UPPSALA
UNIVERSITET

Digital Comprehensive Summaries of Uppsala Dissertations from the Faculty of Social Sciences 90

Likelihood-Based Tests for Common and Idiosyncratic Unit Roots in the Exact Factor Model

MARTIN SOLBERGER

ACTA
UNIVERSITATIS
UPSALIENSIS
UPPSALA
2013

ISSN 1652-9030
ISBN 978-91-554-8754-6
urn:nbn:se:uu:diva-207296

Dissertation presented at Uppsala University to be publicly examined in Hörsal 2, Ekonomikum, Kyrkogårdsgatan 10, Uppsala, Friday, November 8, 2013 at 10:00 for the degree of Doctor of Philosophy. The examination will be conducted in English.

Abstract
Solberger, M. 2013. Likelihood-Based Tests for Common and Idiosyncratic Unit Roots in the Exact Factor Model. Acta Universitatis Upsaliensis. *Digital Comprehensive Summaries of Uppsala Dissertations from the Faculty of Social Sciences* 90. 51 pp. Uppsala. ISBN 978-91-554-8754-6.

Dynamic panel data models are widely used by econometricians to study over time the economics of, for example, people, firms, regions, or countries, by pooling information over the cross-section. Though much of the panel research concerns inference in stationary models, macroeconomic data such as GDP, prices, and interest rates are typically trending over time and require in one way or another a nonstationary analysis. In time series analysis it is well-established how autoregressive unit roots give rise to stochastic trends, implying that random shocks to a dynamic process are persistent rather than transitory. Because the implications of, say, government policy actions are fundamentally different if shocks to the economy are lasting than if they are temporary, there are now a vast number of univariate time series unit root tests available. Similarly, panel unit root tests have been designed to test for the presence of stochastic trends within a panel data set and to what degree they are shared by the panel individuals. Today, growing data certainly offer new possibilities for panel data analysis, but also pose new problems concerning double-indexed limit theory, unobserved heterogeneity, and cross-sectional dependencies. For example, economic shocks, such as technological innovations, are many times global and make national aggregates cross-country dependent and related in international business cycles.

Imposing a strong cross-sectional dependence, panel unit root tests often assume that the unobserved panel errors follow a dynamic factor model. The errors will then contain one part which is shared by the panel individuals, a common component, and one part which is individual-specific, an idiosyncratic component. This is appealing from the perspective of economic theory, because unobserved heterogeneity may be driven by global common shocks, which are well captured by dynamic factor models. Yet, only a handful of tests have been derived to test for unit roots in the common and in the idiosyncratic components separately. More importantly, likelihood-based methods, which are commonly used in classical factor analysis, have been ruled out for large dynamic factor models due to the considerable number of parameters.

This thesis consists of four papers where we consider the exact factor model, in which the idiosyncratic components are mutually independent, and so any cross-sectional dependence is through the common factors only. Within this framework we derive some likelihood-based tests for common and idiosyncratic unit roots. In doing so we address an important issue for dynamic factor models, because likelihood-based tests, such as the Wald test, the likelihood ratio test, and the Lagrange multiplier test, are well-known to be asymptotically most powerful against local alternatives.

Our approach is specific-to-general, meaning that we start with restrictions on the parameter space that allow us to use explicit maximum likelihood estimators. We then proceed with relaxing some of the assumptions, and consider a more general framework requiring numerical maximum likelihood estimation. By simulation we compare size and power of our tests with some established panel unit root tests. The simulations suggest that the likelihood-based tests are locally powerful and in some cases more robust in terms of size.

Keywords: panel unit root, dynamic factors, maximum likelihood, Lagrange multiplier, likelihood ratio, factor analysis

Martin Solberger, Uppsala University, Department of Statistics, SE-751 20 Uppsala, Sweden.

ISSN 1652-9030
ISBN 978-91-554-8754-6
urn:nbn:se:uu:diva-207296 (http://urn.kb.se/resolve?urn=urn:nbn:se:uu:diva-207296)

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. (2012) On a Likelihood Ratio Test for Idiosyncratic Unit Roots in the Exact Factor Model with Nonstationary Common Shocks. Submitted.

III Solberger M., and Zhou, X. (2012) LM-Type Tests for Idiosyncratic and Common Unit Roots in the Exact Factor Model with AR(1) Dynamics.

IV Zhou, X., and Solberger, M. (2013) An LM-Type Test for Idiosyncratic Unit Roots in the Exact Factor Model under Misspecification.

Contents

1. Notation

The following abbreviations and symbols will be used in this summary:

E	Expected value
Var	Variance
Cov	Covariance
$\sup$	Supremum
$\arg\max$	Argument of the maximum
$\arg\min$	Argument of the minimum
$\lim$	Limit
$\sim$	"Distributed as"
$\overset{app.}{\sim}$	"Distributed approximately as"
$\forall$	"For all"
$\exists$	"There exists"
$\in$	"Is an element of"
$\subset$	"Is a subset of"
$\mathbb{R}$	The set of real numbers
$\mathbb{Z}$	The set of integers
N	Number of panel individuals
T	Time series end point
$T \to$	Limit taken over T
$N,T \to$	Joint limit, with limit taken over T and N simultaneously
$(T,N)_s \to$	Sequential limit, with limit taken over T followed by over N
$\overset{p}{\to}$	Convergence in probability
$\overset{d}{\to}$	Convergence in distribution
$N_p(\boldsymbol{\mu},\boldsymbol{\Omega})$	A p-variate normal distribution with mean $\boldsymbol{\mu}$ and variance $\boldsymbol{\Omega}$
χ^2_d	A chi-square distribution with d degrees of freedom

We will also use the following notation; For a matrix $\mathbf{A} \in \mathbb{R}^{n\times n}$, $\mathbf{A} > 0$ denotes positive definiteness, $|\mathbf{A}|$ denotes the determinant, $\mathrm{tr}(\mathbf{A})$ denotes the trace, and $\varphi_1(\mathbf{A}) \geq \varphi_2(\mathbf{A}) \geq \cdots \geq \varphi_n(\mathbf{A})$ denote the eigenvalues. For a matrix $\mathbf{A} \in \mathbb{R}^{n\times m}$,

$\text{rk}(\mathbf{A})$ denotes the rank, $||\mathbf{A}|| = [\text{tr}(\mathbf{A}\mathbf{A}')]^{1/2}$ denotes the Frobenius norm, and $\mathbf{A}_v = \text{vec}(\mathbf{A})$ is the vectorization which stacks the columns in an $nm \times 1$ vector. Unless specified differently, $[a_{i,j}]_{n\times m}$ is an $n \times m$ matrix with element $a_{i,j}$ corresponding to the ith row and jth column, $\text{diag}\,(a_1, a_2, \ldots, a_n)$ is an $n \times n$ diagonal matrix with entries $a_1, a_2, \ldots, a_n$ and $\mathbf{I}_n$ is the $n \times n$ identity matrix. For a stochastic process $X_{N,T}$ and deterministic sequences $a_{N,T}$ and $b_{N,T}$, we say that $X_{N,T} = O_p(a_{N,T})$ if $X_{N,T}/a_{N,T}$ is bounded in probability, $X_{N,T} = o_p(a_{N,T})$ if $X_{N,T}/a_{N,T}$ converges in probability to zero, and $b_{N,T} = O(a_{N,T})$ if $b_{N,T}/a_{N,T}$ is bounded for all N and T.

2. Introduction

Panel unit root tests were first introduced to gain power by exploiting the cross-sectional dimension for otherwise relatively short time series. However, the increasing amount of available data has pushed the framework to consider very large panels with both the time dimension and cross-sectional dimension large. Also, the assumption of cross-sectional independence imposed by the first generation of tests has been deemed unrealistic for many panels in practice and documented to result in size distortions (e.g. O'Connel, 1998; Maddala and Wu, 1999; Strauss and Yigit, 2003; Urbain and Westerlund, 2006).

A popular approach to model the cross-sectional dependence among the second generation of panel unit root tests is to impose unobserved common factors. This is convenient for at least two reasons; (i) the unobserved common factors reduce the inference to concern the factor space, which generally is much smaller than the cross-sectional space; and (ii) this type of dependence is natural for a large set of economic models, as economic theory often predicts that shocks to a system will affect its individuals (e.g. countries) proportionally, and that a relatively small set of shocks may drive a large system (see e.g. Breitung and Eickmeier, 2006).

In this thesis we consider the exact factor model with orthogonal idiosyncratic components and propose likelihood based tests for common and idiosyncratic unit roots. To our knowledge this has not been done before, and the results should contribute to the challenging task of developing likelihood based unit root tests in dynamic factor models.

3. The likelihood trinity

Three likelihood based tests are frequently used in econometrics; the Likelihood ratio test, the Wald test, and the Lagrange multiplier test (Score test).[1] Here we give a short summary of these tests. More extensive reviews can be found in e.g. Breusch and Pagan (1980), Buse (1982), Engle (1984), and Bera and Bilias (2001).

Let $\mathbf{x}$ be a real-valued $n \times 1$ random vector drawn from a joint density $f(\mathbf{x}|\boldsymbol{\theta})$ where $\boldsymbol{\theta}$ is a $k \times 1$ vector of parameters that lie in some parameter space $\boldsymbol{\Theta} \subset \mathbb{R}^k$. Suppose that under the null hypothesis $\boldsymbol{\theta} \in \boldsymbol{\Theta}_0 \subset \boldsymbol{\Theta}$, and that under the alternative hypothesis $\boldsymbol{\theta} \in \boldsymbol{\Theta}_1 \subset \boldsymbol{\Theta}$, where $\boldsymbol{\Theta}_0 \cup \boldsymbol{\Theta}_1 = \boldsymbol{\Theta}$. For a statistical test, the performance crucially depends on the size and the power of the test-statistic.[2] Let C denote the critical region for some test-statistic $\mathscr{T} : \mathbf{x} \mapsto \mathbb{R}$. The size is given by the probability of rejecting a true null hypothesis,

$$P(\mathscr{T} \in C|\boldsymbol{\theta} \in \boldsymbol{\Theta}_0),$$

and the power is given by the probability of rejecting a false null hypothesis,

$$P(\mathscr{T} \in C|\boldsymbol{\theta} \in \boldsymbol{\Theta}_1).$$

Usually the null hypothesis only concerns a subset of the parameters such that the parameter vector may be partitioned as $\boldsymbol{\theta} = (\boldsymbol{\theta}_1', \boldsymbol{\theta}_2')'$, where $\boldsymbol{\theta}_2$ $(k_2 \times 1)$ is subject to restrictions and $\boldsymbol{\theta}_1$ $(k_1 \times 1)$ is left unrestricted. The null hypothesis may then be specified as

$$H_0 : \boldsymbol{\theta}_2 = \boldsymbol{\theta}_2^0, \quad \boldsymbol{\theta}_1 \text{ unrestricted.} \tag{3.1}$$

Because any well-behaved test will have power tending to one when the sample size increases, it is common practice to look at the asymptotic and finite behavior under a shrinking neighborhood of $\boldsymbol{\theta}_2^0$. The alternative hypothesis may then be specified as a sequence of local alternatives,

$$H_1 : \boldsymbol{\theta}_2^n = \boldsymbol{\theta}_2^0 + \frac{\mathbf{c}}{\sqrt{n}}, \quad \boldsymbol{\theta}_1 \text{ unrestricted,} \tag{3.2}$$

for some vector $\mathbf{c} \in \mathbb{R}^{k_2}$. A test that performs equally well in any direction $\mathbf{c}$, is an *invariant* test.

[1] The Likelihood ratio test was introduced by Neyman and Pearson (1928), Wald tests are based on the analysis of Wald (1943), and the Lagrange Multiplier test is based on the two papers by Aitchison and Silvey (1958) and Silvey (1959) and is equivalent to the Score test of Rao (1948).

[2] A test is a function, $\mathscr{H}$ say, which maps the values of a test-statistic, $\mathscr{T}$ say, to the set consisting of the decisions "not reject H_0" (d_0) and "reject H_0" (d_1); $\mathscr{H} \circ \mathscr{T} : \mathbf{x} \mapsto \{d_0, d_1\}$.

Let $\mathscr{L}(\boldsymbol{\theta}|\mathbf{x}) = f(\mathbf{x}|\boldsymbol{\theta})$ denote the likelihood as a function of $\boldsymbol{\theta}$, conditional on $\mathbf{x}$, and let $l(\boldsymbol{\theta}|\mathbf{x}) = \log \mathscr{L}(\boldsymbol{\theta}|\mathbf{x})$ denote the corresponding log-likelihood. Because the logarithmic function is monotonic, $l(\boldsymbol{\theta}|\mathbf{x})$ and $\mathscr{L}(\boldsymbol{\theta}|\mathbf{x})$ share maximizing arguments, but the log-likelihood is usually much easier to maximize. Assuming certain regularity conditions are met, the log-likelihood maximizes at a point satisfying

$$\mathbf{V}(\widehat{\boldsymbol{\theta}}) = \left.\frac{\partial l(\boldsymbol{\theta}|\mathbf{x})}{\partial \boldsymbol{\theta}}\right|_{\boldsymbol{\theta}=\widehat{\boldsymbol{\theta}}} = \mathbf{0},$$

where $\mathbf{V}(\boldsymbol{\theta}) = \partial l(\boldsymbol{\theta}|\mathbf{x})/\partial \boldsymbol{\theta}$ is the score and $\widehat{\boldsymbol{\theta}}$ is the maximum likelihood estimator (MLE) with variance given by the inverse of the Fisher information, $Var(\widehat{\boldsymbol{\theta}}) = \mathbf{J}(\boldsymbol{\theta})^{-1}$, where

$$\mathbf{J}(\boldsymbol{\theta}) \equiv -E\left(\frac{\partial^2 l(\boldsymbol{\theta})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}'}\right).$$

Let $\widetilde{\boldsymbol{\theta}}$ denote the MLE of $\boldsymbol{\theta}$ subject to an $m \times 1$ $(m < k)$ vector of constraints $\mathbf{h}(\boldsymbol{\theta}) = \mathbf{0}$, where $\mathbf{h}(\boldsymbol{\theta}) : \mathbb{R}^k \mapsto \mathbb{R}^m$. The Lagrange multiplier test aims to maximize the Lagrangian function

$$F = l(\boldsymbol{\theta}) - \boldsymbol{\kappa}'\mathbf{h}(\boldsymbol{\theta}),$$

with respect to $\widetilde{\boldsymbol{\theta}}$, and $\boldsymbol{\kappa}$, an $m \times 1$ vector of Lagrange multipliers. The first-order conditions are

$$\left.\frac{\partial F}{\partial \boldsymbol{\theta}}\right|_{\boldsymbol{\theta}=\widetilde{\boldsymbol{\theta}}} = \mathbf{V}(\widetilde{\boldsymbol{\theta}}) - \mathbf{H}(\widetilde{\boldsymbol{\theta}})\widetilde{\boldsymbol{\kappa}} = \mathbf{0},$$
$$\left.\frac{\partial F}{\partial \boldsymbol{\kappa}}\right|_{\boldsymbol{\theta}=\widetilde{\boldsymbol{\theta}}} = \mathbf{h}(\widetilde{\boldsymbol{\theta}}) = \mathbf{0},$$

where $\mathbf{H}(\boldsymbol{\theta}) = \partial \mathbf{h}(\boldsymbol{\theta})'/\partial \boldsymbol{\theta}$ is a $k \times m$ matrix. It can be shown that, under certain regularity conditions, $\mathbf{C_n}^{-1/2}\mathbf{V}(\widetilde{\boldsymbol{\theta}}) \xrightarrow{d} N_k(\mathbf{0}, \boldsymbol{\Sigma})$, where $\mathbf{C_n}^{-1/2}$ is some suitable scaling matrix, and $\boldsymbol{\Sigma} = \lim_{n\to\infty} \mathbf{C_n}^{-1/2}\mathbf{J}(\boldsymbol{\theta})\mathbf{C_n}^{-1/2}$. The resulting statistic is

$$LM = \widetilde{\boldsymbol{\kappa}}'\mathbf{H}(\widetilde{\boldsymbol{\theta}})'\mathbf{J}(\boldsymbol{\theta})^{-1}\mathbf{H}(\widetilde{\boldsymbol{\theta}})\widetilde{\boldsymbol{\kappa}} = \mathbf{V}(\widetilde{\boldsymbol{\theta}})'\mathbf{J}(\widetilde{\boldsymbol{\theta}})^{-1}\mathbf{V}(\widetilde{\boldsymbol{\theta}}), \tag{3.3}$$

where the right part is known as the score statistic. If the constraints $\mathbf{h}(\boldsymbol{\theta}) = \mathbf{0}$ are true, then the score evaluated at the restricted MLE, $\mathbf{V}(\widetilde{\boldsymbol{\theta}})$, will be close to the score evaluated at the unrestricted MLE, $\mathbf{V}(\widehat{\boldsymbol{\theta}}) = \mathbf{0}$ (and likewise the lagrange multiplier $\widetilde{\boldsymbol{\kappa}}$ will be close to zero), and so the intuition behind the LM-test is to reject the constraints for large values of (3.3).

When restrictions are imposed only on a subset of the parameters, the statistic is simplified. Suppose we have a null hypothesis of the form (3.1) and let $\widetilde{\boldsymbol{\theta}} = (\widetilde{\boldsymbol{\theta}}_1', \boldsymbol{\theta}_2^{0\prime})'$ denote the constrained MLE, where $\widetilde{\boldsymbol{\theta}}_1$ is the restricted MLE of $\boldsymbol{\theta}_1$ under the null hypothesis and $\boldsymbol{\theta}_2^0$ is the restriction imposed under the null hypothesis. The null hypothesis may then be formulated as $H_0 : \mathbf{h}(\boldsymbol{\theta}) = \mathbf{0}$, where $\mathbf{h}(\boldsymbol{\theta}) = (\mathbf{0}\;\; \mathbf{I}_{k_2})\boldsymbol{\theta} - \boldsymbol{\theta}_2^0 = \boldsymbol{\theta}_2 - \boldsymbol{\theta}_2^0$ is a $k_2 \times 1$ vector. Partition the information matrix and its inverse as

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

where $\mathbf{J}_{11} = -E\left(\partial^2 l(\boldsymbol{\theta}) / \partial\boldsymbol{\theta}_1 \partial\boldsymbol{\theta}_1'\right)$, $\mathbf{J}_{12} = -E\left(\partial^2 l(\boldsymbol{\theta}) / \partial\boldsymbol{\theta}_1 \partial\boldsymbol{\theta}_2'\right) = \mathbf{J}_{21}'$, and $\mathbf{J}_{22} = -E\left(\partial^2 l(\boldsymbol{\theta}) / \partial\boldsymbol{\theta}_2 \partial\boldsymbol{\theta}_2'\right)$. Under maximum likelihood estimation we have that $\mathbf{V}(\boldsymbol{\theta}) = \left[\mathbf{0}, (\partial l(\boldsymbol{\theta}) / \partial\boldsymbol{\theta}_2)'\right]'$. Hence the LM statistic (3.3) becomes (in score form)

$$LM = \left(\frac{\partial l(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}_2}\right)' \mathbf{J}^{22} \left(\frac{\partial l(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}_2}\right)\Bigg|_{\boldsymbol{\theta}=\widetilde{\boldsymbol{\theta}}}, \tag{3.4}$$

where, from a well-known result, $\mathbf{J}^{22} = (\mathbf{J}_{22} - \mathbf{J}_{21}\mathbf{J}_{11}^{-1}\mathbf{J}_{12})^{-1}$. If $\mathbf{J}(\boldsymbol{\theta})$ is block-diagonal, then $\mathbf{J}^{22} = \mathbf{J}_{22}^{-1}$. Although the statistic (3.4) is actually the score statistic by Rao (1948), we will continue to refer to it as an LM statistic.

By Taylor expansion of $\mathbf{h}(\widehat{\boldsymbol{\theta}})$ around the hypothesized value $\boldsymbol{\theta}_0$ it follows that, under the null hypothesis, $\sqrt{n}\mathbf{h}(\widehat{\boldsymbol{\theta}}) \xrightarrow{d} N_m(\mathbf{0}, \mathbf{H}(\widehat{\boldsymbol{\theta}})'\mathbf{J}(\widehat{\boldsymbol{\theta}})^{-1}\mathbf{H}(\widehat{\boldsymbol{\theta}}))$. This leads to the Wald statistic, which is based only on the unrestricted MLEs, and is defined as

$$W = \mathbf{h}(\widehat{\boldsymbol{\theta}})' \left[\mathbf{H}(\widehat{\boldsymbol{\theta}})'\mathbf{J}(\widehat{\boldsymbol{\theta}})^{-1}\mathbf{H}(\widehat{\boldsymbol{\theta}})\right]^{-1} \mathbf{h}(\widehat{\boldsymbol{\theta}}). \tag{3.5}$$

If the constraints $\mathbf{h}(\boldsymbol{\theta}) = \mathbf{0}$ are true, then $\mathbf{h}(\widehat{\boldsymbol{\theta}}) = \widehat{\boldsymbol{\theta}}_2 - \boldsymbol{\theta}_2^0$ will be close to zero, and so we reject also the Wald test for large values of (3.5).

The Likelihood ratio statistic is

$$LR = \frac{\sup\limits_{\boldsymbol{\theta}\in\boldsymbol{\Theta}_0} \mathscr{L}(\boldsymbol{\theta}|\mathbf{x})}{\sup\limits_{\boldsymbol{\theta}\in\boldsymbol{\Theta}_1} \mathscr{L}(\boldsymbol{\theta}|\mathbf{x})}. \tag{3.6}$$

If the constraints imposed under the null hypothesis are true, then the ratio (3.6) should be close to 1, and otherwise it will be less than 1. The statistic is usually expressed in logarithms, $LR = \sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}_0} l(\boldsymbol{\theta}|\mathbf{x}) - \sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}_1} l(\boldsymbol{\theta}|\mathbf{x})$, where a large spread between the restricted likelihood and the unrestricted likelihood indicates a rejection of the constraints.

It is well known that, in well-behaved problems, the LM statistic, the Wald statistic, and the LR statistic have the same asymptotic distribution (sometimes after appropriate scaling or transformation) and equivalent asymptotic power characteristics. Especially, a test based on any of the three statistics is an asymptotically locally most powerful invariant test. However, the statistics might perform differently in small samples.

4. Background

4.1 Unit root tests

The interest here is on the autoregressive class of dynamic models.[1] For the univariate case, the autoregressive process of order k, AR(k), is

$$x_t = \pi + \rho_1 x_{t-1} + \rho_2 x_{t-2} + \cdots + \rho_k x_{t-k} + \varepsilon_t, \quad t \in \mathbb{Z}, \tag{4.1}$$

where π is a constant, $\rho_1, \rho_2, \ldots, \rho_k$ are coefficients, and ε_t is identically and independently distributed (iid) over time. The process (4.1) can be efficiently summarized as $\rho(L)x_t = \pi + \varepsilon_t$, where $\rho(L)$ is a lag-polynomial such that $\rho(L) = 1 - \rho_1 L - \rho_2 L^2 - \cdots - \rho_k L^k$, where L is the lag-operator with the property $\{\forall q \in \mathbb{Z} : L^q x_t = x_{t-q}\}$. The process is *strictly* stationary if, for any values $s_1, s_2, \ldots, s_n$, the joint probability distribution of $(x_t, x_{t+s_1}, x_{t+s_2} \ldots, x_{t+s_n})$ depends only on the lag lengths $s_1, s_2, \ldots, s_n$ and not on the time (t). If the first two moments are finite, then strict stationarity implies *weak* stationarity (covariance stationarity), meaning, similarly, for any value s, that $E(x_t) = E(x_{t-s})$ and that the covariance $E\left(x_t - E(x_t)\right)\left(x_{t-s} - E(x_{t-s})\right)$ is independent of time and only depends on the lag-length s. If all the roots of the characteristic polynomial $A(z) = 1 - \rho_1 z - \rho_2 z^2 - \cdots - \rho_k z^k = 0$ lie outside the unit circle, then x_t is weakly stationary. The assumption of a weakly stationary process is standard for many models, and is often necessary for estimation and prediction. Every weakly stationary AR(k) process has an invertible infinite moving average representation, MA(∞), defined for the process (4.1) as

$$x_t = \rho(L)^{-1}\pi + \rho(L)^{-1}\varepsilon_t = \mu + \sum_{j=0}^{\infty} \psi_j \varepsilon_{t-j},$$

where $\rho(L)^{-1} = 1 + \psi_1 L + \psi_2 L^2 + \ldots$ has the property $\rho(L)^{-1}\rho(L) = 1$, μ is a constant, and the MA-coefficients are square summable, $\sum_{j=0}^{\infty} \psi_j^2 < \infty$.[2] The restrictions on the MA coefficients are sometimes defined in terms of absolute summability, like $\sum_{j=0}^{\infty} |\psi_j| < \infty$, or $\sum_{j=0}^{\infty} j|\psi_j| < \infty$, which both imply square summability.

We have assumed here that the time series has started in the infinite past. However, in practice, a time series is only observed over a finite number of

[1] Basic summaries of these models can be found in most time series textbooks, e.g. Hamilton (1994).

[2] In fact, from a famous theorem by Wold (1938), every zero-mean weakly stationary process $\{w_t\}$ can be decomposed into an MA part and a deterministic part.

time periods, say $t = 1, 2, \ldots, T$, where assumptions on the past will have an impact on inference. In this case, for a weakly stationary process, it is a necessary, but not sufficient, condition that the roots of $A(z)$ lie outside the unit circle. We also need the distribution of the starting values $x_0, x_{-1}, \ldots, x_{-k+1}$ (characterizing here the very immediate past) to coincide with the particular stationary distribution of x_t.

Whenever the process (4.1) is nonstationary, the roots of the characteristic polynomial could either lie inside the unit circle (explosive), which is generally not an interesting case for economic time series, or they could lie on the unit circle. For the latter case we say that the process has one or more unit roots, where the number of unit roots decides of what order the process is integrated. We say that a process is integrated of order d, denoted $\sim I(d)$, if d is the minimum number of differences required to make the process stationary. If we define the difference operator as $\Delta = (1-L)$, where L is the lag-operator, then a process x_t is $I(d)$ if $\Delta^d x_t \sim I(0)$. It follows from the characteristic polynomial that $A(1) = 0 \Leftrightarrow 1 - \rho_1 - \rho_2 - \cdots - \rho_k = 0$. Hence if $\sum_{m=1}^{k} \rho_m = 1$, then x_t has at least one unit root. A process that is $I(1)$ is often referred to as difference-stationary. Higher orders of integration ar rarely of empirical interest, and $I(0)$ processes and $I(1)$ processes cover a wide range of variables of consideration for economic analysis.[3]

Of special interest in this thesis is the AR(1) process, $x_t = \pi + \rho x_{t-1} + \varepsilon_t$, which is stationary for $|\rho| < 1$. If $\rho = 1$, then the AR(1) process has a unit root, which will give rise to a stochastic trend (as opposed to a deterministic trend) because the random shocks ε_t will cumulate over time. The AR(1) process with a unit root is called a random walk if $\pi = 0$, and a random walk with drift if $\pi \neq 0$. The random walk with drift will have both a deterministic trend and a stochastic trend, which can be seen from rewriting the process recursively as

$$\begin{aligned}
x_t &= \pi + x_{t-1} + \varepsilon_t \\
&= \pi + (\pi + x_{t-2} + \varepsilon_{t-1}) + \varepsilon_t \\
&= \pi + \pi + (\pi + x_{t-3} + \varepsilon_{t-2}) + \varepsilon_{t-1} + \varepsilon_t \\
&= \ldots \\
&= x_0 + \pi t + \sum\nolimits_{s=1}^{t} \varepsilon_s,
\end{aligned}$$

where x_0 is some fixed starting value, πt is a deterministic trend of order $O(t)$, and $\sum_{s=1}^{t} \varepsilon_s$ is a stochastic trend of order $O_p(t^{1/2})$. Hence, for the random walk with drift, the deterministic trend will asymptotically dominate the stochastic trend. Note that both the deterministic trend and the stochastic trend give rise to nonstationarity, but that they both stem from that $\rho = 1$, and x_t is still $I(1)$. If $|\rho| < 1$, then there is neither a deterministic nor a stochastic trend,

[3] We will refer to econometric applications through out this summary. Naturally, if the models we consider here are of interest for other areas, then we could simply change the economic interpretation into which ever interpretation is suitable.

and x_t is $I(0)$. Similarly, a pure random walk will have a stochastic trend only, $x_t = x_0 + \sum_{s=1}^{t} \varepsilon_s$, and is $I(1)$. Due to the cumulative property of the autoregressive process, it is usually fruitful to treat deterministic components outside the process, and think of a drift-less $I(1)$ process. As described by, e.g., Granger (1986), series that are $I(1)$ without drift have variances that depend on time and go to infinity as time goes to infinity, wander widely (appear to be trending), and have indefinitely long memory the (random shocks have permanent effects), while series that are $I(0)$ have finite variances, tend to fluctuate around the mean (mean-reverting), and have finite memory (random shocks have transitory effects).

A convenient property of the AR(k) process is that it may be decomposed as an AR(1) process with an AR($k-1$) error term (see e.g. Karanasos, 1998), such that (leaving out the constant)

$$x_t = \rho x_{t-1} + z_t, \tag{4.2}$$

$$z_t = \phi_1 z_{t-1} + \phi_2 z_{t-2} + \cdots + \phi_{k-1} z_{t-k+1} + \varepsilon_t, \tag{4.3}$$

where ε_t is iid. By expressing (4.2) in terms of z_{t-1}, z_{t-2} etc., and substituting into (4.3) we find that

$$z_t = \phi_1 (x_{t-1} - \rho x_{t-2}) + \cdots + \phi_{k-1} (x_{t-k+1} - \rho x_{t-k}) + \varepsilon_t, \tag{4.4}$$

and substituting (4.4) back into (4.2) it follows that

$$\begin{aligned} x_t = \ & (\rho + \phi_1) x_{t-1} + (\phi_2 - \rho \phi_1) x_{t-2} + \dots \\ & \cdots + (\phi_{k-1} - \rho \phi_{k-2}) x_{t-k+1} + (-\phi_{k-1} \rho) x_{t-k} + \varepsilon_t. \end{aligned}$$

From the characteristic polynomial we then have that

$$\begin{aligned} A(1) &= 1 - (\rho + \phi_1) - (\phi_2 - \rho \phi_1) - \cdots - (\phi_{k-1} - \rho \phi_{k-2}) - (-\phi_{k-1} \rho) \\ &= 1 - \rho + (\rho - 1) \sum_{m=1}^{k-1} \phi_m. \end{aligned}$$

Hence if $\rho = 1$, then $A(1) = 0$, which corresponds to a unit root. Because, assuming weak stationarity, the AR($k-1$) error term in (4.3) has an MA(∞) representation, a common decomposition is of the type

$$\begin{aligned} x_t &= \gamma + \delta t + v_t, \qquad t = 1, 2, \dots, T, \\ v_t &= \rho v_{t-1} + u_t, \\ u_t &= C(L) \varepsilon_t, \end{aligned} \tag{4.5}$$

where γ is a constant, δ and ρ are coefficients, ε_t is iid, $C(L) = \sum_{j=0}^{\infty} c_j L^j$ with the sequence $\{c_j\}_0^{\infty}$ square summable, and the initial condition is set at $t = 0$. That is, x_t is decomposed into a deterministic part ($\gamma + \delta t$) and a stochastic part (v_t) with some initial value (v_0), and where the stochastic part is AR(1) with an autocorrelated error (u_t). Of course, other types of dynamics, such as higher order trend polynomials, could be considered in a similar fashion. Note that treating the deterministic components outside the AR process leave them unaffected of the autoregressive parameter, so that when there is a unit root they will not cumulate into higher order deterministic components.

The process x_t in (4.5) is nonstationary if $\delta \neq 0$ or $\rho = 1$. However, the source of non-stationarity matters. If $\delta \neq 0$ and $|\rho| < 1$, then x_t is nonstationary due to a deterministic trend, but the deviations from this trend are stationary because $v_t \sim I(0)$. The process is then defined as trend-stationary. If $\delta = 0$ and $\rho = 1$, then x_t is nonstationary due to the stochastic trend imposed by the nonstationary process v_t. These two cases could very well look similar in a graph. If $\delta \neq 0$ and $\rho = 1$, then x_t has a deterministic linear trend where the deviations from this trend are nonstationary because $v_t \sim I(1)$. The classical hypotheses for this model would be $H_0 : \rho = 1, \delta = 0$ versus $H_0 : |\rho| < 1$, i.e. testing for a unit root versus a trend-stationary process. For statistics designed to test such a null hypothesis, the asymptotic properties will usually be dependent on nuisance parameters that are functions of the long-run variance of u_t.[4] This is usually solved by using parametric or non-parametric methods (e.g. by the principles of Newey and West, 1987) to consistently estimate the nuisance parameters (see e.g. Phillips and Xiao, 1998, for a review on unit root tests).

Why are unit-roots important? In econometrics, time series are mainly used for forecasting and hypothesis testing, i.e. testing whether economic theory is supported by empirical data. If the data is nonstationary, but this property is ignored or overseen, then standard asymptotic theory relying on conventional laws of large numbers and central limit theorems does not apply, and inference will be misleading. Further, the potential benefits from policy actions, such as government spending, are fundamentally different if shocks to the economic system have non-transitory effects than if they have transitory effects. Applications of unit roots may be found within a number of economically oriented fields concerning variables of interest such as GDP, interest rates, inflation, exchange rates, asset prices, bond yields, unemployment, wages etc. (see e.g. Dolado, Jenkinson, and Sosvilla-Rivero, 1990, and references therein). Although there seems to be a general consensus on that most macroeconomic variables are trending, a great deal of the discussion has concerned, and still concerns, whether or not the trends are stochastic or deterministic. It could be argued that a time series is trend-stationary, perhaps with multiple trend-shifts, rather than exhibiting a stochastic trend (see e.g. Nelson and Plosser, 1982; Perron, 1989; Zivot and Andrews, 1992; Rudebusch, 1992; Lumsdaine and Papell, 1997; Papell and Prodan, 2007; Charles and Darné, 2012). As an illustration on these issues, Figure 4.1a-d show some different stationary and nonstationary series where we have simulated data from (4.5). For all series we let $\gamma = 20$ and $\varepsilon_t \sim N(0, \sqrt{50})$, where; for (a) $\delta = 0$, $\rho = 0.3$ (stationary);

[4]The long-run variance of a stationary process z_t is $\omega_z^2 = Var(z_t) + 2\sum_{h=1}^{\infty} Cov(z_t, z_{t-h})$.

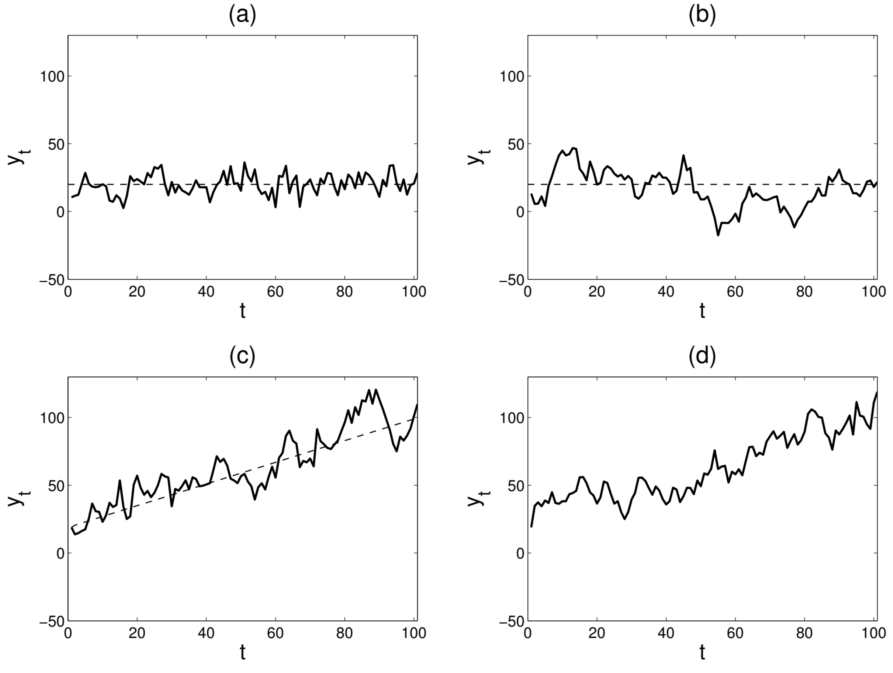

Figure 4.1. Some different stationary and nonstationary series based on (4.5).

for (b) $\delta = 0$, $\rho = 0.9$ (stationary); for (c) $\delta = 0.8$, $\rho = 0.8$ (nonstationary, but trend-stationary); and for (d) $\delta = 0$, $\rho = 1$ (nonstationary). Note how the stationary series (a) and (b) are clearly mean reverting, but also that the larger autoregressive parameter in (b) gives the series a longer swing because the shocks ε_t have more persistent effects over time. The series in (c) and (d) are clearly different from the series in (a) and (b), and though (c) and (d) happen to look similar, they are fundamentally different. The series in (c) has a deterministic trend and could potentially be predicted around this trend, while the series in (d) has a stochastic trend, and its best prediction is never anything else but the last observed value.

In empirical macroeconomics it is common practice to isolate the cyclical (stationary) component of the model (see e.g. Campbell and Perron, 1991). That is, referring to model (4.5), we would seek either to locate v_t, or if $v_t \sim I(1)$, then we would seek to locate u_t. This will require to remove the nonstationarity parts, e.g. by differencing and/or de-trending in the spirit of Box and Jenkins (see e.g. Box, Jenkins, and Reinsel, 2008). Unit root tests may be considered key in this process. To see this, note that taking first-differences of (4.5) yields $\Delta x_t = \gamma - \gamma + \delta t - \delta(t-1) + \Delta v_t = \delta + \Delta v_t$. If $v_t \sim I(1)$, then $\Delta v_t = u_t$, and so $\Delta x_t = \delta + u_t$ is stationary because $u_t \sim I(0)$. However, if

$v_t \sim I(0)$, then we will have over-differenced the cyclical part, which will have led to a loss of information. Also, differencing stationary series leads to inferential problems related to noninvertible AR and MA representations (see e.g. Stock, 1994). In that case a proper procedure would be to only remove the deterministic trend by some de-trending procedure. Conversely, if $v_t \sim I(1)$, then simply de-trending will leave us with a nonstationary component. Also, applying de-trending methods (parametric and non-parametric) to integrated processes are generally subject to inducing spurious cyclical behavior (see e.g. Harvey and Jaeger, 1993; Cogley and Nason, 1995, and references therein).

When independent nonstationary series are regressed on each other we will typically find nonsense correlations, i.e. that t-tests suggest that there is a significant statistical relationship between these variables when in fact there is not. This is called spurious regression and was found by simulation in Granger and Newbold (1974) and then theorized by, among others, Nelson and Kang (1984), Phillips (1986, 1998), Durlauf and Phillips (1988) and Marmol (1995).[5] However, two or more nonstationary series may share stochastic trends (hence be dependent) such that a linear combination of them is stationary. The series are then cointegrated, where the cointegrating relationship represents a long-run steady-state solution and is closely related to the principles of economic equilibria (see e.g. Banerjee, Dolado, Galbraith, and Hendry, 1993). In the sense of Engle and Granger (1987), two series integrated of order 1, say $x_{1,t} \sim I(1)$ and $x_{2,t} \sim I(1)$, are cointegrated if for some β, $z_t = x_{1,t} - \beta x_{2,t}$ is stationary, $z_t \sim I(0)$. If such a relationship exists, then any divergence of the two integrated variables from this steady-state will be stochastically bounded. This is appealing, because economic theory often predicts that even though certain variables drift apart in the short run, market mechanisms and policy actions typically bring them together again. Examples of cointegration can be found in the relationship between relative prices and nominal exchange rates, between disposable income and consumption, between money velocity and interest rates etc. (see e.g. Engle and Granger, 1991; Hatanaka, 1996; Juselius, 2006). Finding cointegrating relationships may significantly improve forecasting, and allows for hypothesis testing within error-correction models, which, as optional representations to the AR and MA representations, capture both short-run and long-run dynamics.

Unit root tests play an important role in the analysis of cointegration, because the existence of a cointegrating relationship within a set of variables requires at least two of the variables to be $I(1)$ (excluding the trivial linear combination of I(0) variables). Conventionally, the individual series are pretested for unit roots. Also, a basic cointegration tests may be executed by testing the residuals from an ordinary least squares (OLS) regression for a unit root (residual based cointegration tests), where the null hypothesis "unit

[5] Though Yule (1926) described this phenomenon long before. A nice review on spurious regression is provided by Ventosa-Santaulària (2009).

root" would be equivalent with "no cointegration". The OLS estimator of β is in fact super-consistent under the alternative of "cointegration" as shown by Stock (1987), meaning that $|\widehat{\beta}-\beta| = O_p(T^{-1})$ rather than the usual rate $O_p(T^{-1/2})$. However, due to the spurious regression under the null hypothesis, the critical values have to be adjusted (see e.g. Phillips and Ouliaris, 1990).

Classical unit-root tests are the Augmented Dickey-Fuller (ADF) tests developed primarily in a series of papers by Dickey and Fuller (1979, 1981) and Said and Dickey (1984), which are based on coefficient estimates and corresponding t-tests from an OLS regression. However, these tests suffer from low power, especially in finite samples and close to the unit root. Several extensions and alternatives to the Dickey-Fuller tests have been proposed, notably Sargan and Bhargava (1983), Solo (1984), Said and Dickey (1985), Bhargava (1986), Phillips (1987), Phillips and Perron (1988), Perron (1989), Zivot and Andrews (1992), Schmidt and Phillips (1992), Robinson (1994), Ng and Perron (1995, 2001), Perron and Ng (1996), Elliot, Rothenberg, and Stock (1996), Lumsdaine and Papell (1997), Park (2003), Carrion-I-Silvestre and Sansó (2006), and Kim and Perron (2009).

The *vector* autoregressive process is a common multivariate extension to the univariate autoregressive process. It has been extensively used for forecasting and hypotheses testing in macroeconomics ever since Sims (1980) introduced it in econometrics (see e.g. Stock and Watson, 2001). Let $\mathbf{x}_t = (x_{1,t}, x_{2,t}, \dots, x_{p,t})'$ be a $p \times 1$ vector of variables (these could be GDP, inflation, interest rates etc.). The vector autoregressive model of order k, VAR(k), with fixed starting values $\mathbf{x}_0, \mathbf{x}_{-1}, \dots, \mathbf{x}_{-k+1}$, is (excluding deterministic terms)

$$\mathbf{x}_t = \mathbf{P}_1\mathbf{x}_{t-1} + \mathbf{P}_2\mathbf{x}_{t-1} + \cdots + \mathbf{P}_k\mathbf{x}_{t-1} + \varepsilon_t,$$

where $\mathbf{P}_1, \mathbf{P}_2, \dots, \mathbf{P}_p$ are $p \times p$ matrices of autoregressive coefficients and $\varepsilon_t = (\varepsilon_{1,t}, \varepsilon_{2,t}, \dots, \varepsilon_{p,t})'$ is a $p \times 1$ vector of iid errors, with $E(\varepsilon_t) = \mathbf{0}$ and $E(\varepsilon_t\varepsilon_s') = \mathbf{\Sigma}_\varepsilon > 0$ if $t = s$ and $\mathbf{0}$ if $t \neq s$. The VAR(k) is stationary if, for a suitable choice of starting values, the roots of $|\mathbf{A}(z)| = 0$ lie outside the unit circle, where $\mathbf{A}(z) = \mathbf{I}_p - \sum_{m=1}^{k} \mathbf{P}_m z^m$ is the characteristic polynomial. If $z = 1$ is a root, then $\mathbf{x}_t$ has a unit root, meaning that at least one variable in $\mathbf{x}_t$ is integrated. The VAR(1) model is of some interest later on in this summary. It has a unit root if $|\mathbf{A}(1)| = |\mathbf{I}_p - \mathbf{P}_1| = 0$. For the VAR(1) case this is equivalent to that $\mathbf{P}_1$ has at least one eigenvalue that is equal to 1. Taking differences yields

$$\Delta\mathbf{x}_t = \mathbf{\Gamma}\mathbf{x}_{t-1} + \varepsilon_t, \tag{4.6}$$

where $\mathbf{\Gamma} = \mathbf{P}_1 - \mathbf{I}_p = -\mathbf{A}(1)$. If $\mathbf{x}_t \sim I(1)$, then $\Delta\mathbf{x}_t \sim I(0)$, implying that also the right side of equation (4.6) must be stationary. In this case, (4.6) is the error-correction representation where $\mathbf{\Gamma}$ is the error-correction term.[6]

[6]For VARs of higher order, or including deterministic components, there will be additional terms in the error-correction representation.

The analysis of interest lies in the rank of $\mathbf{\Gamma}$. Let $q = \text{rk}(\mathbf{\Gamma})$. If $q = 0$, then $\mathbf{\Gamma} = \mathbf{0}$, and the elements of $\mathbf{x}_t$ are independent $I(1)$ processes because $\Delta\mathbf{x}_t = \varepsilon_t$ is white noise. If $q = p$, i.e. $\mathbf{\Gamma}$ has full rank, then we have that $|\mathbf{\Gamma}| = |-\mathbf{A}(1)| \neq 0 \Rightarrow |\mathbf{A}(1)| \neq 0$, and $\mathbf{x}_t$ is stationary (assuming no explosive roots). Cointegration exists in between these two cases. If $1 \leq q < p$, then $\mathbf{\Gamma}$ may be decomposed as $\mathbf{\Gamma} = \boldsymbol{\alpha}\boldsymbol{\beta}'$, where $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ are both full rank $p \times q$ matrices. In this case $\mathbf{x}_t$ is $I(1)$, and there are q linear combinations that are stationary, provided by $\boldsymbol{\beta}'\mathbf{x}_t \sim I(0)$. Thus $\boldsymbol{\beta}$ holds the q linearly independent cointegrating vectors as columns. Also, $\boldsymbol{\alpha}$ is often called the adjustment matrix, because it specifies the speed by which the changes in $\mathbf{x}_t$ adjust to the last periods equilibrium error. Conveniently, if $\mathbf{\Gamma}$ has reduced rank, then the number of cointegrating vectors, q, plus the number of stochastic trends, say h, equals the number of $I(1)$ variables in $\mathbf{x}_t$, $p = q + h$. A popular approach for cointegration analysis in VAR models is that of Johansen (1995), who provides likelihood based methods to estimate the spaces spanned by $\boldsymbol{\beta}$ and $\boldsymbol{\alpha}$. Another popular procedure, which will be referred to in Section 4.3.3, is that of Stock and Watson (1988), who derive statistics to decide the number of stochastic trends. Suppose $\mathbf{x_t}$ is fully integrated, i.e. $x_{i,t} \sim I(1)$ for all $i = 1, 2, \ldots, p$. Then the change in $\mathbf{x}_t$ can also be given an MA(∞) representation, $\Delta\mathbf{x}_t = \mathbf{C}(L)\varepsilon_t$, where $\mathbf{C}(L) = \sum_{j=1}^{\infty} \mathbf{C}_j L^j$ with the sequence $\{\mathbf{C}_j\}_0^{\infty}$ absolutely summable and $\mathbf{C}(0) = \mathbf{I}_p$. By recursive substitution in the MA representation we may find the representation

$$\mathbf{x}_t = \mathbf{x}_0 + \mathbf{C}(1)\boldsymbol{\omega}_t + \mathbf{D}(L)\varepsilon_t, \tag{4.7}$$

where $\mathbf{D}(L) = (1-L)^{-1}[\mathbf{C}(L) - \mathbf{C}(1)]$, and where $\boldsymbol{\omega}_t = \sum_{s=1}^{t} \varepsilon_s$ is a vector of (independent) random walks and therefore constitutes the vector of stochastic trends.[7] If $\boldsymbol{\beta}'\mathbf{x}_t \sim I(0)$, then we must have that $\boldsymbol{\beta}'\mathbf{C}(1) = 0$, implying that every cointegrating vector lies in the left null space of $\mathbf{C}(1)$, and that $\mathbf{C}(1)$ has reduced rank, $\text{rk}(\mathbf{C}(1)) = p - q = h$, the number of stochastic trends. Under the null hypothesis $H_0 : h = m$ versus the alternative hypothesis $H_1 : h < m$ for some m, Stock and Watson (1988) provide procedures to consistently estimate a matrix $\widetilde{\boldsymbol{\beta}}$ with the properties $\widetilde{\boldsymbol{\beta}}'\boldsymbol{\beta} = \mathbf{0}$ and $\widetilde{\boldsymbol{\beta}}'\widetilde{\boldsymbol{\beta}} = \mathbf{I}_m$, and propose to regress the transformation $\mathbf{W}_t = \widetilde{\boldsymbol{\beta}}'\mathbf{x}_t = \widetilde{\boldsymbol{\beta}}'\mathbf{x}_0 + \widetilde{\boldsymbol{\beta}}'\mathbf{C}(1)\boldsymbol{\omega}_t + \widetilde{\boldsymbol{\beta}}'\mathbf{D}(L)\varepsilon_t$ onto its own lagged value $\mathbf{W}_{t-1}$. If the autocorrelation from $\mathbf{D}(L)\varepsilon_t$ is controlled for, then under the null hypothesis the regression coefficient matrix should asymptotically have m unit eigenvalues, while under the alternative it will have less than m unit eigenvalues.

[7]The representation (4.7) is sometimes called a multivariate Beveridge-Nelson decomposition due to a similar univariate decomposition by Beveridge and Nelson (1981).

4.2 Panel unit root tests

A panel time series model is essentially one or more variables measured over time for different individuals (e.g. countries, regions, firms or industries). The main motivation for constructing a panel unit root test is to pool the information of the individuals (e.g. GDP measured over time for different countries) and thereby increase power against a stationary alternative. Let $i = 1, 2, \ldots, N$ denote the number of individuals and $t = 1, 2, \ldots, T$ the number of time points. The panel autoregressive model of order k has the form

$$x_{i,t} = \rho_{i,1}x_{i,t-1} + \rho_{i,2}x_{i,t-2} + \cdots + \rho_{i,k}x_{i,t-k} + \varepsilon_{i,t}, \tag{4.8}$$

where $\rho_{i,1}, \rho_{i,2}, \ldots, \rho_{i,k}$ are the autoregressive coefficients for the ith individual and $\varepsilon_{i,t}$ is iid. As with the simple time series case we may include deterministic trends.

Panel unit root tests are broadly divided into homogenous and heterogenous tests (see e.g. Banerjee, 1999; Breitung and Pesaran, 2008, for reviews). Consider, for simplicity, the AR(1) panel model with a homogenous autoregressive parameter, $x_{i,t} = \rho x_{i,t-1} + \varepsilon_{i,t}$. For suitable choices of starting values $\{x_{i,0}, i = 1, 2, \ldots, N\}$, the panel is then stationary if $|\rho| < 1$. In principle, a homogenous panel unit root test (e.g. Harris and Tzavalis, 1999; Levin, Lin, and Chu, 2002; Herwartz and Siedenburg, 2008; Lopez, 2009) would be constructed around the hypotheses

$$\begin{aligned} &H_0 : \rho = 1, \\ &H_1 : \rho < 1. \end{aligned}$$

That is, to test that every panel individual has a unit root versus that all individuals are stationary with the same autoregressive estimator. Consider next the analogous AR(1) panel model allowing for autoregressive heterogeneity, $x_{i,t} = \rho_i x_{i,t-1} + \varepsilon_{i,t}$. If the panel individuals are independent, then this is just N different univariate time series, albeit on the same variable of interest where the conditions for stationarity on the ith series coincide with those in Section 4.1. However, the set of variables $\{x_{i,t}, i = 1, 2, \ldots, N\}$ will be nonstationary if the conditions are broken for any of the individual series. Let $\mathbb{W}$ denote the set $\{1,2,...,N\}$ and let $\mathbb{M} \subseteq \mathbb{W}$ denote the set $\{1, 2, \ldots, M\}$ where $M = m(N)$ is some increasing function of N satisfying $\lim_{N\to\infty} \frac{m(N)}{N} = \kappa$, $0 < \kappa \leq 1$. A heterogenous test would, in principle, be constructed to test $H_0 : \forall i \in \mathbb{W}, \rho_i = 1$, i.e. the same null hypothesis as for a homogenous test, where at least the following alternative hypotheses may be encountered in the panel unit root literature:

$$\begin{aligned} &H_{1a} : \forall i \in \mathbb{W}, \rho_i < 1, \\ &H_{1b} : \exists i \in \mathbb{W}, \rho_i < 1, \\ &H_{1c} : \forall i \in \mathbb{M}, \rho_i < 1. \end{aligned}$$

That is, the alternative hypothesis is specified to allow the autoregressive coefficients to be different. The first alternative would under rejection of the null hypothesis lead to the conclusion that all individuals are stationary. This type of alternative is specified by e.g. Maddala and Wu (1999). Analogously, the second alternative would lead to the conclusion that at least one of the individuals is stationary (e.g. Chang, 2004). However, as argued by Pesaran (2012), this type of alternative will only make sense for a fixed number of panel individuals, because tests adopting H_{1b} will be subject to a loss of power as $N \to \infty$. It may be more natural to adopt the third alternative (e.g. Im, Pesaran, and Shin, 2003), and perhaps also try to estimate the proportion of stationary individuals (see e.g. Smeekes, 2010). For most panel unit root tests it seems these matters only concern the interpretation of a rejection of the null hypothesis rather than the mathematical execution. For instance, a homogenous test will generally have power also against a heterogenous alternative even when only a fraction of the individuals are stationary. However, just as much as this property has been highlighted it has also been questioned if it is desirable for a homogenous test. For discussions on homogeneity contra heterogeneity, see e.g. Karlsson and Löthgren (2000), Phillips and Sul (2003), Jönsson (2005), Pesaran (2012), and Westerlund and Breitung (2013).

The technical distinction between homogeneous and heterogenous tests lies in the way information is pooled. Homogenous tests typically exploit the procedures available in the dynamic panel literature around a pooled homogenous ρ (see e.g. Arellano, 2003; Baltagi, 2008), and then build statistics from these procedures. Heterogenous tests generally rely on individual unit root tests, such as ADF, and then base the statistics on pooling over N. For example, Im, Pesaran, and Shin (2003) propose to estimate N individual ADF t-statistics and then compute the average of these, and Maddala and Wu (1999) and Choi (2001) propose to pool the p-values of individual unit root tests in the sense of Fisher (1932). Appropriately standardized, these tests will converge to standard normal distributions as both T and N tend to infinity. Here there are mainly two ways to derive the asymptotic distribution under the null hypothesis (see Phillips and Moon, 1999, 2000). We may use *sequential limits*, where one index, perhaps N, is fixed, and the other, in this case T, is passed to infinity to arrive at an intermediate limit. By letting subsequently N pass to infinity, we find a sequential limit. Here we denote this approach $(T,N)_s \to \infty$.

A more robust approach is a *joint limit*, where we let N and T pass to infinity simultaneously, perhaps with some restriction, such as $N/T \to 0$, or with some specified functional relation $T = f(N)$. Here we denote joint limit $N,T \to \infty$. When studying the power performance of panel unit root tests, analytically or by simulation, it is conventional to look at power under a sequence of local alternatives as specified in (3.2). For panels this usually takes the form $\rho_i = 1 - \frac{c}{T\sqrt{N}}$, where c is some constant.

What is now called the first generation panel unit root tests assume that the panel individuals are mutually independent, implying in the model (4.8) that, for $\varepsilon_t = (\varepsilon_{1,t}, \varepsilon_{2,t}, \ldots, \varepsilon_{N,t})'$, $Var(\varepsilon_t)$ is diagonal. This has been deemed unrealistic for most panels of economic data. For instance, nearby countries are likely to have economies that are interrelated. A second generation of tests has therefore focused on overcoming this issue, where a popular approach is to assume an error factor structure. This is done by e.g. Phillips and Sul (2003), Moon and Perron (2004) and Pesaran (2007), such that, in principle, the unobservable error term follows a dynamic factor structure

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

where, for $j = 1, 2, \ldots, r$ and $i = 1, 2, \ldots, N$, $f_{j,t}$ are unobservable dynamic factors, $\lambda_{i,j}$ are the corresponding factor loadings and $v_{i,t}$ are unobservable dynamic idiosyncratic (individual-specific) components. Because the panel individuals share the latent factors, $x_{i,t}$ is cross-sectionally dependent. This imposes what Chudik, Pesaran, and Tosetti (2011) define as a *strong* cross-sectional dependence since the largest eigenvalue of the error covariance matrix will diverge to infinity as $N \to \infty$.

A related issue, which is also the subject of this thesis, is to study the dynamic properties of the factor model itself. For instance Bai and Ng (2004, 2010) and Sul (2009) study raw factor models and derive unit root tests for the factors and the idiosyncratic components. We cover these topics in more detail in the following sections. Other second generation panel unit root tests, related to what we have discussed here, may be found in, e.g., Choi (2002), Chang (2002, 2004), Breitung and Das (2005), Shin and Kang (2006), Cerrato and Sarantis (2007), Chang and song (2009), Wang, Wang, Yang, and Li (2010), Shin and Park (2010), and Palm, Smeekes, and Urbain (2011).

4.3 Factor models

Factor models have a long history in the statistical science and were originally developed as means to measure unobserved effects in psychology. In the classical factor model, p observable variables $\mathbf{x} = (x_1, x_2, \ldots, x_p)'$ are determined by a set of linear relationships through q unobservable common factors, $\mathbf{f} = (f_1, f_2, \ldots, f_q)'$, and p unobservable specific factors (idiosyncratic disturbances), $\mathbf{u} = (u_1, u_2, \ldots, u_p)'$, such that

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

where $\boldsymbol{\Lambda}$ is a $p \times q$ matrix of coefficients (factor loadings). The standard assumptions are that $\mathbf{f}$ and $\mathbf{u}$ are mutually independent with $E(\mathbf{f}) = \mathbf{0}$ and $E(\mathbf{u}) = \mathbf{0}$, in which case

$$\boldsymbol{\Sigma} = Var(\mathbf{x}) = \boldsymbol{\Lambda}\boldsymbol{\Theta}\boldsymbol{\Lambda}' + \boldsymbol{\Upsilon}, \tag{4.9}$$

where $\boldsymbol{\Theta} = Var(\mathbf{f})$ and $\boldsymbol{\Upsilon} = Var(\mathbf{u})$. In exploratory factor analysis $\boldsymbol{\Lambda}$ is unrestricted (a priori) while $\boldsymbol{\Theta}$ is normalized, $\boldsymbol{\Theta} = \mathbf{I}_q$, and $\boldsymbol{\Upsilon}$ is assumed diagonal.

For any given $\mathbf{\Sigma}$ and $\mathbf{\Upsilon}$ the factor loadings are then only uniquely defined up to postmultiplication by an orthogonal matrix $\mathbf{G}$, since $(\mathbf{\Lambda G})(\mathbf{\Lambda G})' + \mathbf{\Upsilon} = \mathbf{\Lambda \Lambda}' + \mathbf{\Upsilon}$. This is usually solved by imposing arbitrary identifying linear restrictions on $\mathbf{\Lambda}$ (see e.g. Jöreskog, 1967). In confirmatory factor analysis $\mathbf{\Lambda}$ and $\mathbf{\Theta}$ are subject to prior linear restrictions (see e.g. Jöreskog, 1969, and Bekker, 1989).

4.3.1 Dynamic factor models

The idea that the behavior of a set of potentially many variables is driven by a relatively small set of common factors and some remaining individual-specific shocks, is also attractive in economic modeling with time series. For instance, in macroeconomics nearby countries of similar size are likely to be affected in similar ways from (unobservable) economic shocks appearing over time. Geweke (1977) therefore considered the one-factor multivariate time series model

$$\mathbf{x}_t = (x_{1,t}, x_{2,t}, \ldots, x_{N,t})' = \mathbf{\Lambda} f_t + \mathbf{u}_t, \tag{4.10}$$

where $\mathbf{u}_t = (u_{1,t}, u_{2,t}, \ldots, u_{N,t})'$ are zero-mean time-dependent (dynamic) idiosyncratic components, f_t is a single zero-mean dynamic factor and $\mathbf{\Lambda}$ is an $N \times 1$ matrix of factor loadings. Here the time index suggests that we need to make assumptions of how the time series behave over time. Assuming mutually independent and covariance stationary components, i.e. $E(f_t u_{i,\tau}) = 0$ for all i, t, τ, and $Cov(f_t, f_{t+s})$ and $Cov(u_{i,t}, u_{i,t+s})$ only depend on s, and using the Wold decomposition theorem (see note 2 of this chapter), the model (4.10) may be given the dynamic representation

$$\mathbf{x}_t = \mathbf{a}(L) w_t + \mathbf{B}(L) \mathbf{e}_t, \tag{4.11}$$

where $\mathbf{a}(L)$ is a finite lag polynomial vector, w_t is a scalar white noise process, $\mathbf{B}(L)$ is a finite diagonal lag polynomial matrix, and $\mathbf{e}_t$ is a vector white noise process. Geweke (1977) then showed that exploratory likelihood based inference, similar to Jöreskog (1967), can be made in the frequency domain by the use of Fourier transforms. Parallel, Sargent and Sims (1977) considered the multi-factor version of (4.11) and, subsequently, Geweke and Singleton (1981) considered confirmatory dynamic factor analysis. The dynamic factor model has since then been developed under many different assumptions and restrictions, e.g. by Chamberlain (1983), Chamberlain and Rothschild (1983), Molenaar (1985), Connor and Korajcyk (1986), Peña and Box (1987), Molenaar, Goijer, and Schmitz (1992) and Forni, Hallin, Lippi, and Reichlin (2000, 2004).[8]

Forni, Hallin, Lippi, and Reichlin (2000) state the generalized dynamic factor model,

$$\mathbf{x}_t = \mathbf{A}(L) \boldsymbol{w}_t + \boldsymbol{\nu}_t, \tag{4.12}$$

[8] See also Peña and Poncela (2004) for an overview.

where $\mathbf{A}(L)$ is a lag polynomial matrix with square summable coefficients, $\boldsymbol{w}_t = (w_{1,t}, w_{2,t}, \ldots, w_{q,t})'$ is a q-dimensional orthonormal white noise vector, where $w_{j,t}$ $(j = 1, 2, \ldots, q)$ are usually referred to as common shocks, and $\boldsymbol{\nu}_t = (\nu_{1,t}, \nu_{2,t}, \ldots, \nu_{N,t})'$ is a zero-mean stationary vector process. Their model is generalized in the sense that it incorporates many of the preceding models as special cases. They too consider estimation in the frequency domain with the use of spectral analysis.

Conversely, working from (4.12), if the lags are finite, then the model can be given the representation (4.10) (but with potentially several factors). This is then coined the *static* representation, in the sense that the loadings only have contemporaneous effects. However, the factors may still be dynamic, e.g. AR processes, which would then add a dynamic structure to the static model. It can be shown (see e.g. Bai and Ng, 2008) that if the generalized dynamic factor model (4.12) with s lags has q common shocks, then the static representation has $r = q(s+1)$ factors. In general the dynamic factor model and the static factor model produce quite similar forecasts, but the static representation is often used for estimation purposes, because it allows for estimation in the time domain instead of the more complicated frequency domain.

4.3.2 The static representation

The static factor model (or static representation) with dynamic factors and dynamic idiosyncratic components is the focus of this thesis, so we go through it in some more detail. In its simplest form, without any deterministic components, it is

$$\mathbf{x}_t = \boldsymbol{\Lambda}\mathbf{f}_t + \mathbf{u}_t, \tag{4.13}$$

where $\mathbf{f}_t = (f_{1,t}, f_{2,t}, \ldots, f_{r,t})'$ is an $r \times 1$ vector of unobservable dynamic factors, $\mathbf{u}_t = (u_{1,t}, u_{2,t}, \ldots, u_{N,t})'$ is an $N \times 1$ vector of unobservable dynamic idiosyncratic components and $\boldsymbol{\Lambda} = (\boldsymbol{\lambda}_1, \boldsymbol{\lambda}_2, \ldots, \boldsymbol{\lambda}_N)'$ is the $N \times r$ matrix of static factor loadings, where $\boldsymbol{\lambda}_i = (\lambda_{i,1}, \lambda_{i,2}, \ldots, \lambda_{i,r})'$.

As with the classical factor model the usual assumption is that $\mathbf{f}_t$ and $\mathbf{u}_t$ are independently distributed, in which case the contemporaneous covariances are $\boldsymbol{\Sigma}_{xx} = Var(\mathbf{x}_t) = \boldsymbol{\Lambda}\boldsymbol{\Sigma}_{ff}\boldsymbol{\Lambda}' + \boldsymbol{\Sigma}_{uu}$, where $\boldsymbol{\Sigma}_{ff} = Var(\mathbf{f}_t)$ and $\boldsymbol{\Sigma}_{uu} = Var(\mathbf{u}_t)$. Without loss of generality we may assume that $\boldsymbol{\Sigma}_{ff} = \mathbf{I}_r$, because unless we impose additional restrictions, the factors and the factor loadings are not separately identified. Also, as $N \to \infty$ the covariance structure is not identified unless we impose some restrictions on $\boldsymbol{\Sigma}_{xx}$ such that the covariances coming from the common factors are distinguished from the idiosyncratic covariances. The typical assumptions are; (a) that the factors are pervasive, in the sense that as N grows large, $N^{-1}\boldsymbol{\Lambda}'\boldsymbol{\Lambda}$ converges to some positive definite matrix implying that the eigenvalues of $\boldsymbol{\Lambda}'\boldsymbol{\Lambda}$ (likewise the non-zero eigenvalues of $\boldsymbol{\Lambda}\boldsymbol{\Lambda}'$) are $O(N)$; and (b) that the eigenvalues of $\boldsymbol{\Sigma}_{uu}$ are uniformly bounded. This is the essence of the *approximate* factor model by Chamberlain and Rothschild

(1983), which allows the idiosyncratic components to be stationary dynamic processes. If $\mathbf{\Sigma}_{uu}$ is restricted to be diagonal, then we have the *exact* (or *strict*) factor model. Hence, the exact factor model is nested within the approximate factor model (see e.g. Bai, 2003).

The standard procedure to estimate the factors and the factor loadings is by the method of principal components (PC). Let $\mathbf{X} = (\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_T)$ be the $N \times T$ matrix of the observed panel data, $\mathbf{F} = (\mathbf{f}_1, \mathbf{f}_2, \ldots, \mathbf{f}_T)$ be the $r \times T$ matrix of latent factors, and $\mathbf{\Lambda}$ be defined as before. The principal components estimators are the arguments that minimize the total sum of squares,

$$\left\{\widehat{\mathbf{F}}, \widehat{\mathbf{\Lambda}}\right\}_{PC} = \underset{\mathbf{F},\mathbf{\Lambda}}{\arg\min}\, \mathrm{tr}[(\mathbf{X} - \mathbf{\Lambda F})(\mathbf{X} - \mathbf{\Lambda F})'], \tag{4.14}$$

subject to the constraint $T^{-1}\mathbf{F}\mathbf{F}' = \mathbf{I}_r$. The solutions are given by $\widehat{\mathbf{F}}_{PC} = T^{1/2}\widehat{\mathbf{W}}'$ and $\widehat{\mathbf{\Lambda}}_{PC} = T^{-1/2}\mathbf{X}\widehat{\mathbf{W}}$, where $\widehat{\mathbf{W}}$ is the $T \times r$ matrix with columns consisting of the orthonormal eigenvectors associated with the r largest eigenvalues of the $T \times T$ matrix $\mathbf{X}'\mathbf{X}$ (see e.g. Stock and Watson, 2002; Bai, 2003; Breitung and Tenhofen, 2011). Bai (2003) shows that $\mathbf{f}_t$ and $\mathbf{\Lambda}$ can be consistently estimated by principal components, up to some invertible $r \times r$ transformation, such that for all t, $\widehat{\mathbf{f}}_{t,PC} \xrightarrow{p} \mathbf{H}'\mathbf{f}_t$ as $N, T \to \infty$, if $\sqrt{N}/T \to 0$, and for all i, $\widehat{\boldsymbol{\lambda}}_{i,PC} \xrightarrow{p} \mathbf{H}^{-1}\boldsymbol{\lambda}_i$ as $N, T \to \infty$, if $\sqrt{T}/N \to 0$, for some invertible matrix $\mathbf{H}$. A stronger result is that the common component $c_{i,t} = \mathbf{f}_t'\boldsymbol{\lambda}_i$ is consistently estimated by $\widehat{c}_{i,t,PC} = \widehat{\mathbf{f}}'_{t,PC}\widehat{\boldsymbol{\lambda}}_{i,PC}$, as $N, T \to \infty$ without any further restrictions on the relationship between N and T.

We may also consider maximum likelihood estimation, though full maximum likelihood estimation has long been seen as intractable and computationally infeasible due to the large number of parameters. Assuming $\mathbf{f}_t$ is iid $N(\mathbf{0}, \mathbf{I}_r)$ and $\mathbf{u}_t$ is iid $N(\mathbf{0}, \mathbf{\Sigma}_{\mathbf{uu}})$, the log-likelihood is

$$l\left(\boldsymbol{\theta}|\mathbf{X}\right) = -\frac{NT}{2}\log 2\pi - \frac{T}{2}\log|\mathbf{\Sigma}_{xx}| - \frac{T}{2}\,\mathrm{tr}\left(\mathbf{\Sigma}_{xx}^{-1}\mathbf{S}\right), \tag{4.15}$$

where $\mathbf{\Sigma}_{xx}$ was defined before and $\mathbf{S} = T^{-1}\mathbf{X}\mathbf{X}'$ is the $N \times N$ sample covariance matrix. A special case which is of interest in this thesis is when $\mathbf{\Sigma}_{uu} = \sigma^2\mathbf{I}_N$ (the exact factor model with spherical noise), in which case the factor model breaks down to the classical factor model with covariance of the form (4.9). It is well-known (see e.g. Anderson, 2003, p. 69) that (4.15) maximizes when $\widehat{\mathbf{\Sigma}}_{xx} = \mathbf{S}$, which under the normalization that $\mathbf{\Lambda}'\mathbf{\Lambda}$ is a diagonal matrix yields the explicit solutions

$$\widehat{\sigma}^2_{ML} = \tfrac{1}{N}\,\mathrm{tr}\left(\mathbf{S} - \widehat{\mathbf{\Lambda}}_{ML}\widehat{\mathbf{\Lambda}}'_{ML}\right), \tag{4.16}$$

$$\widehat{\mathbf{\Lambda}}_{ML} = \widehat{\mathbf{A}}\left(\widehat{\mathbf{\Phi}} - \widehat{\sigma}^2_{ML}\mathbf{I}_r\right)^{1/2}, \tag{4.17}$$

$$\widehat{\mathbf{F}}_{ML} = \mathbf{X}\widehat{\mathbf{A}}\left(\widehat{\mathbf{\Phi}} - \widehat{\sigma}^2_{ML}\mathbf{I}_r\right)^{1/2}\widehat{\mathbf{\Phi}}^{-1},$$

where $\widehat{\mathbf{\Phi}} = \text{diag}\left[\widehat{\varphi}_1(\mathbf{S}), \widehat{\varphi}_2(\mathbf{S}), \ldots, \widehat{\varphi}_r(\mathbf{S})\right]$ are the r largest eigenvalues of $\mathbf{S}$ and $\widehat{\mathbf{A}}$ are the associated normalized eigenvectors, implying $\widehat{\mathbf{A}}'\widehat{\mathbf{A}} = \mathbf{I}_r$ (see e.g. Doz, Giannone, and Reichlin, 2012; Stoica and Jansson, 2009). Also, it is well-known (see e.g. Fernández-Macho, 2000) that the static representation has a state space representation that allows for maximum likelihood estimation to be achieved based on iterative procedures from the Kalman filter, which may be executed using the EM algorithm. This is necessary if we do not have spherical noise in the idiosyncratic component, i.e. if $\mathbf{\Sigma}_{uu} \neq \sigma^2 \mathbf{I}_N$. The EM algorithm is carried out by alternating the following two steps until convergence:

1 Expectation (E) step: the conditional expectation of the log-likelihood is calculated using the previous iteration, $\boldsymbol{\theta}(j)$:

$$l^*(\boldsymbol{\theta}, \boldsymbol{\theta}(j)) = E\left[l(\mathbf{X}|\boldsymbol{\theta})|\boldsymbol{\theta}(j)\right].$$

2 Maximization (M) step: the parameters are estimated again by maximizing the expected log-likelihood:

$$\boldsymbol{\theta}(j+1) = \arg\max_{\boldsymbol{\theta}} l^*(\boldsymbol{\theta}, \boldsymbol{\theta}(j)).$$

The iteration needs to be initiated to find $\boldsymbol{\theta}(1)$. For dynamic factor models, the typical choice of initial estimation is the PC estimator (4.14). The iteration is repeated until $||\boldsymbol{\theta}(j+1) - \boldsymbol{\theta}(j)||$ reaches some sufficiently small tolerance level.

Recently, Doz, Giannone, and Reichlin (2012) have considered (quasi) maximum likelihood estimation in a misspecified exact factor model, where the true model is the approximate factor model. That is, they falsely treat the idiosyncratic components as orthogonal. Assuming that the idiosyncratic components are iid and that the factors follow a stationary VAR model, they show that the factors are consistently estimated with quasi-MLE up to an invertible transformation as $N, T \to \infty$. Bai and Li (2012b,a) extend the results in Doz, Giannone, and Reichlin (2012) by considering a more general quasi maximum likelihood approach. Let $\mathbf{\Phi}_t = E(\mathbf{u}_t \mathbf{u}_t')$, which allows for time heteroscedasticity. Also, suppose the factors are fixed constants and let $\mathbf{M}_{ff} = \frac{1}{T-1}\sum_{t=1}^{T}(\mathbf{f}_t - \bar{\mathbf{f}}_t)(\mathbf{f}_t - \bar{\mathbf{f}}_t)'$, where $\bar{\mathbf{f}}_t = T^{-1}\sum_{t=1}^{T}\mathbf{f}_t$ and suppose that $\lim_{T\to\infty} \mathbf{M}_{ff}$ exists and is positive definite. They consider maximizing the objective function

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

where $\mathbf{\Sigma}_{zz} = \mathbf{\Lambda}\mathbf{M}_{ff}\mathbf{\Lambda}' + \widetilde{\mathbf{\Phi}}$, where $\widetilde{\mathbf{\Phi}}$ is a diagonal matrix holding the diagonal elements of $T^{-1}\sum_{t=1}^{T}\mathbf{\Phi}_t$. Bai and Li (2012a) show that, under quite general forms of idiosyncratic autocorrelation, time heteroscedasticity and cross-sectional correlations (admitting an approximate factor model), the quasi-MLEs are consistent and have the following convergence rates:

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

which are faster than the PC estimators. If the exact factor model is the true factor model, then the convergence rates are $O_p(T^{-1})$ as shown by Bai and Li (2012b), i.e. the $O_p(N^{-2})$ term then falls out. The results hold also if the factors have arbitrary dynamic properties with $E||\mathbf{f}_t||^4<\infty$.

Similar setups have been considered by e.g. Engle and Watson (1981), Watson and Engle (1983), Stock and Watson (1989), Camba-Mendez, Kapetanios, Smith, and Weale (2001), and Jungbacker and Koopman (2008). However, they consider only estimation and forecasting in stationary dynamic factor models. To our knowledge unit root testing in the static factor model, exact or approximate, based on maximum likelihood or quasi-maximum likelihood have not been pursued.

4.3.3 The Bai and Ng procedure

The main motivation to assume a factor structure in panel data may have been to control for a cross-sectional dependence. However, Bai and Ng (2004, 2010) point out that the non-stationarity in the model might also arise from stochastic trends in the factors. They assume the following model (excluding deterministic trends)

$$\begin{aligned}
x_{i,t} &= \mu_i+\boldsymbol{\lambda}_i'\mathbf{f}_t+u_{i,t}, \qquad (4.18)\\
(1-L)\mathbf{f}_t &= \mathbf{C}(L)\mathbf{v}_t,\\
u_{i,t} &= \rho_i u_{i,t-1}+\varepsilon_{i,t},
\end{aligned}$$

where $\varepsilon_{i,t}$ is allowed to be autocorrelated and cross-sectionally dependent, $\mathbf{v}_t$ is an $r\times 1$ iid vector, and $\mathbf{C}(L)=\sum_{m=0}^{\infty}\mathbf{C}_m L^m$ is an $r\times r$ lag polynomial matrix, with $\sum_{m=0}^{\infty}m||\mathbf{C}_m||<\infty$ (implies square summability). Under the additional assumption that $\mathrm{rk}(\mathbf{C}(1))=r_1$, where $0\leq r_1\leq r$, this setup allows for r_1 stochastic trends among the factors.

By taking first-differences the constant term in (4.18) is removed, $\Delta x_{i,t}=\boldsymbol{\lambda}_i'\Delta\mathbf{f}_t+\Delta u_{i,t}$, and because $\Delta\mathbf{f}_t=(\Delta f_{1,t},\Delta f_{2,t},\ldots,\Delta f_{r,t})'\sim I(0)$ and $\Delta u_{i,t}\sim I(0)$, the model also becomes stationary and therefore allows for estimation by principal components. For notational convenience, let $y_{i,t}=\Delta x_{i,t}$, $z_{i,t}=\Delta u_{i,t}$, and $\mathbf{g}_t=\Delta\mathbf{f}_t$. We may solve (4.14) with respect to the differenced panel data to find $\{\widehat{\mathbf{g}}_t,\widehat{\boldsymbol{\Lambda}}\}_{PC}$ whereby the idiosyncratic components, in first-differences, are left as residuals, $\widehat{z}_{i,t} = y_{i,t} - \widehat{\boldsymbol{\lambda}}_i' \widehat{\mathbf{g}}_t$.[9] Unit root tests are then performed on the re-accumulated factors and idiosyncratic components, $\widehat{\mathbf{f}}_t = \sum_{s=2}^{T} \widehat{\mathbf{g}}_t$ and $\widehat{u}_{i,t} = \sum_{s=2}^{T} \widehat{z}_{i,t}$, which tests are shown by Bai and Ng (2004) to be asymptotically independent. To test for idiosyncratic unit roots they suggest to pool individual unit root tests and propose the Fisher-type statistic

$$P_{\hat{u}}^{c} = \frac{-2\sum_{i=1}^{N} \log p_{\hat{u}}^{c}(i) - 2N}{\sqrt{4N}} \xrightarrow{d} N(0,1), \text{ as } N \to \infty, \tag{4.19}$$

where $p_{\hat{u}}^{c}(i)$ are p-values of idiosyncratic ADF tests.

To test for non-stationarity in the factors they suggest a univariate ADF test in case $r = 1$. For $r > 1$ they propose a sequential procedure to determine the number of stochastic trends among the factors based on Stock and Watson (1988) (see Section 4.1). The most general statistic is MQ_c^c (Bai and Ng, 2004, p. 1133), which nonparametrically corrects for error autocorrelation of arbitrary form. This statistic is calculated as follows: Start with the hypothesis $H_0 : r_1 = m = r$, where m is the hypothesized number of stochastic trends and r is the number of factors. Demean the estimated factors as $\widehat{\mathbf{f}}_t^c = \widehat{\mathbf{f}}_t - \overline{\widehat{\mathbf{f}}}_t$, where $\overline{\widehat{\mathbf{f}}}_t = (T-1)^{-1} \sum_t^T \widehat{\mathbf{f}}_t$ and let $\widehat{\mathbf{B}}_t^c = \widehat{\boldsymbol{\beta}}_\perp' \widehat{\mathbf{f}}_t^c$, where $\widehat{\boldsymbol{\beta}}_\perp$ are the m eigenvectors associated with the m largest eigenvalues of $T^{-2} \sum_{t=2}^{T} \widehat{\mathbf{f}}_t^c \widehat{\mathbf{f}}_t^{c\prime}$. Next, let $\widehat{e}_t^c$ be the residuals from estimating a VAR(1) model for $\widehat{\mathbf{B}}_t^c$, and calculate

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

where, for $j = 0, 1, \ldots, J$, $K(j) = 1 - j/(J+1)$ are Bartlett Kernel weights with truncation lag J chosen such that $J/\min(\sqrt{N}, \sqrt{T}) \to 0$ as $J, N, T \to \infty$. Finally, let $v_c^c(m)$ be the smallest eigenvalue of

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

and calculate the test-statistic $MQ_c^c(m) = T\left[v_c^c(m) - 1\right]$. The idea here is to use the Kernel estimation to remove the difference between the long-run residual variance and the short-run residual variance, and then test if the eigenvalues of the VAR(1) coefficient matrix are significantly different from unity, suggesting that there are less than m stochastic trends in the space spanned by the factors. The critical values are found from Bai and Ng (2004, Table 1, p. 1136). If $H_0 : r_1 = m$ is rejected, then repeat for $H_0 : r_1 = m - 1$ until the number of stochastic trends is decided.

[9] Before extracting the factors (or factor loadings), the number of factors has to be chosen. Bai and Ng (2002) propose decision criteria that has become very popular. See also Bai (2004), Peña and Poncela (2006), Bai and Ng (2007), Hallin and Liska (2007), Onatski (2010) and Kapetanios (2010).

4.3.4 The Moon and Perron tests

Moon and Perron (2004) consider the model

$$
\begin{aligned}
x_{i,t} &= \mu_i + x_{i,t}^0, \\
x_{i,t}^0 &= \rho_i^0 x_{i,t-1}^0 + \xi_{i,t}, \\
\xi_{i,t} &= \boldsymbol{\lambda}_i^{0\prime} \mathbf{f}_t^0 + u_{i,t}^0,
\end{aligned} \tag{4.20}
$$

with initial condition $x_{i,0}^0 = 0$. They propose two statistics to test $H_0 : \rho_i^0 = 1$ for all i, by projecting the panel data onto the space orthogonal to the factor loadings $\boldsymbol{\Lambda}^0$. In what follows, let $\overline{\mathbf{X}} = (\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_N)$ be the $T \times N$ matrix with columns $\mathbf{x}_i = (x_{i,1}, x_{i,2}, \ldots, x_{i,T})'$, and let $\overline{\mathbf{X}}_{-1} = (\mathbf{x}_{-1,1}, \mathbf{x}_{-1,2}, \ldots, \mathbf{x}_{-1,N})$ be the $T \times N$ matrix with columns $\mathbf{x}_{-1,i} = (x_{i,0}, x_{i,1}, \ldots, x_{i,T-1})'$. Define analogously the $T \times N$ matrix of errors in (4.20) as $\boldsymbol{\Xi} = (\boldsymbol{\xi}_1, \boldsymbol{\xi}_2, \ldots, \boldsymbol{\xi}_N)$, where $\boldsymbol{\xi}_i = (\xi_{i,1}, \xi_{i,2}, \ldots, \xi_{i,T})'$, and consider the estimator $\widehat{\boldsymbol{\Xi}} = \overline{\mathbf{X}} - \widehat{\rho}_{\text{pool}}^0 \overline{\mathbf{X}}_{-1}$, where $\widehat{\rho}_{\text{pool}}^0 = \operatorname{tr}(\overline{\mathbf{X}}_{-1}' \overline{\mathbf{X}}) / \operatorname{tr}(\overline{\mathbf{X}}_{-1}' \overline{\mathbf{X}}_{-1})$ is the pooled autoregressive estimator. Here the errors follow a factor model. As such, we may apply principal components and solve (4.14) with respect to the errors $\widehat{\boldsymbol{\Xi}}$ and estimate the factors $\widehat{\mathbf{f}}_{PC}^0$ and the factor loadings $\widehat{\boldsymbol{\Lambda}}_{PC} = (\widehat{\boldsymbol{\lambda}}_1, \widehat{\boldsymbol{\lambda}}_2, \ldots, \widehat{\boldsymbol{\lambda}}_N)'$. From this, Moon and Perron (2004) define the re-scaled estimated loadings $\widehat{\boldsymbol{\Lambda}} = \widehat{\boldsymbol{\Lambda}}_{PC} (\widehat{\boldsymbol{\Lambda}}_{PC}' \widehat{\boldsymbol{\Lambda}}_{PC})^{1/2} N^{-1/2}$. Next, let $u_{i,t}^0$ have variance $\sigma_{u^0,i}^2$, long-run variance $\omega_{u^0,i}^2$ (see footnote 4 of this chapter), and one-sided long-run variance $\zeta_{u^0,i} = (\omega_{u^0,i}^2 - \sigma_{u^0,i}^2)/2$, and define the average moment estimators

$$
\widehat{\omega}_{u^0}^2 = \tfrac{1}{N} \textstyle\sum_{i=1}^N \widehat{\omega}_{u^0,i}^2, \tag{4.21}
$$

$$
\widehat{\kappa}_{u^0}^4 = \tfrac{1}{N} \textstyle\sum_{i=1}^N \widehat{\omega}_{u^0,i}^4, \tag{4.22}
$$

$$
\widehat{\zeta}_{u^0} = \tfrac{1}{N} \textstyle\sum_{i=1}^N \widehat{\zeta}_{u^0,i}, \tag{4.23}
$$

where $\widehat{\omega}_{u^0,i}^4 = (\widehat{\omega}_{u^0,i}^2)^2$. Also, define the pooled (de-factored and autocorrelation-adjusted) autoregressive estimator

$$
\rho_{\text{pool}}^* = \frac{\operatorname{tr}\left(\overline{\mathbf{X}}_{-1} \mathbf{Q}_{\widehat{\lambda}} \overline{\mathbf{X}}'\right) - NT \widehat{\zeta}_{u^0}}{\operatorname{tr}\left(\overline{\mathbf{X}}_{-1} \mathbf{Q}_{\widehat{\lambda}} \overline{\mathbf{X}}_{-1}'\right)},
$$

where $\mathbf{Q}_{\widehat{\lambda}} = \mathbf{I}_N - \mathbf{P}_{\widehat{\lambda}}$, where $\mathbf{P}_{\widehat{\lambda}} = \widehat{\boldsymbol{\Lambda}} (\widehat{\boldsymbol{\Lambda}}' \widehat{\boldsymbol{\Lambda}})^{-1} \widehat{\boldsymbol{\Lambda}}'$ is the matrix that projects onto the space orthogonal to $\widehat{\boldsymbol{\Lambda}}$. Moon and Perron (2004) propose the statistics

$$
t_a = \frac{T \sqrt{N} (\widehat{\rho}_{\text{pool}}^* - 1)}{\sqrt{2 \widehat{\kappa}_{u^0}^4 / \widehat{\omega}_{u^0}^4}}, \tag{4.24}
$$

and

$$t_b = T\sqrt{N}(\widehat{\rho}^*_{\text{pool}} - 1)\left(\frac{\widehat{\omega}_{u^0}}{\widehat{\kappa}^2_{u^0}}\right)\sqrt{\tfrac{1}{T^2N}\operatorname{tr}\left(\overline{\mathbf{X}}_{-1}\mathbf{Q}_{\widehat{\lambda}}\overline{\mathbf{X}}'_{-1}\right)}, \tag{4.25}$$

where $\widehat{\omega}_{u^0} = \sqrt{\widehat{\omega}^2_{u^0}}$, $\widehat{\omega}^4_{u^0} = (\widehat{\omega}^2_{u^0})^2$, and $\widehat{\kappa}^2_{u^0} = \sqrt{\widehat{\kappa}^4_{u^0}}$. Under the null hypothesis $\rho_i^0 = 1$ for all i, the statistics (4.24) and (4.25) tend to the standard normal distribution as $N, T \to \infty$ with $N/T \to 0$.

The model (4.20) considered by Moon and Perron (2004) and the model (4.18) considered by Bai and Ng (2004) are closely related. If $x_{i,0} = x^0_{i,0} = 0$ for all i, and if we impose the restriction $\rho_i = \rho_i^0 = \rho$ for all i, then the models are parametrically equivalent with $(1-\rho L)\mathbf{f}_t = \mathbf{f}_t^0$ and $(1-\rho L)u_{i,t} = u^0_{i,t}$. That is, if the factors and the idiosyncratic components are restricted to have the same order of integration under homogeneity. If we impose heterogeneity, then the models are only approximately equivalent in the parameters. However, the assumptions made on the processes are somewhat different. In principle, Moon and Perron (2004) make the same assumption for $\mathbf{f}_t^0$ as Bai and Ng (2004) do for $(1-L)\mathbf{f}_t$.

4.3.5 Bai and Ng revisited

Based on the procedures of Moon and Perron (2004), Bai and Ng (2010) propose two pooled unit root tests applied to the idiosyncratic terms in (4.18) estimated as $\widehat{u}_{i,t} = \sum_{s=2}^T \widehat{z}_{i,s}$, where $\widehat{z}_{i,t}$ is found from the procedures explained in Section 4.3.3. Let $\widehat{\mathbf{u}}_{-1}$ and $\widehat{\mathbf{u}}$ be the $(T-2)\times N$ matrices defined analogously to $\overline{\mathbf{X}}_{-1}$ and $\overline{\mathbf{X}}$ in Section 4.3.4 (or if the index in levels starts at $t = 0$, then $\widehat{\mathbf{u}}_{-1}$ and $\widehat{\mathbf{u}}$ are $(T-1)\times N$ matrices), and let $\widehat{\rho}^+$ be the bias-adjusted pooled OLS estimator of ρ in $\widehat{u}_{i,t} = \rho\widehat{u}_{i,t-1} + \varepsilon_{i,t}$,

$$\widehat{\rho}^+ = \frac{\operatorname{tr}(\widehat{\mathbf{u}}'_{-1}\widehat{\mathbf{u}}) - NT\widehat{\zeta}_\varepsilon}{\operatorname{tr}(\widehat{\mathbf{u}}'_{-1}\widehat{\mathbf{u}}_{-1})},$$

where $\widehat{\zeta}_\varepsilon = N^{-1}\sum_{i=1}^N \widehat{\zeta}_{\varepsilon,i}$ is the analog to (4.23) based on the residuals $\widehat{\varepsilon}_{i,t} = \widehat{u}_{i,t} - \widehat{\rho}\widehat{u}_{i,t-1}$, where $\widehat{\rho} = \operatorname{tr}(\widehat{\mathbf{u}}'_{-1}\widehat{\mathbf{u}})/\operatorname{tr}(\widehat{\mathbf{u}}'_{-1}\widehat{\mathbf{u}}_{-1})$ is the standard pooled OLS estimator. Based on the same residuals, $\widehat{\varepsilon}_{i,t}$, define analogously to (4.21) and (4.22) the average moment estimators $\widehat{\omega}^2_\varepsilon = \frac{1}{N}\sum_{i=1}^N \widehat{\omega}^2_{\varepsilon,i}$ and $\widehat{\kappa}^4_\varepsilon = \frac{1}{N}\sum_{i=1}^N \widehat{\omega}^4_{\varepsilon,i}$. Bai and Ng (2010) propose the following two statistics:

$$P_a = \frac{T\sqrt{N}(\widehat{\rho}^+ - 1)}{\sqrt{2\widehat{\kappa}^4_\varepsilon/\widehat{\omega}^4_\varepsilon}}, \tag{4.26}$$

$$P_b = T\sqrt{N}(\widehat{\rho}^+ - 1)\left(\frac{\widehat{\omega}_\varepsilon}{\widehat{\kappa}^2_\varepsilon}\right)\sqrt{\tfrac{1}{T^2N}\operatorname{tr}(\widehat{\mathbf{u}}'_{-1}\widehat{\mathbf{u}}_{-1})}, \tag{4.27}$$

where as before $\widehat{\omega}_\varepsilon = \sqrt{\widehat{\omega}_\varepsilon^2}$, $\widehat{\omega}_\varepsilon^4 = (\widehat{\omega}_\varepsilon^2)^2$, and $\widehat{\kappa}_\varepsilon^2 = \sqrt{\widehat{\kappa}_\varepsilon^4}$. Under the null hypothesis $\rho_i = 1$ for all i, the statistics (4.26) and (4.27) tend to the standard normal distribution as $N, T \to \infty$ with $N/T \to 0$.

4.4 Contributions of this thesis

We consider the factor model (4.13) (including a constant term) with cross-sectionally independent idiosyncratic terms, and derive some likelihood-based tests for unit roots in the common and idiosyncratic components. To our knowledge, unit root tests based on maximum likelihood in this setting has not been attempted before. An attractive property of likelihood-based tests is that they are generally powerful against local alternatives. The simulation studies in this thesis support this, and suggest also that the tests are robust in terms of size compared with existing second generation panel unit root tests.

We propose that more general tests that allow for moderate idiosyncratic cross-sectional dependencies, time heteroscedasticity, and other type of dynamics, can be treated in a quasi-likelihood setting following the same principles as in this thesis. As such, the results presented here should contribute to the development of unit root tests in dynamic factors models based on maximum likelihood.

5. Main results

5.1 Framework

We consider the multivariate process

$$\mathbf{x}_t = (x_{1,t}, x_{2,t}, \ldots, x_{N,t})' = \boldsymbol{\mu} + \boldsymbol{\Lambda}\mathbf{f}_t + \mathbf{u}_t, \qquad (5.1)$$

where $\boldsymbol{\mu} = (\mu_1, \mu_2, \ldots, \mu_N)'$ is a vector of constants, $\mathbf{u}_t = (u_{1,t}, u_{2,t}, \ldots, u_{N,t})'$ is a vector of independent idiosyncratic components, $\mathbf{f}_t = (f_{1,t}, f_{2,t}, \ldots, f_{r,t})'$ is a vector of independent dynamic common factors and $\boldsymbol{\Lambda}$ is the $N \times r$ matrix of factor loadings. Only $\mathbf{x}_t$ is observed, while all components on the right side of Equation (5.1) are unobservable. In panel notation, the model (5.1) may be written as

$$x_{i,t} = \mu_i + \boldsymbol{\lambda}_i'\mathbf{f}_t + u_{i,t}, \qquad i = 1, 2, \ldots, N; t = 1, 2, \ldots, T,$$

where μ_i is the individual-specific constant, $\boldsymbol{\lambda}_i = (\lambda_{i,1}, \lambda_{i,2}, \ldots, \lambda_{i,r})'$ is the $r \times 1$ vector of factor loadings associated with the ith individual, and $u_{i,t}$ is the dynamic idiosyncratic component. Here we make the following general assumptions:

Assumption 5.1 The idiosyncratic components are Gaussian AR(1), $u_{i,t} = \rho_i u_{i,t-1} + \varepsilon_{i,t}$, where $\rho_i \in (-1, 1]$, $\varepsilon_{i,t} \sim N(0, \sigma^2_{\varepsilon,i})$ are iid with $\sigma^2_{\varepsilon,i} < \infty$, and $E(\varepsilon_{i,t}\varepsilon_{l,s}) = 0$ for all $i \neq l$ $(i, l = 1, 2, \ldots, N)$ and all $t, s = 1, 2, \ldots, T$.
Assumption 5.2 The factors are Gaussian AR(1), $f_{j,t} = \alpha_j f_{j,t-1} + v_{j,t}$, where $\alpha_j \in (-1, 1]$, $v_{j,t} \sim N(0, 1)$ are iid, and $E(v_{j,t}v_{q,s}) = 0$ for all $j \neq q$ $(j, q = 1, 2, \ldots, r)$ and all $t, s = 1, 2, \ldots, T$.
Assumption 5.3 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 5.4 The errors $v_{j,t}$ and $\varepsilon_{i,t}$ are mutually independent at all leads and lags, such that $E(v_{j,t}\varepsilon_{i,s}) = 0$ for all $j = 1, 2, \ldots, r$, all $i = 1, 2, \ldots, N$ and all $t, s = 1, 2, \ldots, T$.
Assumption 5.5 For stationary processes $f_{j,t}$ and $u_{i,t}$, the starting values $f_{j,0}$ and $u_{i,0}$ come from the stationary distributions, while for non-stationary processes they are $O_p(1)$.

Under Assumption 5.1 the idiosyncratic components are allowed to be cross-sectionally heteroscedastic but time heteroscedasticity is ruled out. Also, because the errors $\varepsilon_{i,t}$ are independent over $i = 1, 2, \ldots, N$ we have the exact factor model. By assumption 5.2 the factors are defined analogously to the idiosyncratic components, where the unit variance in the errors $v_{j,t}$ can be made without loss of generality as explained in Section 4.3. The normality assumptions will allow us to obtain the likelihood. Assumption 5.3 implies that $\mathbf{\Lambda}$ has rank r (full rank) and that the eigenvalues of $\mathbf{\Lambda}'\mathbf{\Lambda}$ are $O(N)$, which together with Assumption 5.4 is standard in factor analysis to identify the covariance structure (see Section 4.3.2). Assumption 5.5 assures us that the time series processes are well-defined with respect to their initial values.

Assumptions 5.1 and 5.2 will be specified somewhat differently in the respective papers due to that we will consider different hypotheses. In Paper I and II we fix $\alpha = 1$ and test for idiosyncratic unit roots, while in Paper II we propose two additional tests where we, respectively, fix $\rho = 1$ and test for common unit roots, and then leave the alternative unrestricted and test for "not $\alpha = 1, \rho = 1$". In Paper IV we allow the factors to have quite general forms of dynamics and test for idiosyncratic unit roots, where the test is shown to be asymptotically independent of the distribution of the factors.

Our likelihood based analysis rests on the same observation as is made by Kruiniger (2008). Let, for notational convenience, $T^* = T - 1$, and let $\mathbf{D}$ be the first-difference-matrix

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

Also, for the ith idiosyncratic component, let $\mathbf{u}_i = (u_{i,1}, u_{i,2}, \ldots, u_{i,T})'$. Consider first the case when $\mathbf{u}_i$ has a unit root, i.e. $\rho_i = 1$. After taking differences we have that $\mathbf{D}\mathbf{u}_i \sim N_T(\mathbf{0}, \sigma^2_{\varepsilon,i}\mathbf{I})$. Consider next the case when $\mathbf{u}_i$ is stationary, i.e. $|\rho_i| < 1$. If the starting value is chosen such that $E(u_{i,0}) = 0$ and $Var(u_{i,0}) = \sigma^2_{\varepsilon,i}/(1-\rho_i^2)$, then we have that, in levels, $\mathbf{u}_i \sim N_T(\mathbf{0}, \sigma^2_{\varepsilon,i}\mathbf{\Pi}(\rho_i))$, where $\mathbf{\Pi}(\rho_i) = [\Pi^i_{k,m}]_{T\times T}$ is an autocovariance matrix with

$$\Pi^i_{k,m} = \begin{cases} \frac{1}{1-\rho_i^2} & \text{for } k = m \\ \frac{\rho_i^{|k-m|}}{1-\rho_i^2} & \text{for } k \neq m \end{cases}.$$

This matrix may be found explicitly from, e.g., van der Leeuw (1994) or Karanasos (1998). It is now straightforward to show that $\mathbf{D}\mathbf{u}_i \sim N_{T^*}(\mathbf{0}, \sigma^2_{\varepsilon,i}\mathbf{\Psi}(\rho_i))$, where $\mathbf{\Psi}(\rho_i) = [\Psi^i_{k,m}]_{T^*\times T^*}$ with

$$\Psi^i_{k,m} = \begin{cases} \frac{2}{1+\rho_i} & \text{for } k = m \\ -\frac{\rho_i^{||k-m|-1|}(1-\rho_i)}{1+\rho_i} & \text{for } k \neq m \end{cases}. \tag{5.2}$$

The autocovariance matrix $\boldsymbol{\Psi}$ has the following convenient properties; (a) $\boldsymbol{\Psi}(1) = \mathbf{I}_{T^*}$, i.e. it is correctly defined in the non-stationary point; and (b) it is infinitely many times continuously differentiable at and in the neighborhood of the unit root $\rho_i = 1$ (see Kruiniger, 2008, Lemma 7). It is thus well-defined for a likelihood based analysis.

Likewise, for the jth factor, let $\mathbf{f}_j = (f_{j,1}, f_{j,2}, \ldots, f_{j,T})'$. If there is a unit root in $\mathbf{f}_j$, then $\mathbf{Df}_j \sim N_{T^*}(\mathbf{0}, \mathbf{I})$, and if $|\alpha_j| < 1$, then $\mathbf{Df}_j \sim N_{T^*}(\mathbf{0}, \boldsymbol{\Psi}(\alpha_j))$, where $\boldsymbol{\Psi}(\alpha_j) = [\Psi^j_{k,m}]_{T^* \times T^*}$ has the same elements as (5.2), but with parameter α_j.

Assumption 5.6 The idiosyncratic components and the factors are separate homogenous groups; $\rho_i = \rho$ for all i, and $\alpha_j = \alpha$ for all j.

Assumption 5.6, which may be relaxed, simplifies the analysis. Imposing this assumption will affect power, but not size. The findings in this thesis are that the impact on power is small close to the unit root, which is generally of most interest.

Consider now the following null hypothesis:

$$H_0 : \rho = 1, \alpha = 1,$$

which under Assumptions 5.1-5.6 may be equivalently stated as

$$H_0 : u_{i,t} \sim I(1), f_{j,t} \sim I(1), \text{for all } i \text{ and all } j.$$

Let $\mathbf{X} = (\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_T)$ be the $N \times T$ matrix of observed panel data and let $\mathbf{Y} = \mathbf{XD}'$ be the $N \times T^*$ matrix of differenced panel data. In stacked form, $\mathbf{Y}_v = \mathrm{vec}(\mathbf{Y}) = (\mathbf{y}_2', \mathbf{y}_3', \ldots, \mathbf{y}_T')'$, the differenced data has covariance matrix

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

where $\otimes$ denotes the Kronecker product and $\boldsymbol{\Sigma}_{\varepsilon\varepsilon} = \mathrm{diag}(\sigma^2_{\varepsilon,1}, \sigma^2_{\varepsilon,2}, \ldots, \sigma^2_{\varepsilon,N})$. Also, let $\boldsymbol{\theta} = (\boldsymbol{\Lambda}_v', \sigma^2_{\varepsilon,1}, \sigma^2_{\varepsilon,2}, \ldots, \sigma^2_{\varepsilon,N}, \alpha, \rho)'$ be the parameter vector holding the $K = Nr + N + 2$ parameters imposed under Assumptions 5.1-5.6, and let $\widetilde{\boldsymbol{\theta}} = (\widetilde{\boldsymbol{\Lambda}}_v', \widetilde{\sigma}^2_{\varepsilon,1}, \widetilde{\sigma}^2_{\varepsilon,2}, \ldots, \widetilde{\sigma}^2_{\varepsilon,N}, 1, 1)'$ be the vector of restricted MLEs under the null hypothesis. Assuming normality, the log-likelihood with respect to the differenced and stacked data is

$$l(\boldsymbol{\theta}) = -\frac{NT^*}{2} \log 2\pi - \frac{1}{2} \log |\boldsymbol{\Sigma}| - \frac{1}{2} \mathbf{Y}_v' \boldsymbol{\Sigma}^{-1} \mathbf{Y}_v. \tag{5.3}$$

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

$$\boldsymbol{\Sigma} = \left(\mathbf{I}_{T^*} \otimes \boldsymbol{\Lambda}\boldsymbol{\Lambda}'\right) + \left(\mathbf{I}_{T^*} \otimes \boldsymbol{\Sigma}_{\varepsilon\varepsilon}\right) = \mathbf{I}_{T^*} \otimes \boldsymbol{\Omega}, \tag{5.4}$$

where $\boldsymbol{\Omega} = \boldsymbol{\Lambda}\boldsymbol{\Lambda}' + \boldsymbol{\Sigma}_{\varepsilon\varepsilon}$, such that, using some well-known results for the Kronecker product (see e.g Magnus and Neudecker, 2001, p. 28-31),

$$\begin{aligned} l(\boldsymbol{\theta}) &= -\frac{NT^*}{2}\log 2\pi - \frac{1}{2}\log|\boldsymbol{\Omega}|^{T^*} - \frac{1}{2}\operatorname{vec}(\mathbf{Y})'\left(\mathbf{I}_{T^*}\otimes\boldsymbol{\Omega}^{-1}\right)\operatorname{vec}(\mathbf{Y}) \\ &= -\frac{NT^*}{2}\log 2\pi - \frac{T^*}{2}\log|\boldsymbol{\Omega}| - \frac{T^*}{2}\operatorname{tr}\left(\boldsymbol{\Omega}^{-1}\tfrac{1}{T^*}\mathbf{Y}\mathbf{Y}'\right). \end{aligned}$$

That is, the likelihood is the same as (4.15). If we impose yet another assumption that the idiosyncratic components are cross-sectionally homoscedastic, i.e. $\sigma^2_{\varepsilon,i} = \sigma^2_{\varepsilon}$ for all $i = 1, 2, \ldots, N$ implying $\boldsymbol{\Sigma}_{\varepsilon\varepsilon} = \sigma^2_{\varepsilon}\mathbf{I}_N$, then, under the restrictions of the null hypothesis, we have an explicit MLE of $\boldsymbol{\Omega}$, $\widetilde{\boldsymbol{\Omega}}_{ML} = \boldsymbol{\Omega}(\widetilde{\boldsymbol{\theta}}) = \widetilde{\boldsymbol{\Lambda}}\widetilde{\boldsymbol{\Lambda}}' + \widetilde{\sigma}^2_{\varepsilon}\mathbf{I}_N$, where $\widetilde{\sigma}^2_{\varepsilon}$ and $\widetilde{\boldsymbol{\Lambda}}$ are the MLEs defined in (4.16) and (4.17) respectively. If we let $\boldsymbol{\Sigma}_{\varepsilon\varepsilon}$ be a diagonal matrix with distinct elements, then a quasi-MLE of $\boldsymbol{\Omega}$ under the null hypothesis may be found from the EM algorithm described in Section 4.3.2. For future reference we define the MLE of $\boldsymbol{\Omega}$ under the null hypothesis, under both idiosyncratic homoscedasticity and heteroscedasticity, as

$$\mathbf{S}_{01} = \boldsymbol{\Omega}(\widetilde{\boldsymbol{\theta}}). \tag{5.5}$$

Assumption 5.7 The idiosyncratic components are cross-sectionally homoscedastic; $\sigma^2_{\varepsilon,i} = \sigma^2_{\varepsilon}$ for all i.

Assumption 5.7 is maintained through Papers I-III. Imposing this assumption simplifies the analysis and makes the derivations of the asymptotic distributions of the proposed unit root tests rather straightforward. The results are still important, because relaxing Assumption 5.7 (and Assumption 5.6) is likely to result in test-statistics with similar asymptotic distributions. In Paper IV we relax Assumption 5.7.

Lastly, consider the sample covariances $\mathbf{Y}_v\mathbf{Y}_v'$, which consists of the T^{*2} blocks $\mathbf{S}_{t,s} = \mathbf{y}_t\mathbf{y}_s'$ for $t, s = 2, 3, \ldots, T$, where each block is of size $N \times N$. For these blocks we define the key products

$$\mathbf{S}_0 = \sum_{t=2}^{T}\mathbf{S}_{t,t} = \sum_{t=2}^{T}\mathbf{y}_t\mathbf{y}_t' = \mathbf{Y}\mathbf{Y}', \tag{5.6}$$

and

$$\mathbf{S}_{00} = \sum_{t=2}^{T}\sum_{s=2}^{T}\mathbf{S}_{t,s} = \sum_{t=2}^{T}\sum_{s=2}^{T}\mathbf{y}_t\mathbf{y}_s' = \left(\sum_{t=2}^{T}\mathbf{y}_t\right)\left(\sum_{t=2}^{T}\mathbf{y}_t\right)'. \tag{5.7}$$

5.2 Paper I

We fix $\alpha_j = 1$ for all $j = 1,2,\ldots,r$, and consider the hypotheses

$$H_0 : \rho = 1 | \forall j, f_{j,t} \sim I(1),$$
$$H_1 : \rho < 1 | \forall j, f_{j,t} \sim I(1).$$

That is, the parameter vector may be partitioned as $\boldsymbol{\theta} = (\boldsymbol{\theta}_1', \boldsymbol{\theta}_2')'$, where $\boldsymbol{\theta}_1 = \left(\boldsymbol{\Lambda}_v', \sigma_\varepsilon^2\right)'$ and $\boldsymbol{\theta}_2 = \rho$. Let $\widetilde{\boldsymbol{\theta}} = (\widetilde{\boldsymbol{\Lambda}}_v', \widetilde{\sigma}_\varepsilon^2, 1)'$ be the constrained MLE under the null hypothesis. We show that the lower right part of the information matrix (see Chapter 3) is the scalar $J^{22} = J_{22}^{-1}(1 - \frac{1}{T^*})$, where $J_{22} = -E\left(\partial^2 l(\boldsymbol{\theta})/\partial \rho^2\right)$, and derive the "square-root" LM-statistic

$$\vartheta_a = \frac{\partial l}{\partial \rho}\sqrt{J^{22}}\bigg|_{\boldsymbol{\theta}=\widetilde{\boldsymbol{\theta}}} = \frac{T^* \operatorname{tr}\left(\mathbf{S}_{01}^{-1}\right) - 2\operatorname{tr}\left(\mathbf{S}_{01}^{-1}\mathbf{S}_0\mathbf{S}_{01}^{-1}\right) + \operatorname{tr}\left(\mathbf{S}_{01}^{-1}\mathbf{S}_{00}\mathbf{S}_{01}^{-1}\right)}{\sqrt{2T^*\left(T^*-1\right)\operatorname{tr}\left(\mathbf{S}_{01}^{-1}\mathbf{S}_{01}^{-1}\right)}},$$

where $\mathbf{S}_{01}$, $\mathbf{S}_0$ and $\mathbf{S}_{00}$ are given by (5.5), (5.6) and (5.7) respectively. For any fixed $N > r$ the asymptotic distribution of this statistic as T tends to infinity is a weighted sum of independent χ_1^2 (chi-square with one degrees of freedom) variables:

$$\vartheta_a \xrightarrow{d} \frac{1}{\sqrt{2\sum_{i=1}^N w_i^2}}\left(\sum_{i=1}^N w_i \chi_{1,i}^2 - \sum_{i=1}^N w_i\right), \text{ as } T \to \infty, \tag{5.8}$$

where $\chi_{1,i}^2$ are independent over $i = 1,2,\ldots,N$ with weights $w_i = \sigma_\varepsilon^2/(\eta_i + \sigma_\varepsilon^2)$ for $1 \leq i \leq r$ and $w_i = 1$ for $r < i \leq N$, where σ_ε^2 is the idiosyncratic error variance imposed under Assumption 5.7 and $\eta_i = \varphi_i(\boldsymbol{\Lambda}'\boldsymbol{\Lambda})$. This weighted χ^2-distribution is non-standard, and the critical values will depend on the estimated weights $\widehat{w}_i$. However, we show that it can be suitably approximated by

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

with $d_1 = \left[\operatorname{tr}(\boldsymbol{\Omega}^{-1})\right]^2 / \operatorname{tr}\left(\boldsymbol{\Omega}^{-1}\boldsymbol{\Omega}^{-1}\right)$, where $\boldsymbol{\Omega}$ is the contemporaneous covariance matrix in (5.4). Letting sequentially also N tend to infinity, the distribution of the test-statistic is standard normal, $\vartheta_a \xrightarrow{d} N(0,1)$ as $(T,N)_s \to \infty$.

By simulation we show that ϑ has nice properties even for finite samples except when N is large and T is small, or if N is not large enough relative to r. We also show that ϑ has higher power than the pooled Fisher-type test (4.19) when the factors are integrated. Especially, our test statistic seems to have quite substantially higher local power, where we compare the rejection rate versus a sequence of local alternatives as described in Section 4.2. Further, the simulations suggest that the test-statistic is robust against the conditioning on integrated factors, and that it has nice size and power properties also for stationary factors as long as the panel dimensions are large enough.

5.3 Paper II

We consider the same restrictions as in Paper I, and evaluate, by simulation, the power of the log-likelihood ratio test

$$LR = \sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}_0} l(\boldsymbol{\theta}|\mathbf{Y}) - \sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}_1} l(\boldsymbol{\theta}|\mathbf{Y}),$$

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 of the log-likelihood (5.3). Under the null hypothesis the MLE of $\boldsymbol{\Omega}$ is found from (5.5) with the explicit MLEs (4.16) and (4.17), while under the alternative hypothesis it has to be found numerically.

We compare power of the likelihood ratio test (LR) with the analogous LM-test in Paper I (LM) and the Fisher-type test (4.19) proposed by Bai and Ng (2004) (BN). Our results suggest that LR and LM have higher power than BN, and while LR and LM are equivalent in terms of power for large samples, their performance in small samples depend on which dimension is small. When the time dimension is small, LM is more powerful than LR, while the benefit of LR seems to be for a small cross-section and large enough T. Figure 5.1 shows the size-adjusted power of the three tests for some finite samples.

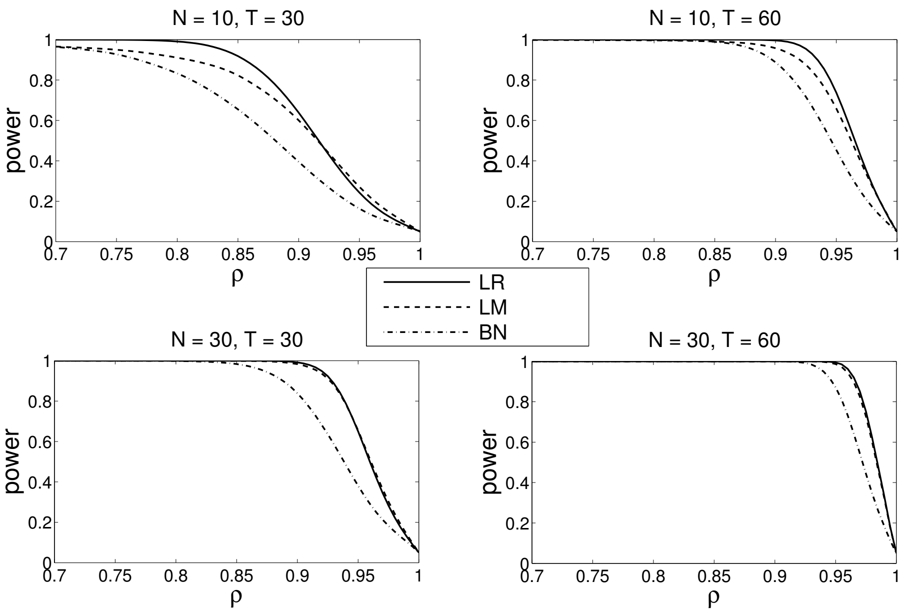

Figure 5.1. Size-adjusted power of LR, LM and BN when $r = 1$ for $T = 30, 60$ and $N = 10, 30$.

5.4 Paper III

In Paper III we consider, opposite to paper I and II, a homogenous test for unit roots in the factors by keeping $\rho_i = 1$ fixed for all i. The hypotheses are then

$$H_0 : \alpha = 1 | \forall i, u_{i,t} \sim I(1),$$
$$H_1 : \alpha < 1 | \forall i, u_{i,t} \sim I(1).$$

That is, the parameter vector may be partitioned as $\boldsymbol{\theta} = (\boldsymbol{\theta}_1', \boldsymbol{\theta}_2')'$, where $\boldsymbol{\theta}_1 = \left(\boldsymbol{\Lambda}_v', \sigma_\varepsilon^2\right)'$ and $\boldsymbol{\theta}_2 = \alpha$. We show that for any fixed N as $T \to \infty$, the information matrix is block-diagonal such that the "square root" LM-statistic is, approximately,

$$\vartheta_b = \frac{\partial l}{\partial \alpha} \sqrt{J^{22}} \Bigg|_{\boldsymbol{\theta}=\widetilde{\boldsymbol{\theta}}} \approx \vartheta_b^* = \frac{\partial l}{\partial \alpha} \sqrt{J_{22}^{-1}} \Bigg|_{\boldsymbol{\theta}=\widetilde{\boldsymbol{\theta}}},$$

in the sense that $|\vartheta_b - \vartheta_b^*| = o_p(1)$, where $J_{22} = -E\left(\partial^2 l(\boldsymbol{\theta})/\partial \alpha^2\right)$. Deriving the score and information we have that

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

where $\mathbf{M} = \boldsymbol{\Lambda}\boldsymbol{\Lambda}'$, and $\mathbf{S}_{01}$, $\mathbf{S}_0$ and $\mathbf{S}_{00}$ are given by (5.5), (5.6) and (5.7) respectively. Similar to Paper I, the asymptotic distribution of ϑ_b^* as $T \to \infty$ is a weighted chi-square distribution,

$$\vartheta_b^* \xrightarrow{d} \frac{1}{\sqrt{2\sum_{j=1}^r \overline{w}_j^2}} \left(\sum_{j=1}^r \overline{w}_j \chi_{1,j}^2 - \sum_{j=1}^r \overline{w}_j \right), \tag{5.9}$$

where $\chi_{1,j}^2$ are independent over $j = 1, \ldots, r$ with weights $\overline{w}_j = \eta_j/(\sigma_\varepsilon^2 + \eta_j)$, where σ_ε^2 is the idiosyncratic error variance imposed under Assumption 5.7 and $\eta_j = \varphi_j(\boldsymbol{\Lambda}'\boldsymbol{\Lambda})$. Here the weights $\overline{w}_j$ have the properties that $\overline{w}_j + w_j = 1$ for $j = 1, 2, \ldots, r$, where w_j are the first r weights in (5.8). From Assumption 5.3 it follows that the eigenvalues of $\boldsymbol{\Lambda}'\boldsymbol{\Lambda}$ (and likewise the r non-zero eigenvalues of $\boldsymbol{\Lambda}\boldsymbol{\Lambda}'$) are $O(N)$, with the result that, as $(T,N)_s \to \infty$, $\vartheta_b^* \xrightarrow{d} \frac{1}{\sqrt{2r}}(\chi_r^2 - r)$. For small N the distribution (5.9) is suitably approximated with

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

where $d_2 = \left[\operatorname{tr}\left(\boldsymbol{\Lambda}\boldsymbol{\Lambda}'\boldsymbol{\Omega}^{-1}\right)\right]^2 / \operatorname{tr}\left(\boldsymbol{\Lambda}\boldsymbol{\Lambda}'\boldsymbol{\Omega}^{-1}\boldsymbol{\Lambda}\boldsymbol{\Lambda}'\boldsymbol{\Omega}^{-1}\right)$.

In addition, we propose two tests with the hypotheses $H_0 : \alpha = 1, \rho = 1$ versus $H_1 : \alpha < 1$ and/or $\rho < 1$, which are essentially combinations of ϑ_a in Paper I and ϑ_b^* in Paper III. By simulation we show that the proposed statistics have higher local power than the MQ_c^c-statistic of Bai and Ng (2004) presented in Section 4.3.3, which appears to have virtually no power against locally stationary factors. Further, though we do not provide a formal proof, the simulation results suggest that ϑ_b^* is robust to the conditioning on $I(1)$-idiosyncratic components.

5.5 Paper IV

We consider a similar setup as in Paper I, but relax the assumption of cross-sectional homoscedasticity in the idiosyncratic components. That is, we let $\mathbf{\Sigma}_{\varepsilon\varepsilon} = \text{diag}(\sigma^2_{\varepsilon,1}, \sigma^2_{\varepsilon,2}, \dots, \sigma^2_{\varepsilon,N})$. Also, we treat the $I(1)$ factors as a misspecification, which we show our statistic to be robust against in large panels. Let $\mathbf{\Sigma}_{gg} = Var(\mathbf{g}_t)$, where $\mathbf{g}_t = \Delta\mathbf{f}_t$. Assumption 5.2 is replaced with the more general assumption:

Assumption 5.2* (difference-stationary factors) The factors admit the representation $(1-L)\mathbf{f}_t = \mathbf{C}(L)\mathbf{v}_t$, where $\mathbf{C}(L) = \sum_{m=0}^{\infty} \mathbf{C}_m L^m$ is an $r \times r$ lag polynomial matrix, and where

(i) $\mathbf{v}_t \sim N_r(\mathbf{0}, \mathbf{\Sigma}_{vv})$ is iid, where $E||\mathbf{v}_t||^4 < \infty$,

(ii) $\mathbf{\Sigma}_{gg} = \sum_{m=0}^{\infty} \mathbf{C}_m \mathbf{\Sigma}_{vv} \mathbf{C}_m' > 0$,

(iii) $\sum_{m=0}^{\infty} m||\mathbf{C}_m|| < \infty$,

(iv) $\text{rk}(\mathbf{C}(1)) = r_1$, where $0 \le r_1 \le r$.

Assumption 5.2* is the same dynamic assumption made on the factors in Bai and Ng (2004), except we assume normally distributed shocks. It allows for r_1 stochastic trends among the factors, where $0 \le r_1 \le r$. We also impose the following misspecification:

Misspecification 5.1 The differenced factors are multivariate white noise, drawn from the normal distribution, $\Delta\mathbf{f}_t \sim N_r(\mathbf{0}, \mathbf{I})$.

Under Misspecification 5.1 the parameter vector may be partitioned as $\boldsymbol{\theta} = (\boldsymbol{\theta}_1', \boldsymbol{\theta}_2')'$, where $\boldsymbol{\theta}_1 = (\mathbf{\Lambda}_v', \sigma^2_{\varepsilon,1}, \sigma^2_{\varepsilon,2}, \dots, \sigma^2_{\varepsilon,N})'$ and $\boldsymbol{\theta}_2 = \rho$. We then consider the hypotheses

$$H_0 : \rho = 1,$$
$$H_1 : \rho < 1.$$

We show that as $T \to \infty$, the information matrix is block-diagonal, and we consider again the approximate "square-root" LM-statistic

$$\vartheta = \frac{\partial l}{\partial \rho} \sqrt{J^{22}} \bigg|_{\boldsymbol{\theta} = \widetilde{\boldsymbol{\theta}}} \approx \vartheta^* = \frac{\partial l}{\partial \rho} \sqrt{J_{22}^{-1}} \bigg|_{\boldsymbol{\theta} = \widetilde{\boldsymbol{\theta}}},$$

in the sense that $|\vartheta-\vartheta^*|=o_p(1)$, where $J_{22}=-E\left(\partial^2 l(\boldsymbol{\theta})/\partial\rho^2\right)$. Here ϑ^* has the decomposition

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

where $\mathbf{S}_0$ and $\mathbf{S}_{00}$ are given by (5.6) and (5.7), and $\mathbf{S}_{01}=\widetilde{\boldsymbol{\Lambda}}\widetilde{\boldsymbol{\Lambda}}'+\widetilde{\boldsymbol{\Sigma}}_{\varepsilon\varepsilon}$ with $\widetilde{\boldsymbol{\Sigma}}_{\varepsilon\varepsilon}=\operatorname{diag}(\widetilde{\sigma}^2_{\varepsilon,1},\widetilde{\sigma}^2_{\varepsilon,2},\ldots,\widetilde{\sigma}^2_{\varepsilon,N})$, where $\widetilde{\boldsymbol{\Lambda}}$ and $\widetilde{\boldsymbol{\Sigma}}_{\varepsilon\varepsilon}$ can be found from the EM algorithm. The asymptotic distribution of the test-statistic is

$$\vartheta^*\xrightarrow{d}\frac{1}{\sqrt{2\sum_{i=1}^N w_i^2}}\left(\sum_{i=1}^N w_i\chi^2_{1,i}-\sum_{i=1}^N w_i\right),\tag{5.10}$$

where $\chi^2_{1,i}$ are independent over $i=1,2,\ldots,N$ with associated weights

$$w_i=\begin{cases}(1+\eta_i)^{-1} & \text{if } 1\le i\le r\\ 1 & \text{if } r<i\le N\end{cases},$$

where $\eta_i=\varphi_i\left(\boldsymbol{\Lambda}'\boldsymbol{\Sigma}_{\varepsilon\varepsilon}^{-1}\boldsymbol{\Lambda}\right)$ $(\eta_1\ge\eta_2\ge\cdots\ge\eta_r\ge 0)$. As before, for small N the distribution (5.10) is suitably approximated with

$$\vartheta^*\overset{app.}{\sim}\frac{1}{\sqrt{2d_3}}(\chi^2_{d_3}-d_3),$$

where $d_3=\left[\operatorname{tr}(\boldsymbol{\Sigma}_{\varepsilon\varepsilon}\boldsymbol{\Omega}^{-1})\right]^2/\operatorname{tr}(\boldsymbol{\Sigma}_{\varepsilon\varepsilon}\boldsymbol{\Omega}^{-1}\boldsymbol{\Sigma}_{\varepsilon\varepsilon}\boldsymbol{\Omega}^{-1})$, while for large N it is standard normal, $\vartheta^*\xrightarrow{d}N(0,1)$, as $(T,N)_s\to\infty$.

A key result in this paper is that we prove that the LM-type test is asymptotically independent of the distribution of the factors as long as they fall under Assumption 5.2*, meaning that if Misspecification 5.1 is violated, then it still holds that $\vartheta^*\xrightarrow{d}N(0,1)$ as $(T,N)_s\to\infty$.

In a simulation study we compare size and power with the Fisher-type test (4.19) proposed by Bai and Ng (2004), the statistics (4.24) and (4.25) of Moon and Perron (2004), and the statistics (4.26) and (4.27) of Bai and Ng (2010). The results suggest that; (a) the LM-statistic and the Fisher-type statistic are quite robust in terms of size, while the other statistics at times are grossly over-sized; (b) when the factors are stationary, the LM-statistic is under-sized for small samples, but tend to nominal size as the sample size grows, verifying that the LM-statistic is robust against Misspecification 5.1; and (c) the LM-statistic is locally most powerful of the considered statistics in a shrinking neighborhood of the null hypothesis as N and T grow large.

6. Future research

The results provided in this thesis are derived under rather strict assumptions. As such, there is a need to generalize the tests. For instance, it is desirable to relax the assumption of AR(1) dynamics, to allow for parameter heterogeneity, and to allow for deterministic trend components. I will make a few short remarks concerning these issues:

- AR(p) processes: To allow for higher order dynamics a common way is to non-parametrically adjust for autocorrelation by estimation of long-run variances. However, we could also consider a parametric approach using a similar procedure as in this thesis. For example, consider the AR(2) case, which, assuming Gaussian errors, can be decomposed as

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

  where $|\gamma_i| < 1$ and $\nu_{i,t}$ is iid $N(0, \sigma^2_{\nu,i})$, and where the unit root hypothesis corresponds to $\rho_i = 1, \forall i$. If we let $\mathbf{\Psi}(\rho_i, \gamma_i)$ denote the autocovariance matrix for the differenced AR(2) process, then $\mathbf{\Psi}(1, \gamma_i) = \mathbf{\Pi}(\gamma_i)$, where $\mathbf{\Pi}$ is the autocovariance matrix for a stationary AR(1) process in levels. Here the N nuisance parameters γ_i will have to be concentrated out. For the AR(p) case there will be $N(p-1)$ such nuisance parameters. The recent results in Bai and Li (2012b,a) offer a potential way of consistently estimating these nuisance parameters.
- Heterogeneity: Consider for simplicity the tests derived in Paper I and Paper IV. If we allow for heterogeneity, the effective part of the information matrix will no longer be a scalar, but will instead be an $N \times N$ matrix. That is, we need to take the inverse of a matrix of infinite size as $T, N \to \infty$. This will require some special attention to derive the asymptotic distributions when considering sequential, or perhaps joint, limits. Also, the "square-root" LM-type statistics will no longer be possible to consider. Instead we will have to look at the full LM test.
- Trend components: The demeaning procedure in Bai and Ng (2004), $\Delta x_{i,t} - \overline{\Delta x_i}$, where $\overline{\Delta x_i} = 1/(T-1)\sum_{t=2}^{T} \Delta x_{i,t}$, will remove a trend component from the data, but would not work for the LM-type tests considered here, because in the limit as $T \to \infty$, the stochastic part of the statistics comes from $\mathbf{S}_{00}$, which will then sum to zero. The trend component would have to be treated as a constant in the differenced data and be included in the likelihood function.

7. Acknowledgements

Many people have in one way or another contributed to the completion of this thesis. It is impossible to thank all of them here. The list would be too long. There are however a few people whom I would especially like to thank.

I wish to express my most sincere gratitude to my supervisor, Professor Rolf Larsson. Your enthusiasm and way of reasoning has been an inspiration to me. Thank you for your genuine support and for answering all of my, roughly one million, questions at any inconvenience. Although you were at most times right, please remember that APA is certainly foremost an animal, while only secondarily perhaps some kind of matrix multiplication. I also wish to express my appreciation to my second supervisor, Professor Johan Lyhagen. Your assistance and guidance, already when I was an undergraduate student, has been extremely valuable. I am further obliged to the rest of the personnel associated with the Department of Statistics, for their support and friendship. Lisbeth Hansson, you are a rock! Eva Enefjord, I have simply lost count of the number of times you have helped me.

For the last two years I have had the great pleasure to work closely with the outstanding Xingwu Zhou. I have truly enjoyed it. To you, and to all my other fellow PhD candidates, some of which have already graduated, thank you for a splendid time!

A large part of the PhD was funded by the Jan Wallander and Tom Hedelius Foundation through the research grants P2005-0117:1 and P2009-0189:1. This is gratefully acknowledged. I also thank Thun's foundation for supporting me with a travel grant to go to Maastricht for six months.

To my friends at the Department of Quantitative Economics, Maastricht University, thank you for making my visit a very pleasant one. I am immensely grateful to Professor Jean-Pierre Urbain. You always made me feel at home. Karin van den Boorn, thank you for all the help, concerning basically anything. If not for you, I think the planet might explode!

I would not have got far without the endless support of my parents. Thank you for making my life easier whenever in need, and for making it brilliant at any other times. The rest of my family; my brother, my two sisters, and Helene - wherever I go, you are on my mind.

Martin Solberger
August 14, 2013, on the train between Stockholm and Uppsala, constantly, but no more.

Bibliography

AITCHISON, J., AND S. D. SILVEY (1958): "Maximum likelihood estimation of parameters subject to restraints," *The Annals of Mathematical Statistics*, 29, 813–828.

ANDERSON, T. W. (2003): *An Introduction to Multivariate Statistical Analysis*. Wiley, New York.

ARELLANO, M. (2003): *Panel Data Econometrics*. Oxford University Press, Oxford.

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

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

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

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

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

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

——— (2007): "Determining the number of primitive shocks in factor models," *Journal of Business and Economic Statistics*, 25, 52–60.

——— (2008): "Large dimensional factor analysis," *Foundations and Trends in Econometrics*, 3, 89–168.

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

BALTAGI, B. H. (2008): *Econometric Analysis of Panel Data*. Wiley, New york, 4th edn.

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

BANERJEE, A., J. DOLADO, J. W. GALBRAITH, AND D. F. HENDRY (1993): *Co-integration, Error-correction, and the Econometric Analysis of Non-stationary Data*. Oxford University Press, Oxford.

BEKKER, P. A. (1989): "Identification in restricted factor models and the evaluation of rank conditions," *Journal of Econometrics*, 41, 5–16.

BERA, A. K., AND Y. BILIAS (2001): "Rao's score, Neyman's $C(\alpha)$ and Silvey's LM tests: an essay on historical developments and some new results," *Journal of Statistical Planning and Inference*, 97, 9–44.

BEVERIDGE, S., AND C. R. NELSON (1981): "A new approach to decomposition of economic time series into permanent and transitory components with particular attention to measurment of the 'business cycle'," *Journal of Monetary Economics*, 7, 151–174.

BHARGAVA, A. (1986): "On the theory of testing for unit roots in observed time series," *Review of Economic Studies*, 53, 369–384.

BOX, G. E., G. M. JENKINS, AND G. C. REINSEL (2008): *Time Series Analysis: Forecasting and Control*. Wiley, New Jersey, 4th edn.

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

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

BREITUNG, J., AND H. M. PESARAN (2008): "Unit roots and cointegration in panels," in *The Econometrics of Panel Data: Fundamentals and Recent Developments in Theory and Practice*, ed. by L. Matyas, and P. Sevestre, pp. 279–322. Kluwer Academic Publishers, Berlin, 3rd edn.

BREITUNG, J., AND J. TENHOFEN (2011): "GLS estimation of dynamic factor models," *Journal of the American Statistical Association*, 106, 1150–1166.

BREUSCH, T. S., AND A. R. PAGAN (1980): "The Lagrange multiplier test and its applications to model specification in econometrics," *Review of Economic Studies*, XLVII, 239–253.

BUSE, A. (1982): "The likelihood ratio, Wald, and Lagrange multiplier tests: an expository note," *The American Statistician*, 36, 153–157.

CAMBA-MENDEZ, G., G. KAPETANIOS, R. J. SMITH, AND M. R. WEALE (2001): "An automatic leading indicator of economic activity: forecasting GDP growth for European countries," *Econometrics Journal*, 4, S56–S90.

CAMPBELL, J. Y., AND P. PERRON (1991): "Pitfalls and opportunities: what macroeconomists should know about unit roots," *NBER Macroeconomics Annual*, 6, 141–200.

CARRION-I-SILVESTRE, J. L., AND A. SANSÓ (2006): "Joint hypothesis specification for unit root tests with a structural break," *Econometrics Journal*, 9, 196–224.

CERRATO, M., AND N. SARANTIS (2007): "A bootstrap panel unit root test under cross-sectional dependence, with an application to PPP," *Computational Statistics and Data Analysis*, 51, 4028–4037.

CHAMBERLAIN, G. (1983): "Funds, factors, and diversification in arbitrage pricing models," *Econometrica*, 51, 1305–1323.

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.

CHANG, Y., AND W. SONG (2009): "Testing for unit roots in small panels with short-run and long-run cross-sectional dependencies," *The Review of Economic Studies*, 76, 903–935.

CHARLES, A., AND O. DARNÉ (2012): "Trends and random walks in macroeconomic time series: a reappraisal," *Journal of Macroeconomics*, 34, 167–180.

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

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

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.

COGLEY, T., AND J. M. NASON (1995): "Effects of the Hodrick-Prescot filter on trend and difference stationary time series: implications for business cycle research," *Journal of Economic Dynamics and Control*, 19, 253–278.

CONNOR, G., AND R. A. KORAJCYK (1986): "Performance measurment with the arbitrage pricing theory," *Journal of financial Economics*, 15, 373–394.

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 tests for autoregressive time series with a unit root," *Econometrica*, 49, 1057–1072.

DOLADO, J. J., T. JENKINSON, AND S. SOSVILLA-RIVERO (1990): "Cointegration and unit roots," *Journal of Economic Surveys*, 4, 249–273.

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

DURLAUF, S. N., AND P. C. B. PHILLIPS (1988): "Trends versus random walks in time series analysis," *Econometrica*, 56, 1433–1354.

ELLIOT, 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 X. Griliches, and M. D. Intriligator, vol. 2, pp. 775–826. North-Holland, Amsterdam.

ENGLE, R. F., AND C. W. J. GRANGER (1987): "Co-integration and error correction: representation, estimation, and testing," *Econometrica*, 55, 251–276.

——— (eds.) (1991): *Long-run Economic Relationships: Readings in Cointegration*. Oxford University Press, Oxford.

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

FERNÁNDEZ-MACHO, F. J. (2000): "A dynamic factor model for economic time series," *Kybernetika*, 33, 583–606.

FISHER, R. A. (1932): *Statistical Methods for Research Workers*. Oliver and Bond, Edinburgh, 4th edn.

FORNI, M., M. HALLIN, M. LIPPI, AND L. REICHLIN (2000): "The generalized dynamic-factor model: identification and estimation," *The Review of Economics and Statistics*, 82, 540–554.

——— (2004): "The generalized dynamic factor model consistency and rates," *Journal of Econometrics*, 119, 231–255.

GEWEKE, J. (1977): "The dynamic factor analysis of economic time-series models," in *Latent Variables in Socio- Economic Models*, ed. by D. J. Aigner, and A. S. Goldberger, pp. 365–383. North-Holland, Amsterdam.

GEWEKE, J., AND K. J. SINGLETON (1981): "Maximum likelihood 'confirmatory' factor analysis of economic time series," *International Economic Review*, 22, 37–54.

GRANGER, C. W. J. (1986): "Developments in the study of sointegrated economic variables," *Oxford Bulletin of Economics and Statistics*, 48, 213–228.

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

HALLIN, M., AND R. LISKA (2007): "Determining the number of factors in the general dynamic factor model," *Journal of the American Statistical Association*, 102, 603–617.

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

HARRIS, R. D. F., AND E. TZAVALIS (1999): "Inference for unit roots in dynamic panels where the time dimension is fixed," *Journal of Econometrics*, 91, 201–226.

HARVEY, A. C., AND A. JAEGER (1993): "Detrending, stylized facts and the business cycle," *Journal of Applied Econometrics*, 8, 231–247.

HATANAKA, M. (1996): *Time-Series-Based Econometrics: Unit Roots and Cointegration*. Oxford University Press, Oxford.

HERWARTZ, H., AND F. SIEDENBURG (2008): "Homogenous panel unit root tests under cross-sectional dependence: finite sample modifications and the wild bootstrap," *Computational Statistics and Data Analysis*, 53, 137–150.

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

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.

JOHANSEN, S. (ed.) (1995): *Likelihood-Based Inference in Cointegrated Vector Autoregressive Models*. Oxford University Press, Oxford.

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

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

JUNGBACKER, B., AND S. J. KOOPMAN (2008): "Likelihood-based analysis for dynamic factor models," Tinbergen Institute Discussion Papers 2008-007/4.

JUSELIUS, K. (2006): *The Cointegrated VAR Model: Methodology and Applications*. Oxford University Press, Oxford.

KAPETANIOS, G. (2010): "A testing procedure for determining the number of factors in approximate factor models with large datasets," *Journal of Business and Economic Statistics*, 28, 397–409.

KARANASOS, M. (1998): "A new method for obtaining the autocovariance of an ARMA model: an exact form solution," *Econometric Theory*, 14, 622–640.

KARLSSON, S., AND M. LÖTHGREN (2000): "On the power and interpretation of panel unit root tests," *Economics Letters*, 66, 249–255.

KIM, D., AND P. PERRON (2009): "Unit root tests allowing for a break in the trend function at an unknown time under both the null and alternative hypotheses," *Journal of Econometrics*, 148, 1–13.

KRUINIGER, H. (2008): "Maximum likelihood estimation and inference methods for the covariance stationary panel AR(1)/unit root model," *Journal of Econometrics*, 144, 447–464.

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.

LOPEZ, C. (2009): "A panel unit root test with good power in small samples," *Econometric Reviews*, 28, 295–313.

LUMSDAINE, R. L., AND D. PAPELL (1997): "Multiple trend breaks and the unit-root hypothesis," *The Review of Economics and Statistics*, 79, 212–218.

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.

MAGNUS, J. R., AND H. NEUDECKER (2001): *Matrix Differential Calculus with Applications in Statistics and Econometrics*. Wiley, Chichester.

MARMOL, F. (1995): "Spurious regressions between I(d) processes," *Journal of Time Series Analysis*, 16, 313–321.

MOLENAAR, P. C. M. (1985): "A dynamic factor model for the analysis of multivariate time series," *Psychometrica*, 50, 181–202.

MOLENAAR, P. C. M., J. G. D. GOIJER, AND B. SCHMITZ (1992): "Dynamic factor analysis of nonstationary multivariate time series," *Psychometrica*, 57, 333–349.

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

NELSON, C. R., AND H. KANG (1984): "Pitfalls in the use of time as an explanatory variable in regression," *Journal of Business and Economic Statistics*, 2, 73–82.

NELSON, C. R., AND C. I. PLOSSER (1982): "Trends and random walks in macroeconomic time series," *Journal of Monetary Economics*, 10, 139–162.

NEWEY, W. K., AND K. D. WEST (1987): "A simple, positive semi-definite, heteroscedasticity and autocorrelation consistent covariance matrix," *Econometrica*, 55, 703–708.

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.

NG, S., AND P. PERRON (1995): "Unit root tests in ARMA models with data-dependent methods for the selection of the truncation lag," *Journal of the American Statistical Association*, 90, 268–281.

——— (2001): "Lag length selection and the construction of unit root tests with good size and power," *Econometrica*, 69, 1519–1554.

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

ONATSKI, A. (2010): "Determining the number of factors from empirical distribution of eigenvalues," *The Review of Economic and Statistics*, 92, 1004–1016.

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

PAPELL, D. H., AND R. PRODAN (2007): "Restricted structural change and the unit root hypothesis," *Economic Inquiry*, 45, 834–853.

PARK, J. Y. (2003): “Bootstrap unit root tests,” *Econometrica*, 71, 1845–1895.
PEÑA, D., AND E. P. BOX (1987): “Identifying a simplifying structure in time series,” *Journal of the American Statistical Association*, 82, 836–843.
PEÑA, D., AND P. PONCELA (2004): “Forecasting with nonstationary dynamic factor models,” *Journal of Econometrics*, 119, 291–321.
——— (2006): “Nonstationary dynamic factor analysis,” *Journal of Statistical Planning and Inference*, 136, 1237–1257.
PERRON, P. (1989): “The great crash, the oil price shock, and the unit root hypothesis,” *Econometrica*, 57, 1361–1401.
PERRON, P., AND S. NG (1996): “Useful modifications to some unit root tests with dependent errors and their local asymptotic properties,” *Review of Economic Studies*, 63, 435–463.
PESARAN, M. H. (2007): “A simple panel unit root test in the presence of cross-section dependence,” *Journal of Applied Econometrics*, 22, 265–312.
——— (2012): “On the interpretation of panel unit root tests,” *Economics Letters*, 116, 545–546.
PHILLIPS, P. C. B. (1986): “Understanding spurious regressions in econometrics,” *Journal of Econometrics*, 33, 311–340.
——— (1987): “Time series regression with a unit root,” *Econometrica*, 55, 277–301.
——— (1998): “New tools for understanding spurious regressions,” *Econometrica*, 66, 1299–1325.
PHILLIPS, P. C. B., AND H. R. MOON (1999): “Linear regression limit theory for nonstationary panel data,” *Econometrica*, 67, 1057–1111.
——— (2000): “Nonstationary panel data analysis: an overview of some recent developments,” *Econometric Reviews*, 19, 263–286.
PHILLIPS, P. C. B., AND S. OULIARIS (1990): “Asymptotic properties of residual based tests for cointegration,” *Econometrica*, 58, 165–193.
PHILLIPS, P. C. B., AND P. PERRON (1988): “Testing for a unit root in time series,” *Biometrika*, 75, 335–346.
PHILLIPS, P. C. B., AND D. SUL (2003): “Dynamic panel estimation and homogeneity testing under cross section dependence,” *Econometrics Journal*, 6, 217–259.
PHILLIPS, P. C. B., AND Z. XIAO (1998): “A primer on unit root testing,” *Journal of Economic Surveys*, 12, 423–469.
RAO, C. R. (1948): “Large sample tests of statistical hypotheses concerning several parameters with application to problems of estimation,” *Preceedings of the Cambridge Philosophical Society*, 44, 50–57.
ROBINSON, P. M. (1994): “Efficient tests of nonstationary hypotheses,” *Journal of the American Statistical Association*, 89, 1420–1437.
RUDEBUSCH, G. D. (1992): “Trends and random walks in macroeconomic time series: a re-examination,” *International Economic Review*, 33, 661–680.
SAID, S. E., AND D. A. DICKEY (1984): “Testing for unit roots in autoregressive-moving average models of unknown order,” *Biometrika*, 71, 599–607.
——— (1985): “Hypothesis testing in ARIMA(p, 1, q) models,” *Journal of the American Statistical Association*, 80, 369–374.

SARGAN, J. D., AND A. BHARGAVA (1983): “Testing residuals from least squares regression for being generated by the Gaussian random walk,” *Econometrica*, 51, 153–174.
SARGENT, T. J., AND C. A. SIMS (1977): “Business cycle modeling without pretending to have too much a priori economic theory,” in *New Methods in Business Cycle Research: Proceedings from a Conference*, ed. by C. A. Sims, pp. 45–110. Federal Reserve Bank of Minneapolis.
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.
SHIN, D. W., AND S. KANG (2006): “An instrumental variable approach for panel unit root tests under cross-sectional dependence,” *Journal of Econometrics*, 134, 215–234.
SHIN, D. W., AND S. PARK (2010): “Robust panel unit root tests for cross-sectionally dependent multiple time series,” *Computational Statistics and Data Analysis*, 54, 2801–2813.
SILVEY, S. D. (1959): “The Lagrange multiplier test,” *The Annals of Mathematical Statistics*, 30, 389–407.
SIMS, C. A. (1980): “Macroeconomics and reality,” *Econometrica*, 48, 1–48.
SMEEKES, S. (2010): “Bootstrap sequential tests to determine the stationary units in a panel,” METEOR Research Memoranda 11-003.
SOLO, V. (1984): “The order of differencing in ARIMA models,” *Journal of the American Statistical Association*, 79, 916–921.
STOCK, J. H. (1987): “Aymptotic properties of least-squares estimators of cointegrating vectors,” *Econometrica*, 55, 1035–1056.
——— (1994): “Unit roots, structural breaks and trends,” in *Handbook of Econometrics*, ed. by R. F. Engle, and D. L. McFadden, vol. 4, pp. 2739–2841. Elsevier, Amsterdam.
STOCK, J. H., AND M. W. WATSON (1988): “Testing for common trends,” *Journal of the American Statistical Association*, 83, 1097–1107.
——— (1989): “New indexes of coincident and leading economic indicator,” *NBER Macroeconomics Journal*, 4, 351–394.
——— (2001): “Vector autoregressions,” *Journal of Economic Perspectives*, 15, 101–115.
——— (2002): “Macroeconomic forecasting using diffusion indexes,” *Journal of Business and Economic Statistics*, 20, 147–162.
STOICA, P., AND M. JANSSON (2009): “On maximum likelihood estimation in factor analysis - an algebraic derivation,” *Signal Processing*, 89, 1260–1262.
STRAUSS, J., AND T. YIGIT (2003): “Shortfalls of panel unit root testing,” *Economics Letters*, 81, 309–313.
SUL, D. (2009): “Panel unit root tests under cross section dependence with recursive mean adjustment,” *Economics Letters*, 105, 123–126.
URBAIN, J.-P., AND J. WESTERLUND (2006): “Spurious regression in nonstationary panels with cross-unit cointegration,” METEOR Research Memoranda 06-057.
VAN DER LEEUW, J. (1994): “The covariance matrix of ARMA errors in closed form,” *Journal of Econometrics*, 63, 397–405.

VENTOSA-SANTAULÀRIA, D. (2009): "Spurious regression," *Journal of Probability and Statistics*, 2009, 1–27.
WALD, A. (1943): "Tests of statistical hypotheses concerning several parameters when the number of observations is large," *Transactions of the 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.
WATSON, M. W., AND R. F. ENGLE (1983): "Alternative algorithms for the estimation of dynamic factor, mimic and varying coefficient regression models," *Journal of Econometrics*, 23, 385–400.
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 the Analysis of Stationary Time Series*. Almqvist & Wiksell, Uppsala.
YULE, G. U. (1926): "Why do we sometimes get nonsense-correlations between time series? - a study in sampling and the nature of time series," *Journal of the Royal Statistical Society*, 89, 1–63.
ZIVOT, E., AND D. W. K. ANDREWS (1992): "Further evidence on the great crash, the oil-price shock, and the unit-root hypothesis," *Journal of Business and Economic Statistics*, 10, 251–270.

Acta Universitatis Upsaliensis

Digital Comprehensive Summaries of Uppsala Dissertations from the Faculty of Social Sciences 90

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.

Distribution: publications.uu.se
urn:nbn:se:uu:diva-207296

ACTA
UNIVERSITATIS
UPSALIENSIS
UPPSALA
2013