

Testing non-stationary panel data with persistent cross-sectional dependence

Citation for published version (APA):

Gengenbach, C. (2009). *Testing non-stationary panel data with persistent cross-sectional dependence*. Universitaire Pers Maastricht. <https://doi.org/10.26481/dis.20090903cg>

Document status and date:

Published: 01/01/2009

DOI:

[10.26481/dis.20090903cg](https://doi.org/10.26481/dis.20090903cg)

Document Version:

Publisher's PDF, also known as Version of record

Please check the document version of this publication:

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

[Link to publication](#)

General rights

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

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

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

www.umlib.nl/taverne-license

Take down policy

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

repository@maastrichtuniversity.nl

providing details and we will investigate your claim.

Testing Non-Stationary Panel Data
with
Persistent Cross-Sectional Dependence

Christian Gengenbach

© Christian Gengenbach, 2009

All rights reserved. No part of this publication may be reproduced, stored in a retrieval system, or transmitted in any form, or by any means, electronic, mechanical, photocopying, recording or otherwise, without the prior permission in writing from the author.

This book was typeset by the author using L^AT_EX.

Published by Universitaire Pers Maastricht

ISBN: 978-90-5278-854-8

Printed in The Netherlands by Datawyse Maastricht

Testing Non-Stationary Panel Data
with
Persistent Cross-Sectional Dependence

PROEFSCHRIFT

ter verkrijging van de graad van doctor
aan de Universiteit Maastricht,
op gezag van de Rector Magnificus,
Prof. mr. G.P.M.F. Mols,
volgens het besluit van het College van Decanen,
in het openbaar te verdedigen op
donderdag 3 september 2009 om 16.00 uur

door

Christian Carl Gengenbach

Promotores:

Prof. dr. F.C. Palm

Prof. dr. J.R.Y.J Urbain

Beoordelingscommissie:

Prof. dr. M.A. Carree (voorzitter)

Prof. dr. J. Breitung (University of Bonn)

Prof. dr. P. Mohnen

Dit onderzoek werd gefinancierd door de Maastricht Research School of Economics of Technology and Organizations (METEOR).

Acknowledgements

Working on my Ph.D. over the past years has been an exciting experience. There has been the joy when interesting results had been obtained after some hard work and when they were accepted for presentation or publication.. But there has also been frustration when results turned out uninteresting or no results could be obtained at all, and I realized that I had wasted time and energy running around in circles. However, that is also part of research, and looking back I have to say that the positive moments by far outnumber the negative ones.

I would like to take the opportunity here to thank a number of people whose continued support helped me a lot over the recent years. First and foremost I would like to thank my supervisors, Franz Palm and Jean-Pierre Urbain. I have enjoyed our collaboration tremendously over the years, not only during our stimulating and at times even merry meetings. I am also grateful for the fact that both your doors are always open to discuss questions, problems or ideas. Furthermore, I would like to mention that Jean-Pierre not only has to accept main responsibility for me pursuing a Ph.D., since he contacted me about that possibility towards the end of my master, but his teaching was influential in my choice of specializing in econometrics in the first place.

I would like to thank the members of the reading committee, Martin Carree, Jörg Breitung and Pierre Mohnen, for their time and careful reading of the manuscript. Jörg also deserves thanks for comments and discussion at earlier stages of the work. I thank all my colleagues at the Department of Quantitative Economics for providing an extremely pleasant working atmosphere. In particular, I would like to thank Karin van den Boorn and Haydeé Hallmanns for their administrative support and Stephan Smeekes, who personally contributed to this thesis in form of the Dutch summary. I am also grateful to all colleagues I met at conferences and workshops and who provided stimulating discussion and comments. Especially, I would like to mention Joakim Westerlund, who is more than a brilliant researcher, and who contributed to this thesis as a coauthor.

Last but not least I would like to thank my family and friends. My parents Rolf and Renate and my sister Ulrike have always supported me in all possible ways. I would like to thank my grand mother Ruth and my grand aunt Christel, who both passed away before the completion of this thesis. I am not going to try to name all friends who deserve credit here. However, I would like to thank Sascha Wolf, who studied with me in Maastricht since the start. Finally, I would like to thank Timm and Lukas Roegler for their long-lasting and close friendship.

Christian Gengenbach

Maastricht, 2009

Contents

Acknowledgements	v
1 Introduction	1
1.1 Objective of the thesis	1
1.2 Motivation	2
1.3 Contributions and Thesis Structure	4
2 Panel Unit Root Tests in the Presence of Cross-Sectional Dependencies: Comparison and Implications for Modelling	9
2.1 Introduction	9
2.2 Testing for unit roots in panel data with unobserved common factors	12
2.2.1 Pesaran (2007): A dynamic panel model with one common factor	13
2.2.2 Moon and Perron (2004): A dynamic panel model with K common factors	14
2.2.3 Bai and Ng (2004b): A common factor model with unobserved common and idiosyncratic components of unknown order of integration.	16
2.2.4 Alternative panel unit root tests in the presence of cross-sectional dependencies	20
2.2.5 Differences and similarities	23
2.3 Small sample performance: Monte Carlo results	26
2.3.1 Monte Carlo simulation setup	26
2.3.2 Monte Carlo results	30
2.4 An illustrative application: Testing for PPP using the new approaches	33
2.5 Conclusion	36
2.A Tables	38

3	Cointegration Testing in Panels with Common Factors	45
3.1	Introduction	45
3.2	The Model	47
3.3	The behavior of panel residual based tests	52
3.3.1	Kao (1999)	52
3.3.2	Pedroni (1999)	54
3.4	A two-step procedure to test for (no)cointegration in the presence of common factors .	57
3.5	Some Monte Carlo Evidence	60
3.6	Conclusions	62
3.A	Tables	64
3.B	Appendix	66
3.B.1	Lemma 3.B.1 to 3.B.3	66
3.B.2	Proof of Propositions 3.1 and 3.2	69
4	Panel Error Correction Testing with Global Stochastic Trends	77
4.1	Introduction	77
4.2	Model representation	79
4.3	Individual tests for no error correction	84
4.3.1	Observed factors	84
4.3.2	Unobserved factors	90
4.3.3	Critical values	92
4.4	Panel tests for no error correction	93
4.4.1	The tests	93
4.4.2	Critical values	95
4.5	Monte Carlo simulations	95
4.6	Empirical Applications	98
4.6.1	The Fisher effect	99
4.6.2	The monetary exchange rate model	101
4.7	Conclusions	103
4.A	Tables	104
4.B	Appendix	118
4.B.1	Proof of Theorem 4.1.	118
4.B.2	Proof of Theorem 4.2.	120

4.B.3	Proof of Theorem 4.3	124
5	A Panel Cointegration Study of the Euro Effect on Trade:	
	Revisiting Bun and Klaassen (2007)	129
5.1	Introduction	129
5.2	Data and previous studies	131
5.3	Panel unit root and cointegration tests	133
5.3.1	Panel Unit Root Tests	134
5.3.2	Panel Cointegration Tests	138
5.4	Estimation of the gravity equation	140
5.5	Conclusion	145
6	Conclusion and Further Research	147
6.1	Summary, limitations and further research	147
	Bibliography	151
	Nederlandse samenvatting	159
	Curriculum Vitae	163

List of Tables

2.1	Finite sample (average) rejection rates for DGP (2.24) with I(1) common factor(s) and I(1) idiosyncratic components.	39
2.2	Finite sample (average) rejection rates for DGP (2.24) with I(1) common factor(s) and I(0) idiosyncratic components.	40
2.3	Finite sample (average) rejection rates for DGP (2.24) with I(0) common factor(s) and I(1) idiosyncratic components.	41
2.4	Finite sample (average) rejection rates for DGP (2.24) with I(0) common factor(s) and I(0) idiosyncratic components.	42
2.5	Finite sample (average) rejection rates for DGP (2.25) with I(0) common factor(s) I(0) idiosyncratic components.	43
2.6	Unit root test statistics for individual series of real exchange rates.	44
2.7	Pooled unit root test statistics panels of real exchange rates.	44
3.1	$k = 2$ common factors; Non-stationary common factors F_t with I(0) or I(1) idiosyncratic component $E_{i,t}$	64
3.2	$k = 2$ common factors; Cointegration in either F_t or $E_{i,t}$ or both. NC denotes no cointegration, C cointegration.	65
4.1	Critical values and moments for the individual t -tests.	104
4.2	Critical values and moments for the individual Wald tests.	105
4.3	Critical values for the panel t -tests.	106
4.4	Critical values for the pooled Wald tests.	107
4.5	Size and size-adjusted power in Model 1 with no deterministic components.	109

4.6	Size and size-adjusted power in Model 2 with an unrestricted constant.	111
4.7	Size and size-adjusted power in Model 3 with unrestricted constant and trend terms. .	113
4.8	Cointegration test results for the Fisher effect.	115
4.9	Cointegration test results for the monetary exchange rate model.	116
4.10	Descriptive statistics for the common factors in the monetary exchange rate model. . .	117
5.1	Estimation results Bun and Klaassen (2007)	132
5.2	Panel unit root and cointegration tests	139
5.3	Parameter estimates for static and dynamic models	143

Introduction

1.1 Objective of the thesis

The analysis of non-stationary panel data is a relatively young field of econometric research. It emerged with the growing availability of large macro data sets which for example contain information on countries over several decades. This contrasts with traditional (micro) panel data sets where the time dimension usually has been small. The larger time dimension made it possible to adopt methods advanced for the analysis of time-series data to panels. While early studies did not pay much attention to cross-sectional dependence in the data, it was soon pointed out that it is a prevalent feature of economic panel data and that cross-sectional dependence can have an adverse effect on the proposed econometric methods if not properly accounted for. Several alternatives have been proposed in the literature to allow for cross-sectional dependence.

In this thesis we focus on approaches that employ a common factor structure to model cross-sectional dependence. The objectives are to study tests for non-stationary panels with persistent common factors. In particular, we investigate the implications of non-stationary common factors on panel data models. They provide an elegant way to model cross-member cointegration (Chapters 2 and 3), which has been shown to be a feature of some economic panels. We study differences and communalities of several recently advance panel unit root tests which allow for common factors. The considered methods differ in the way the common factors are allowed to influence the dynamic properties of the data and are shown to test different data components (Chapter 2). We also consider spurious regression in panels with non-stationary common factors. We show that methods developed for cross-sectionally independent panels are not suited to test for no cointegration in that case and propose a testing procedure based on defactoring the data (Chapter 3). Furthermore, we derive a representation theorem for a cointegrated panel with non-stationary common factors. We use the conditional error-correction model as a base to propose an alternative no cointegration test (Chapter 4). Finally, we apply discussed methods in an empirical study to investigate the effect of the introduction of the Euro on bilateral trade (Chapter 5).

1.2 Motivation

In recent years an increasing number of researchers in economics and business administration have relied on panel data methods to analyze the dynamic behavior of individuals, firms, sectors, countries, etc. One recently developed research area examines the properties of non-stationary panel data in which both the number of entities (N) and the number of time periods (T) are large. In traditional panel data N is usually large but T is small. When T is large, it becomes feasible to study more interesting (and more complex) issues, in particular to learn about the dynamic features of the data and allow for more general heterogeneity. Two examples from macroeconomics for the application of panel unit root and cointegration tests are the analysis of purchasing power parity (PPP) and the study of growth convergence. To test the validity of purchasing power parity, one should examine the properties of the (log) real exchange rate (strong PPP), or the joint behavior of the (log) nominal exchange rate and (log) price differentials (weak PPP). The former is done by testing for a unit root in the series, while the later requires tests for cointegration to examine whether the stochastic trends in the data series are the same. Tests of growth convergence study the behavior of real per capita output growth.

Methods used to analyze data when N and T are large combine elements of traditional panel data/cross-sectional analysis and of non-stationary time-series analysis. In many empirical application pooling information of several individuals arises naturally in the context of the underlying theory. PPP for example is at least defined as a bilateral relationship but should also hold for larger groups of countries. In the theory of growth convergence, countries with similar underlying fundamentals should converge to similar steady state growth paths. Initially, much attention in the literature has focused on the potential gains from pooling cross-sectional observations under the assumption of cross-sectional independence, while loosening the assumption of parameter homogeneity. In unit root/cointegration cases, the gains appear to be substantial because pooling converts non-standard into more traditional limit theory by cross-sectional averaging, as shown for example by Phillips and Moon (1999). This favorable property can be preserved when a deterministic common time effect (CTE) is included to capture some contemporaneous dependencies. More general forms of cross-sectional correlation (dynamic, long-run) are harder to model due to the lack of natural ordering, a problem known from traditional panel data theory which is amplified in dynamic panels.

In early approaches to panel unit root testing, the often unrealistic assumption of cross-sectional independence is made. For instance, the tests proposed in Levin, Lin, and Chu (2002) and Im, Pesaran, and Shin (2003), denoted respectively as LLC and IPS, assume cross-sectional independence,

but allow for heterogeneity of the form of individual deterministic effects (constant and/or linear time trend) and heterogenous serial correlation structure of the error terms. Both methods test the same null hypothesis of non-stationarity, but differ in terms of the considered alternative and hence, in the way information is pooled. Levin et al. (2002) study balanced panels with N cross-sectional units and T time series observations. They assume a homogenous first order autoregressive parameter and their test is based on the pooled t -statistic of the estimator. Im et al. (2003) allow unbalanced panels with N cross sectional units and T_i time series observations for each $i = 1, \dots, N$. They propose a standardized average of individual ADF statistics to test the pooled unit root null hypothesis against a heterogenous alternative. Maddala and Wu (1999) and Choi (2001) propose Fisher type tests based on pooling p -values. The methods assume cross-sectional independence among panel units except for a common time effect. In that case, the derived results remain valid if cross-sectional averages are subtracted from the data.

The effect of persistent cross-sectional dependence on panel unit root tests has been recently analyzed and documented in some detail in the literature. As shown by Monte Carlo simulations (Banerjee, Marcellino, and Osbat, 2005) or by asymptotic analysis (Lyhagen, 2000; Pedroni and Urbain, 2001), the standard (LLC or IPS) panel unit root tests are severely affected in that either they display dramatic size distortions or even worse can be shown to diverge with the cross-sectional dimension of the panel. Several studies have addressed the issue of cross-sectional correlation in non-stationary panels. Bai and Ng (2004b), Moon and Perron (2004) and Pesaran (2007) all use common factor structures to model cross-sectional correlation in a panel, but the assumed data generating processes (DGP), as well as the developed estimation and testing procedures differ in important ways. Moon and Perron (2004) and Pesaran (2007) use an autoregressive representation for the observed data and assume common factors to be present in the unobserved (stationary) error terms. They suggest different methods to effectively remove the common factors from the statistics used to test a pooled unit root hypothesis. The DGP assumed by Bai and Ng (2004b) models the observed data as the sum of unobserved common and idiosyncratic components with (possibly) heterogeneous dynamic properties. They propose a procedure to estimate the unobserved components and test them for unit roots separately. Breitung and Das (2008) propose an FGLS and robust t test to test for a unit root in a cross-sectionally dependent panel. Sul (2007) proposes a recursive mean adjusted FGLS test for a panel unit root and a recursive mean adjusted t test for the cross-sectional average to test for a unit root in the common component if the data permits a common factor structure. While the later two studies allow for persistent cross-sectional dependence they do not model this explicitly as

a common factor structure.

The case of tests for the null of no-cointegration, has seen some attention recently. Kao (1999) and Pedroni (1999, 2004a) propose residual based panel cointegration tests for independent panels, while Westerlund (2007) investigates panel error-correction tests. Banerjee, Marcellino, and Osbat (2004) conduct an extensive Monte Carlo study where they conclude that while all statistics investigated (residual-based tests or likelihood based trace-type test) are affected, the presence of cross-member cointegration appears much less harmful for single-equation tests than for the panel version of the Johansen test. In many cases, the tests are affected by the presence of cointegration between members in such a way that these tests cannot discriminate between cointegration across members and cointegration within, that is for a single member of the panel. Banerjee and Carrion-i Silvestre (2005, 2006) and Bai and Kao (2006) study tests for panel no-cointegration with cross-sectional dependence. These studies consider residual-based tests for a single cointegration relationship, where the error term of the cointegrating equation follows a common factor structure as in Bai and Ng (2004b). Urbain and Westerlund (2008) on the other hand studies analytically the issue of spurious regression in panels when the units are cointegrated along the cross-sectional dimension, i.e. when there is cross-member cointegration. In contrast to the spurious regression result for independent panel studied by Phillips and Moon (1999), Pedroni (1995) or Kao (1999), in most of the cases considered these estimators are not consistent and actually converge to non-degenerate limiting distribution once the observed non-stationarity is generated by a reduced number of common stochastic trends.

In this thesis we focus on approaches that model the cross-sectional dependence by using a common factor structure. Alternatively, Chang (2002) proposes a non-linear IV panel unit root test allowing for general cross-sectional dependence. Based on the non-linear IV estimator of Chang (2002), Demetrescu and Tarcolea (2005) propose a residual based no cointegration test. Bootstrap panel unit root tests are considered by for example Chang (2004) or more recently Palm, Smeekes, and Urbain (2008). As each chapter of this thesis contains a discussion of the relevant literature, no review is provided here. Extensive overviews of the recent literature are provided by Breitung and Pesaran (2008) and in Baltagi (2008, Chapter 12). Earlier literature overviews are provided by Banerjee (1999) and Baltagi and Kao (2000).

1.3 Contributions and Thesis Structure

This thesis contributes to the econometric literature on non-stationary panel data in several regards. In Chapter 2 we provide a detailed comparison of several second generation panel unit root tests that

model cross-sectional dependence in the data using a common factor structure. The cross-sectionally augmented unit root tests proposed by Pesaran (2007) are designed for cases where cross-sectional dependence is due to a single common factor. The Moon and Perron (2004) tests which use defactored data is similar in spirit but can account for multiple common factors. The Bai and Ng (2004b) tests allow to determine the source of non-stationarity by testing for unit roots in the common factors and the idiosyncratic factors separately. Breitung and Das (2008) and Sul (2007) propose panel unit root tests when cross-section dependence is present possibly due to common factors, but the common factor structure is not fully exploited. We compare the testing procedures in terms of similarities and difference in the data generation process, tests, null and alternative hypotheses considered and compare the small sample properties of the tests in models with up to two common factors using Monte Carlo results. While most considered tests allow to detect a unit root in the idiosyncratic component of the data, only the approaches of Bai and Ng (2004b) and Sul (2007) can detect stochastic trends in the common factors. Consequently, only the later two approaches allow to detect cross-member cointegration in a panel. Furthermore, we provide an application which illustrates the use of the tests and finally it discusses the use of the tests in modelling in general.

In Chapter 3, we consider a spurious regression model for a panel with non-stationary common factors. We assume that the observed variables permit a common factor representation as the one proposed by Bai and Ng (2004b) for panel unit root tests. Our model allows to distinguish between two different cases are considered that we believe are of theoretical and empirical relevance: (i) the case where the observed non-stationarity in the variables originates from cross-sectional common trends only (*cross-member cointegration*); (ii) the case where we have both cross-sectional common and idiosyncratic stochastic trends. In the later case rejection of the null for both data components is a necessary but not a sufficient condition for cointegration and we discuss the required homogeneity restrictions on the cointegrating vectors resulting from the presence of common factor cointegration. Furthermore, we study the asymptotic behavior of some existing, residual-based panel no-cointegration, as suggested by Kao (1999) and Pedroni (1999, 2004a). Under the DGP used, the test statistics are shown to be no longer asymptotically normal, and convergence occurs at rate T rather than \sqrt{NT} as for independent panels. Following the work of Bai and Ng (2004b), we then examine the possibilities to test for various forms of no-cointegration by extracting the common factors and individual components from the observed data directly and then test the estimated components separately.

Chapter 4 considers a cointegrated panel data model with common factors. We develop alternative

representations of a cointegrated panel that allows for the possibility of non-stationary common factors. Starting from the triangular representation of the system used by for example Bai, Kao, and Ng (2009), we derive a Granger type representation theorem similar the the one obtained by Cappuccio and Lubian (1996) for a single time series. The conditional error correction representation is obtained, which is used as a basis for developing two new tests for the null hypothesis of no error correction. In particular we consider panel versions of the t -test as proposed by Banerjee, Dolado, and Mestre (1998) and Boswijk (1994) and the Wald test of Boswijk (1994). We show that the individual specific tests are asymptotically nuisance parameter free and only depend on the number of non-stationary variables in the system. However, they are not cross-sectionally independent due to the common factors. Nevertheless, the average of the test statistics converges to a random variable with a distribution which, while not analytically tractable, can be simulated. This makes pooling possible in spite of the cross-sectional dependence. We investigate the finite sample performance of the proposed tests in a Monte Carlo experiment and compare them to the tests proposed by Westerlund (2007). We also present two empirical applications of the new tests.

This thesis also contributes to the empirical literature of gravity models of bilateral trade and common currency effect on trade. In Chapter 5 we revisit Bun and Klaassen (2007) for an investigation of the impact of the introduction of the Euro on bilateral trade. Although there is strong evidence of a positive common currency effect on trade (see e.g. Rose and Stanley, 2005, for a meta analysis of published studies), there is an ongoing discussion of the actual magnitude of the effect in the empirical literature, starting with the pioneering work of Rose (2000). Accounting for deterministic trends in the residuals of the gravity equation Bun and Klaassen (2007) estimate an Euro effect of about 3%, smaller than previous estimates in the range of 5% to 40%. We revisit their data, which contains observations on the 15 members of the European Union prior to the 2004 expansion of the union as well as Norway, Switzerland, Canada, Japan and the US over a time span from 1967 to 2002, using methods recently advanced in the analysis of non-stationary panel data with cross-sectional dependence. Using several panel unit root tests we find strong evidence that (the log of) bilateral trade, as well as the product of GDP and GDP per capita have unit roots. However, we find cointegration between these variables using the cointegration test of Gengenbach, Palm, and Urbain (2006) and the error correction tests proposed by Gengenbach, Westerlund, and Urbain (2008). Employing the common correlated effects (CCEP) estimator of Pesaran (2006) and the continuously updated (CUP) estimator of Bai et al. (2009), we obtain estimates of the cointegrating vector and estimates of the Euro effect on bilateral trade. Our estimates vary between models and estimators but seem

to support the findings of Bun and Klaassen (2007).

Chapter 6 summarizes the main findings of the thesis, discusses the limit of the current work and presents some outlines for further research.

Panel Unit Root Tests in the Presence of Cross-Sectional Dependencies: Comparison and Implications for Modelling¹

2.1 Introduction

For many economic applications it is important to know whether an observed time series is stationary or non-stationary. For example, to test the validity of Purchasing Power Parity (PPP) one should examine the properties of the real exchange rates. One needs to look at the behavior of differences in real per capita output growth to test for growth convergence. Therefore, unit root tests are an important tool for econometric analysis. However, univariate unit root tests are known to lack power for samples of small or medium size. Unfortunately, for many macroeconomic variables data is available only for a small sample span. But, since studies investigating for example PPP or growth convergence are concerned with the behavior of similar data series from several countries, a natural attempt is to pool the information contained in a data panel. Indeed, that is the general idea of panel unit root tests, and they only differ in the way the information is pooled. Unfortunately, simple pooling is only valid if the units of the panel are independent of each other and sufficiently homogenous. Independence however is unlikely to hold in most applications of panel unit root tests. In cross-country analysis there might be common influences to all panel members, e.g. in PPP-studies one usually uses a common numeraire country to calculate real exchange rates.

In early approaches to panel unit root testing, the often unrealistic assumption of cross-sectional independence is made. For instance, the tests proposed in Levin et al. (2002) and Im et al. (2003), denoted respectively as LLC and IPS, assume cross-sectional independence, but allow for heterogeneity of the form of individual deterministic effects (constant and/or linear time trend) and heterogeneous

¹This Chapter is based on Gengenbach, Palm, and Urbain (2009).

serial correlation structure of the error terms. Both methods test the same null hypothesis of non-stationarity, but differ in terms of the considered alternative and hence, in the way information is pooled. Levin et al. (2002) study balanced panels with N cross-sectional units and T time series observations. They assume a homogenous first order autoregressive parameter and their test is based on the pooled t-statistic of the estimator. Im et al. (2003) allow unbalanced panels with N cross sectional units and T_i time series observations for each $i = 1, \dots, N$. They propose a standardized average of individual ADF statistics to test the pooled unit root null hypothesis against a heterogenous alternative. Both methods assume cross-sectional independence among panel units except for a common time effect. In that case, the derived results remain valid if cross-sectional averages are subtracted from the data.

Attention has been drawn recently to the assumption of cross-sectional independence on which the asymptotic results of both procedures rely. Among the first to analyze the effect of cross-sectional correlation on panel unit root tests was O'Connell (1998). Using Monte Carlo simulations he shows that the LLC test severely suffers from cross-correlation in terms of increased size and reduced power. He suggests using FGLS estimation to overcome this problem. However, estimation of the error covariance matrix becomes infeasible as N and T grow large. Flôres, Jorion, Preumont, and Szafarz (1999) use SUR estimation of the (possibly heterogenous) AR parameter, and determine critical values for their test via Monte Carlo simulations. Their methodology has the disadvantage that it requires extensive simulations to determine critical values and does only account for contemporaneous cross-sectional correlation. In simulation studies, Banerjee et al. (2004, 2005) assess the finite sample performance of panel unit root and cointegration tests when panel members are cross-correlated or even cross-sectionally cointegrated². Their finding is, that all methods experience size distortions when panel members are cointegrated. This means that procedures such as the LLC or IPS test would over-reject the non-stationarity null when there are common sources of non-stationarity. This is analytically confirmed by Lyhagen (2000).

Recently, panel unit root tests have been proposed model cross-sectional correlation using a common factor representation of the data, or robust methods allowing for a general form of cross-sectional dependence, e.g. Chang (2002). The purpose of this chapter is to study some of the new methods which assume a factor structure and compare them in terms of modelling, assumptions and statistical properties of the test statistics. A Monte Carlo study assesses the finite sample properties of the test statistics in terms of size and power in order to compare them.

²The notation of panel cointegration tests refers to tests for cointegration between several variables of one panel member, in contrast to cointegration between panel members.

Three different newly proposed unit root tests will be considered. Pesaran (2007) suggests a cross-sectionally augmented Dickey-Fuller (CADF) test where the standard DF regressions are augmented with cross-sectional averages of lagged levels and first differences of the individual series. He also considers a cross-sectional augmented IPS (CIPS) test, which is a simple average of the individual CADF-tests. The data generating process (DGP) is a simple dynamic linear heterogeneous panel data model. The error term is assumed to have an unobserved one-common-factor structure accounting for cross-sectional correlation and an idiosyncratic component.

A second type of panel unit root tests has been proposed by Moon and Perron (2004). We consider two feasible t-statistics proposed by them to test for unit roots in a dynamic panel model allowing for fixed effects. The stationary error term follows a K -unobserved-common-factor model to which an idiosyncratic shock is added. The t-statistics are based on appropriately standardized pooled estimators of the first order serial correlation coefficients of the data series.

The third type of panel unit root tests has been proposed by Bai and Ng (2004b). In their “Panel Analysis of Non-stationarity in Idiosyncratic and Common Components” (PANIC) approach the space spanned by the unobserved common factors and idiosyncratic disturbances is consistently estimated without knowing whether they are stationary or integrated. Next, the number of independent stochastic trends driving the common factors is determined. Both individual and pooled individual statistics are proposed to test separately for unit roots in the unobserved common and idiosyncratic components of the data instead of the observed series. Both common and idiosyncratic components may be stationary or integrated.

These three panel unit root tests have been selected for the following reasons. First of all, the model specifications are sufficiently close to each other and some are partly nested to allow for comparison. At the same time, the test procedures differ in important ways to make it interesting to compare their properties and provide some guidelines for the empirical analysis of non-stationary panel data. Second, in all the approaches an unobserved common factor structure is assumed to explain cross-sectional correlation. Common factor structures have several advantages. Statistical estimation and testing methods, and selection procedures for the number of factors are at the disposal of the empirical researcher. The statistical properties of these procedures are in general well-understood. These method recently experienced a revival in the common features literature. Using common factors to explain cross-sectional correlation allows to deal with the curse of dimensionality problem in a natural way, which has been found to work well in empirical econometrics. Finally, common factor structures often result from theoretical considerations in economics. For instance the

CAPM and the APT models used in finance are common factor models, and many intertemporal microeconomic models imply factor structures for the data.

The chapter is organized as follows: In Section 2.2 we present the DGPs used in the three approaches mentioned above. Wherever one DGP is nested in another this will be pointed out. Also, the testing procedures used will be described in some detail. We briefly discuss which features of the three approaches will be compared. In Section 2.3, we present the results of an extensive simulation study which compares the three approaches to panel unit root testing for models with factor structures and two panel unit root tests proposed by Breitung and Das (2008) and by Sul (2007) which do not fully exploit factor structure. A PPP test using the described methods is presented in Section 2.4 as an illustrative example. Section 2.5 is devoted to conclusions. In particular, the implications of the findings for modeling in practice will be discussed.

2.2 Testing for unit roots in panel data with unobserved common factors

This section describes three approaches to panel unit root testing in the presence of cross-sectional correlation which employ factor models. In particular, the methods proposed by Pesaran (2007), Moon and Perron (2004) and Bai and Ng (2004b) will be presented. For reasons of comparison, it also briefly describes the panel unit root tests by Breitung and Das (2008) and by Sul (2007) which assume a factor structure but do not fully exploit it.

The factor structure used by all approaches is a convenient form to model cross-correlation, or even cointegration between panel members. Therefore, the (for pooled testing necessary) assumption of independence between the individual specific components of the data is far less restrictive than the assumption of independent cross-sections, underlying the IPS and LLC test.

A note on notation: Throughout this chapter, M is used to denote a finite, generic constant. For a matrix A , $A > 0$ denotes that A is positive definite. Common factors which are denoted by f_t are always assumed to be stationary. Common factors denoted by F_t result from an autoregressive transformation of f_t . F_t has a unit root when there is a unit root in the autoregression. Whenever we refer to nonstationary common factors, this means nonstationarity of F_t .

2.2.1 Pesaran (2007): A dynamic panel model with one common factor

For a panel of observed data with N cross-sectional units and T time series observations, Pesaran (2007) uses a simple dynamic linear heterogenous model

$$Y_{i,t} = (1 - \delta_i)\mu_i + \delta_i Y_{i,t-1} + u_{i,t}, \quad i = 1, \dots, N, \quad t = 1, \dots, T, \quad (2.1)$$

with given initial values $Y_{i,0}$ and a one-factor structure for the disturbance

$$u_{i,t} = \lambda_i f_t + e_{i,t}. \quad (2.2)$$

Considering serially uncorrelated disturbances, the idiosyncratic components, $e_{i,t}$, $i = 1, \dots, N$, $t = 1, \dots, T$ are assumed to be independently distributed both across i and t , have zero mean, variance σ_i^2 , and finite fourth-order moment. The common factor f_t is serially uncorrelated with mean zero and constant variance σ_f^2 , and finite fourth-order moment. Without loss of generality, σ_f^2 is set equal to one. $e_{i,t}$, λ_i and f_t are assumed to be mutually independent for all i and t .

It is convenient to write (2.1) and (2.2) as

$$\Delta Y_{i,t} = \alpha_i - (1 - \delta_i)Y_{i,t-1} + \lambda_i f_t + e_{i,t}, \quad (2.3)$$

where $\alpha_i = (1 - \delta_i)\mu_i$ and $\Delta Y_{i,t} = Y_{i,t} - Y_{i,t-1}$. The unit root hypothesis considered by Pesaran (2007), $\delta_i = 1$ for all i is tested against the possibly heterogenous alternative $\delta_i < 1$ for $i = 1, \dots, N_1$, $\delta_i = 1$ for $i = N_1 + 1, \dots, N$. Pesaran (2007) assumes that $\frac{N_1}{N}$, the fraction of the individual processes that is stationary, is non-zero and tends to some fixed value κ such that $0 < \kappa \leq 1$ as $N \rightarrow \infty$.

It is important to notice that any non-stationarity of the observations $Y_{i,t}$ in the setting considered by Pesaran (2007) is due to the presence of a unit root in the autoregressive part of (2.1), i.e. $\delta_i = 1$. For the unit root null hypothesis considered by Pesaran (2007), he proposes a test based on the t-ratio of the OLS estimate \hat{b}_i in the following cross-sectionally augmented DF (CADF) regression

$$\Delta Y_{i,t} = a_i + b_i Y_{i,t-1} + c_i \bar{Y}_{t-1} + d_i \Delta \bar{Y}_t + \epsilon_{i,t}, \quad (2.4)$$

where $\bar{Y}_t = \frac{1}{N} \sum_{i=1}^N Y_{i,t}$, $\Delta \bar{Y}_t = \frac{1}{N} \sum_{i=1}^N \Delta Y_{i,t}$, and $\epsilon_{i,t}$ is the regression error.

The cross-sectional averages, \bar{Y}_{t-1} and $\Delta \bar{Y}_t$, are included into (2.4) as a proxy for the unobserved common factor f_t . For analytical convenience when deriving the asymptotic properties, Pesaran

(2007) replaces the usual estimator for σ_i^2 in the t-value for b_i by a slightly modified and also consistent one. He derives the asymptotic distribution of the modified t-statistic and shows that it is free of nuisance parameters as $N \rightarrow \infty$ for any fixed $T > 3$, as well as for the case where $N \rightarrow \infty$ followed by $T \rightarrow \infty$.

In line with Im et al. (2003), Pesaran (2007) proposes a cross-sectional augmented version of the IPS-test

$$CIPS = \frac{1}{N} \sum_{i=1}^N CADF_i, \quad (2.5)$$

where $CADF_i$ is the cross-sectionally augmented Dickey-Fuller statistic for the i -th cross-sectional unit given by the t-ratio of b_i in the CADF regression (2.4). Due to the presence of the common factor, the $CADF_i$ statistics will not be cross-sectionally independent³. Thus, a central limit theorem cannot be applied to derive the limiting distribution of the $CIPS$ statistic, and it is shown to be non-standard even for large N . Furthermore, to ensure the existence of moments for the distribution of $CADF_i$ in finite samples, Pesaran (2007) advocates the use of a truncated version of the $CIPS$ test, where for positive constants K_1 and K_2 such that $Pr[-K_1 < CADF_i < K_2]$ is sufficiently large, values of $CADF_i$ smaller than $-K_1$ or larger than K_2 are replaced by the respective bounds. Pesaran (2007) provides values for K_1 and K_2 obtained by simulations.

The presentation above outlines the procedure for serially uncorrelated disturbances. If there is serial correlation present in the common factors or idiosyncratic errors, additional lags of $\Delta Y_{i,t}$ and its cross-sectional average $\Delta \bar{Y}_t$ have to be included in the ADF regression (2.4).

2.2.2 Moon and Perron (2004): A dynamic panel model with K common factors

For a panel of observed data with N cross-sectional units and T time series observations, Moon and Perron (2004) model the DGP for $Y_{i,t}$ as an AR(1) process and assume, similar to Pesaran (2007), that common factors are present in the error term. They assume a K -factor model for the error term $u_{i,t}$

$$Y_{i,t} = (1 - \delta_i)\mu_i + \delta_i Y_{i,t-1} + u_{i,t}, \quad (2.6)$$

$$u_{i,t} = \lambda_i' f_t + e_{i,t}, \quad (2.7)$$

³Under the null hypothesis of a unit root, $CADF_i$ converges to a functional of Brownian motions, say $G(W_f, W_i)$, where W_f and W_i are Brownian motions driven by the common factor and idiosyncratic error, respectively.

for $i = 1, \dots, N$ and $t = 1, \dots, T$, where f_t is a $(K \times 1)$ vector of common factors, λ_i is the corresponding vector of factor loadings for cross-section i , and $e_{i,t}$ is an idiosyncratic disturbance term.

The DGPs considered by Pesaran (2007) and Moon and Perron (2004) are identical if a single common factor is present in the composite error term. For the components of the composite error term in (2.7) similar assumptions are made as by Pesaran (2007). The idiosyncratic part $e_{i,t}$ follows a stationary and invertible infinite MA process, and is cross-sectionally uncorrelated, so that $e_{i,t} = \Gamma_i(L)\varepsilon_{i,t}$, where $\Gamma_i(L) = \sum_{j=0}^{\infty} \gamma_{i,j}L^j$ and $\varepsilon_{i,t} \sim i.i.d.(0, 1)$ across i and t . Also the common factors f_t are assumed to have a stationary, invertible $MA(\infty)$ representation, i.e. $f_t = \Phi(L)\eta_t$. Here, $\Phi(L) = \sum_{j=0}^{\infty} \phi_jL^j$ is a K -dimensional lag polynomial and $\eta_t \sim i.i.d.(0, I_K)$. Furthermore, the covariance matrix of f_t is (asymptotically) positive definite. Although more than one common factor are permitted to be present in the data, some maximum number $\bar{K}(\geq K)$ is supposed to be known. Also, redundant factors, i.e. factors that asymptotically influence only a finite number of observed series, are excluded by imposing $\frac{1}{N} \sum_{i=1}^N \lambda_i \lambda_i' \rightarrow^p \Sigma_\lambda > 0$. Furthermore, short-run and long-run variances, $\sigma_{e_i}^2 (= \sum_{j=0}^{\infty} \gamma_{i,j}^2)$ and $\omega_{e_i}^2 (= (\sum_{j=0}^{\infty} \gamma_{i,j})^2)$, as well as the one sided long-run covariance $\varphi_{e_i} (= \sum_{l=1}^{\infty} \sum_{j=0}^{\infty} \gamma_{i,j} \gamma_{i,j+l})$ are supposed to exist for all idiosyncratic disturbances $e_{i,t}$. Additionally, these parameters are assumed to have non-zero cross-sectional averages, $\sigma_e^2 = \frac{1}{N} \sum_{i=1}^N \sigma_{e_i}^2$, $\omega_e^2 = \frac{1}{N} \sum_{i=1}^N \omega_{e_i}^2$ and $\varphi_e^2 = \frac{1}{N} \sum_{i=1}^N \varphi_{e_i}^2$.

The unit root null hypothesis considered by Moon and Perron (2004) is $H_0 : \delta_i = 1$ for all $i = 1, \dots, N$, which is tested against the heterogenous alternative $H_1 : \delta_i < 1$ for some i^4 . To test this hypothesis, two modified t-statistics are suggested, based on pooled estimation of the first-order serial correlation coefficient of the data. The estimation and testing procedure relies on de-factoring the data by a projection onto the space orthogonal to that spanned by the common factors. For that purpose, the matrix of factor loading $\Lambda = (\lambda_1, \dots, \lambda_N)'$ has to be estimated to construct a projection matrix $Q_\Lambda = I_N - \Lambda(\Lambda'\Lambda)^{-1}\Lambda'$.

Imposing $\delta_i = \delta$ for all i , the pooled OLS estimator, denoted as $\hat{\delta}_{pooled}$, is T -consistent for 1 under the unit root null, as well as under the local alternative considered by Moon and Perron (2004). The usual t-ratio to test this hypothesis has a non-standard limiting distribution, due to the persistent cross-sectional correlation introduced by the common factors. From the residuals of the

⁴To analyze local power properties of their test, Moon and Perron (2004) consider the following local alternative hypothesis:

$$\delta_i = 1 - \frac{\theta_i}{\sqrt{NT}},$$

where θ_i is a random variable with mean μ_θ on finite support $[0, \bar{M}]$. The considered null hypothesis is $H_0' : \mu_\theta = 0$, which is tested against the local alternative $H_1' : \mu_\theta > 0$.

pooled regression (under the null where the intercept is equal to zero)

$$\hat{u}_{i,t} = Y_{i,t} - \hat{\delta}_{pooled} Y_{i,t-1}, \quad (2.8)$$

the matrix of factor loadings is estimated by the method of principal components⁵. With the estimator $\hat{\Lambda}$ one can then construct an estimator of the projection matrix denoted as $Q_{\hat{\Lambda}_K}$. Additionally, consistent estimates of the above defined nuisance parameters can be obtained non-parametrically from the de-factored residuals $\hat{e} = \hat{u}Q_{\hat{\Lambda}_K}$, where $\hat{u} = (\hat{u}_1, \dots, \hat{u}_N)$ with $\hat{u}_i = (\hat{u}_{i,1}, \dots, \hat{u}_{i,T})'$. Denote the estimates as $\hat{\phi}_{e_i}$ and $\hat{\omega}_{e_i}^2$, and their cross-sectional averages as $\hat{\phi}_e$ and $\hat{\omega}_e^2$. Then the modified pooled estimator of δ suggested by Moon and Perron (2004) is

$$\delta_{pooled}^* = \frac{\sum_{t=2}^T Y'_{t-1} Q_{\hat{\Lambda}_K} Y_t - NT \hat{\phi}_e}{\sum_{t=2}^T Y'_{t-1} Q_{\hat{\Lambda}_K} Y_{t-1}}, \quad (2.9)$$

where $Y_t = (Y_{1,t}, \dots, Y_{N,t})'$. Based on this estimator, the following two t-statistics can be used to test the pooled unit root null hypothesis,

$$t_a^* = \frac{\sqrt{NT}(\hat{\delta}_{pooled}^* - 1)}{\sqrt{\frac{2\hat{\phi}_e^4}{\hat{\omega}_e^4}}} \quad (2.10)$$

and

$$t_b^* = \sqrt{NT}(\hat{\delta}_{pooled}^* - 1) \sqrt{\frac{1}{NT^2} \sum_{t=2}^T Y'_{t-1} Q_{\hat{\Lambda}_K} Y_{t-1} \begin{pmatrix} \hat{\omega}_e \\ \hat{\phi}_e^2 \end{pmatrix}}, \quad (2.11)$$

where $\hat{\phi}_e^4 = \frac{1}{N} \sum_{i=1}^N \hat{\phi}_{e_i}^4$, $\hat{\phi}_{e_i}^4 = \hat{\omega}_{e_i}^4$. Moon and Perron (2004) analyze the asymptotic behavior of the two statistics as $N \rightarrow \infty$ and $T \rightarrow \infty$ with⁶ $\liminf_{(N,T \rightarrow \infty)} \frac{\log T}{\log N} > 1$. Both test statistics have a limiting standard normal distribution under the null, and diverge under the stationary alternative.

2.2.3 Bai and Ng (2004b): A common factor model with unobserved common and idiosyncratic components of unknown order of integration.

In contrast to Pesaran (2007) or Moon and Perron (2004), the PANIC model of Bai and Ng (2004b) permits the non-stationarity in a panel of observed data to come either from a common source, or

⁵The principal component estimator is in general not unique. Moon and Perron (2004) use the normalization $\frac{1}{T} \sum_{t=1}^T f_t f_t' = I_K$ and re-scale the obtained estimate.

⁶The restriction on the relative divergence rate of N and T is necessary, as f_t and $e_{i,t}$ are unobserved.

from the idiosyncratic errors, or from both⁷. Therefore, they focus on consistent estimation of the common factors and error terms, to test the properties of these series separately.

The model Bai and Ng (2004b) consider describes the observed data $Y_{i,t}$ as the sum of a deterministic part, a common (stochastic) component, and the idiosyncratic error. In particular,

$$Y_{i,t} = D_{i,t} + \lambda_i' F_t + E_{i,t} \quad i = 1, \dots, N, t = 1, \dots, T, \quad (2.12)$$

where as before λ_i is a $(K \times 1)$ vector of factor loadings, F_t is a $(K \times 1)$ vector of common factors⁸, and $E_{i,t}$ is an error term. The deterministic component $D_{i,t}$ contains either a constant α_i or a linear trend $\alpha_i + \beta_i t$. As the two aforementioned approaches, Bai and Ng (2004b) consider a balanced panel with N cross-sectional units and, T time series observations.

The common factors are assumed to follow an $AR(1)$ process, such that

$$F_t = F_{t-1} + f_t, \quad (2.13)$$

where $f_t = \Phi(L)\eta_t$, $\Phi(L) = \sum_{j=1}^{\infty} \phi_j L^j$ is a K -dimensional lag polynomial and $\text{rk}(\Phi(1)) = k_1$. So, F_t contains $k_1 \leq K$ independent stochastic trends and consequently $K - k_1$ stationary components. The shock η_t is assumed to be *i.i.d.*(0, Σ_η) with finite fourth-order moment. The idiosyncratic terms are allowed to be either $I(0)$ and $I(1)$, and are also modelled as $AR(1)$ processes

$$E_{i,t} = \delta_i E_{i,t-1} + e_{i,t}, \quad (2.14)$$

where $e_{i,t}$ follows a mean zero, stationary, invertible MA process, such that $e_{i,t} = \Gamma_i(L)\varepsilon_{i,t}$ with $\varepsilon_{i,t} \sim \text{i.i.d.}(0, \sigma_{\varepsilon_i}^2)$. Bai and Ng (2004b) do not assume cross-sectional independence of the idiosyncratic term⁹ from the outset, but impose it later to validate pooled testing. The assumption that Σ_η is not (necessarily) a diagonal matrix is more general than the corresponding assumption in Moon and Perron (2004), where the innovations of the common factors are assumed to be uncorrelated. The short-run covariance matrix of ΔF_t has full rank while the long-run covariance matrix has reduced rank and hence permits cointegration among the common factors. As in Moon and Perron (2004), (asymptotically) redundant factors are ruled out.

⁷Under the unit root null the data in the Pesaran (2007) or Moon and Perron (2004) model contains a common, as well as an idiosyncratic stochastic trend.

⁸ K is assumed to be known here.

⁹Bai and Ng (2004b) allow for some weak cross-sectional dependence of the shock terms driving the $e_{i,t}$. The full set of assumptions can be found in their paper.

In this setup, the goal of PANIC is to determine the number of non-stationary factors k_1 , and to test for each $i = 1, \dots, N$, whether $\delta_i = 1$. Bai and Ng (2004b) suggest using principal components to consistently estimate the unobserved components F_t and $E_{i,t}$. However, to derive consistent estimates even if some elements of F_t and $E_{i,t}$ are $I(1)$, a suitable transformation of $Y_{i,t}$ is used. In particular, if the DGP does not contain a deterministic linear trend, the first differences of the data are employed, while in the presence of a deterministic linear trend, demeaned first-differences are used. So, in the former case $y_{i,t} = \Delta Y_{i,t} = Y_{i,t} - Y_{i,t-1}$, while in the latter $y_{i,t} = \Delta Y_{i,t} - \Delta \bar{Y}_{i,t}$, where $\Delta \bar{Y}_{i,t} = \frac{1}{T-1} \sum_{t=2}^T \Delta Y_{i,t}$. As the estimated common factors and idiosyncratic errors, denoted as \hat{f}_t and $\hat{e}_{i,t}$ respectively, are derived applying the method of principal components to first-differenced or de-trended data, Bai and Ng (2004b) propose to re-accumulate them to remove the effect of possible over-differencing. This yields

$$\hat{F}_t = \sum_{s=2}^t \hat{f}_s, \quad (2.15)$$

$$\hat{E}_{i,t} = \sum_{s=2}^t \hat{e}_{i,s}. \quad (2.16)$$

These estimates are now individually tested for unit roots.

For the idiosyncratic components, Bai and Ng (2004b) suggest to compute an ADF statistic based on up to p lags. Denote the t-statistic to test the unit root hypothesis for each $\hat{E}_{i,t}$ as $ADF_{\hat{E}_i}^c$ or $ADF_{\hat{E}_i}^\tau$, depending on whether a constant, or a constant and linear trend is included in the DGP. Bai and Ng (2004b) derive the limiting distributions, which are non-standard. For the case where a constant is present in the DGP given by (2.12), the distribution coincides with the usual Dickey-Fuller (DF) distribution where no constant is included in the estimation. The 5% critical value is -1.95 . If the DGP in (2.12) contains a constant and a linear trend, the limiting distribution is proportional to the reciprocal of a Brownian bridge. Critical values for this distribution are not tabulated yet, and have to be simulated.

Both ADF statistics given above do not have the advantage of a standard normal limiting distribution, as do the other panel unit root tests described so far. That is due to the fact that the panel information has only been used to consistently estimate $E_{i,t}$, but not to analyze its dynamic properties. Only if independence among the error terms is assumed, pooled testing is valid. In that case, Bai and Ng (2004b) propose a Fisher-type test¹⁰ as suggested in Maddala and Wu (1999), using

¹⁰In principal, also an IPS-type test using a standardized average of the above described t-statistics should be possible. See also Bai and Ng (2007).

the correction proposed by Choi (2001). The test statistic, denoted as $P_{\hat{E}}^c$ or $P_{\hat{E}}^\tau$ depending on the deterministic specification, is given by

$$P_{\hat{E}}^c, P_{\hat{E}}^\tau = \frac{-2 \sum_{i=1}^N \log \pi_i - 2N}{\sqrt{4N}}, \quad (2.17)$$

where π_i is the p-value of the ADF test for the i -th cross-section. These two panel unit root test statistics have standard normal limiting distributions.

Depending on whether there is just one, or several common factors, Bai and Ng (2004b) suggest to use either an ADF test based on up to p lags, or a rank test for \hat{F}_t . Denote the t-statistic for the unit root hypothesis as $ADF_{\hat{F}}^c$ when only a constant is accounted for, and as $ADF_{\hat{F}}^\tau$ in the linear trend case. Then, Bai and Ng (2004b) derive their limiting distributions, which coincide with the DF distributions for the cases where only a constant, or a constant and a linear trend are included in the ADF estimation. The asymptotic 5% critical values are -2.86 and -3.41, respectively.

If there are $K > 1$ common factors, Bai and Ng (2004b) suggest an iterative procedure, comparable to the Johansen trace test for cointegration to select k_1 . They use demeaned or de-trended factor estimates, depending on whether (2.12) contains just a constant, or a constant and linear trend. Define $\tilde{F}_t = \hat{F}_t - \bar{\hat{F}}_t$ with $\bar{\hat{F}}_t = \frac{1}{T-2} \sum_{t=2}^T \hat{F}_t$ in the former case. In the latter, let \tilde{F}_t denote the residuals from a regression of \hat{F}_t on a constant and linear trend. Using \tilde{F}_t , the following steps describe the proposed test.

Starting with $m = K$,

1. Let $\hat{\beta}_\perp$ be the m eigenvectors associated with the m largest eigenvalues of $\frac{1}{T^2} \sum_{t=2}^T \tilde{F}_t \tilde{F}_t'$. Let $\hat{X}_t = \hat{\beta}_\perp' \tilde{F}_t$. Two statistics can be considered:
 2. (a) Let $K(j) = 1 - \frac{j}{J+1}$, $j = 1, \dots, J$;
 - i. Let $\hat{\xi}_t$ be the residuals from estimating a VAR(1) in \hat{X}_t , and let

$$\hat{\Sigma}_1 = \sum_{j=1}^J K(j) \left(\frac{1}{T} \sum_{t=2}^T \hat{\xi}_{t-j} \hat{\xi}_t' \right).$$

- ii. Let $\hat{\nu}_c(m)$ be the smallest eigenvalue of

$$\hat{\Phi}_c(m) = \frac{1}{2} \left[\sum_{t=2}^T (\hat{X}_t \hat{X}_{t-1}' + \hat{X}_{t-1} \hat{X}_t') - T(\hat{\Sigma}_1 + \hat{\Sigma}_1') \right] \left(\sum_{t=2}^T \hat{X}_{t-1} \hat{X}_{t-1}' \right)^{-1}.$$

- iii. Denote $T[\hat{\nu}_c(m) - 1]$ as $MQ_c^c(m)$ in the constant only case, or as $MQ_c^\tau(m)$ in the

linear trend case.

(b) For p fixed that does not depend on N or T ,

- i. Estimate a $\text{Var}(p)$ in $\Delta\hat{X}_t$ in order to obtain $\hat{\Pi}(L) = I_m - \hat{\Pi}_1 L - \dots - \hat{\Pi}_p L^p$. Filter \hat{X}_t by $\hat{\Pi}(L)$ to get $\hat{x}_t = \hat{\Pi}(L)\hat{X}_t$.
- ii. Let $\hat{\nu}_f(m)$ be the smallest eigenvalue of

$$\hat{\Phi}_f(m) = \frac{1}{2} \left[\sum_{t=2}^T (\hat{x}_t \hat{x}'_{t-1} + \hat{x}_{t-1} \hat{x}'_t) \right] \left(\sum_{t=2}^T \hat{x}_{t-1} \hat{x}'_{t-1} \right)^{-1}.$$

- iii. Denote $T[\hat{\nu}_f(m) - 1]$ as $MQ_f^c(m)$ in the constant only case, or as $MQ_f^\tau(m)$ in the linear trend case.

3. If $H_0 : k_1 = m$ is rejected, set $m = m - 1$ and return to Step 1. Otherwise, set $\hat{k}_1 = m$ and stop.

For the $MQ_c^{c,\tau}$ and $MQ_f^{c,\tau}$ statistics described above, Bai and Ng (2004b) derive limiting distributions, which are again non-standard, and they provide 1%, 5%, and 10% critical values for all four statistics and for various values of m .

The PANIC procedure has the advantage that the estimated common factors and idiosyncratic components are consistent whether they are stationary or non-stationary. This is due to the practice of estimating the unobserved components from the first-differenced (or de-trended) data, and re-accumulating the estimates to remove the effect of possible over-differencing if the factors or errors are stationary. Hence, the obtained estimates could also be used for stationarity tests, which is discussed in Bai and Ng (2004a).

2.2.4 Alternative panel unit root tests in the presence of cross-sectional dependencies

The three approaches to panel unit root testing presented in the previous sections explicitly account for the common factors employed to model the cross-sectional dependence in the data by using methods that require large N to be valid. In this section we introduce alternative panel unit root tests which do not necessarily exploit the common factor structure, and could provide alternatives to the aforementioned tests in small N panels. In particular, we will consider two test statistics proposed by Breitung and Das (2008) and the tests proposed by Sul (2007).

Breitung and Das (2008)

Breitung and Das (2008) study the behaviour of several panel unit root tests when cross-sectional dependence in the data is present in the form of a common factor. The DGP they employ is similar to that of Bai and Ng (2004b) presented in Section 2.2.3, Equations (2.12) to (2.14). However, Breitung and Das (2008) focus on the special case where (2.13) can be replaced by

$$F_t = \rho F_{t-1} + f_t,$$

with the scalar first order autoregressive parameter $|\rho| \leq 1$. They consider test statistics on the “reduced form” regression equation below, which is obtained when $\delta_i = \delta$ for all i and $\rho = \delta$:

$$\Delta Y_t = \phi Y_{t-1} + u_t, \quad (2.18)$$

where $\Delta Y_t = (\Delta Y_{1,t}, \dots, \Delta Y_{N,t})'$, $Y_{t-1} = (Y_{1,t-1}, \dots, Y_{N,t-1})'$, $u_t = (u_{1,t}, \dots, u_{N,t})'$ with $u_{i,t} = \lambda_i f_t + e_{i,t}$ and $\phi = (\delta - 1)$. Breitung and Das (2008) present their analysis for a DGP and model without individual specific constant or time trend. The deterministic component in (2.12) has been assumed to be zero in this case. If a model with individual specific constant is employed, Breitung and Das (2008) suggest to remove it by considering data in deviation from the first observation, $Y_{i,t}^* = Y_{i,t} - Y_{i,0}$.

Breitung and Das (2008) particularly consider a robust OLS t-statistic t_{rob} and a GLS t-statistic t_{gls} to test for the unit root null hypothesis $\phi = 0$ against the homogenous alternative $\phi < 0$. The robust OLS statistic is given by

$$t_{rob} = \frac{\sum_{t=1}^T Y'_{t-1} \Delta Y_t}{\left(\sum_{t=1}^T Y'_{t-1} \hat{\Omega} Y_{t-1} \right)^{\frac{1}{2}}},$$

with $\hat{\Omega} = \sum_{t=1}^T \hat{u}_t \hat{u}'_t$ where $\hat{u}_t = \Delta Y_t - \hat{\phi} Y_{t-1}$ are the OLS residuals. The GLS statistic, t_{gls} , is given by

$$t_{gls} = \frac{\sum_{t=1}^T Y'_{t-1} \hat{\Omega}^{-1} \Delta Y_t}{\left(\sum_{t=1}^T Y'_{t-1} \hat{\Omega}^{-1} Y_{t-1} \right)^{\frac{1}{2}}}.$$

Note that this statistic can only be computed for $T > N$, as otherwise $\hat{\Omega}$ is singular. Also, if a common factor structure is assumed for the data, one could exploit this in for the GLS statistic by taking the factor structure into account when estimating the covariance matrix Ω . For the static factor model with orthonormal factors, $\Omega = \Lambda \Lambda' + \Sigma$, where Λ is the $N \times k$ matrix of factor loadings

and Σ is the covariance matrix of the idiosyncratic innovations. Estimates of Λ and Σ can be obtained using a principal component approach as in Bai and Ng (2004b) or Moon and Perron (2004). If there is higher order serial correlation present in the residuals, a Newey-West type estimator for Ω can be employed, or an ADF regression estimated in the first step.

Breitung and Das (2008) consider 3 cases in their analysis, where the reduced form (2.18) is misspecified in cases 2 and 3, namely an I(1) common factor combined with I(1) idiosyncratic components, an I(1) common factor and I(0) idiosyncratic components (cross-member cointegration) and the case where a unit root is present in the idiosyncratic component but the common factor is I(0). If $\frac{N^3}{T} \rightarrow 0$, t_{gls} is asymptotically normally distributed in the first and third case, while it diverges in the second case. t_{rob} converges to a Dickey-Fuller distribution in the first case if there is a single common factor. It is equivalent to an ADF test on the first principal component of Y_t in that case. In the other cases, the test is not valid.

Sul (2007)

Sul (2007) proposes to use recursive mean adjustment for panel unit root tests to increase their power. Similar to Moon and Perron (2004), Sul (2007) models cross-sectional dependence by employing a common factor structure for the error term. The DGP is similar to that given in Equations (2.6) and (2.7). To account for the cross-sectional dependence, Sul (2007) suggests a (feasible) GLS statistic to test for the unit root null hypothesis $\delta_i = 1$ for all i against the heterogenous alternative $\delta_i < 1$ for some i in

$$Y_{i,t} = (1 - \delta_i)\mu_i + \delta_i Y_{i,t-1} + u_{i,t}, \quad (2.19)$$

The test procedure follows multiple steps, where the regression can be augmented by lagged first differences of $Y_{i,t}$ to account for higher order serial correlation in the residuals:

1. Run the following regression for each unit individually

$$Y_{i,t} - c_{i,t-1} = \delta_i(Y_{i,t-1} - c_{i,t-1}) + \sum_{j=1}^{p_i} \varphi_{ij} \Delta Y_{i,t-j} + \epsilon_{i,t}, \quad (2.20)$$

where $c_{i,t-1} = (t-1)^{-1} \sum_{s=1}^{t-1} Y_{i,s}$ is the recursive mean, to obtain the LS estimator $\hat{\delta}_i$.

2. If $\hat{\delta}_i > 1$ set $\hat{\delta}_i = 1$ and run the regression

$$Y_{i,t} - \hat{\delta}_i Y_{i,t-1} = a_i + \sum_{j=1}^p \varphi_{ij} \Delta Y_{i,t-j} + \epsilon_{i,t}. \quad (2.21)$$

Construct the sample covariance matrix $\hat{\Omega} = (T - p - 1)^{-1} \sum_{t=p+1}^T \hat{\varepsilon}_t \hat{\varepsilon}_t'$, where $\hat{\varepsilon}_t = (\hat{\varepsilon}_{1,t}, \dots, \hat{\varepsilon}_{N,t})'$ are the vectors of residuals from the previous regression.

3. Project $(Y_{i,t} - c_{i,t-1})$ and $(Y_{i,t-1} - c_{i,t-1})$ on the lagged first differences

$$\begin{aligned} (Y_{i,t} - c_{i,t-1}) &= \sum_{j=1}^p \phi_{ij} \Delta Y_{i,t-j} + \xi_{i,t}, \\ (Y_{i,t-1} - c_{i,t-1}) &= \sum_{j=1}^p \zeta_{ij} \Delta Y_{i,t-j} + \xi_{i,t-1}. \end{aligned}$$

4. Define $\hat{\omega}'_{ij}$ as the ij^{th} element of $\hat{\Omega}^{-1}$, one can now obtain the pooled FGLS estimator of δ and the associated t-statistic as

$$\hat{\delta}_{fglsrma} = \frac{\sum_{i=1}^N \sum_{j=1}^N \hat{\omega}'_{ij} \sum_{t=p+1}^T \hat{\xi}_{i,t-1} \hat{\xi}_{j,t}}{\sum_{i=1}^N \sum_{j=1}^N \hat{\omega}'_{ij} \sum_{t=p+1}^T \hat{\xi}_{i,t-1}^2}, \quad (2.22)$$

$$t_{fglsrma} = \frac{\hat{\delta}_{fglsrma} - 1}{\sqrt{\sum_{i=1}^N \sum_{j=1}^N \hat{\omega}'_{ij} \sum_{t=p+1}^T \hat{\xi}_{i,t-1}^2}}. \quad (2.23)$$

Sul (2007) shows that the $t_{fglsrma}$ converges to a Dickey-Fuller distribution, and he provides finite sample critical values to account for finite sample bias.

Similar to Breitung and Das (2008), Sul's $t_{fglsrma}$ effectively tests for a unit root in the idiosyncratic component of the data if the error term $u_{i,t}$ in (2.19) permits a common factor structure. To test for a unit root in the common component, Sul (2007) proposes to apply a recursive mean adjusted covariate augmented DF test to the cross-sectional averages of the data, $\bar{Y}_t = N^{-1} \sum_{i=1}^N Y_{i,t}$. The steps of the procedure are similar to the ones outlined above, and the resulting t-statistic is denoted as t_{crma} . Sul (2007) provides some evidence that his test is precise and powerful, especially when T is larger than N , a case for which it has been designed.

2.2.5 Differences and similarities

This section discusses differences and similarities of the panel unit root tests relying on a factor structure, presented in the previous subsections. For all considered DGPs, we can write the data as the sum of the deterministic component ($D_{i,t}$), a ‘‘common component’’ ($CC_{i,t}$) and an ‘‘idiosyncratic component’’ ($IC_{i,t}$) such that

$$Y_{i,t} = D_{i,t} + CC_{i,t} + IC_{i,t}.$$

For the DGP of Bai and Ng (2004b), we have $CC_{i,t} = \lambda'_i F_t$ and $IC_{i,t} = E_{i,t}$. For a DGP as assumed by Pesaran (2007) or Moon and Perron (2004) given in e.g. (2.6)-(2.7) where the common factor structure is assumed for the error term, we obtain for a simple AR(1),

$$Y_{i,t} = (1 - \delta_i^t)\mu_i + \lambda'_i \sum_{s=0}^{t-1} \delta_i^{t-s} f_{t-s} + \sum_{s=0}^{t-1} \delta_i^{t-s} e_{i,t-s}.$$

Hence, $CC_{i,t} = \lambda'_i \sum_{s=0}^{t-1} \delta_i^{t-s} f_{t-s}$ and $IC_{i,t} = \sum_{s=0}^{t-1} \delta_i^{t-s} e_{i,t-s}$ for those DGPs. The approaches to panel unit root testing presented above may differ in terms of assumptions made which place restriction on the DGP, in particular whether the order of integration is allowed to differ between $CC_{i,t}$ and $IC_{i,t}$ and thus whether the possibility of cross-member cointegration is excluded or not, and the number of common factors. Furthermore, the presented test statistics are applied to different data components. For example, the Moon and Perron (2004) tests apply to the idiosyncratic component only, as has been shown by Breitung and Das (2008) and forcefully argued by Bai and Ng (2007).

DGP

The DGP assumed by Pesaran (2007) for a single common factor and Moon and Perron (2004) for $K \geq 1$ restrict the common and idiosyncratic component to have the same order of integration. Bai and Ng (2004b) explicitly allow the order of integration to differ between $CC_{i,t}$ and $IC_{i,t}$ and they allow for the presence of $K \geq 1$ factors. Sul (2007) considers a DGP similar to Bai and Ng (2004b) as well and proposes to proxy a single common factor with the cross-sectional average of the data. Breitung and Das (2008) analyze the behaviour of their tests in DGPs as assumed by Bai and Ng (2004b).

Null Hypothesis and Tested Data Component

All considered tests have non-stationarity as null hypothesis. The statistics proposed by Pesaran (2007) and Moon and Perron (2004) test defactored data ($IC_{i,t}$) for a unit root. The common component is not tested, although it is non-stationary if $IC_{i,t}$ is non-stationary given that the assumptions on the DGP are true. Bai and Ng (2004b) suggest test statistics for the idiosyncratic and common component separately, where the null hypothesis is non-stationarity of the given component. Breitung and Das (2008) formulate the null hypothesis in terms of the reduced form regression (2.18) as a unit root in the observed data. However, they show that their FGLS statistic effectively tests for a unit root in the idiosyncratic component, while their robust OLS statistic is equivalent to an ADF

test for the first principal component only if both $CC_{i,t}$ and $IC_{i,t}$ are non-stationary. Sul's FGLS statistic also tests for a unit root in the idiosyncratic component, while cross-sectional averages are used as a proxy for a single common factor and tested for a unit root with the t_{crma} test.

The *CIPS* test of Pesaran (2007), the tests of Moon and Perron (2004), the $P_{\hat{E}}^{c,\tau}$ statistics of Bai and Ng (2004b) and the FGLS statistics proposed by Breitung and Das (2008) and Sul (2007) are pooled tests for the null hypothesis that the defactored data are unit root processes for all i . All approaches except Breitung and Das (2008) use a heterogenous alternative, namely that some series have a unit root and some do not. Moon and Perron (2004) use a pooled estimator of the first order autoregressive coefficient δ_i in the construction of their statistics. Similarly, the FGLS tests of Breitung and Das (2008) and Sul (2007) are based on pooled estimators $\hat{\delta}$. The individual specific *CADF* statistic of Pesaran (2007) and the $ADF_{\hat{E}}^{c,\tau}$ statistic of Bai and Ng (2004b) test for a unit root in the idiosyncratic component for a given i , and the alternative hypothesis is stationarity of that component.

Bai and Ng's $ADF_{\hat{F}}^{c,\tau}$ statistic and Sul's t_{crma} test for a unit root in a single common factor. Also, Bai and Ng (2004b) allow for more than one common factor and the $MQ_c^{c,\tau}$ and $MQ_f^{c,\tau}$ statistics are designed to determine the number of independent stochastic trends in F_t .

Panel dimensions N and T

The three type of tests proposed by Pesaran (2007), Moon and Perron (2004) and Bai and Ng (2004b) are designed for large N and T due to the estimation of the common factor(s) either by using principal components or by including the cross-sectional mean as proposed by Pesaran (2007). The FGLS tests of Breitung and Das (2008) and Sul (2007) on the other hand can only be constructed if $T > N$.

Cointegration

While Pesaran (2007) and Moon and Perron (2004) exclude the possibility of cointegration among the $Y_{i,t}$, as well as between the observed data and the common factors, Bai and Ng (2004b) include both possibilities in their model. In particular, if $k_1 \geq 1$ and $E_{i,t}(= IC_{i,t})$ is stationary for some i , then the observed data and the common factors are cointegrated for those i with cointegrating vector $(1, -\lambda'_i)'$. Furthermore, if all idiosyncratic errors are $I(0)$, then the orthogonalization matrix used by Moon and Perron (2004) to eliminate the common factors, Q_Λ , serves as cointegration matrix for the $Y_{i,t}$. So, Bai and Ng's procedure can be used as a cointegration test¹¹, by investigating the

¹¹What is meant here is a cointegration test between panel members, in contrast to panel cointegration tests. The latter ones are used to test for cointegration between several variables for the same i .

hypotheses $k_1 \geq 1$ and all idiosyncratic errors are stationary¹². Breitung and Das (2008) consider the case of cross-member cointegration in their analysis, however their tests are not able to detect it. Sul's tests could be used to detect cross-member cointegration, namely if the $t_{fglsrma}$ statistic rejects a unit root for the idiosyncratic component while the t_{crma} test fails to reject the unit root for the cross-sectional averages.

Common Factors and Estimation of K

For the tests proposed by Moon and Perron (2004) and Bai and Ng (2004b), an important aspect in application is the selection of the number of common factors K . Consistent estimation of K is discussed in Bai and Ng (2002) for a factor model as given by (2.12) with stationary errors, and also briefly treated in Moon and Perron (2004). It should be noted that while the information criteria designed to estimate the number of common factors work well in simulations, their application in practice is difficult as they are usually observed to select the maximum number of common factors allowed.

In terms of computational burden, all procedures are rather easy to implement. Pesaran (2007) provides tables with critical values for his tests. The PANIC procedure of Bai and Ng (2004b) also requires some tabulated critical values for the rank test statistics $MQ_{(\cdot)}^c$ and $MQ_{(\cdot)}^f$, as well as for the $ADF_{\hat{E}_{i,t}}^\tau$ statistic. Also, a procedure to calculate the p-values of $ADF_{\hat{E}_{i,t}}^c$ and $ADF_{\hat{E}_{i,t}}^\tau$ is needed to implement the suggested pooled tests. Sul (2007) also provides simulated finite sample critical values for his test statistics.

2.3 Small sample performance: Monte Carlo results

2.3.1 Monte Carlo simulation setup

In this section we study the small sample performance of the tests proposed by Pesaran (2007), Moon and Perron (2004) and Bai and Ng (2004b) for various types of DGPs. Furthermore, we consider the robust OLS t-test t_{rob} and the FGLS t-test t_{GLS} described in Breitung and Das (2008) and the recursive mean adjusted FGLS test $t_{\rho fglsrma}$ and the recursive mean adjusted test for the average data proposed by Sul (2007). All considered DGPs with one exception have the following structure

¹²Note that the null hypothesis for the ADF tests using the estimated error terms remains that of non-stationarity. Rejecting the unit root hypothesis for all i is thus one part of not rejecting cointegration between panel members.

which corresponds to the framework of Bai and Ng (2004b):

$$\begin{aligned} Y_{i,t} &= \lambda'_i F_t + E_{i,t}, \\ F_{m,t} &= \varphi F_{m,t-1} + f_{m,t}, \\ E_{i,t} &= \delta_i E_{i,t-1} + e_{i,t}, \end{aligned} \tag{2.24}$$

with $i = 1, \dots, N$, $t = 1, \dots, T$ and $m = 1, \dots, K$. We consider three different values for N and T each, namely 20, 50 and 100¹³. The method of principle components estimates the space spanned by the common factors when N is large. We have chosen N and T at least equal to 20 to assure that common factors are estimated with sufficient precision or approximated reasonably well by cross-sectional averages. Notice that the regularity condition $N \neq T$ needed for some tests is not satisfied in some cases. First a single common factor is considered, which is generated by a first order autoregression, or a random walk when $\varphi = 1$. We also consider the case of two common factors which are generated using the same parameter values for φ and σ_f^2 , but different drawings for the error terms. The idiosyncratic terms $E_{i,t}$ are also generated by a first order autoregression or random walk with first order moving average, depending on whether or not $\delta_i = 1$.

In addition, a DGP as assumed by Pesaran (2007) and Moon and Perron (2004) is used:

$$\begin{aligned} Y_{i,t} &= \delta_i Y_{i,t-1} + u_{i,t}, \\ u_{i,t} &= \lambda_i f_t + e_{i,t}. \end{aligned} \tag{2.25}$$

In (2.24) and (2.25) the error terms are generated as MA(1) processes such that

$$\begin{aligned} f_{m,t} &= \eta_{m,t} + \gamma_m \eta_{m,t-1}, \\ e_{i,t} &= \varepsilon_{i,t} + \rho_i \varepsilon_{i,t-1}. \end{aligned}$$

The shocks are drawn from independent normal distributions, such that $\eta_t \sim i.i.d.N(0, \Sigma_f^2)$, with $\Sigma_f^2 = \sigma_f^2 I_K$, and $\varepsilon_{i,t} \sim i.i.d.N(0, 1)$. We consider three different values for the signal-to-noise ratio, such that $\sigma_f^2 = 0.5, 1$ and 2^{14} . The MA parameters γ_m and ρ_i are independently, uniformly distributed

¹³Pesaran (2007) reports Monte Carlo results for his tests with $N, T = 10, 20, 30, 50, 100$, Moon and Perron (2004) choose $N = 10, 20$ and $T = 100, 300$, Bai and Ng (2004b) report results for $N = 40, T = 100$ while Bai and Ng (2007) choose $N, T = 20, 50, 100$, Sul (2007) performs simulations with $N = 5, 10, 15, 20$ and $T = 50, 100, 200$, and Breitung and Das (2008) select $N = 10, 20, 50$ and $T = 20, 50, 100$.

¹⁴In the tables we only report the values for $\sigma_f^2 = 1$. The other results are available at <http://www.personeel.unimaas.nl/J.Urbain/>.

on $[0.2, 0.5]$. The factor loading λ_i are uniformly distributed on $[-1, 3]$ ¹⁵.

Three different types of non-stationarity are considered as null hypothesis, as well as different settings for the stationary alternative hypothesis. In particular, we consider the following 5 cases, where 1 to 4 use the DGP given by (2.24) and 5 uses DGP (2.25)¹⁶:

1. Common and idiosyncratic unit roots

$$H_0^A : \varphi = 1, \text{ and } \delta_i = 1 \text{ for all } i.$$

2. Common unit root, nearly stationary idiosyncratic components

$$H_0^B : \varphi = 1, \text{ and } \delta_i \sim U[0.8, 1] \text{ for all } i,$$

3. Stationary common component, integrated idiosyncratic components

$$H_0^C : \varphi = 0.95, \text{ and } \delta_i = 1 \text{ for all } i,$$

4. Stationary common and idiosyncratic components

$$H_A^A : \varphi = 0.95 \text{ and } \delta_i \sim U[0.8, 1].$$

5. Stationary data using a DGP as given by (2.25) with heterogenous roots

$$H_A^E : \delta_i \sim U[0.8, 1] \text{ for all } i.$$

The results are obtained with GAUSS 8.0 using 1000 replications. The reported rejection frequencies are based on 5% nominal size. All power results are size unadjusted. For Pesaran's *CADF* and *CIPS* we use the critical values reported in Pesaran (2007, Tables 1b and 3b). Results for Moon and Perron's statistics, Bai and Ng's $P_{\hat{E}}^c$ statistic and Breitung and Das's t_{rob} and t_{gls} are based on a critical value from the standard normal distribution. Rejection frequencies of the $ADF_{\hat{E}}^c$ and $ADF_{\hat{F}}^c$ statistics are obtained using the critical values from DF distributions for the no intercept

¹⁵Consistency of the test procedure of Pesaran (2007) requires a non-zero mean for the factor loadings. This assumption is not necessary for the other approaches.

¹⁶Please note that under setup 1 (2.24) and (2.25) are equivalent. In cases 4 and 5 we have stationarity provided $\delta_i \neq 1$.

and intercept only cases, respectively. Critical values for the MQ_c^c and MQ_f^c are provided in Bai and Ng (2004b, Table 1). For Sul's $t_{\rho f g l s r m a}$ test we use finite sample critical values reported in Sul (2007, Table 5) and for the $t_{c r m a}$ we use the asymptotic critical value of -1.88 . When obtaining the $t_{c r m a}$ statistic we use $\Delta Y_{1,t}$ as stationary covariate and calculate the cross-sectional averages over the remaining $N - 1$ panel members such that $\bar{Y}_t = (N - 1)^{-1} \sum_{i=2}^N Y_{i,t}$.

Similar to Moon and Perron (2004), we use the Andrews and Monahan (1992) estimator employing the quadratic spectral kernel in the estimation of the nuisance parameters for the t_a^* and t_b^* statistics. For Bai and Ng's $ADF_{\hat{E}}^c$ and $ADF_{\hat{F}}^c$ and Pesaran's $CADF$ and $CIPS$ we use the Akaike information criterion (AIC) to determine the lag length, starting with a maximum lag length of $p_{max} = 6$. For the test of Sul (2007) and Breitung and Das (2008) we use the Bayesian information criterion (BIC). For the MQ_c^c statistic we use the Bartlett kernel with a bandwidth as suggested in Andrews (1991). The lag length for the MQ_f^c statistic is determined using the criteria proposed by Aznar and Salvador (2002). We do not estimate the number of common factors K but assume it known. However, in addition to results where the number of common factors is correctly specified we also report results for a case where this number is misspecified. We do so to show sensitivity of the tests to this form of misspecification and not to be unfair to authors of any of these tests when evaluating the adequacy of the asymptotic approximations since the theory does not incorporate sampling variability due to the number of common factors¹⁷.

Although the considered DGPs do not include deterministic components, we do account for individual fixed effects in the simulation by including constants in the regressions. Following the advise of Breitung and Das (2008) for the $t_{r o b}$ and $t_{g l s}$ test we consider data in deviation from the initial observation to remove the effect of an individual specific constant¹⁸.

The finite sample performance of the considered test statistics depend on these choices. For reasons of comparison, we follow the original authors with the choices they report or we select a procedure that performs better in terms of size in our simulations.

¹⁷Bai and Ng (2007) also obtain Monte Carlo results assuming the true number of common factors to be known. Bai and Ng (2004b) and Moon and Perron (2004) provide Monte Carlo simulations for their respective tests where the number of common factors is estimated.

¹⁸As already noted by Breitung and Das (2008), applying the tests to demeaned data leads to dependence on nuisance parameter unless applied to the GLS transformed data, and severe finite sample size distortions.

2.3.2 Monte Carlo results

A general finding is that the presence of serial correlation¹⁹ leads to size distortions for almost all statistics when T is small, which can be quite strong in some cases and even persist for $T = 100$. For a single common factor, the signal-to-noise ratio seems to have little to no effect on the tests proposed by Pesaran (2007) and Bai and Ng (2004b). For two common factors in the DGP, Bai and Ng's MQ_c^c and MQ_f^c statistics usually select maximum possible number of common stochastic trends, leading to low size and low power for these tests when the auto-regressive root is close to unity. The FGLS statistics of Breitung and Das (2008) and Sul (2007) behave quite similarly in terms of size and power. Sul's t_{crma} statistic applied to the cross-sectional averages of the data has similar size properties as Bai and Ng's $ADF_{\hat{F}}^c$. Power properties of the two tests are similar too for most cases.

The results in Table 2.1 are obtained for the case where a unit root is present in the common factors and in all idiosyncratic errors. Both statistics proposed by Pesaran (2007), the $CADF$ ²⁰ and the $CIPS$ test show size distortions when T is small (20), which are stronger for the $CIPS$ test. For a single common factor those size distortions are reduced as T increases and for $T = 100$ the tests are only slightly over-sized. For $K = 2$, size distortions increasing in the signal-to-noise ratio remain even for large T , in particular for the $CIPS$ test. Both statistics proposed by Moon and Perron (2004) show slight size distortions which seem to increase with the signal-to-noise ratio when K is correctly specified. The size distortions are decreasing in T and higher for t_a^* than for t_b^* . The later is actually undersized for small signal-to-noise ratios. If K is misspecified, both statistics show strong size distortions increasing in the signal-to-noise ratio, but size distortions are lower when K is over-estimated. Bai and Ng's $ADF_{\hat{E}}^c$ and $ADF_{\hat{F}}^c$ statistics for the extracted individual idiosyncratic error series and the single common factor respectively, are oversized for small $T(= 20)$ but size distortions decrease as T gets large. The pooled statistic $P_{\hat{E}}^c$ has strong size distortions when T is small and size increases in N . For $T = 100$, size ranges from 0.12 to 0.18 for the different values of N . Similar to Moon and Perron's tests, size distortions are less severe when the number of common factors is over-specified if $\hat{K} \neq K$. Both rank statistics MQ_c^c and MQ_f^c usually pick the maximum number of possible common stochastic trends, leading to good properties when K is specified correctly but failure to estimate the correct number of common factors if $\hat{K} = 3$ is used. Breitung and Das's t_{rob} is under-sized for small T with rejection frequencies increasing in T but decreasing with N , leading

¹⁹Results for the case of i.i.d. $N(0, 1)$ error terms $e_{i,t}$ and f_t in (2.24) are not included here. They are available at <http://www.personeel.unimaas.nl/J.Urbain/>.

²⁰Entries for the $CADF$ -statistics are average rejection frequencies of the individual unit root tests.

to rejection frequencies between 0.00 and 0.11. The t_{gls} test has a size of about 0.05 for $N = 20$ and is under-sized for $N = 50$, similarly to Sul's $t_{\rho fglrma}$ tests. The t_{crma} test for the cross-sectional averages is slightly oversized with size distortions decreasing in T . All four statistics behave similarly whether a single or two common factors are present in the data.

Table 2.2 considers the case of a unit root in the common factors and near-unit roots in the idiosyncratic factors, i.e. the case of cross-member cointegration. For $K = 1$, Pesaran's $CADF$ statistic has an average rejection frequency of about 0.32 for $T = 20$ and between 0.17 and 0.21 for larger T . The rejection frequencies of the $CIPS$ test are high and go to 1 for large N and T . For $K = 2$, rejection frequencies are reduced, in particular for $CIPS$ where they also decrease as the signal-to-noise ratio increases. Both statistics proposed by Moon and Perron (2004) have rejection increasing to 1 in N and T , with rejection frequencies for t_a^* slightly higher than those for t_b^* when the correct number of common factors is employed. When K is under-estimated, rejection frequencies are strongly reduced. Bai and Ng's $ADF_{\hat{E}}^c$ statistic has an average power increasing from about 0.23 to 0.48 as both N and T increase. The pooled $P_{\hat{E}}^c$ test has a power of 1 for almost all combinations of N and T considered, when K is correctly specified or over-specified. When a single common factor is extracted but two common factors are present in the data, rejection frequencies are reduced. The $ADF_{\hat{F}}^c$ tests has some size distortions, but rejection frequencies decrease from about 0.40 for $T = 20$ to 0.07 to 0.10 for $T = 100$. The MQ_c^c and MQ_f^c statistics again pick the maximum number of possible trends, leading to good properties only when K is correctly specified. Rejection frequencies for Breitung and Das's t_{rob} statistics decrease for higher signal-to-noise ratios, whereas they increase with T . The t_{gls} statistic has rejection frequencies between 0.44 and 0.64, increasing with T . Sul's $t_{\rho fglrma}$ statistic has similar rejection frequencies ranging between 0.38 and 0.66, which also increase in T . The t_{crma} test is slightly oversized with size distortions lower for $T = 100$.

Table 2.3 covers the case of integrated idiosyncratic errors combined with a stationary common factor. The statistics proposed by Pesaran (2007) behave similar to the case of I(1) idiosyncratic and common component (Table 2.1), but size is slightly reduced for the $CIPS$ test, which is now under-sized for $T = 100$ and $K = 1$. Moon and Perron's t_a^* and t_b^* also behave similar to Table 2.1 but have slightly higher rejection frequencies, increasing in the signal-to-noise ratio in particular for $K = 2$. When K is misspecified, rejection frequencies for both statistics increase in N , T and the signal-to-noise ratio. Bai and Ng's $ADF_{\hat{E}}^c$ and $P_{\hat{E}}^c$ tests have sizes close to the one shown in Table 2.1. The power of the $ADF_{\hat{F}}^c$ is smaller than 0.20 for $T \geq 50$. The MQ_c^c and MQ_f^c statistics fail to detect the correct number of common stochastic trends. Breitung and Das's t_{rob} test has size increasing

in T but decreasing in N . The t_{gls} and Sul's $t_{fglsrma}$ tests are slightly over-sized for $N = 20$ and under-sized for $N = 50$, with size increasing in the signal-to-noise ratio. The t_{crma} test has rejection frequencies ranging from 0.15 to 0.24, increasing in N and T but decreasing as the signal-to-noise ratio increases. Also, rejection frequencies for the tests of Breitung and Das (2008) and Sul (2006) are slightly larger when $K = 2$.

Tables 2.4 and 2.5 consider stationary data. For Table 2.4 the DGP is given by (2.24) with $I(0)$ idiosyncratic and common components. Pesaran's $CADF$ has low power while the power of the $CIPS$ test is relatively high and increasing in N , reaching 1 for $N, T = 100$ when $K = 1$. For $K = 2$, power is reduced and furthermore decreasing in the signal-to-noise ratio. Moon and Perron's tests both have power increasing to 1 as N and T increase. The average power of Bai and Ng's $ADF_{\hat{E}}^c$ is relatively low (0.52 for $N, T = 100$) while the pooled test $P_{\hat{E}}^c$ has a power of 1 for $N > 20$ or $T > 20$. The power of the $ADF_{\hat{F}}^c$ is low and both rank tests MQ_c^c and MQ_f^c fail to select the correct number of common stochastic trends. Breitung and Das's t_{rob} test has power increasing in T but decreasing the signal-to-noise ratio. The t_{gls} test has a power between 0.55 and 0.84, increasing in T but decreasing in N , similar to Sul's $t_{fglsrma}$ test which has power between 0.51 and 0.87. Power for these 3 tests is increased for $K = 2$. Rejection frequencies for the t_{crma} are 0.10 and 0.27 when $K = 1$ and 0.07 and 0.30 for $K = 2$, increasing in T but decreasing in N and the signal-to-noise ratio.

Table 2.5 considers stationary data generated using (2.25). Rejection frequencies for most tests are reduced and now decrease as the signal-to-noise ration increases, in particular for Moon and Perron's t_a^* and t_b^* and Breitung and Das's t_{gls} and Sul's $t_{fglsrma}$ tests. Bai and Ng's $ADF_{\hat{F}}^c$ has a higher power now, but it is still relatively low. Sul's t_{crma} test also has an increased power now, increasing in N , T and the signal-to-noise ratio.

From the Monte Carlo simulations, several general conclusions can be drawn. The presence of serial correlation in the error term leads to size distortions which can be quite large in small samples. The Moon and Perron (2004) tests, the tests of Pesaran (2007), the $P_{\hat{E}}^c$ and $ADF_{\hat{E}}^c$ statistics of Bai and Ng (2004b) and the FGLS statistics proposed by Breitung and Das (2008) and Sul (2007) indeed test for a unit root in the idiosyncratic component, and reject a unit root if it is present in the common factor alone. The pooled CIPS test of Pesaran (2007) and $P_{\hat{E}}^c$ test of Bai and Ng (2004b) are more powerful than the individual test statistics they are based on, $CADF$ and $ADF_{\hat{E}}^c$ respectively. However, the pooled tests show higher size distortions for small T . The CIPS test has good size and power for large N and T if a single common factor is present. However, an additional common

factor leads to size distortions and reduced power. The $P_{\hat{E}}^c$ statistic has high power, but some size distortions remain even for $N, T = 100$. The t_a^* statistic has slightly larger size distortions than the t_b^* test, with power being high for both statistics. The later three statistics are not distorted by the presence of a second common factor if K is correctly specified in the estimation. If K is misspecified, the statistics exhibit size distortions, but over-estimating K seems to be less harmful in terms of power. The two FGLS statistics are slightly undersized for $N = 50$ but have a high power. Also, their performance remains good in terms of size and power if two factors are included in the data.

Bai and Ng's $ADF_{\hat{F}}^c$ statistic and Sul's t_{crma} statistic have been proposed test whether there is a unit root in a single common factor. The $ADF_{\hat{F}}^c$ has low power and some size distortions even for large N, T . Sul's t_{crma} test shows similar size and power in most cases, but has a higher power when the DGP given in (2.25) is used. Bai and Ng's MQ_c^c and MQ_f^c are designed to test for the number of common stochastic trends if more than one common factor is present, but have very low power against alternatives close to a unit root.

We have not studied the issue of which test to choose if the common factor model representation is not appropriate to describe cross-sectional dependence. Bootstrap unit root tests might be used in such an instance, but this question is left for future research. Furthermore, we do not consider DGPs with idiosyncratic linear deterministic trends. Moon and Perron (2004) show that their tests have no local power in that case, but all other authors consider propose their tests for such DGPs as well. Bai and Ng (2007) provide simulation results for some tests for DGPs including idiosyncratic linear deterministic trends.

2.4 An illustrative application: Testing for PPP using the new approaches

This section presents an application of the new panel unit root tests described in Section 2.2 to illustrate their use in an empirical study of the validity of purchasing power parity (PPP). For this purpose we consider the potential existence of a unit root in real exchange rate series that are constructed as

$$Y_{i,t} = s_{i,t} - p_t^* + p_{i,t}, \quad (2.26)$$

where $s_{i,t}$ is the ln of country i 's nominal exchange rate versus some numeraire currency, p_t^* is the ln of the aggregate price level in the numeraire country, and $p_{i,t}$ is the ln of country i 's domestic aggregate price level.

The numerous analyzes of PPP in the literature do not come to a common conclusion with respect to PPP. Some studies report stronger rejection of the unit root null, if the German Mark instead of the US Dollar is used as a numeraire currency. Also, studies using univariate unit root or cointegration tests reject PPP, while tests using panel methods as the LLC or IPS test tend to find evidence in favor of it, see for example Oh (1996). However, as was already discussed in the introduction, several studies have analyzed the properties of early panel unit root tests in the presence of cross-sectional dependence since then, and argued against their use for PPP tests. Lyhagen (2000) analytically derives the cross-correlation structure in a panel of real exchange rates, constructed with a common numeraire country. He also derives the effect of the common stochastic trend in the data introduced by the numeraire on the limiting distributions of various panel statistics. In Monte Carlo simulations, he finds size distortions similar to those reported by Banerjee et al. (2004, 2005)).

In the analysis presented in this section, monthly data from 14 European countries is considered. The data set includes information on the nominal exchange rates of local currency versus US Dollar (\$US) for Austria, Denmark, Finland, France, Germany, Greece, Italy, the Netherlands, Norway, Portugal, Spain, Sweden, Switzerland and the UK. Furthermore, the Consumer Price Index (CPI) as a proxy for aggregate price levels is included for those 14 countries and the US. The sample includes monthly observations on all variables for the period from February 1986 to September 2002, so 200 observations. For such a sample, one can expect to find high correlation between panel units, due to a high degree of economic integration and political co-operation. As far as monetary policy is concerned, the most important mechanism of co-operation is the European Monetary System (EMS), to which some panel members belong, and which finally led to the introduction of the Euro as a common currency in some countries.

As a starting point of the analysis, the real exchange rate series are individually tested for a unit root using an ADF test. The lag length p is set to 12 for all countries. The individual ADF statistics are shown in Table 2.6. Only for the UK, the unit root null can be rejected for both real exchange rate series. Using the real exchange rate versus DM, also for Switzerland the ADF test rejects at a 5% significance level. These findings are representative for those of studies using univariate tests. The problem here is that it remains unclear whether the non-rejection of the unit root is due to a failure of PPP, or the low power of the ADF test against near unit root alternative.

Next, the panel unit root tests described in Section 2.2 are performed. For each test, it is assumed that a single common factor is present in the data. Given that the real exchange rate series are constructed using a common base currency, this assumption seems reasonable. For the tests of

Pesaran (2007) and Moon and Perron (2004), the data representation in (2.1)-(2.2) is assumed to be valid. The results of the $CADF_i$ tests suggested by Pesaran (2007) are given in Table 2.6, and Table 2.7 presents Pesaran's $CIPS$ statistic and those proposed by Moon and Perron (2004). Except for the French real exchange rate when measured against the German Mark, the $CADF_i$ statistics fail to reject the unit root null. Also, the pooled $CIPS$ test does not reject the null in both panels. This provides some evidence against PPP. The t_a^* and t_b^* statistics of Moon and Perron (2004) do not provide such a clear picture. While the former one rejects PPP in both panels, the latter one does not reject it when the US Dollar is used as a numeraire currency.

The results for the panel unit root tests proposed by Breitung and Das (2008) and by Sul (2007) are given in Table 2.7. While the unit root null hypothesis is not rejected by any test for real exchange rates constructed with the US as base country, the t_{gls} test of Breitung and Das (2008) and the $t_{fglsrma}$ test of Sul (2007) reject the unit root when real exchange rates are constructed with Germany as base country.

For the application of the Bai and Ng (2004b) procedure, it is assumed that the data can be represented as in (2.12). With this representation, there is an interpretive problem. Clearly, if both F_t and $E_{i,t}$ are stationary, the real exchange rate is stationary and PPP holds in the long run at least. Also, if both common and idiosyncratic components are $I(1)$, PPP can be rejected. But, if just the common factors are non-stationary the real exchange rate series are pairwise cointegrated along the cross-section but individually non-stationary, so that PPP in the usual sense does not hold between panel members and the base country. However, in the special case $\lambda_i = \lambda_j$, the cointegrating vector for $Y_{i,t}^B$ and $Y_{j,t}^B$ is $[1, -1]$, where the superscript B denotes the base country. Then PPP holds between countries i and j , since

$$Y_{i,t}^B - Y_{j,t}^B = s_{i,t}^B - s_{j,t}^B + p_{i,t} - p_{j,t} = s_{i,t}^j + p_{i,t} - p_{j,t} = Y_{i,t}^j \sim I(0). \quad (2.27)$$

The results for the test statistics suggested by Bai and Ng (2004b) are presented in Table 2.7. Most of the individual tests for the idiosyncratic errors, as well as the test for the common factor reject the unit root. Also, the pooled error test rejects the unit root for both panels of real exchange rates. This provides some evidence in favor of PPP.

2.5 Conclusion

In this chapter several panel unit root tests that account for cross section dependence assuming or using a common factor structure have been compared, notably Pesaran (2007), Moon and Perron (2004), Bai and Ng (2004b), Breitung and Das (2008) and Sul (2007). There are often valid theoretical and empirical reasons why a common factor structure can be expected to yield sensible results. Therefore, panels with dynamic factors are of interest in economic modelling.

We have studied these approaches to unit root testing in panels with dynamic factors, compared them in terms of DGP, tests, null and alternative hypotheses. We have studied the small sample behavior of the tests proposed in a common framework and discussed their use in econometric modelling. In addition, we have applied them in an empirical study of purchasing power parity.

The main conclusions are:

- In the case where the observed non-stationarity is only due to a non-stationary common factor, the individual series are pairwise cointegrated along the cross sectional dimension. Only the Bai and Ng (2004b) and Sul (2007) tests allow for this type of structure to be detected, if the unit root is rejected for the idiosyncratic component but not for the common factor.
- The $ADF_{\hat{F}}^c$ for testing for the presence of unit roots in a single common factor is found to have low power. Similarly, in a multi-factor setting, the MQ_c^c and MQ_f^c tests fail to distinguish high but stationary serial correlation from non-stationarity in the common factors. For the one factor model, Bai and Ng's $ADF_{\hat{F}}^c$ test has similar size and power than Sul's t_{crma} test in most cases, except when a DGP as given in (2.25) is employed in which case the later test is more powerful.
- Testing the idiosyncratic component for a unit root: Pesaran's $CADF$ and $CIPS$ tests are indeed designed for testing for unit roots when cross-sectional dependence is due to a single common factor, and size and power are adversely affected by a second common factor. The pooled $CIPS$ test has better power properties than the individual specific $CADF$ tests. Similarly, Bai and Ng's pooled $P_{\hat{E}}^c$ tests is more powerful than the individual specific $ADF_{\hat{E}}^c$ in detecting unit roots in the idiosyncratic components, although the former can have strong size distortion when the time dimension of the panel is small. However, the $P_{\hat{E}}^c$ and $ADF_{\hat{E}}^c$ statistic can accommodate to more than one common factor. The Moon and Perron (2004) tests can also account for multiple common factors. The two tests proposed by Moon and Perron (2004) are found to have similar small sample power, but the t_a^* statistic is found to have slightly larger

size distortions than the t_b^* . When the FGLS tests considered by Breitung and Das (2008) and Sul (2007) can be computed, i.e. when $N < T$, they provide good alternatives to test for unit roots in the idiosyncratic component.

- When the number of common factors is unknown and has to be selected, it is less harmful in terms of power to include too many factors than too few in the test procedures of Bai and Ng (2004b) and Moon and Perron (2004). The statistics exhibits size distortions if the number of common factors is misspecified.

2.A Tables

Table 2.1: Finite sample (average) rejection rates for DGP (2.24) with I(1) common factor(s) and I(1) idiosyncratic components.

K	N	T	t_a^*	t_b^*	$P_{\hat{E}}^c$	$ADF_{\hat{E}}^c$	$ADF_{\hat{F}}^c$	CADF	CIPS	t_{rob}	t_{gls}	$t_{fglsrma}$	t_{crma}	
1	20	20	0.09	0.04	0.64	0.16	0.37	0.28	0.52	0.03	-	-	-	
1	20	50	0.08	0.05	0.15	0.07	0.15	0.12	0.16	0.09	0.09	0.07	0.12	
1	20	100	0.09	0.05	0.12	0.06	0.08	0.07	0.07	0.10	0.06	0.05	0.07	
1	50	20	0.11	0.06	0.86	0.16	0.37	0.28	0.59	0.00	-	-	-	
1	50	50	0.06	0.04	0.20	0.07	0.14	0.12	0.16	0.04	-	-	-	
1	50	100	0.06	0.04	0.14	0.06	0.07	0.07	0.07	0.08	0.01	0.01	0.08	
1	100	20	0.09	0.05	0.96	0.16	0.35	0.28	0.64	0.00	-	-	-	
1	100	50	0.05	0.04	0.28	0.07	0.13	0.12	0.16	0.00	-	-	-	
1	100	100	0.05	0.04	0.18	0.06	0.09	0.07	0.05	0.02	-	-	-	
K	N	T	t_a^*	t_b^*	$P_{\hat{E}}^c$	MQ_c^c	MQ_f^c	CADF	CIPS	t_{rob}	t_{gls}	$t_{fglsrma}$	t_{crma}	
2	20	20	0.11	0.05	0.63	1.00	0.97	0.32	0.53	0.04	-	-	-	
2	20	50	0.08	0.05	0.16	1.00	1.00	0.15	0.30	0.09	0.09	0.06	0.12	
2	20	100	0.10	0.05	0.12	1.00	1.00	0.09	0.24	0.09	0.06	0.07	0.10	
2	50	20	0.14	0.10	0.82	1.00	0.98	0.31	0.59	0.01	-	-	-	
2	50	50	0.05	0.03	0.20	1.00	1.00	0.15	0.34	0.03	-	-	-	
2	50	100	0.06	0.04	0.14	1.00	1.00	0.08	0.25	0.09	0.01	0.01	0.08	
2	100	20	0.12	0.09	0.96	1.00	0.99	0.31	0.60	0.00	-	-	-	
2	100	50	0.06	0.04	0.31	1.00	1.00	0.14	0.33	0.01	-	-	-	
2	100	100	0.04	0.03	0.15	1.00	1.00	0.08	0.28	0.01	-	-	-	
$\hat{K} = 1$														
K	N	T	t_a^*	t_b^*	$P_{\hat{E}}^c$	$ADF_{\hat{E}}^c$	$ADF_{\hat{F}}^c$			t_a^*	t_b^*	$P_{\hat{E}}^c$	MQ_c^c	MQ_f^c
2	20	20	0.31	0.22	0.58	0.16	0.40			0.25	0.22	0.61	0.00	0.01
2	20	50	0.32	0.23	0.25	0.07	0.12			0.18	0.14	0.16	0.00	0.00
2	20	100	0.32	0.23	0.21	0.06	0.08			0.15	0.11	0.13	0.00	0.00
2	50	20	0.38	0.32	0.75	0.16	0.39			0.37	0.36	0.82	0.00	0.01
2	50	50	0.38	0.33	0.32	0.07	0.13			0.25	0.26	0.19	0.00	0.00
2	50	100	0.41	0.37	0.26	0.06	0.09			0.22	0.21	0.12	0.00	0.00
2	100	20	0.45	0.42	0.78	0.16	0.38			0.37	0.36	0.92	0.00	0.01
2	100	50	0.48	0.46	0.42	0.07	0.14			0.30	0.29	0.30	0.00	0.00
2	100	100	0.50	0.49	0.38	0.06	0.10			0.22	0.21	0.18	0.00	0.00
$\hat{K} = 3$														

Finite sample (average) rejection rates for Pesaran's (2007) CADF and CIPS statistics, Moon and Perron's (2004) t_a^* and t_b^* statistics, Bai and Ng's (2004b) $ADF_{\hat{E}}^c$, $P_{\hat{E}}^c$, and $ADF_{\hat{F}}^c$ statistics, Breitung and Das's (2008) t_{rob} and t_{gls} statistics, and Sul's (2007) $t_{fglsrma}$ and t_{crma} statistics. Proportions of repetitions when Bai and Ng's (2004b) MQ_c^c and MQ_f^c statistics chose the correct number of common stochastic trends. Finite sample (average) rejection rates for Moon and Perron's (2004) t_a^* and t_b^* statistics, and Bai and Ng's (2004b) $ADF_{\hat{E}}^c$, $P_{\hat{E}}^c$, and $ADF_{\hat{F}}^c$ statistics, the proportions of repetitions when Bai and Ng's (2004b) MQ_c^c and MQ_f^c statistics chose the correct number of common stochastic trends, when the number of common factors is misspecified. K denotes the number of common factors in the DGP. \hat{K} specifies the number of common factors used when testing if \hat{K} is different from K . Rejection frequencies are based on 5% cutoff values from Pesaran (2007, Tables 1b and 3b), Sul (2007, Table 5), Bai and Ng (2004b, Table 1) Table 1, 5% cutoff values of the standard normal distribution, or 5% Dickey-Fuller critical values for the test statistics as specified in the text. Results are obtained with GAUSS 8.0 using 1000 replications.

Table 2.2: Finite sample (average) rejection rates for DGP (2.24) with I(1) common factor(s) and I(0) idiosyncratic components.

K	N	T	t_a^*	t_b^*	$P_{\hat{E}}^c$	$ADF_{\hat{E}}^c$	$ADF_{\hat{F}}^c$	CADF	CIPS	t_{rob}	t_{gls}	$t_{fglsrma}$	t_{crma}
1	20	20	0.57	0.40	0.94	0.23	0.40	0.32	0.68	0.06	-	-	-
1	20	50	0.85	0.76	1.00	0.25	0.14	0.17	0.55	0.19	0.45	0.39	0.10
1	20	100	0.92	0.85	1.00	0.44	0.07	0.18	0.85	0.16	0.62	0.65	0.09
1	50	20	0.75	0.67	1.00	0.23	0.38	0.31	0.79	0.02	-	-	-
1	50	50	0.98	0.96	1.00	0.26	0.13	0.18	0.71	0.11	-	-	-
1	50	100	1.00	0.99	1.00	0.46	0.09	0.19	0.98	0.17	0.56	0.59	0.07
1	100	20	0.99	0.85	1.00	0.24	0.38	0.32	0.86	0.02	-	-	-
1	100	50	1.00	1.00	1.00	0.27	0.14	0.19	0.81	0.07	-	-	-
1	100	100	1.00	1.00	1.00	0.48	0.09	0.21	1.00	0.11	-	-	-
K	N	T	t_a^*	t_b^*	$P_{\hat{E}}^c$	MQ_c^c	MQ_f^c	CADF	CIPS	t_{rob}	t_{gls}	$t_{fglsrma}$	t_{crma}
2	20	20	0.48	0.37	0.91	1.00	0.98	0.33	0.59	0.05	-	-	-
2	20	50	0.84	0.75	0.99	1.00	1.00	0.16	0.44	0.13	0.42	0.35	0.11
2	20	100	0.93	0.88	1.00	1.00	1.00	0.13	0.48	0.12	0.64	0.62	0.09
2	50	20	0.72	0.62	1.00	1.00	0.98	0.33	0.63	0.04	-	-	-
2	50	50	0.98	0.96	1.00	1.00	1.00	0.17	0.51	0.06	-	-	-
2	50	100	1.00	0.99	1.00	1.00	1.00	0.12	0.55	0.11	0.47	0.56	0.06
2	100	20	0.85	0.81	1.00	1.00	0.98	0.34	0.68	0.03	-	-	-
2	100	50	1.00	1.00	1.00	1.00	1.00	0.16	0.51	0.04	-	-	-
2	100	100	1.00	1.00	1.00	1.00	1.00	0.12	0.54	0.07	-	-	-
			$\hat{K} = 1$					$\hat{K} = 3$					
K	N	T	t_a^*	t_b^*	$P_{\hat{E}}^c$	$ADF_{\hat{E}}^c$	$ADF_{\hat{F}}^c$		t_a^*	t_b^*	$P_{\hat{E}}^c$	MQ_c^c	MQ_f^c
2	20	20	0.46	0.40	0.69	0.19	0.38		0.41	0.31	0.84	0.00	0.01
2	20	50	0.56	0.49	0.56	0.14	0.15		0.78	0.69	0.96	0.00	0.00
2	20	100	0.60	0.54	0.61	0.18	0.08		0.90	0.86	1.00	0.00	0.00
2	50	20	0.58	0.56	0.79	0.19	0.39		0.54	0.49	0.99	0.00	0.03
2	50	50	0.57	0.56	0.64	0.14	0.14		0.95	0.92	1.00	0.00	0.00
2	50	100	0.65	0.62	0.70	0.19	0.09		0.99	0.99	1.00	0.00	0.00
2	100	20	0.64	0.62	0.83	0.20	0.37		0.68	0.64	1.00	0.00	0.03
2	100	50	0.67	0.66	0.66	0.13	0.12		1.00	0.99	1.00	0.00	0.00
2	100	100	0.70	0.69	0.73	0.18	0.08		1.00	1.00	1.00	0.00	0.00

See notes Table 2.1.

Table 2.3: Finite sample (average) rejection rates for DGP (2.24) with I(0) common factor(s) and I(1) idiosyncratic components.

K	N	T	t_a^*	t_b^*	$P_{\hat{E}}^c$	$ADF_{\hat{E}}^c$	$ADF_{\hat{F}}^c$	CADF	CIPS	t_{rob}	t_{gls}	$t_{fglsrma}$	t_{crma}
1	20	20	0.13	0.06	0.61	0.16	0.41	0.27	0.47	0.04	-	-	-
1	20	50	0.11	0.06	0.12	0.07	0.19	0.11	0.10	0.22	0.11	0.10	0.18
1	20	100	0.12	0.06	0.12	0.06	0.19	0.06	0.02	0.43	0.09	0.09	0.20
1	50	20	0.17	0.10	0.84	0.16	0.40	0.27	0.55	0.00	-	-	-
1	50	50	0.09	0.06	0.20	0.07	0.17	0.11	0.08	0.08	-	-	-
1	50	100	0.08	0.06	0.12	0.06	0.19	0.05	0.01	0.31	0.01	0.01	0.23
1	100	20	0.13	0.08	0.95	0.16	0.40	0.27	0.58	0.00	-	-	-
1	100	50	0.08	0.06	0.25	0.07	0.16	0.11	0.06	0.01	-	-	-
1	100	100	0.07	0.06	0.15	0.06	0.19	0.06	0.01	0.09	-	-	-
K	N	T	t_a^*	t_b^*	$P_{\hat{E}}^c$	MQ_c^c	MQ_f^c	CADF	CIPS	t_{rob}	t_{gls}	$t_{fglsrma}$	t_{crma}
2	20	20	0.16	0.08	0.59	0.00	0.00	0.32	0.54	0.06	-	-	-
2	20	50	0.13	0.08	0.14	0.00	0.00	0.14	0.24	0.30	0.10	0.11	0.21
2	20	100	0.15	0.09	0.10	0.00	0.00	0.08	0.15	0.57	0.11	0.14	0.29
2	50	20	0.23	0.16	0.78	0.00	0.00	0.31	0.59	0.02	-	-	-
2	50	50	0.12	0.08	0.14	0.00	0.00	0.14	0.26	0.13	-	-	-
2	50	100	0.14	0.10	0.11	0.00	0.00	0.07	0.12	0.53	0.01	0.04	0.24
2	100	20	0.25	0.19	0.94	0.00	0.00	0.31	0.59	0.00	-	-	-
2	100	50	0.15	0.11	0.21	0.00	0.00	0.14	0.31	0.06	-	-	-
2	100	100	0.12	0.09	0.12	0.00	0.00	0.08	0.18	0.28	-	-	-
			$\hat{K} = 1$					$\hat{K} = 3$					
K	N	T	t_a^*	t_b^*	$P_{\hat{E}}^c$	$ADF_{\hat{E}}^c$	$ADF_{\hat{F}}^c$		t_a^*	t_b^*	$P_{\hat{E}}^c$	MQ_c^c	MQ_f^c
2	20	20	0.45	0.33	0.63	0.17	0.42		0.42	0.38	0.57	0.00	0.00
2	20	50	0.62	0.48	0.39	0.10	0.17		0.52	0.45	0.14	0.00	0.00
2	20	100	0.77	0.64	0.48	0.10	0.15		0.59	0.52	0.10	0.00	0.00
2	50	20	0.56	0.50	0.82	0.17	0.41		0.62	0.60	0.79	0.00	0.00
2	50	50	0.74	0.66	0.50	0.09	0.16		0.76	0.75	0.15	0.00	0.00
2	50	100	0.89	0.83	0.63	0.09	0.18		0.89	0.89	0.11	0.00	0.00
2	100	20	0.64	0.60	0.87	0.17	0.41		0.67	0.67	0.90	0.00	0.00
2	100	50	0.85	0.84	0.68	0.10	0.19		0.81	0.81	0.24	0.00	0.00
2	100	100	0.97	0.95	0.82	0.10	0.20		0.94	0.94	0.14	0.00	0.00

See notes Table 2.1.

Table 2.4: Finite sample (average) rejection rates for DGP (2.24) with I(0) common factor(s) and I(0) idiosyncratic components.

K	N	T	t_a^*	t_b^*	$P_{\hat{E}}^c$	$ADF_{\hat{E}}^c$	$ADF_{\hat{F}}^c$	CADF	CIPS	t_{rob}	t_{gls}	$t_{fglsrma}$	t_{crma}
1	20	20	0.66	0.49	0.94	0.23	0.43	0.31	0.64	0.08	-	-	-
1	20	50	0.91	0.83	1.00	0.26	0.17	0.16	0.48	0.36	0.56	0.51	0.18
1	20	100	0.95	0.90	1.00	0.48	0.17	0.16	0.74	0.63	0.81	0.83	0.24
1	50	20	0.86	0.78	1.00	0.24	0.41	0.30	0.73	0.03	-	-	-
1	50	50	0.99	0.98	1.00	0.27	0.17	0.17	0.59	0.25	-	-	-
1	50	100	1.00	1.00	1.00	0.49	0.20	0.17	0.95	0.62	0.69	0.75	0.13
1	100	20	0.95	0.94	1.00	0.24	0.41	0.31	0.82	0.04	-	-	-
1	100	50	1.00	1.00	1.00	0.29	0.18	0.17	0.69	0.20	-	-	-
1	100	100	1.00	1.00	1.00	0.52	0.15	0.19	1.00	0.51	-	-	-
K	N	T	t_a^*	t_b^*	$P_{\hat{E}}^c$	MQ_c^c	MQ_f^c	CADF	CIPS	t_{rob}	t_{gls}	$t_{fglsrma}$	t_{crma}
2	20	20	0.62	0.48	0.90	0.00	0.00	0.33	0.59	0.10	-	-	-
2	20	50	0.91	0.86	1.00	0.00	0.00	0.17	0.47	0.43	0.67	0.60	0.22
2	20	100	0.97	0.93	1.00	0.00	0.00	0.14	0.62	0.77	0.89	0.88	0.29
2	50	20	0.87	0.80	1.00	0.00	0.00	0.34	0.67	0.06	-	-	-
2	50	50	1.00	0.99	1.00	0.00	0.00	0.18	0.53	0.32	-	-	-
2	50	100	1.00	1.00	1.00	0.00	0.00	0.14	0.68	0.77	0.70	0.88	0.07
2	100	20	0.96	0.94	1.00	0.00	0.00	0.35	0.69	0.07	-	-	-
2	100	50	1.00	1.00	1.00	0.00	0.00	0.17	0.54	0.27	-	-	-
2	100	100	1.00	1.00	1.00	0.00	0.00	0.15	0.67	0.66	-	-	-
			$\hat{K} = 1$					$\hat{K} = 3$					
K	N	T	t_a^*	t_b^*	$P_{\hat{E}}^c$	$ADF_{\hat{E}}^c$	$ADF_{\hat{F}}^c$		t_a^*	t_b^*	$P_{\hat{E}}^c$	MQ_c^c	MQ_f^c
2	20	20	0.63	0.57	0.78	0.21	0.42		0.58	0.50	0.83	0.00	0.00
2	20	50	0.89	0.85	0.87	0.23	0.17		0.93	0.87	0.99	0.00	0.00
2	20	100	0.99	0.98	1.00	0.40	0.19		0.98	0.96	1.00	0.00	0.00
2	50	20	0.75	0.73	0.88	0.22	0.43		0.77	0.72	0.99	0.00	0.00
2	50	50	0.96	0.95	0.94	0.22	0.17		1.00	1.00	1.00	0.00	0.00
2	50	100	1.00	1.00	1.00	0.40	0.18		1.00	1.00	1.00	0.00	0.00
2	100	20	0.82	0.82	0.90	0.21	0.41		0.87	0.84	1.00	0.00	0.00
2	100	50	0.97	0.97	0.95	0.21	0.16		1.00	1.00	1.00	0.00	0.00
2	100	100	1.00	1.00	1.00	0.39	0.17		1.00	1.00	1.00	0.00	0.00

See notes Table 2.1.

Table 2.5: Finite sample (average) rejection rates for DGP (2.25) with I(0) common factor(s) I(0) idiosyncratic components.

K	N	T	t_a^*	t_b^*	$P_{\hat{E}}^c$	$ADF_{\hat{E}}^c$	$ADF_{\hat{F}}^c$	CADF	CIPS	t_{rob}	t_{gls}	$t_{fglsrma}$	t_{crma}
1	20	20	0.49	0.35	0.90	0.22	0.45	0.30	0.60	0.09	-	-	-
1	20	50	0.59	0.52	0.97	0.22	0.26	0.16	0.39	0.37	0.37	0.31	0.39
1	20	100	0.55	0.51	1.00	0.39	0.31	0.18	0.80	0.54	0.54	0.37	0.63
1	50	20	0.60	0.53	1.00	0.23	0.41	0.29	0.70	0.03	-	-	-
1	50	50	0.68	0.65	1.00	0.22	0.20	0.16	0.54	0.20	-	-	-
1	50	100	0.69	0.67	1.00	0.40	0.29	0.18	0.97	0.39	0.47	0.28	0.87
1	100	20	0.77	0.73	1.00	0.23	0.42	0.30	0.77	0.04	-	-	-
1	100	50	0.82	0.80	1.00	0.24	0.25	0.16	0.59	0.18	-	-	-
1	100	100	0.81	0.80	1.00	0.41	0.34	0.20	1.00	0.40	-	-	-
K	N	T	t_a^*	t_b^*	$P_{\hat{E}}^c$	MQ_c^c	MQ_f^c	CADF	CIPS	t_{rob}	t_{gls}	$t_{fglsrma}$	t_{crma}
2	20	20	0.41	0.29	0.83	0.00	0.00	0.33	0.57	0.11	-	-	-
2	20	50	0.45	0.38	0.87	0.00	0.00	0.17	0.44	0.37	0.28	0.21	0.27
2	20	100	0.48	0.41	1.00	0.00	0.00	0.18	0.73	0.55	0.37	0.25	0.40
2	50	20	0.61	0.53	0.98	0.00	0.00	0.33	0.68	0.07	-	-	-
2	50	50	0.62	0.56	1.00	0.00	0.00	0.18	0.56	0.29	-	-	-
2	50	100	0.61	0.57	1.00	0.00	0.00	0.19	0.84	0.57	0.25	0.18	0.95
2	100	20	0.69	0.65	1.00	0.00	0.00	0.33	0.69	0.07	-	-	-
2	100	50	0.70	0.67	1.00	0.00	0.00	0.17	0.57	0.24	-	-	-
2	100	100	0.67	0.65	1.00	0.00	0.00	0.18	0.81	0.45	-	-	-
			$\hat{K} = 1$					$\hat{K} = 3$					
K	N	T	t_a^*	t_b^*	$P_{\hat{E}}^c$	$ADF_{\hat{E}}^c$	$ADF_{\hat{F}}^c$		t_a^*	t_b^*	$P_{\hat{E}}^c$	MQ_c^c	MQ_f^c
2	20	20	0.51	0.43	0.83	0.22	0.44		0.53	0.46	0.75	0.00	0.00
2	20	50	0.57	0.50	0.94	0.25	0.27		0.66	0.60	0.84	0.00	0.00
2	20	100	0.62	0.55	1.00	0.45	0.40		0.67	0.62	0.99	0.00	0.00
2	50	20	0.69	0.65	0.93	0.47	0.46		0.70	0.67	0.96	0.00	0.00
2	50	50	0.82	0.78	0.99	0.27	0.27		0.88	0.87	1.00	0.00	0.00
2	50	100	0.85	0.82	1.00	0.37	0.36		0.87	0.86	1.00	0.00	0.00
2	100	20	0.75	0.74	0.96	0.22	0.43		0.80	0.78	0.99	0.00	0.00
2	100	50	0.82	0.81	0.99	0.24	0.22		0.94	0.93	1.00	0.00	0.00
2	100	100	0.86	0.84	1.00	0.47	0.31		0.94	0.93	1.00	0.00	0.00

See notes Table 2.1.

Table 2.6: Unit root test statistics for individual series of real exchange rates.

Country	ADF_i		$CADF_i$		$ADF_{\hat{E}}^c$	
	$Y_{i,t}^{\$}$	$Y_{i,t}^{DM}$	$Y_{i,t}^{\$}$	$Y_{i,t}^{DM}$	$Y_{i,t}^{\$}$	$Y_{i,t}^{DM}$
Austria	-1.4546	-1.9706	-1.4507	-1.7930	-2.0148**	-1.0179
Denmark	-1.7667	-1.8571	-2.6133	-2.3326	-2.5865**	-12.3785**
Finland	-1.4707	-1.7878	-0.8203	-1.5242	-1.7758*	-14.8259**
France	-1.3803	-1.9697	-2.1514	-3.5277**	-5.3264**	-7.9788**
Germany	-1.3537	-	-1.6298	-	-8.1950**	-
Greece	-1.3602	-2.0247	-1.2353	-2.4531	-0.1947	-0.5564
Italy	-1.1971	-1.9952	-1.5815	-2.2758	-1.9926**	-3.4528**
NL	-1.8189	-1.1789	-1.8504	0.4805	-12.4864**	-2.4592**
Norway	-1.6895	-1.8398	-1.4439	-2.4683	-3.2075**	-2.6420**
Portugal	-1.9609	-1.9522	-0.6468	-1.5935	-1.3352	-2.2390**
Spain	-1.0189	-1.9702	-0.6811	-1.2196	-1.6539*	-1.5992
Sweden	-1.3288	-1.9717	-2.3255	-1.5175	-13.3224**	-10.7084**
Switzerland	-2.0188	-3.0893**	-1.9543	-2.4054	-2.8777**	-14.4867**
UK	-3.2172**	-2.8032*	-1.7646	-2.7754	-3.7003**	-12.4749**
factor	-	-	-	-	-7.6314**	-7.749**

* indicates rejection at 10% significance level;

** indicates rejection at 5% significance level.

Table 2.7: Pooled unit root test statistics panels of real exchange rates.

	Pesaran (2007)	Moon and Perron (2004)		Bai and Ng (2004b)
	CIPS	t_a^*	t_b^*	$P_{\hat{E}}^c$
$q_{i,t}^{\$}$	-1.5821	-0.1214	-2.9890**	16.6123**
$q_{i,t}^{DM}$	-1.9543	-0.0358	-0.2076	18.4179**

	Breitung and Das (2008)		Sul (2007)	
	t_{rob}	t_{gls}	$t_{fglsrma}$	t_{crma}
$q_{i,t}^{\$}$	-0.9687	0.1789	0.5708	-0.6639
$q_{i,t}^{DM}$	-1.0903	-3.3914**	-4.0404**	0.6835

* indicates rejection at 10% significance level;

** indicates rejection at 5% significance level.

Cointegration Testing in Panels with Common Factors¹

3.1 Introduction

The effect of persistent cross-sectional dependence on panel unit root tests has been analyzed in some detail in Monte Carlo simulations (Banerjee, Marcellino, and Osbat, 2005) or by asymptotic analysis (Lyhagen, 2000; Pedroni and Urbain, 2001). First generation panel unit root tests are found to display dramatic size distortions or even worse to diverge with the cross-sectional dimension of the panel. To overcome these problems, new panel unit root tests have been proposed that model the possibly persistent cross-sectional dependency using common factor models (see Breitung and Pesaran, 2008, for a recent overview).

For tests for the null of no-cointegration, not much work has been done yet. Banerjee et al. (2004) conduct an extensive Monte Carlo study where they conclude that while all statistics investigated (residual-based tests or likelihood-based trace-type test) are affected, the presence of cross-member cointegration appears much less harmful for single-equation tests than for the panel version of the Johansen test. In many cases, in the presence of cointegration between members these tests cannot discriminate between cointegration among members and cointegration for a single member of the panel. Bai and Kao (2006) and Banerjee and Carrion-i Silvestre (2005) study tests for panel no-cointegration with cross-sectional dependence using residual-based tests for a single cointegration relationship. The error term of the cointegrating equation follows a common factor structure as in Bai and Ng (2004b). Urbain and Westerlund (2008) on the other hand studies analytically the issue of spurious regression in panels when the units are cointegrated along the cross-sectional dimension, i.e. when there is cross-member cointegration. In contrast to the spurious regression result for independent panels studied by Phillips and Moon (1999), Pedroni (1995) or Kao (1999), these estimators are often not consistent and actually converge to non-degenerate limiting distributions once the observed non-stationarity is generated by a reduced number of common stochastic trends.

¹This Chapter is based on Gengenbach et al. (2006).

This chapter builds on these results to study panel tests for no-cointegration when the cross-sectional dependence in the panel is modelled by a common factor structure as in Bai and Ng (2004b). Two different cases are considered that we believe are of theoretical and empirical relevance: (i) the case where the observed non-stationarity in the variables originates from cross-sectional common trends only; (ii) the case where we have both cross-sectional common and idiosyncratic stochastic trends. The spurious regression analysis for the first case reported in Urbain and Westerlund (2008) corresponds to the cross-member cointegration case. The second case is considered by Moon and Perron (2004) and Pesaran (2007) in the context of panel unit root analysis and excludes the existence of cross-unit cointegration in the panel since both components are $I(1)$.

For both classes of DGP's, we discuss the homogeneity restrictions on the cointegrating vectors resulting from the presence of common factor cointegration. These implications of the common factor cointegration are important reasons to propose a sequential approach whereby the data are decomposed into common and idiosyncratic components and the (no-)cointegration is tested for these components separately. Then, we study analytically the behavior of several tests for panel cointegration including residual-based panel no-cointegration tests proposed by Kao (1999) and Pedroni (1999, 2004a) that have been widely used in empirical work. For example, when the number of common factors generating the non-stationarity in the panel is kept fixed while the cross-sectional dimension of the panel increases, the Gaussian limiting results derived for the independent case are not valid anymore. Tests that are based on pooled or LSDV estimation of the underlying panel cointegration static regression in some cases diverge with \sqrt{N} and hence important size distortions can occur for moderate values of N . Group mean statistics are also affected and not asymptotically Gaussian anymore. These results complement and help to have a better understanding of some of the Monte Carlo results reported by Banerjee et al. (2004). We then examine the possibilities to test for no-cointegration, using residual-based panel tests applied to the defactored data.

The chapter is organized as follows: In Section 3.2 we present our model for panel no-cointegration with a common factor structure. In Section 3.3 we examine the asymptotic behavior of some residual-based panel no-cointegration tests when the data is generated by our DGP. Section 3.4 discusses defactoring the data prior to testing for various forms of no-cointegration when the data contains unobserved common factors. The finite sample behavior of the proposed approach is analyzed in Section 3.5. Conclusions are drawn in Section 3.6.

A note on notation: throughout the text, M is used to denote a generic positive number, not depending on T or N . For a matrix \mathbf{A} , $\mathbf{A} > 0$ denotes that \mathbf{A} is positive definite. Furthermore, $\|\mathbf{A}\| =$

$\text{trace}(\mathbf{A}'\mathbf{A})^{\frac{1}{2}}$. We write the integral $\int_0^1 B(r)\mathbf{d}r$ as $\int B$, and $\int_0^1 B(r)B(r)'\mathbf{d}r$ as $\int BB'$. Furthermore, \implies denotes weak convergence, and \xrightarrow{p} denotes convergence in probability. For any number x , $[x]$ denotes the largest integer smaller than x . For any variable $X_{i,t}$, $\tilde{X}_{i,t} = X_{i,t} - \frac{1}{T} \sum_{s=1}^T X_{i,s}$. Similarly, for any Brownian motion B , $\tilde{B} = B - \int B$. Throughout the chapter we employ sequential limit theory², where we consider $T \rightarrow \infty$ followed by $N \rightarrow \infty$.

3.2 The Model

We consider balanced panels with N cross-sectional units and T time-series observations, indexed by $i = 1, \dots, N$ and $t = 1, \dots, T$ respectively. For each unit in the panel we observe a $(1 + m)$ -dimensional vector of variables $Z_{i,t} = (Y_{i,t}, X'_{i,t})'$, where $Y_{i,t}$ is a scalar time series and $X_{i,t}$ is a m -vector time series³. We assume that the DGP for $Z_{i,t}$ has a common factor structure as e.g. in Bai and Ng (2004b), and we assume the presence of k common factors in the data. Furthermore, we assume the number of common factors to be fixed as $T, N \rightarrow \infty$ throughout the chapter. Our model is given by

$$Z_{i,t} = D_{i,t} + \Lambda_i F_t + E_{i,t}, \quad (3.1)$$

$t = 1, \dots, T, i = 1, \dots, N$. $D_{i,t}$ is an unobserved deterministic component such that either $D_{i,t} = 0$ for all i and t if there are no deterministic components present, $D_{i,t} = d_{0i}$ for all t if the data contains individual specific fixed effects, or $D_i = d_{0i} + d_{1i}t$ if the data contains individual specific deterministic linear time trends, where the coefficients d_{0i} and d_{1i} depend on i only. For the remainder of the chapter we assume $D_{i,t} = 0$ unless mentioned otherwise. The common component in $Z_{i,t}$ is given by F_t in (3.1). F_t is a k -vector of common $I(1)$ factors given by

$$F_t = F_{t-1} + f_t, \quad (3.2)$$

where $f_t = \Phi(L)\eta_t$, η_t is a sequence of $(k \times 1)$ $iid(0, I_k)$ random vectors, $\Phi(L) = \sum_{j=0}^{\infty} \Phi_j L^j$. The $(1 + m) \times k$ matrix of factor loadings Λ_i is assumed to be of full rank and block-diagonal, with block-diagonality corresponding to the partition of $Z_{i,t}$, and diagonal blocks denotes a λ'_{1i} and λ'_{2i} for the upper left and lower right block respectively.

²Although sequential limits are sometimes restrictive, they correspond to joint limits under certain restrictions (see e.g. Phillips and Moon, 1999). Furthermore, sequential asymptotic theory is well established in the literature.

³We assume that e.g. economic theory leads to a natural choice of Y in such a way that swapping some X for Y would not make sense. Nevertheless, the choice of Y is an interesting topic in cointegration analysis, but beyond the scope of this work.

As for the vectors of observations $Z_{i,t}$, we have partitions for the unobserved vector of common factors $F_t = (F_t^{Y'}, F_t^{X'})'$ where F_t^Y and F_t^X have k_Y and k_X elements respectively, and the partition of F_t corresponds to the structure of Λ_i , such that λ_{1i} is $k_Y \times 1$ and λ_{2i} is $k_X \times m$. The block diagonal structure for the factor loadings is necessary to ensure that $Y_{i,t}$ and $X_{i,t}$ are not cointegrated when the non-stationarity in the data is driven by the common factors alone. When the idiosyncratic components are non-stationary as well, this assumption on Λ_i might be relaxed and a more general structure can be considered.

For the idiosyncratic component in (3.1), $E_{i,t}$, we distinguish two cases, namely stationary and non-stationary idiosyncratic components. For the former case we have

$$E_{i,t} = e_{i,t}, \quad (3.3)$$

while in the latter case we assume

$$E_{i,t} = E_{i,t-1} + e_{i,t}, \quad (3.4)$$

where the stationary vector $e_{i,t} = \Gamma_i(L)\varepsilon_{i,t}$ with $\varepsilon_{i,t}$ being a sequence of random $iid(0, \Sigma_i)$ random vectors, $\Gamma_i(L) = \sum_{j=0}^{\infty} \Gamma_{ij}L^j$. Again, we partition $E_{i,t}$ conformable with the data $Z_{i,t}$, such that $E_{i,t} = (E_{i,t}^Y, E_{i,t}^{X'})'$, where $E_{i,t}^Y$ is a scalar time series and $E_{i,t}^X$ has m elements.

For the above given model we specify the following assumptions, where M denotes a generic positive real number:

Assumption 3.1 *Common factors: (i) $\eta_t \sim iid(0, I_k)$ with finite 4th moments, (ii) there is an M such that $\sum_{j=0}^{\infty} j \cdot \|\Phi_j\| < M$, (iii) $\text{rk}(\Phi(1)) = k$, (iv) $E\|F_0\| \leq M$.*

Assumption 3.2 *Factor loadings: (i) for non-random λ_{1i} and λ_{2i} , $\|\lambda_{1i}\| \leq M$ and $\|\lambda_{2i}\| \leq M$; for random λ_{1i} and λ_{2i} , $E\|\lambda_{1i}\|^4 \leq M$ and $E\|\lambda_{2i}\|^4 \leq M$, (ii) $N^{-1} \sum_{i=1}^N \Lambda_i' \Lambda_i \xrightarrow{p} \Sigma_{\Lambda} > 0$, (iii) for non-random λ_{1i} and λ_{2i} , $N^{-1} \sum_{i=1}^N \lambda_{1i} \neq 0$ and $N^{-1} \sum_{i=1}^N \lambda_{2i} \neq 0$; for random λ_{1i} and λ_{2i} , $E(\lambda_{1i}) \neq 0$ and $E(\lambda_{2i}) \neq 0$.*

Assumption 3.3 *Idiosyncratic components: for each $i = 1, \dots, N$, (i) $\varepsilon_{i,t} \sim iid(0, \Sigma_i)$ with finite 8th moments, and $\varepsilon_{i,t}$ and $\varepsilon_{j,s}$ are independent for any t, s and $i \neq j$, (ii) $E\|\varepsilon_{i,0}\| < M$, (iii) $\Gamma_i(L)$ fulfills the random coefficients and summability conditions from Phillips and Moon (1999), Assumptions 1 and 2 on p.1060 and p.1061 respectively, (iv) $\text{rk}(\Gamma_i(1)) = m + 1, \forall i$.*

Assumption 3.4 *The errors, η_t , $\varepsilon_{i,t}$, and the factor loadings Λ_i form mutually independent groups.*

Under the conditions of Assumption 3.1, the common factors F_t form a k -dimensional I(1) process and the possibility of cointegration between the common factors is excluded. The full rank assumption on the long-run covariance matrix of F_t could in fact be relaxed, as long as the diagonal blocks corresponding to the long-run covariances of F_t^Y and F_t^X have at least rank 1 each. The long-run covariance matrix of the common factors is given by $\Omega = \Phi(1)\Phi(1)' = \Xi + \Theta + \Theta'$, where $\Xi = \lim_{T \rightarrow \infty} \frac{1}{T} \sum_{t=1}^T \mathbf{E}(f_t f_t')$ and $\Theta = \lim_{T \rightarrow \infty} \frac{1}{T} \sum_{t=1}^T \mathbf{E}(f_t f_{t-1}')$ (see e.g. Phillips and Durlauf, 1986). Furthermore, an invariance principle holds such that

$$T^{-1/2} F_{\lfloor rT \rfloor} \Longrightarrow B_F(r) \quad \text{as } T \rightarrow \infty, \quad (3.5)$$

where B_F is a k -vector Brownian motion with covariance matrix Ω . Assumptions 3.2(i) and 3.2(ii) are standard assumptions for factor models and ensure that the factor loadings are identifiable. Assumption 3.2(iii) is needed for the spurious regression results when the non-stationarity in the data is only driven by the common factors. Assumption 3.3(iii) specifies that a panel functional central limit theorem holds for $S_{i,t} = \sum_{s=1}^t e_{i,t}$, which corresponds to $E_{i,t}$ in case the idiosyncratic components are non-stationary as in (3.4), or to its cumulative sum if (3.3) is true. The long-run covariance of $S_{i,t}$ is given by $\Psi_i = \Gamma_i(1)\Sigma_i\Gamma_i(1)' = \Upsilon_i + \Delta_i + \Delta_i'$, where $\Upsilon_i = \lim_{T \rightarrow \infty} \frac{1}{T} \sum_{t=1}^T \mathbf{E}(e_{i,t} e_{i,t}')$ and $\Delta_i = \lim_{T \rightarrow \infty} \frac{1}{T} \sum_{t=1}^T \mathbf{E}(e_{i,t} S'_{i,t-1})$, and an invariance principle ensures that

$$T^{-1/2} S_{i,\lfloor rT \rfloor} \Longrightarrow B_i(r) \quad \text{as } T \rightarrow \infty, \quad (3.6)$$

where B_i is a randomly scaled $(1+m)$ -vector Brownian motion with covariance matrix Ψ_i . Assumption 3.3(iv) ensures that the idiosyncratic terms do not cointegrate in case these are I(1) vectors.

The implications of these assumptions are best understood by considering the Beveridge-Nelson (BN) decomposition for F_t and for $E_{i,t} = \sum_{s=1}^t e_{i,s}$:

$$F_t = \Phi(1) \sum_{s=1}^t \eta_s + \Phi^*(L)(\eta_t - \eta_0) + F_0, \quad (3.7)$$

$$E_{i,t} = \Gamma_i(1) \sum_{s=1}^t \varepsilon_{i,s} + \Gamma_i^*(L)(\varepsilon_{i,t} - \varepsilon_{i,0}) + E_{i,0}, \quad (3.8)$$

where $\Phi^*(L) = \sum_{j=0}^{\infty} \Phi_j^* L^j$ with $\Phi_j^* = -\sum_{l=j+1}^{\infty} \Phi_l$, $\Gamma_i^*(L) = \sum_{j=0}^{\infty} \Gamma_{i,j}^* L^j$ with $\Gamma_{i,j}^* = -\sum_{l=j+1}^{\infty} \Gamma_{i,l}$, $\Phi^*(L)(\eta_t - \eta_0)$ and $\Gamma_i^*(L)(\varepsilon_{i,t} - \varepsilon_{i,0})$ are stationary with finite fourth order moments and F_0 and $E_{i,0}$ are $O_p(1)$ by assumption.

If (3.3) is true the idiosyncratic data components are $I(0)$, and the $I(1)$ trends of the common factors contained in $\Lambda_i \Phi(1) \sum_{s=1}^t \eta_s$ drive the non-stationarity in the data. Then, we might observe *cross-member cointegration* between some $Y_{i,t}$ and $Y_{j,t}$, and between some $X_{i,t}$ and $X_{j,t}$ for some i, j , $i \neq j$, the exact cointegration structure depending on the individual loadings. The assumption on the block-diagonal structure of the factor loadings Λ_i in turn implies that we have separation in a cointegrating system, see Hecq, Palm, and Urbain (2002). Note that the assumption of cointegration between $Y_{i,t}$ and $X_{i,t}$ would only be possible if the common factors F_t^Y and F_t^X would cointegrate, which is ruled out by Assumption 3.1 from which the full rank of the long-run covariance matrix of F_t follows.

When $E_{i,t}$ is given by (3.4), both common and idiosyncratic data components are non-stationary. Furthermore, the idiosyncratic components do not cointegrate along the cross-section. Hence, we do not have cointegration “within” units, e.g. between $Y_{i,t}$ or $X_{i,t}$. The BN decomposition of the $Z_{i,t}$ is easily obtained from (1) and (3.7-3.8) and shows that the non-stationarity of $Z_{i,t}$ stems from the term $\Lambda_i \Phi(1) \sum_{s=1}^t \eta_s + \Gamma_i(1) \sum_{s=1}^t \varepsilon_{i,s}$.

Remark 1. To investigate tests for no-cointegration we need to maintain the assumption that there does not exist a full column rank matrix β'_i such that $\beta'_i Z_{it} \sim I(0)$. Different cases can be considered. Two cases are important, namely one with cross-member cointegration where we have $I(1)$ common factors and $I(0)$ idiosyncratic terms and one where the panel units contain common stochastic trends, but do not cointegrate even along the cross-sectional dimension so that both the common and the idiosyncratic components are $I(1)$.

Remark 2: Heterogeneity and cross-sectional dependence. With $I(1)$ common factors as well as $I(1)$ idiosyncratic components, we actually have two different sets of possible cointegrating vectors that would annihilate the idiosyncratic and the common $I(1)$ stochastic trends respectively, see also the discussion in Gregoir (2005) and Breitung and Pesaran (2008). Combining (1) and (3.7)-(3.8), the resulting BN representation of $Z_{i,t}$ shows that it will not be easy to annihilate both. In particular, cointegrating vector(s), say δ , that annihilate the common $I(1)$ components should lie in the left null space of Λ_i , that is $\delta \Lambda_i \Phi(1) = 0$ as $\Phi(1)$ is of full rank by Assumption 3.1, while those for the idiosyncratic components, say γ'_i would have to lie in the left null space of $\Gamma_i(1)$, i.e. $\gamma'_i \Gamma_i(1) = 0$. If the intersection of these left null spaces is empty, there does not exist a cointegrating relationship that annihilates both the unit roots from the common stochastic trends and those of the idiosyncratic terms. In this case none of the $Z_{i,t}$ vectors are cointegrated. The components taken in isolation could be cointegrated though.

In fact, there is an important trade-off between the degree of heterogeneity that can be allowed for and the existence of cross-sectional dependence modeled by common factors.

Without loss of generality, consider the following simple bivariate DGP where we have a single I(1) common factor in Y and a single I(1) common factor in X :

$$Y_{i,t} = \lambda_{1,i}F_t^Y + E_{i,t}^Y, \quad (3.9)$$

$$X_{i,t} = \lambda_{2,i}F_t^X + E_{i,t}^X, \quad (3.10)$$

from which we see that any linear combination can be written as

$$Y_{i,t} - \beta_i X_{i,t} = \lambda_{1,i} \left(F_t^Y - \frac{\beta_i \lambda_{2,i}}{\lambda_{1,i}} F_t^X \right) + E_{i,t}^Y - \beta_i E_{i,t}^X. \quad (3.11)$$

For the linear combination $(1, -\beta_i)$ to be a cointegrating vector such that $Y_{i,t} - \beta_i X_{i,t} \sim I(0)$, two conditions need to hold, namely (i) $(F_t^Y - \frac{\beta_i \lambda_{2,i}}{\lambda_{1,i}} F_t^X) \sim I(0)$ (ii) $(E_{i,t}^Y - \beta_i E_{i,t}^X) \sim I(0)$. Given that here we have only two I(1) common factors, there can be at most a single linear cointegrating combination between these factors and hence $\frac{\beta_i \lambda_{2,i}}{\lambda_{1,i}}$ should be the same $\forall i$. Three different cases are compatible with a constant (over i) ratio:

1. With *homogeneity* of the factor loadings and of the vector β_i the ratio $\frac{\beta_i \lambda_{2,i}}{\lambda_{1,i}} = \frac{\beta \lambda_2}{\lambda_1}$ does not depend on i . A similar restriction is considered by Gregoir (2005). Another possibility is homogeneity of β_i and constancy of the ratios of the factor loadings $\frac{\lambda_{2,i}}{\lambda_{1,i}}$ for all i which is also excluded by Assumptions 3.1-3.4.
2. The second case allows for some degree of heterogeneity: the factor loadings vary with β_i such that the ratio $\frac{\beta_i \lambda_{2,i}}{\lambda_{1,i}}$ is constant across i . This is excluded by Assumptions 3.1-3.4 where the loadings and Ψ_i are assumed to vary independently from each other.
3. A third case arises when for all i the variables $Y_{i,t}$ and $X_{i,t}$ have a single common source of nonstationarity F_t only. The idiosyncratic component is assumed to be stationary (or could be cointegrated with cointegrating vector β_i). In this case, $Y_{i,t}$ and $X_{i,t}$ are cointegrated with $\beta_i = \frac{\lambda_{1,i}}{\lambda_{2,i}}$. It is ruled out by the assumption of block-diagonality of Λ_i , but it would be a natural alternative hypothesis to the null of no-cointegration. Homogeneity of the cointegrating vector then arises if $\frac{\lambda_{1,i}}{\lambda_{2,i}}$ is constant across entities i .

To conclude, if we allow for almost unrestricted (under Assumptions 3.1-3.4) heterogeneity, the existence of cointegrating relations that annihilate *both* the common and idiosyncratic I(1) stochastic

trends is very unlikely. The consequences of this for testing of the null of no-cointegration in this factor set-up will be mentioned in Section 3.4.

Remark 3. A similar framework is also, independently of the present work, proposed by Dees, di Mauro, Pesaran, and Smith (2007) for the study of macroeconomic linkages within the Euro area. The purpose of their work is however different as no attempt to discuss tests for cointegration is made. This work is thus complementary to theirs.

3.3 The behavior of panel residual based tests

The purpose of this section is to study, given the set-up introduced in the preceding section, the asymptotic behavior of some standard and popular panel tests for no-cointegration. The statistics we consider are designed to test for the presence of a single cointegration relationship between $Y_{i,t}$ and $X_{i,t}$.⁴ Kao (1999) considers a homogenous cointegrating vector, whereas Pedroni (1999) allows for heterogeneity. However, both rely on the cross-sectional independence of the panel unit to derive asymptotic normality for their test statistics.

3.3.1 Kao (1999)

Kao (1999) proposes to estimate the homogeneous cointegrating relationship by pooled regression allowing for individual fixed effects. The regression equation is given by

$$Y_{i,t} = \alpha_i + \beta X_{i,t} + u_{i,t}, \quad (3.12)$$

where β and $X_{i,t}$ are row and column vectors respectively, and $u_{i,t}$ is a regression error. The least squared dummy variable (LSDV) estimator for β is

$$\tilde{\beta} = \left(\sum_{i=1}^N \sum_{t=1}^T \tilde{Y}_{i,t} \tilde{X}'_{i,t} \right) \left(\sum_{i=1}^N \sum_{t=1}^T \tilde{X}_{i,t} \tilde{X}'_{i,t} \right)^{-1},$$

where $\tilde{Y}_{i,t} = Y_{i,t} - \frac{1}{T} \sum_{s=1}^T Y_{i,s}$ and $\tilde{X}_{i,t} = X_{i,t} - \frac{1}{T} \sum_{s=1}^T X_{i,s}$. The residuals from this first stage regression $\tilde{u}_{i,t} = \tilde{Y}_{i,t} - \tilde{\beta} \tilde{X}_{i,t}$ will still contain a unit root under the null hypothesis of no cointegration.

⁴This is a restrictive assumption that we however will make in the sequel by assuming the existence of a single cointegrating vector. Approaches that allow for more than one cointegrating vector, are reviewed in Breitung and Pesaran (2008).

We now estimate a pooled DF regression

$$\Delta \tilde{u}_{i,t} = (\rho - 1)\tilde{u}_{i,t-1} + v_{i,t}, \quad (3.13)$$

where the pooled ordinary least squares (POLS) estimator of $(\rho - 1)$ is given by

$$(\tilde{\rho} - 1) = \left(\sum_{i=1}^N \sum_{t=2}^T \Delta \tilde{u}_{i,t} \tilde{u}_{i,t-1} \right) \left(\sum_{i=1}^N \sum_{t=2}^T \tilde{u}_{i,t-1}^2 \right)^{-1}.$$

Kao's tests are based on $\tilde{\rho}$ and the corresponding t -statistic

$$t_{\tilde{\rho}} = (\tilde{\rho} - 1) \left(\hat{s}_{\tilde{u}}^2 \left(\sum_{i=1}^N \sum_{t=2}^T \tilde{u}_{i,t-1}^2 \right)^{-1} \right)^{-\frac{1}{2}},$$

where $\hat{s}_{\tilde{u}}^2 = N^{-1}T^{-1} \sum_{i=1}^N \sum_{t=2}^T (\Delta \tilde{u}_{i,t-1} - (\tilde{\rho} - 1)\tilde{u}_{i,t-1})^2$, corrected for endogeneity and serial correlation. When the panel units are cross-sectionally independent, the test statistics are asymptotically normally distributed as $T \rightarrow \infty$ followed by $N \rightarrow \infty$. However, for the model given by (3.1), (3.2) and (3.3) or (3.4), this assumption is clearly violated. Using the results reported in Lemmas 3.B.1-3.B.3 in Appendix A, we obtain the following limit results, where $\text{vec}(\int \mathbf{d}B_{F\Lambda} B'_{F\Lambda}) = \check{\mathbf{A}} \text{vec}(\int \mathbf{d}B_F B'_F)$, $\text{vec}(\Theta_{F\Lambda}) = \check{\mathbf{A}} \text{vec}(\Theta)$, $\text{vec}(\int B_{F\Lambda} B'_{F\Lambda}) = \check{\mathbf{A}} \text{vec}(\int B_F B'_F)$, $\text{vec}(\int \mathbf{d}B_{F\Lambda} \tilde{B}'_{F\Lambda}) = \check{\mathbf{A}} \text{vec}(\int \mathbf{d}B_F \tilde{B}'_F)$ and $\text{vec}(\int \tilde{B}_{F\Lambda} \tilde{B}'_{F\Lambda}) = \check{\mathbf{A}} \text{vec}(\int \tilde{B}_F \tilde{B}'_F)$, and $\check{\mathbf{A}} = \text{plim}_{N \rightarrow \infty} \frac{1}{N} \sum_{i=1}^N (\Lambda_i \otimes \Lambda_i)$, Ψ^{YX} is the average long run covariance between the idiosyncratic errors in $Y_{i,t}$ and $X_{i,t}$, Ψ^{XX} is the average long run covariance matrix of the idiosyncratic errors in $X_{i,t}$, and B_F and B_i are given in Equations (3.5) and (3.6), respectively.

Proposition 3.1 *Given Assumptions 3.1, 3.2, 3.3 and 3.4:*

(A) Consider the model given by (3.1), (3.2) and (3.3),

$$(a) \tilde{\beta} \implies (\int \tilde{B}_{F\Lambda}^Y \tilde{B}_{F\Lambda}^{X'}) (\int \tilde{B}_{F\Lambda}^X \tilde{B}_{F\Lambda}^{X'})^{-1} = \tilde{\mathbf{b}}_A \text{ as } T, N \rightarrow \infty \text{ sequentially,}$$

$$(b) T(\tilde{\rho} - 1) \implies \frac{(1, -\tilde{\mathbf{b}}_A) (\int \mathbf{d}B_{F\Lambda} \tilde{B}'_{F\Lambda} + \Theta_{F\Lambda} + \gamma_1 - \Upsilon) (1, -\tilde{\mathbf{b}}_A)'}{(1, -\tilde{\mathbf{b}}_A) (\int \tilde{B}_{F\Lambda} \tilde{B}'_{F\Lambda}) (1, -\tilde{\mathbf{b}}_A)'} \text{ as } T, N \rightarrow \infty \text{ sequentially, where}$$

$$\gamma_1 = \mathbf{E}(\gamma_{i1}) \text{ and } \gamma_{i1} = \mathbf{E}(\tilde{e}_{i,t-1} \tilde{e}'_{i,t}),$$

$$(c) t_{\tilde{\rho}} \text{ diverges at rate } \sqrt{N} \text{ as } T, N \rightarrow \infty \text{ sequentially.}$$

(B) Consider the model given by (3.1), (3.2) and (3.4),

$$(a) \tilde{\beta} \implies (\int \tilde{B}_{F\Lambda}^Y \tilde{B}_{F\Lambda}^{X'} + \frac{1}{6} \Psi^{YX}) (\int \tilde{B}_{F\Lambda}^X \tilde{B}_{F\Lambda}^{X'} + \frac{1}{6} \Psi^{XX})^{-1} = \tilde{\mathbf{b}}_B \text{ as } T, N \rightarrow \infty \text{ sequentially,}$$

- (b) $T(\tilde{\rho} - 1) \implies \frac{(1, -\tilde{\mathbf{b}}_B)(\int \mathbf{d}B_{F\Lambda}\tilde{B}'_{F\Lambda} + \Theta_{F\Lambda} - \frac{1}{2}\Psi + \Delta)(1, -\tilde{\mathbf{b}}_B)'}{(1, -\tilde{\mathbf{b}}_B)(\int \tilde{B}_{F\Lambda}\tilde{B}'_{F\Lambda} + \frac{1}{6}\Psi)(1, -\tilde{\mathbf{b}}_B)'}$ as $T, N \rightarrow \infty$ sequentially,
- (c) $t_{\tilde{\rho}}$ diverges at rate \sqrt{N} as $T, N \rightarrow \infty$ sequentially.

Proof: see Appendix 3.B.

The results summarized in Proposition 3.1 are clearly in contrast to the asymptotic normality Kao (1999) derives for the tests statistics for independent panels, although we have not yet considered corrections for serial correlation and endogenous regressors. Results $A(a)$ and $B(a)$ are similar to those derived by Urbain and Westerlund (2008) for the pooled least squares estimator (PLS). This is in sharp contrast with the \sqrt{N} consistency of the LSDV estimator in the case of a spurious regression estimated from independent panel data, see Phillips and Moon (1999). The statistics proposed by Kao (1999) rely on this consistency, namely on the fact that $\tilde{\beta} \xrightarrow{p} \Psi^{YX}\Psi^{XX-1}$ where Ψ^{YX} is the average long-run covariance between the errors driving $X_{i,t}$ and those driving $Y_{i,t}$ and Ψ^{XX} is the average long covariance matrix of the $X_{i,t}$'s. The presence of common factors destroys this property and consequently the asymptotic normality of these estimators and of the statistics relying on this result. For the case of stationary idiosyncratic components, our findings are similar to the spurious regression results from time-series analysis. With non-stationary idiosyncratic components we obtain some mixture of time-series and panel spurious regression results in the limiting distributions. The tests are inconsistent when the data has a common factor structure, and size distortions have to be expected which will increase with N . The nuisance parameters in the limiting distributions given in Proposition 3.1 introduced by the serial correlation in the common factors and idiosyncratic components can be corrected for non-parametrically, i.e. the composite effect of $\Theta_{F\Lambda} + \gamma_1 - \Upsilon$ or $\Theta_{F\Lambda} + \Delta$ can be accounted for. However, it is not possible to identify nuisance parameters associated with the common factors or the idiosyncratic components individually. So, the covariance of $\tilde{B}_{F\Lambda}$ as well as the average long-run covariance matrix of idiosyncratic stochastic trends, Ψ , will in general remain in the limits. The limit of $t_{\tilde{\rho}}$ will be the product of \sqrt{N} , the limit of $(\tilde{\rho} - 1)$ and the limit of the standard deviation of $(\tilde{\rho} - 1)$. Whereas the latter is positive, the driving factor of the limiting distribution of $(\tilde{\rho} - 1)$ is $\frac{\int \mathbf{d}B_{F\Lambda}\tilde{B}'_{F\Lambda}}{\int \tilde{B}_{F\Lambda}\tilde{B}'_{F\Lambda}}$ which has a negative expected value. Thus, $t_{\tilde{\rho}}$ can be expected to diverge to $-\infty$.

3.3.2 Pedroni (1999)

Pedroni (1999) allows for heterogeneity of the slope coefficient β in the cointegration relationship (3.12), which becomes β_i . He proposes to estimate a first stage regression individually for each panel

member to obtain an estimate of β_i ,

$$\tilde{\beta}_i = \left(\sum_{t=1}^T \tilde{Y}_{i,t} \tilde{X}'_{i,t} \right) \left(\sum_{t=1}^T \tilde{X}_{i,t} \tilde{X}'_{i,t} \right)^{-1}. \quad (3.14)$$

Pedroni (1999) proposes two classes of statistics, namely those based on the within-dimension denoted as “panel” statistics, and those based on the between-dimension denoted as “group mean” statistics. For the former group, the residuals from the first stage regression, $\tilde{u}_{i,t} = \tilde{Y}_{i,t} - \tilde{\beta}_i \tilde{X}_{i,t}$, are stacked and a pooled DF regression is estimated as in (3.13).⁵ The group mean statistics are based on averages of individual unit root statistics, derived from

$$\Delta \tilde{u}_{i,t} = (\rho_i - 1) \tilde{u}_{i,t-1} + v_{i,t}, \quad (3.15)$$

to obtain

$$(\tilde{\rho}_i - 1) = \left(\sum_{t=2}^T \Delta \tilde{u}_{i,t} \tilde{u}_{i,t-1} \right) \left(\sum_{t=2}^T \tilde{u}_{i,t-1}^2 \right)^{-1}.$$

Consider now the panel-rho statistic denoted by $Z_{\tilde{\rho}_{NT-1}}$ and the group-mean rho statistic $\tilde{Z}_{\tilde{\rho}_{NT-1}}$ given by

$$Z_{\tilde{\rho}_{NT-1}} = \left(\sum_{i=1}^N \sum_{t=2}^T (\Delta \tilde{u}_{i,t} \tilde{u}_{i,t-1} - \hat{\lambda}_i) \right) \left(\sum_{i=1}^N \sum_{t=2}^T \tilde{u}_{i,t-1}^2 \right)^{-1}, \quad (3.16)$$

and

$$\tilde{Z}_{\tilde{\rho}_{NT-1}} = \sum_{i=1}^N \left(\left(\sum_{t=2}^T (\Delta \tilde{u}_{i,t} \tilde{u}_{i,t-1} - \hat{\lambda}_i) \right) \left(\sum_{t=2}^T \tilde{u}_{i,t-1}^2 \right)^{-1} \right), \quad (3.17)$$

with $\hat{\lambda}_i = T^{-1} \sum_{s=1}^J \omega_{sJ} \sum_{t=s+1}^T \tilde{v}_{i,t} \tilde{v}_{i,t-s}$ where $\tilde{v}_{i,t}$ are the residuals of the second stage regression, and J and ω_{sJ} are suitable bandwidth and kernel functions, respectively. For these 2 statistics, we obtain the following limiting results:

Proposition 3.2 *Given Assumptions 3.1, 3.2, 3.3 and 3.4:*

(A) *Consider the model given by (3.1), (3.2) and (3.3),*

$$(a) \tilde{\beta}_i \implies (\lambda'_{1i} (\int \tilde{B}_F^Y \tilde{B}_F^{X'} \lambda_{2i}) (\lambda'_{2i} (\int \tilde{B}_F^X \tilde{B}_F^{X'}) \lambda_{2i})^{-1}) = \tilde{\mathbf{b}}_{iA} \text{ as } T \rightarrow \infty,$$

$$(b) TZ_{\tilde{\rho}_{NT-1}} \implies \frac{\sum_{i=1}^N \lambda'_{1i} L'_{11} \int \mathbf{d}Q_F \tilde{Q}'_F L_{11} \lambda_{1i}}{\sum_{i=1}^N \lambda'_{1i} L'_{11} \int \tilde{Q}_F \tilde{Q}'_F L_{11} \lambda_{1i}} \text{ as } T \rightarrow \infty,$$

$$(c) T\tilde{Z}_{\tilde{\rho}_{NT-1}} \implies \sum_{i=1}^N \frac{\lambda'_{1i} L'_{11} \int \mathbf{d}Q_F \tilde{Q}'_F L_{11} \lambda_{1i}}{\lambda'_{1i} L'_{11} \int \tilde{Q}_F \tilde{Q}'_F L_{11} \lambda_{1i}} \text{ as } T \rightarrow \infty,$$

⁵Note that although the estimated DF equation is the same for Kao (1999) and Pedroni (1999), the residuals used in the estimation are obtained from individual regressions instead of a pooled one.

where $\tilde{Q} = \tilde{W}_F^Y - (\int \tilde{W}_F^Y \tilde{W}_F^{X'}) (\int \tilde{W}_F^X \tilde{W}_F^{X'})^{-1} \tilde{W}_F^X$, \tilde{W}_F is a demeaned k -vector standard Brownian motion, and L_{11} is upper left element of L , the block triangular decomposition of $\Omega = L'L$.

(B) Consider the model given by (3.1), (3.2) and (3.4),

$$(a) \quad \begin{aligned} \tilde{\beta}_i &\implies (\lambda'_{1i} \int \tilde{B}_F^Y \tilde{B}_F^{X'} \lambda_{2i} + \int \tilde{B}_i^Y \tilde{B}_i^{X'} + \lambda'_{1i} \int \tilde{B}_F^Y \tilde{B}_i^{X'} + \int \tilde{B}_i^Y \tilde{B}_F^{X'} \lambda_{2i}) \\ &(\lambda'_{2i} \int \tilde{B}_F^X \tilde{B}_F^{X'} \lambda_{2i} + \int \tilde{B}_i^X \tilde{B}_i^{X'} + \lambda'_{2i} \int \tilde{B}_F^X \tilde{B}_i^{X'} + \int \tilde{B}_i^X \tilde{B}_F^{X'} \lambda_{2i})^{-1} \\ &= \tilde{\mathbf{b}}_{iB} \end{aligned}$$

as $T \rightarrow \infty$,

$$(b) \quad TZ_{\tilde{\rho}_{NT-1}} \implies \frac{\sum_{i=1}^N (1 - \tilde{\mathbf{b}}_{iB}) (\Lambda'_i (\int \mathbf{dB}_F \tilde{B}'_F) \Lambda'_i + \int \mathbf{dB}_i \tilde{B}'_i + \Lambda_i \int \mathbf{dB}_F \tilde{B}'_i + \int \mathbf{dB}_i \tilde{B}'_F \Lambda'_i) (1 - \tilde{\mathbf{b}}_{iB})'}{\sum_{i=1}^N (1 - \tilde{\mathbf{b}}_{iB}) (\Lambda_i \int \tilde{B}_F \tilde{B}'_F \Lambda'_i + \int \tilde{B}_i \tilde{B}'_i + \Lambda_i \int \tilde{B}_F \tilde{B}'_i + \int \tilde{B}_i \tilde{B}'_F \Lambda'_i) (1 - \tilde{\mathbf{b}}_{iB})'} \quad \text{as } T \rightarrow \infty,$$

$$(c) \quad T\tilde{Z}_{\tilde{\rho}_{NT-1}} \implies \sum_{i=1}^N \frac{(1 - \tilde{\mathbf{b}}_{iB}) (\Lambda'_i (\int \mathbf{dB}_F \tilde{B}'_F) \Lambda'_i + \int \mathbf{dB}_i \tilde{B}'_i + \Lambda_i \int \mathbf{dB}_F \tilde{B}'_i + \int \mathbf{dB}_i \tilde{B}'_F \Lambda'_i) (1 - \tilde{\mathbf{b}}_{iB})'}{(1 - \tilde{\mathbf{b}}_{iB}) (\Lambda_i \int \tilde{B}_F \tilde{B}'_F \Lambda'_i + \int \tilde{B}_i \tilde{B}'_i + \Lambda_i \int \tilde{B}_F \tilde{B}'_i + \int \tilde{B}_i \tilde{B}'_F \Lambda'_i) (1 - \tilde{\mathbf{b}}_{iB})'} \quad \text{as } T \rightarrow \infty.$$

Proof: see Appendix 3.B.

For the panel-rho and group-mean-rho statistics Pedroni (1999, 2004a) derives asymptotic normality when they are properly standardized. In particular, $\sqrt{NT}Z_{\tilde{\rho}_{NT-1}} - \sqrt{N}\theta_2\theta_1^{-1}$ and $N^{-\frac{1}{2}}T\tilde{Z}_{\tilde{\rho}_{NT-1}} - \sqrt{N}\tilde{\theta}_1$ are asymptotically normally distributed for independent panels, where θ_1 , θ_2 and $\tilde{\theta}_1$ are means of functionals of Brownian motions (for details see Pedroni, 2004a). The results from Proposition 3.2 indicate that under the DGP we consider, $TZ_{\tilde{\rho}_{NT-1}}$ and $N^{-1}T\tilde{Z}_{\tilde{\rho}_{NT-1}}$ converge, so that the two test-statistics diverge at rate \sqrt{N} when standardized as above. Furthermore, due to the presence of the common factors, the individual statistics will not be independent along the cross-section, so that the use of a CLT to derive asymptotic normality of the average statistic will be invalid. The result is similar to that derived by Lyhagen (2000) for the Im et al. (2003) (IPS) statistics. Also, for independent panels the distributions of $Z_{\tilde{\rho}_{NT-1}}$ and $\tilde{Z}_{\tilde{\rho}_{NT-1}}$ will be nuisance parameter free. For the DGP we consider, this is not true in general. Although the composite effect of serial correlation in the common factors and idiosyncratic components can be corrected for non-parametrically, nuisance parameters coming only from the common factors or from the idiosyncratic components cannot be identified. So, the limiting distributions will in general depend on the long-run covariances of the common and/or idiosyncratic stochastic trends. A special case arises when there is a single common factor in $Y_{i,t}$ and the idiosyncratic components are stationary. Then, $\lambda_{1i}L_{11}$ will cancel from the limits given in Proposition 3.2 A (b) and (c).

3.4 A two-step procedure to test for (no)cointegration in the presence of common factors

As shown in Section 3.3 standard panel tests for the null of no-cointegration suffer from serious problems when applied to data with a common factor structure. To tackle the problem we propose a simple approach based on the Bai and Ng (2004b) PANIC methodology.⁶

A related, albeit different, idea is exploited in the work of Banerjee and Carrion-i Silvestre (2005), who assume a factor structure for the *disturbance* of a panel static regression model:

$$\begin{aligned} Y_{i,t} &= \alpha_i + \beta_i X_{i,t} + u_{i,t} \\ u_{i,t} &= \gamma_i' F_t + E_{i,t}, \end{aligned}$$

where F_t and $E_{i,t}$ are the common factors and the idiosyncratic components respectively that can be either $I(1)$ or $I(0)$. A similar framework is used by Bai and Kao (2004) for the estimation of a cointegrating relationship in the presence of common factors. Under some conditions that bound the possible heterogeneity, this framework leads to panel statistics for the null of no-cointegration that have the same distribution as panel unit root tests and hence are not affected by the number of regressors.⁷

Consider the simple bivariate DGP (3.9)-(3.10)⁸ and address the issue of no-cointegration at three different levels.

- (i) Testing for idiosyncratic component no-cointegration. This would mean to test the null hypothesis that $(E_{i,t}^Y - \beta_i E_{i,t}^X) \sim I(1)$ against $(E_{i,t}^Y - \beta_i E_{i,t}^X) \sim I(0)$,
- (ii) Testing for common factor no-cointegration. This would boil down to testing the null that $(F_t^Y - \delta F_t^X) \sim I(1)$ against $(F_t^Y - \delta F_t^X) \sim I(0)$,
- (iii) Testing for panel no-cointegration, that is testing the null that $Y_{i,t} - \beta_i X_{i,t} \sim I(1)$ against $Y_{i,t} - \beta_i X_{i,t} \sim I(0)$. Rejecting the null of no-cointegration requires evidence of idiosyncratic component cointegration with cointegrating vector $(1, -\beta_i)$ as well as of common factor cointegration with cointegrating vector $(1, -\frac{\beta_i \lambda_{2,i}}{\lambda_{1,i}})$ which should be constant across the individuals i .

⁶Wagner and Müller-Fürstenberger (2004) use similar ideas in an empirical study of the Kuznets curve.

⁷A similar set-up is retained by Westerlund (2005) who proposes Durbin-Hausman tests for cointegration in panels.

⁸The discussion extends to a more general set-up.

Provided the components have been extracted from the data, case (i) is tested using standard panel tests for no-cointegration given in (3.16) and (3.17). Case (ii) can be investigated using standard time series no-cointegration tests such as the Johansen rank test. Case (iii) is slightly more problematic since rejecting the null of panel no-cointegration requires not only factor and idiosyncratic cointegration, but also cointegrating vector(s) for the factors of a very specific form. The restrictions between the cointegrating coefficients result from the common factor structure and from the condition that the left null spaces of the common factor and idiosyncratic component cointegration must have a non-empty intersection.

There is however a useful indirect way of addressing this question. Consider (3.11) and write $(F_t^Y - \frac{\beta_i \lambda_{2,i}}{\lambda_{1,i}} F_t^X) \equiv G_t$ and $(E_{i,t}^Y - \beta_i E_{i,t}^X) \equiv E_{i,t}^*$ such that (3.11) becomes:

$$Y_{i,t} - \beta_i X_{i,t} = \lambda_{1,i} G_t + E_{i,t}^* \quad (3.18)$$

which is nothing but the parametrization considered in Banerjee and Carrion-i Silvestre (2005). Under this parametrization, $(1, -\frac{\beta_i \lambda_{2,i}}{\lambda_{1,i}})$ will be a cointegrating vector for the common factors if and only if $G_t \sim I(0)$. One may consequently investigate the hypothesis of panel cointegration using the approach proposed by these authors.

Now we shall outline a sequential testing procedure based on the factor structure under (3.1), (3.2) and (3.3) or (3.4) that does not restrict the heterogeneity. The approach starts with a decomposition of the data into common factors and idiosyncratic components as in Bai and Ng (2004b). It investigates the cointegration properties of the extracted factors and components.

- Step 1. Conduct a PANIC analysis of each variable $X_{i,t}$ and $Y_{i,t}$ individually to extract the common factors, e.g. using the principal components approach advocated by Bai and Ng (2004b). Test for unit roots in both the factors and the idiosyncratic components using the Bai and Ng (2004b) or the Breitung and Das (2008) approach.
- Step 2. a. If $I(1)$ common factors and $I(0)$ idiosyncratic components are detected, we face the situation of *cross-member cointegration* and consequently the nonstationarity in the panel is entirely due to a reduced number of common stochastic trends. Cointegration between $Y_{i,t}$ and $X_{i,t}$ can only occur if the common factors for $Y_{i,t}$ cointegrate with those of $X_{i,t}$. The null of no-cointegration between these estimated factors can be tested using a Johansen type of likelihood ratio test for example.
- b. If $I(1)$ common factors and $I(1)$ idiosyncratic components are detected, we carry out step

a on the estimated common factors and we will work with *defactored* series. In contrast to Banerjee and Carrion-i Silvestre (2005) however, who defactor the residuals from a static regression (11) we defactor separately $Y_{i,t}$ and $X_{i,t}$. The defactored $Y_{i,t}$ (e.g. the estimated idiosyncratic components) is simply obtained as $\widehat{E}_{i,t}^Y = \sum_{s=1}^t \widehat{e}_{i,s}^Y = \sum_{s=1}^t (\Delta Y_{i,s} - \widehat{\lambda}'_{1,i} \widehat{f}_s)$ where \widehat{f}_s is a consistent factor estimate of f_t in (3.2) and $\widehat{\lambda}'_{1,i}$ a consistent estimate of the loading. Testing for no-cointegration between the defactored data can be conducted using standard panel tests for no-cointegration such as those of Pedroni (1999, 2004a) given in (3.16) and (3.17).

The rejection of no-cointegration between $Y_{i,t}$ and $X_{i,t}$ only occurs if the tests for both common factor and idiosyncratic no-cointegration reject. However, this is a necessary condition. If the three restrictions mentioned under (iii) hold as well for the cointegrating vectors, panel cointegration will hold. If the outcome of step 2.b is that both the common factors and the idiosyncratic components cointegrate one might want to jointly or sequentially test the restrictions on the cointegrating vectors. The required tests are not available with the exception of a homogeneity test on the idiosyncratic component cointegrating vectors proposed by Pedroni (2004b). Comparing point estimates of the parameters involved could yield further insight into the structure of the model. Formal testing of panel no-cointegration could be done using the Banerjee and Carrion-i Silvestre (2005) test.

Remark 4. The sequential panel no-cointegration test outlined in Steps 1-2 is a multiple comparison procedure. Panel no-cointegration is rejected if both the hypotheses of common factor no-cointegration and idiosyncratic component no-cointegration are rejected and the restrictions between the cointegrating vector parameters are not rejected. An approximate test of the joint hypothesis could use the Bonferroni procedure (see e.g. Savin, 1980). In a Monte Carlo simulation, the joint hypothesis test of factor and idiosyncratic component (no-)cointegration is found to be undersized due to the idiosyncratic component (no-)cointegration. Its power properties are shown to be fine. The results are available upon request.

Remark 5. The theoretical justification for this sequential procedure is analogous to that of the PANIC panel unit root analysis. Since the DGP implies that all series have a Bai and Ng (2004b) representation, we proceed by analogy with the results derived in Bai and Ng (2004b). Provided the number of common factors is known or consistently selected using one of the consistent selection procedures discussed in Bai and Ng (2004b), then it holds that $T^{-1/2} \sum_{s=2}^t \widehat{e}_{i,s}^Y =$

$T^{-1/2} \sum_{s=2}^t e_{i,s}^Y + O_p(C_{NT}^{-1})$ where $\hat{e}_{i,t}^Y$ is the estimated idiosyncratic component, $\hat{e}_{i,t}^Y = \Delta Y_{i,s} - \hat{\lambda}'_{1,i} \hat{f}_s$, \hat{f}_s a consistent factor estimate of f_t , $\hat{\lambda}'_{1,i}$ a consistent estimate of the loading and $C_{NT}^{-1} = \min(N^{1/2}, T^{1/2})$. Consequently, $\frac{1}{\sqrt{T}} \sum_{t=1}^{\lfloor Tr \rfloor} \hat{e}_{i,t}^Y \implies B_i^Y(r)$, $\forall i$ where $B_i^Y(r)$ is the first element of the $(1+m)$ -vector Brownian motion $B_i(r)$. $B_i^Y(r)$ and $B_j^Y(r)$ are uncorrelated Brownian motions for $i \neq j$. The same holds for $X_{i,t}$. Consequently, standard panel no-cointegration tests derived under the maintained assumption of independent panel unit, such as those proposed by Pedroni (2004a), can be used on the defactored observations.

Remark 6. This approach requires both large N and T which is one of the important limitations. Also, this approach will have finite sample properties that can, *at best*, be close to those observed for the tests when applied to a panel data set with independent units.

Remark 7. If the rank of the long-run covariance matrix of the factors turns out to be smaller than k , that is if the factors cointegrate, then a further step is needed to assess overall lack of cointegration between $Y_{i,t}$ and $X_{i,t}$. No cointegration then requires separability in cointegration as discussed and analyzed in details in Hecq et al. (2002).

3.5 Some Monte Carlo Evidence

The theoretical foundation of the approach proposed in the preceding section requires both large N and T which is not always met in typical applications of panel cointegration techniques. A Monte Carlos analysis of some of its finite sample properties is called for. We focus on the empirical size properties of the proposed approach, namely testing for no-cointegration using defactored data, as it was shown that tests designed for cross-sectionally independent data may suffer from dramatic size distortions when applied to panels with cross-member cointegration for example as pointed out by Banerjee et al. (2004). The DGP is a simple bivariate process (i.e. $m = 1$) with $k = 2$ common factors that obeys the representation (3.1)-(3.4).

$$\begin{aligned} Z_{i,t} &= \Lambda_i F_t + E_{i,t}, & E_{i,t} &= e_{i,t} & \text{or} & E_{i,t} &= E_{i,t-1} + e_{i,t}, \\ e_{i,t} &= \varepsilon_{i,t} + \Gamma_i \varepsilon_{i,t-1}, \\ F_t &= F_{t-1} + f_t, & f_t &= \eta_t + \Phi_1 \eta_{t-1}, \end{aligned}$$

where $\varepsilon_{i,t} \sim \text{i.i.d. } N(0, \Sigma_i)$, $\eta_t \sim \text{i.i.d. } N(0, I_2)$. The loading matrix has a diagonal structure with diagonal elements $\lambda_{1i}, \lambda_{2i} \sim U[-1, 3]$ where U denotes uniform distributions. The remaining parameters are also drawn from independent uniform distributions to allow for some degree of heterogeneity:

$\Phi_{11,22} \sim U[0.5, 0.7]$, $\Phi_{12,21} \sim U[0, 0.5]$, $\Sigma_{i,11,22} \sim U[1, 1.4]$, $\Gamma_{11,22} \sim U[0.5, 0.7]$, $\Phi_{12,21} \sim U[0, 0.5]$ and $\Sigma_{i,12,21} \sim U[0, 0.2]$. The sample size has been set to $T \in \{50, 100, 250\}$ and the number of units in the panel is set to $N \in \{25, 50, 100\}$. We consider the rejection frequencies based on 1000 replications⁹ for Kao's pooled normalized coefficient (the ρ test) and pooled ADF test, and Pedroni's panel- t , panel- ρ , group mean t and group mean ρ statistics based on raw data. Furthermore, we consider Pedroni's panel ρ and Pedroni's group mean ρ statistics applied to the defactored data, and Johansen trace test for the estimated common factors, using the information criterion of Aznar and Salvador (2002) to select the lag length of the VECM.

For the last two statistics based on the defactored data, we estimate the number of common factors k using the IC_1 criterion of Bai and Ng (2002) with $k_{max} = 4$. For the ADF type tests the lag length is selected using the AIC. For the non-parametric correction for serial correlation, we use a quadratic spectral kernel with a bandwidth of $3.21T^{\frac{1}{3}}$ (see Andrews, 1991).

The two polar cases that we consider in the simulations are the cases discussed earlier, namely the case of cross-member cointegration in which the common factors are $I(1)$ and the idiosyncratic components are $I(0)$, and the case where both common factors and idiosyncratic components are $I(1)$. In addition, we consider cases where only the common factors are cointegrated but the idiosyncratic components are not cointegrated, not cointegrated common factors are combined with cointegrated idiosyncratic components, and cointegration in both the common factors and the idiosyncratic components.

Tables 3.1 and 3.2 present simulation results for the 5 cases with MA(1) dynamics in the error terms and $k = 2$ common factors, one common factor in $Y_{i,t}$ and one in $X_{i,t}$. Furthermore, the number of common factors is estimated using the IC_1 criterion of Bai and Ng (2002) with $k_{max} = 4$. Note that the criterion always picks the correct number of common factors. Both Kao test statistics show strong size distortions when either the common factors or the idiosyncratic components (or both) cointegrate. The Pedroni tests exhibit very strong size distortions in the cross-member cointegration case (Table 3.1). When non-stationary idiosyncratic components are combined with non-cointegrated or cointegrated common factors (Tables 3.1 and 3.2) size distortions are reduced. The tests are even undersized for some combinations of N and T . When both the common factors and the idiosyncratic components cointegrate (Table 3.2), the Pedroni tests have rejection frequencies of up to 1. However, as the factor loadings are heterogenous, panel cointegration is not present (see the discussion in Section 3.4).

⁹All experiments are carried out using GAUSS 6.0.

The tests applied to the estimated idiosyncratic components show rejection frequencies of (close to) 1 when those are stationary or cointegrated. When the idiosyncratic components are not cointegrated, the Idiosyncratic Panel- ρ and Idiosyncratic Group- ρ tests are undersized. The Aznar/Johansen test applied to the estimated common factors is slightly oversized when the common factors do not cointegrate with rejection frequencies between 8% and 15%. When there is cointegration among the common factors, the test has a power between 61% and 92%.

We also perform simulations where we have introduced a second factor in $X_{i,t}$, such that $k = 3$ now.¹⁰ Again estimating the number of common factors using the IC_1 criterion of Bai and Ng (2002), we note that the second common factor of $X_{i,t}$ is not picked up¹¹. Nevertheless, simulation results for the Kao and Pedroni tests applied to the raw data and the Aznar/Johansen test applied to the extracted common factors do not change qualitatively compared to the results obtained for $k = 2$. However, the Idiosyncratic Panel- ρ and Idiosyncratic Group- ρ applied to the estimated common components exhibit a reduced power when the common components are cointegrated, in particular when $T=50$.

3.6 Conclusions

We have considered the problem of testing for (no-)cointegration in panel data characterized by strong cross-sectional dependencies resulting from common factors as in the work of Bai and Ng (2004b). We focus on two polar cases that we believe are of empirical relevance.

For both classes of DGP's, we discuss the homogeneity restrictions for the cointegrating vectors resulting from the presence of common factor cointegration. We study analytically the behavior of several test for panel cointegration including the residual-based panel no-cointegration tests of Kao (1999) and Pedroni (1999, 2004a) that have been widely used in empirical work in the recent years. The results complement and help to understand some of the Monte Carlo results reported by Banerjee et al. (2004), such as the loss of Gaussian limiting results and occurrence of size distortions resulting from the presence of cross-sectional dependence.

These observations provide sufficient reason to propose a two-step procedure for testing for no-cointegration in panels with common factors. Our procedure is similar in spirit and complementary to the work of Banerjee and Carrion-i Silvestre (2005). It has the advantages of covering many sub-cases of interest and allowing to get a clear picture of the common and idiosyncratic components in

¹⁰Tables with the results for these simulations are included in the working paper version of this chapter Gengenbach et al. (2006).

¹¹Similarly, the PC_1 or BIC_3 criteria from Bai and Ng (2002) only select a single common factor for $X_{i,t}$

the panel and about the homogeneity requirements for common factor cointegration. The procedure is simple to apply and makes use of existing tools. Simulation results show the procedure to have reasonable size properties.

While being attractive due, among other things, to its ease of application and nice properties, some limitations are inherent in this approach. The theoretical validity of the proposed procedure, and that of Banerjee and Carrion-i Silvestre (2005), relies on both large N and large T which may be unrealistic for applications with "moderate" N and large T . The performance of the proposed procedures, in particular the power properties, in such situations needs to be further studied even if the size properties reported in Monte Carlo section are promising. If a large N analysis is inappropriate for the problem under study, an alternative could be to adopt the non-linear IV testing approach of Demetrescu and Tarcolea (2005) or use bootstrapping techniques that seem to work well from an empirical point of view (see Fachin, 2005). Future work should study the merits of these alternative approaches both theoretically and empirically.

A second limitation lies in the fact that the approach is residual-based and hence it suffers from the usual critiques against residual-based tests such as the maintained assumptions of a single cointegrating relationship (if it exists) as well as the imposition of the common factor restriction. Nothing however precludes conceptually to extend the idea developed in this chapter to other cointegration techniques that would not suffer from these drawbacks.

3.A Tables

Table 3.1: $k = 2$ common factors; Non-stationary common factors F_t with I(0) or I(1) idiosyncratic component $E_{i,t}$.

$E_{i,t}$	N	25	50	100	25	50	100	25	50	100
Raw data										
	T	Kao - ρ^*			Kao - ADF			Pedroni - <i>Panel</i> - ρ		
I(0)	50	0.27	0.32	0.35	0.48	0.50	0.54	0.68	0.90	0.88
	100	0.39	0.47	0.49	0.54	0.62	0.62	0.84	0.96	0.95
	250	0.52	0.54	0.55	0.64	0.67	0.69	0.93	1.00	0.96
I(1)	50	0.17	0.17	0.23	0.59	0.65	0.69	0.00	0.00	0.00
	100	0.23	0.28	0.36	0.63	0.74	0.75	0.02	0.02	0.03
	250	0.34	0.39	0.45	0.74	0.81	0.80	0.10	0.08	0.14
Estimated components										
	T	Pedroni - <i>Panel</i> - t			Pedroni - <i>Group</i> - ρ			Pedroni - <i>Group</i> - t		
I(0)	50	0.76	0.92	0.91	0.33	0.67	0.62	0.52	0.79	0.77
	100	0.83	0.96	0.94	0.67	0.94	0.89	0.67	0.89	0.85
	250	0.92	0.99	0.95	0.88	1.00	0.94	0.78	0.95	0.88
I(1)	50	0.03	0.02	0.04	0.00	0.00	0.00	0.03	0.02	0.04
	100	0.06	0.04	0.08	0.00	0.00	0.01	0.04	0.03	0.06
	250	0.13	0.10	0.18	0.02	0.01	0.04	0.07	0.05	0.10
	T	Idiosyncratic- <i>Panel</i> - t			Idiosyncratic - <i>Group</i> - ρ			Aznar/Johansen		
I(0)	50	1.00	1.00	1.00	1.00	1.00	1.00	0.12	0.12	0.11
	100	1.00	1.00	1.00	1.00	1.00	1.00	0.09	0.11	0.09
	250	1.00	1.00	1.00	1.00	1.00	1.00	0.08	0.10	0.08
I(1)	50	0.00	0.00	0.00	0.00	0.00	0.00	0.12	0.14	0.12
	100	0.00	0.00	0.00	0.00	0.00	0.00	0.10	0.12	0.09
	250	0.02	0.01	0.00	0.00	0.00	0.00	0.11	0.10	0.09

Rejection frequencies are based on 5% asymptotic critical values.

Table 3.2: $k = 2$ common factors; Cointegration in either F_t or $E_{i,t}$ or both. NC denotes no cointegration, C cointegration.

F_t	$E_{i,t}$	N	25	50	100	25	50	100	25	50	100
Raw data											
		T	Kao- ρ^*			Kao- ADF			Pedroni $Panel - \rho$		
C	NC	50	0.18	0.17	0.24	0.55	0.60	0.65	0.01	0.00	0.01
		100	0.25	0.29	0.39	0.59	0.69	0.73	0.07	0.03	0.11
		250	0.37	0.40	0.50	0.70	0.77	0.78	0.18	0.10	0.32
NC	C	50	0.26	0.28	0.33	0.57	0.60	0.64	0.05	0.13	0.10
		100	0.36	0.43	0.46	0.62	0.72	0.72	0.16	0.39	0.32
		250	0.46	0.52	0.55	0.71	0.77	0.78	0.31	0.53	0.48
C	C	50	0.41	0.43	0.47	0.57	0.58	0.61	0.58	0.63	0.69
		100	0.50	0.56	0.59	0.60	0.67	0.69	0.84	0.91	0.97
		250	0.60	0.63	0.63	0.69	0.72	0.74	0.96	0.99	1.00
Estimated components											
		T	Idiosyncratic $Panel - \rho$			Idiosyncratic $Group - \rho$			Aznar/Johansen		
C	NC	50	0.00	0.00	0.00	0.00	0.00	0.00	0.61	0.61	0.74
		100	0.00	0.00	0.00	0.00	0.00	0.00	0.70	0.78	0.85
		250	0.02	0.00	0.00	0.00	0.00	0.00	0.68	0.85	0.87
NC	C	50	0.92	0.99	1.00	0.94	1.00	1.00	0.12	0.12	0.11
		100	1.00	1.00	1.00	1.00	1.00	1.00	0.10	0.11	0.10
		250	1.00	1.00	1.00	1.00	1.00	1.00	0.09	0.10	0.08
C	C	50	0.98	1.00	1.00	0.99	1.00	1.00	0.67	0.65	0.75
		100	1.00	1.00	1.00	1.00	1.00	1.00	0.79	0.83	0.89
		250	1.00	1.00	1.00	1.00	1.00	1.00	0.78	0.87	0.92

If F_t is cointegrated, $F_t^Y = \sum_{s=1}^t f_s^Y$, $F_t^X = F_t^Y + f_t^X$. If $E_{i,t}$ is cointegrated, $E_{i,t}^Y = \sum_{s=1}^t e_{i,s}^Y$, $E_{i,t}^X = E_{i,t}^Y + e_{i,t}^X$, where f_t and $e_{i,t}$ are MA processes generated as described in Section 3.5. Rejection frequencies are based on 5% asymptotic critical values.

3.B Appendix

3.B.1 Lemma 3.B.1 to 3.B.3

Given Assumptions 3.1 to 3.4, we can summarize some convergence results. In the following lemmas, M is used to denote a generic positive number, not depending on T or N . For a matrix \mathbf{A} , $\mathbf{A} > 0$ denotes that \mathbf{A} is positive definite. Furthermore, $\|\mathbf{A}\| = \text{trace}(\mathbf{A}'\mathbf{A})^{\frac{1}{2}}$. We write the integral $\int_0^1 B(r)\mathbf{d}r$ as $\int B$, and $\int_0^1 B(r)B(r)'\mathbf{d}r$ as $\int BB'$. Furthermore, \Rightarrow denotes weak convergence, and \xrightarrow{p} denotes convergence in probability. For any number x , $[x]$ denotes the largest integer smaller than x . For any variable $X_{i,t}$, $\tilde{X}_{i,t} = X_{i,t} - \frac{1}{T} \sum_{s=1}^T X_{i,s}$. Similarly, for any Brownian motion B , $\tilde{B} = B - \int B$. Throughout the chapter we employ sequential limit theory, where we consider $T \rightarrow \infty$ followed by $N \rightarrow \infty$. Furthermore, for non-random factor loadings, $\bar{\Lambda} = \lim_{N \rightarrow \infty} \frac{1}{N} \sum_{i=1}^N \Lambda_i$, while for random factor loadings $\bar{\Lambda} = \mathbf{E}(\Lambda_i)$, $\bar{\Psi} = \mathbf{E}(\Psi_i)$ and $\bar{\Delta} = \mathbf{E}(\Delta_i)$.

Lemma 3.B.1 presents convergence results for the common data component $\Lambda_i F_t$. The limiting distributions are functionals of Brownian motions weighted by the factor loadings, even as $N \rightarrow \infty$. These results are intuitive, as we assume a fixed number of common factors. Lemma 3.B.2 summarizes the convergence for the idiosyncratic components, where we recover the panel spurious regression results for Phillips and Moon (1999). In Lemma 3.B.3, the limits for the cross-products of the common and individual specific components are given. It is evident that these cross-products will only affect limiting distributions for finite N , but as $N \rightarrow \infty$ these effects will vanish due to the independence of the shock driving F_t and $E_{i,t}$.

Lemma 3.B.1: Common Component

Lemma 3.B.1 *Common Component: Given Assumptions 3.1, 3.2 and 3.4,*

$$\begin{aligned}
 (a) \quad & \frac{1}{T} \sum_{t=1}^T \Lambda_i f_t F_{t-1}' \Lambda_i' \Rightarrow \Lambda_i (\int \mathbf{d}B_F B_F' + \Theta) \Lambda_i' \quad \text{as } T \rightarrow \infty, \text{ and} \\
 & \frac{1}{N} \sum_{i=1}^N \Lambda_i (\int \mathbf{d}B_F B_F' + \Theta) \Lambda_i' \xrightarrow{p} \int \mathbf{d}B_{F\Lambda} B_{F\Lambda}' + \Theta_{F\Lambda} \quad \text{as } N \rightarrow \infty, \\
 (b) \quad & \frac{1}{T^2} \sum_{t=1}^T \Lambda_i F_t F_t' \Lambda_i' \Rightarrow \Lambda_i (\int B_F B_F') \Lambda_i' \quad \text{as } T \rightarrow \infty, \text{ and} \\
 & \frac{1}{N} \sum_{i=1}^N \Lambda_i (\int B_F B_F') \Lambda_i' \xrightarrow{p} \int B_{F\Lambda} B_{F\Lambda}' \quad \text{as } N \rightarrow \infty, \\
 (c) \quad & \frac{1}{T} \sum_{t=1}^T \Lambda_i f_t \tilde{F}_{t-1}' \Lambda_i' \Rightarrow \Lambda_i (\int \mathbf{d}B_F \tilde{B}_F' + \Theta) \Lambda_i' \quad \text{as } T \rightarrow \infty, \text{ and} \\
 & \frac{1}{N} \sum_{i=1}^N \Lambda_i (\int \mathbf{d}B_F \tilde{B}_F' + \Theta) \Lambda_i' \xrightarrow{p} \int \mathbf{d}B_{F\Lambda} \tilde{B}_{F\Lambda}' + \Theta_{F\Lambda} \quad \text{as } N \rightarrow \infty, \\
 (d) \quad & \frac{1}{T^2} \sum_{t=1}^T \Lambda_i \tilde{F}_t \tilde{F}_t' \Lambda_i' \Rightarrow \Lambda_i (\int \tilde{B}_F \tilde{B}_F') \Lambda_i' \quad \text{as } T \rightarrow \infty, \text{ and} \\
 & \frac{1}{N} \sum_{i=1}^N \Lambda_i (\int \tilde{B}_F \tilde{B}_F') \Lambda_i' \xrightarrow{p} \int \tilde{B}_{F\Lambda} \tilde{B}_{F\Lambda}' \quad \text{as } N \rightarrow \infty,
 \end{aligned}$$

where $\text{vec}(\int \mathbf{d}B_{F\Lambda} B_{F\Lambda}') = \tilde{\Lambda} \text{vec}(\int \mathbf{d}B_F B_F')$, $\text{vec}(\Theta_{F\Lambda}) = \tilde{\Lambda} \text{vec}(\Theta)$, $\text{vec}(\int B_{F\Lambda} B_{F\Lambda}') = \tilde{\Lambda} \text{vec}(\int B_F B_F')$, $\text{vec}(\int \mathbf{d}B_{F\Lambda} \tilde{B}_{F\Lambda}') = \tilde{\Lambda} \text{vec}(\int \mathbf{d}B_F \tilde{B}_F')$ and $\text{vec}(\int \tilde{B}_{F\Lambda} \tilde{B}_{F\Lambda}') = \tilde{\Lambda} \text{vec}(\int \tilde{B}_F \tilde{B}_F')$, and $\tilde{\Lambda} = \text{plim}_{N \rightarrow \infty} \frac{1}{N} \sum_{i=1}^N (\Lambda_i \otimes \Lambda_i)$.

Proof of Lemma 3.B.1

For the common factors given in (3.2) we find the following Beveridge-Nelson (BN) decomposition:

$$F_t = \Phi(1) \sum_{s=1}^t \eta_s + \Phi^*(L)(\eta_t - \eta_0) + F_0, \quad (3.B.1)$$

where $\Phi^*(L) = \sum_{j=0}^{\infty} \Phi_j^* L^j$ with $\Phi_j^* = -\sum_{l=j+1}^{\infty} \Phi_l$. Now, $\frac{1}{\sqrt{T}} \Phi(1) \sum_{s=1}^{\lfloor rT \rfloor} \eta_s \implies \Phi(1)W_F(r) \equiv B_F(r)$ by the FCLT, where W_F is standard Brownian Motion. Furthermore, $\Phi^*(L)(\eta_t - \eta_0)$ is stationary with finite fourth order moments such that $\frac{1}{\sqrt{T}} \Phi^*(L)(\eta_t - \eta_0) \xrightarrow{p} 0$, and F_0 is $O_p(1)$ by assumption.

- (a) We have $\frac{1}{T} \sum_{t=1}^T \Lambda_i f_t F'_{t-1} \Lambda'_i = \Lambda_i (\frac{1}{T} \sum_{t=1}^T f_t F'_{t-1}) \Lambda'_i$. Now, $\frac{1}{T} \sum_{t=1}^T f_t F'_{t-1} \implies \int \mathbf{d}B_F B'_F + \Theta$ as $T \rightarrow \infty$ as shown in e.g. Davidson and de Jong (2000), and the result of Lemma 3.B.1 (a) follows immediately. Furthermore, $\text{vec}(\Lambda_i (\int \mathbf{d}B_F B'_F + \Theta) \Lambda'_i) = (\Lambda_i \otimes \Lambda_i) \text{vec}(\int \mathbf{d}B_F B'_F + \Theta)$. As $\mathbb{E}\|(\Lambda_i \otimes \Lambda_i)\|^2 = \mathbb{E}\|\Lambda_i\|^4 \leq M$ by Assumption 3.2 (i), we can apply a LLN to $\frac{1}{N} \sum_{i=1}^N (\Lambda_i \otimes \Lambda_i)$. Denote $\text{plim}(\Lambda_i \otimes \Lambda_i) = \tilde{\Lambda}$ to obtain the second result of Lemma 3.B.1 (a).
- (b) The proof of Lemma 3.B.1 (b) is similar to that of (a), except that $\frac{1}{T^2} \sum F_t F'_t \implies \int B_F B'_F$ as shown in e.g. Phillips and Durlauf (1986) in the first step.
- (c) $\frac{1}{T} \sum_{t=1}^T f_t \tilde{F}'_{t-1} = \frac{1}{T} \sum_{t=1}^T f_t F'_{t-1} - (\sum_{t=1}^T \frac{f_t}{\sqrt{T}}) (\frac{1}{T^{\frac{3}{2}}} \sum_{s=1}^T F'_s)$. Now, $\frac{1}{T} \sum_{t=1}^T f_t F'_{t-1} \implies \int \mathbf{d}B_F B'_F + \Theta$ while $(\sum_{t=1}^T \frac{f_t}{\sqrt{T}}) (\frac{1}{T^{\frac{3}{2}}} \sum_{s=1}^T F'_s) \implies \int \mathbf{d}B_F (\int B_F)'$ as $T \rightarrow \infty$, so that $\frac{1}{T} \sum_{t=1}^T f_t \tilde{F}'_{t-1} \implies \int \mathbf{d}B_F \tilde{B}'_F + \Theta$. The remainder of the proof follows the same arguments as above.
- (d) Now, $\sum_{t=1}^T \tilde{F}_t \tilde{F}'_t \implies \int \tilde{B}_F \tilde{B}'_F$ as shown in Phillips and Moon (1999), and the limit as $N \rightarrow \infty$ follows as above. □

Lemma 3.B.2: Idiosyncratic Components

Lemma 3.B.2 *Idiosyncratic Components: Given Assumption 3.3,*

- (a) $\frac{1}{T} \sum_{t=1}^T e_{i,t} S'_{i,t-1} \implies (\int \mathbf{d}B_i B'_i + \Delta_i)$ as $T \rightarrow \infty$, and
 $\frac{1}{N} \sum_{i=1}^N (\int \mathbf{d}B_i B'_i + \Delta_i) \xrightarrow{p} \Delta$ as $N \rightarrow \infty$,
- (b) $\frac{1}{T^2} \sum_{t=1}^T S_{i,t} S'_{i,t} \implies \int B_i B'_i$ as $T \rightarrow \infty$, and
 $\frac{1}{N} \sum_{i=1}^N \int B_i B'_i \xrightarrow{p} \frac{1}{2} \Psi$ as $N \rightarrow \infty$,
- (c) $\frac{1}{T} \sum_{t=1}^T e_{i,t} \tilde{S}'_{i,t-1} \implies (\int \mathbf{d}B_i \tilde{B}'_i + \Delta_i)$ as $T \rightarrow \infty$, and
 $\frac{1}{N} \sum_{i=1}^N (\int \mathbf{d}B_i \tilde{B}'_i + \Delta_i) \xrightarrow{p} -\frac{1}{2} \Psi + \Delta$ as $N \rightarrow \infty$,
- (d) $\frac{1}{T^2} \sum_{t=1}^T \tilde{S}_{i,t} \tilde{S}'_{i,t} \implies \int \tilde{B}_i \tilde{B}'_i$ as $T \rightarrow \infty$, and
 $\frac{1}{N} \sum_{i=1}^N \int \tilde{B}_i \tilde{B}'_i \xrightarrow{p} \frac{1}{6} \Psi$ as $N \rightarrow \infty$.

Proof of Lemma 3.B.2

For the partial sum process $S_{i,t} = \sum_{s=1}^t e_{i,s}$ we obtain a BN decomposition

$$S_{i,t} = \Gamma_i(1) \sum_{s=1}^t \varepsilon_{i,s} + \Gamma_i^*(L)(\varepsilon_{i,t} - \varepsilon_{i,0}) + E_{i,0}, \quad (3.B.2)$$

where $\Gamma_i^*(L) = \sum_{j=0}^{\infty} \Gamma_{i,j}^* L^j$ with $\Gamma_{i,j}^* = -\sum_{l=j+1}^{\infty} \Gamma_{i,l}$. Now, $\frac{1}{\sqrt{T}} S_{i, \lfloor rT \rfloor} \Rightarrow \Gamma_i(1) \Sigma_i^{\frac{1}{2}} W_i(r) \equiv B_i(r)$ as $T \rightarrow \infty$ for all i , where W_i is standard Brownian motion and $\Sigma_i^{\frac{1}{2}}$ is the Cholesky decomposition of Σ_i such that $\Sigma_i^{\frac{1}{2}} \Sigma_i^{\frac{1}{2}'} = \Sigma_i$ as shown in Phillips and Moon (1999). Furthermore, B_i and B_j are *i.i.d* over the i -dimension.

- (a) We have $\frac{1}{T} \sum_{t=1}^T e_{i,t} S'_{i,t-1} \Rightarrow \int \mathbf{d}B_i B'_i + \Delta_i$ as $T \rightarrow \infty$ as shown in Davidson and de Jong (2000). Now, $\int \mathbf{d}B_i B'_i$ are *i.i.d* across the i -dimension with $\mathbb{E}(\int \mathbf{d}B_i B'_i) = 0$ and $E\|\text{vec}(\int \mathbf{d}B_i B'_i)\|^2 < M$. So, we can apply a LLN to find $\frac{1}{N} \sum_{i=1}^N \int \mathbf{d}B_i B'_i \xrightarrow{p} 0$. Furthermore, a LLN also applies such that $\frac{1}{N} \sum_{i=1}^N \Delta_i \xrightarrow{p} \Delta \equiv \mathbb{E}(\Delta_i)$, which proves the first result.
- (b) This result is proven in Phillips and Moon (1999).
- (c) $\frac{1}{T} \sum_{t=1}^T e_{i,t} \tilde{S}'_{i,t-1} = \frac{1}{T} \sum_{t=1}^T e_{i,t} S'_{i,t-1} - \frac{1}{T} \sum_{t=1}^T e_{i,t} \bar{S}'_i$, where $\bar{S}'_i = \frac{1}{T} \sum_{t=1}^T S_{i,t}$. Now, $\frac{1}{T} \sum_{t=1}^T e_{i,t} S'_{i,t-1} \Rightarrow \int \mathbf{d}B_i B'_i + \Delta_i$, while $\frac{1}{T} \sum_{t=1}^T e_{i,t} \bar{S}'_i = \frac{1}{\sqrt{T}} S_T \frac{1}{T^{\frac{3}{2}}} \sum_{t=1}^T S'_{i,t} \Rightarrow B_i(1) \int B_i$ as $T \rightarrow \infty$. So, $\frac{1}{T} \sum_{t=1}^T e_{i,t} \tilde{S}'_{i,t-1} \Rightarrow \int \mathbf{d}B_i \tilde{B}'_i + \Delta_i$ as $T \rightarrow \infty$. Furthermore, $\mathbb{E}(\int \mathbf{d}B_i \tilde{B}'_i) = -\frac{1}{2} \Psi$ and hence, using similar arguments as in (a) $\frac{1}{N} \sum_{i=1}^N \int \mathbf{d}B_i \tilde{B}'_i + \Delta_i \xrightarrow{p} -\frac{1}{2} \Psi + \Delta$ as $N \rightarrow \infty$.
- (d) See Phillips and Moon (1999).

□

Lemma 3.B.3

Lemma 3.B.3 *Given Assumptions 3.1, 3.2, 3.3 and 3.4*

- (a) $\frac{1}{T} \sum_{t=1}^T \Lambda_i F_{t-1} e'_{i,t} \Rightarrow \Lambda_i \int B_F \mathbf{d}B'_i$ as $T \rightarrow \infty$, and
 $\frac{1}{N} \sum_{i=1}^N \Lambda_i \int B_F \mathbf{d}B'_i \xrightarrow{p} 0$ as $N \rightarrow \infty$,
- (b) $\frac{1}{T} \sum_{t=1}^T \Lambda_i f_t S'_{i,t-1} \Rightarrow \Lambda_i \int \mathbf{d}B_F B'_i$ as $T \rightarrow \infty$, and
 $\frac{1}{N} \sum_{i=1}^N \Lambda_i \int \mathbf{d}B_F B'_i \xrightarrow{p} 0$ as $N \rightarrow \infty$,
- (c) $\frac{1}{T^2} \sum_{t=1}^T \Lambda_i F_t S'_{i,t} \Rightarrow \Lambda_i \int B_F B'_i$ as $T \rightarrow \infty$, and
 $\frac{1}{N} \sum_{i=1}^N \Lambda_i \int B_F B'_i \xrightarrow{p} 0$ as $N \rightarrow \infty$,
- (d) $\frac{1}{T} \sum_{t=1}^T \Lambda_i \tilde{F}_{t-1} e'_{i,t} \Rightarrow \Lambda_i \int \tilde{B}_F \mathbf{d}B'_i$ as $T \rightarrow \infty$, and
 $\frac{1}{N} \sum_{i=1}^N \Lambda_i \int \tilde{B}_F \mathbf{d}B'_i \xrightarrow{p} 0$ as $N \rightarrow \infty$,
- (e) $\frac{1}{T} \sum_{t=1}^T \Lambda_i \tilde{F}_{t-1} \tilde{e}'_{i,t} \Rightarrow \Lambda_i \int \tilde{B}_F \mathbf{d}B'_i$ as $T \rightarrow \infty$, and
 $\frac{1}{N} \sum_{i=1}^N \Lambda_i \int \tilde{B}_F \mathbf{d}B'_i \xrightarrow{p} 0$ as $N \rightarrow \infty$,

$$\begin{aligned}
(f) \quad & \frac{1}{T} \sum_{t=1}^T \Lambda_i f_t \tilde{S}'_{i,t-1} \implies \Lambda_i \int \mathbf{d}B_F \tilde{B}'_i \quad \text{as } T \rightarrow \infty, \text{ and} \\
& \frac{1}{N} \sum_{i=1}^N \Lambda_i \int \mathbf{d}B_F \tilde{B}'_i \xrightarrow{p} 0 \quad \text{as } N \rightarrow \infty, \\
(g) \quad & \frac{1}{T^2} \sum_{t=1}^T \Lambda_i \tilde{F}_t \tilde{S}'_{i,t} \implies \Lambda_i \int \tilde{B}_F \tilde{B}'_i \quad \text{as } T \rightarrow \infty, \text{ and} \\
& \frac{1}{N} \sum_{i=1}^N \Lambda_i \int \tilde{B}_F \tilde{B}'_i \xrightarrow{p} 0 \quad \text{as } N \rightarrow \infty,
\end{aligned}$$

Proof of Lemma 3.B.3

For each i , the stacked error vector $w_{i,t} = (f'_t, e'_{i,t})'$ and the corresponding partial sum process $W_{i,t} = \sum_{s=1}^t w_{i,s} = (F'_t, S'_{i,t})'$ fulfill the conditions for a FCLT, such that $\frac{1}{\sqrt{T}} W_{i, \lfloor rT \rfloor} \implies B_{w_i}(r) = (B_F(r)', B_i(r))'$. Due to the independence of f_t and $e_{i,t}$, the covariance matrix of B_{w_i} will have zero off-diagonal blocks. Now, for every panel unit i we obtain time series spurious regression results as $T \rightarrow \infty$. Furthermore, the functionals of B_F and B_i we obtain in the first step have zero mean and finite variance, and are uncorrelated across the i -dimension of the panel. So, we can apply a LLN to the average to find the limits as $N \rightarrow \infty$. We present the proof for (a), (b)-(g) are obtained using a similar line of argumentation.

- (a) The limit as $T \rightarrow \infty$ follows from applying a spurious regression result as above and noting that $\mathbb{E}(F_{t-1} e_{i,t}) = 0$ for all i and t . Now, taking expectations we find $\mathbb{E}(\Lambda_i \int B_F \mathbf{d}B_i) = 0$, while $\mathbb{E} \|\text{vec}(\Lambda_i \int B_F \mathbf{d}B_i)\|^2 < M$. for all i . Applying a LLN, we find
- $$\frac{1}{N} \sum_{i=1}^N \Lambda_i \int B_F \mathbf{d}B_i \xrightarrow{p} 0.$$

□

3.B.2 Proof of Propositions 3.1 and 3.2

Proof of Proposition 3.1 (a): Convergence of $\tilde{\beta}$

The LSDV estimator of β is given by $\tilde{\beta} = (\sum_{i=1}^N \sum_{t=1}^T \tilde{Y}_{i,t} \tilde{X}'_{i,t}) (\sum_{i=1}^N \sum_{t=1}^T \tilde{X}_{i,t} \tilde{X}'_{i,t})^{-1}$. Consider the numerator

$$\sum_{i=1}^N \sum_{t=1}^T \tilde{Y}_{i,t} \tilde{X}'_{i,t} = \sum_{i=1}^N \sum_{t=1}^T (\lambda'_{1i} \tilde{F}_t^Y \tilde{F}_t^{X'} \lambda_{21} + \tilde{E}_{i,t}^Y \tilde{E}_{i,t}^{X'} + \lambda'_{1i} \tilde{F}_t^Y \tilde{E}_{i,t}^{X'} + \tilde{E}_{i,t}^Y \tilde{F}_t^{X'} \lambda_{21}). \quad (3.B.3)$$

If the idiosyncratic term is given by (3.3), we have $\sum_{i=1}^N (O_p(T^2) + O_p(T) + O_p(T) + O_p(T))$ in (3.B.3). So, as $T \rightarrow \infty$, $\sum_{i=1}^N \frac{1}{T^2} \sum_{t=1}^T \tilde{Y}_{i,t} \tilde{X}'_{i,t} \implies \sum_{i=1}^N \lambda'_{1i} \int \tilde{B}_F^Y \tilde{B}_F^{X'} \lambda_{21}$ from the first result of Lemma 3.B.1 (d). Now, using the second result we obtain $\frac{1}{N} \sum_{i=1}^N \lambda'_{1i} \int \tilde{B}_F^Y \tilde{B}_F^{X'} \lambda_{21} \xrightarrow{p} \int \tilde{B}_{F\Lambda}^Y \tilde{B}_{F\Lambda}^{X'}$ as $N \rightarrow \infty$, where $\int \tilde{B}_{F\Lambda}^Y \tilde{B}_{F\Lambda}^{X'}$ is the $1 \times m$ upper right block of $\int \tilde{B}_{F\Lambda} \tilde{B}'_{F\Lambda}$ defined in Lemma 3.B.1.

If the idiosyncratic terms are also I(1), such that the DGP includes (3.4), all terms in (3.B.3) are $O_p(T^2)$ when summed over T . However, the cross-products of the common factors and idiosyncratic components will vanish in the limit as $N \rightarrow \infty$. Using Lemmas 3.B.1 (d), 3.B.2 (d) and 3.B.3 (g) we find as $T \rightarrow \infty$ followed

by $N \rightarrow \infty$,

$$\frac{1}{N} \sum_{i=1}^N \frac{1}{T^2} \sum_{t=1}^T \tilde{Y}_{i,t} \tilde{X}'_{i,t} \implies \int \tilde{B}_{F\Lambda}^Y \tilde{B}_{F\Lambda}^{X'} + \frac{1}{6} \Psi^{YX},$$

where Ψ^{YX} is the upper right $1 \times m$ block of Ψ .

Now the denominator of $\tilde{\beta}$ is given by

$$\sum_{i=1}^N \sum_{t=1}^T \tilde{X}_{i,t} \tilde{X}'_{i,t} = \sum_{i=1}^N \sum_{t=1}^T (\lambda'_{2i} \tilde{F}_t^X \tilde{F}_t^{X'} \lambda_{2i} + \tilde{E}_{i,t}^X \tilde{E}_{i,t}^{X'} + \lambda'_{2i} \tilde{F}_t^X \tilde{E}_{i,t}^{X'} + \tilde{E}_{i,t}^X \tilde{F}_t^{X'} \lambda_{2i}). \quad (3.B.4)$$

Similar to the results for the numerator, the terms in (3.B.4) are $\sum_{i=1}^N (O_p(T^2) + O_p(T) + O_p(T) + O_p(T))$, if the DGP contains (3.3). Hence, $\sum_{i=1}^N \frac{1}{T^2} \sum_{t=1}^T \tilde{X}_{i,t} \tilde{X}'_{i,t} \implies \sum_{i=1}^N \lambda'_{2i} \int \tilde{B}_F^X \tilde{B}_F^{X'} \lambda_{2i}$ as $T \rightarrow \infty$. Furthermore, the remaining term is $O_p(N)$, and we obtain $\frac{1}{N} \sum_{i=1}^N \frac{1}{T^2} \sum_{t=1}^T \tilde{X}_{i,t} \tilde{X}'_{i,t} \implies \int \tilde{B}_{F\Lambda}^X \tilde{B}_{F\Lambda}^{X'}$ as $T \rightarrow \infty$ followed by $N \rightarrow \infty$, where $\int \tilde{B}_{F\Lambda}^X \tilde{B}_{F\Lambda}^{X'}$ is the lower right $m \times m$ block of $\int \tilde{B}_{F\Lambda} \tilde{B}'_{F\Lambda}$.

If the true DGP contains (3.4), all terms in the summation over T in (3.B.4) are $O_p(T^2)$ and as above, the cross-products between common and idiosyncratic components will vanish in the cross-sectional average as $N \rightarrow \infty$. We find as $T \rightarrow \infty$ followed by $N \rightarrow \infty$,

$$\frac{1}{N} \sum_{i=1}^N \frac{1}{T^2} \sum_{t=1}^T \tilde{X}_{i,t} \tilde{X}'_{i,t} \implies \int \tilde{B}_{F\Lambda}^X \tilde{B}_{F\Lambda}^{X'} + \frac{1}{6} \Psi^{XX},$$

where Ψ^{XX} is the lower right $m \times m$ block of Ψ .

Combining the results given above yields Proposition 3.1 A(a) and B(a). \square

Proof of Proposition 3.1 (b): Convergence of $\tilde{\rho}$

The residuals from the first stage PLS regression are given by $\tilde{u}_{i,t} = (1 - \tilde{\beta})Z_{i,t} = Y_{i,t} - \tilde{\beta}X_{i,t}$. For the pooled regression given in (3.13) we have

$$(\tilde{\rho} - 1) = \left(\sum_{i=1}^N \sum_{t=2}^T (1 - \tilde{\beta}) \Delta Z_{i,t} \tilde{Z}'_{i,t-1} (1 - \tilde{\beta})' \right) \left(\sum_{i=1}^N \sum_{t=2}^T (1 - \tilde{\beta}) \tilde{Z}_{i,t-1} \tilde{Z}'_{i,t-1} (1 - \tilde{\beta})' \right)^{-1}. \quad (3.B.5)$$

For the numerator consider

$$\sum_{i=1}^N \sum_{t=2}^T \Delta Z_{i,t} \tilde{Z}'_{i,t-1} = \sum_{i=1}^N \sum_{t=2}^T (\Lambda_i f_t \tilde{F}'_{t-1} \Lambda'_i + \Delta E_{i,t} \tilde{E}'_{i,t-1} + \Lambda_i f_t \tilde{E}'_{i,t-1} + \Delta E_{i,t} \tilde{F}'_{t-1} \Lambda'_i). \quad (3.B.6)$$

From Lemma 3.B.1 (c), $\frac{1}{N} \sum_{i=1}^N \frac{1}{T} \sum_{t=2}^T \Lambda_i f_t \tilde{F}'_{t-1} \Lambda'_i \implies \int \mathbf{d}B_{F\Lambda} \tilde{B}'_{F\Lambda} + \Theta_{F\Lambda}$ as $T \rightarrow \infty$ followed by $N \rightarrow \infty$.

If the idiosyncratic terms are I(0), i.e. the true DGP is given by (3.3),

$$\sum_{i=1}^N \sum_{t=2}^T \Delta E_{i,t} \tilde{E}'_{i,t-1} = \sum_{i=1}^N \sum_{t=2}^T ((e_{i,t} - e_{i,t-1})e'_{i,t-1} - (e_{i,t} - e_{i,t-1})\bar{e}'_i),$$

where $\bar{e}_i = \frac{1}{T} \sum_{t=1}^T e_{i,t}$. Now, $\frac{1}{T} \sum_{t=2}^T e_{i,t} e'_{i,t-1} \xrightarrow{p} \gamma_{i1}$ as $T \rightarrow \infty$, with $\gamma_{i1} = \lim_{T \rightarrow \infty} \frac{1}{T} \sum_{i=1}^T \mathbf{E}(e_{i,t} e'_{i,t-1})$, and $\frac{1}{N} \sum_{i=1}^N \gamma_{i1} \xrightarrow{p} \gamma_1$ as $N \rightarrow \infty$, with $\gamma_1 \equiv \mathbf{E}(\gamma_{i1})$. Also, $\frac{1}{T} \sum_{t=2}^T e_{i,t-1} e'_{i,t-1} \xrightarrow{p} \Upsilon_i$ as $T \rightarrow \infty$ and $\frac{1}{N} \sum_{i=1}^N \Upsilon_i \xrightarrow{p} \Upsilon$ as $N \rightarrow \infty$. Furthermore, $\frac{1}{T} \sum_{t=2}^T e_{i,t} \bar{e}'_i \xrightarrow{p} 0$ and $\frac{1}{T} \sum_{t=2}^T e_{i,t-1} \bar{e}'_i \xrightarrow{p} 0$ as $T \rightarrow \infty$. Hence, $\frac{1}{N} \sum_{i=1}^N \frac{1}{T} \sum_{t=2}^T \Delta E_{i,t} \tilde{E}'_{i,t-1} \xrightarrow{p} \gamma_1 - \Upsilon$ as $T \rightarrow \infty$ followed by $N \rightarrow \infty$.

For the third term in (3.B.6) we have as $T \rightarrow \infty$,

$$\frac{1}{T} \sum_{t=2}^T \Lambda_i f_t \tilde{E}'_{i,t-1} = \frac{1}{T} \sum_{t=2}^T \Lambda_i f_t e'_{i,t-1} - \frac{1}{T} \sum_{t=2}^T \Lambda_i f_t \bar{e}'_i \xrightarrow{p} 0,$$

Finally, as $T \rightarrow \infty$,

$$\frac{1}{T} \sum_{t=2}^T \Delta E_{i,t} \tilde{F}'_{t-1} \Lambda'_i = \frac{1}{T} e_{i,T} \tilde{F}'_{T-1} \Lambda'_i - \frac{1}{T} e_{i,1} \tilde{F}'_1 \Lambda'_i - \frac{1}{T} \sum_{t=2}^T e_{i,t-1} f'_{t-1} \Lambda'_i \xrightarrow{p} 0.$$

Hence, as $T \rightarrow \infty$ followed by $N \rightarrow \infty$,

$$\frac{1}{N} \sum_{i=1}^N \frac{1}{T} \sum_{t=2}^T \Delta Z_{i,t} \tilde{Z}'_{i,t-1} \implies \int \mathbf{d}B_{F\Lambda} \tilde{B}'_{F\Lambda} + \Theta_{F\Lambda} + \gamma_1 - \Upsilon.$$

If the idiosyncratic components are l(1) and their true DGP includes (3.4), such that $\Delta E_{i,t} = e_{i,t}$ and $\tilde{E}_{i,t-1} = \tilde{S}_{i,t-1}$, using Lemmas 3.B.1 (c), 2 (c) and 3 (d) and (f), we obtain as $T \rightarrow \infty$ followed by $N \rightarrow \infty$,

$$\frac{1}{N} \sum_{i=1}^N \frac{1}{T} \sum_{t=2}^T \Delta Z_{i,t} \tilde{Z}'_{i,t-1} \implies \int \mathbf{d}B_{F\Lambda} \tilde{B}_{F\Lambda} + \Theta_{F\Lambda} - \frac{1}{2} \Psi + \Delta.$$

For the denominator in (3.B.5) consider

$$\begin{aligned} \sum_{i=1}^N \sum_{t=2}^T \tilde{Z}_{i,t-1} \tilde{Z}'_{i,t-1} &= \sum_{i=1}^N \sum_{t=2}^T (\Lambda_i \tilde{F}_{t-1} \tilde{F}'_{t-1} \Lambda'_i + \tilde{E}_{i,t-1} \tilde{E}'_{i,t-1} \\ &\quad + \Lambda_i \tilde{F}_{t-1} \tilde{E}'_{i,t-1} + \tilde{E}_{i,t-1} \tilde{F}'_{t-1} \Lambda'_i). \end{aligned} \quad (3.B.7)$$

If the idiosyncratic components are given by (3.3), we find, $\frac{1}{N} \sum_{i=1}^N \frac{1}{T^2} \sum_{t=2}^T \tilde{Z}_{i,t-1} \tilde{Z}'_{i,t-1} \implies \int \tilde{B}_{F\Lambda} \tilde{B}'_{F\Lambda}$ as $T \rightarrow \infty$ followed by $N \rightarrow \infty$.

For l(1) idiosyncratic components given by (3.4), we find using Lemmas 3.B.1 (d), 3.B.2 (d) and 3.B.3(g)

$$\frac{1}{N} \sum_{i=1}^N \frac{1}{T^2} \sum_{t=2}^T \tilde{Z}_{i,t-1} \tilde{Z}'_{i,t-1} \implies \int \tilde{B}_{F\Lambda} \tilde{B}'_{F\Lambda} + \frac{1}{6} \Psi$$

as $T \rightarrow \infty$ followed by $N \rightarrow \infty$. Combining the above given results with those of A (a) or B(a) yields Proposition 3.1 A (b) and B(b). \square

Proof of Proposition 3.1 (c): Divergence of $t_{\tilde{\rho}}$

The t -statistic for $\tilde{\rho} = 1$ is given by

$$t_{\tilde{\rho}} = (\tilde{\rho} - 1)s^{-1} \left(\sum_{i=1}^N \sum_{t=2}^T \tilde{u}_{i,t-1}^2 \right)^{\frac{1}{2}},$$

where

$$s^2 = \frac{1}{N} \sum_{i=1}^N \frac{1}{T} \sum_{t=2}^T (\Delta \tilde{u}_{i,t}^2 + O_p(1)).$$

As

$$\frac{1}{N} \sum_{i=1}^N \frac{1}{T} \sum_{t=2}^T \Delta \tilde{u}_{i,t}^2 = \frac{1}{N} \sum_{i=1}^N \frac{1}{T} \sum_{t=2}^T (1 - \tilde{\beta}) \Delta Z_{i,t} \Delta \tilde{Z}'_{i,t} (1 - \tilde{\beta})',$$

which is $O_p(1)$ whether the idiosyncratic components are I(0) or I(1), s^2 is $O_p(1)$. Furthermore, $T(\tilde{\rho} - 1)$ and $\frac{1}{N} \sum_{i=1}^N \frac{1}{T^2} \sum_{t=2}^T \tilde{u}_{i,t-1} \tilde{u}'_{i,t-1}$ are $O_p(1)$ as well whether $E_{i,t}$ is given by (3.3) or (3.4), as shown above. Hence,

$$t_{\tilde{\rho}} = \sqrt{NT}(\tilde{\rho} - 1)s^{-1} \left(\frac{1}{N} \sum_{i=1}^N \frac{1}{T^2} \sum_{t=2}^T \tilde{u}_{i,t-1} \tilde{u}'_{i,t-1} \right)^{\frac{1}{2}} = \sqrt{N}O_p(1),$$

which diverges at rate \sqrt{N} as $T \rightarrow \infty$ followed by $N \rightarrow \infty$. \square

Proof of Proposition 3.2 (a): Convergence of $\tilde{\beta}_i$

For each panel unit i , the estimator of β_i is given by $\tilde{\beta}_i = (\sum_{t=1}^T \tilde{Y}_{i,t} \tilde{X}'_{i,t}) (\sum_{t=1}^T \tilde{X}_{i,t} \tilde{X}'_{i,t})^{-1}$. Consider the numerator

$$\sum_{t=1}^T \tilde{Y}_{i,t} \tilde{X}'_{i,t} = \sum_{t=1}^T (\lambda'_{1i} \tilde{F}_t^Y \tilde{F}_t^{X'} \lambda_{2i} + \tilde{E}_{i,t}^Y \tilde{E}_{i,t}^{X'} + \lambda'_{1i} \tilde{F}_t^Y \tilde{E}_{i,t}^{X'} + \tilde{E}_{i,t}^Y \tilde{F}_t^{X'} \lambda_{2i}). \quad (3.B.8)$$

If the idiosyncratic term is given by (3.3), we have $O_p(T^2) + O_p(T) + O_p(T) + O_p(T)$ in (3.B.8). So, as $T \rightarrow \infty$, $\frac{1}{T^2} \sum_{t=1}^T \tilde{Y}_{i,t} \tilde{X}'_{i,t} \Rightarrow \lambda'_{1i} \int \tilde{B}_F^Y \tilde{B}_F^{X'} \lambda_{2i}$ from the first result of Lemma 3.B.1 (d).

If the idiosyncratic terms are also I(1), such that the DGP includes (3.4), all terms in (3.B.8) are $O_p(T^2)$ when summed over T . Using Lemmas 3.B.1 (d), 3.B.2 (d) and 3.B.3 (g) we find as $T \rightarrow \infty$,

$$\frac{1}{T^2} \sum_{t=1}^T \tilde{Y}_{i,t} \tilde{X}'_{i,t} \Rightarrow (\lambda'_{1i} \int \tilde{B}_F^Y \tilde{B}_F^{X'} \lambda_{2i} + \int \tilde{B}_i^Y \tilde{B}_i^{X'} + \lambda'_{1i} \int \tilde{B}_F^Y \tilde{B}_i^{X'} + \int \tilde{B}_i^Y \tilde{B}_F^{X'} \lambda_{2i}).$$

Now the denominator of $\tilde{\beta}_i$ is given by

$$\sum_{t=1}^T \tilde{X}_{i,t} \tilde{X}'_{i,t} = \sum_{t=1}^T (\lambda'_{2i} \tilde{F}_t^X \tilde{F}_t^{X'} \lambda_{2i} + \tilde{E}_{i,t}^X \tilde{E}_{i,t}^{X'} + \lambda'_{2i} \tilde{F}_t^X \tilde{E}_{i,t}^{X'} + \tilde{E}_{i,t}^X \tilde{F}_t^{X'} \lambda_{2i}). \quad (3.B.9)$$

Similar to the results for the numerator, the terms in (3.B.9) are $O_p(T^2) + O_p(T) + O_p(T) + O_p(T)$, if the DGP contains (3.3). Hence, $\frac{1}{T^2} \sum_{t=1}^T \tilde{X}_{i,t} \tilde{X}'_{i,t} \implies \lambda'_{2i} \int \tilde{B}_F^X \tilde{B}_F^{X'} \lambda_{2i}$ as $T \rightarrow \infty$.

If the true DGP contains (3.4), all terms in (3.B.9) are $O_p(T^2)$ and we have, as $T \rightarrow \infty$,

$$\frac{1}{T^2} \sum_{t=1}^T \tilde{X}_{i,t} \tilde{X}'_{i,t} \implies (\lambda'_{2i} \int \tilde{B}_F^X \tilde{B}_F^{X'} \lambda_{2i} + \int \tilde{B}_i^X \tilde{B}_i^{X'} + \lambda'_{2i} \int \tilde{B}_F^X \tilde{B}_i^{X'} + \int \tilde{B}_i^X \tilde{B}_F^{X'} \lambda_{2i}).$$

Combining the results given above yields Proposition 3.2 A(a) and B(a). \square

Proposition 3.2 (b): Convergence of $Z_{\tilde{\rho}_{NT-1}}$ and $\tilde{Z}_{\tilde{\rho}_{NT-1}}$

The residuals from the individual first stage regression are given by $\tilde{u}_{i,t} = (1, -\tilde{\beta}_i)Z_{i,t} = Y_{i,t} - \tilde{\beta}_i X_{i,t}$. Consider first

$$\sum_{t=2}^T \Delta \tilde{u}_{i,t} \tilde{u}'_{i,t-1} = \sum_{t=2}^T (1, -\tilde{\beta}_i) \Delta Z_{i,t} \tilde{Z}'_{i,t-1} (1, -\tilde{\beta}_i)'. \quad (3.B.10)$$

Now,

$$\begin{aligned} \sum_{t=2}^T \Delta Z_{i,t} \tilde{Z}'_{i,t-1} &= \sum_{t=2}^T (\Lambda_i f_t + \Delta E_{i,t}) (\Lambda_i \tilde{F}_{t-1} + \tilde{E}_{i,t-1})' \\ &= \sum_{t=2}^T (\Lambda_i f_t \tilde{F}'_{t-1} \Lambda'_i + \Delta E_{i,t} \tilde{E}'_{i,t-1} + \Lambda_i f_t \tilde{E}'_{i,t-1} + \Delta E_{i,t} \tilde{F}'_{t-1} \Lambda'_i). \end{aligned} \quad (3.B.11)$$

From Lemma 3.B.1 (c), $\frac{1}{T} \sum_{t=2}^T \Lambda_i f_t \tilde{F}'_{t-1} \Lambda'_i \implies \int \Lambda_i (dB_F \tilde{B}_F + \Theta) \Lambda'_i$ as $T \rightarrow \infty$. If the idiosyncratic terms are I(0), i.e. the true DGP is given by (3.3),

$$\sum_{t=2}^T \Delta E_{i,t} \tilde{E}'_{i,t-1} = \sum_{t=2}^T ((e_{i,t} - e_{i,t-1}) e'_{i,t-1} - (e_{i,t} - e_{i,t-1}) \bar{e}_i),$$

where $\bar{e}_i = \frac{1}{T} \sum_{t=1}^T e_{i,t}$. Now, $\frac{1}{T} \sum_{t=2}^T e_{i,t} e'_{i,t-1} \xrightarrow{p} \gamma_{i1}$ as $T \rightarrow \infty$, with $\gamma_{i1} = \lim_{T \rightarrow \infty} \frac{1}{T} \sum_{i=1}^T \mathbf{E}(e_{i,t} e_{i,t-1})$. Also, $\frac{1}{T} \sum_{t=2}^T e_{i,t-1} e'_{i,t-1} \xrightarrow{p} \Upsilon_i$ as $T \rightarrow \infty$. Furthermore, $\frac{1}{T} \sum_{t=2}^T e_{i,t} \bar{e}'_i \xrightarrow{p} 0$ and $\frac{1}{T} \sum_{t=2}^T e_{i,t-1} \bar{e}'_i \xrightarrow{p} 0$ as $T \rightarrow \infty$. Hence, $\frac{1}{T} \sum_{t=2}^T \Delta E_{i,t} \tilde{E}'_{i,t-1} \xrightarrow{p} \gamma_{i1} - \Upsilon_i$ as $T \rightarrow \infty$.

For the third term in (3.B.11) we have, as $T \rightarrow \infty$,

$$\frac{1}{T} \sum_{t=2}^T \Lambda_i f_t \tilde{E}'_{i,t-1} = \frac{1}{T} \sum_{t=2}^T \Lambda_i f_t e'_{i,t-1} - \frac{1}{T} \sum_{t=2}^T \Lambda_i f_t \bar{e}'_i \xrightarrow{p} 0.$$

Finally, as $T \rightarrow \infty$,

$$\frac{1}{T} \sum_{t=2}^T \Delta E_{i,t} \tilde{F}'_{t-1} \Lambda'_i = \frac{1}{T} e_{i,T} \tilde{F}'_{T-1} \Lambda'_i - \frac{1}{T} e_{i,1} \tilde{F}'_1 \Lambda'_i - \frac{1}{T} \sum_{t=2}^T e_{i,t-1} f'_{t-1} \Lambda'_i \xrightarrow{p} 0.$$

Hence, as $T \rightarrow \infty$,

$$\frac{1}{T} \sum_{t=2}^T \Delta Z_{i,t} \tilde{Z}'_{i,t-1} \implies \Lambda_i \left(\int \mathbf{d}B_F \tilde{B}_F + \Theta \right) \Lambda'_i + \gamma_{i1} - \Upsilon_i.$$

If the idiosyncratic components are I(1) and their true DGP includes (3.4), such that $\Delta E_{i,t} = e_{i,t}$ and $\tilde{E}_{i,t-1} = \tilde{S}_{i,t-1}$, using Lemmas 3.B.1 (c), 3.B.2 (c) and 3.B.3 (d) and (f), we obtain, as $T \rightarrow \infty$,

$$\frac{1}{T} \sum_{t=2}^T \Delta Z_{i,t} \tilde{Z}'_{i,t-1} \implies (\Lambda'_i \left(\int \mathbf{d}B_F \tilde{B}'_F + \Theta \right) \Lambda'_i + \int \mathbf{d}B_i \tilde{B}'_i + \Delta_i + \Lambda_i \int \mathbf{d}B_F \tilde{B}'_i + \int \mathbf{d}B_i \tilde{B}'_F \Lambda'_i).$$

Furthermore, note that the residuals $\tilde{v}_{i,t} = \Delta \tilde{u}_{i,t} + o_p(1)$ regardless of whether they were obtained from the pooled regression (3.13) or the individual regression (3.15). Now,

$$\begin{aligned} \hat{\lambda}_i &= T^{-1} \sum_{s=1}^J \omega_{sJ} \sum_{t=s+1}^T \tilde{v}_{i,t} \tilde{v}_{i,t-s} \\ &= T^{-1} \sum_{s=1}^J \omega_{sJ} \sum_{t=s+1}^T \Delta \tilde{u}_{i,t} \Delta \tilde{u}_{i,t-s} + o_p(1) \\ &= T^{-1} \sum_{s=1}^J \omega_{sJ} \sum_{t=s+1}^T (1, -\tilde{\beta}_i) \Delta \tilde{Z}_{i,t} \Delta \tilde{Z}'_{i,t-s} (1, -\tilde{\beta}_i)' + o_p(1). \end{aligned}$$

Expanding $\Delta \tilde{Z}_{i,t} \Delta \tilde{Z}'_{i,t-s}$ in terms of the common factors and unobserved components we obtain the following four terms and convergence results for suitable choices of bandwidth J and kernel function ω_{sJ} . First,

$$T^{-1} \sum_{s=1}^J \omega_{sJ} \sum_{t=s+1}^T \Lambda_i \tilde{f}_{i,t} \tilde{f}'_{i,t-s} \Lambda'_i \xrightarrow{p} \Lambda_i \Omega \Lambda'_i. \quad (3.B.12)$$

Next,

$$T^{-1} \sum_{s=1}^J \omega_{sJ} \sum_{t=s+1}^T \Lambda_i \tilde{f}_{i,t} \Delta \tilde{E}'_{i,t-s} \xrightarrow{p} 0, \quad (3.B.13)$$

and

$$T^{-1} \sum_{s=1}^J \omega_{sJ} \sum_{t=s+1}^T \Delta \tilde{E}_{i,t} \tilde{f}'_{i,t-s} \Lambda_i \xrightarrow{p} 0, \quad (3.B.14)$$

due to the independence of common factors and idiosyncratic components. Finally,

$$T^{-1} \sum_{s=1}^J \omega_{sJ} \sum_{t=s+1}^T \Delta \tilde{E}_{i,t} \Delta \tilde{E}'_{i,t-s} \xrightarrow{p} \lim_{T \rightarrow \infty} \frac{1}{T} \sum_{t=1}^T \mathbb{E}(e_{i,t} \tilde{E}_{i,t}), \quad (3.B.15)$$

which is $\gamma_{i1} - \Upsilon_i$ if the idiosyncratic components are stationary, and Δ_i if they are I(1).

Now consider

$$\sum_{t=2}^T \Delta \tilde{u}_{i,t} \tilde{u}_{i,t-1} = \sum_{t=2}^T (1, -\tilde{\beta}_i) \Delta Z_{i,t} \tilde{Z}'_{i,t-1} (1, -\tilde{\beta}_i)'. \quad (3.B.16)$$

We have

$$\begin{aligned} \sum_{t=2}^T \tilde{Z}_{i,t-1} \tilde{Z}'_{i,t-1} &= \sum_{t=2}^T (\Lambda_i \tilde{F}_{t-1} \tilde{F}'_{t-1} \Lambda'_i + \tilde{E}_{i,t-1} \tilde{E}'_{i,t-1} \\ &\quad + \Lambda_i \tilde{F}_{t-1} \tilde{E}'_{i,t-1} + \tilde{E}_{i,t-1} \tilde{F}'_{t-1} \Lambda'_i). \end{aligned} \quad (3.B.17)$$

If the idiosyncratic components are given by (3.3), when summed over T the first term in (3.B.17) is $O_p(T^2)$, while the remaining three are $O_p(T)$. So, $\frac{1}{T^2} \sum_{t=2}^T \tilde{Z}_{i,t-1} \tilde{Z}'_{i,t-1} \implies \Lambda_i \int \tilde{B}_F \tilde{B}'_F \Lambda'_i$ as $T \rightarrow \infty$.

For I(1) idiosyncratic components given by (3.4), we find using Lemmas 3.B.1 (d), 3.B.2 (d) and 3.B.3(g), as $T \rightarrow \infty$,

$$\frac{1}{T^2} \sum_{t=2}^T \tilde{Z}_{i,t-1} \tilde{Z}'_{i,t-1} \implies (\Lambda_i \int \tilde{B}_F \tilde{B}'_F \Lambda'_i + \int \tilde{B}_i \tilde{B}'_i + \Lambda_i \int \tilde{B}_F \tilde{B}'_i + \int \tilde{B}_i \tilde{B}'_F \Lambda'_i).$$

We use the block-triangular decomposition of the long-run covariance matrix of the common non-stationary factors Ω , such that $\Omega = L'L$ with $L_{11} = \Omega_{11} - \Omega'_{21} \Omega_{22}^{-1} \Omega_{21}$, $L_{21} = \Omega_{22}^{-\frac{1}{2}} \Omega_{21}$, and $L_{22} = \Omega_{22}^{\frac{1}{2}}$, where blocks are conformable to the partition of $B_F = (B_F^Y, B_F^X)'$. Note that $\Omega_{22} > 0$ by Assumption 3.1.

Now, $\tilde{B}_F = L' \tilde{W}_F$, where \tilde{W}_F is a demeaned k -vector standard Brownian motion. Furthermore, denote $\boldsymbol{\eta}'_i = (1, -\tilde{\mathbf{b}}_{iA})$, and $\boldsymbol{\kappa}' = (\mathbf{I}_{k_Y}, -(\int \tilde{W}_F^Y \tilde{W}_F^{X'}) (\int \tilde{W}_F^X \tilde{W}_F^{X'})^{-1})$. Then, $L \Lambda'_i \boldsymbol{\eta}_i = \boldsymbol{\kappa} L_{11} \lambda_{1i}$, and $\boldsymbol{\eta}'_i \tilde{B}_F = \lambda'_{1i} L'_{11} \tilde{Q}_F$, with $\tilde{Q}_F = \tilde{W}_F^Y - (\int \tilde{W}_F^Y \tilde{W}_F^{X'}) (\int \tilde{W}_F^X \tilde{W}_F^{X'})^{-1} \tilde{W}_F^X$. Finally,

$$\boldsymbol{\eta}'_i \int \mathbf{d}B_F \tilde{B}'_F \boldsymbol{\eta}_i = \lambda'_{1i} L'_{11} \int \mathbf{d}Q_F \tilde{Q}'_F L_{11} \lambda_{1i},$$

and

$$\boldsymbol{\eta}'_i \int \tilde{B}_F \tilde{B}'_F \boldsymbol{\eta}_i = \lambda'_{1i} L'_{11} \int \tilde{Q}_F \tilde{Q}'_F L_{11} \lambda_{1i}.$$

Combining the above given results with those of A (a) or B (a) yields the convergence results for $Z_{\tilde{\rho}_{NT-1}}$ and $\tilde{Z}_{\tilde{\rho}_{NT-1}}$. \square

Panel Error Correction Testing with Global Stochastic Trends¹

4.1 Introduction

Consider two non-stationary panel data variables $X_{i,t}$ and $Y_{i,t}$, where $i = 1, \dots, N$ and $t = 1, \dots, T$ indexes the cross-sectional and time series dimensions, respectively. The analysis of such variables has been a growing field of econometric research in recent years. See for example Breitung and Pesaran (2008) for an overview. In particular, in many economic applications it is an important question whether $X_{i,t}$ and $Y_{i,t}$ are cointegrated, that is whether there exists a meaningful long-run relationship between them, or whether the relationship is spurious.

Kao (1999) and Pedroni (1999, 2004a) were among the first to propose residual-based tests for the null hypothesis of no cointegration in cross-sectionally independent panels. But cross-sectional independence is a restrictive assumption that is unlikely to be met in practice, in which case the properties of this kind of tests become suspect. In fact, in a recent paper, Gengenbach et al. (2006) show that the presence of cross-section dependence in the form of non-stationary common factors can actually cause the residual-based tests of Kao (1999) and Pedroni (2004a) to become divergent. As a response to this, they propose to estimate separately the common and idiosyncratic components of $X_{i,t}$ and $Y_{i,t}$ using the principal components method of Bai and Ng (2004b), and then to test for cointegration in the resulting component estimates.

Banerjee and Carrion-i Silvestre (2006) propose a similar test but instead of applying the Bai and Ng (2004b) approach to $X_{i,t}$ and $Y_{i,t}$ directly, they apply it to the residuals of a first-stage regression of $Y_{i,t}$ onto $X_{i,t}$. Cointegration requires that both the common and idiosyncratic components of the residuals are stationary. The tests of Bai and Carrion-i Silvestre (2007), Westerlund (2007) and Westerlund and Edgerton (2008) are basically the same in the sense that they are also based on applying the Bai and Ng (2004b) approach to the residuals of a first-stage regression.

¹This Chapter is based on Gengenbach et al. (2008).

However, although very popular, this testing approach has at least two major drawbacks. One lies with the use of residual rather than structural dynamics, which makes it subject to the common factor critique of Kremers, Ericsson, and Dolado (1992), that may lead to tests with low power. The second drawback is that the testing must be carried out in steps, with the estimation error from one step being imported into subsequent steps, and it is not fully clear what effect this has on the final test, see Westerlund and Larsson (2008).

By contrast to the test proposed by Pedroni for example, the tests of Westerlund (2007) are not based on residuals but rather on the significance of the error correction term in a conditional panel error correction model (ECM), and therefore do not impose any common factor restriction. However, the tests are derived under cross-sectional independence, and the use of the bootstrap in case of violations does not fit well with the otherwise parametric flavor of the tests. Another drawback is that the bootstrap used is not equipped to handle the case with non-stationary common factors.

The current chapter can be seen as an attempt to overcome the drawbacks of both these approaches. We begin by developing alternative representations of a cointegrated panel that allows for the possibility of non-stationary common factors. In particular, starting from the triangular representation of the system used by for example Bai et al. (2009), we derive a Granger type representation theorem that is similar to the one obtained by Cappuccio and Lubian (1996) in the case of a single time series.

The Granger representation theorem provides not only moving average (MA) and autoregressive moving average (ARMA) representations of the system, but also the conditional ECM representation, which we use as a basis for developing tests for the null hypothesis of no error correction. In particular, paralleling the development of the time series literature in this field, as pioneered by Banerjee et al. (1998) and Boswijk (1994), we consider both a t -ratio type test, as well as a Wald type test. Besides eliminating the need for a common factor assumption and a stepwise testing procedure, as shown by Pesavento (2004), these tests are not only more powerful than most residual-based tests around, but are also not worse in terms of size distortions.

It is shown that at the level of the individual unit the asymptotic distribution of the Wald tests is free of nuisance parameters and only depends on the number of non-stationary variables in the system. For the t -ratio an appropriate correction has to be employed to remove the nuisance parameter dependence from the limiting distribution. Nevertheless, because of the common factors, the individual tests are not independent, which of course makes pooling, or cross-sectional averaging, difficult, as it invalidates the use of the conventional limit theory. However, although not analytically

tractable, the average still converges to a random variable with a distribution that can be easily simulated, which makes pooling possible in spite of the dependence. We begin by considering the case when the common factors are known, and then we show how the results extend to the case when the factors are approximated by means of cross-sectional averages of the observed data, as suggested by Pesaran (2007).

The rest of this chapter is organized as follows. Section 4.2 presents the model of interest and our version of the Granger representation theorem. Sections 4.3 and 4.4 then present the error correction tests and their asymptotic properties, which are verified using both simulated and real data in Sections 4.5 and 4.6, respectively. Section 4.7 concludes.

A word on notation. The symbols \xrightarrow{w} and \xrightarrow{p} will be used to signify weak convergence and convergence in probability, respectively. As usual, $X_T = O_p(T^r)$ will be used to signify that X_T is at most order T^r in probability, while $X_T = o_p(T^r)$ will be used in case X_T is of smaller order in probability than T^r . In the case of a double indexed sequence $X_{N,T}$, $N, T \rightarrow \infty$ will be used to signify that the limit has been taken while passing both indices to infinity jointly. For a square matrix A , $rk(A)$, $adj(A)$ and $\|A\|$ will denote its rank, adjoint and Euclidian norm, respectively. For simplicity, the Brownian motion $B(s)$ defined on the interval $s \in [0, 1]$ will be written B , with the measure of integration omitted. We write the integral $\int_0^1 B(s)ds$ as $\int B$ and $\int_0^1 B(s)dB(s)'$ as $\int BdB'$. Finally, $[x]$ will be used to denote the integer part of x .

4.2 Model representation

In this section we discuss the model under consideration, and some alternative representations thereof. We start from the triangular representation for a single unit i , which is the same as the one used by Bai et al. (2009). However, these authors focus on how to conduct inference if the variables are in fact long-run related, and do not consider the problem of how to test for cointegration. Moreover, the triangular representation is taken as given, and there is no consideration of other alternatives. Thus, the results reported herein can in many ways be seen as complementary to those reported in Bai et al. (2009).

The data generating process has two basic building blocks, a $(r + m)$ -dimensional vector of idiosyncratic variables, which is denoted by $Z_{i,t} = (Y'_{i,t}, X'_{i,t})'$, where $Y_{i,t}$ is $r \times 1$ while $X_{i,t}$ is $m \times 1$, and a k -dimensional vector of common factors, which is denoted by F_t . The grand vector containing all three variables is denoted $Z_{i,t}^+ = (Z'_{i,t}, F'_t)'$, and for later use we will also let $V_{i,t} = (X'_{i,t}, F'_t)'$ denote the augmented $X_{i,t}$ vector.

The data generating process can be written in the following way

$$Y_{i,t} - \pi'_{1i} G_t = b'_i X_{i,t} + \lambda'_{1i} F_t + u_{1i,t}, \quad (4.2.1)$$

$$\Delta X_{i,t} - \pi'_{2i} g_t = \lambda'_{2i} \Delta F_t + u_{2i,t}, \quad (4.2.2)$$

$$\Delta F_t - \pi'_{3i} g_t = f_t, \quad (4.2.3)$$

where G_t and g_t are vectors of deterministic components such that $g_t = \Delta G_t$ with associated coefficients $\pi_i = \begin{pmatrix} \pi_{1i} & \pi_{2i} & \pi_{3i} \end{pmatrix}$.

We further assume that the vector $u_{i,t}^+ = (u'_{1i,t}, u'_{2i,t}, f'_t)'$ is a stationary linear process given by

$$\begin{aligned} u_{i,t}^+ &= \begin{pmatrix} \Gamma_{11i}(L) & \Gamma_{12i}(L) & 0 \\ \Gamma_{21i}(L) & \Gamma_{22i}(L) & 0 \\ 0 & 0 & \Psi(L) \end{pmatrix} \begin{pmatrix} \varepsilon_{1i,t} \\ \varepsilon_{2i,t} \\ \eta_t \end{pmatrix} = \begin{pmatrix} \Gamma_i(L) & 0 \\ 0 & \Psi(L) \end{pmatrix} \begin{pmatrix} \varepsilon_{i,t} \\ \eta_t \end{pmatrix} \\ &= \Gamma_i^+(L) \varepsilon_{i,t}^+, \end{aligned} \quad (4.2.4)$$

where $\Psi(L) = I_k - \sum_{j=1}^{\infty} \Gamma_{33j} L^j$ and L is the lag operator. Similarly,

$$\Gamma_i(L) = I_{(r+m)} - \sum_{j=1}^{\infty} \Gamma_{ij} L^j.$$

Equations (4.2.1) to (4.2.4) constitute the triangular representation of the model. The rest of the assumptions can be summarized in the following way, where $M < \infty$ denotes a generic positive real number.

Assumption 4.1 (i) $\eta_t \sim i.i.d.(0, I_k)$ with finite fourth moments, (ii) $\sum_{j=0}^{\infty} j \cdot \|\Psi_j\| < M$, (iii) $rk(\Psi(1)) = k$.

Assumption 4.2 (i) $\varepsilon_{i,t} \sim i.i.d.(0, \Sigma_i)$ with finite eighth moments and

$$\Sigma_i = \begin{pmatrix} \Sigma_{11i} & \Sigma_{12i} \\ \Sigma_{21i} & \Sigma_{22i} \end{pmatrix} = \text{cov}(\varepsilon_{i,t}),$$

(ii) $E(\varepsilon_{i,t} \varepsilon_{j,s}) = 0$ for all $i \neq j$ and $t \neq s$, (iii) $\Gamma_i(L)$ fulfils the random coefficient and summability conditions of Phillips and Moon (1999, Assumptions 1 and 2), (iv) $rk(\Gamma_i(1)) = r + m$.

Assumption 4.3 (i) $\Lambda_i = (\lambda_{1i}, \lambda_{2i})'$ is a random matrix such that $\|\Lambda_i\| < M$, (ii) $\bar{\Lambda} = \frac{1}{N} \sum_{i=1}^N \Lambda_i \rightarrow E(\Lambda_i) = \Lambda < M$ as $N \rightarrow \infty$, (iii) $rk(\bar{\Lambda}) = k \leq r + m$.

Assumption 4.4 η_t , $\varepsilon_{i,t}$ and Λ_i are mutually independent.

Assumptions 4.1, 4.2 and 4.4 imply that for any i , $\varepsilon_{i,t}^+ \sim i.i.d.(0, \Sigma_i)$ with

$$\Sigma_i^+ = \begin{pmatrix} \Sigma_i & 0 \\ 0 & I_k \end{pmatrix} = \text{cov}(\varepsilon_{i,t}^+).$$

They also imply that $rk(\Gamma_i^+(1)) = r + m + k$. Under these assumptions, it is easy to see that the system has r cointegrating relationships $\beta_i' Z_{i,t}^+$, where by assumption

$$\beta_i = \begin{pmatrix} I_r & -b_i' & -\lambda_{1i}' \end{pmatrix}'$$

is the cointegrating matrix.

Similar to the time series case considered by Cappuccio and Lubian (1996), given the triangular representation in (4.2.1) to (4.2.4), we can derive a Granger type representation theorem for a given panel member. This provides us with alternative model representations that are better suited for testing the hypothesis of no cointegration.

Theorem 4.1 *Given the triangular representation in (4.2.1) to (4.2.4), $Z_{i,t}^+$ is non-stationary with cointegration rank r .*

(a) *The MA representation of $\Delta Z_{i,t}^+$ is*

$$\Delta Z_{i,t}^+ - (\pi_i^*)' g_t = C_i(L) \varepsilon_{i,t}^+, \quad (4.2.5)$$

where $C_i(L)$ is given in the appendix, $rk(C_i(1)) = m + k$ and

$$\pi_i^* = \begin{pmatrix} \pi_{1i} + \pi_{2i} b_i + \pi_3 (\lambda_{1i} + \lambda_{2i} b_i) & \pi_{2i} + \pi_3 \lambda_{2i} & \pi_3 \end{pmatrix}.$$

(b) *The ARMA representation of $Z_{i,t}^+$ is given by*

$$A_i(L) (Z_{i,t}^+ - (\pi_i^*)' g_t) = c_i(L) \varepsilon_{i,t}^+, \quad (4.2.6)$$

where $c_i(L) = |\Gamma_i^+(L)|$ is a scalar lag polynomial, and where the blocks of

$$A_i(L) = \begin{pmatrix} A_{11i}(L) & A_{12i}(L) & A_{13i}(L) \\ A_{21i}(L) & A_{22i}(L) & A_{23i}(L) \\ 0 & 0 & A_{33i}(L) \end{pmatrix}$$

are given by

$$\begin{aligned} A_{11i}(L) &= |\Psi(L)||\Gamma_{22i}(L)|adj(\Gamma_{11\cdot 2i}(L)), \\ A_{12i}(L) &= -|\Psi(L)||\Gamma_{22i}(L)|adj(\Gamma_{11\cdot 2i}(L))((1-L)\Gamma_{12i}(L)\Gamma_{22i}(L)^{-1} + b'_i), \\ A_{13i}(L) &= |\Psi(L)||\Gamma_{22i}(L)|adj(\Gamma_{11\cdot 2i}(L))((1-L)\Gamma_{12i}(L)\Gamma_{22i}(L)^{-1}\lambda'_{2i} - \lambda'_{1i}), \\ A_{21i}(L) &= -|\Psi(L)|adj(\Gamma_{22i}(L))\Gamma_{21i}(L)adj(\Gamma_{11\cdot 2i}(L)), \\ A_{22i}(L) &= |\Psi(L)|adj(\Gamma_{22i}(L))(\Gamma_{21i}(L)adj(\Gamma_{11\cdot 2i}(L))((1-L)\Gamma_{12i}(L)\Gamma_{22i}(L)^{-1} + b'_i) \\ &\quad + (1-L)|\Gamma_{11\cdot 2i}(L)|), \\ A_{23i}(L) &= -|\Psi(L)|adj(\Gamma_{22i}(L))(\Gamma_{21i}(L)adj(\Gamma_{11\cdot 2i}(L)) \\ &\quad \times ((1-L)\Gamma_{12i}(L)\Gamma_{22i}(L)^{-1}\lambda'_{2i} - \lambda'_{1i}) + (1-L)|\Gamma_{11\cdot 2i}(L)|\lambda'_{2i}), \\ A_{33i}(L) &= (1-L)|\Gamma_{22i}(L)||\Gamma_{11\cdot 2i}(L)|adj(\Psi(L)), \end{aligned}$$

with $\Gamma_{11\cdot 2i}(L) = \Gamma_{11i}(L) - \Gamma_{12i}(L)\Gamma_{22i}(L)^{-1}\Gamma_{21i}(L)$.

(c) $A_i(1)$ has reduced rank r and can be decomposed as $A_i(1) = \alpha_i^* \beta_i'$, where

$$\alpha_i^* = \begin{pmatrix} |\Psi(1)||\Gamma_{22i}(1)|adj(\Gamma_{11\cdot 2i}(1)) \\ -|\Psi(1)|adj(\Gamma_{22i}(1))\Gamma_{21i}(1)adj(\Gamma_{11\cdot 2i}(1)) \\ 0 \end{pmatrix}.$$

(d) The vector ECM representation is

$$A_i^*(L)(\Delta Z_{i,t}^+ - (\pi_i^*)' \Delta g_t) = -\alpha_i^* \beta_i' (Z_{i,t-1}^+ - (\pi_i^*)' g_{t-1}) + c_i(L) \varepsilon_{i,t}^+, \quad (4.2.7)$$

where $A_i^*(L) = A_i^+(L) + A_i(1)$ with $A_i^+(L)$ satisfying $A_i(L) = A_i(1) + (1-L)A_i^+(L)$, $A_i^+(L) = \sum_{j=0}^{\infty} A_{ij}^+ L^j$ and $A_{ij}^+ = -\sum_{l=j+1}^{\infty} A_{il}$.

(e) $\xi'_{i,t} = (Z_{i,t}^+)' \beta_i$ has the following representation

$$\begin{aligned}\xi_{i,t} &= \beta'_i(\pi_i^*)' G_t + \begin{pmatrix} \Gamma_{11i}(L) & \Gamma_{12i}(L) \end{pmatrix} \varepsilon_{i,t}^+, \\ \Delta \xi_{i,t} - K_i(L)(\pi_i^*)' g_t &= -\beta'_i \alpha_i^* (\xi_{i,t-1} - \beta'_i(\pi_i^*)' g_t) + J_i(L) \varepsilon_{i,t}^+, \end{aligned}$$

where $K_i(L)$ and $J_i(L)$ can be obtained as in Engle and Granger (1987).

Proof: see Appendix 4.B.1

From the vector ECM representation given in (4.2.7) we can obtain the conditional ECM for $Y_{i,t}$ and the marginal ECM for $V_{i,t}$. Towards this end, let $\alpha_i = -A_i(0)^{-1} \alpha_i^*$ and $\tilde{A}_i^*(L) = A_i(0)^{-1} A_i^*(L)$, where $\tilde{A}_i^{**}(L) = \sum_{j=1}^{\infty} \tilde{A}_{ij}^{**} L^j$ with $\tilde{A}_{ij}^{**} = -\tilde{A}_{i,j+1}^*$ such that

$$\Delta Z_{i,t}^+ - \tilde{A}_i^*(L)(\pi_i^*)' \Delta g_t = \alpha_i \beta'_i (Z_{i,t-1}^+ - (\pi_i^*)' g_{t-1}) + \tilde{A}_i^{**}(L) \Delta Z_{i,t-1}^+ + c_i(L) \varepsilon_{i,t}^+.$$

Defining $B_i^* = (\Sigma_{12i} \Sigma_{22i}^{-1} + b'_i, -\Sigma_{12i} \Sigma_{22i}^{-1} \lambda'_{2i} + \lambda'_{1i})$ and $\kappa_i = (I_r, -B_i^*)$, the conditional ECM for $Y_{i,t}$ is given by

$$\begin{aligned} \Delta Y_{i,t} - \kappa_i \tilde{A}_i^*(L)(\pi_i^*)' \Delta g_t &= B_i^* \Delta V_{i,t} + \kappa_i \alpha_i \beta'_i (Z_{i,t-1}^+ - (\pi_i^*)' g_{t-1}) + \kappa_i \tilde{A}_i^{**}(L) \Delta Z_{i,t-1}^+ \\ &+ c_i(L) \varepsilon_{1.2i,t}, \end{aligned} \quad (4.2.8)$$

where $\varepsilon_{1.2i,t} = \varepsilon_{1i,t} - \Sigma_{12i} \Sigma_{22i}^{-1} \varepsilon_{2i,t}$, while the marginal models for $X_{i,t}$ and F_t are

$$\Delta X_{i,t} - \tilde{A}_{2i}^*(L)(\pi_i^*)' \Delta g_t = \alpha_{2i} \beta'_i (Z_{i,t-1}^+ - (\pi_i^*)' g_{t-1}) + \tilde{A}_{2i}^{**}(L) \Delta Z_{i,t-1}^+ + c_i(L) \varepsilon_{2i,t}^*, \quad (4.2.9)$$

$$\Delta F_t - \tilde{A}_{33i}^*(L) \pi_3' \Delta g_t = A_{33i}^{**}(L) \Delta F_{t-1} + c_i(L) \eta_t, \quad (4.2.10)$$

where $\tilde{A}_{2i}^*(L)$ and $\tilde{A}_{2i}^{**}(L)$ are the second rows of $\tilde{A}_i^*(L)$ and $\tilde{A}_i^{**}(L)$, respectively, and where $\varepsilon_{2i,t}^* = \varepsilon_{2i,t} + \lambda'_{2i} \eta_t$.

Some remarks can be made here.

Remark: What this theorem shows is that alternative representations may lead naturally to alternative approaches to cointegration testing. In particular, while the triangular representation is better suited for developing residual-based tests, the vector ECM, and more precisely its factorization into conditional and marginal models, is more suitable for developing tests based on error correction.

Remark: If $\Gamma_i^+(L)$ is a unimodular matrix polynomial, the MA part in the vector ECM in (4.2.7) vanishes. Furthermore, if $\Gamma_i^+(L)$ is of order p_i , $A_i(L)$ is of order $q_i \leq (r + m + k - 1)p_i$.

Remark: The common factor F_t is by assumption strongly exogenous for β_i , see for example Urbain (1992) for weak and strong exogeneity conditions in this class of models. Similarly, $X_{i,t}$ is weakly exogenous for β_i if $\alpha_{2i}^* = 0$, which will be the case when $\Gamma_{21i}(1) = 0$. It is strongly exogenous if in addition $\Gamma_{21i}(L) = 0$. The relevance of the two latter assumptions will be discussed later.

Remark: Depending on the specification of the deterministic component g_t , we can distinguish at least five variations of the ECM in (4.2.8) to (4.2.10). If $g_t = 0$, henceforth referred to as Model 1, then there are no deterministic components present. If $\pi_{1i} = 0$, then $\beta_i'(\pi_i^*)' = 0$ and hence g_t do not appear in the error correction term. If in addition $g_t = (1, t)'$, then a constant should be included, while if $g_t = (1, t, t^2)'$, then a linear trend should also be included. These specifications are henceforth referred to as Models 2 and 3, respectively. Moreover, if $\pi_{1i} \neq 0$, we have a constant restricted to the error correction term if $g_t = 1$, henceforth referred to as Model 4, or an unrestricted constant and a linear trend in the error correction term if $g_t = (1, t)'$, henceforth referred to as Model 5. Although higher order trend terms are certainly possible, such models are rarely used in practice, and we therefore restrict our attention to these five.

4.3 Individual tests for no error correction

In this section we show how the conditional ECM in (4.2.8) can be used as a basis for constructing cointegration tests. In particular, we propose two test statistics that are designed to test the null hypothesis that unit i is not error correcting versus the alternative that it is error correcting. We begin by considering the baseline case with known factors, and then we show how the testing can be carried out in the more realistic case when F_t is no longer observed.

4.3.1 Observed factors

Assumptions 4.1 to 4.4 are quite relaxed in the sense that even at the level of the individual unit, the models they imply are multivariate, which makes a full-blown system approach necessary. However, the purpose of this section is not to devise the most general test possible, but rather to derive tests that are simple, and easy to implement. This requires more assumptions.

Assumption 4.5 (i) $r = 1$, (ii) $c_i(L) = c_i$ for some constant $c_i < M$, (iii) $X_{i,t}$ is weakly exogenous for α_{1i} and β_i .

Remark: Assumption 4.5 implies that the r -dimensional conditional model in (4.2.8) can be written as a well-specified single equation, with no serial correlation and with the scalar coefficient

α_{1i} measuring the extent of the error correcting behavior in $Y_{i,t}$.

Under Assumption 4.5, and omitting any deterministic component for now, the conditional ECM in (4.2.8) reduces to

$$\Delta Y_{i,t} = \alpha_{1i} \beta_i' Z_{i,t-1}^+ + B_{11i}(L) \Delta Y_{i,t-1} + B_{12i}(L) \Delta X_{i,t} + B_{13i}(L) \Delta F_t + \varepsilon_{1 \cdot 2i,t}, \quad (4.3.1)$$

while the marginal models for $X_{i,t}$ and F_t become

$$\Delta X_{i,t} = B_{21i}(L) \Delta Y_{i,t-1} + B_{22i}(L) \Delta X_{i,t-1} + B_{23i}(L) \Delta F_{t-1} + \varepsilon_{2i,t}^*, \quad (4.3.2)$$

$$\Delta F_t = B_{33i}(L) \Delta F_{t-1} + \eta_t, \quad (4.3.3)$$

where the lag polynomials $B_{jli}(L)$ are obtained by simply collecting the appropriate terms from (4.2.8) to (4.2.10).

Assumptions 4.1 to 4.5 ensure that the following functional central limit theorem holds as $T \rightarrow \infty$

$$\frac{1}{\sqrt{T}} \sum_{t=1}^{\lfloor sT \rfloor} \begin{pmatrix} \varepsilon_{1 \cdot 2i,t} \\ \varepsilon_{2i,t}^* \\ \eta_t \end{pmatrix} \xrightarrow{w} B_i,$$

where $s \in [0, 1]$ and $B_i = (B_{1i}, B'_{2i}, B'_3)'$ is a $(1 + m + k)$ -dimensional vector Brownian motion, which can be partitioned as $B_i = (B_{1i}, B'_{2 \cdot i})'$ with $B_{2 \cdot i} = (B'_{2i}, B'_3)'$ having dimension $m + k$. The covariance matrix of B_i is given by

$$\Xi_i = \begin{pmatrix} \sigma_i^2 & 0 & 0 \\ 0 & \Sigma_{22i} + \lambda'_{2i} \lambda_{2i} & \lambda'_{2i} \\ 0 & \lambda_{2i} & I_k \end{pmatrix} = \text{cov}(B_i),$$

where $\sigma_i^2 = \Sigma_{11i} - \Sigma_{12i} \Sigma_{22i}^{-1} \Sigma_{21i}$. Thus, $B_i = \Xi_i^{\frac{1}{2}} W_i$, where $W_i = (W_{1i}, W'_{2i}, W'_3)'$ is a $(1 + m + k)$ -dimensional standard Brownian motion that is partitioned conformably with B_i . Furthermore, the long-run covariance matrix of $Z_{i,t}^+$ is given by

$$\Omega_i = \tilde{B}_i(1) \Xi_i \tilde{B}_i(1)' = \Omega_i^{\frac{1}{2}} (\Omega_i^{\frac{1}{2}})',$$

where the lag polynomial $\tilde{B}_i(L)$ is obtained from collecting the appropriate terms from (4.3.1) to (4.3.3) and $\Omega_i^{\frac{1}{2}} = \tilde{B}_i(1) \Xi_i^{\frac{1}{2}}$.

For later reference it is useful to consider the continuous time regression of W_{1i} , the first element of W_i , onto some vector X_i ,

$$W_{1i} = P_i(X_i)'X_i + Q_X W_{1i},$$

where

$$P_i(X_i) = \left(\int X_i X_i' \right)^{-1} \int X_i W_{1i} = V(X_i) p_i(X_i) \quad (4.3.4)$$

is the ordinary least squares (OLS) projection with $Q_X W_{1i}$ being the associated projection error. For example, if $X_i = 1$, then $P_i(X_i) = \int W_{1i}$ in which case $Q_1 W_{1i} = W_{1i} - \int W_{1i}$ is the demeaned version of W_{1i} .

As (4.3.1) makes clear, as long as F_t is observed, the problem of testing the null of no error correction is equivalent to testing

$$H_{0i} : \alpha_{1i} = 0$$

against

$$H_{1i} : \alpha_{1i} < 0.$$

The problem is that, unless one resorts to nonlinear techniques, this parameter is not easily estimated. One way to get around this is to assume that β_i is known, and to estimate α_{1i} using OLS. However, as shown by Boswijk (1994) and Zivot (2000), apart from the obvious drawback that β_i is almost never known in practice, tests based on a prespecified β_i are generally not similar and depend on nuisance parameters, even asymptotically.

As an alternative approach, note that (4.3.1) can be reparameterized as

$$\begin{aligned} \Delta Y_{i,t} &= \alpha_{1i} Y_{i,t-1} + \gamma'_{1i} X_{i,t-1} + \gamma'_{2i} F_{t-1} + B_{11i}(L) \Delta Y_{i,t-1} + B_{12i}(L) \Delta X_{i,t} \\ &+ B_{13i}(L) \Delta F_t + \varepsilon_{1 \cdot 2i,t}, \end{aligned} \quad (4.3.5)$$

where $\gamma'_{1i} = -\alpha_{1i} b'_{1i}$ and $\gamma'_{2i} = -\alpha_{1i} \lambda'_{1i}$. The advantage of rewriting (4.3.1) in this way is that because γ_{1i} and γ_{2i} are unrestricted, the cointegrating vector is implicitly estimated under the alternative hypothesis. Hence, as long as we are not interested in β_i , all the parameters of (4.3.5) can be consistently estimated by simple OLS, which in turn suggests the OLS estimator of α_{1i} as a natural candidate for constructing asymptotically similar tests of the null hypothesis of no error correction. In this section we propose two such tests, whose construction is described next.

One obvious candidate is the t -test. Suppose that the lag polynomial $B_{1ji}(L)$ is of order q_i , and let

$$W_{i,t} = (\Delta Y_{i,t-1}, \dots, \Delta Y_{i,t-q_i}, \Delta X'_{i,t}, \dots, \Delta X'_{i,t-q_i}, \Delta F'_t, \dots, \Delta F'_{t-q_i})'$$

denote the vector of stationary, first-differenced, regressors, while $V_{i,t}$ again denotes the vector of weakly exogenous non-stationary, level, variables, then (4.3.5) can be written as

$$\begin{aligned} \Delta Y_{i,t} &= \alpha_{1i} Y_{i,t-1} + \gamma'_i V_{i,t-1} + \Pi'_i W_{i,t} + \varepsilon_{1 \cdot 2i,t} \\ &= \alpha_{1i} Y_{i,t-1} + \Phi'_i S_{i,t} + \varepsilon_{1 \cdot 2i,t}, \end{aligned} \quad (4.3.6)$$

where $\Phi_i = (\gamma'_i, \Pi'_i)'$, $S_{i,t} = (V'_{i,t-1}, W'_{i,t})'$, $\gamma_i = (\gamma'_{1i}, \gamma'_{2i})'$ and Π_i is the vector stacking the coefficient vectors of the lag polynomials $B_{11i}(L)$, $B_{12i}(L)$ and $B_{13i}(L)$. This equation can in turn be written as

$$\Delta(Q_S Y_{i,t}) = \alpha_{1i}(Q_S Y_{i,t-1}) + Q_S \varepsilon_{1 \cdot 2i,t},$$

where again Q_S is the OLS projection error operator, with

$$Q_S Y_{i,t} = Y_{i,t} - \sum_{t=2}^T Y_{i,t-1} S'_{i,t} \left(\sum_{t=2}^T S_{i,t} S'_{i,t} \right)^{-1} S_{i,t}$$

being the residual from projecting $Y_{i,t}$ onto $S_{i,t}$.

In this notation, the OLS estimator of α_{1i} is given by

$$\hat{\alpha}_{1i} = \left(\sum_{t=2}^T (Q_S Y_{i,t-1})^2 \right)^{-1} \sum_{t=2}^T Q_S Y_{i,t-1} \Delta(Q_S Y_{i,t}),$$

whose estimated variance is given by

$$\text{var}(\hat{\alpha}_{1i}) = \hat{\sigma}_i^2 \left(\sum_{t=2}^T (Q_S Y_{i,t-1})^2 \right)^{-1},$$

where $\hat{\sigma}_i^2 = \frac{1}{T} \sum_{t=2}^T (\Delta(Q_S Y_{i,t}) - \hat{\alpha}_{1i}(Q_S Y_{i,t-1}))^2$. The t -statistic for testing H_{0i} can now be written as

$$\tau_{\hat{\alpha}_{1i}} = \frac{\hat{\alpha}_{1i}}{\sqrt{\text{var}(\hat{\alpha}_{1i})}}.$$

Another possibility is to follow Boswijk (1994), and to use a Wald statistic to test if α_{1i} and γ_i are jointly zero. In so doing, note that (4.3.6) can be rewritten as

$$\Delta Y_{i,t} = \delta'_{1i} Z_{i,t-1}^+ + \Pi'_i W_{i,t} + \varepsilon_{1 \cdot 2i,t},$$

where $\delta_{1i} = (\alpha_{1i}, \gamma'_i)'$, or in terms of projection residuals,

$$\Delta(Q_W Y_{i,t}) = \delta'_{1i} (Q_W Z_{i,t-1}^+) + Q_W \varepsilon_{1 \cdot 2i,t}.$$

The Wald statistic for testing the restriction that $\delta_{1i} = 0$ is given by

$$w_{\hat{\delta}_{1i}} = \hat{\delta}'_{1i} (\text{var}(\hat{\delta}_{1i}))^{-1} \hat{\delta}_{1i},$$

where

$$\hat{\delta}_{1i} = \left(\sum_{t=2}^T Q_W Z_{i,t-1}^+ (Q_W Z_{i,t-1}^+)' \right)^{-1} \sum_{t=2}^T Q_W Z_{i,t-1}^+ \Delta(Q_W Y_{i,t})$$

is the OLS estimator of δ_{1i} , and

$$\text{var}(\hat{\delta}_{1i}) = \hat{\sigma}_i^2 \left(\sum_{t=2}^T Q_W Z_{i,t-1}^+ (Q_W Z_{i,t-1}^+)' \right)^{-1}$$

is the associated variance.

The t -statistic $\tau_{\hat{\alpha}_{1i}}$ and the Wald statistic $w_{\hat{\delta}_{1i}}$ are the two test statistics considered in this chapter. Their limiting distributions under the no error correction null are given in the following theorem.

Theorem 4.2 *Under H_{0i} and Assumptions 4.1 to 4.5, as $T \rightarrow \infty$*

$$(a) \quad w_{\hat{\delta}_{1i}} \xrightarrow{w} D_{i,w} = p_i(W_i)' P_i(W_i),$$

$$(b) \quad \tau_{\hat{\alpha}_{1i}} \xrightarrow{w} D_{i,\tau}^\Omega = \frac{d_i}{\sqrt{D_i}},$$

where $p_i(\cdot)$ and $P_i(\cdot)$ are defined in (4.3.4),

$$\begin{aligned} D_i &= \sigma_i^2 \omega_{11 \cdot 2i}^{-2} V(U_i) \\ &+ \sigma_i^2 \omega_{11 \cdot 2i}^{-2} \omega_{11i} V(U_i) (\rho'_i (\Omega'_{22i})^{-1} P_i(W_{2 \cdot i}) + V(W_{2 \cdot i}) p_i(W_{2 \cdot i})' \Omega_{22i}^{-1} \rho_i) \\ &+ \sigma_i^2 \omega_{11 \cdot 2i}^{-2} \omega_{11i}^2 \rho'_i (\Omega'_{22i})^{-1} (V(W_{2 \cdot i}) + P_i(W_{2 \cdot i}) V(U_i) V(W_{2 \cdot i}) p_i(W_{2 \cdot i})') \Omega_{22i}^{-1} \rho_i, \\ d_i &= \sigma_i \omega_{11 \cdot 2i}^{-1} P_i(U_i) + \omega_{11 \cdot 2i}^{-1} \omega_{11i} \rho'_i (\Omega'_{22i})^{-1} P_i(W_{2 \cdot i}) (P_i(U_i) - 1), \end{aligned}$$

where $U_i = W_{1i} - \int W_{1i}W'_{2i} (\int W_{2i}W'_{2i})^{-1} W_{2i}$ with $W_{2i} = (W'_{2i}, W'_3)'$, and with ω_{112i} , ω_{11i} , Ω_{22i} and ρ_i depending on the parameters of $\Omega_i^{\frac{1}{2}}$, as defined in the appendix.

Proof: see Appendix 4.B.2

The asymptotic distribution of $\tau_{\hat{\alpha}_{1i}}$ simplifies substantially if $X_{i,t}$ is strongly exogenous.

Assumption 4.6 $X_{i,t}$ is strongly exogenous for α_{1i} and β_i .

This is shown in the following corollary.

Corollary 4.1 Under Assumption 4.6 and the conditions of Theorem 4.2, as $T \rightarrow \infty$,

$$\tau_{\hat{\alpha}_{1i}} \xrightarrow{w} D_{i,\tau} = \frac{P_i(U_i)}{\sqrt{V(U_i)}}.$$

Theorem 4.2 shows that the distribution of $w_{\hat{\delta}_{1i}}$ as $T \rightarrow \infty$ is nuisance parameter free and only depends on $m + k$, the number of non-stationary exogenous variables in the system. By contrast, the distribution of $\tau_{\hat{\alpha}_{1i}}$ depends on several nuisance parameters, and although these vanish under Assumption 4.6, strong exogeneity is quite restrictive. Fortunately, as Zivot (2000) points out relying on results obtained by for example Saikkonen (1991), there is a simple modification available that eliminates the nuisance parameters that are there under Assumption 4.5 (iii). The idea is to model these parameters by making the lag polynomial $B_{12i}(L)$ double-sided, as in

$$B_{12i}(L + L^{-1}) = \sum_{j=-\infty}^{\infty} B_{12ij}L^j,$$

where L^{-1} is the lead operator, which in turn requires augmenting (4.3.6) not only by the lags, but also by the leads of $\Delta X_{i,t}$. If the number of leads is large enough, then the asymptotic distribution of the resulting test statistic is given in Corollary 4.1.

In this sense, the results in Theorem 4.2 are basically the same as those provided by Banerjee et al. (1998) and Boswijk (1994) for the pure time series case. The proof is therefore very similar. The difference lies with the presence of F_t , which has two effects. One is that the number of unit roots increases from $1 + m$ to $1 + m + k$, which is reflected through W_3 in the asymptotic distribution of the test. The second effect is that the test statistics across units are no longer independent of each other, although the degree of the dependence between all pairs of units is the same.

In the presence of nonzero deterministic constant and trend terms, as in Models 2 to 5, the above theorem needs to be modified in order to obtain similar tests. This requires replacing U_i in (a) and

W_i in (b) by their appropriately detrended counterparts. Specifically, U_i and W_i should be demeaned in Model 1, and demeaned and detrended in Model 2. The t -test cannot be used in Models 4 and 5, and so for these models there is only the Wald test. In Model 4, W_i is replaced by $(W'_i, 1)'$, while in Model 5, W_i is replaced by $(W'_i, 1, s)'$, where s is the limiting trend function, see Boswijk (1994).

Furthermore, under the alternative hypothesis of cointegration, $\tau_{\hat{\alpha}_{1i}} \rightarrow -\infty$ whereas $w_{\hat{\delta}_{1i}} \rightarrow \infty$ as $T \rightarrow \infty$, suggesting that the tests are consistent. A proof of this is provided by Boswijk (1994).

4.3.2 Unobserved factors

So far we have assumed F_t to be observed, an assumption which is generally not true. To account for this, in a recent unit root paper Bai and Ng (2004b) propose using the method of principal components to estimate F_t , and then to use this estimate in place of F_t in the subsequent analysis. This approach has proven very fruitful, and has also been extended to the case of cointegration, see for example Bai et al. (2009), Banerjee and Carrion-i Silvestre (2006), Gengenbach et al. (2006) and Westerlund (2008). The problem with this approach is that, regardless of whether one considers unit roots or cointegration, the analysis must be carried out in steps, which means that the estimation error from one step is imported in subsequent steps.

As a response to this, Pesaran (2007) proposes a joint approach, which is based on using cross-sectional averages of the observed variables as proxies for the unobserved common factors. Apart from the advantage that it eliminates the need for a two-step estimation procedure, this approach fits very well with the parametric flavor of our conditional ECM, and it will therefore be used in this chapter.

Part (b) of Theorem 4.1 implies that $Z_{i,t}$ can be written as

$$Z_{i,t} = \Lambda_i F_t + E_{i,t},$$

where Λ_i is the $(1 + m) \times k$ matrix of factor loadings, and where $E_{i,t}$ is a vector representing the idiosyncratic component of $Z_{i,t}$. Denoting by \bar{Z}_t , $\bar{\Lambda}$ and \bar{E}_t the cross-sectional averages of $Z_{i,t}$, Λ_i and $E_{i,t}$, respectively, it is clear that

$$\bar{Z}_t = \bar{\Lambda} F_t + \bar{E}_t,$$

which, via Assumption 4.3 (iii) and the fact that $E_{i,t}$ is cross-sectionally independent, suggests that

F_t can be written as

$$F_t = (\bar{\Lambda}'\bar{\Lambda})^{-1}\bar{\Lambda}'\bar{Z}_t + (\bar{\Lambda}'\bar{\Lambda})^{-1}\bar{\Lambda}'\bar{E}_t = (\bar{\Lambda}'\bar{\Lambda})^{-1}\bar{\Lambda}'\bar{Z}_t + O_p\left(\frac{1}{\sqrt{N}}\right).$$

The implication is that the common factors can be approximated by the cross-sectional averages \bar{Z}_t , and that the resulting approximation error should become negligible as $N \rightarrow \infty$. Following this argument, we propose using \bar{Z}_t to approximate F_t . In so doing, it is convenient to let $\tilde{W}_{i,t}$, $\tilde{V}_{i,t}$ and $\tilde{Z}_{i,t}^+$ denote $W_{i,t}$, $V_{i,t}$ and $Z_{i,t}^+$, respectively, with \bar{Z}_t in place of F_t . Starting with (4.3.6) the approximate test regression can now be written as

$$\Delta Y_{i,t} = \alpha_{1i}Y_{i,t-1} + \Phi_i'\tilde{S}_{i,t} + \tilde{\varepsilon}_{1.2i,t},$$

or equivalently,

$$\Delta(Q_{\tilde{S}}Y_{i,t}) = \alpha_{1i}(Q_{\tilde{S}}Y_{i,t-1}) + Q_{\tilde{S}}\tilde{\varepsilon}_{1.2i,t},$$

where the error $\tilde{\varepsilon}_{1.2i,t}$ depends on the accuracy of the approximation. Nevertheless, by regressing $\Delta(Q_{\tilde{S}}Y_{i,t})$ on $Q_{\tilde{S}}Y_{i,t-1}$, we obtain another OLS estimator of α_{1i} , which we will henceforth denote by $\tilde{\alpha}_{1i}$. The associated t -statistic of H_{0i} can be written in an obvious notation as

$$\tau_{\tilde{\alpha}_{1i}} = \frac{\tilde{\alpha}_{1i}}{\sqrt{\text{var}(\tilde{\alpha}_{1i})}},$$

while the Wald statistic can be written as

$$w_{\tilde{\delta}_{1i}} = (\tilde{\delta}_{1i})'(\text{var}(\tilde{\delta}_{1i}))^{-1}\tilde{\delta}_{1i},$$

where $\tilde{\delta}_{1i}$ and $\text{var}(\tilde{\delta}_{1i})$ are defined just as in Section 4.3.1 but with $Q_{\tilde{W}}$ in place of Q_W .

Theorem 4.3 provides the limiting null distributions of these test statistics.

Theorem 4.3 *Under the conditions of Theorem 4.2, as $N, T \rightarrow \infty$,*

- (a) $w_{\tilde{\delta}_{1i}} \xrightarrow{w} D_{i,w},$
- (b) $\tau_{\tilde{\alpha}_{1i}} \xrightarrow{w} \tilde{D}_{i,\tau}^\Omega = \frac{\tilde{d}_i}{\sqrt{\tilde{D}_i}},$

where \tilde{d}_i and \tilde{D}_i are defined analogously to d_i and D_i but depending on the parameters of $\tilde{\Omega}_i^{\frac{1}{2}}$, as defined in the appendix.

Proof: see Appendix 4.B.3

Theorem 4.3 shows that the asymptotic distributions of $w_{\hat{\delta}_{1i}}$ is the same as that of $w_{\hat{\delta}_{1i}}$ provided in Theorem 4.2, which is based on observed factors. The limiting distribution of $\tau_{\hat{\alpha}_{1i}}$ is similar to that of $\tau_{\hat{\alpha}_{1i}}$ but depending on different nuisance parameter due to the approximation of F_t by \bar{Z}_t . The difference is that Theorem 4.2 only requires that $T \rightarrow \infty$. If F_t is not observed, we require $N \rightarrow \infty$ as well to ensure that \bar{Z}_t provides a sufficiently good approximation for F_t .

Similarly to the case of observed factors, if $X_{i,t}$ is strongly exogenous the asymptotic distribution of $\tau_{\hat{\alpha}_{1i}}$ simplifies and is the same as that of $\tau_{\hat{\alpha}_{1i}}$. This is shown in Corollary 4.2.

Corollary 4.2 *Under Assumption 4.6 and the conditions of Theorem 4.3, as $N, T \rightarrow \infty$,*

$$\tau_{\hat{\alpha}_{1i}} \xrightarrow{w} D_{i,\tau}.$$

4.3.3 Critical values

As in the simple case with cross-sectionally independent units, our tests are one-sided. The t -test is left-tailed, while the Wald test is right-tailed. The difference is that in our case the asymptotic test distribution, and hence also the simulation of the critical values, is complicated by the dependence across i . However, conditional on W_3 , the Brownian motion associated with F_t , the random variables $D_{1,w}, \dots, D_{N,w}$, are identically and independently distributed for all values of N . We say that $D_{1,w}, \dots, D_{N,w}$ form an exchangeable sequence, similar to for example Pesaran (2007) and Gregoir (2005). Thus, since $D_{i,w}$ is the same for all N , we can just as well set $N = 1$ in the simulations, a finding also confirmed by our results. The same argument applies to $D_{i,\tau}$. However, this is only valid for the limiting distribution of the t -test under strong exogeneity of $X_{i,t}$, or if an appropriate correction is employed to remove the nuisance parameter dependence. Otherwise, the individual test statistics are not identically distributed across i .

The simulated critical values at the 1%, 5% and 10% significance levels are reported in Table 4.1 for the t -test, and in Table 4.2 for the Wald test. These are based on making 1,000,000 draws from the limiting test distributions, with normal random walks of length $T = 1,000$. The results are reported for all five deterministic model specifications, and for $m = 1, \dots, 5$.

4.4 Panel tests for no error correction

In this section we build on the results of Section 4.3, and show how these can be used to construct pooled tests for the null of no error correction at the overall panel level. As an example, we will consider the t -statistic in the most simple case with known factors.

4.4.1 The tests

There are many ways in which one can combine a set of individual test statistics into a pooled test. The by far most common way is to follow Im *et al.* (2003) and to take the average, which for the t -statistic in case of known factors amounts to computing

$$\bar{\tau}_{\hat{\alpha}_1} = \frac{1}{N} \sum_{i=1}^N \tau_{\hat{\alpha}_{1i}}.$$

This is a test of the null of no error correction against the alternative that there is a non-vanishing fraction of error correcting units. Formally, the null and alternative hypotheses are formulated as

$$H_0 : \alpha_{1i} = 0 \text{ for all } i$$

against

$$H_1 : \alpha_{1i} < 0 \text{ for } i = 1, \dots, N_1 \text{ with } \frac{N_1}{N} \rightarrow \delta > 0$$

as $N_1, N \rightarrow \infty$. However, due to the dependence across i , in our case it is not possible to follow the usual practice in applying a central limit theorem to obtain a normal distribution for \sqrt{N} times $\bar{\tau}_{\hat{\alpha}_1}$.

One possibility is to look directly at the average. Following similar arguments as Pesaran (2007), because $D_{1,\tau}, \dots, D_{N,\tau}$ are identically and independently distributed given W_3 , a law of large numbers applies to the conditional average of these random variables. That is, we have that as $N \rightarrow \infty$

$$\bar{D}_\tau = \frac{1}{N} \sum_{i=1}^N D_{i,\tau} \xrightarrow{p} \mathbb{E}(D_\tau | W_3),$$

where the i index in the expectation has been suppressed because all $D_{i,\tau}$ have the same conditional expectation. Thus, unconditionally the average converges to some random distribution. However, unless $\tau_{\hat{\alpha}_{1i}}$ has finite moments for all N and T , this distribution is not necessarily the same as the one that applies to $\bar{\tau}_{\hat{\alpha}_1}$.

In order to get around this technical difficulty, we follow Pesaran (2007) and base our pooled test

on a truncated version of $\tau_{\hat{\alpha}_{1i}}$. Because this test has finite moments by construction, the associated cross-sectional average converges to the same asymptotic distribution as \bar{D}_τ .

The truncated statistic is defined as

$$\tau_{\hat{\alpha}_{1i}}^* = \begin{cases} K_l & \text{if } \tau_{\hat{\alpha}_{1i}} \leq K_l \\ \tau_{\hat{\alpha}_{1i}} & \text{if } K_l < \tau_{\hat{\alpha}_{1i}} < K_u \\ K_u & \text{if } \tau_{\hat{\alpha}_{1i}} \geq K_u \end{cases} ,$$

where the thresholds K_l and K_u are such that the probability of observing $\tau_{\hat{\alpha}_{1i}} \leq K_l$ and $\tau_{\hat{\alpha}_{1i}} \geq K_u$ is sufficiently small. In particular, by using the normal approximation of $\tau_{\hat{\alpha}_{1i}}$, $K_l = \mathbf{E}(D_\tau) - \Phi^{-1}(1 - \frac{\varepsilon}{2})\sqrt{\text{var}(D_\tau)}$ and $K_u = \mathbf{E}(D_\tau) + \Phi^{-1}(1 - \frac{\varepsilon}{2})\sqrt{\text{var}(D_\tau)}$, where $\varepsilon > 0$ is a small number, while Φ is the standard normal cumulative distribution function.

The corresponding truncated version of $\bar{\tau}_{\hat{\alpha}_1}$ is given by

$$\bar{\tau}_{\hat{\alpha}_1}^* = \frac{1}{N} \sum_{i=1}^N \tau_{\hat{\alpha}_{1i}}^* .$$

Making use of Theorem 4.2, it is not difficult to see that as $T \rightarrow \infty$

$$\bar{\tau}_{\hat{\alpha}_1}^* \xrightarrow{w} \bar{D}_\tau^* = \frac{1}{N} \sum_{i=1}^N D_{i,\tau}^* ,$$

where

$$D_{i,\tau}^* = \begin{cases} K_l & \text{if } D_{i,\tau} \leq K_l \\ D_{i,\tau} & \text{if } K_l < D_{i,\tau} < K_u \\ K_u & \text{if } D_{i,\tau} \geq K_u \end{cases} .$$

But all moments of $D_{i,\tau}^*$ exist, so by conditioning on W_3 , as $N \rightarrow \infty$

$$\bar{D}_\tau^* \xrightarrow{p} \mathbf{E}(D_\tau^* | W_3) ,$$

where

$$\begin{aligned} \mathbf{E}(D_\tau^* | W_3) &= K_l \cdot \text{Prob}(D_\tau \leq K_l | W_3) + K_u \cdot \text{Prob}(D_\tau \geq K_u | W_3) \\ &+ \mathbf{E}(D_\tau | W_3, K_l < D_\tau < K_u) \rightarrow \mathbf{E}(D_\tau | W_3) \end{aligned}$$

as $K_l, K_u \rightarrow \infty$, and so we get the same result as for \bar{D}_τ . This suggests that $\bar{\tau}_{\hat{\alpha}_1}^*$ can be used for

the test of H_0 versus H_1 . Another possibility is to use $\bar{w}_{\delta_1}^*$, the average of the truncated Wald test statistics.

4.4.2 Critical values

The above results show that if $K_l, K_u \rightarrow \infty$, $\bar{\tau}_{\hat{\alpha}_1}^*$ converges to a distribution that only depends on number of non-stationary variables in the system. With K_l and K_u finite, however, then there is not just this dependence, but also a dependence on the specific threshold values. Similarly, if N is finite, then there is also a dependence on the size of the cross-section. The generation of the critical values has to account for all these dependencies.

We begin by simulating values of $E(D_\tau)$ and $\text{var}(D_\tau)$ for all five deterministic model specifications, and for $m = 1, \dots, 5$. These are needed in order to compute K_l and K_u . Just as in Section 4.3.3 we make 1,000,000 draws from the limiting test distribution, with normal random walks of length $T = 1,000$. The results for the t -test are reported in Table 4.1, while the results for the Wald test are reported in Table 4.2.

The next step is to simulate N -tuples $D_{1,\tau}^*, \dots, D_{N,\tau}^*$ using $\varepsilon = \frac{1}{10^6}$, and the first-step moments to compute K_l and K_u . The average is then taken, which yields one simulated value of \bar{D}_τ^* . By repeating this exercise 10,000 times, we obtain the simulated distribution of \bar{D}_τ^* . The critical values at the 1%, 5% and 10% levels are reported in Table 4.3 for the t -test and in Table 4.4 for the Wald test, in which case \bar{D}_τ^* is replaced by \bar{D}_w^* , the average of the truncated Wald test distributions.

4.5 Monte Carlo simulations

In this section we report the findings of a small set of simulations. We do not intend to give a comprehensive account of all the merits and drawbacks of the tests, but rather we want to convey a rough idea of their relative performance, also when compared to some of the more conventional tests from the literature.

The data generating conditional ECM is given by

$$\Delta Y_{i,t} = \alpha_1 (Y_{i,t-1} - X_{i,t-1} - \iota_2' F_{t-1}) + \Delta X_{i,t} + B_{13i} \Delta F_t + \varepsilon_{1.2i,t},$$

while the marginal models for $X_{i,t}$ and F_t are generated as

$$\begin{aligned}\Delta X_{i,t} &= B_{23i}\Delta F_{t-1} + \varepsilon_{2i,t}^*, \\ \Delta F_t &= \eta_t,\end{aligned}$$

where the elements of B_{23i} and B_{13i} are drawn from $N(1, 1)$, while $\iota_2 = (1, 1)'$ is a two-dimensional vector of ones. Thus, in this setup $X_{i,t}$ is a scalar, while F_t is two-dimensional. For simplicity, we assume that there are no deterministic components in the data generating process, and that there is a common error correction parameter α_1 , which is equal to zero under the null hypothesis, and equal to -0.05 under the alternative.

The results are organized in four parts depending on whether there is any serial correlation present or not. If there is no serial correlation, then $\varepsilon_{1,2i,t}$, $\varepsilon_{2i,t}^*$ and η_t are drawn from the standard normal distribution, while if there is serial correlation, then one of these errors is specified as a first-order autoregressive (AR) process with standard normal innovations, and a common AR coefficient of magnitude 0.5, while the remaining two errors are again drawn from the standard normal distribution.

All experiments are based on generating 5,000 panels with N individual and $T + 50$ time series observations, where the first 50 observations for each series are discarded in order to attenuate the effect of the initial conditions, which are all set to zero.

For comparison, the error correction tests of Westerlund (2007) are also simulated. Two are based on the group mean, or between, principle and are denoted G_τ and G_ρ , while the corresponding panel, or within, type statistics are denoted P_τ and P_ρ . Analogous to $\bar{\tau}_{\hat{\alpha}_1}^*$, G_τ and P_τ are constructed as t -ratios, while G_ρ and P_ρ are coefficient type statistics.

The problem with these tests is that they are based on assuming cross-sectional independence, as explained earlier, and are therefore not expected to work in a setup as general as this one. Therefore, for better comparability, we follow the suggestion of Gengenbach et al. (2006), and run the tests on the defactored data. Specifically, we begin by estimating separately the common component of $X_{i,t}$ and $Y_{i,t}$ using the method of Bai and Ng (2004b), which involves applying the principal components method to the variables in their first differences. The estimated common component is then removed, and the defactored data are cumulated back to levels again. The number of factors are determined using the IC_1 information criterion of Bai and Ng (2002) with a maximum of five factors.

For the number of lags and leads to use in the conditional ECM, we used the Schwarz Bayesian information criterion, which facilitates a data dependent choice. Consistent with the results of Ng

and Perron (1995), the maximum number of lags and leads is permitted to grow with T at rate $4\left(\frac{T}{100}\right)^{2/9}$. The same rate is used for picking the bandwidth needed for constructing G_ρ and P_ρ . Also, for better comparability across all tests, we do not consider Models 4 and 5 when the deterministic constant and trend terms are restricted to the error correction term. All tests are performed at the 5% significance level, and all powers are adjusted for size.

The results for the case with no deterministic components are reported in Table 4.5. The first thing to note is the relative performance of the new t -tests, which is very good. This is especially true when the data are serially correlated, in which case there are only one other test with roughly the same performance as ours, G_τ . The overall best performance is obtained by using the individual $\tau_{\hat{\alpha}_{1i}}$ and $\tau_{\hat{\alpha}_{1i}}$ tests, which seem to maintain the nominal level very well in all cases considered. At the other end of the scale we have the $\bar{w}_{\hat{\delta}_1}$ test, which generally suffers from severe distortions, even if it is based on the true factors.

Pesavento (2004) reports some results for the original Wald test of Boswijk (1994), and find it to be oversized when the serial correlation is of the positive AR type considered here. The overall poor performance of the new Wald tests is therefore not very surprising. On the other hand, unreported results suggest that the relative performance of these tests is much improved if the serial correlation is of the negative MA type, which is also what Pesavento (2004) finds in her simulation study. In any case, the size distortions generally decrease substantially as T increases, which corroborates our asymptotic results.²

Among the different versions of the new tests considered, the best size accuracy is not surprisingly obtained by using the true factors. The tests based on using the cross-sectional averages of the observed data as proxies for the factors are, however, almost as accurate, and perform only slightly less well. Thus, the approximation seem to be effective even when N is as small as 10. The defactored versions of the tests of Westerlund (2007) also seem to perform quite well, which is in agreement with consistency of the principal components method, as shown by Bai and Ng (2004b). However, although improving in N , we also see that the size accuracy is basically unaffected by T , which is unexpected because theoretically the precision of the principal components estimator should get better as T grows.

Consider next the results reported in Table 4.5 for the power of the tests, which can be summarized as follows. Firstly, the power increases rapidly as T and N increase, which is presumably a reflection of the consistency of the tests. Secondly, the Westerlund (2007) tests generally suffer from poor

²One possibility here is to follow Palm, Smeekes, and Urbain (2007), and to use bootstrap methods to eliminate the size distortions of the Wald test.

power, especially when $\varepsilon_{1,2i,t}$ is serially correlated, in which case the power is only rarely in excess of the size. The G_ρ test suffers most, and can actually be less powerful than some of the individual tests. Thirdly, as expected, the power of the new tests is generally greatly improved by pooling. Similarly, the tests based on the true factors are generally more powerful than those based on the cross-sectional averages of the observed data.

The results for the models with a constant, and constant and trend reported in Tables 4.6 and 4.7, respectively, are very similar to those reported in Table 4.5. Nevertheless, there are still a few differences that are noteworthy. One difference is the magnitude of the size distortions, which has a slight tendency to increase as more deterministic components are added. Similarly, we see that inclusion of more deterministic components reduces the power of the tests, especially for the individual ones. Another difference is that the relative power of the Wald tests is generally much higher in Tables 4.6 and 4.7 than in Table 4.5.

We also examined the effects of a violation of the weak exogeneity assumption. We used the same data generating process as before but this time we allowed the equation for $\Delta X_{i,t}$ to be error correcting. The results, which are not reported but available from the corresponding author upon request, conforms well with our expectations. In particular, while the size of the tests is not effected, the power can be very low in cases when it is mainly $\Delta X_{i,t}$ that is error correcting. Thus, even though the tests continue to perform well in some setups, in general we need the weak exogeneity assumption to ensure that they work properly.³

The above results are all based on the truncated panel statistics. We carried out the same simulations for their non-truncated versions, and obtained identical results. In fact, the two types of statistics differ only for very small values of T , and are basically indistinguishable for $T > 20$. Thus, although little is gained in the present case, the truncation of the extreme test statistics seem to pay out when T is very small. This effect is particularly strong when the number parameters of the underlying ECM regressions is large.

4.6 Empirical Applications

In this section we present two empirical applications of the tests developed in this chapter. The first is concerned with the Fisher effect, while the second is concerned with the monetary exchange rate model.

³Zivot (2000) examines the performance of the time series tests of Banerjee et al. (1998) and Boswijk (1994), and reach the same conclusions.

4.6.1 The Fisher effect

There are very few theoretical economic relationships with as much intuitive appeal as the Fisher effect, which states that a one-time permanent shock in monetary variables has no long-run effect on the real economy. A simple implication of this theory is that changes in inflation should be reflected fully in subsequent movements of the nominal interest rate, thus leaving the real interest rate constant over time. Yet, oddly, for a theory so widely accepted, the postulated long-run relationship between inflation and nominal interest rates has proven extremely difficult to establish empirically. In fact, most studies are unable to reject the null hypothesis of no cointegration between inflation and nominal interest rates.

Westerlund (2008) argues that this lack of empirical support can be partly explained by the poor precision of the routinely applied time series approach, and that the use of panel data can produce more accurate tests. Consistent with this story, drawing upon a panel of 20 OECD countries between the first quarter of 1980 and the fourth quarter of 2004, the author shows that while the null hypothesis of no cointegration cannot be rejected at conventional significance levels when using data on individual countries, panel testing leads to a safe rejection. Low power in the tests is therefore one possible explanation for why cointegration has been so difficult to find.

Our findings suggest that there is an alternative interpretation of these results. Namely, that inflation and nominal interest rates are cross-sectionally correlated via the presence of non-stationary common factors, which then invalidates the use of conventional critical values.⁴ Thus, according to this view, it is the factors, and not a lack of power, that make the tests unable to reject the no cointegration null at the individual country level.⁵

In this section, we therefore apply our new tests to the same data to reevaluate the cointegration test results reported by Westerlund (2008). In so doing, we will assume that his unit root test results hold, and hence that the rates of inflation and nominal interest are non-stationary. Hence, in this application $Y_{i,t} = i_{i,t}$ and $X_{i,t} = \pi_{i,t}$, where $i_{i,t}$ is the nominal interest rate for country i in quarter t , while $\pi_{i,t}$ is inflation.

The tests are constructed in the same way as in Section 4.5, using the Schwarz Bayesian information criterion with the same maximum to determine the number of lags and leads. One difference in comparison to the simulations is that the common factors are no longer observed, which means

⁴Although the panel tests of Westerlund (2008) are immune to the presence of common factors, his time series tests are not. This means that the two sets of results are not really comparable in the sense that the observed non-rejections at the individual country level could be due to the factors.

⁵One rationale for these factors is that they represent in part oil price shocks and other unanticipated changes in inflation.

that we cannot evaluate the tests at the true factors. Therefore, as a feasible alternative, in this section we consider replacing the factors by their first differenced and cumulated principal components estimates, which are consistent even if the factors are non-stationary, see Bai and Ng (2004b). In agreement with the so-called full Fisher effect, the estimation is carried out while imposing a unit slope coefficient on inflation. That is, the factors are estimated from the real interest rate, $i_{i,t} - \pi_{i,t}$, which is consistent with the idea of the existence of a world real interest rate, see for example Lee (2002).

The principal components method is implemented as described in Section 4.5, but with the number of factors restricted to two, which ensures that the rank condition in Assumption 4.3 (iii) is fulfilled. As in the simulations, the defactored versions of the error correction tests of Westerlund (2007) are also considered. We focus on the results for Model 2 with an unrestricted constant, but include the results for Model 3 with both constant and trend for comparison.

The results reported in Table 4.8 suggest that there is strong evidence against the no cointegration null, even at the individual country level, which goes against the power argument of Westerlund (2008). Indeed, looking at the baseline specification with no trend, we end up rejecting the null for 13 out of the 20 countries when using the $\tau_{\hat{\alpha}_{1i}}$ test, and for 11 countries when using the $w_{\hat{\delta}_{1i}}$ test. Similarly, the pooled tests are way out in the critical region and lead to a safe rejection, even at the conservative 1% level. In other words, there is not much evidence against the Fisher effect. This conclusion is not altered by the inclusion of a linear trend.

In fact, the no cointegration null is rejected even when the factors are estimated with the slope on inflation fixed at unity. Specifically, although weaker at the individual level, the evidence at the overall panel level is still strong. Thus, we also have some evidence of the full Fisher effect.

To formally test for the presence of unit roots in the estimated factors, we follow the recommendation of Bai and Ng (2004b) and use the augmented Dickey and Fuller (1979) test, ADF henceforth.⁶ The estimated first order AR coefficient for the two factors are 0.81 and 0.89, indicating that there is considerable persistency in the factors. This evidence is reinforced by the associated ADF test values, -1.69 and -1.79 , respectively, which lead to an acceptance of the unit root null for both factors. Thus, if these factors are to be interpreted as emanating from the world real interest rate, then this rate must be non-stationary.

The lesson we draw from these results is that a failure to reject the null of no cointegration at the individual country level need not be taken as an indication of low power, as the possibility remains

⁶The test allow for an intercept and the lag orders are determined using the Schwarz Bayesian information criterion.

that it can be due to the presence of non-stationary common factors.

4.6.2 The monetary exchange rate model

In this section we take a closer look at the monetary exchange rate model, which postulates a strong link between the nominal exchange rate and a set of monetary fundamentals. The by far most scrutinized proposition being that the nominal exchange rate between the domestic and the foreign reference country, usually the United States, should cointegrate with the relative money supply and relative output of these countries.

However, as with the Fisher effect, despite its strong theoretical appeal, the empirical success of the monetary model has been rather limited, to say the least. Westerlund (2008), Mark and Sul (2001) and Rapach and Wohar (2004) for example argue that this is due to low power. They then proceed to show that the use of panel data leads to a much more favorable picture, with strong evidence of cointegration at the aggregate panel level. Therefore, since the countries appear to be cointegrated, the authors proceed to estimate the cointegration vector.

The problem is that since all variables are measured relative to the United States, this means that the common factors are there by construction. Furthermore, both money supply and output are generally believed to possess unit roots, even for the United States, such that the common factors must be non-stationary as well. The potential consequences of unattended non-stationary factors on residual-based panel cointegration tests have been studied by Banerjee et al. (2004) and Gengenbach et al. (2006). The effects may lead to size distortions in small samples or even divergence in large samples. While Mark and Sul (2001) employ a block bootstrap to correct for some weak cross section dependence among the error term. It is not clear whether their test can correct for strong cross sectional dependence induced by non-stationary common factors. Rapach and Wohar (2004) only allow for cross section dependence in form of a common time effect.

In this section we revisit the results of Mark and Sul (2001) and Rapach and Wohar (2004). The data are taken directly from Mark and Sul (2001), and cover 18 countries between the first quarter of 1973 and the first quarter of 1997. Thus, in this application, $Y_{i,t} = e_{i,t}$ and

$$X_{i,t} = \begin{pmatrix} m_t^* - m_{i,t} \\ y_t^* - y_{i,t} \end{pmatrix},$$

where $e_{i,t}$, $m_{i,t}$ and $y_{i,t}$ are the logarithm of the nominal exchange rate, money supply and real income for country i at quarter t , respectively. Asterisks denote the United States.

The average-based tests are computed in the same ways as before, but now we consider two new versions of the factor-based tests. The first is based on using m_t^* and y_t^* as observed factors, which is very interesting in the sense that it provides an example of the scenario considered in Section 4.3.1. The second version is based on pre-specifying the cointegrating relationship as in Mark and Sul (2001). In particular, it is assumed that the relationship can be written as

$$\beta' Z_{i,t} = \begin{pmatrix} 1, & -1, & 1 \end{pmatrix} \begin{pmatrix} Y_{i,t} \\ X_{i,t} \end{pmatrix} = e_{i,t} - (m_t^* - m_{i,t}) + (y_t^* - y_{i,t}),$$

which imposes monetary neutrality and a unit negative income elasticity.⁷ Three factors are estimated from this relationship, which again ensures that Assumption 4.6 is satisfied. Once again we focus on Model 2 with an unrestricted constant as the deterministic component. For simplicity, in this section we drop the Westerlund (2007) tests.

The results are reported in Table 4.9. The first thing to note is that for the first 11 countries there is almost no evidence of cointegration at all, except possibly for Belgium, where we count four rejections at the 5% level. The pooled tests are generally much more significant, especially the Wald tests, displaying evidence of cointegration for all five panels. Just as before the results show almost no variation at all depending on whether the trend is included or not.

These two sets of results suggest that the evidence at the aggregate panel level could very well be due to only a few cointegrated countries. Indeed, a closer look at the different panel members reveals that the significance at the aggregate panel level is mainly due to three individually cointegrated countries, Italy, Spain and Korea. Although these differences could of course also be due to the relatively low power of the individual tests, they nevertheless show that one should take caution in interpreting test results at the aggregate panel level. Indeed, based on the results reported here it seems very hazardous, and borderline erroneous, to treat all five panels as cointegrated, and to proceed with the analysis as if all members are individually cointegrated.

When we compare the results from across the different tests, in agreement with our simulations, we see that the average-based Wald test leads to most rejections. As a final piece of evidence, Table 4.10 reports some summary statistics for the estimated factors. As in the Fisher application, we see that the estimated AR coefficients are very close to one, indicating the presence of unit roots, which is again supported by the ADF test results.

⁷In order to avoid the problem with nuisance parameter dependency discussed in Section 4.3.1, the cointegrating relationship is only pre-specified for the purpose of estimating the factors. In other words, although restricted in the factor estimation, in the implementation of the error correction tests the cointegrating relationship is still unrestricted.

4.7 Conclusions

In this chapter we consider the issue of testing for cointegration in a panel data model with non-stationary common factors. We begin by showing that the model admits to an ECM representation, a result that is then used for developing two new cointegration tests based on the significance of the error correction term.

It is shown that under the null of no error correction the asymptotic distributions of the tests are free of nuisance parameters, and that they only depend on the number of non-stationary variables in the system. However, the individual tests are not independent along the cross-sectional dimension, which makes pooling difficult. Nonetheless, the cross-sectional averages of these tests are shown to converge to well-defined distributions. These results hold regardless of whether the factors are treated as known or if they are estimated using the averages of the observed data. Some simulation evidence is also provided showing that the tests behave quite well in small samples.

A number of concluding remarks can be made. Firstly, the assumption of weak exogeneity of the regressors in the ECM is crucial for correct interpretation of the tests. This assumption is clearly a weakness in comparison to the residual-based test approach, in which the regressors can be fully endogenous by means of a non-parametric correction. However, it should be pointed out that in principle there is nothing that precludes the use of a similar correction in the current setup. An alternative approach would be to pre-test the validity of the weak exogeneity assumption using panel extensions of the Lagrange multiplier tests proposed by Boswijk and Urbain (1997).

Secondly, the simulations show that the new tests can still be distorted in some cases when the factors are treated as unknown. One possibility towards this end would be to follow Palm et al. (2007), and to consider bootstrap versions of our tests, which are expected to have better size properties in small samples.

Finally, a crucial assumption is that of a single cointegrating vector under the alternative. This is obviously an important limitation of our tests that is shared with most existing residual-based tests. When the dimension of the cointegrating space is unknown it is probably best to analyze the data using system-based approaches, see for example Larsson, Lyhagen, and Lötgren (2001).

4.A Tables

Table 4.1: Critical values and moments for the individual t -tests.

Model	m	Critical values			Moments	
		10%	5%	1%	$E(D_\tau)$	$\text{var}(D_\tau)$
1	1	-2.985	-3.315	-3.932	-1.709	1.069
	2	-3.484	-3.819	-4.434	-2.212	1.044
	3	-3.883	-4.219	-4.848	-2.617	1.026
	4	-4.233	-4.570	-5.191	-2.965	1.020
	5	-4.538	-4.876	-5.503	-3.272	1.012
2	1	-3.426	-3.744	-4.339	-2.250	0.884
	2	-3.845	-4.168	-4.775	-2.644	0.915
	3	-4.199	-4.528	-5.138	-2.985	0.931
	4	-4.512	-4.841	-5.454	-3.287	0.943
	5	-4.792	-5.123	-5.747	-3.564	0.947
3	1	-3.814	-4.122	-4.697	-2.704	0.779
	2	-4.175	-4.488	-5.078	-3.024	0.837
	3	-4.494	-4.815	-5.411	-3.316	0.872
	4	-4.780	-5.103	-5.703	-3.589	0.892
	5	-5.043	-5.370	-5.973	-3.841	0.904

Notes: Model 1 refers to the specification with no deterministic component, while Models 2 and 3 refer to the specifications with an unrestricted constant, and unrestricted constant and trend, respectively. The value m refers to the number of regressors contained in $X_{i,t}$.

Table 4.2: Critical values and moments for the individual Wald tests.

Model	m	Critical values			Moments	
		10%	5%	1%	$E(D_w)$	$\text{var}(D_w)$
1	1	12.209	14.291	18.726	6.979	15.188
	2	17.399	19.839	24.913	10.937	23.438
	3	22.344	25.010	30.634	14.872	31.381
	4	27.108	30.061	36.132	18.785	39.317
	5	31.795	34.966	41.435	22.709	47.043
2	1	14.821	17.081	21.870	8.944	19.467
	2	19.870	22.460	27.817	12.886	27.601
	3	24.750	27.571	33.400	16.833	35.554
	4	29.484	32.542	38.789	20.756	43.392
	5	34.076	37.329	43.941	24.639	50.867
3	1	17.525	19.940	24.973	11.091	23.396
	2	22.424	25.113	30.674	14.988	31.266
	3	27.190	30.127	36.200	18.891	39.200
	4	31.840	34.992	41.404	22.767	46.941
	5	36.389	39.768	46.581	26.639	54.563
4	1	15.769	18.012	22.789	9.964	18.782
	2	20.781	23.337	28.680	13.898	26.598
	3	25.629	28.422	34.305	17.824	34.284
	4	30.368	33.430	39.625	21.756	42.014
	5	34.995	38.236	44.840	25.648	49.671
5	1	18.412	20.800	25.830	12.093	22.321
	2	23.297	25.968	31.550	15.982	30.128
	3	28.084	31.016	37.034	19.888	37.981
	4	32.708	35.839	42.256	23.757	45.521
	5	37.293	40.612	47.334	27.628	53.097

Notes: Models 4 and 5 refer to the specifications with a constant, and constant and trend in the error correction term, respectively. See Table 4.1 for an explanation of the remaining features of the table.

Table 4.3: Critical values for the panel t -tests.

N	m = 1			m = 2			m = 3			m = 4			m = 5		
	10%	5%	1%	10%	5%	1%	10%	5%	1%	10%	5%	1%	10%	5%	1%
10	-2.181	-2.300	-2.519	-2.655	-2.771	-2.976	-3.060	-3.192	-3.394	-3.400	-3.519	-3.745	-3.688	-3.812	-4.013
15	-2.131	-2.232	-2.417	-2.590	-2.687	-2.866	-2.988	-3.085	-3.275	-3.322	-3.418	-3.597	-3.628	-3.725	-3.904
20	-2.103	-2.200	-2.380	-2.552	-2.644	-2.820	-2.941	-3.024	-3.182	-3.281	-3.368	-3.534	-3.583	-3.677	-3.832
25	-2.067	-2.163	-2.318	-2.531	-2.617	-2.779	-2.924	-3.006	-3.148	-3.254	-3.332	-3.472	-3.557	-3.636	-3.789
30	-2.048	-2.133	-2.287	-2.513	-2.588	-2.733	-2.899	-2.979	-3.114	-3.239	-3.305	-3.448	-3.536	-3.607	-3.744
40	-2.035	-2.121	-2.267	-2.493	-2.566	-2.706	-2.871	-2.947	-3.078	-3.208	-3.271	-3.398	-3.509	-3.574	-3.701
50	-2.019	-2.094	-2.234	-2.476	-2.545	-2.669	-2.859	-2.921	-3.048	-3.187	-3.254	-3.371	-3.491	-3.556	-3.672
70	-2.012	-2.087	-2.212	-2.457	-2.520	-2.627	-2.838	-2.897	-3.004	-3.169	-3.222	-3.332	-3.470	-3.525	-3.623
100	-1.995	-2.065	-2.190	-2.440	-2.505	-2.616	-2.822	-2.876	-2.975	-3.152	-3.204	-3.297	-3.451	-3.503	-3.594
							Model 2								
10	-2.658	-2.772	-2.971	-3.056	-3.167	-3.377	-3.393	-3.513	-3.728	-3.692	-3.804	-4.010	-3.966	-4.082	-4.284
15	-2.603	-2.698	-2.866	-2.993	-3.083	-3.265	-3.325	-3.420	-3.599	-3.624	-3.714	-3.887	-3.897	-3.997	-4.179
20	-2.568	-2.653	-2.796	-2.948	-3.033	-3.196	-3.283	-3.370	-3.525	-3.588	-3.670	-3.820	-3.862	-3.941	-4.096
25	-2.544	-2.623	-2.773	-2.925	-3.000	-3.139	-3.258	-3.331	-3.468	-3.560	-3.643	-3.784	-3.832	-3.906	-4.066
30	-2.530	-2.601	-2.735	-2.909	-2.981	-3.120	-3.238	-3.305	-3.438	-3.538	-3.607	-3.746	-3.810	-3.882	-4.008
40	-2.504	-2.574	-2.694	-2.885	-2.949	-3.071	-3.221	-3.285	-3.404	-3.515	-3.579	-3.702	-3.785	-3.848	-3.956
50	-2.492	-2.554	-2.672	-2.863	-2.924	-3.032	-3.200	-3.259	-3.374	-3.496	-3.551	-3.658	-3.774	-3.830	-3.925
70	-2.477	-2.536	-2.650	-2.848	-2.903	-3.016	-3.181	-3.234	-3.332	-3.475	-3.523	-3.620	-3.751	-3.799	-3.894
100	-2.458	-2.517	-2.611	-2.835	-2.886	-2.983	-3.162	-3.213	-3.301	-3.459	-3.507	-3.584	-3.729	-3.776	-3.862
							Model 3								
10	-3.085	-3.188	-3.382	-3.410	-3.514	-3.707	-3.709	-3.820	-4.021	-3.983	-4.098	-4.299	-4.231	-4.347	-4.555
15	-3.026	-3.112	-3.276	-3.351	-3.441	-3.616	-3.644	-3.737	-3.904	-3.921	-4.009	-4.185	-4.173	-4.263	-4.428
20	-2.980	-3.060	-3.212	-3.306	-3.389	-3.531	-3.595	-3.677	-3.827	-3.878	-3.955	-4.108	-4.130	-4.208	-4.344
25	-2.965	-3.041	-3.189	-3.286	-3.361	-3.490	-3.577	-3.657	-3.784	-3.853	-3.927	-4.064	-4.101	-4.172	-4.299
30	-2.948	-3.013	-3.130	-3.269	-3.337	-3.460	-3.563	-3.632	-3.756	-3.831	-3.898	-4.025	-4.083	-4.147	-4.284
40	-2.921	-2.982	-3.095	-3.244	-3.306	-3.420	-3.535	-3.594	-3.706	-3.805	-3.865	-3.972	-4.053	-4.113	-4.223
50	-2.906	-2.964	-3.067	-3.227	-3.282	-3.395	-3.519	-3.576	-3.681	-3.785	-3.845	-3.953	-4.037	-4.091	-4.202
70	-2.892	-2.946	-3.045	-3.210	-3.258	-3.355	-3.500	-3.548	-3.645	-3.764	-3.817	-3.906	-4.015	-4.063	-4.148
100	-2.875	-2.925	-3.010	-3.193	-3.237	-3.319	-3.485	-3.526	-3.606	-3.749	-3.793	-3.881	-3.996	-4.040	-4.119

Notes: See Table 4.1 for an explanation of the various features of this table.

Table 4.4: Critical values for the pooled Wald tests.

N	m = 1			m = 2			m = 3			m = 4			m = 5		
	10%	5%	1%	10%	5%	1%	10%	5%	1%	10%	5%	1%	10%	5%	1%
	Model 1														
10	8.724	9.296	10.478	13.053	13.689	14.974	17.246	17.988	19.383	21.394	22.227	23.772	25.514	26.366	28.032
15	8.466	8.927	9.785	12.710	13.244	14.282	16.895	17.511	18.625	20.965	21.708	23.022	25.138	25.843	27.185
20	8.311	8.709	9.419	12.521	13.016	13.882	16.671	17.234	18.186	20.728	21.327	22.422	24.829	25.430	26.690
25	8.209	8.570	9.275	12.403	12.858	13.681	16.525	16.979	17.872	20.591	21.114	22.067	24.627	25.218	26.224
30	8.164	8.505	9.211	12.304	12.698	13.571	16.406	16.855	17.690	20.473	21.013	21.908	24.488	25.030	26.023
40	8.087	8.389	8.947	12.163	12.523	13.149	16.219	16.649	17.425	20.293	20.743	21.635	24.265	24.731	25.578
50	7.989	8.307	8.877	12.097	12.430	13.068	16.152	16.528	17.258	20.187	20.627	21.347	24.168	24.595	25.431
70	7.921	8.207	8.719	11.993	12.310	12.913	16.036	16.363	16.975	20.032	20.391	21.065	24.032	24.409	25.102
100	7.883	8.135	8.593	11.911	12.197	12.696	15.913	16.219	16.788	19.867	20.205	20.787	23.891	24.227	24.790
	Model 2														
10	10.881	11.499	12.781	15.158	15.862	17.156	19.386	20.153	21.721	23.543	24.373	26.042	27.693	28.571	30.120
15	10.564	11.051	12.045	14.801	15.344	16.447	18.980	19.603	20.786	23.075	23.722	25.053	27.128	27.853	29.419
20	10.438	10.875	11.728	14.587	15.137	16.077	18.695	19.266	20.364	22.768	23.332	24.393	26.823	27.540	28.805
25	10.275	10.667	11.487	14.394	14.871	15.684	18.506	19.004	20.033	22.576	23.117	24.145	26.644	27.231	28.323
30	10.212	10.584	11.262	14.348	14.767	15.499	18.433	18.897	19.819	22.432	22.910	23.860	26.502	27.063	28.126
40	10.103	10.437	11.036	14.194	14.530	15.248	18.247	18.648	19.446	22.307	22.741	23.542	26.240	26.698	27.550
50	10.042	10.364	10.940	14.083	14.423	15.042	18.133	18.526	19.200	22.156	22.564	23.359	26.136	26.580	27.380
70	9.897	10.183	10.736	14.008	14.322	14.901	17.980	18.332	19.010	22.010	22.363	22.998	25.968	26.355	27.020
100	9.845	10.096	10.585	13.866	14.150	14.727	17.910	18.206	18.755	21.898	22.213	22.813	25.794	26.142	26.754
	Model 3														
10	13.240	13.882	15.123	17.414	18.194	19.572	21.541	22.401	24.024	25.675	26.537	28.143	29.711	30.684	32.492
15	12.837	13.403	14.406	17.002	17.607	18.702	21.051	21.678	22.975	25.144	25.879	27.175	29.221	30.010	31.559
20	12.608	13.072	13.969	16.722	17.273	18.381	20.798	21.358	22.425	24.943	25.624	26.776	28.863	29.572	30.719
25	12.461	12.903	13.748	16.568	17.023	17.870	20.658	21.168	22.180	24.700	25.240	26.273	28.699	29.277	30.319
30	12.398	12.793	13.514	16.500	16.929	17.811	20.499	20.998	21.938	24.469	25.007	25.956	28.509	29.090	30.142
40	12.254	12.585	13.239	16.316	16.674	17.393	20.372	20.804	21.575	24.347	24.806	25.699	28.367	28.875	29.820
50	12.188	12.514	13.124	16.229	16.585	17.313	20.231	20.617	21.389	24.190	24.626	25.453	28.173	28.620	29.488
70	12.099	12.389	12.950	16.096	16.429	17.017	20.103	20.456	21.109	24.035	24.393	25.102	27.987	28.421	29.132
100	11.977	12.243	12.721	15.993	16.268	16.778	19.960	20.265	20.876	23.908	24.229	24.895	27.837	28.203	28.849

Continued overleaf

Table 4.4: Continued.

N	m = 1			m = 2			m = 3			m = 4			m = 5		
	10%	5%	1%	10%	5%	1%	10%	5%	1%	10%	5%	1%	10%	5%	1%
	Model 4														
10	11.913	12.551	13.708	16.139	16.817	18.132	20.376	21.180	22.669	24.463	25.285	26.872	28.600	29.452	31.153
15	11.541	12.033	12.903	15.822	16.386	17.485	19.900	20.509	21.757	24.061	24.685	26.021	28.135	28.969	30.308
20	11.386	11.789	12.690	15.554	16.047	17.081	19.640	20.200	21.258	23.734	24.360	25.445	27.774	28.407	29.622
25	11.294	11.689	12.471	15.404	15.858	16.751	19.497	19.989	20.886	23.544	24.079	25.155	27.629	28.202	29.327
30	11.215	11.556	12.241	15.302	15.715	16.543	19.373	19.840	20.720	23.449	23.932	24.891	27.490	28.024	28.947
40	11.082	11.403	12.080	15.218	15.591	16.334	19.227	19.605	20.445	23.257	23.714	24.520	27.278	27.748	28.660
50	11.014	11.347	11.861	15.098	15.465	16.150	19.124	19.499	20.163	23.163	23.566	24.348	27.135	27.565	28.413
70	10.944	11.223	11.755	14.955	15.264	15.857	18.967	19.299	19.932	22.979	23.323	24.069	26.975	27.361	28.086
100	10.880	11.137	11.633	14.886	15.175	15.696	18.874	19.179	19.719	22.848	23.208	23.781	26.810	27.139	27.771
	Model 5														
10	14.166	14.818	16.222	18.334	19.047	20.452	22.525	23.277	24.967	26.616	27.443	29.169	30.711	31.665	33.368
15	13.813	14.310	15.340	17.970	18.555	19.613	22.095	22.788	24.037	26.075	26.827	28.317	30.177	30.902	32.330
20	13.600	14.088	14.941	17.695	18.183	19.263	21.788	22.358	23.424	25.801	26.437	27.596	29.846	30.514	31.755
25	13.485	13.904	14.624	17.547	18.006	18.937	21.576	22.145	23.133	25.658	26.177	27.292	29.657	30.245	31.412
30	13.377	13.766	14.479	17.458	17.902	18.653	21.507	21.972	22.871	25.502	26.007	26.931	29.563	30.100	31.168
40	13.269	13.620	14.257	17.321	17.699	18.419	21.305	21.729	22.515	25.334	25.758	26.561	29.323	29.811	30.696
50	13.203	13.529	14.155	17.199	17.534	18.167	21.193	21.583	22.284	25.212	25.641	26.503	29.156	29.568	30.452
70	13.068	13.347	13.893	17.082	17.411	18.020	21.058	21.392	22.033	25.047	25.411	26.104	28.985	29.398	30.127
100	12.989	13.251	13.744	16.981	17.276	17.793	20.939	21.267	21.884	24.903	25.226	25.848	28.798	29.140	29.777

Notes: See Tables 4.1 and 4.2 for an explanation of the various features of this table.

Table 4.5: Size and size-adjusted power in Model 1 with no deterministic components.

AR case	N	T	This study								Westerlund (2007)			
			$\bar{\tau}_{\alpha_{1t}}$	$\bar{\tau}_{\alpha_{1t}}$	$w_{\delta_{1t}}^{\hat{\delta}}$	$w_{\delta_{1t}}^{\hat{\delta}}$	$\bar{\tau}_{\alpha_1}^*$	$\bar{\tau}_{\alpha_1}^*$	$\bar{w}_{\delta_1}^*$	$\bar{w}_{\delta_1}^*$	G_τ	G_ρ	P_τ	P_ρ
1	10	100	5.7	6.1	7.5	7.6	5.4	7.1	12.0	15.5	9.5	5.8	11.8	17.0
	10	200	5.3	5.5	6.1	6.3	5.5	6.1	8.0	11.2	7.9	6.1	11.1	16.1
	20	100	5.6	5.8	7.6	7.6	4.4	5.8	14.3	19.5	7.6	4.4	8.0	11.1
	20	200	5.4	5.6	6.2	6.4	4.9	7.1	9.5	14.9	5.8	4.0	8.1	11.2
	10	100	6.1	6.1	7.9	8.4	6.7	6.6	13.3	18.1	2.1	0.9	2.9	3.4
	10	200	5.6	5.1	6.3	6.5	5.9	5.1	8.3	11.7	2.1	1.2	2.5	3.0
2	20	100	6.1	6.0	8.0	8.4	5.7	6.3	16.9	22.8	2.2	0.9	2.6	3.7
	20	200	5.6	5.3	6.4	6.7	5.5	5.2	10.5	16.2	1.4	0.7	2.7	3.1
	10	100	5.1	5.5	23.3	18.6	0.7	1.9	68.1	52.1	4.1	1.7	2.3	1.9
	10	200	5.0	4.8	12.1	9.3	2.7	2.7	28.0	19.2	5.2	2.0	7.3	7.9
	20	100	5.1	5.5	23.7	18.3	0.3	0.6	86.4	69.5	4.0	0.7	1.8	1.6
	20	200	5.0	4.8	12.0	9.1	1.5	2.0	37.7	23.3	4.2	0.6	6.2	5.2
3	10	100	7.0	6.4	8.7	12.3	8.4	6.3	16.0	31.1	9.8	6.6	12.0	17.5
	10	200	5.9	5.3	6.5	10.0	6.7	5.3	9.5	22.6	8.1	6.4	11.0	16.3
	20	100	6.8	6.0	8.9	12.8	7.9	5.5	22.1	38.9	7.9	4.4	8.1	11.5
	20	200	5.9	5.5	6.7	10.6	6.7	5.0	12.6	30.9	6.0	4.2	7.6	11.0
	10	100	5.1	5.5	23.3	18.6	0.7	1.9	68.1	52.1	4.1	1.7	2.3	1.9
	10	200	5.0	4.8	12.1	9.3	2.7	2.7	28.0	19.2	5.2	2.0	7.3	7.9
4	20	100	5.1	5.5	23.7	18.3	0.3	0.6	86.4	69.5	4.0	0.7	1.8	1.6
	20	200	5.0	4.8	12.0	9.1	1.5	2.0	37.7	23.3	4.2	0.6	6.2	5.2
	10	100	7.0	6.4	8.7	12.3	8.4	6.3	16.0	31.1	9.8	6.6	12.0	17.5
	10	200	5.9	5.3	6.5	10.0	6.7	5.3	9.5	22.6	8.1	6.4	11.0	16.3
	20	100	6.8	6.0	8.9	12.8	7.9	5.5	22.1	38.9	7.9	4.4	8.1	11.5
	20	200	5.9	5.5	6.7	10.6	6.7	5.0	12.6	30.9	6.0	4.2	7.6	11.0

Continued overleaf

Table 4.5: Continued.

AR case	N	T	This study						Westerlund (2007)					
			τ_{α_1}	$\bar{\tau}_{\alpha_1}$	w_{δ_1}	$w_{\delta_1}^*$	$\bar{\tau}_{\alpha_1}^*$	$\bar{\tau}_{\alpha_1}^*$	$\bar{w}_{\delta_1}^*$	$\bar{w}_{\delta_1}^*$	G_τ	G_ρ	P_τ	P_ρ
1	10	100	16.4	13.0	28.4	13.1	77.6	60.8	85.9	39.0	28.7	20.3	34.3	45.9
	10	200	49.7	35.6	58.2	30.3	100.0	99.5	99.9	88.4	47.7	45.7	50.9	62.3
	20	100	16.8	13.6	28.4	13.8	95.6	81.8	97.1	50.8	54.5	41.6	55.2	62.8
	20	200	49.9	36.0	58.5	31.2	100.0	100.0	100.0	97.1	76.3	74.1	69.0	74.1
	10	100	15.8	13.1	28.0	12.4	74.7	58.0	85.3	37.5	4.1	2.5	5.7	7.2
	10	200	48.8	33.9	57.7	29.5	100.0	98.2	99.9	86.9	6.6	5.4	7.3	11.1
2	10	100	16.0	13.7	27.9	13.2	94.2	76.6	96.4	48.9	20.0	12.7	21.9	28.6
	20	100	49.2	34.5	58.1	30.3	100.0	99.8	100.0	96.9	34.1	32.2	33.1	41.5
	10	100	12.0	9.9	6.8	3.5	56.2	44.8	14.7	4.8	25.9	13.1	11.6	14.2
	10	200	30.2	23.9	16.4	10.1	98.7	96.1	74.9	40.1	70.9	58.3	57.7	64.2
	20	100	12.5	10.0	6.6	3.6	79.8	68.9	22.0	6.2	67.5	36.9	19.5	19.7
	20	200	30.9	24.1	17.1	11.0	100.0	99.8	94.9	60.6	96.8	91.4	87.4	88.5
3	10	100	25.8	15.5	56.8	22.4	90.7	58.8	99.4	51.2	18.9	13.9	22.6	31.7
	10	200	71.1	46.8	82.6	43.2	100.0	99.0	100.0	91.8	29.2	28.5	32.2	42.2
	20	100	26.6	16.4	56.5	22.6	98.9	76.4	100.0	57.7	32.0	22.7	32.7	38.9
	20	200	71.3	47.6	82.6	43.4	100.0	99.9	100.0	98.1	47.8	47.3	43.0	47.6
	10	100	25.8	15.5	56.8	22.4	90.7	58.8	99.4	51.2	18.9	13.9	22.6	31.7
	10	200	71.1	46.8	82.6	43.2	100.0	99.0	100.0	91.8	29.2	28.5	32.2	42.2
4	20	100	26.6	16.4	56.5	22.6	98.9	76.4	100.0	57.7	32.0	22.7	32.7	38.9
	20	200	71.3	47.6	82.6	43.4	100.0	99.9	100.0	98.1	47.8	47.3	43.0	47.6

Notes: AR case 1 refers to the setup with no serial correlation, while cases 2 to 4 refer to the setup with serial correlation in the errors driving $\Delta Y_{i,t}$, $\Delta X_{i,t}$ and $\Delta F_{i,t}$, respectively. In all cases, the serial correlation is of the first-order AR type with a common AR parameter of magnitude 0.5. The error correction parameter α_1 is zero under the null and -0.05 under the alternative.

Table 4.6: Size and size-adjusted power in Model 2 with an unrestricted constant.

AR case	N	T	This study								Westerlund (2007)			
			$\bar{\tau}_{\alpha_{1i}}$	$\bar{\tau}_{\alpha_{1i}}$	$w_{\delta_{1i}}^{\hat{\epsilon}}$	$w_{\delta_{1i}}^{\bar{\epsilon}}$	$\bar{\tau}_{\alpha_1}^*$	$\bar{\tau}_{\alpha_1}^*$	$\bar{w}_{\delta_1}^*$	$\bar{w}_{\delta_1}^*$	G_τ	G_ρ	P_τ	P_ρ
1	10	100	6.2	6.3	8.1	8.2	6.1	7.4	13.4	16.0	12.5	6.4	13.2	14.9
	10	200	5.7	5.7	6.5	6.4	6.0	6.9	8.6	11.7	9.1	6.5	11.8	14.2
	20	100	6.0	6.2	8.1	8.0	5.6	7.3	15.9	20.5	12.2	4.9	10.5	10.0
	20	200	5.5	5.7	6.5	6.6	5.9	7.8	9.8	15.0	9.0	5.1	10.1	10.4
	10	100	6.9	6.3	8.6	9.1	7.9	7.9	15.5	20.3	4.4	1.0	4.0	1.2
	10	200	6.0	5.3	6.8	7.0	6.9	5.7	9.4	13.1	2.6	0.8	2.9	1.1
2	20	100	6.7	6.3	8.7	8.9	7.9	7.7	19.7	26.1	4.8	1.1	3.9	2.4
	20	200	5.9	5.2	6.8	7.0	7.1	6.1	11.3	18.0	2.5	0.5	3.0	1.3
	10	100	4.7	5.3	19.3	16.3	0.6	1.1	50.1	41.2	6.6	1.3	3.3	1.1
3	10	200	5.1	4.8	11.1	8.7	2.5	2.5	22.1	15.2	7.0	1.9	7.9	4.5
	20	100	4.7	5.1	19.6	16.1	0.1	0.4	68.9	52.7	7.3	0.6	2.5	0.8
	20	200	5.0	4.6	10.8	8.4	1.5	1.8	27.3	18.4	5.9	0.5	7.1	2.0
	10	100	7.8	6.5	9.7	11.2	10.6	6.5	20.2	25.8	12.7	6.6	14.1	15.4
	10	200	6.5	5.4	7.3	8.8	7.8	5.5	11.3	19.2	9.2	6.7	11.7	14.0
	20	100	7.7	6.4	9.8	11.2	12.2	6.6	26.2	32.0	12.9	5.3	11.4	10.9
4	20	200	6.3	5.4	7.2	9.2	8.7	5.5	13.8	24.6	8.4	4.8	9.2	9.4

Continued overleaf

Table 4.6: Continued.

AR case	N	T	This study						Westerlund (2007)					
			T_{α_1}	T_{α_1}	$w_{\hat{\delta}_1}$	$w_{\hat{\delta}_1}$	$\bar{\tau}_{\hat{\alpha}_1}^*$	$\bar{\tau}_{\hat{\alpha}_1}^*$	$\bar{w}_{\hat{\delta}_1}^*$	$\bar{w}_{\hat{\delta}_1}^*$	G_τ	G_ρ	P_τ	P_ρ
1	10	200	9.1	8.7	17.6	10.1	31.7	26.7	60.5	25.1	29.9	21.3	34.2	43.5
	10	200	32.2	23.0	45.7	24.1	98.5	88.7	99.5	75.7	48.4	51.1	52.9	67.0
	20	100	9.4	8.6	18.0	10.3	49.5	39.4	82.2	33.8	55.2	45.2	56.4	65.5
2	20	200	32.7	23.3	46.7	24.5	100.0	98.3	100.0	90.5	79.8	82.8	76.3	83.2
	10	100	8.7	9.0	17.1	9.4	29.4	25.1	59.4	23.2	6.5	2.3	6.5	3.5
	10	200	31.5	22.4	45.6	22.9	98.4	85.7	99.6	73.9	7.4	4.8	7.3	6.9
3	20	100	9.0	9.0	17.4	9.9	45.9	36.3	80.9	31.0	22.8	14.0	22.4	25.2
	20	200	32.0	23.1	46.3	23.4	99.9	96.2	100.0	88.4	34.8	33.7	33.7	41.9
	10	100	8.4	7.3	7.4	5.0	25.3	22.0	15.6	8.5	21.5	11.0	7.5	6.7
4	10	200	19.8	15.6	15.2	9.8	84.8	75.1	68.0	34.9	60.5	53.6	50.5	55.2
	20	100	8.8	7.5	7.2	5.1	41.6	34.0	23.5	11.1	54.1	28.1	17.4	17.5
	20	200	20.2	16.4	15.7	10.3	97.7	93.3	90.1	54.5	95.6	90.8	84.9	85.2
4	10	100	12.4	8.7	40.5	18.3	41.3	23.6	94.6	41.9	23.4	16.0	25.4	30.9
	10	200	53.6	31.9	74.2	40.2	99.8	89.4	100.0	88.9	32.6	32.6	35.3	44.4
	20	100	13.1	8.8	40.7	18.9	61.7	31.5	99.3	49.5	38.0	26.4	37.4	41.5
	20	200	54.3	32.7	74.6	40.1	100.0	97.1	100.0	96.1	52.6	53.9	48.3	52.2

Notes: See Table 4.5 for an explanation of the various features of this table.

Table 4.7: Size and size-adjusted power in Model 3 with unrestricted constant and trend terms.

AR case	N	T	This study								Westerlund (2007)			
			$\bar{\tau}_{\alpha_{1t}}$	$\bar{\tau}_{\alpha_{1t}}$	$w_{\delta_{1t}}^{\hat{\epsilon}}$	$w_{\delta_{1t}}^{\tilde{\epsilon}}$	$\bar{\tau}_{\alpha_1}^*$	$\bar{\tau}_{\alpha_1}^*$	$\bar{w}_{\delta_1}^*$	$\bar{w}_{\delta_1}^*$	G_r	G_p	P_r	P_p
1	10	100	6.6	6.8	8.4	8.4	6.3	8.2	14.4	17.8	16.6	5.4	14.9	11.2
	10	200	5.9	5.9	6.6	6.5	6.1	7.4	9.2	12.6	9.8	6.0	11.6	10.9
	20	100	6.4	6.6	8.7	8.5	5.8	8.8	19.0	23.7	15.0	3.3	11.1	6.6
	20	200	5.8	6.0	6.8	6.9	6.2	9.2	11.7	18.4	8.7	3.6	8.3	6.5
	10	100	7.3	6.7	9.1	9.7	8.5	7.5	17.2	22.7	3.8	0.7	1.7	0.4
	10	200	6.3	5.4	6.9	7.2	7.2	5.3	10.5	15.0	1.7	0.2	0.9	0.2
2	20	100	7.2	6.6	9.4	9.7	9.0	8.8	24.2	31.6	4.2	0.9	2.1	0.6
	20	200	6.3	5.4	7.2	7.4	7.5	6.7	13.6	21.8	1.2	0.2	1.0	0.3
	10	100	4.3	5.1	15.7	14.1	0.5	1.0	34.8	30.3	6.9	1.2	2.2	1.0
	10	200	5.1	4.7	9.7	7.8	1.8	2.1	16.6	12.8	5.7	1.6	5.7	2.7
	20	100	4.3	5.0	16.1	14.2	0.1	0.2	50.3	42.6	7.2	0.5	2.3	0.6
	20	200	5.1	4.5	10.0	8.1	1.4	1.6	22.6	17.5	4.8	0.2	5.1	0.6
4	10	100	8.4	6.9	10.4	10.2	12.3	6.9	23.7	22.4	17.0	6.2	15.9	12.0
	10	200	6.9	5.4	7.5	7.7	8.5	5.1	12.1	15.6	10.3	5.5	12.0	10.7
	20	100	8.4	6.7	10.7	10.3	14.1	7.8	35.2	28.3	16.8	4.1	11.6	6.9
	20	200	6.8	5.5	7.8	8.1	10.0	6.4	17.0	21.5	8.1	3.4	7.6	5.8

Continued overleaf

Table 4.7: Continued.

AR case	N	T	This study						Westerlund (2007)					
			$\bar{\tau}_{\alpha_i}$	τ_{α_i}	w_{δ_i}	$w_{\delta_i}^*$	$\bar{\tau}_{\hat{\alpha}_1}$	$\tau_{\hat{\alpha}_1}^*$	$\bar{w}_{\hat{\delta}_1}^*$	$w_{\hat{\delta}_1}^*$	G_τ	G_ρ	P_τ	P_ρ
1	10	200	6.2	6.4	11.7	8.0	10.5	12.3	37.1	16.3	26.3	11.5	26.2	22.4
	10	200	18.5	14.1	35.5	19.1	80.2	61.2	96.4	59.4	39.7	32.3	44.9	50.3
	20	100	6.4	6.6	11.9	8.2	14.9	15.6	53.5	19.3	40.7	15.9	35.7	28.3
	20	200	18.8	14.5	36.0	19.2	96.9	81.1	99.8	73.4	69.1	61.1	69.5	72.0
	10	100	6.1	6.6	11.5	7.4	9.6	13.9	36.0	14.3	4.6	1.3	2.4	0.8
	10	200	17.7	14.6	35.0	17.5	78.7	58.9	96.3	55.0	4.4	2.0	2.9	1.6
2	20	100	6.2	6.9	11.6	7.7	13.0	16.2	52.3	16.3	13.9	3.8	8.9	4.9
	20	200	18.2	15.0	35.8	18.0	96.5	75.2	99.8	70.2	19.8	12.4	18.6	17.5
	10	100	6.5	6.1	7.3	5.9	13.0	10.1	14.3	9.1	13.1	3.5	3.7	2.3
	10	200	12.9	11.1	13.9	9.3	54.3	45.4	55.7	29.3	39.3	20.8	36.8	31.5
	20	100	6.7	6.2	7.3	5.8	17.3	14.0	20.4	11.1	28.5	3.9	11.1	4.9
	20	200	13.1	11.5	14.1	9.4	76.5	65.8	78.6	42.2	79.9	43.0	77.8	67.3
4	10	100	5.7	5.4	27.4	14.4	6.8	7.2	79.3	31.2	23.4	9.8	21.8	18.0
	10	200	33.2	18.4	65.6	36.7	95.3	60.5	100.0	82.4	27.6	20.5	30.2	32.8
	20	100	5.9	5.5	27.6	14.9	8.1	7.0	92.1	34.8	29.8	10.9	25.1	19.0
	20	200	33.8	18.6	65.6	36.0	99.8	75.2	100.0	91.9	42.3	33.1	40.0	39.5

Notes: See Table 4.5 for an explanation of the various features of this table.

Table 4.8: Cointegration test results for the Fisher effect.

Country	Model 2				Model 3			
	$\tau_{\alpha_{1i}^*}$	$\tau_{\alpha_{1i}}$	$w_{\delta_{1i}^*}$	$w_{\delta_{1i}}$	$\tau_{\alpha_{1i}^*}$	$\tau_{\alpha_{1i}}$	$w_{\delta_{1i}^*}$	$w_{\delta_{1i}}$
Australia	-3.43*	-4.22**	5.19	13.95	-2.01	-2.82	17.77*	18.34*
Belgium	-1.84	-3.36	6.40	12.98	-2.82	-3.69	10.80	15.63
Canada	-2.33	-4.51***	9.16	22.38***	-3.26	-4.10*	15.89	24.82***
Switzerland	-0.70	-2.18	3.74	5.72	-1.09	-1.85	8.51	5.50
Germany	-2.93	-2.27	6.25	7.69	-3.30	-2.70	8.61	7.92
Denmark	-4.90***	-5.55***	28.43***	28.43***	-5.06***	-4.69**	32.46***	32.46***
Spain	-3.72*	-4.57***	14.57	22.46***	-4.48**	-4.62**	21.16**	23.09**
Finland	-2.74	-3.43***	9.97	13.27	-2.89	-3.32	10.83	13.04
France	-1.82	-2.54	7.49	7.70	-2.12	-3.18	11.01	12.95
United Kingdom	-4.39	-4.57***	24.07***	25.54***	-4.41**	-4.36**	23.28**	24.89**
Ireland	-6.40***	-4.46	28.43***	28.43***	-6.60***	-4.45**	32.46***	32.46***
Italy	-3.14	-4.16**	18.84**	25.10***	-3.72	-3.82*	22.96**	24.38**
Japan	-2.83	-5.60***	24.39***	28.43***	-3.30	-5.76***	22.40**	32.46***
Luxembourg	-2.83	-4.18**	11.46	23.01***	-3.84*	-4.37*	18.35*	23.89**
Netherlands	-0.31	-2.39	4.57	10.96	-0.97	-2.89	6.19	13.17
Norway	-2.70	-3.00	8.83	11.71	-4.15*	-3.93*	8.67	11.23
New Zealand	-6.40***	-6.40***	28.43***	28.43***	-6.14***	-6.19***	32.46***	32.46***
Portugal	-3.27	-3.47*	28.43***	26.94***	-3.46	-3.20	32.18***	25.26***
Sweden	-2.49	-4.44***	13.77	22.08***	-3.82*	-4.49*	23.40**	22.28**
United States	-1.46	-1.72	2.99	5.19	-6.06***	-6.32***	32.46***	32.46***
All 20 ^a	-3.03	-3.85	14.27***	18.52***	-3.68	-4.04	19.59***	21.43***
All 20 ^b	G_τ -8.74	G_ρ -9.30	P_τ -7.43***	P_ρ -10.01	G_τ -8.72***	G_ρ -6.64***	P_τ -7.00	P_ρ -6.12***

Notes: The superscripts (***), (**) and (*) denote significance at the 1%, 5% and 10% levels, respectively. The superscript (*) in $\tau_{\alpha_{1i}^*}$ and $w_{\delta_{1i}^*}$ indicates that these tests have been computed using principal component estimates of the factors instead of cross-sectional averages. See Table 4.1 for an explanation of the two models.

^aThe reported values are for the average of the truncated test statistics.
^bThe reported values are obtained by applying the Westerlund (2007) tests to the defactored data.

Table 4.9: Cointegration test results for the monetary exchange rate model.

Country	Model 2							Model 3				
	$T\alpha_{1i}^+$	$T\alpha_{1i}^{++}$	$T\alpha_{1i}$	$w\delta_{1i}^+$	$w\delta_{1i}^{++}$	$w\delta_{1i}^z$	$T\alpha_{1i}^+$	$T\alpha_{1i}^{++}$	$T\alpha_{1i}$	$w\delta_{1i}^+$	$w\delta_{1i}^{++}$	$w\delta_{1i}^z$
United Kingdom	-2.38	-3.19	-1.41	12.70	20.36*	12.16	-2.32	-3.60	-1.48	12.76	27.75**	13.71
Austria	-2.54	-2.62	-1.59	9.59	10.88	17.46	-1.62	-2.49	-2.46	6.68	10.28	16.22
Belgium	-4.62**	-2.60	-1.63	23.26**	10.10	15.94	-4.72**	-3.05	-1.53	27.91**	12.57	18.45
Denmark	-2.20	-2.66	-1.03	15.81	23.68**	16.04	-3.07	-2.18	-1.14	21.02	23.12*	16.48
France	-1.58	-2.87	-1.98	12.26	14.23	6.55	-2.52	-2.36	-3.23	14.14	14.12	13.25
Germany	-2.11	-3.10	-2.23	9.27	12.63	23.19**	-3.84	-3.18	-2.58	19.92	12.63	22.46*
Netherlands	-2.66	-2.53	-1.04	10.15	9.19	15.33	-2.54	-2.58	-2.38	8.88	11.11	19.04
Canada	-2.46	-1.16	-2.94	9.30	11.56	17.31	-2.56	-1.38	-2.89	9.87	9.47	17.57
Japan	-4.06**	-2.38	-0.67	22.17*	17.44	9.36	-2.68	-2.33	-1.23	22.24	16.47	10.07
Finland	-1.86	-2.29	-2.56	19.95*	7.40	16.49	-2.67	-2.83	-2.40	24.41*	11.77	17.07
Greece	-2.89	-3.53	-2.55	14.67	16.69	14.40	-2.99	-3.51	-3.60	15.14	17.83	21.51
Spain	-4.46**	-3.71	-5.84***	24.88**	31.11***	36.09***	-4.49**	-3.76	-5.80***	25.78**	31.28***	39.69***
Australia	-3.27	-2.90	-3.40	15.87	12.07	20.78*	-3.44	-2.91	-2.82	15.63	11.28	19.11
Italy	-1.99	-2.19	-4.54**	13.33	17.32	23.24**	-3.54	-2.00	-4.62**	23.12*	15.73	23.54*
Switzerland	-2.22	-3.55	-3.05	10.26	15.52	36.09***	-2.07	-3.42	-2.77	5.86	12.89	32.35***
Korea	-1.07	-3.79	-3.37	30.07***	36.09***	35.01***	-1.82	-4.11	-3.47	28.10**	38.30***	32.46***
Norway	-3.20	-2.62	-4.48**	18.79	10.48	29.89***	-2.96	-2.60	-4.39*	18.53	10.33	30.37**
Sweden	-1.68	-2.26	-4.17**	6.91	15.54	23.20**	-2.74	-2.33	-4.15	11.36	15.51	22.95*
European community ^a	-2.73*	-2.93**	-2.50	16.38***	17.26***	15.90***	-3.17**	-2.91	-3.00	19.29***	18.46***	18.61***
European monetary system ^b	-1.59	-2.70*	-2.42	13.84**	14.53**	13.92**	-2.67	-2.56	-3.04	18.82***	14.88	15.21*
Group of 6 ^c	-2.42	-2.48	-1.98	14.31**	15.59***	12.28	-2.56	-2.48	-2.45	16.32**	16.03**	14.78
Group of 10 ^d	-2.48	-2.58	-2.48	14.10**	14.39**	15.18***	-2.71	-2.62	-3.19**	14.90	14.82	19.86***
All 18	-2.63*	-2.78**	-2.69**	15.51***	16.24***	20.48***	-2.92	-2.81	-2.94	17.30***	16.80***	21.46***

Notes: The individual tests are based on the full panel averages. The superscript (*) in $T\alpha_{1i}^+$ and $w\delta_{1i}^z$ indicates that these tests have been computed using the observed factors, while (**) indicates that the tests are based on applying the principal components method to a pre-specified cointegrating relationship. See Table 4.8 for an explanation of the remaining features of the table.

^aBelgium, Denmark, France, Germany, Greece, Italy, Netherlands, Spain and United Kingdom.

^bBelgium, Denmark, France, Germany, Italy, Netherlands, Spain.

^cCanada, France, Germany, Italy, Japan and United Kingdom.

^dBelgium, Canada, France, Germany, Italy, Japan, Netherlands, Sweden, Switzerland and United Kingdom.

Table 4.10: Descriptive statistics for the common factors in the monetary exchange rate model.

Value	Principal components			Observed	
	Factor 1	Factor 2	Factor 3	m_t^*	y_t^*
AR	0.97	0.98	0.97	1.00	1.00
SE	0.03	0.05	0.02	0.00	0.01
ADF	-1.09	-0.34	-1.79	-3.06**	-0.42

Notes: AR refers to the estimated first order AR coefficient, SE refers to its standard error and ADF refers to the augmented Dickey and Fuller (1979) test. The autoregressions are fitted with an intercept and the lag orders are determined using the Schwarz Bayesian criterion. See Table 4.8 for an explanation of the remaining features of the table.

4.B Appendix

4.B.1 Proof of Theorem 4.1.

Consider (a). From (4.2.3) and (4.2.4) we have $\Delta F_t - \pi_3' g_t = \Psi(L)\eta_t$. Substituting for ΔF_t in (4.2.2) and using (4.2.4) we obtain

$$\Delta X_{i,t} - \pi_{2i}' g_t = \Gamma_{21i}(L)\varepsilon_{1i,t} + \Gamma_{22i}(L)\varepsilon_{2i,t} + \lambda_{2i}'\Psi(L)\eta_t. \quad (4.B.1)$$

Taking first differences of (4.2.1) and substituting for ΔF_t and $\Delta X_{i,t}$ from (4.B.1) we obtain the following MA representation

$$\begin{aligned} \Delta Y_{i,t} - \pi_{1i}' g_t &= ((1-L)\Gamma_{11i} + b_i'\Gamma_{21i}(L))\varepsilon_{1i,t} \\ &+ ((1-L)\Gamma_{21i} + b_i'\Gamma_{22i}(L))\varepsilon_{2i,t} + (\lambda_{1i}' + b_i'\lambda_{2i}')\Psi(L)\eta_t. \end{aligned} \quad (4.B.2)$$

Combining the results above we find

$$C_i(L) = \begin{pmatrix} (1-L)\Gamma_{11i}(L) + b_i'\Gamma_{21i}(L) & (1-L)\Gamma_{12i}(L) + b_i'\Gamma_{22i}(L) & (\lambda_{1i}' + b_i'\lambda_{2i}')\Psi(L) \\ \Gamma_{21i}(L) & \Gamma_{22i}(L) & \lambda_{2i}'\Psi(L) \\ 0 & 0 & \Psi(L) \end{pmatrix}.$$

Furthermore,

$$C_i(1) = \begin{pmatrix} b_i' & (\lambda_{1i}' + b_i'\lambda_{2i}') \\ I_m & \lambda_{2i}' \\ 0 & I_k \end{pmatrix} \begin{pmatrix} \Gamma_{21i}(1) & \Gamma_{22i}(1) & 0 \\ 0 & 0 & \Psi(1) \end{pmatrix}$$

such that $C_i(1)$ has rank $m + k$. This establishes part (a) of the theorem.

Next, consider (b). Partition $C_i(L)$ such that the diagonal blocks $C_{11i}(L)$ and $C_{22i}(L)$ are of dimension $r \times r$ and $(m + k) \times (m + k)$, respectively. Since

$$C_{22i}(L) = \begin{pmatrix} \Gamma_{22i}(L) & \lambda_{2i}'\Psi(L) \\ 0 & \Psi(L) \end{pmatrix}$$

is invertible, we can decompose $C_i(L)$ as

$$C_i(L) = \begin{pmatrix} I_r & C_{12i}(L)C_{22i}(L)^{-1} \\ 0 & I_{(m+k)} \end{pmatrix} \begin{pmatrix} C_{11\cdot 2i}(L) & 0 \\ C_{21i}(L) & C_{22i}(L) \end{pmatrix}.$$

As $C_{11 \cdot 2i} = (1 - L)\Gamma_{11 \cdot 2i}(L)$ we can further write

$$\begin{aligned} C_i(L) &= \begin{pmatrix} I_r & C_{12i}(L)C_{22i}(L)^{-1} \\ 0 & I_{(m+k)} \end{pmatrix} \begin{pmatrix} (1-L)I_r & 0 \\ 0 & I_{(m+k)} \end{pmatrix} \begin{pmatrix} \Gamma_{11 \cdot 2i}(L) & 0 \\ C_{21i}(L) & C_{22i}(L) \end{pmatrix} \\ &= U_i(L)^{-1}M(L)V_i(L)^{-1}, \end{aligned} \quad (4.B.3)$$

where the lag polynomials

$$V_i(L) = \begin{pmatrix} \Gamma_{11 \cdot 2i}(L)^{-1} & 0 & 0 \\ -\Gamma_{22i}(L)^{-1}\Gamma_{21i}(L)\Gamma_{11 \cdot 2i}(L)^{-1} & \Gamma_{22i}(L)^{-1} & -\Gamma_{22i}(L)^{-1}\lambda'_{2i} \\ 0 & 0 & \Psi(L)^{-1} \end{pmatrix}$$

and

$$U_i(L) = \begin{pmatrix} I_r & -((1-L)\Gamma_{12i}(L)\Gamma_{22i}(L)^{-1} + b'_i) & (1-L)\Gamma_{12i}(L)\Gamma_{22i}(L)^{-1}\lambda'_{2i} - \lambda'_{1i} \\ 0 & I_m & 0 \\ 0 & 0 & I_k \end{pmatrix}$$

are invertible.

Substituting (4.B.3) for $C_i(L)$ in the MA representation of $\Delta Z_{i,t}^+$ and pre-multiplying by $U_i(L)$ and $\bar{M}(L)$, where

$$\bar{M}(L) = \begin{pmatrix} I_r & 0 \\ 0 & (1-L)I_{(m+k)} \end{pmatrix},$$

we obtain

$$\bar{M}(L)U_i(L)(1-L)(Z_{i,t}^+ - (\pi_i^*)'g_t) = (1-L)V_i(L)^{-1}\varepsilon_{i,t}^+.$$

Eliminating $(1-L)$ from both sides and pre-multiplying by $V_i(L)$ yields the following possibly infinite AR representation for $Z_{i,t}^+$

$$V_i(L)\bar{M}(L)U_i(L)(Z_{i,t}^+ - (\pi_i^*)'g_t) = \varepsilon_{i,t}^+.$$

Using that $\Gamma_{11 \cdot 2i}(L)^{-1} = |\Gamma_{11 \cdot 2i}(L)|^{-1}adj(\Gamma_{11 \cdot 2i}(L))$, $\Gamma_{22i}(L)^{-1} = |\Gamma_{22i}(L)|^{-1}adj(\Gamma_{22i}(L))$, $\Psi(L)^{-1} = |\Psi(L)|^{-1}adj(\Psi(L))$ and $|\Gamma_i^+(L)| = |\Gamma_{11 \cdot 2i}(L)||\Gamma_{22i}(L)||\Psi(L)|$, we can recover both the scalar lag polynomial $c_i(L) = |\Gamma_i^+(L)|$ and the lag polynomial matrix $A_i(L)$ given in the theorem. This establishes part (b).

Consider (c). Direct computation of $A_i(1)$ yields

$$A_i(1) = \begin{pmatrix} |\Psi(1)||\Gamma_{22i}(1)|adj(\Gamma_{11 \cdot 2i}(1)) \\ -|\Psi(1)|adj(\Gamma_{22i}(1))\Gamma_{21i}(1)adj(\Gamma_{11 \cdot 2i}(1)) \\ 0 \end{pmatrix} \begin{pmatrix} I_r & -b'_i & -\lambda'_{1i} \end{pmatrix} = \alpha_i^* \beta'_i.$$

Since

$$C_i(1) = \begin{pmatrix} b'_i & (\lambda'_{1i} + b'_i \lambda'_{2i}) \\ I_m & \lambda'_{2i} \\ 0 & I_k \end{pmatrix} \begin{pmatrix} \Gamma_{21i}(1) & \Gamma_{22i}(1) & 0 \\ 0 & 0 & \Psi(1) \end{pmatrix} = \tilde{\beta}_i (\tilde{\alpha}_i^*)',$$

where $\tilde{\alpha}_i^*$ and $\tilde{\beta}_i$ denote the matrices orthogonal to α_i^* and β_i , respectively. It follows that $\beta_i' C_i(1) = 0$ and $C_i(1) \alpha_i = 0$, and so the proof of (c) is complete.

Parts (d) and (e) follow by manipulating of the lag polynomial matrix $A_i(L)$ and rearranging terms, as in Engle and Granger (1987).

4.B.2 Proof of Theorem 4.2.

Before we come to the proof of the theorem we need some preliminary results, which are summarized in Lemma 4.B.1.

Lemma 4.B.1 *Under H_{0i} and Assumptions 4.1, 4.2, 4.4 and 4.5, as $T \rightarrow \infty$*

- (a) $T^{-\frac{1}{2}} Z_{i,t-1}^+ \xrightarrow{w} \Omega_i^{\frac{1}{2}} W_i$,
- (b) $T^{-2} \sum_{t=2}^T Q_W Z_{i,t-1}^+ (Q_W Z_{i,t-1}^+)' \xrightarrow{w} \Omega_i^{\frac{1}{2}} \int W_i W_i' (\Omega_i^{\frac{1}{2}})',$
- (c) $T^{-1} \sum_{t=2}^T Q_W Z_{i,t-1}^+ Q_W \varepsilon_{1 \cdot 2i,t} \xrightarrow{w} \sigma_i \Omega_i^{\frac{1}{2}} p_i(W_i)$,

where

$$\begin{aligned} \Omega_i^{\frac{1}{2}} &= \begin{pmatrix} \sigma_i \tilde{B}_{11 \cdot 2i}^{-1} & \tilde{B}_{11 \cdot 2i}^{-1} M_{1i} M_{2i} \\ \sigma_i \tilde{B}_{11 \cdot 2i}^{-1} \rho_i & (M_{2i} + \tilde{B}_{11 \cdot 2i}^{-1} \rho_i M_{1i} M_{2i}) \end{pmatrix} \\ &= \begin{pmatrix} \omega_{11i} & \Omega_{12i} \\ \Omega_{21i} & \Omega_{22i} \end{pmatrix}, \end{aligned}$$

where $\tilde{B}_{jji}(L) = I - B_{jji}(L)L$, $B_{jli} = B_{jli}(1)$, $\tilde{B}_{jli} = \tilde{B}_{jli}(1)$, $\tilde{B}_{11 \cdot 2i}^{-1} = (\tilde{B}_{11i} - B_{12i} \tilde{B}_{22i}^{-1} B_{21i})^{-1}$, $\rho_i' = \begin{pmatrix} B_{21i}' (\tilde{B}_{22i}^{-1})' & 0 \end{pmatrix}$, $M_{1i} = \begin{pmatrix} B_{12i} & B_{13i} \end{pmatrix}$ and

$$M_{2i} = \begin{pmatrix} \tilde{B}_{22i}^{-1} \Sigma_{22i}^{\frac{1}{2}} & \tilde{B}_{22i}^{-1} (B_{23i} \tilde{B}_{33i}^{-1} + \lambda'_{2i}) \\ 0 & \tilde{B}_{33i}^{-1} \end{pmatrix}.$$

Note that $\Omega_{21i} = \omega_{11i} \rho_i$ and define for future use $\omega_{11 \cdot 2i} = \omega_{11i} - \omega_{11i} \Omega_{12} \Omega_{22i}^{-1} \rho_i$.

Proof of Lemma 4.B.1.

Consider (a). Note that under H_{0i} ,

$$\begin{pmatrix} \tilde{B}_{11i}(1) & -B_{12i}(1) & -B_{13i}(1) \\ -B_{21i}(1)L & \tilde{B}_{22i}(1) & -B_{23i}(1)L \\ 0 & 0 & \tilde{B}_{33i}(1) \end{pmatrix} \Delta Z_{i,t}^+ = \begin{pmatrix} \varepsilon_{1 \cdot 2i,t} \\ \varepsilon_{2i,t}^* \\ \eta_t \end{pmatrix},$$

such that

$$\begin{aligned} Z_{i,t}^+ &= \begin{pmatrix} \tilde{B}_{11i}(1) & -B_{12i}(1) & -B_{13i}(1) \\ -B_{21i}(1) & \tilde{B}_{22i}(1) & -B_{23i}(1) \\ 0 & 0 & \tilde{B}_{33i}(1) \end{pmatrix}^{-1} \begin{pmatrix} t \\ \sum_{s=1}^t \end{pmatrix} \begin{pmatrix} \varepsilon_{1 \cdot 2i,t} \\ \varepsilon_{2i,t}^* \\ \eta_t \end{pmatrix} \\ &+ \begin{pmatrix} -\tilde{B}_{11i}^+(L) & B_{12i}^+(L) & B_{13i}^+(L) \\ B_{21i}^+(L)L & -\tilde{B}_{22i}^+(L) & B_{23i}^+(L)L \\ 0 & 0 & -\tilde{B}_{33i}^+(L) \end{pmatrix} \Delta Z_{i,t}^+, \end{aligned} \quad (4.B.4)$$

where $B_{jli}^+(L)$ and $\tilde{B}_{jli}^+(L)$ are obtained from the Beveridge-Nelson decompositions of $B_{jli}(L)$ and $\tilde{B}_{jli}(L)$ as $B_{jli}(L) = B_{jli}(1) + B_{jli}^+(L)(1-L)$ and $\tilde{B}_{jli}(L) = \tilde{B}_{jli}(1) + \tilde{B}_{jli}^+(L)(1-L)$, respectively.

Substituting $\varepsilon_{2i,t}^* = \varepsilon_{2i,t} + \lambda'_{2i}\eta_t$ into (4.B.4) we obtain

$$\begin{aligned} T^{-\frac{1}{2}} Z_{i,t}^+ &= \begin{pmatrix} \tilde{B}_{11i}(1) & -B_{12i}(1) & -B_{13i}(1) \\ -B_{21i}(1) & \tilde{B}_{22i}(1) & -B_{23i}(1) \\ 0 & 0 & \tilde{B}_{33i}(1) \end{pmatrix}^{-1} \begin{pmatrix} \sigma_i & 0 & 0 \\ 0 & \Sigma_{22i}^{\frac{1}{2}} & \lambda'_{2i} \\ 0 & 0 & I_k \end{pmatrix} \\ &\times T^{-\frac{1}{2}} \sum_{s=1}^t \begin{pmatrix} \sigma_i^{-1} \varepsilon_{1 \cdot 2i,t} \\ \Sigma_{22i}^{-\frac{1}{2}} \varepsilon_{2i,t} \\ \eta_t \end{pmatrix} + o_p(1), \end{aligned}$$

such that

$$T^{-\frac{1}{2}} Z_{i,t}^+ \xrightarrow{w} \Omega_i^{\frac{1}{2}} W_i \quad (4.B.5)$$

as $T \rightarrow \infty$, proving (a).

Now, by using the rules for projections, $\sum_{t=2}^T Q_W Z_{i,t-1}^+ (Q_W Z_{i,t-1}^+)'$ can be written as

$$\begin{aligned} \sum_{t=2}^T Q_W Z_{i,t-1}^+ (Q_W Z_{i,t-1}^+)' &= \sum_{t=2}^T Z_{i,t-1}^+ (Z_{i,t-1}^+)' \\ &- \sum_{t=2}^T Z_{i,t-1}^+ W_{i,t}' \left(\sum_{t=2}^T W_{i,t} W_{i,t}' \right)^{-1} \sum_{t=2}^T W_{i,t} (Z_{i,t-1}^+)' . \end{aligned} \quad (4.B.6)$$

By Lemma 2.1 of Park and Phillips (1988), $\sum_{t=2}^T Z_{i,t-1}^+ W_{i,t}' = O_p(T)$, $\sum_{t=2}^T Z_{i,t-1}^+ (Z_{i,t-1}^+)' = O_p(T^2)$ and

$\sum_{t=2}^T W_{i,t}W'_{i,t} = O_p(T)$ such that (4.B.6) reduces to

$$\begin{aligned} T^{-2} \sum_{t=2}^T Q_W Z_{i,t-1}^+ (Q_W Z_{i,t-1}^+)' &= T^{-2} \sum_{t=2}^T Z_{i,t-1}^+ (Z_{i,t-1}^+)' + T^{-2} O_p(T) O_p(T^{-1}) O_p(T) \\ &= T^{-2} \sum_{t=2}^T Z_{i,t-1}^+ (Z_{i,t-1}^+)' + O_p(T^{-1}), \end{aligned}$$

where we can make use of (a) to show that as $T \rightarrow \infty$

$$T^{-2} \sum_{t=2}^T Z_{i,t-1}^+ (Z_{i,t-1}^+)' \xrightarrow{w} \Omega_i^{\frac{1}{2}} \int W_i W_i' (\Omega_i^{\frac{1}{2}})'$$

This proves (b).

Finally, consider (c). By definition,

$$\sum_{t=2}^T Q_W Z_{i,t-1}^+ (Q_W \varepsilon_{1 \cdot 2i,t}) = \sum_{t=2}^T Z_{i,t-1}^+ \varepsilon_{1 \cdot 2i,t} - \sum_{t=2}^T Z_{i,t-1}^+ W'_{i,t} \left(\sum_{t=2}^T W_{i,t} W'_{i,t} \right)^{-1} \sum_{t=2}^T W_{i,t} \varepsilon_{1 \cdot 2i,t},$$

where $\sum_{t=2}^T W_{i,t} \varepsilon_{1 \cdot 2i,t} = O_p(\sqrt{T})$. Thus, by using the same arguments as above,

$$T^{-1} \sum_{t=2}^T Q_W Z_{i,t-1}^+ (Q_W \varepsilon_{1 \cdot 2i,t}) = T^{-1} \sum_{t=2}^T Z_{i,t-1}^+ \varepsilon_{1 \cdot 2i,t} + O_p(T^{-\frac{1}{2}}), \quad (4.B.7)$$

where the limit of the first term on the right-hand side is given by

$$T^{-1} \sum_{t=2}^T Z_{i,t-1}^+ \varepsilon_{1 \cdot 2i,t} \xrightarrow{w} \sigma_i \Omega_i^{\frac{1}{2}} \int W_i dW_{1i} = \sigma_i \Omega_i^{\frac{1}{2}} p_i(W_i).$$

This establishes (c), and hence the proof of Lemma 4.B.1 is complete. \square

Now, since under the null hypothesis,

$$\Delta(Q_W Y_{i,t}) = \delta'_{1i} (Q_W Z_{i,t-1}^+) + Q_W \varepsilon_{1 \cdot 2i,t} = Q_W \varepsilon_{1 \cdot 2i,t},$$

we have

$$\hat{\delta}_{1i} = \left(\sum_{t=2}^T Q_W Z_{i,t-1}^+ (Q_W Z_{i,t-1}^+)' \right)^{-1} \sum_{t=2}^T Q_W Z_{i,t-1}^+ (Q_W \varepsilon_{1 \cdot 2i,t}).$$

From Lemma 4.B.1 (b) and (c) we have that

$$T \hat{\delta}_{1i} \xrightarrow{w} \sigma_i \left(\Omega_i^{\frac{1}{2}} \int W_i W_i' (\Omega_i^{\frac{1}{2}})' \right)^{-1} \Omega_i^{\frac{1}{2}} \int W_i dW_{1i} = \sigma_i (\Omega_i^{-\frac{1}{2}})' P_i(W_i). \quad (4.B.8)$$

Similarly, under the null the Wald statistic is given by

$$w_{\hat{\delta}_{1i}} = \hat{\sigma}_i^{-2} \sum_{t=2}^T Q_W \varepsilon_{1.2i,t} (Q_W Z_{i,t}^+)' \left(\sum_{t=2}^T Q_W Z_{i,t-1}^+ (Q_W Z_{i,t-1}^+)' \right)^{-1} \sum_{t=2}^T Q_W Z_{i,t-1}^+ (Q_W \varepsilon_{1.2i,t}).$$

Consider $\hat{\sigma}_i^2 = T^{-1} \sum_{t=2}^T (\Delta(Q_W Y_{i,t}) - \hat{\delta}'_{1i} (Q_W Z_{i,t-1}^+))'^2$. By making use of Lemma 4.B.1, and the fact that under the null, $\Delta(Q_W Y_{i,t}) = Q_W \varepsilon_{1.2i,t}$, we get

$$\begin{aligned} \hat{\sigma}_i^2 &= T^{-1} \sum_{t=2}^T (\Delta(Q_W Y_{i,t}) - \hat{\delta}'_{1i} (Q_W Z_{i,t-1}^+))'^2 \\ &= T^{-1} \sum_{t=2}^T (Q_W \varepsilon_{1.2i,t})'^2 - 2\hat{\delta}'_{1i} T^{-1} \sum_{t=2}^T Q_W Z_{i,t-1}^+ (Q_W \varepsilon_{1.2i,t}) \\ &\quad + \hat{\delta}'_{1i} T^{-1} \sum_{t=2}^T Q_W Z_{i,t-1}^+ (Q_W Z_{i,t-1}^+)' \hat{\delta}_{1i} \\ &= T^{-1} \sum_{t=2}^T (Q_W \varepsilon_{1.2i,t})'^2 + O_p(T^{-1}) O_p(1) + T^{-1} O_p(T^{-1}) O_p(T^2) O_p(T^{-1}) \\ &= T^{-1} \sum_{t=2}^T (Q_W \varepsilon_{1.2i,t})'^2 + O_p(T^{-1}). \end{aligned} \tag{4.B.9}$$

As for the first term on the right-hand side, we have

$$\begin{aligned} \sum_{t=2}^T (Q_W \varepsilon_{1.2i,t})'^2 &= \sum_{t=2}^T \varepsilon_{1.2i,t}^2 - \sum_{t=2}^T \varepsilon_{1.2i,t} W'_{i,t} \left(\sum_{t=2}^T W_{i,t} W'_{i,t} \right)^{-1} \sum_{t=2}^T W_{i,t} \varepsilon_{1.2i,t} \\ &= \sum_{t=2}^T \varepsilon_{1.2i,t}^2 + O_p(\sqrt{T}) O_p(T^{-1}) O_p(\sqrt{T}). \end{aligned}$$

Hence, by inserting this into (4.B.9), and then taking the limit as $T \rightarrow \infty$, we obtain

$$\hat{\sigma}_i^2 = T^{-1} \sum_{t=2}^T \varepsilon_{1.2i,t}^2 + O_p(T^{-1}) \xrightarrow{p} \sigma_i^2. \tag{4.B.10}$$

Combining this result with Lemma 4.B.1 (b) and (c) we get the following limit as $T \rightarrow \infty$

$$\begin{aligned} w_{\hat{\delta}_{1i}} &\xrightarrow{w} \sigma_i^{-2} \sigma_i \int dW_{1i} W'_i (\Omega_i^{\frac{1}{2}})' \left(\Omega_i^{\frac{1}{2}} \int W_i W'_i (\Omega_i^{\frac{1}{2}})' \right)^{-1} \sigma_i \Omega_i^{\frac{1}{2}} \int W_i dW_{1i} \\ &= p_i(W_i)' P_i(W_i), \end{aligned} \tag{4.B.11}$$

which establishes part (a) of the theorem.

Consider (b). Under the null hypothesis, $\Delta(Q_S Y_{i,t}) = Q_S \varepsilon_{1.2i,t}$. By using this result, (4.B.8) and the rules

for partitioned regressions, we obtain as $T \rightarrow \infty$

$$\begin{aligned} T\hat{\alpha}_{1i} &= \left(T^{-2} \sum_{t=2}^T (Q_S Y_{i,t-1})^2 \right)^{-1} T^{-1} \sum_{t=2}^T Q_S Y_{i,t-1} (Q_S \varepsilon_{1,2i,t}) \\ &\xrightarrow{w} \sigma_i \omega_{11 \cdot 2i}^{-1} P_i(U_i) + \omega_{11 \cdot 2i}^{-1} \omega_{11i} \rho_i' (\Omega'_{22i})^{-1} P_i(W_{2 \cdot i}) (P_i(U_i) - 1) = d_i. \end{aligned} \quad (4.B.12)$$

Next, consider

$$\text{var}(\hat{\alpha}_{1i}) = \hat{\sigma}_i^2 \left(T^{-2} \sum_{t=2}^T (Q_S Y_{i,t-1})^2 \right)^{-1}.$$

We have already shown that $\hat{\sigma}_i^2 \xrightarrow{p} \sigma_i^2$ as $T \rightarrow \infty$. From this result and arguments similar to those used in the proof of Lemma 4.B.1 we obtain as $T \rightarrow \infty$

$$\begin{aligned} T^2 \text{var}(\hat{\alpha}_{1i}) &\xrightarrow{w} \sigma_i^2 \omega_{11 \cdot 2i}^{-2} V(U_i) \\ &+ \sigma_i^2 \omega_{11 \cdot 2i}^{-2} \omega_{11i} V(U_i) (\rho_i' (\Omega'_{22i})^{-1} P_i(W_{2 \cdot i}) + V(W_{2 \cdot i}) p_i(W_{2 \cdot i})' \Omega_{22i}^{-1} \rho_i) \\ &+ \sigma_i^2 \omega_{11 \cdot 2i}^{-2} \omega_{11i}^2 \rho_i' (\Omega'_{22i})^{-1} \\ &\times (V(W_{2 \cdot i}) + P_i(W_{2 \cdot i}) V(U_i) V(W_{2 \cdot i}) p_i(W_{2 \cdot i})') \Omega_{22i}^{-1} \rho_i = D_i. \end{aligned} \quad (4.B.13)$$

The proof is completed by noting that

$$D_{i,\tau}^\Omega = \lim_{T \rightarrow \infty} \frac{T\hat{\alpha}_{1i}}{\sqrt{T^2 \text{var}(\hat{\alpha}_{1i})}} = \frac{d_i}{\sqrt{D_i}}. \quad (4.B.14)$$

□

Proof of Corollary 4.1

If $X_{i,t}$ is strongly exogenous, $B_{21i} = 0$ such that $\rho_i = 0$. Thus, (4.B.14) simplifies to

$$D_{i,\tau} = \frac{d_i}{\sqrt{D_i}} = \frac{\tilde{B}_{11i} P_i(U_i)}{\sqrt{\tilde{B}_{11i} V(U_i)}} = \frac{P_i(U_i)}{\sqrt{V(U_i)}}. \quad (4.B.15)$$

This completes the proof. □

4.B.3 Proof of Theorem 4.3

We begin with the following lemma.

Lemma 4.B.2 *Under H_{0i} and Assumptions 4.1 to 4.6, as $N, T \rightarrow \infty$*

- (a) $T^{-\frac{1}{2}} \bar{Z}_t \xrightarrow{w} M_3 W_3,$
- (b) $T^{-\frac{1}{2}} \tilde{Z}_{i,t-1}^+ \xrightarrow{w} \tilde{\Omega}_i^{\frac{1}{2}} W_i,$

$$(c) \quad T^{-2} \sum_{t=2}^T Q_{\tilde{W}} \tilde{Z}_{i,t}^+ (Q_{\tilde{W}} \tilde{Z}_{i,t}^+)' \xrightarrow{w} \tilde{\Omega}_i^{\frac{1}{2}} \left(\int W_i W_i' \right) (\tilde{\Omega}_i^{\frac{1}{2}})',$$

$$(d) \quad T^{-1} \sum_{t=2}^T Q_{\tilde{W}} \tilde{Z}_{i,t-1}^+ (Q_{\tilde{W}} \tilde{\varepsilon}_{1 \cdot 2i,t}) \xrightarrow{w} \sigma_i \tilde{\Omega}_i^{\frac{1}{2}} p_i(W_i),$$

where $M_3 = \lim_{N \rightarrow \infty} \overline{M}_3$, $\overline{M}_3 = \frac{1}{N} \sum_{i=1}^N M_{3i}$,

$$M_{3i} = \begin{pmatrix} \tilde{B}_{11 \cdot 2i}^{-1} (B_{12i} \tilde{B}_{22i}^{-1} (B_{23i} \tilde{B}_{33i}^{-1} + \lambda'_{2i}) + B_{13i} \tilde{B}_{33i}^{-1}) \\ \tilde{B}_{22i}^{-1} (B_{23i} \tilde{B}_{33i}^{-1} + \lambda'_{2i}) + \tilde{B}_{22i}^{-1} B_{21i} \tilde{B}_{11 \cdot 2i}^{-1} (B_{12i} \tilde{B}_{22i}^{-1} (B_{23i} \tilde{B}_{33i}^{-1} + \lambda'_{2i}) + B_{13i} \tilde{B}_{33i}^{-1}) \end{pmatrix},$$

$$M_{4i} = \begin{pmatrix} \tilde{B}_{22i}^{-1} \Sigma_{22i}^{\frac{1}{2}} & \tilde{B}_{22i}^{-1} (B_{23i} \tilde{B}_{33i}^{-1} + \lambda'_{2i}) \\ 0 & M_3 \end{pmatrix},$$

$\tilde{\rho}_i = \begin{pmatrix} B'_{21i} (\tilde{B}_{22i}^{-1})' & 0 \end{pmatrix}$ and

$$\begin{aligned} \tilde{\Omega}_i^{\frac{1}{2}} &= \begin{pmatrix} \sigma_i \tilde{B}_{11 \cdot 2i}^{-1} & \tilde{B}_{11 \cdot 2i}^{-1} M_{1i} M_{2i} \\ \sigma_i \tilde{B}_{11 \cdot 2i}^{-1} \tilde{\phi}_i & (M_{4i} + \tilde{B}_{11 \cdot 2i}^{-1} \tilde{\phi}_i M_{1i} M_{2i}) \end{pmatrix} \\ &= \begin{pmatrix} \tilde{\omega}_{11i} & \tilde{\Omega}_{12i} \\ \tilde{\Omega}_{21i} & \tilde{\Omega}_{22i} \end{pmatrix}. \end{aligned}$$

Furthermore, we have $\tilde{\Omega}_{21i} = \tilde{\omega}_{11i} \tilde{\rho}_i$ and we define $\tilde{\omega}_{11 \cdot 2i} = \tilde{\omega}_{11i} - \tilde{\omega}_{11i} \tilde{\Omega}_{12i} \tilde{\Omega}_{22i}^{-1} \tilde{\rho}_i$.

Proof of Lemma 4.B.2

Letting $\varphi_i = \tilde{B}_{11 \cdot 2i}^{-1} B_{12i}(1) \tilde{B}_{22i}(1)^{-1}$, we have

$$\begin{aligned} T^{-\frac{1}{2}} \bar{Z}_t &= \overline{M}_3 T^{-\frac{1}{2}} \sum_{s=1}^t \varepsilon_{3s} + \frac{1}{N} \sum_{i=1}^N \begin{pmatrix} \tilde{B}_{11 \cdot 2i}^{-1} & \varphi_i \\ \varphi_i' & (\tilde{B}_{22i}(1)^{-1} + \tilde{B}_{22i}(1)^{-1} B_{21i}(1) \varphi_i) \end{pmatrix} \\ &\times T^{-\frac{1}{2}} \sum_{s=1}^t \begin{pmatrix} \varepsilon_{1 \cdot 2i,s} \\ \varepsilon_{2i,s} \end{pmatrix} + O_p \left(\frac{1}{\sqrt{NT}} \right) + O_p \left(\frac{1}{\sqrt{T}} \right), \end{aligned}$$

from which it follows that as $N, T \rightarrow \infty$

$$\begin{aligned} T^{-\frac{1}{2}} \bar{Z}_t &= \overline{M}_3 T^{-\frac{1}{2}} \sum_{s=1}^t \varepsilon_{3s} + O_p \left(\frac{1}{\sqrt{N}} \right) + O_p \left(\frac{1}{\sqrt{NT}} \right) + O_p \left(\frac{1}{\sqrt{T}} \right) \\ &\xrightarrow{w} M_3 W_3. \end{aligned} \tag{4.B.16}$$

This proves (a).

Moreover, by combining (a) with (4.B.5), we find that as $N, T \rightarrow \infty$

$$T^{-\frac{1}{2}} \tilde{Z}_{i,t}^+ \xrightarrow{w} \tilde{\Omega}_i^{\frac{1}{2}} W_i,$$

which proves (b).

Analogous to the prove of Lemma 4.B.1, we have

$$\begin{aligned} \sum_{t=2}^T Q_{\tilde{W}} \tilde{Z}_{i,t-1}^+ (Q_{\tilde{W}} \tilde{Z}_{i,t-1}^+)' &= \sum_{t=2}^T \tilde{Z}_{i,t-1}^+ (\tilde{Z}_{i,t-1}^+)' - \sum_{t=2}^T \tilde{Z}_{i,t-1}^+ (\tilde{W}_{i,t})' \left(\sum_{t=2}^T \tilde{W}_{i,t} (\tilde{W}_{i,t})' \right)^{-1} \\ &\times \sum_{t=2}^T \tilde{W}_{i,t} (\tilde{Z}_{i,t-1}^+)' \\ &= \sum_{t=2}^T \tilde{Z}_{i,t-1}^+ (\tilde{Z}_{i,t-1}^+)' + O_p(T). \end{aligned}$$

Combining this with (b) we obtain as $N, T \rightarrow \infty$

$$\begin{aligned} T^{-2} \sum_{t=2}^T Q_{\tilde{W}} \tilde{Z}_{i,t-1}^+ (Q_{\tilde{W}} \tilde{Z}_{i,t-1}^+)' &= T^{-2} \sum_{t=2}^T \tilde{Z}_{i,t-1}^+ (\tilde{Z}_{i,t-1}^+)' + O_p(T^{-1}) \\ &\xrightarrow{w} \tilde{\Omega}_i^{\frac{1}{2}} \left(\int W_i W_i' \right) (\tilde{\Omega}_i^{\frac{1}{2}})'. \end{aligned} \quad (4.B.17)$$

This proves (c).

Finally,

$$\begin{aligned} \sum_{t=2}^T Q_{\tilde{W}} \tilde{Z}_{i,t-1}^+ (Q_{\tilde{W}} \tilde{\varepsilon}_{1 \cdot 2i,t}) &= \sum_{t=2}^T \tilde{Z}_{i,t-1}^+ \tilde{\varepsilon}_{1 \cdot 2i,t} - \sum_{t=2}^T \tilde{Z}_{i,t-1}^+ (\tilde{W}_{i,t})' \left(\sum_{t=2}^T \tilde{W}_{i,t} (\tilde{W}_{i,t})' \right)^{-1} \\ &\times \sum_{t=2}^T \tilde{W}_{i,t} (\tilde{\varepsilon}_{1 \cdot 2i,t}) \\ &= \sum_{t=2}^T \tilde{Z}_{i,t-1}^+ \tilde{\varepsilon}_{1 \cdot 2i,t} + O_p(T) O_p(T^{-1}) O_p(\sqrt{T}). \end{aligned}$$

Thus,

$$\begin{aligned} T^{-1} \sum_{t=2}^T Q_{\tilde{W}} \tilde{Z}_{i,t-1}^+ (Q_{\tilde{W}} \tilde{\varepsilon}_{1 \cdot 2i,t}) &= T^{-1} \sum_{t=2}^T \tilde{Z}_{i,t-1}^+ \tilde{\varepsilon}_{1 \cdot 2i,t} + O_p(T^{-\frac{1}{2}}) \\ &\xrightarrow{w} \sigma_i \tilde{\Omega}_i^{\frac{1}{2}} p_i(W_i) \end{aligned} \quad (4.B.18)$$

as $N, T \rightarrow \infty$. This proves (d) and hence the proof of Lemma 4.B.2 is complete. \square

The proof of Theorem 4.3 follows similar arguments as the proof of Theorem 4.2. However, if $k < m + 1$, $\tilde{\Omega}_i^{\frac{1}{2}}$ and $\tilde{\Omega}_{22i}$ are no longer square matrices such that we have to make use of generalized inverse in that case.

The Wald statistic $w_{\tilde{\delta}_{1i}}$ is given by

$$w_{\tilde{\delta}_{1i}} = \check{\sigma}_i^{-2} \sum_{t=2}^T Q_{\tilde{W}} \tilde{\varepsilon}_{1 \cdot 2i,t} (Q_{\tilde{W}} \tilde{Z}_{i,t}^+)' \left(\sum_{t=2}^T Q_{\tilde{W}} \tilde{Z}_{i,t-1}^+ (Q_{\tilde{W}} \tilde{Z}_{i,t-1}^+)' \right)^{-1} \quad (4.B.19)$$

$$\times \sum_{t=2}^T Q_{\tilde{W}} \tilde{Z}_{i,t-1}^+ (Q_{\tilde{W}} \tilde{\varepsilon}_{1 \cdot 2i,t}), \quad (4.B.20)$$

where $\tilde{\sigma}_i^2$ is $\hat{\sigma}_i^2$ with $Q_{\tilde{W}}$ in place of Q_W . By using the same steps as for $\hat{\sigma}_i^2$ in Theorem 4.2, we obtain $\tilde{\sigma}_i^2 \xrightarrow{p} \sigma_i^2$ as $N, T \rightarrow \infty$. This result, together with Lemma 4.B.2 (c) and (d), implies that as $N, T \rightarrow \infty$

$$\begin{aligned} w_{\tilde{\delta}_{1i}} &\xrightarrow{w} \sigma_i^{-2} \sigma_i \int dW_{1i} W_i' (\tilde{\Omega}_i^{\frac{1}{2}})' \left(\tilde{\Omega}_i^{\frac{1}{2}} \int W_i W_i' (\tilde{\Omega}_i^{\frac{1}{2}})' \right)^{-1} \sigma_i \tilde{\Omega}_i^{\frac{1}{2}} \int W_i dW_{1i} \\ &= p_i(W_i)' P_i(W_i), \end{aligned} \quad (4.B.21)$$

which establishes the required result for (a).

Furthermore, similarly to the prove of Theorem 4.2, by the rules for partitioned regressions, $T\tilde{\alpha}_{1i} \xrightarrow{w} \tilde{d}_i$ and $T^2 \text{var}(\tilde{\alpha}_{1i}) \xrightarrow{w} \tilde{D}_i$ as $N, T \rightarrow \infty$, where \tilde{d}_i and \tilde{D}_i are defined similarly to d_i and D_i above, but replacing ω_{11i} , $\omega_{11 \cdot 2i}$ and Ω_{22i} with $\tilde{\omega}_{11i}$, $\tilde{\omega}_{11 \cdot 2i}$ and $\tilde{\Omega}_{22i}$ respectively. This yields the required result for (b). \square

Proof of Corollary 4.2

The proof of Corollary 4.2 is completed by noting that $\tilde{\rho}_i = 0$ if $X_{i,t}$ is strongly exogenous. \square

A Panel Cointegration Study of the Euro Effect on Trade: Revisiting Bun and Klaassen (2007)

5.1 Introduction

The gravity model of trade has been widely used in the empirical literature to study the effect of various policy decisions (see e.g. Tinbergen, 1962, for an early application). In its original form the gravity model explains trade flows between country pairs as being proportional to their national income and inverse proportional to their distance. It has been successful in empirical studies of the effect of various variables on trade and also received several theoretical foundations (see e.g. Anderson, 1979; Bergstrand, 1985, 1989). The original gravity model has been augmented by numerous additional variables. Population size or GDP per capita have been added as additional measures of mass. Features of the geographic location such as longitude/latitude or dummy variables for landlocked or island nations. Furthermore, variables measuring whether country pairs share certain aspects have been added, such as common border, common language or a common colonial history.

One particular area of interest is the impact of currency unions on trade. Rose (2000) first estimates the effect of currency unions on trade in a cross-sectional study and finds that adopting a currency union leads to a 200% increase in bilateral trade. Glick and Rose (2002) and Frankel and Rose (2002) perform panel data studies of the common currency effect on trade and obtain similarly large estimates. It is also found that the increase in trade between countries sharing a common currency is not due to a deviation of trade from other partners but due to an increase in total trade. Frankel and Rose (2002) also investigate the effect of trade on income and find that a 1% increase in total trade leads to an increase in income per capita by about one-third of a percent. These very high estimates have led to a controversy in the empirical literature. In particular, most observations on currency unions in Rose's data comes from poorer or small countries or dependencies. This has led to the question whether the result applies to bigger countries such as the members of the European Monetary Union (EMU) (see the discussion in e.g. Glick and Rose, 2002; Frankel and Rose, 2002).

In 1999, 11 countries of the EMU (Austria, Belgium, Finland, France, Germany, Ireland, Italy, Luxembourg, the Netherlands, Portugal and Spain) introduced the Euro as a common currency. Greece joined the Euro in 2001, after the initial launch but before the introduction of Euro notes and coins. Since then, also Slovenia, Cyprus, Malta and Slovakia have joined the Euro. At the same time, other member states of the European Union (EU) have decided not to adopt the Euro, namely Denmark, Sweden and the United Kingdom. It is therefore an important question if there is a Euro effect on bilateral trade, i.e if there is an economic reward for joining the common currency. Furthermore, the introduction of the Euro itself was preceded by the European Monetary System (EMS) aimed at aligning the European exchange rates. The EMS might have had a mitigating effect on the introduction of the Euro on bilateral trade. But it might be interesting to investigate whether there is an additional benefit of a common currency over (relative) exchange rate stability. Early studies report an Euro effect between 5% and 40%. Micco, Stein, and Ordoñez (2003) estimate an increase in bilateral trade due to the introduction of the Euro between 8% and 16% when compared to trade between non-EMU countries. Flam and Nordström (2003) estimate the Euro effect between 5% and 15%. Bun and Klaassen (2002) estimate a dynamic panel data model of the gravity equation and find a long-run Euro effect of about 38%, where the immediate effect is estimated at about 3.9% increase in 1999. Rose and Stanley (2005) perform a meta analysis of the results of 34 studies of the effect of currency unions on trade. Although they find evidence of publication bias, they also find evidence that currency unions have a significant positive effect on bilateral trade, and obtain a combined estimate of the trade effect between 30% and 90%.

Bun and Klaassen (2007) estimate the gravity equation allowing for country pair specific time trends to account trending behavior observed in the residuals. This reduces the estimate of the Euro effect on bilateral trade to about 3%. They also investigate whether the data is non-stationary and find unit roots in the (log of) bilateral trade, GDP and GDP per capita using the panel unit root test of Harris and Tzervalis (1999) and the panel stationarity test of Hadri (2000). Furthermore, they find cointegration between these variables using the panel cointegration test of Pedroni (1999) and estimate the coefficients using the dynamic OLS (DOLS) estimator of Mark and Sul (2003). However, the employed methods assume that the data is cross-sectionally independent, an assumption unlikely to hold bilateral trade data.

Cross-sectional dependence has received much attention recently in the literature on non-stationary panel data as the assumption of cross-sectional independence is unlikely to hold in many data sets. It is of particular interest here as bilateral trade data by construction is highly cross-

sectionally related. Furthermore, the gravity model itself implies spatial dependence in the data due to the hypothesized effect of distance on trade. Several new panel unit root or cointegration tests have been proposed that allow from cross-sectional dependence in the form of common factors. See for example Breitung and Pesaran (2008) for an overview of the literature and Gengenbach et al. (2009) for a comparison of panel unit root tests. We use the panel unit root tests proposed by Pesaran (2007), Moon and Perron (2004), Breitung and Das (2008), Sul (2007), Bai and Ng (2004b) and Palm et al. (2008) to test whether the variables entering the gravity model are non-stationary. We then test for cointegration between the variables using the panel no-cointegration test proposed by Gengenbach et al. (2006) and the panel no error-correction tests suggested by Gengenbach et al. (2008) to investigate whether the variables are cointegrated. Both approaches allow for persistent cross-sectional dependence in the data in form of unobserved common factors. We estimate the cointegrating vector using the CUP estimator of Bai et al. (2009) and the CCEP estimator of Pesaran (2006) and obtain an estimate of the Euro effect on bilateral trade.

The remainder of this chapter is organized as follows. Section 5.2 describes the data, summarizes the main findings of Bun and Klaassen (2007) and presents a brief overview of other studies of the Euro effect on trade. Section 5.3 presents the results of the panel unit root and panel cointegration tests. We obtain estimates of parameters of the cointegrated gravity model in Section 5.4. Section 5.5 concludes.

5.2 Data and previous studies

We use the data set of Bun and Klaassen (2007) which contains data on all bilateral combinations in a panel of 19 countries, namely the 15 member countries of the EU prior to the 2004 expansion as well as Norway, Switzerland, Canada, Japan and the US. The data for Belgium and Luxembourg is combined because trade data for these countries is only available for the Belgium-Luxembourg Economic Union. The data spans the time period between 1967 and 2002. Thus we have a balance panel with $N=171$ country pairs and $T=36$ time series observations.

The data set includes the following variables. $TRADE_{ijt}$ is the log of real bilateral trade between countries i and j at time t , where real bilateral trade is measured as the sum of nominal bilateral exports and imports in US dollars divide by the US producer price index. GDP_{ijt} is the log of the product of countries' real GDP. $GDPCAP_{ijt}$ measures the log of the product of the countries' real GDP per capita. Furthermore, 2 dummy variables are included in the data, namely $EURO_{ijt}$ which is 1 if both countries have adopted the Euro at time t and FTA_{ijt} which is 1 if both countries have

a free trade agreement at time t . The model estimated by Bun and Klaassen (2007) is given by the following equation.

$$TRADE_{ijt} = \beta_1 GDP_{ijt} + \beta_2 GDPCAP_{ijt} + \delta_1 EURO_{ijt} + \delta_2 FTA_{ijt} + \eta_{ij} + \tau_{ij} \cdot t + \lambda_t + \varepsilon_{ijt}, \quad (5.2.1)$$

where η_{ij} is a country pair specific fixed effect, λ_t is a common time effect, $t_{ij} \cdot t$ is a country pair specific time trend and ε_{ijt} is the error term. The current model does not include distance between countries as a dependent variable. Nevertheless, country pair specific fixed effects will account for part of the distance effect, and any time invariant measure of distance would be removed by the within transformation. Furthermore, Pesaran and Tosetti (2009) show that cross-sectional averages are well suited to account for spatial dependence.

Table 5.1: Estimation results Bun and Klaassen (2007)

	LSDV		DOLS	
	no trends	trends	no trends	trends
δ_1	0.410	0.032	0.374	0.034
δ_2	0.41	0.06	0.38	0.05
β_1	1.41	0.70	0.59	0.94
β_2	-0.68	-0.23	0.20	-0.49

Notes: “No trends” indicates that τ_{ij} is set to 0 in (5.2.1). LSDV gives results for LSDV-type estimates from Bun and Klaassen (2007, Table 2, p. 480). DOLS give the estimates from Bun and Klaassen (2007, Table 5, p. 491).

Table 5.1 presents the estimates of the parameters of Equation (5.2.1) obtained by Bun and Klaassen (2007). When not allowing for time trends the estimated Euro effects are 51% and 45% for the LSDV and DOLS estimators, respectively. These estimates are above effects reported in earlier studies. However, the effect is reduced to only about 3% for both estimators when time trends are included in the model. The reduced estimate is robust to various other specifications employed by Bun and Klaassen (2007).

Similar to the Euro effect, the estimated effect of a free trade agreement between countries is also reduced when accounting for time trends in the estimation. The coefficient of GDP_{ijt} is positive, but there are some differences in the size of the estimated effect, ranging from 0.59 to 1.41. The coefficient of $GDPCAP_{ijt}$ is negative for 3 out of the 4 case.

Bun and Klaassen (2002) estimate a dynamic gravity model for real exports. They have annual data ranging from 1965 to 2001 for the 15 EU member countries before the 2004 expansion as well as on Canada, Japan and the US. They estimate an ADL model using the LSDV estimator and obtain

a Euro effect estimate of about 4% in the first year with an estimated long-run effect of about 40%. However, they do not consider models with deterministic or stochastic trends. Flam and Nordström (2003) estimate the Euro effect in a panel of 20 OECD countries with annual observations from 1989 to 2002. The panel includes data on the 10 countries (combining the data on Belgium and Luxembourg) that entered the EMU in 1999 as well as Australia, Canada, Denmark, Japan, New Zealand, Norway, Sweden, Switzerland, the UK and the US. They detect a break associated with the introduction of the Euro already in 1998. For the period of 1998 to 2002, they estimate an average increase in exports between EMU countries of about 15% compared to average exports in the 1989 to 1997 period. Flam and Nordström (2003) also detect a spill-over effect of the Euro. They estimate an increase in exports from the Euro zone to non-Euro countries of about 8% and an increase in exports to the Euro zone from non-Euro countries of approximately 5%. Flam and Nordström (2003) also analyze sector specific data and find significant Euro concentrated to a few sectors. Micco et al. (2003) consider two data sets in their analysis: a panel of 22 industrialized countries included in the IMF's Directions of Trade Statistics data set and a panel including only the 15 EU member countries prior to the 2004 expansion. Using different specification of the gravity equation, Micco et al. (2003) estimate a Euro effect between 4% and 26%. Furthermore, they also identify the Euro having an effect on trade starting in 1998, similar to Flam and Nordström (2003). They also estimate a spill-over effect of the Euro, increasing trade between Euro zone countries and non-Euro countries by up to 9%.

5.3 Panel unit root and cointegration tests

In this section we test for unit roots in the variables of the model and then test for cointegration between the non-stationary variables. We employ the tests proposed by Pesaran (2007), Moon-Perron (2004), Breitung and Das (2008), Sul (2007), Palm et al. (2008) and Bai and Ng (2004b) to test for unit roots. We then use the no cointegration test advanced by Gengenbach et al. (2006) and the no error correction tests proposed by Gengenbach et al. (2008) to investigate whether the variables are cointegrated. We briefly outline the test procedures before presenting the results.

5.3.1 Panel Unit Root Tests

Considers a heterogenous, linear model for a balanced panel with N cross sectional units and T times series observations. In particular,

$$Y_{i,t} = (1 - \rho_i)\mu_i + \rho_i Y_{i,t-1} + u_{i,t}, \quad (5.3.1)$$

where the error term u_{it} has a common factor structure, such that

$$u_{i,t} = \gamma_i f_t + e_{i,t}.$$

Here, f_t is an unobserved common factor, γ_i is the corresponding factor loading and $e_{i,t}$ is an idiosyncratic error term independent across i and independent of the common factor. It is convenient to re-write (5.3.1) as

$$\Delta Y_{i,t} = \alpha_{0i} + \alpha_{1i} Y_{i,t-1} + \gamma_i f_t + e_{i,t}, \quad (5.3.2)$$

where $\alpha_{01} = (1 - \rho_i)\mu_i$ and $\alpha_{1i} = (\rho_i - 1)$. Pesaran (2007) suggests to cross-sectionally augment the test equation (5.3.2) with cross-sectional averages of the first differences and the lagged levels to account for the cross-sectional dependence induced by a single common factor. The cross-sectionally augmented (CA)DF equation is then given by

$$\Delta Y_{i,t} = a_i + b_i Y_{i,t-1} + c_i \bar{Y}_{t-1} + d_i \Delta \bar{Y}_t + \varepsilon_{i,t}, \quad (5.3.3)$$

where $\bar{Y}_{t-1} = \sum_{i=1}^N Y_{i,t-1}$, $\Delta \bar{Y}_t = \sum_{i=1}^N \Delta Y_{i,t}$ and $\varepsilon_{i,t}$ is the regression error. The individual specific test statistic for the hypothesis $H_{0i} : \rho_i = 1$ for a given i is now the t-statistic of b_i in (5.3.3), denoted by $CADF_i$. The panel unit root for the hypothesis $H_0 : \rho_i = 1$ for all i against the heterogenous alternative $H_1 : \rho_i < 1$ for some i is given by the cross-sectional average of the $CADF_i$ tests, such that

$$\overline{CADF} = N^{-1} \sum_{i=1}^N CADF_i.$$

For computational reasons, Pesaran (2007) advocates the use of a truncated version, \overline{CADF}^* , where for positive constants K_1 and K_2 such that $Pr[-K_1 < CADF_i < K_2]$ is sufficiently large values of $CADF_i$ smaller than $-K_1$ or larger than K_2 are replaced by the respective bound. Pesaran (2007) provides values for K_1 and K_2 as well as critical values for the test statistics obtained via stochastic simulation. In case the error terms or the common factor are serially the CADF equation (5.3.3) can

be augmented by additional lags of \bar{Y}_{t-1} , $\Delta\bar{Y}_t$ and $\Delta Y_{i,t-1}$.

Moon and Perron (2004) propose two test statistics for the null hypothesis $H_0 : \rho_i = 1$ for all i against the heterogenous alternative $H_1 : \rho_i < 1$ for some i . They allow for k common factors in $u_{i,t}$. Their method relies on de-factoring the data by a projection onto the space orthogonal to that spanned by the factor loadings. They propose to estimate the factor loadings by method of principle components from the residuals of a pooled first stage regression,

$$\hat{u}_{i,t} = Y_{i,t} - \hat{\rho}_{pols} Y_{i,t-1},$$

where $\hat{\rho}_{pols}$ is the pooled OLS estimator of ρ_i in (5.3.1). The de-factored data is now given by

$$Y_{i,t}^* = Y_{i,t} - \hat{\gamma}_i \left(\sum_{j=1}^N \hat{\gamma}'_j \hat{\gamma}_j \right)^{-1} \sum_{j=1}^N \hat{\gamma}'_j Y_{j,t}.$$

The two test statistics suggested by Moon and Perron (2004) are based on a modified pooled estimator of ρ ,

$$\hat{\rho}^* = \left(\sum_{i=1}^N \sum_{t=2}^T (Y_{i,t-1}^*)^2 \right)^{-1} \left(\sum_{i=1}^N \sum_{t=2}^T Y_{i,t-1}^* Y_{i,t}^* - NT \hat{\varphi}_e \right),$$

where $\hat{\varphi}_e$ is the average estimated one-sided long-run covariance. The tests are given by

$$t_a^* = \frac{\sqrt{NT}(\hat{\rho}^* - 1)}{\sqrt{\frac{2\hat{\phi}_e^4}{\hat{\omega}_e^4}}} \quad (5.3.4)$$

and

$$t_b^* = \sqrt{NT}(\hat{\rho}^* - 1) \sqrt{\frac{1}{NT^2} \sum_{i=1}^N \sum_{t=2}^T (Y_{i,t-1}^*)^2 \left(\frac{\hat{\omega}_e}{\hat{\phi}_e^2} \right)}, \quad (5.3.5)$$

where $\hat{\omega}_e^2$ is the average estimated long-run covariance and $\hat{\phi}_e^4 = N^{-1} \sum_{i=1}^N \hat{\omega}_e^4$. Moon and Perron (2004) show that both test statistics have a standard normal limiting distribution.

Breitung and Das (2008) propose two tests for a unit root in (5.3.1), namely a generalized least squares (GLS) t-test, which is only feasible if $N < T$, and a robust t-test, t_{rob} . The later is given by

$$t_{rob} = \left(\sum_{t=2}^T Y'_{t-1} \hat{\Omega} Y_{t-1} \right)^{-\frac{1}{2}} \left(\sum_{t=2}^T Y_{t-1} \Delta Y_t \right), \quad (5.3.6)$$

where $Y_{t-1} = (Y_{1,t-1}, \dots, Y_{N,t-1})'$, $\Delta Y_t = (\Delta Y_{1,t}, \dots, \Delta Y_{N,t})'$ and $\hat{\Omega} = \sum_{t=2}^T \hat{u}_t \hat{u}_t'$, with $\hat{u}_t = (\hat{u}_{1,t}, \dots, \hat{u}_{N,t})'$ being the pooled OLS residuals. Breitung and Das (2008) show that t_{rob} converges to a Dickey-Fuller (DF) distribution under the null hypothesis $H_0 : \rho_i = 1$ for all i .

Palm et al. (2008) propose several bootstrap panel unit roots. They consider pooled Levin et al. (2002) type tests based on the pooled OLS estimate of ρ_i in (5.3.1) and group mean Im et al. (2003) type tests based on individual specific estimates of ρ_i . In particular, the pooled statistic is defined as

$$\tau_p = T(\hat{\rho}_{pols} - 1). \quad (5.3.7)$$

The group mean statistic is given by the following equation,

$$\tau_{gm} = N^{-1} \sum_{i=1}^N T(\hat{\rho}_i - 1), \quad (5.3.8)$$

where

$$\rho_i = \left(\sum_{t=2}^T Y_{i,t-1}^2 \right)^{-1} \left(\sum_{t=2}^T Y_{i,t-1} Y_{i,t} \right).$$

Palm et al. (2008) also consider τ_{med} which is given by T times the median of $(\hat{\rho}_i - 1)$, as the median might be more robust to outliers. Palm et al. (2008) propose a block bootstrap and show that it is asymptotically valid for a number of cross-sectional correlation models.

Bai and Ng (2004b) consider a more general model than (5.3.1). In particular,

$$Y_{i,t} = \gamma_i F_t + E_{i,t}, \quad (5.3.9)$$

where F_t is a k -vector common factor and $E_{i,t}$ is the idiosyncratic component. They allow either F_t or $E_{i,t}$ to be non-stationary and propose to test them separately. As both common and idiosyncratic components are unobserved, Bai and Ng (2004b) propose a consistent estimator. They apply the methods of principle components to the (demeaned) first differences of the data and re-accumulate the estimates to preserve the order of integration.

For the estimated idiosyncratic component, $\hat{E}_{i,t}$, Bai and Ng (2004b) propose an ADF test to test for individual unit roots. To test the pooled unit root hypothesis that all $\hat{E}_{i,t}$ are non-stationary, Bai and Ng (2004b) suggest a Fisher-type, using the correction proposed by Choi (2001) for the test of

Maddala and Wu (1999). In particular, the test statistic is given by

$$P_{\hat{E}}^{c,\tau} = \frac{-2 \sum_{i=1}^N \log \pi_i - 2N}{\sqrt{4N}}, \quad (5.3.10)$$

where π_i is the p-value of the ADF test for the i -th cross section and c and τ denote the constant only or linear deterministic trend case, respectively. Bai and Ng (2004b) show that $P_{\hat{E}}^{c,\tau}$ has a standard normal limiting distribution.

Bai and Ng (2004b) propose several tests to select the number of independent stochastic trends, k_1 in the estimated common factors, \hat{F}_t . If a single common factor is estimated, they propose an ADF test, $ADF_{\hat{F}}^{c,\tau}$. Bai and Ng (2004b) show that the limiting distribution of $ADF_{\hat{F}}^{c,\tau}$ coincides with the Dickey-Fuller distribution for the respective cases. If more than one common factor is estimated, Bai and Ng (2004b) propose an iterative procedure to select k_1 , similar to Johansen trace test for cointegration. Bai and Ng (2004b) propose two modified Q statistics to test the hypothesis of $k_1 = m$ against the alternative $k_1 < m$ for m starting from \hat{k} . The procedure terminates if at any step $k_1 = m$ cannot be rejected. The two test statistics are denoted as $MQ_e^{c,\tau}$ and $MQ_f^{c,\tau}$, where the former uses a non-parametric correction to account for additional serial correlation while the later employs a parametric correction. Both statistics have a non standard limiting distribution and Bai and Ng (2004b) provide critical values for several m .

Sul (2007) proposes recursive mean adjusted panel unit roots. He proposes a GLS test to test the hypothesis $H_0 : \rho_i = 1$ for all i against the heterogenous alternative $H_1 : \rho_i < 1$ for some i . However, the GLS test is not feasible if $T < N$. In case the data permits a Bai and Ng type representation as in (5.3.9), Sul (2007) proposes a recursive mean adjusted unit root test applied to the cross-sectional average of the data to test for a unit root in the common component. The test statistic is given by the FGLS t -test for $H_0 : \rho = 1$ in the following regression

$$\bar{Y}_t - \bar{C}_{t-1} = \rho(\bar{Y}_{t-1} - \bar{C}_{t-1}) + \sum_{j=1}^p \phi_j \Delta \bar{Y}_{t-j} + \varepsilon_{i,t}, \quad (5.3.11)$$

where $\bar{C}_{t-1} = \sum_{i=1}^N C_{i,t-1}$ with $C_{i,t-1} = (t-1)^{-1} \sum_{s=1}^{t-1} Y_{i,s}$. Sul (2007) provides simulated critical values for the test statistics, t_{crma} .

We apply the panel unit root tests described above to test for unit roots in $TRADE_{ijt}$, GDP_{ijt} and $GDPCAP_{ijt}$. The appropriate lag-lengths for tests is selected using the Akaike information criterion with a maximum $p = 4$. We use the Andrews and Monahan (1992) estimator employing the quadratic spectral kernel to estimate the nuisance parameters for the Moon and Perron (2004)

tests. The number of common factors for the Moon and Perron (2004) and Bai and Ng (2004b) test is estimated using the BIC_3 criterion of Bai and Ng (2002)¹ allowing for at most $k_{max} = 4$ factors. For the bootstrap tests of Palm et al. (2008)² we draw 10000 bootstrap samples. We use a fixed block length of $b = 6^3$. We allow for a linear trend in data.

The critical value for the \overline{CADF}^* test from is -2.56 at 5% level (see Pesaran, 2007, Table II(c)). With test statistic of -2.360, -1.778 and -1.827 for $TRADE_{ijt}$, GDP_{ijt} and $GDPCAP_{ijt}$, respectively, we cannot reject the null hypothesis for all 3 panels. Using the asymptotic critical value of -1.645, the t_a^* test of Moon and Perron (2004) can reject the unit root null for $GDPCAP_{ijt}$ with a statistic of -9.620. The t_b^* test reject the null in all three panels with values of -3.795, -11.402 and -11.982, respectively. The Breitung and Das (2008) t_{rob} test rejects the unit root null for $TRADE_{ijt}$ with a statistic of -4.606, using the asymptotic critical value of -3.41. Given the 5% asymptotic critical value of -1.86, the t_{crma} test of Sul (2007) rejects the unit root for all three panels. The $P_{\hat{E}}^r$ test of Bai and Ng (2004b) cannot reject the unit root null using the asymptotic critical value of 1.645 for the estimated idiosyncratic component of either $TRADE_{ijt}$, GDP_{ijt} or $GDPCAP_{ijt}$. Estimating a single common factor for $TRADE_{ijt}$, the $ADF_{\hat{F}}^r$ test does not reject the unit root. Estimating 4 common factors in each panel for GDP_{ijt} or $GDPCAP_{ijt}$, both MQ_c^r and MQ_f^r cannot reject the null hypothesis that there are 4 independent stochastic trends. The critical values for the two statistics are -40.442 and -48.421, respectively (see Bai and Ng, 2004b, Table I). The bootstrap panel unit root tests of Palm et al. (2008) cannot reject the unit root null in either of the three panels. For a block length of $b = 6$, the 5% bootstrap critical values for $TRADE_{ijt}$ are -12.491, -13.815 and -13.120 for τ_p , τ_{gm} and τ_{med} , respectively. For GDP_{ijt} , we obtain bootstrap critical values of -12.894, -13.627 and -13.475, while the critical values for the $GDPCAP_{ijt}$ panel are -12.777, -13.565 and -13.453.

As only the t_b^* of Moon and Perron (2004) and the t_{crma} of Sul (2007) are able to reject the unit root null for all three panels, there is strong evidence that the data is non-stationary⁴.

5.3.2 Panel Cointegration Tests

Gengenbach et al. (2006) consider the problem of testing for no cointegration in a balanced panel with N cross-sections and T time series observations. For each cross-sectional unit, a $1 + m$ vector $Z_{i,t} = (Y_{i,t}, X'_{i,t})'$ is observed. Gengenbach et al. (2006) assume that both $Y_{i,t}$ and $X_{i,t}$ allow a Bai

¹The results are robust to using other selecting criterions and selecting different numbers of common factors.

²The author would like to thank Stephan Smeekes for providing the GAUSS codes for the test procedures.

³The results are robust for various block lengths $b = 1, \dots, \frac{T}{2}$.

⁴Bayer and Hanck (2009) consider the possibility of combining several no cointegration tests. In principle, it should be possible to apply their method also to panel unit root tests.

Table 5.2: Panel unit root and cointegration tests

Panel unit root tests								
variable	\overline{CADF}^*	τ_p	τ_{gm}	τ_{med}	t_a^*	t_b^*	t_{rob}	t_{crma}
$TRADE_{ijt}$	-2.360	-8.936	-9.438	-8.309	-1.225	-3.795 [†]	-4.606 [†]	-6.315 [†]
GDP_{ijt}	-1.778	-5.730	-5.975	-5.842	-1.539	-11.402 [†]	-2.154	-2.322 [†]
$GDPCAP_{ijt}$	-1.827	-5.841	-6.064	-5.899	-9.620 [†]	-11.982 [†]	-2.120	-2.361 [†]
variable	$P_{\hat{E}}^\tau$	$ADF_{\hat{F}}^\tau$	MQ_c^τ	MQ_f^τ				
$TRADE_{ijt}$	-10.49	-2.560	-	-				
GDP_{ijt}	-11.12	-	-17.07	-20.83				
$GDPCAP_{ijt}$	-11.07	-	-16.23	-21.67				
Panel cointegration tests								
	Z_ρ^p	Z_ρ^{gm}	Z_t^p	Z_t^{gm}	L_{trace}^F	$\bar{\tau}_{\alpha_i}^*$	$\bar{w}_{\delta_i}^*$	
	0.501	5.249	-4.949 [†]	-5.425 [†]	52.87 [†]	-5.051 [†]	62.60 [†]	

Notes: [†] denotes rejection at 5% level. Critical values are -2.56 for \overline{CADF}^* , -1.645 for t_a^* and t_b^* , -1.86 for -4.040 for t_{crma} , -3.41 for t_{rob} and $ADF_{\hat{F}}^\tau$, 1.645 for $P_{\hat{E}}^\tau$, -40.442 for MQ_c^τ and -48.421 for MQ_f^τ . Bootstrap 5% critical values for τ_p are -12.491, -12.894 and -12.777 for $TRADE_{ijt}$, GDP_{ijt} and $GDPCAP_{ijt}$, respectively. Bootstrap 5% critical values are -13.815, -13.627 and -13.565 for τ_{gm} and -13.120, -13.475 and -13.453 for τ_{med} .

The critical value for Z_ρ^p , Z_ρ^{gm} , Z_t^p and Z_t^{gm} is -1.645. The critical value for L_{trace}^F is 27.169. Critical values are -4.040 for $\bar{\tau}_{\alpha_i}^*$ and 28.203 for $\bar{w}_{\delta_i}^*$.

and Ng type representation as in (5.3.9). They propose to estimate the common and idiosyncratic components of the panels using the principle component estimator of Bai and Ng (2004b). To test for no cointegration between the estimated idiosyncratic components, $\hat{E}_{i,t}^Y$ and $\hat{E}_{i,t}^X$, they suggest to use the panel no cointegration tests of Pedroni (1999). In particular, we consider the pooled and group mean coefficient test, Z_ρ^p and Z_ρ^{gm} , as well as the pooled and group mean t -test, Z_t^p and Z_t^{gm} . Gengenbach et al. (2006) propose the Johansen trace test to test for cointegration between the estimated common factors \hat{F}_t^Y and \hat{F}_t^X . Denote the statistic as L_{trace}^F . As discussed in Gengenbach et al. (2006), rejection of the null hypothesis of no cointegration for both the idiosyncratic component and the common factor is a necessary but not a sufficient condition for cointegration between $Y_{i,t}$ and $X_{i,t}$.

Gengenbach et al. (2008) propose tests for panel (no) error correction. Starting from the triangular representation of a cointegrated panel with non-stationary (unobserved) common factors, they derive the following conditional error correction model (ECM).

$$\Delta Y_{i,t} = \alpha_i Y_{i,t-1} + \gamma_{1i} X_{i,t-1} + \gamma_{2i} F_{t-1} + \sum_{s=1}^{p_i} \pi_{1is} \Delta Y_{i,t-s} + \sum_{s=0}^{p_i} \pi_{2is} \Delta X_{i,t-s} + \sum_{s=0}^{p_i} \pi_{3is} \Delta F_{t-s} + \varepsilon_{i,t}, \quad (5.3.12)$$

where $\gamma_{1i} = -\alpha_i \theta_{1i}$ and $\gamma_{2i} = -\alpha_i \theta_{2i}$ such that $(1, -\theta'_{1i}, -\theta'_{2i})'$ is the cointegrating vector, $X_{i,t}$ is a $m \times 1$ vector of idiosyncratic weakly exogenous variables, F_t is a $k \times 1$ vector of possibly unobserved,

strongly exogenous common factors and $\varepsilon_{i,t}$ is an *i.i.d.* error term. Gengenbach et al. (2008) consider 2 tests for the individual specific null hypothesis of no error correction, namely the t-statistic for $H_{0i}^t : \alpha_i = 0, \tau_{\alpha_i}$, and the Wald test for $H_{0i}^w : \delta_i = 0, w_{\delta_i}$, where $\delta_i = (\alpha_i, \gamma'_{1i}, \gamma'_{2i})'$. Following the ideas of Pesaran (2007), they propose to augment the conditional ECM (5.3.12) with cross-sectional averages of $Y_{i,t-1}$ and $X_{i,t-1}$ as well as the contemporaneous and lagged averages of $\Delta Y_{i,t}$ and $\Delta X_{i,t}$. The panel test statistics for the null hypothesis of no error correction for all i are given by the (truncated) averages of the individual specific statistics, denoted as $\bar{\tau}_{\alpha_i}^*$ and $\bar{w}_{\delta_i}^*$ for the t and Wald test, respectively. Gengenbach et al. (2008) provide critical values for both test statistics obtained via stochastic simulations.

We test for cointegration between $TRADE_{ijt}$, GDP_{ijt} and $GDPCAP_{ijt}$ using both test procedures outlined above. Using the BIC_3 criterion of Bai and Ng (2002)⁵ we find one common factor in $TRADE_{ijt}$ and three in the joint panel of GDP_{ijt} and $GDPCAP_{ijt}$. The coefficient based tests Z_ρ^p and Z_ρ^{gm} cannot reject the null of no cointegration between the estimated common factors. However, both t -tests, Z_t^p and Z_t^{gm} , can reject the null. The Johansen trace test, L_{trace}^F , applied to the 4 estimated common factors finds a single cointegrating relationship⁶. For the tests of panel no error correction, $\bar{\tau}_{\alpha_i}^*$ and $\bar{w}_{\delta_i}^*$, allowing for a constant and linear trend in the ECM, we select the lag length p_i using the Akaike information criterion with $pmax = 4$. The critical value for $\bar{\tau}_{\alpha_i}^*$ is -4.040 at 5% significance while the corresponding critical value for $\bar{w}_{\delta_i}^*$ is 28.203 (see Gengenbach et al., 2008, Tables 3 and 4). With statistics of -4.958 and 62.21 for the t and Wald test respectively we can reject the null of no error correct. As only the coefficient based tests for the estimated idiosyncratic components cannot reject the null of no cointegration, there is evidence that $TRADE_{ijt}$, GDP_{ijt} and $GDPCAP_{ijt}$ are cointegrated.

5.4 Estimation of the gravity equation

In the previous section we have found evidence that the variables entering the gravity equation are non-stationary and cointegrated. Therefore, equation (5.2.1) describes a long-run equilibrium relationship between $TRADE_{ijt}$, GDP_{ijt} and $GDPCAP_{ijt}$. In this section, we use the CUP estimator of Bai et al. (2009) and the CCEP estimator of Pesaran (2006) to obtain estimates of the parameters of the static long-run model given in (5.2.1). Furthermore, we estimate a dynamic error correction model with a CCEP estimator. All considered estimators allow for a heterogenous effect of the

⁵The results are qualitatively robust to using different criterions.

⁶We also select the cointegration rank using an information criteria advanced by Aznar and Salvador (2002) which finds 3 cointegrating relation ships.

common factors. Furthermore, we allow for heterogenous short-run dynamics when estimating the ECM.

Bai et al. (2009) consider the problem of estimating the cointegrating vector in a cointegrated panel data model with non stationary common factors. They consider the following model,

$$Y_{i,t} = \beta X_{i,t} + \gamma_i F_t + e_{i,t}, \quad (5.4.1)$$

where F_t is a k vector of common factors, γ_i is the corresponding vector of factor loadings and e_{it} is an idiosyncratic error term. Bai and Kao (2006) propose a 2-stage fully modified (FM) estimator of β in the case of stationary F_t . However, if F_t is non-stationary the least-squares (LS) estimator of β is inconsistent. Bai et al. (2009) propose a bias corrected (BC) and fully modified (FM) estimator for β for the case of observed F_t . However, those are infeasible in the case of unobserved common factors. The proposed solution is an iterative procedure where F_t is estimated given an estimate of β and then β is re-estimated given the estimate of F_t . The objective function of the procedure is

$$S_{NT}(\beta, F, \Gamma) = \sum_{i=1}^N \sum_{t=1}^T (Y_{i,t} - \beta X_{i,t} - \gamma_i F_t)^2, \quad (5.4.2)$$

which is minimized subject to the constraints $T^{-2} \sum_{t=1}^T F_t F_t' = I_k$ and $\Gamma' \Gamma$ is positive definite, where $\Gamma = (\gamma_1', \dots, \gamma_N)'$.

Now, given F_t the LS estimator of β is given by

$$\hat{\beta} = \left(\sum_{i=1}^N \sum_{t=1}^T Q_F X_{it} (Q_F X_{it})' \right)^{-1} \left(\sum_{i=1}^N \sum_{t=1}^T Q_F X_{it} Q_F Y_{it} \right), \quad (5.4.3)$$

where Q_F is the OLS projection error operator such that

$$Q_F X_{i,t} = X_{i,t} - \sum_{t=1}^T X_{i,t} F_t' \left(\sum_{t=1}^T F_t F_t' \right)^{-1} F_t.$$

Define

$$W_{i,t} = Y_{i,t} - \beta X_{i,t},$$

and $W_i = (W_{i,1}, \dots, W_{i,T})'$. The estimator of F_t given $\hat{\beta}$ is the given by the k largest eigenvectors of

the matrix $(NT^2)^{-1} \sum_{i=1}^N W_i W_i'$. Concentrating out Γ , the objective function can be rewritten as

$$S_{NT}(\beta, F) = \sum_{i=1}^N \sum_{t=1}^T (Q_F W_{i,t})^2.$$

The continuously updated (CUP) estimator is then given by

$$(\hat{\beta}_{CUP}, \hat{F}_{CUP}) = \arg \min S_{NT}(\beta, F).$$

The procedure outlined above requires that k , the number of common factors, is known. In general that is not the case and k has to be estimated. Bai and Ng (2002) propose several information criteria which can be used to obtain consistent estimates.

We obtain CUP estimates of β_1 and β_2 in (5.2.1) after concentrating out the fixed effects. We estimate the number of common factors using the BIC_3 criterion of Bai and Ng (2002) which performs well in empirical studies. With $k_{max} = 4$ we estimate a single common factor, such that $\hat{k} = 1$. We then obtain estimates of the long run parameters β_1 and β_2 and estimates of the coefficient of the two dummy variables $EURO_{ijt}$ and FTA_{ijt} which minimize (5.4.2).

Pesaran (2006) proposes a consistent estimator for the slope parameter β_i in a heterogenous panel data model similar to (5.4.1). He allows for both observed and unobserved common factors, D_t and F_t respectively. Furthermore, he assumes that $X_{i,t}$ also permits a common factor structure. In particular, his model is given by the following equations:

$$Y_{i,t} = \beta_i X_{i,t} + \alpha_i D_t + \gamma_i F_t + e_{i,t}, \quad (5.4.4)$$

$$X_{i,t} = A_i D_t + \Gamma_i F_t + v_{i,t}, \quad (5.4.5)$$

where $e_{i,t}$ and $v_{i,t}$ are idiosyncratic errors. Pesaran (2006) suggests to use the cross-sectional average of $Z_{i,t} = (Y_{i,t}, X'_{i,t})'$ as a proxy for the unobserved common factors. The pooled estimator for β is now give by

$$\hat{\beta}_{CCEP} = \left(\sum_{i=1}^N \sum_{t=1}^T Q_{D, \bar{Z}} X_{it} (Q_{D, \bar{Z}} X_{i,t})' \right)^{-1} \left(\sum_{i=1}^N \sum_{t=1}^T Q_{D, \bar{Z}} X_{it} Q_{D, \bar{Z}} Y_{i,t} \right). \quad (5.4.6)$$

Pesaran (2006) shows that β_{CCEP} is consistent for the mean of β_i . However, he only considers weakly stationary variables. Kapetanios, Pesaran, and Yamagata (2008) expand the analysis to allow for non-stationary common factors and show that the CCEP estimator remains consistent.

Table 5.3: Parameter estimates for static and dynamic models

	Static model estimates							
	Bun and Klaassen (2007)				current study			
	LSDV		DOLS		CCEP		CUP	
	no trends	trends	no trends	trends	no trends	trends	no trends	trends
$\hat{\beta}_1$	1.410	0.700	0.590	0.940	0.014	0.334	1.351	0.359
$\hat{\beta}_2$	-0.680	-0.230	0.200	-0.490	0.354	0.091	-0.710	0.179
$\hat{\delta}_1$	0.410	0.032	0.374	0.034	0.043	0.006	0.417	0.075
$\hat{\delta}_2$	0.410	0.060	0.380	0.050	0.033	0.011	0.418	0.016
Dynamic model estimates								
	Bun and Klaassen (2002)				CCEP for ECM			
	LSDV for ADL				$p = 0$		$\hat{p} = 4$	
	no trends	trends			no trends	trends	no trends	trends
$\hat{\alpha}$	-	-			-0.407	-0.487	-0.782	-0.785
$\hat{\gamma}_1$	-	-			0.303	0.340	0.257	0.021
$\hat{\gamma}_2$	-	-			-0.238	-0.232	0.170	0.574
$\hat{\delta}_1$	0.040	-			0.196	0.016	-0.030	0.012
$\hat{\delta}_2$	0.080	-			0.025	0.008	0.020	0.008
Long-run estimates								
	no trends	trends			no trends	trends	no trends	trends
$\hat{\beta}_1$	-	-			0.744	0.698	0.329	0.027
$\hat{\beta}_2$	-	-			-0.585	-0.476	0.217	0.731
$\hat{\delta}_1$	0.330	-			0.482	0.033	-0.038	0.015
$\hat{\delta}_2$	0.710	-			0.061	0.016	0.026	0.010

Notes: “No trends” indicates that τ_{ij} is set to 0. LSDV gives results for LSDV-type estimates from Bun and Klaassen (2007, Table 2, p. 480). DOLS give the estimates from Bun and Klaassen (2007, Table 5, p. 491). Dynamic model estimates for a stationary ADL are taken from Bun and Klaassen (2002, Table 1, p. 11), where results are reported for the European FTA dummy.

Results for Bai, Kao, and Ng’s CUP estimator are obtained with $\hat{k} = 1$ common factor, as selected by BIC_3 . For the ECM a lag length of $\hat{p} = 4$ is selected by the BIC.

Augmenting the gravity equation (5.2.1) with cross-sectional averages of $TRADE_{ijt}$, GDP_{ijt} and $GDPCAP_{ijt}$ we obtain CCEP estimates of β_1 and β_2 as well as of δ_1 and δ_2 .

Furthermore, we estimate a dynamic ECM as given in (5.3.12). We obtain CCEP-type estimates by including cross-sectional averages the lagged level and (lagged) first-differences of $TRADE_{ijt}$, GDP_{ijt} and $GDPCAP_{ijt}$ as a proxy for the common factors in the regression. Allowing for a maximum lag length of $p_{max} = 6$ we select an appropriate lag length of $\hat{p} = 4$ using the BIC.

Table 5.3 reports the obtained parameter estimates as well as the results obtained by Bun and Klaassen (2007, 2002)⁷ for direct comparison. For the ECM estimates results are sensitive to the

⁷Bun and Klaassen (2002) use a different data set. However, we include their results as a comparison for the

specified lag length. We report findings for the estimated lag length $\hat{p} = 4$ as well as for estimates obtained without allowing for additional short-run dynamics in the model ($p = 0$). Similarly to the findings of Bun and Klaassen (2007), the CUP estimate and the CCEP-ECM estimate for the case without short-run dynamics observed a strong drop in the estimated coefficients when allowing for country pair specific trends. Without trends, the CUP estimate of the Euro effect on trade is about 52% and the CCEP-ECM estimate even 62%. However, these estimates are reduced to 7.8% and 3.4%, respectively, in the trend case. The CCEP estimator for the static model finds a Euro effect on bilateral trade of about 4.4% when not allowing for trends which is reduced to only 0.6% in the trend case. The CCEP-ECM estimate in the model with short-run dynamics is even negative in the no trend case with an estimated long-run effect of about -3.7%. When allowing for trends, the effect is estimated at about 1.5%.

Estimates for β_1 and β_2 also vary between estimators and models. While the estimates of β_1 all have the expected sign, the static CCEP estimate in the no trend case and the CCEP-ECM estimate with short-run dynamics and trend are very small at 0.014 and 0.027, respectively. For these two estimators we observe a stronger positive effect of GDP per capita than GDP on trade, with estimates of β_2 at 0.354 and 0.731, respectively. In 3 cases we obtain negative estimates of β_2 , namely for the CUP estimator without trends and for the CCEP-ECM estimators without short-run dynamics. In those cases the estimate of β_2 is smaller than the estimate of β_1 in absolute value. Our estimates of the trade effect of a free trade area are in general smaller than previously reported estimates. We find a positive effect between 1% and 6.3%. Only the CUP estimator in the no trend case finds a strong effect of about 52%.

While the standard errors of δ_1 and δ_2 are relatively small and the estimated coefficients seem significant in most considered models, the standard errors of β_1 and β_2 are very large relative to the estimates. This is in contrast to (invalid) OLS standard errors, where all coefficients would be significant. The large standard errors of the pooled estimates seem to be due to a strong heterogeneity in the data with regards to the effect of GDP_{ijt} and GDP_{CAP}_{ijt} on $TRADE_{ijt}$. This would indicate that a pooled model is not correctly specified in the considered setting and that further control variables beyond fixed effects have to be included in the model as in e.g. Glick and Rose (2002).

5.5 Conclusion

Using the data of Bun and Klaassen (2007) we have estimated the Euro effect on bilateral trade using a cointegrated panel data model. Bilateral trade data is by construction strongly cross-sectionally correlated and we have allowed for persistent cross-sectional dependencies by allowing for (unobserved) common factors. Using several panel unit root tests, we have found strong evidence that the variables entering the gravity equation are non-stationary. However, $TRADE_{ijt}$, GDP_{ijt} and $GDPCAP_{ijt}$ seem to be cointegrated as indicated by the panel cointegration test of Gengenbach et al. (2006) and the panel error correction tests of Gengenbach et al. (2008). Using the CCEP estimator of Pesaran (2006) and the CUP estimator of Bai et al. (2009) we obtain estimates of the parameters of the static long-run model. We also obtain CCEP-type estimates for the parameter of a dynamic ECM. Our parameter estimates vary between models and estimators but seem to confirm the findings of Bun and Klaassen (2007) of a smaller Euro effect than previously estimated. Only the CUP and CCEP-ECM estimator find strong effects of the Euro on trade when not accounting for country pair specific trends.

Conclusion and Further Research

6.1 Summary, limitations and further research

In this thesis we have studied tests for non-stationary panels with persistent cross-sectional dependence. In particular, Chapter 2 has compared several panel unit root tests that account for cross-section dependence using a common factor structure. We have discussed the differences and similarities of the considered tests and studied their small sample properties in a Monte Carlo experiment. Only the tests proposed by Bai and Ng (2004b) and Sul (2007) allow to detect non-stationarity that is only due to non-stationary common factors. In this case, the individual series are pairwise cointegrated along the cross-sectional dimension. Test designed to detect non-stationarity in the common factors do not gain in power from the cross-sectional dimension of the panel. However, large N is required to obtain consistent estimates of the common and idiosyncratic components. Furthermore, tests that rely on estimating the number common factors are distorted if it is misspecified, but it seems less harmful to overestimate the number of common factors.

In Chapter 3 we considered the problem of testing for no cointegration in a panel with non-stationary common factors. We have shown that panel unit root tests proposed for independent panels (Kao, 1999; Pedroni, 1999, 2004a) diverge at rate \sqrt{N} in this case. We proposed a two-step procedure for testing for no cointegration in panels with common factors. First, the common factors and idiosyncratic components from the data are consistently estimated using the principal component estimator of Bai and Ng (2004b). Then, both data components are separately tested for no cointegration. We can distinguish the case of cross-member cointegration, where the observed non-stationarity is only due to non-stationary common factors and the case of non-stationarity on both data components. However, rejection of the null hypothesis for both common factor and idiosyncratic component is only a necessary, not a sufficient condition for panel cointegration. We discussed the additional homogeneity requirements on the cointegrating vector resulting from the presence of common factor cointegration.

In Chapter 4 we have derived a Granger-type representation theorem for a cointegrated panel with non-stationary common factors. Starting from the triangular representation of the model considered by Bai et al. (2009) we have obtained the MA, ARMA and ECM representations. Based on the conditional ECM, we proposed two tests for no error-correction which are panel extensions of the t and Wald tests proposed by Banerjee et al. (1998) and Boswijk (1994) in the time-series literature. While the individual test statistics are asymptotically nuisance parameter free and only depend on the number of non-stationary variable, they are not independent along the cross-section due to the presence of non-stationary common factors. Nevertheless, their averages converge to a random variable with a distribution that, while not analytically tractable, can be obtained via stochastic simulation. This makes pooled testing possible. We have provided simulated critical values for the proposed tests and obtained some simulation evidence that the tests perform quite well in small samples.

In Chapter 5 we have applied the methods discussed in the previous chapters in a study of the Euro effect on bilateral trade. We use the data set of Bun and Klaassen (2007) containing data on the 15 pre-2004 member states of the European Union, Norway, Switzerland, Canada, Japan and the US from 1967 to 2002. Applying the panel unit root tests discussed in Chapter 2 as well as bootstrap panel unit root tests of Palm et al. (2008) we obtained evidence that the variables entering the gravity equation are non-stationary. However, the no-cointegration and no-error correction tests proposed in Chapters 3 and 4, respectively, indicate that the variables are cointegrated. We obtained estimates of the parameters of the gravity equations using the CCEP estimator of Pesaran (2006) and the CUP estimator of Bai et al. (2009).

Throughout the thesis we have seen the importance of accounting for cross-sectional correlation. Furthermore, the results show the important implications of the way the common factors enter the model. Some approaches differ in terms of restrictions placed on the factor structure and thus in the way the common factors are allowed to influence the dynamic properties of the data. Furthermore, some tests apply to different data components. Consequently not all tests are robust to cross-member cointegration, which might be an important feature of the data in empirical studies.

While the test for the absence of cointegration proposed in Chapter 3 is robust to cross-member cointegration, the parameter restrictions on the cointegrating vector resulting from the presence of common factor cointegration require additional testing. Thus, it is not possible to say whether cointegration holds between the observed variables or the observed variables and the common factors (or conditional on the common factors). The model considered in Chapter 4 and in Bai et al. (2009)

include the common factors into the cointegrating relationship and therefore fall under the later case. The error correction tests proposed in Chapter 4 require weak exogeneity of the explanatory variables. However, this assumption is not tested. A panel extension of the Lagrange multiplier test of Boswijk and Urbain (1997) could be developed to pre-test that assumption. Furthermore, the tests developed in Chapters 3 and 4 test for a single cointegrating relationship under the alternative. The extension of system based approaches to panel data allowing for more than one cointegrating relationship has not seem much attention yet beyond the study of e.g. Larsson et al. (2001).

The methods considered throughout this thesis also differ in the way they estimate or approximate the common factors. One approach is to follow Bai and Ng (2004b) and estimate the common factors using the method of principal components, while an alternative is to follow the suggestion of Pesaran (2006) and use cross-sectional averages as a proxy. While the principal component estimator is consistent, the simple augmentation by cross-sectional averages is shown to work very well in theoretical studies and empirical applications. The analysis of the relationship between principal components and cross-sectional averages thus merits further research.

Estimating a cointegrating relationship in a panel with common factors is a problem that has not been considered in this thesis beyond the empirical study in Chapter 5. While several estimators have been proposed in the literature recently, we obtain varying results which are sensitive to model specifications. The properties of different estimators could be analyzed further. Furthermore, a proper extension to the estimation of dynamic models could be considered. From an empirical point of view it has to be remarked that the data set used in the analysis ends in 2002 and thus only includes 3 observations that have the Euro. Extending the data set should yield better estimates of the trade effect of the common currency.

Bibliography

- Anderson, J. E. (1979). A theoretical foundation for the gravity equation. *The American Economic Review* 69, 106–116.
- Andrews, D. W. K. (1991). Heteroskedasticity and autocorrelation consistent covariance matrix estimation. *Econometrica* 59, 817–858.
- Andrews, D. W. K. and C. Monahan (1992). An improved heteroskedasticity and autocorrelation consistent covariance matrix estimator. *Econometrica* 60, 953–966.
- Aznar, A. and M. Salvador (2002). Selecting the rank of the cointegration space and the form of the intercept using an information criterion. *Econometric Theory* 18, 926–947.
- Bai, J. and J. L. Carrion-i Silvestre (2007). Testing for panel cointegration with unobserved dynamic common factors. Unpublished Manuscript.
- Bai, J. and C. Kao (2006). On the estimation and inference of a panel cointegration model with cross-sectional dependence. In B. Baltagi (Ed.), *Panel Data Econometrics: Theoretical Contributions and Empirical Applications*, pp. 3–30. Amsterdam: Elsevier.
- Bai, J., C. Kao, and S. Ng (2009). Panel cointegration with global stochastic trends. *Journal of Econometrics* 149, 82–99.
- Bai, J. and S. Ng (2002). Determining the number of factors in approximate factor models. *Econometrica* 70, 191–221.
- Bai, J. and S. Ng (2004a). A new look at panel testing of stationarity and the PPP hypothesis. In D. Andrews and J. Stock (Eds.), *Identification and Inference in Econometric Models: Essays in Honor of Thomas J. Rothenberg*. Cambridge University Press.

- Bai, J. and S. Ng (2004b). A PANIC attack on unit roots and cointegration. *Econometrica* 72, 1127–1177.
- Bai, J. and S. Ng (2007). Panel unit root tests with cross-section dependence: A further investigation. Working Paper, Department of Economics, New York University.
- Baltagi, B. H. (2008). *Econometric Analysis of Panel Data* (fourth ed.). Chichester: John Wiley and Sons, Ltd.
- Baltagi, B. H. and C. Kao (2000). Nonstationary panels, cointegration in panels and dynamic panels, a survey. In B. Baltagi (Ed.), *Nonstationary Panels, Panel Cointegration, and Dynamic Panels*, pp. 7–52. Amsterdam: Elsevier.
- Banerjee, A. (1999). Panel data unit roots and cointegration: An overview. *Oxford Bulletin of Economics and Statistics* 61, 607–629.
- Banerjee, A. and J. L. Carrion-i Silvestre (2005). Breaking panel data cointegration. mimeo, University of Barcelona.
- Banerjee, A. and J. L. Carrion-i Silvestre (2006). Cointegration in panel data with breaks and cross-section dependence. Economics Working Papers ECO2006/5, European University Institute.
- Banerjee, A., J. J. Dolado, and R. Mestre (1998). Error-correction mechanism tests for cointegration in a single-equation framework. *Journal of Time Series Analysis* 19, 267–283.
- Banerjee, A., M. Marcellino, and C. Osbat (2004). Some cautions on the use of panel methods for integrated series of macro-economic data. *Econometrics Journal* 7, 322–340.
- Banerjee, A., M. Marcellino, and C. Osbat (2005). Testing for PPP: Should we use panel methods? *Empirical Economics* 30, 77–91.
- Bayer, C. and C. Hanck (2009). Combining non-cointegration tests. METEOR Research Memorandum RM/09/012, Maastricht University.
- Bergstrand, J. H. (1985). The gravity equation in international trade: Some microeconomic foundations and empirical evidence. *The Review of Economics and Statistics* 67, 474–481.
- Bergstrand, J. H. (1989). The generalized gravity equation, monopolistic competition, and the factor-proportions theory in international trade. *The Review of Economics and Statistics* 71, 143–153.

- Boswijk, P. H. (1994). Testing for an unstable root in conditional and structural error correction models. *Journal of Econometrics* 63, 37–60.
- Boswijk, P. H. and J.-P. Urbain (1997). Lagrange multiplier tests for weak exogeneity: A synthesis. *Econometric Reviews* 16, 21–38.
- Breitung, J. and S. Das (2008). Testing for unit roots in panels with a factor structure. *Econometric Theory* 24, 88–108.
- Breitung, J. and M. H. Pesaran (2008). Unit roots and cointegration in panels. In L. Matyas and P. Sevestre (Eds.), *The Econometrics of Panel Data: Fundamentals and Recent Developments in Theory and Practice*, pp. 279–322. Boston: Kluwer Academic Publishers.
- Bun, M. J. G. and F. J. G. M. Klaassen (2002). Has the euro increased trade? Tinbergen Institute Discussion Paper No. 02-108/2, University of Amsterdam.
- Bun, M. J. G. and F. J. G. M. Klaassen (2007). The euro effect on trade is not as large as commonly thought. *Oxford Bulletin of Economics and Statistics* 69, 473–496.
- Cappuccio, N. and D. Lubian (1996). Triangular representation and error correction mechanisms in cointegrated systems. *Oxford Bulletin of Economics and Statistics* 58, 409–415.
- Chang, Y. (2002). Nonlinear IV unit root tests in panels with cross-sectional dependency. *Journal of Econometrics* 110, 261–292.
- Chang, Y. (2004). Bootstrap unit root tests in panels with cross-sectional dependency. *Journal of Econometrics* 120, 263–293.
- Choi, I. (2001). Unit root tests for panel data. *Journal of International Money and Finance* 20, 249–272.
- Davidson, J. and R. de Jong (2000). The functional central limit theorem and weak convergence to stochastic integrals i: Weakly dependent processes. *Econometric Theory* 16, 621–642.
- Dees, S., F. di Mauro, M. H. Pesaran, and L. V. Smith (2007). Exploring the international linkage of the euro area: A global VAR analysis. *Journal of Applied Econometrics* 22, 1–38.
- Demetrescu, M. and A.-I. Tarcolea (2005). Panel cointegration testing using nonlinear instruments. Working Paper, Goethe-University Frankfurt.

- Engle, R. F. and C. W. J. Granger (1987). Co-integration and error correction: Representation, estimation and testing. *Econometrica* 55, 251–276.
- Fachin, S. (2005). Long-run trends in internal migrations in Italy: a study in panel cointegration with dependent units. mimeo, University of Rome ‘La Sapienza’.
- Flam, H. and H. Nordström (2003). Trade volume effects of the euro: Aggregate and sector estimates. mimeo, Institute for International Economic Studies, Stockholm University.
- Flôres, R., P. Jorion, P.-Y. Preumont, and A. Szafarz (1999). Multivariate unit root tests of the PPP hypothesis. *Journal of Empirical Finance* 6, 335–353.
- Frankel, J. and A. K. Rose (2002). An estimate of the effect of common currencies on trade and income. *The Quarterly Journal of Economics* 117, 437–466.
- Gengenbach, C., F. C. Palm, and J.-P. Urbain (2006). Panel cointegration testing in the presence of common factors. *Oxford Bulletin of Economics and Statistics* 68, 683–719.
- Gengenbach, C., F. C. Palm, and J.-P. Urbain (2009). Panel unit root tests in the presence of cross-sectional dependencies: Comparison and implications for modelling. *Econometric Reviews*, forthcoming.
- Gengenbach, C., J. Westerlund, and J.-P. Urbain (2008). Panel error correction testing with global stochastic trends. METEOR Research Memorandum RM/08/051, Maastricht University.
- Glick, R. and A. K. Rose (2002). Does a currency union affect trade? the time-series evidence. *European Economic Review* 46, 1125–1151.
- Gregoir, S. (2005). Representation and statistical analysis of weakly linearly exchangeable dynamic panels. Working Paper, CREST.
- Hadri, K. (2000). Testing for stationarity in heterogeneous panel data. *Econometrics Journal* 3, 148–161.
- Harris, R. D. F. and E. Tzervalis (1999). Inference for unit roots in dynamic panels where the time dimension is fixed. *Journal of Econometrics* 91, 201–226.
- Hecq, A., F. C. Palm, and J.-P. Urbain (2002). Separation, weak exogeneity and P-T decomposition in cointegrated VAR systems with common features. *Econometric Reviews* 21, 273–307.

- Im, K. S., M. H. Pesaran, and Y. Shin (2003). Testing for unit roots in heterogenous panels. *Journal of Econometrics* 115, 53–74.
- Kao, C. (1999). Spurious regression and residual-based tests for cointegration in panel data. *Journal of Econometrics* 90, 1–44.
- Kapetanios, G., M. H. Pesaran, and T. Yamagata (2008). Panels with nonstationary multifactor error structures. mimeo, University of Cambridge.
- Kremers, J. J. M., N. R. Ericsson, and J. J. Dolado (1992). The power of cointegration tests. *Oxford Bulletin of Economics and Statistics* 54, 325–348.
- Larsson, R., J. Lyhagen, and M. Lötgren (2001). Likelihood-based cointegration tests in heterogenous panels. *Econometrics Journal* 4, 109–142.
- Levin, A., C. F. Lin, and C. S. J. Chu (2002). Unit root tests in panel data: Asymptotic and finite-sample properties. *Journal of Econometrics* 108, 1–24.
- Lyhagen, J. (2000). Why not use standard panel unit root test for testing PPP. mimeo, Stockholm School of Economics.
- Maddala, G. S. and S. Wu (1999). A comparative study of unit root tests with panel data and cointegration tests. *Oxford Bulletin of Economics and Statistics* 61, 631–652.
- Mark, N. C. and D. Sul (2001). Nominal exchange rates and monetary fundamentals: Evidence from a small post-Bretton Woods panel. *Journal of International Economics* 53, 29–52.
- Mark, N. C. and D. Sul (2003). Cointegration vector estimation by panel DOLS and long-run money demand. *Oxford Bulletin of Economics and Statistics* 65, 655–680.
- Micco, A., E. Stein, and G. Ordoñez (2003). The currency union effect on trade: Early evidence from EMU. *Economic Policy* 37, 325–356.
- Moon, H. R. and B. Perron (2004). Testing for a unit root in panels with dynamic factors. *Journal of Econometrics* 122, 81–126.
- O’Connell, P. G. J. (1998). The overvaluation of purchasing power parity. *Journal of International Economics* 44, 1–19.

- Oh, K.-Y. (1996). Purchasing power parity and unit root tests using panel data. *Journal of International Money and Finance* 15, 405–418.
- Palm, F. C., S. Smeeke, and J.-P. Urbain (2007). A sieve bootstrap test for cointegration in a conditional error correction model. METEOR Reseach Memorandum RM/07/054, Maastricht University.
- Palm, F. C., S. Smeeke, and J.-P. Urbain (2008). Cross-sectional dependence robust block bootstrap panel unit root tests. METEOR Reseach Memorandum RM/08/048, Maastricht University.
- Park, J. Y. and P. C. B. Phillips (1988). Statistical inference in regressions with integrated processes: Part 1. *Econometric Theory* 4, 468–497.
- Pedroni, P. (1995, June). Panel cointegration; asymptotic and finite sample properties of pooled time series tests, with an application to the PPP hypothesis. Indiana University Working Papers in Economics, No/ 95-013.
- Pedroni, P. (1999). Critical values for cointegration tests in heterogenous panels with multiple regressors. *Oxford Bulletin of Economics and Statistics* 61, 653–670.
- Pedroni, P. (2004a). Panel cointegration; asymptotic and finite sample properties of pooled time series tests, with an application to the PPP hypothesis. *Econometric Theory* 20, 597–625.
- Pedroni, P. (2004b). Social capital, barriers to production and capital shares; implications for the importance of parameter heterogeneity from a non-stationary panel approach. mimeo, Williams College.
- Pedroni, P. and J.-P. Urbain (2001). Cross member cointegration in non-stationary panels. mimeo, Maastricht University.
- Pesaran, M. H. (2006). Estimation and inference in large heterogenous panels with a multifactor error structure. *Econometrica* 74, 967–1012.
- Pesaran, M. H. (2007). A simple panel unit root test in the presence of cross section dependence. *Journal of Applied Econometrics* 22, 365–312.
- Pesaran, M. H. and E. Tosetti (2009). Large panels with spatial correlation and common factors. mimeo, University of Cambridge.

- Pesavento, E. (2004). Analytical evaluation of the power of tests for the absence of cointegration. *Journal of Econometrics* 122, 349–384.
- Phillips, P. C. B. and S. Durlauf (1986). Multiple time series regression with integrated processes. *Reviews of Economic Studies* 53, 473–495.
- Phillips, P. C. B. and H. R. Moon (1999). Linear regression limit theory for nonstationary panel data. *Econometrica* 67, 1057–1111.
- Rapach, D. E. and M. E. Wohar (2004). Testing the monetary model of exchange rate determination: A closer look at panels. *Journal of International Money and Finance* 23, 867–895.
- Rose, A. K. (2000). One money, one market: The effect of common currencies on trade. *Economic Policy* 15, 9–45.
- Rose, A. K. and T. D. Stanley (2005). A meta-analysis of the effect of common currencies on international trade. *Journal of Economic Surveys* 19, 347–365.
- Saikkonen, P. (1991). Asymptotically efficient estimation of cointegration regressions. *Econometric Theory* 7, 1–21.
- Savin, N. E. (1980). The Bonferroni and Scheffé multiple comparison procedures. *Review of Econometric Studies* 47, 255–273.
- Sul, D. (2007). Panel unit root tests under cross section dependence with recursive mean adjustment. mimeo, University of Auckland.
- Tinbergen, J. (1962). *Shaping the World Economy*. New York: Twentieth Century Fund.
- Urbain, J.-P. (1992). On weak exogeneity in error correction models. *Oxford Bulletin of Economics and Statistics* 54, 187–206.
- Urbain, J.-P. and J. Westerlund (2008). Spurious regression in nonstationary panel time series with cross-unit cointegration. METEOR Research Memorandum RM/08/044, Maastricht University.
- Wagner, M. and G. Müller-Fürstenberger (2004). The Carbon Kuznets curve: A cloudy picture emitted by bad econometrics? Discussion Paper 4.18, University of Bern.
- Westerlund, J. (2005). New simple tests for panel cointegration. Working Paper 2005:8, Department of Economics, Lund University.

- Westerlund, J. (2007). Testing for error correction in panel data. *Oxford Bulletin of Economics and Statistics* 69, 709–748.
- Westerlund, J. (2008). Panel cointegration tests of the Fisher effect. *Journal of Applied Econometrics* 23, 193–233.
- Westerlund, J. and D. Edgerton (2008). Simple tests for cointegration in dependent panels with structural breaks. *Oxford Bulletin of Economics and Statistics* 70, 665–704.
- Westerlund, J. and R. Larsson (2008). A note on the pooling of individual PANIC unit root tests. Unpublished manuscript.
- Zivot, E. (2000). The power of single equation tests for cointegration when the cointegration vector is prespecified. *Econometric Theory* 16, 407–439.

Nederlandse samenvatting

De analyse van niet-stationaire panel data is een relatief jong onderzoeksgebied binnen de econometrie. Het begon met de toenemende beschikbaarheid van grote macro-economische databases die bijvoorbeeld informatie bevatten over landen gedurende tientallen jaren. Dit staat in contrast tot traditionele (micro) panel data waar er meestal slechts observaties voor enkele tijdsperiodes zijn. De grotere tijdsdimensie heeft het mogelijk gemaakt om methodes te gebruiken die ontwikkeld zijn voor de analyse van niet-stationaire tijdreeksen. Hoewel vroege studies besteedden niet veel aandacht aan afhankelijkheid tussen eenheden in de data, werd het snel duidelijk dat dit een cruciale eigenschap van economische panel data en dat deze afhankelijkheid een negatief effect kan hebben op de gebruikte econometrische methodes indien genegeerd. Verschillende alternatieven zijn voorgesteld in de literatuur om afhankelijkheid tussen eenheden toe te kunnen staan.

In dit proefschrift richten we onze aandacht op methodes die een structuur van gemeenschappelijke factoren gebruiken om de afhankelijkheid te modelleren. We dragen op verschillende manieren bij aan de econometrische literatuur op het gebied van niet-stationaire panel data. In Hoofdstuk 2 vergelijken we verschillende tweede generatie panel toetsen op eenheidswortels die de afhankelijkheid tussen eenheden modelleren met gemeenschappelijke factoren. De toetsen op eenheidswortels van Pesaran (2007) zijn ontworpen voor gevallen waar de afhankelijkheid tussen de eenheden wordt veroorzaakt door een enkele gemeenschappelijke factor. De toetsen van Moon and Perron (2004), die gedefactoreerde data gebruiken zijn vergelijkbaar maar kunnen meerdere factoren meenemen. De toetsen van Bai and Ng (2004b) kunnen de bron van de niet-stationariteit bepalen door het afzonderlijk toetsen op de gemeenschappelijke factoren en de idiosyncratische componenten. Breitung and Das (2008) en Sul (2007) stellen panel toetsen op eenheidswortels voor wanneer er mogelijk afhankelijkheid tussen eenheden aanwezig is door gemeenschappelijke factoren, maar de structuur

van de factoren niet volledig benut wordt. We vergelijken de methodes op basis van het data genererende proces, de toetsen, de nul en alternatieve hypothese en we vergelijken de eigenschappen in kleine steekproeven door middel van Monte Carlo resultaten. De meeste toetsen kunnen alleen een eenheidswortel in de idiosyncratische componenten detecteren; alleen de methodes van Bai and Ng (2004b) en Sul (2007) kunnen stochastische trends in de gemeenschappelijke factoren waarnemen. Als gevolg hiervan zijn alleen de laatste twee methodes in staat om coïntegratie tussen eenheden waar te nemen. Daarnaast verschaffen we ook nog een toepassing die het gebruik van de toetsen illustreert. Uiteindelijk wordt besproken hoe de toetsen gebruikt kunnen worden voor modelleren in het algemeen.

In Hoofdstuk 3 bekijken we een *spurious* regressie model voor een panel met niet-stationaire gemeenschappelijke factoren. We nemen aan dat de geobserveerde variabelen een factor structuur volgens als die voorgesteld in Bai and Ng (2004b) voor panel toetsen op eenheidswortels. Ons model staat ons toe onderscheid te maken tussen twee verschillende gevallen waarvan wij denken dat ze van theoretisch en empirisch belang zijn: (i) het geval waar de niet-stationariteit in de variabelen enkel voortkomt uit gemeenschappelijke trends (*coïntegratie tussen eenheden*); (ii) het geval waar er zowel gemeenschappelijke als idiosyncratische trends zijn. In het tweede geval is het verwerpen van de nul hypothese voor beide componenten noodzakelijk maar niet voldoende voor coïntegratie and we bespreken de vereiste homogeniteitsrestricties op de coïntegrerende vectoren die resulteren uit de aanwezigheid van coïntegratie van/door gemeenschappelijke factoren. Bovendien bestuderen we het asymptotische gedrag van enkele bestaande toetsen op coïntegratie in panels die gebaseerd zijn op residuen, zoals bekeken door Kao (1999) en Pedroni (1999, 2004a). We laten zien dat voor het gebruikte data genererende proces de toetsingsgrootheden niet langer asymptotisch normaal verdeeld zijn, en dat convergentie plaatsvindt met snelheid T in plaats van \sqrt{NT} , zoals voor onafhankelijke panels. Naar het werk van Bai and Ng (2004b) beschouwen we dan de mogelijkheden om te toetsen op de verschillende vormen van geen-coïntegratie door de factoren en de individuele componenten rechtstreeks aan de geobserveerde data te onttrekken en dan de geschatte componenten afzonderlijk te toetsen.

Hoofdstuk 4 bekijkt een gecoïntegreerd panel data model met gemeenschappelijke factoren. We ontwikkelen alternatieve representaties van een gecoïntegreerd panel dat niet-stationaire factoren toe laat. Vanuit de triangulaire representatie van het systeem, die bijvoorbeeld door Bai et al. (2009) gebruikt wordt, leiden we een Granger-achtige representatie stelling af vergelijkbaar met degene afgeleid door Cappuccio and Lubian (1996) voor een enkelvoudige tijdreeks. De conditionele

error correction representatie wordt verkregen en gebruikt als basis voor het onwikkelen van twee nieuwe toetsen voor de nul hypothese van geen error correction. In het bijzonder bekijken we panel versies van de t -toets zoals voorgesteld door Banerjee et al. (1998) en Boswijk (1994) en de Wald toets van Boswijk (1994). We laten zien dat de toetsen specifiek voor een individu asymptotisch zonder onbekende parameters zijn en alleen afhangen van het aantal niet-stationaire variabelen in het systeem. Echter, ze zijn niet onafhankelijk tussen eenheden door de gemeenschappelijke factoren. Desondanks convergeert het gemiddelde van de toetsingsgrootheden naar een kansvariabele met een distributie die, hoewel analytisch niet achterhaalbaar, wel gesimuleerd kan worden. Dit maakt poolen mogelijk ondanks de afhankelijkheid. We onderzoeken de eigenschappen van de toetsen in kleine steekproeven in een Monte Carlo experiment en vergelijken ze met de toetsen van Westerlund (2007). We bekijken ook twee empirische toepassingen van de nieuwe toetsen.

Dit proefschrift draagt ook bij aan de empirische literatuur over zwaartekracht modellen van bilaterale handel en het effect van een gezamenlijke munteenheid op handel. In Hoofdstuk 5 bekijken we Bun and Klaassen (2007) opnieuw voor een onderzoek naar de impact van de introductie van de euro op bilaterale handel. Hoewel er overtuigend bewijs is van een positief effect van een gezamenlijke munteenheid op handel, is er een voortdurende discussie over de daadwerkelijke grootte van het effect in de empirische literatuur, beginnende met het werk van Rose (2000). Rekening houdend met deterministische trend in de residuen van de zwaartekracht vergelijking schatten Bun and Klaassen (2007) een euro effect van ongeveer 3%, kleiner dan eerdere schattingen tussen 5% en 40%. Wij bekijken hun data opnieuw, die waarnemingen bevat van de 15 lidstaten van de Europese Unie voor de uitbreiding in 2004, en van Noorwegen, Zwitserland, Canada, Japan en de VS over een periode van 1967 to 2002, gebruik makend van nieuwe methodes die recentelijk zijn ontworpen voor de analyse van niet-stationaire panel data met afhankelijkheid tussen de eenheden. Gebruik makend van verschillende panel toetsen op eenheidswortels vinden we sterk bewijs dat (de logaritme van) bilaterale handel, als ook het product van BBP en BBP per capita een eenheidswortel bevat. Echter, we vinden coïntegratie tussen deze variabelen met de coïntegratie toetsen ontwikkeld in Hoofdstuk 3 en 4. Met de CCEP schatter van Pesaran (2006) en the CUP schatter van Bai et al. (2009) verkrijgen we schattingen van de coïntegrerende vector en schattingen van het euro effect op bilaterale handel. Onze schattingen variëren tussen modellen en schatters maar lijken de bevindingen van Bun and Klaassen (2007) te onderschrijven.

Curriculum Vitae

Christian Gengenbach was born on September 1, 1978 in Bonn, Germany. He attended high school at Hardtberg Gymnasium Bonn from 1989 to 1998, obtaining a university qualifying degree (*Abitur*). After high school, he joined the German Air Force for 10 months of compulsory military service. From 1999 to 2003 he studied Econometrics at Maastricht University. During his studies, he worked as a research assistant at the Department of Economics. He spent the spring semester 2003 studying at the Université de Montréal, Canada. In October 2003, he obtained his Master's Degree in Econometrics.

Christian joined the Department of Quantitative Economics at Maastricht University in September 2003. He worked under the supervision of Prof. dr. Franz C. Palm and Prof. dr. Jean-Pierre Urbain. The results of his research has been presented at several international conferences. His work has been published or is forthcoming in international reviewed academic journals.

