon}_t$ are independently distributed such that $E\left(\boldsymbol{\varepsilon}_t\mathbf{e}_t'\right) = \mathbf{0}$ for all t.

Breitung and Das (2008) consider a similar framework, with three null hypotheses: (i) $\mathbf{f}_t \sim I(1), \mathbf{u}_t \sim I(1)$ (ii) $\mathbf{f}_t \sim I(0)$, $\mathbf{u}_t \sim I(1)$ and (iii) $\mathbf{f}_t \sim I(1)$, $\mathbf{u}_t \sim I(0)$. One of their main findings are that none of the test statistics considered seem to handle Case (iii) well, i.e. when there is cross-unit cointegration. From this perspective, our test can be seen as a test for Case (i) versus Case (iii).

Assumptions 1-4 are different from the assumptions in Breitung and Das (2008) on a few accounts. First, we explicitly assume that the factors are independent random walk processes. Second, we are more restrictive since we assume equal variances in the idiosyncratic components, i.e. $E\left(\boldsymbol{\varepsilon}_t\boldsymbol{\varepsilon}_t'\right) = diag\left(\sigma_{1\varepsilon}^2, \sigma_{2\varepsilon}^2, ..., \sigma_{N\varepsilon}^2\right)$ where $\sigma_{i\varepsilon}^2 = \sigma_\varepsilon^2$. This case was also considered by Moon and Perron (2004). Third, we assume homogeneity in the autoregressive parameters, and, fourth, we assume that the stochastic processes are Gaussian, which enables us to formulate the likelihood.

For Assumption 1, the unit variance in $\mathbf{e}_t$ is an identifiability restriction and can be made without loss of generality. Under Assumption 2, the model (2) is also *exact* (or strict) as opposed to *approximate* where some off-diagonal elements of $E\left(\boldsymbol{\varepsilon}_t\boldsymbol{\varepsilon}_t'\right)$ are allowed to be non-zero (c.f. Chamberlain and Rothschild, 1983). Assumptions 3-4 are standard in factor analysis and enable us to identify the covariance structure.

[1] Eickmeier (2005, 2009) offer testing schemes for finding the number of non-stationary factors in the possible presence of non-stationarity in the idiosyncratic components. Feasible procedures are foremost Bai and Ng (2002, 2004). See also general discussion in Bai (2004).

Remark 1 The model (2) is *static* (or restricted), which is a special case of *the generalized dynamic-factor model* of Forni, Hallin, Lippi and Reichlin (2000) where the factor loadings are lag-polynomials. [2] It is well known that the static representation has a state space representation, whereby maximum likelihood estimates can be achieved via iterative methods based on Kalman filters (see e.g. Fernández-Macho, 1997). Such a setup has been considered by e.g. Engle and Watson (1981) and Stock and Watson (1989). In an non-likelihood based setting, the static model has been considered by, among many others, Bai and Ng (2004).

Our interest here is to test the null hypothesis of a unit root in both the factors and the idiosyncratic components versus the alternative hypothesis of stationary idiosyncratics,

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

Because the covariances of the panel data are ill-defined due to the common stochastic trends imposed by the factors, we take first differences. Let Δ be the difference operator such that $\Delta x_{i,t} = x_{i,t} - x_{i,t-1}$, and define, $\mathbf{y}_t \equiv \Delta \mathbf{x}_t = (\Delta x_{1t}, \Delta x_{2t}, ..., \Delta x_{Nt})'$ for $t = 2, 3, ..., T$. Then

$$\mathbf{y}_t = \mathbf{\Lambda} \Delta \mathbf{f}_t + \Delta \mathbf{u}_t, \tag{3}$$

where $\mathbf{y}_t$ and $\Delta \mathbf{u}_t$ are vector stationary processes and $\Delta \mathbf{f_t} = \mathbf{e}_t$ is a i.i.d. vector stationary process. To fully exploit the time-dependence, we stack the panel data. Let $T^* = (T-1)$ and let $\boldsymbol{\xi} = (\mathbf{e}_2, \mathbf{e}_3, ..., \mathbf{e}_T)$ and $\mathbf{V} = (\Delta \mathbf{u}_2, \Delta \mathbf{u}_3, ..., \Delta \mathbf{u}_T)$. Expanding (3) over time we get the differenced panel data, $\mathbf{Y} = (\mathbf{y}_2, \mathbf{y}_3, ..., \mathbf{y}_T) = \mathbf{\Lambda} \boldsymbol{\xi} + \mathbf{V}$, which can be written in stacked form as (see the Appendix)

$$\mathbf{Y}_v = \mathrm{vec}\,(\mathbf{Y}) = (\mathbf{I}_{T^*} \otimes \mathbf{\Lambda})\, \boldsymbol{\xi}_v + \mathbf{V}_v,$$

with covariance matrix, using Assumptions 1-4,

$$\mathbf{\Sigma}_{\mathbf{Y}_v} = \left(\mathbf{I}_{T^*} \otimes \mathbf{\Lambda} \mathbf{\Lambda}' \right) + \mathbf{\Sigma}_{\mathbf{V}_v}. \tag{4}$$

where $\mathbf{\Sigma}_{\mathbf{V}_v} = Cov(\mathbf{V}_v)$. Here the matrix $\mathbf{\Lambda}\mathbf{\Lambda}'$ $(N \times N)$ has at most rank r and is positive semidefinite. Thus, using a reduced singular value decomposition of $\mathbf{\Lambda}$, we may factorize as

$$\mathbf{\Lambda}\mathbf{\Lambda}' = \mathbf{A}\mathbf{H}\mathbf{A}',$$

where $\mathbf{H} = diag\,(\eta_1, \eta_2, ..., \eta_r)$ is positive definite with $\eta_1 \geq \eta_2 \geq ... \geq \eta_r$ and $\mathbf{A}$ $(N \times r)$ is semi-orthogonal, $\mathbf{A}'\mathbf{A} = \mathbf{I}_r$. Under the null hypothesis, also the

[2] Note that the model is still dynamic in the sense that the factors and idiosyncratic components are autoregressive.

differenced idiosyncratic components are i.i.d.-processes, implying $\mathbf{\Sigma}_{\mathbf{V}_v} = \sigma_\varepsilon^2 \mathbf{I}_{NT^*}$. For the stationary case $|\rho| < 1$, let $\mathbf{u}_i = (u_{i1}, u_{i2}, ..., u_{iT})'$ be the $T \times 1$ vector tracking the ith idiosyncratic component over time and in levels. The covariance matrix $Cov(\mathbf{u}_i)$ is well documented and may be found explicitly from Van der Leeuw (1994). If the starting value, u_{i0}, is chosen such that $E(u_{i0}) = 0$ and $Var(u_{i0}) = \frac{\sigma_\varepsilon^2}{1-\rho^2}$, then $Cov(\mathbf{u}_i) = \sigma_\varepsilon^2 \mathbf{\Pi}(\rho)$, where

$$\mathbf{\Pi}(\rho) = \frac{1}{1-\rho^2} \begin{bmatrix} 1 & \rho & \cdots & \rho^{T-1} \\ \rho & 1 & \cdots & \rho^{T-2} \\ \vdots & \vdots & \ddots & \vdots \\ \rho^{T-1} & \rho^{T-2} & \cdots & 1 \end{bmatrix}.$$

Further, let $\mathbf{D}$ be the $T^* \times T$ first-difference-matrix

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

Then $Cov(\mathbf{D}\mathbf{u}_i) = \mathbf{D}E(\mathbf{u}_i\mathbf{u}_i')\mathbf{D}' = \sigma_\varepsilon^2 \mathbf{\Psi}(\rho)$, where

$$\mathbf{\Psi}(\rho) = \frac{1-\rho}{1+\rho} \begin{bmatrix} \frac{2}{1-\rho} & -1 & -\rho & \cdots & -\rho^{T-3} \\ -1 & \frac{2}{1-\rho} & -1 & \cdots & -\rho^{T-4} \\ -\rho & -1 & \ddots & & \vdots \\ \vdots & \vdots & & \ddots & -1 \\ -\rho^{T-3} & -\rho^{T-4} & \cdots & -1 & \frac{2}{1-\rho} \end{bmatrix},$$

and where it is readily verified that $E\left[\text{vec}(\mathbf{V}')\,\text{vec}(\mathbf{V}')'\right] = \left(\mathbf{I}_N \otimes \sigma_\varepsilon^2 \mathbf{\Psi}\right)$. Note that $\mathbf{\Psi}$ properly defines the non-stationary case, as $\mathbf{\Psi}(1) = \mathbf{I}_{T^*}$, and that $\mathbf{\Psi}$ is continuous and twice differentiable around the non-stationary point $\rho = 1$. It is therefore well suited for a likelihood based analysis. Let $\mathbf{K}_{T^*N}$ be a commutation matrix such that $\text{vec}(\mathbf{V}) = \mathbf{K}_{T^*N}\text{vec}(\mathbf{V}')$ (see e.g. Magnus and Neudecker, 2001, p. 47). Then the idiosyncratic variance part in (4) is

$$\begin{aligned} \mathbf{\Sigma}_{\mathbf{V}_v} &= \mathbf{K}_{T^*N} E\left[\text{vec}\left(\mathbf{V}'\right)\text{vec}\left(\mathbf{V}'\right)'\right] \mathbf{K}_{T^*N}' \\ &= \mathbf{K}_{T^*N}\left(\mathbf{I}_N \otimes \sigma_\varepsilon^2 \mathbf{\Psi}\right)\mathbf{K}_{T^*N}' \\ &= \left(\sigma_\varepsilon^2 \mathbf{\Psi} \otimes \mathbf{I}_N\right) \end{aligned}$$

3 The likelihood ratio

Let $L(\boldsymbol{\Theta})$ be the likelihood with the parameter set $\boldsymbol{\Theta} = \left(\text{vec}(\mathbf{A})', \text{vec}(\mathbf{H})', \sigma_\varepsilon^2, \rho\right)$ and let $l(\cdot) = \log L(\cdot)$ be the corresponding log-likelihood. Here we consider the likelihood ratio

$$D(\mathbf{Y}) = \frac{\sup\limits_{\boldsymbol{\Theta}\in\boldsymbol{\Theta_0}} L(\boldsymbol{\Theta}|\mathbf{Y})}{\sup\limits_{\boldsymbol{\Theta}\in\boldsymbol{\Theta}_1} L(\boldsymbol{\Theta}|\mathbf{Y})}\ , \tag{5}$$

where $\boldsymbol{\Theta}_0$ is the parameter space under the null hypothesis, where $\rho = 1$, and $\boldsymbol{\Theta}_1$ is the parameter space under the alternative hypothesis, where $\rho \in (-1, 1)$. Assuming normality, the maximizing arguments may then be found by minimizing the negative log-likelihood

$$-l(\boldsymbol{\Theta}) = \frac{NT^*}{2}\log 2\pi + \frac{1}{2}\log|\boldsymbol{\Sigma}| + \frac{1}{2}\mathbf{Y}_v'\boldsymbol{\Sigma}^{-1}\mathbf{Y}_v\ . \tag{6}$$

To find the maximum likelihood estimates we use the following results (all proofs are in Appendix).

Theorem 1 *Under the null hypothesis the maximum likelihood estimator of* $\mathbf{A}$ *is the set of eigenvectors of* $\mathbf{Z}_0 = \mathbf{Y}\mathbf{Y}'$ *associated with the* r *largest eigenvalues* $\widehat{\varphi}_1(\mathbf{Z}_0) \geq \widehat{\varphi}_2(\mathbf{Z}_0) \geq, ..., \geq \widehat{\varphi}_r(\mathbf{Z}_0)$*, and the maximum likelihood estimators of* σ_ε^2 *and* $\mathbf{H}$ *are*

$$\widehat{\sigma}_\varepsilon^2 = \frac{1}{(N-r)}\frac{1}{T^*}\sum_{i=r+1}^{N}\widehat{\varphi}_i(\mathbf{Z}_0)$$

and

$$\widehat{\mathbf{H}} = \frac{1}{T^*}\widehat{\boldsymbol{\Phi}} - \widehat{\sigma}_\varepsilon^2\mathbf{I}_r$$

where $\widehat{\boldsymbol{\Phi}} = diag\left[\widehat{\varphi}_1(\mathbf{Z}_0), \widehat{\varphi}_2(\mathbf{Z}_0), ..., \widehat{\varphi}_r(\mathbf{Z}_0)\right]$.

Under the alternative hypothesis, the maximum likelihood estimates have to be found numerically. It will then show useful to define

$$\mathbf{C}_1 = \left[\boldsymbol{\Psi}\left(\sigma_\varepsilon^2\mathbf{I}_{T^*} + \sigma_\varepsilon^4\eta^{-1}\boldsymbol{\Psi}\right)\right]^{-1} \tag{7}$$

and

$$\mathbf{C}_2 = \left[\boldsymbol{\Psi}\left(\sigma_\varepsilon^2\mathbf{I}_{T^*} + \sigma_\varepsilon^4\overline{\eta}^{-1}\boldsymbol{\Psi}\right)\right]^{-1}, \tag{8}$$

where $\overline{\eta} = \frac{1}{r}\sum\limits_{i=1}^{r}\eta_i = \frac{1}{r}tr(\mathbf{H})$. When there is only one factor, we have the following result:

Theorem 2 *For* $r = 1$*, the maximum likelihood estimator of* $\mathbf{A}$ *under the alternative hypothesis is given by the estimated eigenvector of* $\mathbf{Z_1} = \mathbf{Y}\mathbf{C_1}\mathbf{Y}'$ *associated with its largest eigenvalue* $\varphi_1(\mathbf{Z}_1)$*, where* $\mathbf{C}_1$ *is defined by (7).*

Theorem 2 puts a restriction on the optimization procedures, which reduces the complexity from evaluating $\mathbf{A}$ as a vector with N parameters to an eigenvector-eigenvalue problem. If we impose this restriction, we then only have to optimize over $\{\mathbf{H}, \sigma_\varepsilon^2, \mathbf{\Psi}\}$. When there are more than one factor, a similar restriction cannot be made. We will then consider the following ML-type estimator:

Proposition 1 *For $r > 1$, a maximum likelihood type estimator of $\mathbf{A}$ under the alternative hypothesis is the set of eigenvectors of $\mathbf{Z_1} = \mathbf{Y}\mathbf{C}_2\mathbf{Y}'$ associated with its r largest eigenvalues $\widehat{\varphi}_1(\mathbf{Z}_1) \geq \widehat{\varphi}_2(\mathbf{Z}_1) \geq ... \geq \widehat{\varphi}_r(\mathbf{Z}_1)$, where $\mathbf{C}_2$ is defined by (8).*

Similar as with Theorem 2, imposing the restriction from Proposition 1 reduces the computational complexity compared to evaluating $\mathbf{A}$ as a matrix with $N \times r$ parameters. This restriction will not provide us with full maximum likelihood estimates, since we have put an arbitrary restriction on the optimization (see the proof in Appendix). However, it is still of interest to see how well this restriction works in terms of power. Let $\widehat{\mathbf{A}}$ denote the MLE of $\mathbf{A}$ for $r = 1$ and likewise the ML-type estimator of $\mathbf{A}$ for $r > 1$. We then have the following result:

Corollary 1 *Let $\widetilde{\mathbf{\Theta}} = (\mathrm{vec}(\widehat{\mathbf{A}})', \mathrm{vec}(\mathbf{H})', \sigma_\varepsilon^2, \rho)$. The log-likelihood (6) with $\mathbf{A} = \widehat{\mathbf{A}}$ is proportional to*

$$\begin{aligned} -l\left(\widetilde{\mathbf{\Theta}}|\mathbf{Y}\right) = \; & T^*(N-r)\log\sigma_\varepsilon^2 + N\log|\mathbf{\Psi}| + \log|\mathbf{R}| \\ & + \frac{1}{\sigma_\varepsilon^2} tr\left(\digamma_1\right) - \frac{1}{\sigma_\varepsilon^2} tr\left(\digamma_2\right) + \digamma_3'\mathbf{R}^{-1}\digamma_3 \end{aligned} \qquad (9)$$

where $\mathbf{R} = \left(\mathbf{\Psi}^{-1} \otimes \mathbf{H}\right) + \sigma_\varepsilon^2 \mathbf{I}_{rT^*}$, $\digamma_1 = \mathbf{Y}\mathbf{\Psi}^{-1}\mathbf{Y}'$, $\digamma_2 = \widehat{\mathbf{A}}'\mathbf{Y}\mathbf{\Psi}^{-1}\mathbf{Y}'\widehat{\mathbf{A}}$ and $\digamma_3 = \mathrm{vec}\left(\widehat{\mathbf{A}}'\mathbf{Y}\mathbf{\Psi}^{-1/2}\right)$.

4 Monte Carlo simulations

We consider the log of the likelihood ratio test (5),

$$\log D(\mathbf{Y}) = \sup_{\mathbf{\Theta} \in \mathbf{\Theta_0}} l\left(\mathbf{\Theta}|\mathbf{Y}\right) - \sup_{\mathbf{\Theta} \in \mathbf{\Theta_1}} l\left(\mathbf{\Theta}|\mathbf{Y}\right). \qquad (10)$$

To implement the test we:

1. Find the explicit solutions for $\mathbf{A}, \mathbf{H}$ and σ_ε^2 under the null from Theorem 1 and calculate the respective log-likelihood.
2. Optimize the conditional likelihood (9) under the alternative and find the respective log-likelihood. The necessary restrictions are, for $r = 1$ (Theorem 2): $\widehat{\mathbf{A}}$ is the left eigenvector associated with the largest eigenvalue of $\mathbf{Y}\mathbf{C}_1\mathbf{Y}'$, and for $r > 1$ (Proposition 1): $\widehat{\mathbf{A}}$ is the set of left eigenvectors associated with the r largest eigenvalues of $\mathbf{Y}\mathbf{C}_2\mathbf{Y}'$.

3. Compute the log-likelihood (10) and compare with corresponding critical values.

Here we find the critical values by simulation.[3] We generate data from equation (1) with

$$f_{j,t} = f_{j,t-1} + e_{j,t}, \quad e_{j,t} \sim \mathcal{N}(0,1), \; f_{j,0} = 0, \forall j, \tag{11}$$

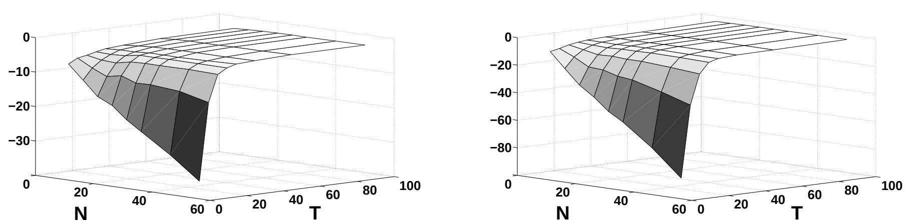
$$u_{i,t} = \rho u_{i,t-1} + \varepsilon_{i,t}, \quad \varepsilon_{i,t} \sim \mathcal{N}(0,1), \; u_{i,0} = 0, \forall i, \tag{12}$$

$$\boldsymbol{\lambda}_i \sim \mathcal{N}_r(\mathbf{0}, \mathbf{I}). \tag{13}$$

where the loadings ($\boldsymbol{\lambda}_i$) are generated only once and are then kept fixed.[4] To find the critical values we let $\rho = 1$ and compute (10) in $10,000$ replications.

Figure 1 shows surface plots of the 5%-critical values for one and two factors for $T \in \{10, 15, 20, 25, 30, 40, 60, 100\}$ and $N \in \{5, 10, 15, 20, 25, 30, 40, 50\}$. Note that the critical values seem to be stabilized for high T. We then perform response surface regressions, where, following e.g. Jönsson (2005), we try different sets of N and T as well as interaction terms. To avoid sensitivity from the choice of loadings and idiosyncratic variance, we start with a rather large T and let $T \in \{30, 40, 60, 100\}$ for $N \in \{5, 10, 15, 20, 25, 30, 40, 50\}$. Let C_α denote the critical value on the significance level α. Here we fit the regression:

$$N^{-1}C_\alpha = \beta_0 + \beta_1 T^{-1} + \beta_2 T^{-1} + \beta_3 N^{-1} + \beta_4 N^{-2} + \beta_5 (NT)^{-1} \tag{14}$$



(a) $r = 1$ (b) $r = 2$

Figure 1. Surface of 5%-critical values for $r = 1$ (a) and $r = 2$ (b).

[3]We use the MATLAB routine *fmincon* from the *optimization toolbox* to find the maximum likelihood estimates under the alternative.

[4]Zhou and Solberger (2012) find that the asymptotic distribution of the LM test is independent of the factor loadings and the idiosyncratic variance. Since the LM test and LR test are asymptotically connected, we conjecture that this independence should hold also here. Initial simulations verify this, but are left out.

Table 1. Estimated coefficients from response surface regressions of critical values

	α	constant	T^{-1}	T^{-2}	N^{-1}	N^{-2}	$(NT)^{-1}$	R^2
$r=1$	1%	-0.043	8.098*	-214.311*	-5.590*	-1.379	-90.872*	0.999
	2.5%	-0.044*	5.631*	-131.628*	-3.594*	-3.159*	-67.865*	1.000
	5%	-0.029*	3.505*	-73.736*	-2.488*	-2.870*	-50.503*	0.999
	10%	-0.032*	2.934*	-52.182*	-1.343*	-2.789*	-36.651*	0.999
$r=2$	1%	-0.068*	12.798*	-406.496*	-5.502*	0.208	-133.564*	0.999
	2.5%	-0.076*	11.124*	-297.861*	-3.807*	-1.287	-97.931*	0.999
	5%	-0.050*	7.057*	-172.698*	-2.623*	-1.366	-74.317*	0.999
	10%	-0.028*	3.697*	-80.184*	-1.531*	-1.929*	-48.407*	0.999

Note: The estimates correspond to the regression (14).* denotes significant on the 1%-level.

Table 1 shows estimated coefficients from the regression (14) for some different α with satisfactory R^2, which may be used to find the critical values for finite samples. Next, we compare size-adjusted power with the Fisher-type test proposed by Bai and Ng (2004),

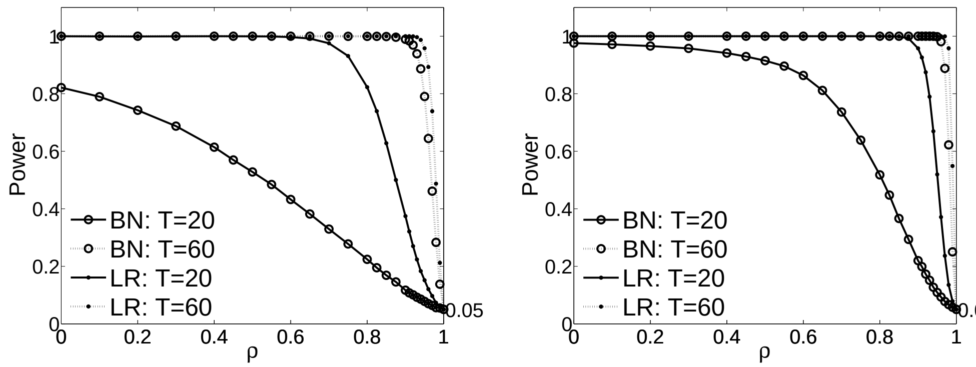
$$P_{\hat{e}}^{c} = \frac{-2\sum_{i=1}^{N}\log p_{\hat{e}}^{c}(i) - 2N}{\sqrt{4N}} \xrightarrow{d} \mathcal{N}(0,1), \text{ as } N \to \infty, \tag{15}$$

where $p_{\hat{e}}^{c}(i)$ are the p-values from individual augmented Dickey-Fuller (ADF) tests.[5] To compare power, we need to locate the p-values for finite dimensions. From Assumption 2, $u_{i,t} = \rho u_{i,t-1} + \varepsilon_{i,t}$ where $\varepsilon_{i,t} \sim \mathcal{N}\left(0, \sigma_{\varepsilon}^{2}\right)$ for all i. For any i, the t-statistic for testing $\rho = 1$ can be written as (see e.g. Hamilton, 1994, pp. 487-489)

$$t_T = \frac{T^{-1}\sum_{t=1}^{T} u_{t-1}\varepsilon_t}{\{T^{-2}\sum_{t=1}^{T} u_{t-1}^2\}^{1/2}\{s_T^2\}^{1/2}} \xrightarrow{d} \frac{(1/2)\{[W(1)]^2 - 1\}}{\{\int_0^1 [W(s)]^2\, ds\}^{1/2}}, \text{ as } T \to \infty, \tag{16}$$

where $s_T^2 = \sum_{t=1}^{T}(u_t - \hat{\rho}u_{t-1})^2/(T-1)$, $\hat{\rho} = \frac{\sum_{t=1}^{T} u_{t-1}u_t}{\sum_{t=1}^{T} u_{t-1}^2}$ is the OLS estimate and $W(s)$ denotes a standard Brownian motion defined on $s \in [0,1]$. Here we simulate the distribution for the univariate ADF-statistic (16) in $10,000$ replications for any specific T, letting $\rho = 1$, $\varepsilon_t \sim \mathcal{N}(0,1)$ and setting the starting value to $u_0 = 0$. We then map the corresponding finite sample p-values. To find the power we then

[5] The procedure in Bai and Ng (2004) is to extract the factors and the factor loadings using principal components on the differenced panel data, leaving the idiosyncratic components as residuals. Dickey-Fuller regressions (without constant) are then run on the re-accumulated idiosyncratics. As we consider AR(1) processes, we choose no lags in the Dickey-Fuller regressions.



(a) $N = 10$ (b) $N = 30$

Figure 2. Size-adjusted power for LR and BN when $r = 1$ for $N = 10$ (a) and $N = 30$ (b).

simulate data from (11-13) with $\rho \in [0, 1]$ in 5,000 replications. Figure 2 shows the size-adjusted power of the log-likelihood ratio test (10) (LR hereafter) and the pooled test (15) (BN hereafter) for $N = \{10, 30\}$ and $T = \{20, 60\}$ when $r = 1$.

Note that, for $T = 20$ and as soon as ρ is less than 1, LR has quite substantially higher power than BN, while it is less obvious for $T = 60$ and close to the unit root. We therefore study the local power using a panel near-unit-root model (see e.g. Breitung and Das, 2005). We let

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

for $c = 1$ and $c = 5$. Zhou and Solberger (2012) found similar power properties of the analogous LM test, as we find here for the LR test. Because the LM test only requires estimation under the null, and its asymptotic distribution is known, we also compare the local power with the LM test.

The LM statistic in Zhou and Solberger (2012) is

$$\vartheta = \frac{(T-1)\, tr(\mathbf{S}_{01}^{-1}) - 2tr(\mathbf{S}_{01}^{-1}\mathbf{S}_0\mathbf{S}_{01}^{-1}) + tr(\mathbf{S}_{01}^{-1}\mathbf{S}_{00}\mathbf{S}_{01}^{-1})}{\sqrt{2\,(T-1)\,(T-2)\, tr(\mathbf{S}_{01}^{-1}\mathbf{S}_{01}^{-1})}} \tag{17}$$

where $\mathbf{S}_0 = \mathbf{Y}\mathbf{Y}'$, $\mathbf{S}_{00} = \Big(\sum_{t=2}^{T}\mathbf{y}_t\Big)\Big(\sum_{t=2}^{T}\mathbf{y}_t\Big)'$ and $\mathbf{S}_{01} = \widehat{\mathbf{A}}_0\widehat{\mathbf{H}}_0\widehat{\mathbf{A}}_0' + \widehat{\sigma}^2_{\varepsilon,0}\mathbf{I}_N$ and where $\widehat{\mathbf{A}}_0$, $\widehat{\mathbf{H}}_0$, $\widehat{\sigma}^2_{\varepsilon,0}$ are the estimators under the null from Theorem 1. For any fixed N, (17) is shown to have an asymptotic distribution as $T \to \infty$ that is a

Table 2. Size-adjusted local power (%) for LR, LM and BN, one factor

N	T	LR		LM		BN	
		$c=1$	$c=5$	$c=1$	$c=5$	$c=1$	$c=5$
5	15	6.4	18.2	9.4	30.2	7.1	19.8
	30	8.3	29.3	10.5	37.8	7.7	25.9
	60	9.3	43.6	10.6	41.6	8.2	32.6
	120	10.1	49.7	10.5	45.8	8.6	35.3
20	15	6.0	21.4	10.9	60.2	8.4	30.3
	30	10.0	64.1	11.8	71.4	8.6	38.8
	60	12.6	77.6	13.1	79.3	8.2	41.4
	120	12.9	80.9	14.4	82.3	9.0	45.2
80	15	5.9	16.8	8.5	58.0	7.7	31.8
	30	11.9	81.3	13.6	85.3	8.6	43.3
	60	13.4	89.3	13.8	89.9	9.5	50.6
	120	14.8	91.5	15.9	92.0	9.6	51.1

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

weighted sum of independent χ_1^2 (chi-square with one degrees of freedom) variables,

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

where $w_i = \begin{cases} \frac{\sigma_\varepsilon^2}{\eta_i + \sigma_\varepsilon^2} & \text{for } 1 \le i \le r \\ 1 & \text{for } r < i \le N \end{cases}$ and where η_i are the diagonal entries of $\mathbf{H}$. When $T, N \to \infty$, the asymptotic distribution of (17) is standard normal.

The size-adjusted local power for LR, LM and BN for one factor is tabulated in Table 2. For large T and N it seems that LR and LM are roughly equivalent in terms of local size-adjusted power, and both are more locally powerful than BN. However, for small T, LM is more powerful than LR. Also, for small T and large N, LR seems to be less powerful than BN. In Table 3, we have tabulated the local size-adjusted power for two factors imposing the restriction from Proposition 1. It seems that this restriction works well in terms power, but the pattern that LM is more powerful for small T, while roughly equivalent for large T, is persistent. This is in favor of LM as it is much faster to compute and the asymptotic distribution is known.

Table 3. Size-adjusted local power (%) for LR, LM and BN, two factors

N	T	LR		LM		BN	
		$c = 1$	$c = 5$	$c = 1$	$c = 5$	$c = 1$	$c = 5$
10	20	6.3	17.8	9.0	38.7	7.3	22.2
	40	9.2	44.9	11.4	54.5	8.2	32.0
	80	10.1	55.8	11.2	58.9	8.4	36.6
25	20	6.5	19.8	11.0	59.1	7.8	29.4
	40	10.8	69.2	12.3	75.3	8.9	39.2
	80	12.9	79.6	13.1	81.7	8.5	41.8
50	20	6.1	17.4	9.0	56.8	7.5	30.9
	40	11.5	77.6	13.3	82.6	8.6	40.5
	80	14.8	87.8	14.8	88.3	8.7	45.3

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

5 Conclusions

In this paper we consider an exact static factor model and propose a likelihood ratio test for idiosyncratic unit roots conditional on r non-stationary factors. Using maximum likelihood to test for non-stationarity in this framework has long been deemed too computationally expensive. The restrictions we impose here makes the test feasible, though under the alternative hypothesis the likelihood has to be found numerically. By simulation, we demonstrate that, for relatively large T, the likelihood ratio test is more powerful than the Fisher-type pooled test of Bai and Ng (2004) when the factors are integrated, and about equally powerful to the LM test proposed by Zhou and Solberger (2012). However, for very small T, the LM test is more powerful than both the LR test and the Fisher-type test. In conclusion, this is in favor of the LM test, over the analogous LR test proposed here, as the LM test does not need to be estimated under the alternative hypothesis, and as such, it has an explicit test-statistic with a known asymptotic distribution.

Acknowledgements

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

Appendix Proofs

We make use of the following results (A1,A2,A3), that can be found in e.g. Magnus and Neudecker (2001, pp. 28-31). For a matrix $\mathbf{A} \in \mathbb{R}^{m\times n}$ with $\mathbf{a}_i$ as its ith column vector, $\text{vec}(\cdot)$ is the vectorization operator such that $\text{vec}(\mathbf{A}) = (\mathbf{a}_1', \mathbf{a}_2', ..., \mathbf{a}_n')' \in \mathbb{R}^{mn\times 1}$. For additional matrices $\mathbf{B} \in \mathbb{R}^{n\times q}$, $\mathbf{C} \in \mathbb{R}^{q\times k}$, $\mathbf{D} \in \mathbb{R}^{k\times r}$, we have that

$$\text{vec}(\mathbf{ABC}) = (\mathbf{C}' \otimes \mathbf{A})\,\text{vec}(\mathbf{B}), \tag{A1}$$

$$tr(\mathbf{ABCD}) = \text{vec}(\mathbf{D}')'(\mathbf{C}' \otimes \mathbf{A})\,\text{vec}(\mathbf{B}), \tag{A2}$$

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

$$(\mathbf{A} \otimes \mathbf{B})^{-1} = \left(\mathbf{A}^{-1} \otimes \mathbf{B}^{-1}\right). \tag{A3}$$

Proof of Theorem 1. For optimization of (6) we may remove any redundant constants and optimize

$$-l(\mathbf{\Theta}) = \log|\mathbf{\Sigma}| + \mathbf{Y}_v'\mathbf{\Sigma}^{-1}\mathbf{Y}_v\ . \tag{A4}$$

Under the null hypothesis we have that $\mathbf{\Psi}(1) = \mathbf{I}_{T^*}$. Hence, then

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

where $\mathbf{\Omega} = \mathbf{\Lambda\Lambda}' + \sigma_\varepsilon^2\mathbf{I}_N$. Thus, using results (A2-A3),

$$\begin{aligned} -l(\mathbf{\Theta}) &= \log|\mathbf{\Omega}|^{T^*} + \text{vec}(\mathbf{Y})'\left(\mathbf{I}_{T^*} \otimes \mathbf{\Omega}^{-1}\right)\text{vec}(\mathbf{Y}) \\ &= T^*\log|\mathbf{\Omega}| + T^* tr\left(\mathbf{\Omega}^{-1}\frac{1}{T^*}\mathbf{YY}'\right). \end{aligned} \tag{A5}$$

It is a standard result that (A5) is minimized for $\widehat{\mathbf{\Omega}} = \frac{1}{T^*}\mathbf{YY}'$ (see e.g. Anderson, 2003, p. 69), and the maximum likelihood estimators are well known to be $\widehat{\sigma}_\varepsilon^2 = \frac{1}{(N-r)}\sum_{i=r+1}^{N}\widehat{\varphi}_i\left(\widehat{\mathbf{\Omega}}\right)$ and $\widehat{\eta}_i = \widehat{\varphi}_i - \widehat{\sigma}_\varepsilon^2$ (see e.g. Stoica and Jansson, 2009). ■

Proof of Theorem 2. Working from (A4), let $\mathbf{M} = (\mathbf{\Psi} \otimes \mathbf{I}_N)$, which is square $NT^* \times NT^*$. Then, using a common identity for determinants (see e.g. Mardia, Kent and Bibby, 1979, p. 457, Result VII),

$$\begin{aligned} |\mathbf{\Sigma}| &= \sigma_\varepsilon^{2NT^*}|\mathbf{M}^{1/2}\left[\left(\mathbf{\Psi}^{-1/2} \otimes \mathbf{H}^{1/2}\mathbf{A}'\right)\left(\mathbf{\Psi}^{-1/2} \otimes \mathbf{AH}^{1/2}\right)/\sigma_\varepsilon^2 + \mathbf{I}_{rT^*}\right]\mathbf{M}^{1/2}| \\ &= \sigma_\varepsilon^{2T^*(N-r)}|\mathbf{M}^{1/2}||\left(\mathbf{\Psi}^{-1} \otimes \mathbf{H}\right) + \sigma_\varepsilon^2\mathbf{I}_{rT^*}||\mathbf{M}^{1/2}| \\ &= \sigma_\varepsilon^{2T^*(N-r)}|\mathbf{M}||\mathbf{R}| \end{aligned}$$

say, with $\mathbf{R} = \left(\mathbf{\Psi}^{-1} \otimes \mathbf{H}\right) + \sigma_\varepsilon^2\mathbf{I}_{rT^*}$. Note that this determinant is not decided by the elements of $\mathbf{A}$. It is straightforward to show that $\mathbf{R}$ is symmetric positive definite. Let $\mathbf{R} = \mathbf{B} + \sigma_\varepsilon^2\mathbf{I}$, where $\mathbf{B} = (\mathbf{\Psi}^{-1} \otimes \mathbf{H})$ is symmetric positive definite. By definition $|\mathbf{B} - \varphi\mathbf{I}|$

$= |\mathbf{B}+\sigma_\varepsilon^2\mathbf{I}-\left(\varphi+\sigma_\varepsilon^2\right)\mathbf{I}| = 0$ implying $\varphi_i\left(\mathbf{B}+\sigma_\varepsilon^2\mathbf{I}\right) = \varphi_i\left(\mathbf{B}\right)+\sigma_\varepsilon^2$, where $\varphi_i\left(\mathbf{B}\right), \sigma_\varepsilon^2 > 0$ $\forall i \in 1, ..., rT^*$. It then follows that $\mathbf{R}$ is invertible, with the inverse symmetric positive definite.

Making repeated use of the matrix inversion lemma (see e.g. Mardia, Kent and Bibby, 1979, p. 459, Result V), we have that

$$\begin{aligned}
\mathbf{\Sigma}^{-1} &= \frac{1}{\sigma_\varepsilon^2}\mathbf{M}^{-1/2}\left[\frac{1}{\sigma_\varepsilon^2}\left(\mathbf{\Psi}^{-1}\otimes\mathbf{AHA}'\right)+\mathbf{I}_{NT^*}\right]^{-1}\mathbf{M}^{-1/2} \\
&= \frac{1}{\sigma_\varepsilon^2}\mathbf{M}^{-1/2}\left[\left(\mathbf{I}_{T^*}\otimes\mathbf{A}\right)\frac{1}{\sigma_\varepsilon^2}\left(\mathbf{\Psi}^{-1}\otimes\mathbf{H}\right)\left(\mathbf{I}_{T^*}\otimes\mathbf{A}'\right)+\mathbf{I}_{NT^*}\right]^{-1}\mathbf{M}^{-1/2} \\
&= \frac{1}{\sigma_\varepsilon^2}\mathbf{M}^{-1/2}\left\{\mathbf{I}_{NT^*}-\left(\mathbf{I}_{T^*}\otimes\mathbf{A}\right)\left[\sigma_\varepsilon^2\left(\mathbf{\Psi}^{-1}\otimes\mathbf{H}\right)^{-1}+\mathbf{I}_{rT^*}\right]^{-1}\left(\mathbf{I}_{T^*}\otimes\mathbf{A}'\right)\right\}\mathbf{M}^{-1/2} \\
&= \frac{1}{\sigma_\varepsilon^2}\mathbf{M}^{-1/2}\left\{\mathbf{I}_{NT^*}-\left(\mathbf{I}_{T^*}\otimes\mathbf{A}\right)\left\{\mathbf{I}_{rT^*}-\left[\mathbf{I}_{rT^*}+\frac{1}{\sigma_\varepsilon^2}\left(\mathbf{\Psi}^{-1}\otimes\mathbf{H}\right)\right]^{-1}\right\}\left(\mathbf{I}_{T^*}\otimes\mathbf{A}'\right)\right\}\mathbf{M}^{-1/2} \\
&= \left(\mathbf{\Psi}^{-1}\otimes\mathbf{I}_N\right)/\sigma_\varepsilon^2-\left(\mathbf{\Psi}^{-1}\otimes\mathbf{AA}'\right)/\sigma_\varepsilon^2+\left(\mathbf{\Psi}^{-1/2}\otimes\mathbf{A}\right)\mathbf{R}^{-1}\left(\mathbf{\Psi}^{-1/2}\otimes\mathbf{A}'\right).
\end{aligned}$$

Inserting these results into (A4) we get

$$\begin{aligned}
-l\left(\mathbf{\Theta}\right) = {} & T^*\left(N-r\right)\log\sigma_\varepsilon^2+N\log|\mathbf{\Psi}|+\log|\mathbf{R}| \\
& +\frac{1}{\sigma_\varepsilon^2}tr\left\{\left[\mathbf{I}_{NT^*}-\left(\mathbf{I}_{T^*}\otimes\mathbf{AA}'\right)\right]\widetilde{\mathbf{S}}\right\} \\
& +tr\left[\mathbf{R}^{-1}\left(\mathbf{I}_{T^*}\otimes\mathbf{A}'\right)\widetilde{\mathbf{S}}\left(\mathbf{I}_{T^*}\otimes\mathbf{A}\right)\right],
\end{aligned} \tag{A6}$$

where $\widetilde{\mathbf{S}} = \text{vec}(\mathbf{Y\Psi}^{-1/2})\text{vec}(\mathbf{Y\Psi}^{-1/2})'$.

The term in (A6) that is dependent on $\mathbf{A}$ is

$$\begin{aligned}
& tr\left[\mathbf{R}^{-1}\left(\mathbf{I}_{T^*}\otimes\mathbf{A}'\right)\widetilde{\mathbf{S}}\left(\mathbf{I}_{T^*}\otimes\mathbf{A}\right)\right]-\frac{1}{\sigma_\varepsilon^2}tr\left[\left(\mathbf{I}_{T^*}\otimes\mathbf{AA}'\right)'\widetilde{\mathbf{S}}\right] \\
&= tr\left[\left(\mathbf{R}^{-1}-\mathbf{I}_{rT^*}/\sigma_\varepsilon^2\right)\left(\mathbf{I}_{T^*}\otimes\mathbf{A}'\right)\widetilde{\mathbf{S}}\left(\mathbf{I}_{T^*}\otimes\mathbf{A}\right)\right] \\
&= -tr\left\{\left[\left(\mathbf{R}-\sigma_\varepsilon^2\mathbf{I}_{rT^*}\right)\mathbf{R}^{-1}/\sigma_\varepsilon^2\right]\left(\mathbf{I}_{T^*}\otimes\mathbf{A}'\right)\widetilde{\mathbf{S}}\left(\mathbf{I}_{T^*}\otimes\mathbf{A}\right)\right\} \\
&= -tr\left(\mathbf{WQ}\right),
\end{aligned} \tag{A7}$$

with $\mathbf{W} = \left(\mathbf{\Psi}^{-1}\otimes\mathbf{H}\right)\mathbf{R}^{-1}/\sigma_\varepsilon^2 = \left[\mathbf{I}_{rT^*}+\sigma_\varepsilon^2\left(\mathbf{\Psi}^{-1}\otimes\mathbf{H}\right)^{-1}\right]^{-1}/\sigma_\varepsilon^2$, which is positive definite, and $\mathbf{Q} = \text{vec}\left(\mathbf{A}'\mathbf{Y\Psi}^{-1/2}\right)\text{vec}\left(\mathbf{A}'\mathbf{Y\Psi}^{-1/2}\right)'$, which is positive semi-definite.

We want to minimize (A6) with respect to $\mathbf{A}$, i.e. maximize $tr\left(\mathbf{WQ}\right)$. If $r = 1$, then $\mathbf{H} = \eta$ is a scalar. Hence, then

$$\begin{aligned}
\mathbf{W} &= \left[\mathbf{I}_{T^*}+\sigma_\varepsilon^2\left(\mathbf{\Psi}\otimes\eta^{-1}\right)\right]^{-1}/\sigma_\varepsilon^2 \\
&= \left(\sigma_\varepsilon^2\mathbf{I}_{T^*}+\sigma_\varepsilon^4\eta^{-1}\mathbf{\Psi}\right)^{-1}.
\end{aligned}$$

Also, because $\mathbf{A}$ is $N \times 1$ and hence $\mathbf{A}'\mathbf{Y}\mathbf{\Psi}^{-1/2}$ is a $1 \times T^*$ vector,

$$\begin{aligned} \mathbf{Q} &= \text{vec}\left(\mathbf{A}'\mathbf{Y}\mathbf{\Psi}^{-1/2}\right)\text{vec}\left(\mathbf{A}'\mathbf{Y}\mathbf{\Psi}^{-1/2}\right)' \\ &= \mathbf{\Psi}^{-1/2}\mathbf{Y}'\mathbf{A}\mathbf{A}'\mathbf{Y}\mathbf{\Psi}^{-1/2}. \end{aligned}$$

Thus

$$\begin{aligned} tr\left(\mathbf{W}\mathbf{Q}\right) &= \mathbf{A}'\mathbf{Y}\mathbf{\Psi}^{-1/2}\left(\sigma_\varepsilon^2\mathbf{I}_{T^*} + \sigma_\varepsilon^4\eta^{-1}\mathbf{\Psi}\right)^{-1}\mathbf{\Psi}^{-1/2}\mathbf{Y}'\mathbf{A} \\ &= \mathbf{A}'\mathbf{C}_1\mathbf{A}, \end{aligned}$$

with $\mathbf{C}_1 = \mathbf{Y}\left[\mathbf{\Psi}\left(\sigma_\varepsilon^2\mathbf{I}_{T^*} + \sigma_\varepsilon^4\eta^{-1}\mathbf{\Psi}\right)\right]^{-1}\mathbf{Y}'$. It then follows from the Poincaré separation theorem (see e.g. Abadir and Magnus, 2005, p. 347ff) that

$$\max_{\mathbf{A}'\mathbf{A}=\mathbf{I}_r} tr(\mathbf{A}'\mathbf{C}_1\mathbf{A}) = \varphi_1\left(\mathbf{C}_1\right),$$

where the maximum is attained when $\mathbf{A}$ is the eigenvector of $\mathbf{C}_1$ associated with the largest eigenvalue $\varphi_1\left(\mathbf{C}_1\right)$. ■

Proof of proposition 1. Following the proof of Theorem 1, and arbitrarily replacing $\mathbf{H}$ with $\overline{\eta} = \frac{1}{r}tr(\mathbf{H})$ in (A7) yields

$$\begin{aligned} \mathbf{W} &= \left[\mathbf{I}_{rT^*} + \sigma_\varepsilon^2\left(\mathbf{\Psi}\otimes\overline{\eta}^{-1}\mathbf{I}_r\right)\right]^{-1}/\sigma_\varepsilon^2 \\ &= \left[\left(\sigma_\varepsilon^4\overline{\eta}^{-1}\mathbf{\Psi} + \sigma_\varepsilon^2\mathbf{I}_{T^*}\right)\otimes\mathbf{I}_r\right]^{-1} \\ &= \left[\left(\sigma_\varepsilon^4\overline{\eta}^{-1}\mathbf{\Psi} + \sigma_\varepsilon^2\mathbf{I}_{T^*}\right)^{-1}\otimes\mathbf{I}_r\right] \\ &= \left(\mathbf{C}\otimes\mathbf{I}_r\right), \end{aligned}$$

where $\mathbf{C} = \left(\sigma_\varepsilon^4\overline{\eta}^{-1}\mathbf{\Psi} + \sigma_\varepsilon^2\mathbf{I}_{T^*}\right)^{-1}$. Then, using (A2),

$$\begin{aligned} tr\left(\mathbf{W}\mathbf{Q}\right) &= \text{vec}\left(\mathbf{A}'\mathbf{Y}\mathbf{\Psi}^{-1/2}\right)'\left(\mathbf{C}\otimes\mathbf{I}_r\right)\text{vec}\left(\mathbf{A}'\mathbf{Y}\mathbf{\Psi}^{-1/2}\right) \\ &= tr\left[\mathbf{A}'\mathbf{Y}\mathbf{\Psi}^{-1/2}\mathbf{C}\mathbf{\Psi}^{-1/2}\mathbf{Y}'\mathbf{A}\right] \\ &= tr\left[\mathbf{A}'\mathbf{C}_2\mathbf{A}\right], \end{aligned}$$

where $\mathbf{C}_2 = \mathbf{Y}\left[\mathbf{\Psi}\left(\sigma_\varepsilon^4\overline{\eta}^{-1}\mathbf{\Psi} + \sigma_\varepsilon^2\mathbf{I}_{T^*}\right)\right]^{-1}\mathbf{Y}'$. Again, using the Poincaré separation theorem,

$$\max_{\mathbf{A}'\mathbf{A}=\mathbf{I}_r} tr(\mathbf{A}'\mathbf{C}_2\mathbf{A}) = \sum_{j=1}^{r}\varphi_j\left(\mathbf{C}_2\right),$$

where the maximum is attained when $\mathbf{A}$ are the eigenvectors of $\mathbf{C}_2$ associated with the r largest eigenvalues $\varphi_1\left(\mathbf{C}_2\right), \varphi_2\left(\mathbf{C}_2\right), ..., \varphi_r\left(\mathbf{C}_2\right)$. ■

Proof of Corollary 1. For the fourth term in (A6) with $\mathbf{A} = \widehat{\mathbf{A}}$ we have that, because $\left(\mathbf{I}_N - \widehat{\mathbf{A}}\widehat{\mathbf{A}}'\right)$ is symmetric idempotent,

$$\begin{aligned}
&\frac{1}{\sigma_\varepsilon^2} tr\left\{\left[\mathbf{I}_{NT^*} - \left(\mathbf{I}_{T^*} \otimes \widehat{\mathbf{A}}\widehat{\mathbf{A}}'\right)\right]\widetilde{\mathbf{S}}\right\} = \frac{1}{\sigma_\varepsilon^2} tr\left\{\left[\mathbf{I}_{T^*} \otimes \left(\mathbf{I}_N - \widehat{\mathbf{A}}\widehat{\mathbf{A}}'\right)\right]\widetilde{\mathbf{S}}\right\} \\
&= \frac{1}{\sigma_\varepsilon^2} \text{vec}\left[\left(\mathbf{I}_N - \widehat{\mathbf{A}}\widehat{\mathbf{A}}'\right)\mathbf{Y}\boldsymbol{\Psi}^{-1/2}\right]' \text{vec}\left[\left(\mathbf{I}_N - \widehat{\mathbf{A}}\widehat{\mathbf{A}}'\right)\mathbf{Y}\boldsymbol{\Psi}^{-1/2}\right] \\
&= \frac{1}{\sigma_\varepsilon^2} tr\left[\boldsymbol{\Psi}^{-1/2}\mathbf{Y}'\left(\mathbf{I}_N - \widehat{\mathbf{A}}\widehat{\mathbf{A}}'\right)\mathbf{Y}\boldsymbol{\Psi}^{-1/2}\right] \\
&= \frac{1}{\sigma_\varepsilon^2}\left[tr\left(\mathbf{Y}\boldsymbol{\Psi}^{-1}\mathbf{Y}'\right) - tr\left(\widehat{\mathbf{A}}'\mathbf{Y}\boldsymbol{\Psi}^{-1}\mathbf{Y}'\widehat{\mathbf{A}}\right)\right].
\end{aligned}$$

For the fifth term in (A6) with $\mathbf{A} = \widehat{\mathbf{A}}$ we have that

$$\begin{aligned}
tr\left[\mathbf{R}^{-1}\left(\mathbf{I}_{T^*} \otimes \widehat{\mathbf{A}}'\right)\widetilde{\mathbf{S}}\left(\mathbf{I}_{T^*} \otimes \widehat{\mathbf{A}}\right)\right] &= tr\left[\mathbf{R}^{-1}\text{vec}\left(\widehat{\mathbf{A}}'\mathbf{Y}\boldsymbol{\Psi}^{-1/2}\right)\text{vec}\left(\widehat{\mathbf{A}}'\mathbf{Y}\boldsymbol{\Psi}^{-1/2}\right)'\right] \\
&= \text{vec}\left(\widehat{\mathbf{A}}'\mathbf{Y}\boldsymbol{\Psi}^{-1/2}\right)'\mathbf{R}^{-1}\text{vec}\left(\widehat{\mathbf{A}}'\mathbf{Y}\boldsymbol{\Psi}^{-1/2}\right).
\end{aligned}$$

Inserting these results into (A6) and rearranging we get (9). ■

References

Abadir, K.M., Magnus, J.R. (2005), *Matrix Algebra,* Cambridge University Press, Cambridge.

Anderson, T.W. (2003), *An Introduction to Multivariate Statistical Analysis*, Third Edition, Wiley, New York.

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

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

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

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

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

Camba-Mendez, G., Kapetanios, G., Smith, R.J., Weale, M.R. (2001), "An automatic leading indicator of economic activity: forecasting GDP growth for European countries", *Econometrics Journal*, 4, S56-S90.

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

Eickmeier, S. (2005), "Common stationary and non-stationary factors in the euro area analyzed in a large-scale factor model", *Bundesbank Discussion Paper*, 02/2005.

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

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

Fernández-Macho, F.J. (1997), "A dynamic factor model for economic time series", *Kybernetika*, 33, 583-606.

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

Hamilton, J.D. (1994), *Time Series Analysis*, Princeton University Press, Princeton, New Jersey.

Jungbacker, B, Koopman S.J. (2008), "Likelihood-based analysis for dynamic factor models", *Tinbergen Institute Discussion paper*, TI 2008-0007/4.

Jönsson, K. (2005), "Cross-sectional dependency and size distortion in a small-sample homogenous panel data unit root test", *Oxford Bulletin of Economics and Statistics*, 67, 369-392.

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

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

Mardia, K.V., Kent, J.T., Bibby, J.M. (1979), *Multivariate Analysis*, Academic press, London.

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

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

Stock, J.H., Watson, M. W. (1989), "New indexes of coincident and leading economic indicators", *NBER Macroeconomics Journal*, 4, 351-394.

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

Watson, M.W., Engle, R.F. (1983), "Alternative algorithms for the estimation of dynamic factor, mimic and varying coefficient regression models", *Journal of Econometrics*, 23, 385-400.

Zhou, X., Solberger, M. (2012), "An LM-type test for idiosyncratic unit roots in a dynamic-factor model with integrated factors", *Mimeo*.