UPPSALA
UNIVERSITET

Digital Comprehensive Summaries of Uppsala Dissertations from the Faculty of Social Sciences 103

Likelihood-Based Panel Unit Root Tests for Factor Models

XINGWU ZHOU

ACTA
UNIVERSITATIS
UPSALIENSIS
UPPSALA
2014

ISSN 1652-9030
ISBN 978-91-554-9049-2
urn:nbn:se:uu:diva-233094

Dissertation presented at Uppsala University to be publicly examined in Hörsal 2, Ekonomikum, Kyrkogårdsgatan 10, Uppsala, Friday, 14 November 2014 at 13:15 for the degree of Doctor of Philosophy. The examination will be conducted in English. Faculty examiner: Professor Joakim Westerlund (Lund University).

Abstract

Zhou, X. 2014. Likelihood-Based Panel Unit Root Tests for Factor Models. *Digital Comprehensive Summaries of Uppsala Dissertations from the Faculty of Social Sciences* 103. 35 pp. Uppsala: Acta Universitatis Upsaliensis. ISBN 978-91-554-9049-2.

The thesis consists of four papers that address likelihood-based unit root tests for panel data with cross-sectional dependence arising from common factors.

In the first three papers, we derive Lagrange multiplier (LM)-type tests for common and idiosyncratic unit roots in the exact factor models based on the likelihood function of the differenced data. Also derived are the asymptotic distributions of these test statistics. The finite sample properties of these tests are compared by simulation with other commonly used unit root tests. The results show that our LM-type tests have better size and local power properties.

In the fourth paper, we estimate the spaces spanned by the common factors and the spaces spanned by the idiosyncratic components of the static factor model by using the quasi-maximum likelihood (ML) method and compare it with the widely used method of principal components (PC). Next, by simulation, we compare the size and power properties of established tests for idiosyncratic unit roots, using both the ML and PC methods. Simulation results show that the idiosyncratic unit root tests based on the likelihood-based residuals generally have better size and higher size-adjusted power, especially when the cross-sectional dimension is small and the time series dimension is large.

Keywords: Panel unit root, Exact factor model, Dynamic factor model, Maximum likelihood, Principal components, Lagrange multiplier

Xingwu Zhou, Department of Statistics, Uppsala University, SE-75120 Uppsala, Sweden.

ISSN 1652-9030
ISBN 978-91-554-9049-2
urn:nbn:se:uu:diva-233094 (http://urn.kb.se/resolve?urn=urn:nbn:se:uu:diva-233094)

List of papers

This thesis is based on the following papers, which are referred to in the text by their Roman numerals.

I Zhou, X. and Solberger, M. (2012) *An LM-Type Test for Idiosyncratic Unit Roots in the Exact Factor Model with Nonstationary Common Shocks*. Submitted.

II Solberger, M. and Zhou, X. (2013) *LM-Type Tests for Idiosyncratic and Common Unit Roots in the Exact Factor Model with AR(1) Dynamics*.

III Zhou, X. and Solberger, M. (2013) *A Lagrange Multiplier-Type Test for Idiosyncratic Unit Roots in the Exact Factor Model*. Submitted.

IV Zhou, X. (2014) *Comparing Idiosyncratic Unit Root Tests Based on the Residuals Estimated from ML and PC Methods in a Static Factor Model*.

Contents

1. Introduction

1.1 Lagrange multiplier test

The Lagrange multiplier (LM) test (Silvey, 1959), the Wald test (Wald, 1943) and the likelihood ratio (LR) test (Neyman and Pearson, 1928) are three commonly used likelihood-based tests for hypotheses testing in econometrics. It is well known that in well-behaved problems these three tests are asymptotically equivalent and asymptotically locally most powerful (e.g., Engle, 1984).

For the LM test, it suffices to estimate the model under the null hypothesis, which is a major advantage since in many cases estimation under the null is computationally convenient. This is in contrast to the Wald test which is based on the estimates of the model under the alternative hypothesis, and the LR test which requires the estimates of parameters for both the restricted and the unrestricted models.

Let $l(\boldsymbol{\theta})$ be the log likelihood function of a $k \times 1$ parameter vector $\boldsymbol{\theta}$, then the score is

$$\boldsymbol{s}(\boldsymbol{\theta}) = \frac{\partial l(\boldsymbol{\theta}|\boldsymbol{x})}{\partial \boldsymbol{\theta}},$$

and the information matrix is

$$\boldsymbol{\mathcal{I}}(\boldsymbol{\theta}) = -\mathbb{E}\left[\frac{\partial^2}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}'} l(\boldsymbol{x};\boldsymbol{\theta})|\boldsymbol{\theta}\right].$$

Let $\widetilde{\boldsymbol{\theta}}$ be the maximum likelihood (ML) estimator of $\boldsymbol{\theta}$ subject to an $r \times 1$ vector of constraints $\boldsymbol{h}(\boldsymbol{\theta}) = \mathbf{0}$. The LM test is derived from the maximization of the Lagrangian function

$$\mathcal{L} = l(\boldsymbol{\theta}) - \boldsymbol{\lambda}' \boldsymbol{h}(\boldsymbol{\theta}),$$

where $\boldsymbol{\lambda}$ is an $r \times 1$ vector of Lagrange multipliers. According to the first order conditions

$$\begin{aligned} \frac{\partial \mathcal{L}}{\partial \boldsymbol{\theta}}|_{\boldsymbol{\theta}=\widetilde{\boldsymbol{\theta}}} &= \boldsymbol{s}(\widetilde{\boldsymbol{\theta}}) - \boldsymbol{H}(\widetilde{\boldsymbol{\theta}})\widetilde{\boldsymbol{\lambda}} = \mathbf{0}, \\ \frac{\partial \mathcal{L}}{\partial \boldsymbol{\lambda}}|_{\boldsymbol{\theta}=\widetilde{\boldsymbol{\theta}}} &= \boldsymbol{h}(\widetilde{\boldsymbol{\theta}}) = \mathbf{0}, \end{aligned}$$

where $\boldsymbol{H}(\widetilde{\boldsymbol{\theta}}) = \partial \boldsymbol{h}(\boldsymbol{\theta})'/\partial \boldsymbol{\theta}$, the LM test statistic can be constructed as

$$LM = \widetilde{\boldsymbol{\lambda}}' \boldsymbol{H}(\widetilde{\boldsymbol{\theta}})' \boldsymbol{\mathcal{I}}^{-1}(\widetilde{\boldsymbol{\theta}}) \boldsymbol{H}(\widetilde{\boldsymbol{\theta}}) \widetilde{\boldsymbol{\lambda}} = \boldsymbol{s}(\widetilde{\boldsymbol{\theta}})' \boldsymbol{\mathcal{I}}^{-1}(\widetilde{\boldsymbol{\theta}}) \boldsymbol{s}(\widetilde{\boldsymbol{\theta}}). \tag{1.1}$$

The score function $s(\theta)$ is exactly equal to zero when θ is estimated under unrestricted ML estimation, whereas $s(\widetilde{\theta})$ is not zero. If the constraints are true, we would expect both $s(\widetilde{\theta})$ and $\widetilde{\lambda}$ to be small, such that large values of (1.1) will lead to the rejection of the null hypothesis $\mathcal{H}_0 : h(\theta) = \mathbf{0}$. Under the null hypothesis, the sample distribution of the LM statistic usually converges with increased sample size to a chi-square distribution with $k - r$ degrees of freedom.

The parameter vector θ may be partitioned as θ_1 $(k_1 \times 1)$ and θ_2 $(k_2 \times 1)$ such that $\theta = (\theta_1', \theta_2')'$, where θ_1 is unrestricted and θ_2 is constrained. The information matrix and its inverse may then be partitioned into the blocks:

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

where $\mathcal{I}_{11} = -\mathbb{E}(\partial^2 l(\theta) / \partial\theta_1 \partial\theta_1')$, $\mathcal{I}_{12} = \mathcal{I}_{21}' = -\mathbb{E}(\partial^2 l(\theta) / \partial\theta_1 \partial\theta_2')$ and $\mathcal{I}_{22} = -\mathbb{E}(\partial^2 l(\theta) / \partial\theta_2 \partial\theta_2')$. Under the null hypothesis

$$\mathcal{H}_0 : \theta_2 = \theta_{20},$$

we have $s(\widetilde{\theta}) = (\mathbf{0}, (\partial l(\theta)/\partial\theta_2)')'$. Hence, the LM statistic becomes

$$LM = \left(\frac{\partial l(\theta)}{\partial \theta_2}\right)' \mathcal{I}^{22} \left(\frac{\partial l(\theta)}{\partial \theta_2}\right). \tag{1.2}$$

If $\mathcal{I}(\theta)$ is block diagonal, then $\mathcal{I}^{22} = \mathcal{I}_{22}^{-1}$. When the null hypothesis only concerns one parameter, by taking into account the appropriate sign, it is more meaningful to use $LM^*(\theta_2) = \left(\frac{\partial l(\theta)}{\partial \theta_2}\right)\sqrt{\mathcal{I}^{22}}$ (e.g., p.315, Cox and Hinkley, 1974).

In this thesis, some LM-type tests are derived to test for panel unit roots in factor models. To the best of our knowledge, the present study is one of the first studies to consider the likelihood-based tests for panel unit roots with cross-sectional dependence. Panel unit root tests and factor models are introduced in the following subsections.

1.2 Unit root tests

Testing for unit root is an important issue in time series econometrics. One reason is that when there exists a unit root, the limiting distributions of the estimators are very different from those in the stationary case. To see this, let us consider an autoregressive process of order one (AR(1)),

$$y_t = \rho y_{t-1} + \varepsilon_t,\ t = 1, \cdots, T, \tag{1.3}$$

where ε_t are identically and independently distributed (*i.i.d*) with mean 0 and variance σ_ε^2. When $|\rho| < 1$, the process is stationary; when $|\rho| > 1$, the process is explosive, which is usually not an interesting case; and when $\rho = 1$, there exists a unit root. The ordinary least squares (OLS) estimator of ρ is

$$\widehat{\rho} = \frac{\sum_{t=2}^{T} y_t y_{t-1}}{\sum_{t=2}^{T} y_{t-1}^2}.$$

Following the central limit theorem (CLT), $\sqrt{T}(\widehat{\rho} - \rho)$ converges to $N(0, 1 - \rho^2)$ when $|\rho| < 1$, whereas it degenerates to zero when $\rho = 1$. If we cannot distinguish between a stationary process and a nonstationary process, misleading conclusions might be drawn. Another reason is that a variable with a unit root behaves completely different from one without (e.g., the difference between a Brownian motion process and a stationary process, e.g., Hamilton, 1994). Additionally, as stated by Granger and Newbold (1974), any regression, unless it is a cointegrating relationship, between two nonstationary processes is spurious. Specifically, the regression of two independent nonstationary series will show significant correlation even though there is none.

For testing the null hypothesis $\mathcal{H}_0 : \rho = 1$ against the alternative $\mathcal{H}_1 : \rho < 1$, the most widely used method is the Dickey-Fuller (DF) test (e.g., Dickey and Fuller, 1979), which is based on OLS estimation of the autoregressive coefficient. By changing the convergence rate from $\sqrt{T}$ to T, the DF ρ test statistic $T(\widehat{\rho} - 1)$ converges to a non-standard distribution,

$$\frac{\frac{1}{2}\{[W(1)]^2 - 1\}}{\int_0^1 [W(r)]^2 dr}, \tag{1.4}$$

where $W(\cdot)$ is the Brownian motion process. The critical values of (1.4) can be tabulated by Monte Carlo simulations. $T(\widehat{\rho} - 1)$ diverges to $-\infty$ under the alternative hypothesis. The DF t-ratio test statistic is defined as

$$t = \frac{\widehat{\rho} - 1}{\widehat{\sigma}_{\widehat{\rho}}}, \tag{1.5}$$

where $\widehat{\sigma}_{\widehat{\rho}}$ is the OLS standard error for $\widehat{\rho}$.

Consider a general AR(p) process,

$$y_t = \mu + \rho_1 y_{t-1} + \rho_2 y_{t-2} + \cdots + \rho_p y_{t-p} + \varepsilon_t. \tag{1.6}$$

Equation (1.6) can be written as $A(L)y_t = \mu + \varepsilon_t$, where $A(L) = 1 - \rho_1 L - \rho_2 L^2 - \cdots - \rho_p L^p$, and L is the lag operator such that for any integer q, $L^q y_t = y_{t-q}$. The process is weakly stationary when both the mean and the autocovariances are independent of t, say $\mathbb{E}(y_t) = \mu^* < \infty$ for all t and $\mathbb{E}((y_t - \mu^*)(y_{t-s} - \mu^*)) = \gamma_s < \infty$ for all t and s. That y_t is stationary implies that all the roots of $A(z) = 0$ are outside of the unit circle, i.e., $|z_i| > 1$ for all i.

The DF test works only for AR(1) processes. For an AR(p) process, the augmented DF (ADF) test was modified by Said and Dickey (1984). Consider an even more general case, say, in Model (1.3), $\varepsilon_t = \sum_{j=0}^{\infty} \psi_j v_{t-j}$, where v_t are *i.i.d* error terms, ε_t is a moving average (MA) process with the coefficients square summable, i.e., $\sum_{j=0}^{\infty} \psi_j^2 < \infty$.[1] In fact, according to the theory of Wold (1938), any weakly stationary process can be represented as the sum of deterministic components and an MA process. For process (1.6), it is straightforward to verify that

$$y_t = A(L)^{-1}(\mu + \varepsilon_t) = \mu^{\dagger} + \psi(L)\varepsilon_t,$$

where $\psi(L) = A(L)^{-1} = \sum_{j=0}^{\infty} \psi_j L^j$ such that $A(L)^{-1}A(L) = 1$. Phillips and Perron (1988) considered a non-parametric method to test unit roots. Their procedure is also based on OLS, adjusted by the short-run covariance, long-run covariance and half long-run covariance, $\sigma_\varepsilon^2 = \mathbb{E}(\varepsilon_t^2) = \sum_{j=0}^{\infty} \psi_j^2$, $\omega_\varepsilon^2 = (\sum_{j=0}^{\infty} \psi_j)^2$ and $\zeta_\varepsilon^2 = (\omega_\varepsilon^2 - \sigma_\varepsilon^2)/2$, respectively.

Some other unit root tests include the LR test by Dickey and Fuller (1981), the LM test by Solo (1984), Schmidt and Phillips (1992) and the generalized least square (GLS) test of Elliott, Rothenberg, and Stock (1996).

The unit root tests for single time series suffer from low power in finite samples. To increase power, econometricians and statisticians often use the increased information in panels.

1.3 First generation panel unit root tests

Since the working paper versions of Quah (1994) and Breitung and Meyer (1994), research about unit root tests in panel data has continuously increased. A recent survey is Breitung and Pesaran (2008). Some early reviews include Banerjee (1999) and Baltagi and Kao (2001). The original motivation for deriving panel unit root tests was to increase power when testing the nonstationary null hypothesis against a stationary alternative by combining the information of the cross-sectional units. Today, the panel unit root analysis has become a research branch in its own right (e.g., Westerlund and Breitung, 2013).

Consider a simple AR(1) panel model,

$$y_{i,t} = \mu_i + \rho_i y_{i,t-1} + \varepsilon_{i,t}, \tag{1.7}$$

where $\varepsilon_{i,t}$ is the error term with mean 0 and variance σ_i^2, $i = 1, \cdots, N$ denote the cross-sectional individuals and $t = 1, \cdots, T$ denote the time

[1]There are other restrictions on the MA coefficients, such as $\sum_{j=0}^{\infty} j|\psi_j| < \infty$, or $\sum_{j=0}^{\infty} |\psi_j| < \infty$, that imply the square summability.

periods. Equation (1.7) can be written in vector form,

$$\boldsymbol{y}_t = \boldsymbol{\mu} + \boldsymbol{\rho}\boldsymbol{y}_{t-1} + \boldsymbol{\varepsilon}_t, \tag{1.8}$$

where $\boldsymbol{y}_t = (y_{1,t}, \cdots, y_{N,t})'$, $\boldsymbol{\mu} = (\mu_1, \cdots, \mu_N)'$, $\boldsymbol{\rho} = \text{diag}\,(\rho_1, \cdots, \rho_N)$ and $\boldsymbol{\varepsilon}_t = (\varepsilon_{1,t}, \cdots, \varepsilon_{N,t})'$. The null hypothesis is

$$\mathcal{H}_0 : \rho_i = 1 \text{ for all i.}$$

There are two types of alternative hypotheses: the *homogeneous* alternative and the *heterogeneous* alternative. A homogeneous alternative tests whether every individual time series has a unit root against the alternative that all time series are stationary with the same autoregressive coefficient, i.e.,

$$\mathcal{H}_{1a} : \rho_i = \rho < 1 \text{ for all i.}$$

The heterogeneous tests share the same null hypothesis as the homogeneous tests, but the autoregressive coefficients are different under the alternative hypotheses, i.e.,

$$\mathcal{H}_{1b} : \rho_i < 1 \text{ for } i = 1, \cdots, N_1 \text{ and all other } \rho_i = 1,$$

where $\frac{N_1}{N} \to \delta_1 > 0$ as N_1, $N \to \infty$.

Depending on whether the cross-sectional individuals are independent of each other or not, the panel unit root tests are roughly classified into two categories: *the first generation* unit root tests, which assume that the cross-sectional individuals are mutually independent (such that the covariance matrix of the error terms, $\text{Var}(\boldsymbol{\varepsilon}_t)$, is diagonal), and *the second generation* unit root tests, which assume that there exists cross-sectional dependence (such that the covariance matrix $\text{Var}(\boldsymbol{\varepsilon}_t)$ is a general positive definite symmetric matrix).

Levin, Lin, and Chu (2002) (hereafter LLC) and Im, Pesaran, and Shin (2003) (hereafter IPS) are two of the most cited papers concerning first generation panel unit root tests. LLC is a pooled OLS test for panel data,

$$\tau_{LLC} = \frac{\sum_{i=1}^{N} \widehat{\sigma}_i^{-2} \Delta \boldsymbol{y}_i' \boldsymbol{M}_\iota \boldsymbol{y}_{i,-1}}{\sqrt{\sum_{i=1}^{N} \widehat{\sigma}_i^{-2} \boldsymbol{y}_{i-1}' \boldsymbol{M}_\iota \boldsymbol{y}_{i,-1}}}. \tag{1.9}$$

In (1.9), $\Delta \boldsymbol{y}_i = (\Delta y_{i,1}, \cdots, \Delta y_{i,T})'$, $\boldsymbol{y}_{i,-1} = (y_{i,0}, \cdots, y_{i,T-1})'$, $\boldsymbol{M}_\iota = \boldsymbol{I}_T - \boldsymbol{\iota}_T(\boldsymbol{\iota}_T'\boldsymbol{\iota}_T)^{-1}\boldsymbol{\iota}_T'$ where $\boldsymbol{\iota}_T$ is a $T \times 1$ vector of ones and

$$\widehat{\sigma}_i^2 = \frac{\Delta \boldsymbol{y}_i' \boldsymbol{M}_i \Delta \boldsymbol{y}_i}{T-2},$$

where $\boldsymbol{M}_i = \boldsymbol{I}_T - \boldsymbol{X}_i(\boldsymbol{X}_i'\boldsymbol{X}_i)^{-1}\boldsymbol{X}_i'$ and $\boldsymbol{X}_i = (\boldsymbol{\iota}_T, \boldsymbol{y}_{i-1})$. IPS uses the average ADF t-ratio test to construct the test statistic

$$\tau_{IPS} = \frac{\sqrt{N}(\bar{\tau} - \mathbb{E}(\tau))}{\sqrt{\text{Var}(\tau)}}, \tag{1.10}$$

where $\bar{\tau} = \frac{1}{N}\sum_{i=1}^{N}\tau_i$ and τ_i are the ADF t-ratio test statistics for each individual.

Another popular approach to deal with heterogeneous alternative hypotheses is the Fisher-type test by Maddala and Wu (1999). This test combines the p-values (e.g., Fisher, 1934) of the usual ADF test for each individual, with N fixed,

$$\tau_{MW} = -2\sum_{i=1}^{N}\log p_i.$$

Choi (2001) extends the analysis to allow $N \to \infty$ by standardization,

$$\frac{-2\sum_{i=1}^{N}\log p_i - 2N}{\sqrt{4N}} \xrightarrow{d} \mathcal{N}(0,1) \text{ as } N \to \infty. \tag{1.11}$$

The first generation panel unit root tests successfully improved power. However, the restriction that individuals should be independent is crucially restrictive and lacks reality. For instance, data in macroeconomics, international trade or international finance are highly dependent (e.g., Pesaran, 2004; Lyhagen, 2008). The dependence could arise through temporal correlations (e.g., O'Connell, 1998) or through common stochastic trends, i.e., when the panel individuals are co-integrated (e.g., Urbain and Westerlund, 2011). As pointed out by O'Connell (1998) and Maddala and Wu (1999), the tests may exhibit severe size distortion when the independence assumption is violated. This potential problem encourages the development of the second generation panel unit root tests.

1.4 Second generation panel unit root tests

The second generation panel unit root tests allow for cross-sectional dependence. Basically, two main approaches have been proposed for this purpose. The first considers general cross-sectional dependence. Some examples include the nonlinear instrumental variables (NIV) test of Chang (2002), the robust GLS test of Breitung and Das (2005) and the bootstrap methods of Chang (2004) and Palm, Smeekes, and Urbain (2011).

Another popular approach models cross-sectional dependence by imposing unobserved common factors. There are two reasons for doing this. First, the cross-sectional space can be reduced to the factor space in which the dimension of the latter is usually much smaller than the dimension of the former. Second, it is natural to consider the dependence of a large set of economic variables arising from common factors: a large system may be driven by a small set of shocks (e.g., Breitung and Eickmeier, 2006). Phillips and Sul (2003) and Pesaran (2007) assume one factor, whereas Moon and Perron (2004), Bai and Ng (2004) and Pesaran, Smith, and Yamagata (2013) allow for a more general factor structure.

In Model (1.7), the unobservable terms can be expressed as

$$\varepsilon_{i,t} = \sum_{j=1}^{r} \lambda_{i,j} f_{j,t} + v_{i,t},$$

where $f_{j,t}$ is the unobservable common factor, $\lambda_{i,j}$ is the corresponding factor loading and $v_{i,t}$ is the idiosyncratic error term. It is obvious that $y_{i,t}$ are cross-sectionally dependent in that they share the same latent variables.

There are several ways to remove the effect of the common factors. Pesaran (2007) uses the mean of the observations as a proxy for the common factor, and then subtracts it from the model. Moon and Perron (2004) use the method of principal components (PC) to estimate the factor loadings and then multiply the observations by the orthogonal projection of the estimated factor loadings to remove the effect of the common factors. Bai and Ng (2004) first use the method of PC to estimate the space spanned by the common factors and the space spanned by the idiosyncratic error terms separately and then treat them as observable.

Since factor structure cross-sectional dependence is so popular in the second generation panel unit root tests, it is necessary to introduce more detailed information about the factor model separately.

1.5 Factor model

Factor analysis is a statistical method to describe the correlation among a large set of observed variables in terms of a handful of unobserved factors. Consider the factor model

$$\boldsymbol{x} = \boldsymbol{\Lambda f} + \boldsymbol{u}, \tag{1.12}$$

where $\boldsymbol{x} = (x_1, \cdots, x_N)'$ is a vector of N observable variables, $\boldsymbol{f} = (f_1, \cdots, f_r)$ is a vector of r unobservable common factors, $\boldsymbol{u} = (u_1, \cdots, u_N)'$ is a vector of N unobservable idiosyncratic components and $\boldsymbol{\Lambda} = (\boldsymbol{\lambda}_i, \cdots, \boldsymbol{\lambda}_N)'$ is a matrix of factor loadings in which $\boldsymbol{\lambda}_i = (\lambda_{i,1}, \cdots, \lambda_{i,r})'$.

The assumptions of the classical factor model are: (i) the number of variables N is fixed; (ii) the factors $\boldsymbol{f}$ and the idiosyncratic component $\boldsymbol{u}$ are *i.i.d* with $\mathbb{E}(\boldsymbol{f}) = \boldsymbol{0}$, $\mathbb{E}(\boldsymbol{ff}') = \boldsymbol{\Sigma}_f$ and $\mathbb{E}(\boldsymbol{u}) = \boldsymbol{0}$, $\mathbb{E}(\boldsymbol{uu}') = \boldsymbol{\Psi}$, where $\boldsymbol{\Psi} = \text{diag}(\psi_1, \cdots, \psi_N)$; and (iii) $\boldsymbol{f}$ and $\boldsymbol{u}$ are mutually independent with

$Cov(\boldsymbol{f},\boldsymbol{u}) = \mathbb{E}(\boldsymbol{f}\boldsymbol{u}') = \mathbf{0}$. The covariance matrix of $\boldsymbol{x}$ is

$$\boldsymbol{\Sigma}_{\boldsymbol{x}} = \mathrm{Var}(\boldsymbol{x}) = \boldsymbol{\Lambda}\boldsymbol{\Sigma}_{\boldsymbol{f}}\boldsymbol{\Lambda}' + \boldsymbol{\Psi}.$$

The parameters to be estimated are $\boldsymbol{\theta} = (\boldsymbol{\Lambda}, \boldsymbol{\Sigma}_{\boldsymbol{f}}, \boldsymbol{\Psi})$. Without additional restrictions, we cannot identify the factor model. To see this, let $\boldsymbol{\Lambda}^* = \boldsymbol{\Lambda}\boldsymbol{G}$ and $\boldsymbol{\Sigma}_{\boldsymbol{f}}^* = \boldsymbol{G}^{-1}\boldsymbol{f}\boldsymbol{G}'^{-1}$, where $\boldsymbol{G}$ is any nonsingular $r \times r$ matrix. It is obvious that $\boldsymbol{\Lambda}\boldsymbol{\Sigma}_{\boldsymbol{f}}\boldsymbol{\Lambda}' = \boldsymbol{\Lambda}^*\boldsymbol{\Sigma}_{\boldsymbol{f}}^*\boldsymbol{\Lambda}^{*\prime}$.

Since the non-singular matrix $\boldsymbol{G}$ has r^2 free parameters, we need r^2 restrictions to identify the model. There are many ways to impose restrictions (e.g., Jöreskog, 1967, 1969). The following are five commonly used identification conditions (IC):

- IC1: $\boldsymbol{\Lambda} = (\boldsymbol{I}_r, \boldsymbol{\Lambda}_2')'$, $\boldsymbol{\Sigma}_{\boldsymbol{f}}$ is unrestricted.
- IC2: $\frac{1}{N}\boldsymbol{\Lambda}'\boldsymbol{\Psi}^{-1}\boldsymbol{\Lambda} = \boldsymbol{I}_r$ and $\boldsymbol{\Sigma}_{\boldsymbol{f}} = \boldsymbol{\mathcal{D}}$, the diagonal matrix, whose diagonal elements are distinct and arranged in descending order.
- IC3: $\frac{1}{N}\boldsymbol{\Lambda}'\boldsymbol{\Psi}^{-1}\boldsymbol{\Lambda} = \boldsymbol{\mathcal{D}}$ and $\boldsymbol{\Sigma}_{\boldsymbol{f}} = \boldsymbol{I}_r$.
- IC4: $\boldsymbol{\Lambda}_1$ is a lower triangular matrix with all diagonal elements being 1 and $\boldsymbol{\Sigma}_{\boldsymbol{f}} = \boldsymbol{\mathcal{D}}$.
- IC5: $\boldsymbol{\Lambda}_1$ is a lower triangular matrix with none of its diagonal elements being 0 and $\boldsymbol{\Sigma}_{\boldsymbol{f}} = \boldsymbol{I}_r$, where $\boldsymbol{\Lambda}_1$ is the upper $r \times r$ submatrix of $\boldsymbol{\Lambda}$.

Under the normality assumption, the most commonly used estimation method is ML (e.g., Lawley and Maxwell, 1963; Anderson, 2003; Johnson and Wichern, 1988). Since the unobserved common factors can be considered as missing data, the expectation maximization (EM) algorithm (e.g., Rubin and Thayer, 1982) is popular. There are two main reasons for its popularity. First, the EM algorithm involves only complete-data ML estimation, which is often computationally simple. Second, the convergence is stable, with each iteration increasing the likelihood.

Let $l(\boldsymbol{x}, \boldsymbol{f}|\boldsymbol{\theta})$ be the complete log-likelihood, and let $p(\boldsymbol{f}|\boldsymbol{x}, \boldsymbol{\theta}^{(t)})$ be the conditional distribution of the unobservable factors $\boldsymbol{f}$ when given $\boldsymbol{x}$ and $\boldsymbol{\theta}^{(t)}$. The E-step is to compute the expectation of the complete log-likelihood based on the conditional distribution $p(\boldsymbol{f}|\boldsymbol{x}, \boldsymbol{\theta}^{(t)})$, which is a function of $\boldsymbol{\theta}$, denoted by $Q(\boldsymbol{\theta}|\boldsymbol{\theta}^{(t)})$,

E-step: Compute $Q(\boldsymbol{\theta}|\boldsymbol{\theta}^{(t)}) = \mathbb{E}_{p(\boldsymbol{f}|\boldsymbol{x},\boldsymbol{\theta}^{(t)})}[l(\boldsymbol{x}, \boldsymbol{f}|\boldsymbol{\theta})]$.

The M-step is to find the $\boldsymbol{\theta}$ which maximizes $Q(\boldsymbol{\theta}|\boldsymbol{\theta}^{(t)})$, denoted by $\boldsymbol{\theta}^{(t+1)}$,

M-step: $\boldsymbol{\theta}^{(t+1)} = \arg\max_{\boldsymbol{\theta}} \mathbb{E}_{p(\boldsymbol{f}|\boldsymbol{x},\boldsymbol{\theta}^{(t)})}[l(\boldsymbol{x},\boldsymbol{f}|\boldsymbol{\theta})]$.

The iteration starts from some initializing values $\boldsymbol{\theta}^{(1)}$, repeating until $||\boldsymbol{\theta}^{(t+1)} - \boldsymbol{\theta}^{(t)}||$ is less than some small tolerance.

1.6 Dynamic factor model and its static representation

A distinction needs to be made between a *static* factor model and a *dynamic* factor model. A *static* factor model is

$$x_{i,t} = \boldsymbol{\lambda}_i' \boldsymbol{f}_t + e_{i,t} = C_{i,t} + e_{i,t}, \tag{1.13}$$

where $\boldsymbol{f}_t$ is a vector of the common factors, $\boldsymbol{\lambda}_i$ is a vector of the factor loadings, $e_{i,t}$ is the idiosyncratic component. Moreover, $C_{i,t} = \boldsymbol{\lambda}_i' \boldsymbol{f}_t$ represent the common components, where i indicates the cross-sectional unit for $i = 1, \cdots, N$ and t indicates the time period for $t = 1, \cdots, T$. The model is defined as static because it specifies a static relationship between $x_{i,t}$ and $\boldsymbol{f}_t$, although $\boldsymbol{f}_t$ could be a dynamic process such that $\boldsymbol{f}_t = \boldsymbol{C}(L)\boldsymbol{v}_t$, where $C(L) = \sum_{i=0}^{\infty} \boldsymbol{C}_j L^j$ and $\boldsymbol{v}_t$ are *i.i.d* errors. In contrast, a *dynamic* factor model is usually defined as

$$x_{i,t} = \boldsymbol{\lambda}_i'(L)\boldsymbol{f}_t + e_{i,t},$$

where $\boldsymbol{\lambda}_i(L) = (\mathbf{1} - \boldsymbol{\lambda}_{i,1}L, \cdots, -\boldsymbol{\lambda}_{i,s}L^s)$ is a vector of dynamic factor loadings of order s and $e_{i,t} = D_i(L)\varepsilon_{i,t}$ (e.g., Bai and Ng, 2008; Forni, Hallin, Lippi, and Reichlin, 2000).

Every dynamic factor model with finite lags of the factor loadings has a static representation. As shown in Bai and Ng (2008), a dynamic factor model with q factors can be written as a static factor model with r factors, where $r = q \times (s+1)$. In this sense, it is sufficient to focus on the static factor model with dynamic factors and dynamic idiosyncratic error terms.

Rewrite (1.13) in vector form,

$$\boldsymbol{x}_t = \boldsymbol{\Lambda}\boldsymbol{f}_t + \boldsymbol{e}_t,$$

where $\boldsymbol{x}_t = (x_{1,t}, \cdots, x_{N,t})'$. The covariance matrix of $\boldsymbol{x}_t$ is

$$\boldsymbol{\Sigma}_x = \mathrm{Var}(\boldsymbol{x}) = \boldsymbol{\Lambda}\boldsymbol{\Sigma}_f\boldsymbol{\Lambda}' + \boldsymbol{\Psi},$$

where $\boldsymbol{\Sigma}_f = \mathbb{E}(\boldsymbol{f}_t - \mathbb{E}(\boldsymbol{f}_t))(\boldsymbol{f}_t - \mathbb{E}(\boldsymbol{f}_t))'$ and $\boldsymbol{\Psi} = \mathbb{E}(\boldsymbol{e}_t - \mathbb{E}(\boldsymbol{e}_t))(\boldsymbol{e}_t - \mathbb{E}(\boldsymbol{e}_t))'$. When $\boldsymbol{\Psi}$ is diagonal, the model in Equation (1.13) is referred to as the *exact* factor model, whereas an *approximate* factor model does not require $\boldsymbol{\Psi}$ to be diagonal (e.g., Chamberlain and Rothschild, 1983).

The eigenvalues of $\boldsymbol{\Psi}$ are usually assumed to be uniformly bounded, but the largest r eigenvalues of $\boldsymbol{\Sigma}_x$ increase with N, i.e., $\mathcal{O}(N)$, and the remaining eigenvalues of $\boldsymbol{\Sigma}_x$ are bounded. According to Chudik, Pesaran, and Tosetti (2011), the dependence among the idiosyncratic terms is *weak* and the dependence among the observations is *strong*.

Many economic and financial problems can be characterized according to a (static) factor model. For instance, in macroeconomics, suppose that $x_{i,t}$ is the GDP growth rate for country i at time t, $\boldsymbol{f}_t$ is a vector of unobserved common shocks with heterogeneous impact of the shocks given by $\boldsymbol{\lambda}_i$ and $e_{i,t}$ is the country-specific growth rate (e.g., Bai and Ng, 2008). The static factor model is different from the classic factor model in the following sense: (i) both the number of individuals N and observations T can be large and (ii) both the factors and the idiosyncratic error terms are usually serially, or cross-sectionally dependent, or both.

In this thesis, we focus on the exact static factor model with high dimensions and with dynamic factors and dynamic idiosyncratic components. Our study will contribute to the more general approximate dynamic factor models.

1.6.1 Estimation of factor model: PC

As mentioned in the previous section, the commonly used method for estimating the classic factor model is ML. In contrast, for factor models of high dimensions (and with dynamic factors and idiosyncratic errors), the commonly used approach is the method of PC (e.g., Connor and Korajzcyk, 1986, 1988; Stock and Waston, 2002; Bai and Ng, 2004).

Rewrite (1.13) in matrix form, $\boldsymbol{x} = \boldsymbol{\Lambda f} + \boldsymbol{e}$, where $\boldsymbol{x}$ is an $N \times T$ matrix, $\boldsymbol{f}$ is a $r \times T$ matrix and $\boldsymbol{e}$ is an $N \times T$ matrix. For a given number of factors r, the PC procedure minimizes the following objective function

$$V(\boldsymbol{\Lambda}, \boldsymbol{f}) = \frac{1}{NT}||\boldsymbol{x} - \boldsymbol{\Lambda f}||^2 = \frac{1}{NT}\sum_i\sum_t(x_{i,t} - \boldsymbol{\lambda}_i'\boldsymbol{f}_t)^2.$$

By concentrating out $\boldsymbol{\Lambda}$, under the condition that $\boldsymbol{f}\boldsymbol{f}'/T = \boldsymbol{I}_r$ (without loss of generality), the optimization problem is identical to maximizing $\mathrm{tr}(\boldsymbol{f}(\boldsymbol{x}'\boldsymbol{x})\boldsymbol{f}')$. Let $v_1, \cdots, v_r$ be the eigenvectors of $\boldsymbol{x}'\boldsymbol{x}$ corresponding to its r largest eigenvalues. Then, the estimated factors based on PC, $\left(\widehat{\boldsymbol{f}}^{pc}\right)'$, is $\sqrt{T}$ times $[v_1, \cdots, v_r]$ such that $\widehat{\boldsymbol{f}}^{pc}\left(\widehat{\boldsymbol{f}}^{pc}\right)'/T = \boldsymbol{I}_r$. Further, we have $\widehat{\boldsymbol{\Lambda}}^{pc} = \boldsymbol{x}\widehat{\boldsymbol{f}}^{pc}/T$. When r is unknown, the methods introduced by Bai and Ng (2002) can be used to estimate it consistently.

The advantages of the PC method are obvious. First, it is easy to compute, and second, it provides consistent estimators for the factors and factor loadings when both N and T are large (Bai, 2003). The estimation of the factor loadings, the space spanned by the common factors and the idiosyncratic error terms are consistent. As shown by Bai (2003), $\widehat{\boldsymbol{f}}_t^{pc} \xrightarrow{p} \boldsymbol{H}'\boldsymbol{f}_t$, as $N,T \to \infty$ if $\sqrt{N}/T \to 0$, and $\widehat{\boldsymbol{\lambda}}_i^{pc} \xrightarrow{p} \boldsymbol{H}^{-1}\boldsymbol{\lambda}_i$ as $N,T \to \infty$ if $\sqrt{T}/N \to 0$, where $\boldsymbol{H}$ is some invertible matrix. $\widehat{C}_{i,t} \xrightarrow{p} C_{i,t}$ as $\min(N,T) \to \infty$ without any further restrictions.

However, the disadvantage of PC is also obvious. As shown in Bai (2003), when N is fixed, the PC estimators of the factor loadings are not consistent when there exists heteroscedasticity, which will impact the unit root tests based on the estimated residuals. ML, on the other hand, is more attractive than PC. Not only is ML consistent when N is fixed but it is also more efficient.

1.6.2 Estimation of factor model: ML

Doz, Giannone, and Reichlin (2012) consider quasi-ML estimation in a misspecified exact factor model in which the true model is the approximate factor model. Bai and Li (2012b,a) consider a more general quasi-ML approach.

Let $\boldsymbol{\Psi}_t = \mathbb{E}(\boldsymbol{e}_t\boldsymbol{e}_t')$, which allows for heteroscedasticity over T, and let $\boldsymbol{\Psi} = \text{diag}\left(\frac{1}{T}\sum_{t=1}^T \boldsymbol{\Psi}_t\right)$. Consider the following objective function

$$l = -\frac{1}{2N}\ln|\boldsymbol{\Sigma}_{\boldsymbol{x}}| - \frac{1}{2N}\text{tr}\left(\boldsymbol{M}_{\boldsymbol{x}}\boldsymbol{\Sigma}_{\boldsymbol{x}}^{-1}\right), \tag{1.14}$$

where $\boldsymbol{M}_{\boldsymbol{x}} = \frac{1}{T}\sum_{t=1}^{T} \dot{\boldsymbol{x}}_t\dot{\boldsymbol{x}}_t'$ denotes the sample variance of $\boldsymbol{x}_t$, and where $\dot{\boldsymbol{x}}_t = \boldsymbol{x}_t - \frac{1}{T}\sum_{t=1}^T \boldsymbol{x}_t$. Also, let $\boldsymbol{\Sigma}_{\boldsymbol{x}} = \boldsymbol{\Lambda}\boldsymbol{\Sigma}_{\boldsymbol{f}}\boldsymbol{\Lambda}' + \boldsymbol{\Psi}$, where $\boldsymbol{\Sigma}_{\boldsymbol{f}}$ is the covariance matrix of $\boldsymbol{f}_t$.

As shown by Bai and Li (2012a), under very general forms of idiosyncratic autocorrelation, time heteroscedasticity and cross-sectional correlation, the quasi-ML estimators of the approximate model, i.e., $\boldsymbol{\Lambda}$, $\boldsymbol{\Sigma}_{\boldsymbol{f}}$ and $\boldsymbol{\Psi}$, are consistent when they are identified by any of the five identification restrictions as stated in Section 1.5, with average convergence rates,

$$\begin{aligned}
\frac{1}{N}\sum_{i=1}^N \frac{1}{\psi_i^2}\left|\left|\widehat{\boldsymbol{\lambda}}_i - \boldsymbol{\lambda}_i\right|\right|^2 &= \mathcal{O}_p\left(T^{-1}\right) + \mathcal{O}_p\left(N^{-2}\right), \\
\frac{1}{N}\sum_{i=1}^N \left(\widehat{\psi}_i^2 - \psi_i^2\right)^2 &= \mathcal{O}_p\left(T^{-1}\right) + \mathcal{O}_p\left(N^{-2}\right), \\
\left|\left|\widehat{\boldsymbol{\Sigma}}_{\boldsymbol{f}} - \boldsymbol{\Sigma}_{\boldsymbol{f}}\right|\right|^2 &= \mathcal{O}_p\left(T^{-1}\right) + \mathcal{O}_p\left(N^{-2}\right).
\end{aligned}$$

When the model is an exact factor model, the $\mathcal{O}_p\left(N^{-2}\right)$ terms vanish. The computation can be done using the EM algorithm as introduced in Section 1.5.

1.6.3 Nonstationary factor model

The nonstationary factor model can now be considered,

$$\begin{aligned} X_{i,t} &= \mu_i + \boldsymbol{\lambda}_i' \boldsymbol{F}_t + e_{i,t}, \\ (1-L)\boldsymbol{F}_t &= C(L)\boldsymbol{v}_t, \\ (1-\rho_i L)e_{i,t} &= D_i(L)\varepsilon_{i,t}. \end{aligned} \tag{1.15}$$

In this model, the non-stationarity of $X_{i,t}$ can arise because $\boldsymbol{F}_t$ is I(1), or because $e_{i,t}$ is I(1), or both. The ability to estimate $\boldsymbol{F}_t$ and $e_{i,t}$ makes it possible to test unit roots in $\boldsymbol{F}_t$ and $e_{i,t}$ separately (e.g., Bai and Ng, 2004).

When the model is nonstationary, estimation based on Bai (2003) is not consistent. However, as shown in Bai and Ng (2004), if we perform the first difference operation, i.e., $\Delta X_{i,t} = \boldsymbol{\lambda}_i' \Delta \boldsymbol{F}_t + \Delta e_{i,t}$, the model becomes stationary. Denote $x_{i,t} = \Delta X_{i,t}$, $\boldsymbol{f}_t = \Delta \boldsymbol{F}_t$, and $z_{i,t} = \Delta e_{i,t}$. Let $(\widehat{\boldsymbol{\lambda}}_1, \cdots, \widehat{\boldsymbol{\lambda}}_N)$, $(\widehat{\boldsymbol{f}}_1, \cdots, \widehat{\boldsymbol{f}}_T)$ and $\widehat{z}_{i,t}$ for all i and t be the PC estimates of $\boldsymbol{\lambda}_i$, $\boldsymbol{f}_t$ and $z_{i,t}$. In other words, $\widehat{\boldsymbol{\lambda}}_i$, $\widehat{\boldsymbol{f}}_t$ and $\widehat{z}_{i,t}$ are consistent for $\boldsymbol{H}^{-1}\boldsymbol{\lambda}_i$, $\boldsymbol{H}'\boldsymbol{f}_t$ and $z_{i,t}$ for some invertible $r \times r$ matrix $\boldsymbol{H}$. Note that, as $N, T \to \infty$, $\widehat{\boldsymbol{F}}_t = \sum_{s=2}^{T} \widehat{\boldsymbol{f}}_t$ is uniformly close to $\boldsymbol{H}\boldsymbol{F}_t$. The results are not dependent on whether $\boldsymbol{F}_t$ or $e_{i,t}$ are I(1) or I(0). Similarly, we can apply ML to the differenced model and then re-accumulate the estimated terms.

After estimating the common factors and the idiosyncratic components, we can test for unit roots of the common factors and the idiosyncratic components separately as they are observed.

2. Summary of papers

2.1 Framework and assumptions

Consider the following factor model

$$X_{i,t} = \mu_i + \boldsymbol{\lambda}_i' \boldsymbol{F}_t + e_{i,t},\ i = 1, \cdots, N; t = 1, \cdots, T, \tag{2.1}$$

where μ_i is the individual specific fixed effect, $\boldsymbol{F}_t = (F_{1,t}, \cdots, F_{r,t})'$ is an $r \times 1$ vector of dynamic factors, $\boldsymbol{\lambda}_i = (\lambda_{i,1}, \cdots, \lambda_{i,r})'$ is a $r \times 1$ vector of factor loadings and $e_{i,t}$ is the idiosyncratic error. In this model, only $X_{i,t}$ is observed. Both $\boldsymbol{F}_t$ and $e_{i,t}$ are dynamic and unobserved. Equation (2.1) can be rewritten in vector form,

$$\boldsymbol{X}_t = (X_{1,t}, \cdots, X_{N,t})' = \boldsymbol{\mu} + \boldsymbol{\Lambda} \boldsymbol{F}_t + \boldsymbol{e}_t, \tag{2.2}$$

where $\boldsymbol{\mu} = (\mu_1, \cdots, \mu_N)'$, $\boldsymbol{e}_t = (e_{1,t}, \cdots, e_{N,t})'$ and $\boldsymbol{\Lambda} = (\boldsymbol{\lambda}_1, \cdots, \boldsymbol{\lambda}_N)'$. The general assumptions made are as follows:

Assumption 2.1 The idiosyncratic components are homogenous AR(1), $e_{i,t} = \rho_i u_{i,t-1} + \varepsilon_{i,t}$, where $\rho_i \in (-1, 1]$, $\varepsilon_{i,t} \sim \mathcal{N}\left(0, \sigma^2_{\varepsilon,i}\right)$ are *i.i.d* with $\sigma^2_{\varepsilon,i} < \infty$ and where $\mathbb{E}(\varepsilon_{i,t}\varepsilon_{l,s}) = 0$ for all $i \neq l$ $(i, l = 1, 2, \ldots, N)$ and all $t, s = 1, 2, \ldots, T$.
Assumption 2.2 The factors are Gaussian AR(1), $F_{j,t} = \alpha_j F_{j,t-1} + v_{j,t}$, where $\alpha_j \in (-1, 1]$, $v_{j,t} \sim \mathcal{N}(0, 1)$ are i.i.d, and $\mathbb{E}(v_{j,t}v_{q,s}) = 0$ $(j, q = 1, \cdots, r)$ for all $j \neq q$ and all $t, s = 1, \cdots, T$.
Assumption 2.3 The factor loadings are non-random with $||\boldsymbol{\lambda}_i|| < \infty$, where $\frac{1}{N}\sum_{i=1}^{N} \boldsymbol{\lambda}_i \boldsymbol{\lambda}_i'$ converges to a positive definite matrix $\boldsymbol{\Sigma}_\Lambda$.
Assumption 2.4 The processes $v_{j,t}$ and $\varepsilon_{i,t}$ are mutually independently distributed such that $\mathbb{E}(v_{j,t}, \varepsilon_{i,t}) = 0$ for all $j = 1, \cdots, r$; $i = 1, \cdots, N$ and $t = 1, \cdots, T$.
Assumption 2.5 For stationary processes $f_{j,t}$ and $e_{i,t}$, the initial values $f_{j,0}$ and $e_{i,0}$ come from the stationary distributions, while for nonstationary processes, they are $\mathcal{O}_p(1)$.
Assumption 2.6 The idiosyncratic components and the factors are separate homogenous groups, i.e., $\rho_i = \rho$ for all i, and $\alpha_j = \alpha$ for all j.

For notational convenience, let $T^* = T - 1$, $\boldsymbol{D}$ be the $T^* \times T$ first-difference matrix,

$$\boldsymbol{D} = \begin{pmatrix} -1 & 1 & 0 & \cdots & 0 \\ 0 & -1 & 1 & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & -1 & 1 \end{pmatrix}, \tag{2.3}$$

and let $\boldsymbol{Y} = \boldsymbol{X}\boldsymbol{D}'$, where $\boldsymbol{X} = (\boldsymbol{X}_1, \cdots, \boldsymbol{X}_T)$ is the observed panel data. Under Assumption 2.6, $\alpha_j = \alpha$ and $\rho_i = \rho$ for all $j = 1, \cdots, r$; $i = 1, \cdots, N$, the differenced and stacked panel data, $\boldsymbol{Y}_v = \text{vec}(\boldsymbol{Y}) = (\boldsymbol{y}_2', \boldsymbol{y}_3', \cdots, \boldsymbol{y}_T')'$, has covariance matrix

$$\boldsymbol{\Sigma} = \mathbb{E}(\boldsymbol{Y}_v \boldsymbol{Y}_v') = [\boldsymbol{\Psi}(\alpha) \otimes \boldsymbol{\Lambda}\boldsymbol{\Lambda}'] + [\boldsymbol{\Psi}(\rho) \otimes \boldsymbol{\Sigma}_{\boldsymbol{\varepsilon}}], \tag{2.4}$$

where $\otimes$ denotes the Kronecker product of two matrices, $\boldsymbol{\Sigma}_{\boldsymbol{\varepsilon}} = \text{diag}(\sigma^2_{\varepsilon,1}, \cdots, \sigma^2_{\varepsilon,N})$ and where

$$\boldsymbol{\Psi}_{k,s}(\rho) = \begin{cases} \frac{2}{1+\rho} \\ -\frac{\rho^{||k-s|-1|}(1-\rho)}{1+\rho} \end{cases}.$$

When $\alpha = 1$ and $\rho = 1$, $\boldsymbol{\Psi}(\alpha) = \boldsymbol{\Psi}(\rho) = \boldsymbol{I}_{T^*}$, (2.4) becomes

$$\boldsymbol{\Sigma}|_{\alpha=1,\rho=1} = \mathbf{I}_{T^*} \otimes \boldsymbol{\Lambda}\boldsymbol{\Lambda}' + \mathbf{I}_{T^*} \otimes \boldsymbol{\Sigma}_{\boldsymbol{\varepsilon}} = \boldsymbol{I}_{T^*} \otimes (\boldsymbol{\Lambda}\boldsymbol{\Lambda}' + \boldsymbol{\Sigma}_{\boldsymbol{\varepsilon}}) = \boldsymbol{I}_{T^*} \otimes \boldsymbol{\Omega},$$

where $\boldsymbol{\Omega} \equiv \boldsymbol{\Lambda}\boldsymbol{\Lambda}' + \boldsymbol{\Sigma}_{\boldsymbol{\varepsilon}}$.

Let $\boldsymbol{\theta} = (\text{vec}(\boldsymbol{\Lambda})', \sigma^2_{\varepsilon,1}, \cdots, \sigma^2_{\varepsilon,N}, \alpha, \rho)'$; under the multivariate normality assumption, the log-likelihood of $\boldsymbol{Y}_v$ is

$$l(\boldsymbol{\theta}) = -\frac{NT^*}{2}\log(2\pi) - \frac{1}{2}\log|\boldsymbol{\Sigma}| - \frac{1}{2}\boldsymbol{Y}_v'\boldsymbol{\Sigma}\boldsymbol{Y}_v. \tag{2.5}$$

Denote $\widetilde{\boldsymbol{\theta}} = (\text{vec}(\widetilde{\boldsymbol{\Lambda}})', \widetilde{\sigma}^2_{\varepsilon,1}, \cdots, \widetilde{\sigma}^2_{\varepsilon,N}, 1, 1)'$ the ML estimators that maximize (2.5). For notational convenience, let

$$\boldsymbol{S}_{01} = \widetilde{\boldsymbol{\Lambda}}\widetilde{\boldsymbol{\Lambda}}' + \widetilde{\boldsymbol{\Sigma}}_{\boldsymbol{\varepsilon}}. \tag{2.6}$$

More strictly, when $\sigma^2_{\varepsilon,i} = \sigma^2_{\varepsilon}$ for all $i = 1, \cdots, N$, we have $\boldsymbol{\Sigma}_{\boldsymbol{\varepsilon}} = \sigma^2_{\varepsilon}\boldsymbol{I}_N$.

Assumption 2.7 The idiosyncratic components are cross-sectionally homoscedastic, i.e., $\sigma^2_{\varepsilon,i} = \sigma^2_{\varepsilon}$ for all i.

In Paper I and II, we follow Assumptions 2.1-2.7. In Paper III, we relax the assumption that the factors are integrated, while allowing the factors to have quite general form of dynamics. Thus, Assumption 2.2 is replaced by Assumption 2.2* in Subsection 2.4. In paper IV, additional to the relaxation of the factors, the restrictions concerning the idiosyncratic components are also relaxed (see Assumption 2.1* in Subsection 2.5).

2.2 Paper I

In paper I, we derive a homogeneous LM-type test for panel unit roots in the idiosyncratic components under the conditions that all factors are nonstationary, i.e., $\boldsymbol{f}_t \sim I(1)$ and that the number of factors are known. We give the asymptotic distribution of the test statistic in two cases: (i) when the number of panel individuals (N) is fixed and the number of time points (T) tends to infinity, and (ii) when T tends to infinity followed by N trending to infinity. We compare the size-adjusted power of the LM-type test with the pooled test of Bai and Ng (2004), and study the robustness to cross-sectional heteroscedasticity and the heterogeneity in the autoregressive parameters.

Under Assumptions 2.1-2.7, we consider the hypotheses

$$\begin{aligned}\mathcal{H}_0 &: \rho = 1 | F_{j,t} \sim I(1), \text{ for all } j = 1, \dots, r, \\ \mathcal{H}_1 &: \rho < 1 | F_{j,t} \sim I(1), \text{ for all } j = 1, \dots, r.\end{aligned}$$

Let $\boldsymbol{\theta} = \left(\text{vec}(\boldsymbol{\Lambda})', \sigma_\varepsilon^2, \rho\right)'$ be the parameter set that can be partitioned as $\boldsymbol{\theta}_1 = (\text{vec}(\boldsymbol{\Lambda})', \sigma_\varepsilon^2)'$, and $\theta_2 = \rho$. Based on (2.5), we derive the LM-type statistic

$$\vartheta_1 = \left[\frac{\partial l}{\partial \rho}\sqrt{\mathcal{I}^{\rho\rho}}\right]\Big|_{\rho=1}, \tag{2.7}$$

where $\frac{\partial l}{\partial \rho}$ is the first derivative of log likelihood corresponding to ρ and $\mathcal{I}^{\rho\rho}$ is the corresponding lower right scalar of the inverse of the information matrix. For any given T, (2.7) is

$$\vartheta_1 = \frac{T^* tr\left(\boldsymbol{S}_{01}^{-1}\right) - 2tr\left(\boldsymbol{S}_{01}^{-1}\boldsymbol{S}_0\boldsymbol{S}_{01}^{-1}\right) + tr\left(\boldsymbol{S}_{01}^{-1}\boldsymbol{S}_{00}\boldsymbol{S}_{01}^{-1}\right)}{\sqrt{2T^*(T^*-1)tr\left(\boldsymbol{S}_{01}^{-2}\right)}}, \tag{2.8}$$

where $\boldsymbol{S}_{01}$ is as shown in (2.6),

$$\boldsymbol{S}_0 = \sum_{t=2}^{T} \boldsymbol{y}_t \boldsymbol{y}_t' = \boldsymbol{Y}\boldsymbol{Y}', \tag{2.9}$$

and

$$\boldsymbol{S}_{00} = \left(\sum_{t=2}^{T} \boldsymbol{y}_t\right)\left(\sum_{t=2}^{T} \boldsymbol{y}_t\right)'. \tag{2.10}$$

Let η_i be the eigenvalues of $\mathbf{\Lambda}'\mathbf{\Lambda}$ with $\eta_1 \geq \eta_2 \geq \cdots \geq \eta_r \geq 0$ for $1 \leq r < N$. We show that when N is fixed and $T \to \infty$, the limiting distribution of (2.8) is a weighted sum of independent χ_1^2-variables,

$$\vartheta_1 \xrightarrow{d} \frac{1}{\sqrt{2\sum_{i=1}^N w_i^2}} \left(\sum_{i=1}^N w_i \chi_{1,i}^2 - \sum_{i=1}^N w_i \right), \tag{2.11}$$

where $\chi_{1,i}^2$ are independent over $i = 1, \ldots, N$, and where $w_i = \frac{\sigma_\varepsilon^2}{\eta_i + \sigma_\varepsilon^2}$ for $1 \leq i \leq r$ and $w_i = 1$ for $r < i \leq N$. Matching the first- and second-order moments, (2.11) can be approximated as

$$\vartheta_1 \overset{app.}{\sim} \frac{1}{\sqrt{2d_1}} \left(\chi_{d_1}^2 - d_1 \right), \tag{2.12}$$

where $d_1 = \frac{\left(\sum_{i=1}^N w_i\right)^2}{\sum_{i=1}^N w_i^2} = \frac{\left[\text{tr}(\mathbf{\Omega}^{-1})\right]^2}{\text{tr}(\mathbf{\Omega}^{-1}\mathbf{\Omega}^{-1})}$, and where $\mathbf{\Omega} = \mathbf{\Lambda}\mathbf{\Lambda}' + \sigma_\varepsilon^2 \boldsymbol{I}_N$.

Furthermore, the number of degrees of freedom of (2.12) can be appropriately approximated by $N - \frac{1}{2}r$, where r is the number of nonstationary factors. When T tends to infinity followed by N tending to infinity, the distribution tends to the standard normal, irrespective of the number of factors, i.e., $\vartheta_1 \xrightarrow{d} \mathcal{N}(0,1)$.

Our Monte Carlo simulations show that ϑ_1 has good size properties if N is not much larger than T, and as long as the number of factors r is not large relative to N. ϑ_1 is also demonstrated to have substantially higher local power than the Fisher-type test in Bai and Ng (2004). Even though the test statistic is derived under integrated factors, the simulations show that as long as the dimensions of the panel are large enough, the test statistic is robust when the factors are stationary or cointegrated. Although the case in this paper is rather special, by using a quasi-likelihood method, the LM statistic can be extended to allow for more general cases (e.g., higher order dynamics and cross sectional heteroscedasticity).

2.3 Paper II

In contrast to Paper I, in which we derive an LM-type test for the idiosyncratic unit roots conditional on $I(1)$ factors, in paper II we consider a homogeneous test for unit roots in the common factors conditional on $I(1)$ idiosyncratic components.

Under Assumptions 2.1-2.7, the hypotheses are

$$\mathcal{H}_0 : \alpha = 1 | u_{i,t} \sim I(1), \text{ for all } i = 1, \cdots, N;$$
$$\mathcal{H}_1 : \alpha < 1 | u_{i,t} \sim I(1), \text{ for all } i = 1, \cdots, N.$$

Similar to Paper 1, the test statistic is

$$\vartheta_2 = \left[\frac{\partial l}{\partial \alpha}\sqrt{\mathcal{I}^{\alpha\alpha}}\right]\Big|_{\alpha=1} \simeq \vartheta_2^* = \left[\frac{\partial l}{\partial \alpha}\sqrt{\mathcal{I}_{\alpha\alpha}^{-1}}\right]\Big|_{\alpha=1}. \tag{2.13}$$

In detail,

$$\vartheta_2^* = \frac{T^* \operatorname{tr}\left(\widetilde{\boldsymbol{M}}\boldsymbol{S}_{01}^{-1}\right) - 2\operatorname{tr}\left(\widetilde{\boldsymbol{M}}\boldsymbol{S}_{01}^{-1}\boldsymbol{S}_0\boldsymbol{S}_{01}^{-1}\right) + \operatorname{tr}\left(\widetilde{\boldsymbol{M}}\boldsymbol{S}_{01}^{-1}\boldsymbol{S}_{00}\boldsymbol{S}_{01}^{-1}\right)}{\sqrt{2T^{*2}\operatorname{tr}\left(\widetilde{\boldsymbol{M}}\boldsymbol{S}_{01}^{-1}\widetilde{\boldsymbol{M}}\boldsymbol{S}_{01}^{-1}\right)}}, \tag{2.14}$$

where $\widetilde{\boldsymbol{M}} = \widetilde{\boldsymbol{\Lambda}}\widetilde{\boldsymbol{\Lambda}}'$ and $\boldsymbol{S}_0, \boldsymbol{S}_{00}, \boldsymbol{S}_{01}$ are given by (2.9),(2.10) and (2.6) respectively.

When N is fixed and $T \to \infty$, the test statistic converges to the weighted sum of χ_1^2 distributions,

$$\vartheta_2^* \xrightarrow{d} \frac{1}{\sqrt{2\sum_{j=1}^{r}\overline{w}_j^2}}\left(\sum_{j=1}^{r}\overline{w}_j\chi_{1,j}^2 - \sum_{j=1}^{r}\overline{w}_j\right), \tag{2.15}$$

where $\chi_{1,j}^2$ are independent over $j = 1, 2, \ldots, r$ and where $\overline{w}_j = \frac{\eta_j}{\sigma_\varepsilon^2 + \eta_j}$. Matching first- and second-order moments of (2.16) yields

$$\vartheta_2^* \overset{app.}{\sim} \frac{1}{\sqrt{2d_2}}\left(\chi_{d_2}^2 - d_2\right),$$

where $d_2 = \frac{\left[\operatorname{tr}\left(\boldsymbol{M}\boldsymbol{\Omega}^{-1}\right)\right]^2}{\operatorname{tr}\left(\boldsymbol{M}\boldsymbol{\Omega}^{-1}\boldsymbol{M}\boldsymbol{\Omega}^{-1}\right)}$. As $T \to \infty$ followed by $N \to \infty$,

$$\vartheta_2^* \xrightarrow{d} \frac{1}{\sqrt{2r}}\left(\chi_r^2 - r\right). \tag{2.16}$$

In addition to the above test, we propose another two LM-type tests that jointly test for unit roots in the factors and the idiosyncratic components and derive their asymptotic distributions. Specifically, for testing $\mathcal{H}_0 : \alpha = 1$ and $\rho = 1$ against $\mathcal{H}_1 : \alpha < 1$ and/or $\rho < 1$, the test statistics are a combination of the test statistic ϑ_1 in Paper I and ϑ_2^* in Paper II.

By simulation, the size and power properties of the proposed statistics are studied. Their performance is also compared with the PANIC procedure of Bai and Ng (2004). The simulation results show the following: (i) the likelihood-based statistics have higher local power than the PANIC procedure based on PC; (ii) PANIC has virtually no power against locally stationary factors; and (iii) our test statistics are robust to the conditioning on $I(1)$ factors and $I(1)$ idiosyncratic components if the panel dimensions are large enough. The results in this paper are derived under the important AR(1) case, which should contribute to the development of likelihood-based unit root tests in dynamic factor models with more general dynamic properties than considered here.

2.4 Paper III

In paper III, we derive a homogeneous LM-type test for idiosyncratic unit roots in the exact factor model based on the likelihood function of the differenced data. The setup in this paper is similar to that in Paper I, but the restrictions are relaxed with respect to two features. First, the idiosyncratic components are assumed to be cross-sectional heteroscedasticity, i.e., $\sigma^2_{\varepsilon,i}$ are distinct for different i. Second, the factors are allowed to be general linear processes. Specifically, we use the Assumption 2.2* in place of the original assumption (i.e., Assumption 2.2).

Assumption 2.2* (factors). The factors admit the representation $(1-L)\boldsymbol{F}_t = \boldsymbol{C}(L)\boldsymbol{v}_t$, where $\boldsymbol{C}(L) = \sum_{m=0}^{\infty} \boldsymbol{C}_m L^m$ is an $r \times r$ lag polynomial matrix and where

(i) $\boldsymbol{v}_t \sim i.i.d\ (\boldsymbol{0}, \boldsymbol{\Sigma}_v)$, where $E||\boldsymbol{v}_t||^4 < \infty$;

(ii) $\boldsymbol{\Sigma}_{\Delta \boldsymbol{F}} = \sum_{m=0}^{\infty} \boldsymbol{C}_m \boldsymbol{\Sigma}_v \boldsymbol{C}'_m > 0$;

(iii) $\sum_{m=0}^{\infty} m||\boldsymbol{C}_m|| < \infty$; and

(iv) $\text{rank}(\boldsymbol{C}(1)) = r_1$, where $0 \leq r_1 \leq r$.

Assumptions 2.2* concerning the factors is the same as Bai and Ng (2004). $r \geq 1$ is the number of factors and $0 \leq r_1 \leq r$ is the number of the common trends among the factors (e.g., Johansen, 1995). To construct the test statistic, we impose the following misspecification:

Misspecification 2.1 The differenced factors are multivariate white noise drawn from the normal distribution, $\Delta \boldsymbol{F}_t \sim i.i.d\ \mathcal{N}_r(\boldsymbol{0}, \boldsymbol{I})$.

Under Assumptions 2.1, 2.3-2.5, Misspecification 2.1 and $\rho_i = \rho$ for all i, the parameter vector is $\boldsymbol{\theta} = (\text{vec}(\boldsymbol{\Lambda})', \sigma^2_{\varepsilon,1}, \cdots, \sigma^2_{\varepsilon,N}, \rho)'$, which can be partitioned as $\boldsymbol{\theta}_1 = (\text{vec}(\boldsymbol{\Lambda})', \sigma^2_{\varepsilon,1}, \cdots, \sigma^2_{\varepsilon,N})'$ and $\theta_2 = \rho$. Consider the null hypothesis and the alternative hypothesis

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

Because the information matrix is block diagonal when $T \to \infty$ no matter if N is fixed or $N \to \infty$, the LM-type test statistic is

$$\vartheta_3 = \left[\frac{\partial l}{\partial \rho} \sqrt{\mathcal{I}^{\rho\rho}} \right] \bigg|_{\rho=1} \simeq \vartheta_3^* = \left[\frac{\partial l}{\partial \alpha} \sqrt{\mathcal{I}_{\rho\rho}^{-1}} \right] \bigg|_{\rho=1}. \tag{2.17}$$

In detail,

$$\vartheta_3^* = \frac{T^* \operatorname{tr}\left(\widetilde{\boldsymbol{\Sigma}}_\varepsilon \boldsymbol{S}_{01}^{-1}\right) - 2\operatorname{tr}\left(\widetilde{\boldsymbol{\Sigma}}_\varepsilon \boldsymbol{S}_{01}^{-1}\boldsymbol{S}_0\boldsymbol{S}_{01}^{-1}\right) + \operatorname{tr}\left(\widetilde{\boldsymbol{\Sigma}}_\varepsilon \boldsymbol{S}_{01}^{-1}\boldsymbol{S}_{00}\boldsymbol{S}_{01}^{-1}\right)}{\sqrt{2T^{*2}\operatorname{tr}\left(\widetilde{\boldsymbol{\Sigma}}_\varepsilon \boldsymbol{S}_{01}^{-1}\widetilde{\boldsymbol{\Sigma}}_\varepsilon \boldsymbol{S}_{01}^{-1}\right)}}, \quad (2.18)$$

where $\boldsymbol{S}_0, \boldsymbol{S}_{00}$ are given by (2.9) and (2.10) respectively, $\boldsymbol{S}_{01} = \widetilde{\boldsymbol{\Lambda}}\widetilde{\boldsymbol{\Lambda}}' + \widetilde{\boldsymbol{\Sigma}}_\varepsilon$, where $\widetilde{\boldsymbol{\Sigma}}_\varepsilon = \operatorname{diag}(\widetilde{\sigma}_{\varepsilon,1}^2, \cdots, \widetilde{\sigma}_{\varepsilon,N}^2)$.

For fixed N, let η_i be the eigenvalues of $\boldsymbol{\Lambda}'\boldsymbol{\Sigma}_\varepsilon^{-1}\boldsymbol{\Lambda}$ with $\eta_1 \geq \eta_2 \geq \cdots \geq \eta_r \geq 0$ for $1 \leq r < N$, where r is the number of factors. As $T \to \infty$, the LM-type test (2.18) has the following asymptotic distribution

$$\vartheta_3^* \xrightarrow{d} \frac{1}{\sqrt{2\sum_{i=1}^N w_i^2}} \left(\sum_{i=1}^N w_i \chi_{1,i}^2 - \sum_{i=1}^N w_i\right), \quad (2.19)$$

where $\chi_{1,i}^2$ are independent over $i = 1, \ldots, N$ and where $w_i = (1+\eta_i)^{-1}$ for $1 \leq i \leq r$ and $w_i = 1$ for $r < i \leq N$. Similar to Paper I and II, ϑ_3 can be approximated by

$$\vartheta_3^* \overset{app.}{\sim} \frac{1}{\sqrt{2d_3}} \left(\chi_{d_3}^2 - d_3\right), \quad (2.20)$$

where $d_3 = \frac{\left[\operatorname{tr}(\boldsymbol{\Sigma}_\varepsilon \boldsymbol{\Omega}^{-1})\right]^2}{\operatorname{tr}(\boldsymbol{\Sigma}_\varepsilon \boldsymbol{\Omega}^{-1}\boldsymbol{\Sigma}_\varepsilon \boldsymbol{\Omega}^{-1})}$.

Under Assumptions 2.1, 2.2*, 2.3-2.5 and $\rho_i = \rho$ for all i, as $T \to \infty$ followed by $N \to \infty$, $\vartheta_3^* \xrightarrow{d} \mathcal{N}(0,1)$. We show that this result is asymptotically independent of the distribution of the factors. The factors may be independent $I(1)$ processes, co-integrated process, or be $I(0)$ processes as long as they satisfy Assumption 2.2*.

By simulation, we compare the size and the local power properties of the LM-type test with some commonly used tests. Those tests include 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 simulation results reveal the following findings: (i) the LM-type test and the Fisher-type test are robust in terms of size, whereas the tests of Moon and Perron (2004) and Bai and Ng (2010) are over-sized; and (ii) the LM-type test has the highest size-adjusted local power as N and T grow.

2.5 Paper IV

To test for idiosyncratic unit roots in Model (2.1), Bai and Ng (2004) use the method of PC to estimate the spaces spanned by the common factors and the idiosyncratic components consistently. The idiosyncratic components can then be treated as if they were observable. In this paper, we replace the PC method by the ML method to get the estimated idiosyncratic components. The size and power properties of some commonly used tests, including Bai and Ng (2004, 2010) and Wang, Wang, Yang, and Li (2010), are compared by simulation based on the residuals estimated by PC and ML.

In this study, we allow the common factors as well as the idiosyncratic components to be general linear processes. That is, in addition to replace Assumption 2.2 with Assumption 2.2*, we also replace Assumption 2.1 with Assumption 2.1*.

Assumption 2.1* The idiosyncratic components are $(1-\rho_i L)e_{i,t} = D_i(L)\varepsilon_{i,t}$, where $\rho_i \in (-1,1]$, $\varepsilon_{i,t} \sim i.i.d(0,1)$ across i and over t, $\mathbb{E}|\varepsilon_{i,t}|^8 < \infty$. $D_i(L) = \sum_{j=0}^{\infty} d_{ij}L^j$ with $\sum_{j=1}^{\infty} j\cdot|d_{ij}| < \infty$ and $D_i(1)^2 > 0$.

Under Assumptions 2.1*-2.2* and 2.3-2.5, applying the method of PC to the first difference of Model (2.1), we have

$$\Delta X_{i,t} = \boldsymbol{\lambda}_i' \Delta \boldsymbol{F}_t + \Delta e_{i,t}.$$

For convenience, let $x_{i,t} = \Delta X_{i,t}, \boldsymbol{f}_t = \Delta \boldsymbol{F}_t$ and $z_{i,t} = \Delta e_{i,t}$. Then

$$x_{i,t} = \boldsymbol{\lambda}_i' \boldsymbol{f}_t + z_{i,t}, \tag{2.21}$$

which mirrors the stationary factor model (1.13). Denote the PC estimators as $\widehat{\boldsymbol{\lambda}}_i^{pc}$ and $\widehat{\boldsymbol{f}}_t^{pc}$ for $i = 1,\cdots,N$ and $t = 1,\cdots,T$. Hence, we have the estimated residuals of the differenced model, $\widehat{z}_{i,t}^{pc} = x_{i,t} - \left(\widehat{\boldsymbol{\lambda}}_i^{pc}\right)' \widehat{\boldsymbol{f}}_t^{pc}$. Estimated factors and the idiosyncratic components of the original model can be obtained by re-accumulating $\widehat{\boldsymbol{f}}_t^{pc}$ and $\widehat{z}_{i,t}^{pc}$,

$$\widehat{e}_{i,t}^{pc} = \sum_{s=2}^{t} \widehat{z}_{i,s}^{pc}, \;\; \widehat{\boldsymbol{F}}_t^{pc} = \sum_{s=2}^{t} \widehat{\boldsymbol{f}}_s^{pc}. \tag{2.22}$$

Using the method introduced in Section 1.6, we get the ML estimators of the differenced model (2.21), $\widetilde{\boldsymbol{\lambda}}_i^{ml}$ and $\widetilde{\boldsymbol{f}}_t^{ml}$, and then

$$\widetilde{z}_{i,t}^{ml} = x_{i,t} - \left(\widetilde{\boldsymbol{\lambda}}_i^{ml}\right)' \widetilde{\boldsymbol{f}}_t^{ml},$$

and

$$\widetilde{e}_{i,t}^{ml} = \sum_{s=2}^{t} \widetilde{z}_{i,s}^{ml}, \; \widetilde{\boldsymbol{F}}_t^{ml} = \sum_{s=2}^{t} \widetilde{\boldsymbol{f}}_s^{ml}. \tag{2.23}$$

The simulation results suggest that the estimations of the factors and factor loadings based on ML are more efficient than those based on PC, regardless of whether the model is stationary or nonstationary. Further, the idiosyncratic unit root tests based on the likelihood-based residuals generally have better size and higher size-adjusted power, particularly when the cross-sectional dimension is small and the time series dimension is large.

3. Further research

In Papers I-III, the LM-type tests are derived directly from the likelihood of the data. In Paper I, a one-sided LM-type statistic is derived to test for unit roots in the idiosyncratic error terms conditional on the integrated factors and the homoscedastic idiosyncratic error terms. In Paper II, we consider the LM-type statistics to test the unit roots in the common factors conditional on the integrated idiosyncratic error terms and homoscedastic factors. In Paper III, we partly relax the assumption of integrated factors, allowing them to be nonstationary, stationary, or co-integrated linear processes. The idiosyncratic components are also allowed to be heteroscedastic over the cross-section though heteroscedasticity in the time dimension is ruled out.

The results from Papers I-III are quite good, especially with reference to the fact that we present the first to attempt a likelihood approach for the model under consideration. However, one can further relax some of the restrictive assumptions as follows.

First, let us consider an AR(p) process. In Papers I-III, we assume that the idiosyncratic error terms $\{e_{i,t}\}$ are AR(1) processes. A common approach to allow for higher-order dynamics is to adjust for autocorrelation non-parametrically by estimating the long-run variance (e.g., Levin, Lin, and Chu, 2002; Moon and Perron, 2004). It is likely that we can use the long-run variance to adjust the LM test to allow for AR(p) processes. Another approach would be to consider a parametric counterpart, where the AR(p) process is decomposed into an AR(1) process with an AR($p-1$) error term. For example, an AR(2) process can be decomposed as

$$e_{i,t} = \rho_i e_{i,t-1} + \varepsilon_{i,t},$$
$$\varepsilon_{i,t} = \gamma_i \varepsilon_{i,t-1} + v_{i,t},$$

where $|\gamma_i| < 1$ and $v_{i,t}$ are *i.i.d.* Here, the nuisance parameters γ_i would have to be concentrated out. Bai and Li (2012a,b) offer a potential path of consistently estimating these nuisance parameters. Yet, a third approach might be the pre-whitening method of Breitung and Das (2005) and extend it to the dynamic factor structure in Breitung and Das (2008).

Second, let us look at the heterogeneity in the error terms. In Papers I-III, we assume $\rho_i = 1$ for all i. This is not a problem when testing the size properties. However, this might lead to a loss of power under heterogeneous alternatives. If we allow for heterogeneity, the inverse of the information matrix will be difficult to derive. More attention will be needed when deriving the asymptotic distributions.

Third, is the issue of the heteroscedasticity in the error terms. In Papers I-III, we also assume that the error terms are *i.i.d* and Gaussian. When applied to financial data, the error terms are, in most cases, non-normally distributed and often heteroscedastic over time. This problem might be resolved by modifying the LM test to allow for GARCH dynamics (e.g., Lee, 1991).

In Paper IV, we use a quasi-ML method based on a misspecified likelihood function (see Bai and Li, 2012a,b; Doz, Giannone, and Reichlin, 2012; Jungbacker and Koopman, 2008) to get the estimated common factors $\widehat{\boldsymbol{F}}_t^{ml}$ and the estimated idiosyncratic error terms $\widehat{e}_{i,t}^{ml}$. This procedure is then compared with the procedure of Bai and Ng (2004), who estimate the factor model by the method of PC. The results show that the ML-based methods are indeed more efficient than the PC-based methods. Paper IV compares the properties of the unit root tests under the two sets of estimated idiosyncratic terms, $\widehat{e}_{i,t}^{ml}$ and $\widehat{e}_{i,t}^{pc}$. It would also be worthwhile to compare the properties of the analogous co-integration tests under the two sets of estimated common factors, $\widehat{\boldsymbol{F}}_t^{ml}$ and $\widehat{\boldsymbol{F}}_t^{pc}$, and perhaps also to compare the performance of the estimated factors in forecasts on financial or macroeconomic data (e.g., Stock and Waston, 2002; Choi, 2012).

4. Acknowledgements

This thesis would not have been completed without the guidance and support of many considerate and exceptional people. I would like to express my gratitude to some of these persons.

First, I wish to express my sincere and deepest appreciation to my supervisors, Professor Johan Lyhagen and Professor Rolf Larsson. To Johan, thank you for having given me endless support and encouragement. Your optimistic attitude inspires me to this day. No matter how small or big the questions have been, it has always been an enlightened and enjoyable experience to discuss these issues with you. To Rolf, thank you for the exhilarating courses you taught and for having been so generous with your many constructive and insightful comments, from all of which I have greatly benefitted. I also want to express my gratitude to Professor Jonas Andersson, the discussant for my licentiate seminar, for his valuable comments that have substantially improved the content and quality of the first paper.

Part of my PhD work was within the project *Solving Macroeconomic Problems Using Nonstationary Panel Data*, which was funded by the Jan Wallander and Tom Hedelius Foundation through research grants P2005-0117:1 and P2009-0189:1. This is gratefully acknowledged. I also wish to thank SNIC for providing valuable resources through the Uppsala Multidisciplinary Center for Advanced Computational Science (UPPMAX) when the simulations of the fourth paper were performed.

Special thanks are extended to Professor Fan Yang Wallentin, who introduced me the opportunity to study statistics in Sweden, starting as a MA student and later as a PhD candidate. I am also grateful to you for your warm care and splendid advice about not only my study, but also about my family.

I want to acknowledge all my colleagues at the Department of Statistics. You created a wonderful working environment, which I truly appreciate. Thank you, Bo Wallentin, for always showing kindness and never-ending support. Special thanks to Thommy Perlinger and Lars Forsberg for your trust and for allowing me serve as your teaching assistant. This job strongly enhanced my self-confidence. Thank you, Lisbeth Hansson, for kindly arranging my teaching tasks. Inger Persson, I am very appreciative for your organizing fun-related activities to make the working atmosphere pleasant and enjoyable. Your optimism and laughter spreads joy everywhere around you. Katrin Kraus, the conversations with you were always so engaging and pleasant. I also would like to express my gratitude to the administrative staff at the department, including Eva Enefjord, Eva Karlsson, Davoud Emamjomeh and Lars-Göran Svensk, thank you all for solving the many problems I faced in a smooth and efficient manner.

To my fellow PhD students, some of whom have already graduated, thank you for your friendship and for making my PhD experience a fruitful and stimulating one. To Martin Solberger, my co-author (Papers I-III), I will always be grateful for the years that we have been working together and enjoying a great time together. You always had plenty of creative and smart ideas. To Ronnie Pingel, thank you for spending working out time and many beer nights together. The night you organized for watching *The PhD Movie* was so exciting. David Kreiberg, you are so knowledgeable: thank you for being open with intriguing discussions and hilarious jokes. Also thank you for always encouraging me to "be more aggressive" with "Eye of the Tiger". To Björn Andersson and Shaobo Jin, thank you for sharing lunchtime with me. To Jianxin Wei and Xijia Liu, thanks for the great time we had spent together in the past several years when we were studying together.

I owe many thanks to my other Chinese classmates: Chengcheng Hao, Dao Li, Deliang Dai, Feng li, Hao luo, Xia Shen, Xin Zhao, Ying Pang, Yuli Liang, some of whom have finished their studies. I am so lucky to have had you people to share the enormous struggle when facing the difficulty and complex subject matter of statistics. You helped me quickly adjust and made my life easier, especially at the very beginning when we just came to Sweden.

No words can express my gratitude to my family. I would not have gotten so far without their unconditional love and support. To dad, though the Cultural Revolution had ruined your college dream, you and mom overcame enormous difficulties feeding five kids under extremely severe and harsh conditions, and made us well educated. Although you could not wait until I completed my PhD, I know you would be very proud of my accomplishment. Thanks to mom who has given me love and care, as well as continuous support over the years. To my two sisters and two brothers and their respective families, thank you so much for taking care of our parents when I was in Sweden, especially during the period when mom and dad stayed at the hospital and the hard times since dad left us.

I especially wish to express my deep gratitude to my wife, Yanwu: thank you for accompanying me in Sweden and giving me two lovely kids, Yoyo and Lele, who are the most important and precious accomplishments of my life. To Yoyo and Lele, daddy loves you so much. It is you who bring the new meanings to my life.

Xingwu Zhou
Uppsala, August 2014

References

ANDERSON, T. W. (2003): *An Introduction to Multivariate Statistical Analysis*. New York: Johh Wiley & sons, 3rd edn.

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.

——— (2008): "Large dimensional factor analysis," *Foundations and Trends in Econometrics*, 3, 89–163.

——— (2010): "Panel unit root tests with cross-section dependence: A further investigation," *Econometric Theory*, 26, 1088–1114.

BALTAGI, B. H., AND C. KAO (2001): "Nonstationary panels, cointegration in panels and dynamic panels: A survey," in *Advances in Econometrics: Nonstationary Panels, Panel Cointegration, and Dynamic Panels*, ed. by B. H. Baltagi, T. B. Fomby, and R. C. Hill, vol. 15, pp. 7–51. Emerald Group Publishing Limited.

BANERJEE, A. (1999): "Panel data unit roots and cointegration: An overview," *Oxford Bulletin of Economics and Statistics*, 61, 607–629.

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

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

BREITUNG, J., AND S. EICKMEIER (2006): "Dynamic factor models," *Allgemeines Statistisches Archiv*, 90, 27–40.

BREITUNG, J., AND W. MEYER (1994): "Testing for unit roots in panel data: are wages on different bargaining levels cointegrated," *Applied Economics*, 26, 353–361.

BREITUNG, J., AND M. H. PESARAN (2008): "Unit roots and cointegration in panels," in *The Econometrics of Panel Data*, ed. by L. Mátyás, and P. Sevestre, vol. 46, pp. 279–322. Springer Berlin Heidelberg, 3rd edn.

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

CHANG, Y. (2002): "Nonlinear IV unit root tests in panels with cross-sectional dependency," *Journal of Econometrics*, 110, 261–292.

——— (2004): "Bootstrap unit root tests in panels with cross sectional dependency," *Journal of Econometrics*, 120, 263–293.

CHOI, I. (2001): "Unit root tests for panel data," *Journal of International Money and Finance*, 20, 249–272.

——— (2012): "Efficient estimation of factor models," *Econometric Theory*, 28, 274–308.

CHUDIK, A., M. H. PESARAN, AND E. TOSETTI (2011): "Weak and strong cross-section dependence and estimation of large panels," *Econometrics Journal*, 14, C45–C90.

CONNOR, G., AND R. A. KORAJZCYK (1986): "Performance measurement with the arbitrage pricing theory: A new framework for analysis," *Journal of Financial Economics*, 15, 373–394.

——— (1988): "Risk and return in an equilibrium APT: Application to a new test methodoloty," *Journal of Financial Economics*, 21, 255–289.

COX, D. R., AND D. V. HINKLEY (1974): *Theoretical Statistics*. Chapman and Hall, London.

DICKEY, D. A., AND W. A. FULLER (1979): "Distribution of the estimators for autoregressive time series with a unit root," *Journal of the American Statistical Association*, 74, 427–431.

——— (1981): "Likelihood ratio statistics for autoregressive time series with a unit root," *Econometrica*, 49, 1057–1072.

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

ELLIOTT, G., T. J. ROTHENBERG, AND J. H. STOCK (1996): "Efficient tests for an autoregressive unit root," *Econometrica*, 64, 813–836.

ENGLE, R. F. (1984): "Wald, likelihood ratio, and Lagrange multiplier tests in econometrics," in *Handbook of Econometrics*, ed. by Z. Griliches, and M. D. lntriligator, vol. 2, pp. 775–826. Elsevier, Amsterdam.

FISHER, R. A. (1934): *Statistical Methods for Research Workers*. Oliver and Bond, Edinburgh, 5th edn.

FORNI, M., M. HALLIN, M. LIPPI, AND L. REICHLIN (2000): "The generalized dynamic-factor model: Identification and estimation," *The review of Econometrics and Statistics*, 82, 540–555.

GRANGER, C. W. J., AND P. NEWBOLD (1974): "Spurious regression in econometrics," *Journal of Econometrics*, 2, 111–120.

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

IM, K. S., M. H. PESARAN, AND Y. SHIN (2003): "Testing for unit roots in heterogeneous panels," *Journal of Econometrics*, 115, 53–74.

JOHANSEN, S. (1995): *Likelihood-Based Inference in Cointegrated Vector Autoregressive Models*. Oxford university press, Oxford.

JOHNSON, R. A., AND D. W. WICHERN (1988): *Applied Multivariate Statistical Analysis*. Premtoce-Hall, New Jersey, 2nd edn.

JÖRESKOG, K. G. (1967): "Some contributions to maximum likelihood factor analysis," *Psychometrika*, 32, 443–482.

——— (1969): "A general approach to confirmatory maximum likelihood factor analysis," *Psychometrika*, 34, 183–202.

JUNGBACKER, B., AND S. J. KOOPMAN (2008): "Likelihood-based analysis for dynamic factor models," Tinbergen Institute Discussion paper 2008-0007/4.

LAWLEY, D. N., AND A. E. MAXWELL (1963): *Factor Analysis as a Statistical Method*. Butterworths, London.

LEE, J. H. H. (1991): "A lagrange multiplier test for GARCH models," *Economics Letters*, 37, 265–271.

LEVIN, A., C.-F. LIN, AND C.-S. J. CHU (2002): "Unit root tests in panel data: asymptotic and finite-sample properties," *Journal of Econometrics*, 108, 1–24.

LYHAGEN, J. (2008): "Why not use standard panel unit root test for testing PPP," *Economics Bulletin*, 3, 1–11.

MADDALA, G. S., AND S. WU (1999): "A comparative study of unit root tests with panel data and a new simple test," *Oxford Bulletin of Economics and Statistics*, 61, 631–652.

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

NEYMAN, J., AND E. S. PEARSON (1928): "On the use and interpretation of certain test criteria for purposes of statistical inference," *Biometrika*, 20, 175–240, 263–294.

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

PALM, F. C., S. SMEEKES, AND J.-P. URBAIN (2011): "Cross-sectional dependence robust block bootstrap panel unit root tests," *Journal of Econometrics*, 163, 85–104.

PESARAN, M. H. (2004): "General diagnostic tests for cross section dependence in panels," Cambridge Working Papers in Economics No. 435, University of Cambridge, and CESifo Working Paper Series No. 1229.

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

PESARAN, M. H., L. V. SMITH, AND T. YAMAGATA (2013): "Panel unit root tests in the presence of a multifactor error structure," *Journal of Econometrics*, 175, 94–115.

PHILLIPS, P. C. B., AND P. PERRON (1988): "Testing for a unit root in time series regression," *Biometrika*, 75, 335–46.

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

QUAH, D. (1994): "Exploiting cross-section variation for unit root inference in dynamic data," *Economics Letters*, 44, 9–19.

RUBIN, D. B., AND D. T. THAYER (1982): "EM algorithms for ML factor analysis," *Psychometrika*, 47, 69–76.

SAID, S. E., AND D. A. DICKEY (1984): "Testing for unit roots in autoregressive-moving average models of unknown order," *Biometrika*, 71, 599–607.

SCHMIDT, P., AND P. C. B. PHILLIPS (1992): "LM tests for a unit root in the presence of deterministic trends," *Oxford Bulletin of Economics and Statistics*, 54, 257–287.

SILVEY, S. D. (1959): "The Lagrangian multiplier test," *The Annals of Mathematical Statistics*, 30, 389–407.

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

URBAIN, J.-P., AND J. WESTERLUND (2011): "Least squares asymptotics in spurious and cointegrated panel regressions with common and idiosyncratic stochastic trends," *Oxford Bulletin of Economics and Statistics*, 73, 119–139.

WALD, A. (1943): "Tests of statistical hypotheses concerning several parameters when the number of observations is large," *Transactions of the American Mathematical Society*, 54, 426–482.

WANG, S., P. WANG, J. YANG, AND Z. LI (2010): "A generalized nonlinear IV unit root test for panel data with cross-sectional dependence," *Journal of Econometrics*, 157, 101–109.

WESTERLUND, J., AND J. BREITUNG (2013): "Lessons from a decade of IPS and LLC," *Econometric Reviews*, 32, 547–591.

WOLD, H. (1938): *A Study in Analysis of Stationary Time Series*. Almqvist & Wiksell, Uppsala.

Acta Universitatis Upsaliensis

Digital Comprehensive Summaries of Uppsala Dissertations from the Faculty of Social Sciences 103

Editor: The Dean of the Faculty of Social Sciences

A doctoral dissertation from the Faculty of Social Sciences, Uppsala University, is usually a summary of a number of papers. A few copies of the complete dissertation are kept at major Swedish research libraries, while the summary alone is distributed internationally through the series Digital Comprehensive Summaries of Uppsala Dissertations from the Faculty of Social Sciences. (Prior to January, 2005, the series was published under the title "Comprehensive Summaries of Uppsala Dissertations from the Faculty of Social Sciences".)

Distribution: publications.uu.se
urn:nbn:se:uu:diva-233094

ACTA
UNIVERSITATIS
UPSALIENSIS
UPPSALA
2014