

Estimation and inference in cross-sectionally dependent panel data models

Citation for published version (APA):

Karabiyik, H. (2015). *Estimation and inference in cross-sectionally dependent panel data models*. [Doctoral Thesis, Maastricht University]. Datawyse / Universitaire Pers Maastricht. <https://doi.org/10.26481/dis.20150121hk>

Document status and date:

Published: 01/01/2015

DOI:

[10.26481/dis.20150121hk](https://doi.org/10.26481/dis.20150121hk)

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.

Estimation and Inference in
Cross-Sectionally Dependent Panel Data Models

Hande Karabiyik

© Hande Karabiyik, Maastricht 2015

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: 978-94-6159-401-3

Printed in The Netherlands by Datawyse Maastricht

Estimation and Inference in Cross-Sectionally Dependent Panel Data Models

PROEFSCHRIFT

ter verkrijging van de graad van doctor
aan de Universiteit Maastricht,
op gezag van Rector Magnificus,
Prof. dr. L.L.G. Soete,
volgens het besluit van het College van Decanen,
in het openbaar te verdedigen
op woensdag 21 januari 2015, om 14:00 uur

door

Hande Karabiyik



Promotores:

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

Prof dr. F. C. Palm

Beoordelingscommissie:

Prof. dr. P. Mohnen (Maastricht University) (voorzitter)

Prof. dr. J. L. Carrion-i-Silvestre (University of Barcelona)

Prof. dr. P. Pedroni (Williams College)

Dr. S.J.M. Smeekes

Dit onderzoek werd financieel mogelijk gemaakt door de Graduate School of Business and Economics (GSBE).

To my family:
Nurşen, Kurtuluş and Acar

Acknowledgements

It is the collection of our experiences that defines us, makes us who we are. In order to learn about the self, one should look back and remember the moments in these experiences together with the people one is accompanied by. When I look back I remember all those who contributed to the person I am now. It is now my turn to express my gratitude to all.

I have been most privileged to have worked under guidance of two great persons: my supervisors Jean-Pierre Urbain and Franz Palm. Jean-Pierre was always there to talk, and surely not only on econometrics. I learned a lot from every second we spent together and I always felt extremely lucky for being supervised by him. When I look back to last four years, I see him as a conductor and the state of my mind as an orchestra that could easily end up in a cacophony if it wasn't for his wisdom to put everything in a melodic order. His contribution to the person I am now is undeniable and for that I am thankful to him. Equally, I would like express my gratitude to Franz Palm for his patience, his wise guidance and the inspiration he provided in every stage of the completion of this thesis. The journey of writing this thesis has been manageable mostly thanks to his attentive support and enlightening advices.

I would also like to express my highest gratitude to Professor Ege Yazgan. He is the person who led me to the path I've been in for years now. Without his support and guidance, I would never be able to start this journey. I am sincerely grateful to him for believing in me and encouraging me.

I am also grateful to Joakim Westerlund for his contributions in the completion process of this thesis. I am thankful to him for helping me in my research. He has been a wise, understanding and helping mentor for me.

ACKNOWLEDGEMENTS

I would like to thank the reading committee, Pierre Mohnen, Peter Pedroni, Josep Lluís Carrion-i-Silvestre, Stephan Smeekes for their time and careful evaluation of this thesis. Among the contributors of this thesis, I would especially like to thank Eric Beutner, Alain Hecq, Stephan Smeekes, Christian Gengenbach and Denis De Crombrughe for the very helpful discussions we had. Also, I am grateful to Karin van den Boorn, Yolanda Paulissen and Haydeé Hallmans for their help in all the administrative matters.

I owe a lot to the welcoming, warm and inspiring environment of the Quantitative Economics Department and I am grateful to its members for that. I would especially like to thank my officemate Thomas Götz for being a good friend and for offering me his help when I needed. Another special thanks go to Hanno Reuvers for helping me in coming up with a Dutch summary for the thesis, for being there to help me in all the difficulties I faced towards the end of the completion process of this thesis and for being such a good friend. Also, I would like to thank Oksana Balabay, Anne Balter, Artem Duplinsky, Ahmad Salahnejhad Ghalehjooghi, Jan Lohmeyer, Kouma Gnameho and Lenard Lieb, among others.

My stay in Maastricht throughout was only joyful, thanks to wonderful friends. They filled my days with cheerful moments and helped me avoid any homesickness by creating an atmosphere that made me feel nowhere but home. For all these reasons I would like to thank Çiğdem Akbulut, İrem Bozbay, Burak Can, Devrim Dumludağ, Bart Dormans, Hande Erkut, Seher Fazlıoğlu, Özge Gökdemir, Mehmet İsmail, Fırat İsmailoğlu, Livingston Merchant, Ingrid Rohde, Merve Özer, Emel Öztürk, Murat Öztürk, Özer Selçuk, Erkan Yönder and Birol Yüceoğlu.

My deepest gratitude goes to my family. My parents, Nurşen and Kurtuluş Karabıyık have supported me in every decision I made. Their presence has given me strength and courage in everything I do. I cannot express enough my gratitude to my brother, Acar Karabıyık, who has always been more than a brother. I thank him for bringing profundity to everything in my life and for being just perfect.

Finally, I would like to thank my companion in life, my best friend, my ultimate help, my husband, Ali. His faith in me, his patience, his care, his never ending support have made everything possible. His presence has always given me strength and inspiration and turned my life into something precious. I thank him for filling my life with such goodness.

Hande Karabıyık
Maastricht, August 2014

Contents

Acknowledgements	vii
Contents	ix
1 Introduction	1
1.1 Objective of the Thesis	1
1.2 Motivation	3
1.3 An Introductory Discussion	5
1.4 Contributions and Thesis Structure	14
1.5 Notation	17
2 CCE Estimation of Factor-Augmented Regression Models with more Factors than Observables	19
2.1 Introduction	19
2.2 Assumptions	22
2.3 Asymptotic Results	24
2.3.1 The C ³ E Estimator	24
2.3.2 Bias Correction	27
2.3.3 Selecting the Combinations	28
2.4 Monte Carlo Simulations	30
2.4.1 Simulation Results	32
2.5 Conclusion	34
2.A Tables	36
2.B Appendix	44

CONTENTS

2.B.1	Preliminary Results	44
2.B.2	Proof of Theorem 2.1	47
2.B.3	Proof of Proposition 2.3.1	61
2.B.4	Proof of Proposition 2.3.2	63
3	A CCE Estimator for Dynamic Panel ECM's with Factors	65
3.1	Introduction	65
3.2	Model and Assumptions	67
3.2.1	Conditional Model	72
3.3	Estimation	73
3.3.1	Observed Factors	75
3.3.2	Unobserved Factors	77
3.3.3	Mean Group Estimator	90
3.4	Monte Carlo Simulations	93
3.4.1	Data Generating Process	94
3.4.2	Simulation Results	95
3.5	Conclusion	97
3.A	Tables	100
3.B	Appendix	114
3.B.1	The Case with Observed Factors	114
3.B.2	The Case with Unobserved Factors	124
3.B.3	The Mean Group Estimator	168
4	Testing Weak Exogeneity in Cross-Sectionally Dependent Cointegrated Panels	177
4.1	Introduction	177
4.2	Model and Assumptions	179
4.2.1	Conditional Model	183
4.3	Weak Exogeneity	185
4.4	A Lagrange Multiplier Test for Weak Exogeneity	186
4.4.1	Observed Factors	187
4.4.2	Unobserved Factors	189
4.4.3	Pooled Test	201
4.5	Monte Carlo Simulations	205
4.5.1	Data Generating Process	206
4.5.2	Monte Carlo Results	207
4.6	Conclusion	210

CONTENTS

4.A Tables	212
4.B Appendix	220
4.B.1 Preliminary Results	220
4.B.2 Proofs of Lemmas 4.4.1-4.4.3	231
5 Conclusion	237
Bibliography	243
Nederlandse samenvatting	251
Valorization	255
Curriculum Vitae	261

Chapter 1

Introduction

1.1 Objective of the Thesis

The analysis of panel data has been popular across many fields of research over the last decades. Early literature assumed parameter homogeneity and independence across cross-section units. Recently, the analysis of panel data sets with large time series and cross-section dimensions has shown that the individual units of the data sets are usually interdependent. Relevant theoretical research has shown that the presence of this interdependence may have serious consequences depending on the nature of dependence, if not accounted for. This further insight led the way to a new branch of theoretical research that takes into account the correlation between cross-section units in the panel data. So far there has been a voluminous amount of effort to account for cross-sectional dependence in different set-ups. The advancements in this branch of the literature are recently surveyed by Sarafidis and Wansbeek (2012) and Chudik and Pesaran (2013d).

In this thesis we focus on cross-sectionally dependent panel data models by employing an unobserved common factor structure to account for cross-correlation between units. In particular, we consider the CCE method that is proposed by Pesaran (2006) to account for cross-sectional dependence in stationary panel data models and propose an extension of that method that works under more general assumptions (Chapter 2). We investigate the validity of the CCE method proposed by Pesaran (2006) within a more general set-up, i.e. a non-stationary dynamic framework with unobserved common factor structure in a cointegrated panel set-

up (Chapter 3). We propose a test to test for weak exogeneity in cointegrated panels with cross-sectional dependence (Chapter 4).

With this work we aim at providing the empirical researcher with the statistical tools to be used in analyzing cross-sectionally dependent panel data under less restrictive assumptions and in more general set-ups. To be more specific, by contributing to the theoretical literature on cross-sectionally dependent panel data models in non-stationary and stationary set-ups we aim to lead the way to a more sophisticated empirical research with panel data sets with large cross-section and time series dimensions.

The method we propose in Chapter 2 aims to allow the researchers to work under less restrictive assumptions regarding the nature of the unobserved common factor structure that drives the data than the assumptions required in the existing literature.

In Chapter 3, we consider the estimation of error correction models under cross-sectional dependence. The error correction model set-up has been frequently used in time series analysis by empirical researchers to investigate the validity of macroeconomic theories in a dynamic setting. Macroeconomic theory suggests the existence of long-run relationships between certain variables. Long-run relationships frequently encountered with macro data include among others the purchasing power parity (PPP), the Fisher equation, the uncovered interest parity. This approach may be preferable depending on the research question, because it allows the researcher to estimate both the long-run relationships and the short-run dynamics simultaneously and is useful when dealing with cointegrated data. The interpretation of the estimated model is straight-forward. It can be used to test for macroeconomic theories, in policy analysis etc. In a panel data case, until now this could only be done by relying on the existing theoretical results that assume cross-sectional independence in the data, for example, see Pesaran et al. (1999). Under cross-sectional dependence the pooled estimation of an error correction model may, however, yield inconsistent estimates and can cause misleading inference if the dependence is ignored. Motivated by this fact we investigate the error correction models for panels with cross-sectional dependence. With the results we obtained in Chapter 3, we purport that the consistent estimation of error correction models is still possible under cross-sectional dependence given that the dependence is properly accounted for.

In Chapter 4, we consider a test for weak exogeneity in cross sectionally dependent, cointegrated panels. Weak exogeneity is a crucial assumption of conditional

error correction model analysis. The results we provide in Chapter 3 also rely on the assumption of weak exogeneity. Assuming weak exogeneity of some regressors for the long-run parameters is useful in error correction model analysis, since it allows for conditioning on the weakly exogenous variables and if assumed correctly, permits efficient estimation of parameters of interest only. Incorrectly assumed weak exogeneity has serious consequences, from efficiency loss to invalid inference. It is a frequently made assumption in empirical studies. For example, the assumption of weak exogeneity of permanent income is crucial in permanent income hypothesis analysis and needs to be tested. By showing that a certain type of test for weak exogeneity works with cross-sectionally dependent panels and proposing a method to overcome the problems of pooling cross-sectionally dependent statistics, we provide the empirical researchers with the tools to test for the assumption of weak exogeneity in cross-correlated panel data.

1.2 Motivation

Empirical researchers from various fields aiming to analyze the behaviour of households, firms, sectors, countries started to rely on methods developed for panel data with large cross-section and time series dimensions, as the availability of such data series increased. In many fields of research like in cross-country studies, finance, macro studies, regional studies, it is now customary to work with panels with large time series and cross-section dimensions. This has created demand for more sophisticated theoretical research on the inference methods for panel data with large cross-section and time dimensions. To be more specific, increasing dimensions of panel data sets has brought about the need for dealing with problems that are encountered both in time series literature and in cross-sectional analysis literature.

Using panel data comes with various advantages when compared to time-series or cross-sectional data. For an extensive discussion of the advantages and limitations of panel data see, for example Hsiao (2007), Baltagi (2008). Firstly, panel data provides a greater ability to capture individual behaviour over time compared to aggregate data, which might be heterogeneous across individuals. In particular, in the presence of heterogeneity, using a panel data might be preferable to using aggregate data, since it makes it possible to include more variables that might be time invariant or state invariant, which explain the heterogeneity of the individuals across the cross-section dimension. This can only be done using dummy variables

with an aggregate data. By using panel data the bias that arises as a consequence of ignored heterogeneity can easily be avoided in a static case. Another advantage of using panel data is that it is more informative and provides more accurate inference of model parameters since it contains more degrees of freedom. It sets the ground for more efficient estimation. Additionally, with panel data, which provide more variability, the collinearity can be reduced compared to the collinearity encountered in time series data. Another advantage of panel data that manifests itself when compared to cross-section data is that panel data has the ability to capture the dynamics of change, whereas cross-section data can only explain the behaviour of entities at a given point in time. With panel data one can focus on inter-individual differences along with the inter-temporal differences. Thus, panel data allows the researcher to formulate more sophisticated hypotheses regarding the behaviour of the entities being considered. The growth in the number applied studies of panel data, as the data become available, can be reasoned by these advantages.

Besides its advantages, panel data has a number of limitations. The limitations that are usually encountered in collecting the data from households (usually encountered in micro panel analysis) are discussed in detail, with examples in Baltagi (2008). Since this topic is out of the scope of our work we do not discuss them here. These problems can partly be overcome by meticulous designs of surveys for households and normally do not arise in macroeconomic or financial data. The second domain of limitations arises from the nature of the data. For instance increased time-dimension introduces time-series related limitations to the analysis, like serial correlation of the errors, non-stationarity or cointegration. Another frequently encountered limitation with macro panels occurs when the cross-section units are interdependent to each other. This phenomenon is called *cross-sectional dependence* and might lead to misleading inference and estimation problems. A detailed discussion of cross-sectional dependence is provided in the next section. This work is an attempt to contribute to the literature of cross-sectionally dependent panel data models.

Modelling approaches that have a static nature have been sufficient for the methods developed for panel data with a small time series dimension. However, a large time series dimension may induce a dynamic nature to the data. With the availability of panel data with an increased time dimension, the need of robust estimation methods for that has increased as well. In a static case, it is possible to obtain unbiased estimates, even when the parameters vary across units,

by pooling, aggregating, averaging group estimates and by cross-section regressions. In the dynamic case with different parameter values across units, which is called *parameter heterogeneity*, these methods give misleading estimates. This issue is addressed in Pesaran and Smith (1995). They examine the four methods mentioned above in a stationary and also in a non-stationary set-up, where the parameters are allowed to be heterogeneous across cross-section units. They show that when the slope parameters are randomly distributed and the regressors are strictly exogenous cross-section regression provides a consistent estimate for the long-run effect. Pesaran et al. (1999) propose a maximum-likelihood based pooled mean group estimator, in which they assume a homogeneous long-run relation and a heterogeneity in the short-run. They provide results for the stationary and the non-stationary case for a fixed cross-section dimension, under the assumption of cross-sectional independence.

A large time-series dimension comes with potentially non-stationary and cointegrated data series and necessitates the robust procedures for testing for unit root. Early literature on this topic assumed cross-sectional independence. The testing procedures that are proposed by, for example, Levin et al. (2002), Im et al. (2003), while allowing for heterogeneity, assumed cross-sectional independence. Maddala and Wu (1999), Choi (2001) also assume cross-sectional independence across units but they allow for the presence of a common time effect. Furthermore, limit theory for panel data with large cross-section and time series dimension is studied by Phillips and Moon (1999). However, their results are not valid under cross-sectional dependence. In the next section after discussing the notion of cross-sectional dependence, we discuss the recent developments in the literature that take into account the correlation across cross-section units.

1.3 An Introductory Discussion

Although each chapter in this thesis has a different set-up, at this stage it is useful to provide some general information about the very general set-up that would play an introductory role for the upcoming chapters. For this purpose, in this section we briefly discuss the relevant econometric issues and introduce the corresponding terminology that have been investigated throughout the chapters of this thesis.

In analysis of panel data the variables have double subscripts. For example, we can denote a double index process as y_{it} where i indicates the cross-section unit with $i = 1, \dots, N$ and where N is the number of cross-section units. Likewise,

t denotes the time points in the data series, where $t = 1, \dots, T$ and where T is the time series dimension. In applications i can denote households, individuals, countries, firms, regions etc. In the literature cross-section units are called with various names, like units, members, individuals, individual units etc. There are two main categories of panel data sets depending on the magnitude of the two dimensions N and T . A micro panel is a panel for which the time dimension T is substantially smaller than the cross-section dimension N and contain information on micro units, like individuals. A macro panel is a panel for which the magnitude of the time dimension T is similar to the magnitude of the cross-section dimension N and often contain aggregated information on macro units, like countries, regions, firms. A panel data set, where both of the dimensions (N and T) are large are called large panels.

Cross-sectional dependence

Cross-sectional dependence in the data occurs when the individual units in the cross-section exhibit statistical dependence. This can arise due to omitted variables, common shocks and/or spatial effects etc. For instance, if there are some omitted variables in the model that are not independently distributed across cross-section units, this would create a dependence between the individual-specific errors. Likewise, a shock that effects all the units would have a similar effect as the omitted variable if the shock effects the units heterogeneously and would imply a dependence between the errors across the individual units. If the effect of the shock is homogeneous then the factors together with their loadings can be considered as a common time effect and can be eliminated by a simple transformation of the data. Depending on the source of the dependence, cross-sectional dependence in panels can take various forms and in return can have diverse consequences.

Economic theory usually suggests that the agents take decisions interdependently among each other. The increasing availability of data and the developments in the relevant theoretical literature supported this claim of economic theory by repeatedly detecting the existence of the ‘interdependence’ among individual units in a panel data. Compared to the dependence across time series units, this form of dependence is difficult to formulate since cross-section data often lacks a natural ordering unlike time series data. Therefore, one often needs to rely on some assumptions regarding the form of the dependence to be able to model the cross-section dependence in a panel data set-up.

Theoretical research has shown that ignoring cross-sectional dependence has severe consequences on the properties of the estimators. These consequences depend on the data generating process of the variables that are being analyzed, the nature and the extent of the dependence and the cross-section and time dimensions of the panel. Ignoring cross-sectional dependence may lead to inconsistent estimates and/or inefficiency problems depending on the factors mentioned above. These consequences create the need for adopting methods to account for cross-sectional dependence in the data.

If the magnitude of the cross-section dimension is small relative to the magnitude of the time series dimension a standard approach is to treat the regressions as they are unrelated and then estimate them by Generalized Least Squares. This method is proposed by Zellner (1962) and named as Seemingly Unrelated Regression Equations (SURE). The limitations of this model lie in two main domains. Firstly, the estimators obtained by this method are consistent only if the regressors are not correlated with the source generating the cross-sectional dependence. Secondly, in the absence of an imposed structure on the regression coefficients, this method only works for panels with a cross-section dimension that is smaller than the time dimension.

To model cross-sectional dependence, methods that are developed for large panels, where the magnitudes of the cross-section and the time dimensions are large, can be classified in two main strands. The first one assumes that the dependence across cross-section units can be characterized by a spatial process that represents the distance among units. As it is mentioned before, cross-sectional dependence is difficult to model also for the reason that it cannot be characterized as the serial correlation in time series, since the data that is being analyzed does not have a natural ordering across cross-section units. In spatial econometric approach, by using an exogenously defined distance metric we can formulate the models similar to models formulated by the time index. In empirical research the *distance metric* that is used to define the cross-correlation across units can be obtained by considering alternative measures of *distance* such as economic distance, social distance, policy distance, depending on the application. The spatial approach requires techniques for estimation other than OLS such as Generalized Methods of Moments (GMM) or approaches based on maximum likelihood. This approach is pioneered by Anselin (1988) and since then there has been a vast literature growth on this branch. The reader is referred to Lee and Yu (2010) for a recent survey. One of the limitations of using this approach is that it typically does not allow

for slope heterogeneity across units. Another limitation is that it requires a priori knowledge of the *distance metric* that might be difficult to come up with in many applications.

Another approach to model the cross-correlation of the errors is called *the factor structure approach*. This approach assumes the existence of a fixed and usually small number of common factors that are affecting the error processes of the individual units heterogeneously. The individual-specific error process for each individual unit can be formulated as the sum of a linear combination of the unobserved common factors and an idiosyncratic term that is usually i.i.d. across cross-section units. The idiosyncratic term can also be weakly correlated across cross-section units, this issue will be discussed later in this section. For a simple mathematical exposition of this approach, consider the following data generating process for the scalar y_{it} ,

$$y_{it} = \beta_i' \mathbf{x}_{it} + e_{it}, \quad (1.1)$$

where β_i is a $m \times 1$ vector of slope coefficients that is allowed to vary across cross-section units and \mathbf{x}_{it} is a $m \times 1$ vector of observed idiosyncratic variables that are the observations on the individual unit i at time t . For our purposes, without loss of generality, the error term of the model, e_{it} is an i.i.d. process across the time dimension but it is cross-sectionally correlated such that at time t

$$\text{Var}(\mathbf{e}_t) = \boldsymbol{\Sigma}_t,$$

where $\mathbf{e}_t = (e_{1t}, \dots, e_{Nt})'$ and $\boldsymbol{\Sigma}_t$ is an $N \times N$ symmetric, nonnegative definite matrix and the (i, j) -th element of $\boldsymbol{\Sigma}_t$ denoted by $\sigma_{ij,t}$ is bounded such that $0 < \sigma_{ij,t} \leq K$, for $i = 1, \dots, N$, where K is a finite constant independent of N . This way of characterizing cross-sectional dependence is used in Chudik et al. (2011) to make the distinction between strong and weak cross-sectional dependence. The assumption we placed on the covariance matrix $\boldsymbol{\Sigma}_t$, imposes strong cross-sectional dependence on \mathbf{e}_t . For a discussion on different ways to characterize cross-sectional dependence, see, for example Chudik and Pesaran (2013d).

The assumed factor structure can be modeled mathematically as

$$e_{it} = \boldsymbol{\lambda}_i' \mathbf{f}_t + \epsilon_{it}, \quad (1.2)$$

where \mathbf{f}_t is a $k \times 1$ vector that consists of the common factors that are unobserved but simultaneously effect all individual units. The $k \times 1$ vector of individual specific

factor loadings are denoted by $\boldsymbol{\lambda}_i$. Intuitively, the factor loadings measure the extent of the exposure of an individual unit to a change in the unobserved common factor. ϵ_{it} is a purely idiosyncratic component with zero mean and $Cov(\epsilon_{it}, \epsilon_{jt}) = 0$ for $i \neq j$ for all $i, j \in \{1, \dots, N\}$ where N is the number of cross-section units. Note that ϵ_{it} can also be weakly correlated across cross-section units. In Chudik et al. (2011), weak cross-section dependence is defined based on weighted cross-sectional averages of the processes in question. Using the notation we introduced, ϵ_{it} can be said to be cross-sectionally weakly dependent if at a given point in time

$$\lim_{N \rightarrow \infty} Var(\mathbf{w}'_t \boldsymbol{\epsilon}_t) = 0,$$

where \mathbf{w}_t is a weighting scheme that satisfies some granularity conditions and $\boldsymbol{\epsilon}_t = (\epsilon_{1t}, \dots, \epsilon_{Nt})'$. For alternative characterizations see Sarafidis (2009), Forni and Lippi (2001).

One of the consequences of ignored cross-sectional dependence are potentially inconsistent estimation procedures. Inconsistency occurs when the regressors are correlated with the factors. In the literature this situation is usually modelled by assuming a data generation process for \mathbf{x}_{it} as the following

$$\mathbf{x}_{it} = \boldsymbol{\Lambda}_i \mathbf{f}_t + \boldsymbol{\eta}_{it},$$

where $\boldsymbol{\Lambda}_i$ is a $m \times k$ matrix of factor loadings with elements that measure the effect of a change in each factor on the elements of the observed idiosyncratic variables, \mathbf{x}_{it} . $\boldsymbol{\eta}_{it}$ is a $m \times 1$ vector and is the idiosyncratic component of \mathbf{x}_{it} . In this simple model presentation of the factor structure approach we use a notation that is frequently encountered in the literature. The various set-ups can be generalized and/or completed by additional assumptions.

There is a growing literature that adopts the factor structure approach to model cross-sectional dependence. The literature varies by the method it uses to approximate the factor structure that has to be accounted for consistent estimation. One of the approaches that proposes an approximation for the unobserved factors is the Principal Components (PC) approach. It is first developed by Coakley et al. (2002). They propose to use a two-step procedure, which is based on the residual principal component analysis. In the first stage they propose to estimate (1.1), by ignoring the factor structure and extract the principal components of the residual covariance matrix that can be denoted by $\hat{\Sigma}_i = \frac{1}{NT} \sum_{i=1}^N \sum_{t=1}^T \hat{e}_{it}^2$. Then, they propose to use k eigenvectors associated with the first largest k eigenvalues of $\hat{\Sigma}_i$

as an approximation for the factors in the second stage. The limitation of this approach arises when the regressors are correlated with the unobserved common factors. If this is the case, the first stage estimation is inconsistent and this invalidates the use of residuals obtained from that regression in the extraction of principal components. Another possible drawback of this approach is that the number of factors k is needed to be estimated. Although there are methods developed to estimate the number of factors consistently, for example see Bai and Ng (2002), this introduces additional uncertainty to the procedure. Other examples in the literature that deal with determining the number of factors can be listed as, Kapetanios (2004), Kapetanios (2010), Onatski (2009), Ahn and Horenstein (2013), Bai and Ng (2007). For a recent survey on this issue see, for example Breitung and Choi (2013). The inconsistency that arises when the regressors are correlated with the factors is addressed by Bai (2009), which proposes an iterative method that can be seen as a quasi maximum likelihood method and yields consistent estimates even if the regressors are correlated with the factors. Note that the number of factors are still needed to be estimated. There exists various extensions of this method in the literature. For instance, Bai et al. (2009) show that when the data is non-stationary and cointegrated with the factors a modification of this method yields consistent estimates. For a review on panel data models and factor analysis, see Chapter 12 of Acemoglu et al. (2013) that is prepared by Jushan Bai, on this topic focusing on the recent works of the author. Song (2013) adopts a dynamic structure for the factors and establishes the consistency of the individual slope parameter estimates obtained by the iterative method proposed by Bai (2009). This method is also adopted for testing for unit roots in cross-sectionally dependent panels, see for example Bai and Ng (2004).

An alternative to the principal components approach is suggested by Pesaran (2006), which is named as the Common Correlated Effects (CCE) method. The author proposes to obtain an approximation for the unobserved factors by taking the cross-sectional averages of the observed variables and to augment the model for the dependent variable by the approximations. Then it is proven in the mentioned paper that the ordinary least squares estimation of the augmented model yields consistent estimates of the individual slope coefficients under certain assumptions and under certain conditions on the growth rates of the cross-section dimension and the time series dimension. Furthermore, the author shows that the mean group estimator based on the individual slope coefficient estimates and the pooled estimator are also consistent and have nuisance parameter free asymptotic distri-

butions. This method can be presented by first defining $\mathbf{z}_{it} = (y_{it}, \mathbf{x}'_{it})'$ that is a $(m+1) \times 1$ vector that consists of the observed variables of the model (dependent variable and regressors). Pesaran (2006) suggests to use $\bar{\mathbf{z}}_t = \frac{1}{N} \sum_{i=1}^N \mathbf{z}_{it}$ as an approximation for the factors and augment the model in (1.1) with the approximations. This approach does not provide an estimator for the factor loadings in the model, it simply treats the loadings as nuisance parameters. The idea is to use the cross-sectional means to approximate the space spanned by the unobserved common factors and to use this approximation to eliminate the factors by projecting them out in the estimation of the augmented model. The attractiveness of this approach comes from its computational simplicity. Another advantage of using this method is that it does not require the estimation of the number of unobserved common factors, which is an unavoidable requirement of the principal components approach. On the other hand the estimation of the individual slope coefficients with this method is possible only if a rank condition related to the factor loadings is satisfied. This rank condition requires that the number of observed variables to be greater than or equal to the number of unobserved common factors. Also, if that rank condition is not satisfied the approach requires the random coefficients assumption for the factor loadings to be able to obtain consistent mean group estimates with a nuisance parameter free asymptotic distribution. This issue is addressed in Chapter 2 of this thesis in detail. In Chapter 2, we propose an extension of the Common Correlated Effects approach, which would not require this rank condition for consistent estimation.

It has been shown that the CCE approach works in various set-ups provided that certain assumptions and conditions are satisfied. For instance, Kapetanios et al. (2011) show that the CCE estimator yields consistent estimates of the slope coefficients even when the data is non-stationary and the observables are cointegrated with the factors. The set-up they adopted in their paper is a static one, in the sense that the short-run dynamics are not parametrically modelled. In Chapter 4 we extend their work to an error correction model set-up by parametrically modelling the short-run dynamics. Westerlund and Urbain (2013a) compare the principal components approach and the common correlated effects method and find that there is no uniform ranking between the two in terms of efficiency, except for some very special cases. Investigating the two methods within a same specific framework, they find that the principal components method yields more efficient estimates. Chudik et al. (2011) show that the method yields consistent and asymptotically normal estimates of the slope coefficient of the model even

when there are a finite number of strong unobserved common factors and infinite number of weak factors. A strong or a weak factor can be characterized based on the factor loadings that correspond to that factor for each cross-section unit. Consider (1.2), according to the characterization provided by Chudik et al. (2011), a factor is said to be strong if

$$\lim_{N \rightarrow \infty} N^{-1} \sum_{i=1}^N |\lambda_{ij}| = K > 0$$

where λ_{ij} is the factor loading of the factor f_{jt} , where $f_{jt} \in \{f_{1t}, \dots, f_{kt}\}$ and K is a finite constant that does not depend on N . A factor is said to be weak if

$$\lim_{N \rightarrow \infty} \sum_{i=1}^N |\lambda_{ij}| = K < \infty.$$

Pesaran and Tosetti (2011) show that the CCE method yields consistent and asymptotically normally distributed estimates of the slope coefficients when ϵ_{it} is not i.i.d. across i , but generated by a spatial process. Chudik and Pesaran (2013a) consider a dynamic model with a factor structure, find that a dynamic augmentation of the model with the cross-sectional averages yields consistent individual and mean group estimates of the slope coefficients. Furthermore, Chudik and Pesaran (2013c) consider the CCE estimation of dynamic panels with a dominant unit, Chudik and Pesaran (2011) consider the CCE estimation of infinite dimensional vector autoregressive models with factors. Pesaran (2007) develops a unit root testing procedure by using the CCE approach in the presence of a single factor structure, Pesaran et al. (2013) extend this work to a multi factor error structure set-up. For other developments in the literature on the CCE approach see, for example Chudik and Pesaran (2013d). Alternatively, Chang (2002) proposes a non-linear IV panel unit root test by allowing for a general cross-section dependence. Another approach to deal with cross-sectional dependence in testing for unit roots in panels is to use a bootstrap procedure for a recent example, see Palm et al. (2011).

The approaches mentioned above constitute a major part of the literature that proposes a solution to the cross-sectional dependency problem in panel data analysis. In this thesis we built on the CCE approach briefly explained above.

Large- N and large- T asymptotics

Recently, by the increasing availability of the data series, asymptotic results which rely on large cross-section (N) and time series (T) dimensions are started to being used by theoretical researchers. The limit theory for panels with large cross-section and time series dimension is pioneered by Phillips and Moon (1999). In that paper, the asymptotic theory as both dimensions grow to infinity is developed for non-stationary data under the assumption of cross-sectional independence (or weak cross-sectional dependence). The purpose of this section is to briefly introduce the approaches for deriving the limits of processes with large- N and large- T dimensions and relate the discussion to the literature on cross-sectionally dependent panel data models.

Deriving the limits when the cross-section dimension and the time series dimension are both assumed to go to infinity can be done by using one of the following approaches. One approach is to consider the limits when one of the dimensions, say N , is kept fixed first while T is allowed to go to infinity. This yields an intermediate limit. Then, allowing N to go to infinity yields the limit of the double indexed process. This limit is called *the sequential limit*. Another approach is to let both dimensions (N and T) go to infinity jointly without any conditions imposed on the growth rates of N and T . The limit result that is obtained by this approach is called *the joint limit*. A third approach that is considered in large- N , large- T asymptotic theory is called *the diagonal path* limit theory. In this approach N and T are let to go to infinity as in the second approach, but this time with restrictions on the growth rates. There is an assumed expansion path, which makes one of the dimensions a monotonically increasing function of the other. Sequential limits are easy to obtain but when both dimensions are let to go infinity simultaneously, they may give misleading results. Diagonal path limits may be useful when the limit theory depend on a specific relation between the growth rates of N and T , since it allows to specify a certain growth path for the N and T dimensions. Joint limits are more robust than the sequential limits and the diagonal path limits but they are usually more difficult to obtain and require more restrictive conditions than the other two. Phillips and Moon (1999) discussed the conditions required to ensure that the limit approaches mentioned above are equivalent to each other in cross-sectionally independent panels.

The asymptotics of cross-sectionally dependent panels are analyzed by, for example Andrews (2005), Bai (2003), Kao et al. (2012). In the literature on

estimating cross-sectionally dependent panel data models when both dimensions of the panel is large, theoretical results usually rely on the sequential path limit results. For instance, Bai (2009) relies on joint limit results for consistency, whereas to obtain nuisance parameter free asymptotic distributions and a consistent bias correction, the author relies on diagonal path limit results. The functional form for the relation between N and T is $\frac{T}{N} \rightarrow \tau$ for the nuisance parameter free distribution of the slope coefficient of the model, where τ is a positive constant. Bai et al. (2009) use sequential limit theory for most of the results provided in their paper. For some of the results like the consistency of the initial estimator that would be continuously updated, they rely on diagonal path limit theory, with $\frac{\sqrt{N}}{T} \rightarrow 0$. On the other hand, the results related to the CCE estimator usually do not rely on sequential limit theory. For instance, in Pesaran (2006) the consistency of the individual slope coefficient estimator is obtained under the condition $\frac{\sqrt{T}}{N} \rightarrow 0$ as $(N, T) \rightarrow \infty$. Moreover, the asymptotic unbiasedness of the mean group estimator is proved only when $N \rightarrow \infty$ and T is kept fixed. As for the asymptotic distribution of the mean group estimator joint asymptotics are used. Kapetanios et al. (2011) use diagonal path for the individual specific slope coefficient estimator and joint asymptotics for the pooled estimates.

In this thesis, to obtain the results, we use joint asymptotics and, when necessary, diagonal path asymptotics. We write the joint asymptotic type limit processes as $(N, T) \rightarrow \infty$ for both cases and indicate the form of the functional dependence additionally when we use diagonal path asymptotics, i.e. we use $N^\alpha T^\beta \rightarrow \tau$ to indicate the dependence, where α and β are constants that depend on the result we obtain and τ is a nonnegative constant.

1.4 Contributions and Thesis Structure

This thesis contributes to the econometric literature on cross-sectionally dependent panel data models in several domains. In Chapter 2 we consider stationary panels, whereas in Chapters 3 and 4 we consider non-stationary, cointegrated panels. In this section we summarize and briefly point out the contributions of the individual works that constitute the chapters of this thesis.

In Chapter 2, we consider a simple, static and cross-sectionally dependent panel data model with homogenous slope coefficients. We assume that a factor structure in the errors of the model explains the dependence across units. Further, we assume that the regressors of the model are correlated with the factors. In a set-up like

this, ignoring the factor structure yields inconsistent estimates. Pesaran (2006) proposes a method to estimate the slope coefficients of such a model, that is to approximate the factors by using cross-sectional averages and then to augment the model with the approximations. This method is discussed also in the section above. As it is also mentioned earlier, for this method to work and yield \sqrt{NT} -consistent pooled estimates for the homogenous case, without requiring further assumptions on the factor loadings of the model, a certain rank condition is required regarding the factor loadings. This rank condition requires the number of observed variables to be greater than or equal to the number of factors. This may be difficult to justify in applications. We propose a modification of this method to eliminate the need for such a rank condition or the need for additional assumptions on the factor loadings. We consider not only one combination of the observed data as in Pesaran (2006) - (that is to use only the weighted cross-sectional averages of the observables) but also other combinations. We propose to use a combination matrix to premultiply the data matrix to obtain the combinations. Under certain assumptions placed on the combination matrix, this approach yields \sqrt{NT} consistent estimates of the slope coefficients of the model, without necessitating to impose additional assumptions on the factor loadings. In addition, we provide a bias correction method for the case where we have $\frac{T}{N} \rightarrow \tau$, where τ is a positive constant. We show that the correction method asymptotically corrects for the bias successfully. Furthermore, we propose a method to select the candidates for the combination matrix, which is similar to the one developed by Bai and Ng (2002). We show that the selection criteria consistently selects the true number of combinations that are required to estimate the slope coefficients consistently. The theoretical findings are also shown to work in small samples by Monte Carlo simulations.

In Chapter 3, we consider an error correction model for panels with cross-sectional dependence. We adopt a dynamic and non-stationary factor structure approach for the cross-sectional dependence. The factors are cointegrated with the observed variables. We assume weak exogeneity for some of the observed variables for the long-run coefficients to be able to estimate the conditional model without loss of efficiency. The short-run parameters are explicitly parameterized. With this property this work can be considered as an extension of Kapetanios et al. (2011), in which they model only the long-run relationship between the observed variables and the factors in a static set-up, by leaving the dynamics in the error term. We first provide the results regarding the estimation of the models for individual units, for the case, in which we assume the factors are observed. The results we

provide for this case are the adaptation of the ones provided in the literature (for example see Boswijk (1995)). The rate of consistency of the short-run coefficients is \sqrt{T} , whereas it is equal to T for the long-run coefficients. As originally suggested by Pesaran (2006), we propose to use the cross-sectional means of the observed variables to account for the cross-correlation between the error terms of the models for individual units. We find that using the levels and the first-differences of the cross-sectional averages to augment the conditional error correction model and estimating the augmented model by OLS (to obtain the estimators of the short-run coefficients) and by indirect least squares (to obtain the estimators of the long-run coefficients) yield consistent estimates of the parameters of interest. The approximation method we use requires the inclusion of further lags (theoretically infinite number of lags) of the first differences of the cross-sectional averages. We show that under certain conditions, by truncating the lag-order, the estimation method we suggest yields consistent estimates. The truncated lag-order is denoted by p_T . As $(N, T, p_T) \rightarrow \infty$, under certain conditions placed on the growth rate of N and T and the truncated lag-order we obtain the following results. (i) The indirect least squares estimation of the individual models yields T -consistent estimates for the long-run coefficients. (ii) The OLS estimation of short-run coefficients of the individual models yields \sqrt{T} -consistent estimates. (iii) Consistent estimators for the covariance matrices of the short-run and the long-run coefficient estimates are proposed. (iv) Mean group estimator for the slope coefficient of the weakly exogenous variables, which is based on averaging the individual estimates is \sqrt{N} -consistent and has a normal distribution, with a covariance that depends on the heterogeneity of the long-run slope coefficient. (v) Following Kapetanios et al. (2011) we suggest to use a non-parametric estimator for the covariance of the mean group estimator. Further, by Monte Carlo simulations, we show that the estimation method works well in small samples.

In Chapter 4, we consider a cross-sectionally dependent, non-stationary and cointegrated panel data set-up. The set-up and assumptions we adopt in this chapter is very much similar to the one in Chapter 3. In this work we develop a test for weak exogeneity. This work can be considered as an extension of, for example Boswijk (1995) to a cross-sectionally dependent panel set-up. The test we propose is a Lagrange multiplier type of test. First, we present the results for the case in which the factors are observed, which suggest the individual test statistic obtained by the Lagrange multiplier test has a χ^2 distribution with a degrees of freedom of m , where m is the number of variables that are being tested

for weak exogeneity. To obtain this result we use the result from Chapter 3, that asserts that the long-run relation can be estimated super-consistently from the conditional model. Then we go back to the case in which the factors are not observed. We use the same method as in Chapter 3 to deal with the unobserved factors. Again, as in the previous chapter, a truncation method is required to make the estimation feasible, since the lag-order for the first-differences of the observed variables induced by the approximation method is infinity. We show that, as $(T, N, p_T) \rightarrow \infty$, the individual Lagrange multiplier test statistics have a $\chi^2(m)$ distribution under the certain conditions that are placed on the relative growth rates of N , T and the truncated lag-order. Moreover, we consider the pooled tests that are based on the mean of the individual test statistics and based on Fisher's p -value combination method (See Fisher (1934)). The dependence of the individual test statistics invalidates the use of standard results for the combined tests. We therefore propose to use a bootstrap method to obtain data dependent, non-standard distributions of the combined test statistics. Finally, by Monte Carlo simulations we show that the tests have the correct size and high power.

In Chapter 5, we summarize the main findings of the thesis, provide a discussion of the limits of the current work and suggest some outlines for further research.

1.5 Notation

The following notation is used invariably throughout the thesis. Let $\mathbf{A} \in \mathbb{M}^{k \times m}$ be a $k \times m$ matrix, where $\mathbb{M}^{k \times m}$ is the space of real-valued $k \times m$ matrices. $rk(\mathbf{A})$ denotes the rank of matrix \mathbf{A} . If $m = k$ then $tr(\mathbf{A})$ is the trace of matrix \mathbf{A} and $\|\mathbf{A}\| \equiv \sqrt{tr(\mathbf{A}'\mathbf{A})}$ denote the Frobenius (Euclidean) norm of the matrix \mathbf{A} . $\|\mathbf{A}\|_\infty \equiv \max_{1 \leq i \leq k} \sum_{j=1}^m a_{ij}$ is the maximum absolute row sum norm and $\|\mathbf{A}\|_1 \equiv \max_{1 \leq j \leq m} \sum_{i=1}^k a_{ij}$ is the maximum absolute column sum norm of the matrix \mathbf{A} . $\langle \mathbf{a}, \mathbf{b} \rangle$ denotes the inner product of vectors \mathbf{a} and \mathbf{b} . \mathbf{A}^+ denotes the Moore-Penrose inverse. If $k > m$ and $rk(\mathbf{A}) = m$ then $\mathbf{A}^- = (\mathbf{A}'\mathbf{A})^{-1}\mathbf{A}'$ denotes the left inverse of \mathbf{A} . If $k < m$ and $rk(\mathbf{A}) = k$ then $\mathbf{A}^- = \mathbf{A}'(\mathbf{A}\mathbf{A}')^{-1}$ denotes the right inverse of \mathbf{A} . If \mathbf{A}_n is a matrix of sequences with deterministic elements, then $\mathbf{A}_n = O(b_n)$, $(\mathbf{A}_n = o(b_n))$ denotes that the elements of \mathbf{A}_n are at most (smaller) of order b_n . If \mathbf{A}_n is a matrix of sequences with random elements, then $\mathbf{A}_n = O_p(b_n)$, $(\mathbf{A}_n = o_p(b_n))$ denotes that the elements of \mathbf{A}_n are at most (smaller) of order b_n in probability. \xrightarrow{w} , \xrightarrow{p} and \xrightarrow{d} signify weak convergence, convergence in

1 INTRODUCTION

probability and convergence in distribution respectively. $(N, T, p) \rightarrow \infty$ signifies the joint convergence in N , T and p . $K < \infty$ denotes a generic positive number. $\mathbf{W}(r)$ is the standard Brownian motion defined on the interval $r \in [0, r]$ and the integral corresponding to the Brownian motion will be denoted as $\int_0^1 \mathbf{W}(r) dr$.

Chapter 2

CCE Estimation of Factor-Augmented Regression Models with more Factors than Observables

2.1 Introduction

Consider the scalar and $m \times 1$ vector of observable panel data variables y_{it} and \mathbf{x}_{it} , where $i = 1, \dots, N$ and $t = 1, \dots, T$ indexes the cross-sectional and time series dimensions, respectively. The data generating process (DGP) of the $T \times 1$ vector $\mathbf{y}_i = (y_{i1}, \dots, y_{iT})'$ is similar to the DGP of Pesaran (2006), and is given by

$$\mathbf{y}_i = \mathbf{X}_i \boldsymbol{\beta} + \mathbf{e}_i, \quad (2.1)$$

$$\mathbf{e}_i = \mathbf{F} \boldsymbol{\lambda}_i + \boldsymbol{\epsilon}_i, \quad (2.2)$$

where $\mathbf{X}_i = (\mathbf{x}'_{i1}, \dots, \mathbf{x}'_{iT})'$ is $T \times m$, $\boldsymbol{\beta}$ is a $m \times 1$ vector of slope coefficients, $\mathbf{F} = (\mathbf{f}'_1, \dots, \mathbf{f}'_T)'$ is a $T \times r$ matrix of common factors with $\boldsymbol{\lambda}_i$ being the associated $r \times 1$ vector of factor loadings, and $\boldsymbol{\epsilon}_i = (\epsilon_{i1}, \dots, \epsilon_{iT})'$ is a $T \times 1$ vector of idiosyncratic

errors. If the model includes unit-specific fixed effects, then \mathbf{y}_i , \mathbf{X}_i , \mathbf{e}_i , \mathbf{F} and $\boldsymbol{\epsilon}_i$ are simply the correspondingly (time)-demeaned variables.

The above model is the prototypical pooled panel regression with a factor error structure, in which $\boldsymbol{\epsilon}_i$ is independent of \mathbf{X}_i .¹ If \mathbf{F} is also independent of \mathbf{X}_i , then (2.1) is nothing but a static panel data regression with exogenous regressors, which can be estimated consistently using least squares (LS). If, however, \mathbf{X}_i is correlated with \mathbf{F} , then consistency will be lost. To allow for this possibility, we follow Pesaran (2006) and assume that

$$\mathbf{X}_i = \mathbf{F}\boldsymbol{\Lambda}'_i + \boldsymbol{\eta}_i, \quad (2.3)$$

where $\boldsymbol{\Lambda}_i$ is a $m \times r$ loading matrix and $\boldsymbol{\eta}_i = (\boldsymbol{\eta}_{i1}, \dots, \boldsymbol{\eta}_{iT})'$ is a $T \times m$ matrix of idiosyncratic errors. By combining (2.1)–(2.3),

$$\mathbf{W}_i = \mathbf{F}\mathbf{C}_i + \mathbf{U}_i, \quad (2.4)$$

where $\mathbf{W}_i = (\mathbf{y}_i, \mathbf{X}_i) = (\mathbf{w}_{i1}, \dots, \mathbf{w}_{iT})'$ is $T \times (m+1)$, $\mathbf{w}_{it} = (y_{it}, \mathbf{x}'_{it})'$ is $(m+1) \times 1$, $\mathbf{C}_i = (\boldsymbol{\Lambda}'_i\boldsymbol{\beta} + \boldsymbol{\lambda}_i, \boldsymbol{\Lambda}'_i)$ is $r \times (m+1)$, and $\mathbf{U}_i = (\mathbf{u}'_{i1}, \dots, \mathbf{u}'_{iT})' = (\boldsymbol{\eta}_i\boldsymbol{\beta} + \boldsymbol{\epsilon}_i, \boldsymbol{\eta}_i)$ is $T \times (m+1)$. Thus, (2.1)–(2.3) can be rewritten equivalently as a static factor model for \mathbf{W}_i , which is convenient because it means that the common component of the data can be estimated using existing methods for such models (see Chudik and Pesaran (2013d), for a recent survey).² In this chapter, however, we focus on the CCE approach of Pesaran (2006), which has become very popular in the empirical literature with a large number of applications. The approach has also attracted much interest in the econometric literature where it has been shown to work under very general conditions, including models with weak factors, dynamic models and even models with non-stationary data (see, for example, Chudik et al. (2011); Chudik and Pesaran (2013b); Kapetanios et al. (2011); Pesaran et al. (2013)).

As is well known from the classical common factor literature, \mathbf{F} and \mathbf{C}_i are not separately identifiable, suggesting that the best that one can hope for is consistent estimation of the space spanned by \mathbf{F} . The idea of Pesaran (2006) is to make use of the cross-section variation to estimate this space. A natural way to accomplish this is to take the cross-section average, giving $\bar{\mathbf{w}}_t = \bar{\mathbf{C}}' \mathbf{f}_t + \bar{\mathbf{u}}_t$, where $\bar{\mathbf{C}}$, $\bar{\mathbf{w}}_t$ and

¹One difference between the above DGP and the one considered by Pesaran (2006) is that here $\boldsymbol{\beta}$ is assumed to be homogenous; see Section 2 for a more detailed discussion.

²In Section 4 of the present chapter we present some Monte Carlo result that enable comparison with the principal components-based estimator of Bai (2009), which is arguably the closest competitor of the CCE approach.

$\bar{\mathbf{u}}_t$ are the cross-section averages of \mathbf{C}_i , \mathbf{w}_{it} and \mathbf{u}_{it} , respectively. Hence, since $\bar{\mathbf{u}}_t = o_p(1)$ (as $N \rightarrow \infty$), we have that $\bar{\mathbf{C}}' \mathbf{f}_t = \bar{\mathbf{w}}_t - \bar{\mathbf{u}}_t = \bar{\mathbf{w}}_t + o_p(1)$. This suggests using $\bar{\mathbf{w}}_t$ as an estimator of $\bar{\mathbf{C}}' \mathbf{f}_t$, a strategy that would seem to require

$$rk(\bar{\mathbf{C}}') = r \leq m + 1, \quad (2.5)$$

where $rk(\bar{\mathbf{C}}')$ denotes the rank of $\bar{\mathbf{C}}'$. Hence, the number of observables must be at least as large as the number of factors. The idea behind the CCE approach is then to estimate β as the LS slope on \mathbf{X}_i in a regression of \mathbf{y}_i onto \mathbf{X}_i and $\bar{\mathbf{w}}_t$.

As Pesaran (2006) points out, the condition in (2.5) can actually be relaxed even within the original CCE framework. However, as we explain in detail in Section 3 of the current chapter, this requires imposing additional restrictive conditions on λ_i and Λ_i , which, if false, may well render the CCE estimator inconsistent. Hence, even if (2.5) can in principle be relaxed, in most situations of practical relevance this is not necessarily so.

The number of regressors, m , is usually a small number that is given by economic theory (and/or previous empirical evidence). Economic theory is, on the other hand, not very informative regarding the number of factors, r (see, for example, Eberhardt et al. (2013)). Therefore, the theoretically implied value of m has typically little or nothing to do with r . This is important because within the original CCE framework choosing m also means restricting r , and in many applications there is little or no reason to believe that this number should be less than or equal to $m + 1$. In empirical work, however, this aspect is almost always ignored.

In the current chapter we take this shortcoming as our starting point. The purpose is to provide a simple modification of the original CCE approach allowing (but not requiring) $r > m + 1$. The idea is to consider not only the equal-weighted average, but also other combinations of $\mathbf{w}_{1t}, \dots, \mathbf{w}_{Nt}$. In particular, by considering k such combinations we can allow for $k(m+1) \geq m+1$ common factors. In addition to the larger number of factors that can be allowed, the new approach also enables one to consider separately the selection of m and r . The asymptotic properties of the proposed combination-augmented CCE (C³E) estimator are studied under the assumption that $N, T \rightarrow \infty$ with $\sqrt{T}/N \rightarrow 0$, which is less restrictive than the $T/N \rightarrow 0$ condition of Pesaran (2006). Some Monte Carlo results are also provided to suggest the asymptotic properties are borne out well in small samples.

The remainder of the chapter is organized as follows. Section 2.2 gives the assumptions, which are used in Section 2.3 to derive the asymptotic distribution of

the proposed C³E estimator. When $T/N \rightarrow \tau < \infty$ but $\sqrt{T}/N \rightarrow 0$ the estimator is biased. As a response to this, we propose using bias correction, a procedure that is shown to be effective provided that $\sqrt{T}/N \rightarrow 0$ and $\sqrt{N}/T \rightarrow 0$. As a solution to the practical problem of how to pick the appropriate combinations, an information criterion (IC)-based selection rule is proposed. Section 2.4 focuses on the finite-sample accuracy of the theory provided in Section 2.3. Section 2.5 offers a conclusion.

2.2 Assumptions

The assumptions placed on $\boldsymbol{\epsilon}_i$, $\boldsymbol{\eta}_i$ and \mathbf{F} may be summarized in the following way.

Assumption 2.1.

- (i) $\boldsymbol{\epsilon}_{it} = \phi_i(L)\boldsymbol{\varepsilon}_{it}$, where $\boldsymbol{\varepsilon}_{it}$ is independent and identically distributed (i.i.d.) with $E(\boldsymbol{\varepsilon}_{it}) = 0$, $E(\boldsymbol{\varepsilon}_{it}^2) = 1$, $E(\boldsymbol{\varepsilon}_{it}^4) \leq M$, $\phi_i(L) = \sum_{j=0}^{\infty} \phi_{ji}L^j$, $\sigma_{\boldsymbol{\epsilon},i}^2 = E(\boldsymbol{\epsilon}_{it}^2) = \sum_{j=0}^{\infty} \phi_{ji}^2 > 0$, and $\sum_{j=0}^{\infty} j|\phi_{ji}| \leq M$;
- (ii) $\boldsymbol{\eta}_{it} = \Phi_i(L)\mathbf{v}_{it}$, where \mathbf{v}_{it} is i.i.d. with $E(\mathbf{v}_{it}) = 0$, $E(\mathbf{v}_{it}\mathbf{v}_{it}') = I_m$, $E(\|\mathbf{v}_{it}\|^4) \leq M$, $\Phi_i(L) = \sum_{j=0}^{\infty} \Phi_{ji}L^j$, $\boldsymbol{\Sigma}_{\boldsymbol{\eta},i} = E(\boldsymbol{\eta}_{it}\boldsymbol{\eta}_{it}') = \sum_{j=0}^{\infty} \Phi_{ji}\Phi_{ji}'$ is positive definite, and $\sum_{j=0}^{\infty} j\|\Phi_{ji}\| \leq M$;
- (iii) \mathbf{f}_t is covariance stationary such that $E(\|\mathbf{f}_t\|^4) \leq M$ and $E(\mathbf{f}_t\mathbf{f}_t') = \boldsymbol{\Sigma}_f$ is positive definite;
- (iv) $\boldsymbol{\lambda}_i$ and $\boldsymbol{\Lambda}_i$ are either random such that $E(\|\boldsymbol{\lambda}_i\|) \leq M$ and $E(\|\boldsymbol{\Lambda}_i\|) \leq M$, or non-random such that $\|\boldsymbol{\lambda}_i\| < \infty$ and $\|\boldsymbol{\Lambda}_i\| < \infty$;
- (v) $\boldsymbol{\varepsilon}_{it}$, \mathbf{v}_{it} and \mathbf{f}_t are mutually independent.

Remark 2.1. Assumption 2.1 is very similar to Assumptions 1–3 in Pesaran (2006), and we therefore refer to this paper for a discussion. The main difference is that while in Pesaran (2006) (Assumption 3) $\boldsymbol{\lambda}_i$ and $\boldsymbol{\Lambda}_i$ are assumed to be i.i.d. with mean $\boldsymbol{\lambda}$ and $\boldsymbol{\Lambda}$, respectively, and positive definite covariance matrices, here $\boldsymbol{\lambda}_i$ and $\boldsymbol{\Lambda}_i$ can be either random in a general (non-i.i.d.) way or non-random.³ Another difference is that while we only consider the case when $\boldsymbol{\beta}$ is homogenous, Pesaran (2006) also considers the more general case when $\boldsymbol{\beta}$ varies randomly across

³Either way, the type of cross-section dependence permitted is of the strong form. However, in analogy with Chudik et al. (2011) (Section 4) some of the factors can also be non-strong without affecting the results.

the cross-section. However, as we explain in Section 2.3, while this random specification of β is certainly more general when it comes to the types of slope behavior that can be permitted, there is also a “cost” to this greater generality.

It is convenient to cast everything in terms of stacked N -dimensional vectors. Let us therefore define the $N(m+1) \times N(m+1)$ matrix $\Sigma_u = \text{diag}(\Sigma_{u,1}, \dots, \Sigma_{u,N})$, where

$$\Sigma_{u,i} = E(\mathbf{u}_{it}\mathbf{u}'_{it}) = \begin{bmatrix} \beta'\Sigma_{\eta,i}\beta + \sigma_{\epsilon,i}^2 & \beta'\Sigma_{\eta,i} \\ \Sigma_{\eta,i}\beta & \Sigma_{\eta,i} \end{bmatrix}.$$

For each column in \mathbf{W}_i , we consider k cross-section combinations, as given by the $T \times (m+1)k$ matrix $N^{-1} \sum_{i=1}^N \mathbf{W}_i \mathbf{Z}_i$, where $\mathbf{Z}_i = (\mathbf{I}_{m+1} \otimes \mathbf{z}'_i)$ is $(m+1) \times (m+1)k$ and \mathbf{z}_i is $k \times 1$. In stacked form, we have $\mathbf{Z} = (\mathbf{Z}'_1, \dots, \mathbf{Z}'_N)'$ is $(\mathbf{Z}_A \otimes \mathbf{I}_{m+1})$, a $N(m+1) \times (m+1)k$ matrix, where $\mathbf{z} = (\mathbf{z}_1, \dots, \mathbf{z}_N)'$ is $N \times k$. The columns of \mathbf{z} will henceforth be referred to as “combinations” with \mathbf{z} being the “combination matrix”. The corresponding $N(m+1) \times r$ matrix stacking \mathbf{C}_i is given by $\mathbf{C} = (\mathbf{C}_1, \dots, \mathbf{C}_N)'$. The columns of \mathbf{Z}_A can be deterministic and/or stochastic, provided that Assumption 2.2 is satisfied, where $\bar{\mathbf{H}} = N^{-1}\mathbf{Z}'\mathbf{C}$ is $(m+1)k \times r$.

Assumption 2.2.

- (i) $rk(\bar{\mathbf{H}}) = r \leq (m+1)k$ for all $N < \infty$ and $\bar{\mathbf{H}} \rightarrow_p \mathbf{H}$ as $N \rightarrow \infty$, where $rk(\mathbf{H}) = r$ and $\|\mathbf{H}\| < \infty$;
- (ii) \mathbf{Z}_i is either deterministic such that $\|\mathbf{Z}_i\| \leq M$, or stochastic such that $E(\|\mathbf{Z}_i\|^2) \leq M$;
- (iii) \mathbf{Z} is independent of \mathbf{u}_{it} ;
- (iv) $N^{-1}\mathbf{Z}'\Sigma_u\mathbf{Z} \rightarrow_p \Theta$ as $N \rightarrow \infty$, where Θ is a $(m+1)k \times (m+1)k$ positive definite matrix.

Remark 2.2. In terms of the above notation, the original CCE approach is obtained by setting $k = 1$, $\bar{\mathbf{H}} = \bar{\mathbf{C}}'$ and $\mathbf{z}_i = 1$, and in this case Assumption 2.2 is the same as in Pesaran (2006). Note in particular how in this case Assumption 2.2 (i) reduces to $rk(\bar{\mathbf{H}}) = rk(\bar{\mathbf{C}}') = r \leq m+1$, requiring that the number of observables must be as least as large as the number of factors. By considering $k \geq 1$ combinations we can accommodate $(m+1)k \geq m+1$ factors.

Remark 2.3. The equal-weighted average is not the only way to combine the data. In fact, as Pesaran (2006) points out, any weighting scheme satisfying granularity condition (14) in his paper will do. However, while recognizing the fact that the weights does not have to be equal, it is still just one combination/weighted average per observable that is being considered. The contribution of the present chapter is the consideration of multiple combinations, which is important, because it relaxes the requirement that $m + 1 \geq r$.

2.3 Asymptotic Results

2.3.1 The C³E Estimator

As already mentioned, since \mathbf{F} and \mathbf{C}_i are not separately identifiable, \mathbf{F} can only be estimated up to a matrix rotation. The proposed estimator $\hat{\mathbf{F}}$ of $\mathbf{F}\bar{\mathbf{H}}$ is given by

$$\hat{\mathbf{F}} = \frac{1}{N} \sum_{i=1}^N \mathbf{W}_i \mathbf{Z}_i = \frac{1}{N} \sum_{i=1}^N (\mathbf{W}_i \otimes \mathbf{z}'_i),$$

whose dimension is $T \times (m + 1)k$. The resulting estimator of $\boldsymbol{\beta}$ is given by

$$\hat{\boldsymbol{\beta}} = \left(\sum_{i=1}^N \mathbf{X}'_i \mathbf{M}_{\hat{\mathbf{F}}} \mathbf{X}_i \right)^{-1} \sum_{i=1}^N \mathbf{X}'_i \mathbf{M}_{\hat{\mathbf{F}}} \mathbf{y}_i.$$

Theorem 2.1. *Under Assumptions 2.1 and 2.2, as $(N, T) \rightarrow \infty$ with $T/N \rightarrow \tau < \infty$,*

$$\sqrt{NT}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}) \sim N(0, \bar{\boldsymbol{\Sigma}}_{\eta}^{-1} \bar{\mathbf{W}} \bar{\boldsymbol{\Sigma}}_{\eta}^{-1}) + \bar{\boldsymbol{\Sigma}}_{\eta}^{-1} \sqrt{\tau} \mathbf{B},$$

where \sim signifies asymptotic equivalence, and

$$\begin{aligned} \bar{\mathbf{W}} &= \frac{1}{N} \sum_{i=1}^N \sigma_{\epsilon, i}^2 \boldsymbol{\Sigma}_{\eta, i}, \\ \mathbf{B} &= \mathbf{B}_1 - \mathbf{B}_2 - \mathbf{B}_3, \\ \mathbf{B}_1 &= \frac{1}{N} \sum_{i=1}^N \boldsymbol{\Lambda}_i (\bar{\mathbf{H}}' \bar{\mathbf{H}})^{-1} \bar{\mathbf{H}}' \boldsymbol{\Theta} \bar{\mathbf{H}} (\bar{\mathbf{H}}' \bar{\mathbf{H}})^{-1} \boldsymbol{\lambda}_i, \\ \mathbf{B}_2 &= \frac{1}{N} \sum_{i=1}^N \boldsymbol{\Sigma}_{\eta, i}(\boldsymbol{\beta}, \mathbf{I}_m) \mathbf{Z}_i \bar{\mathbf{H}} (\bar{\mathbf{H}}' \bar{\mathbf{H}})^{-1} \boldsymbol{\lambda}_i, \end{aligned}$$

$$\mathbf{B}_3 = \frac{1}{N} \sum_{i=1}^N \sigma_{\epsilon,i}^2 \boldsymbol{\Lambda}_i (\overline{\mathbf{H}}' \overline{\mathbf{H}})^{-1} \overline{\mathbf{H}}' \mathbf{z}'_i (1, 0_m)'$$

Remark 2.4. Note that if $k = 1$ and $\mathbf{z}_i = 1$, then $\widehat{\mathbf{F}} = N^{-1} \sum_{i=1}^N (\mathbf{W}_i \otimes \mathbf{z}'_i) = \overline{\mathbf{W}}$, and we are back in the original CCE approach of Pesaran (2006).

Remark 2.5. The fact that the C³E estimator is biased when $T/N \rightarrow \tau > 0$ is not covered by the theory provided by Pesaran (2006), who assumes $T/N \rightarrow 0$. Theorem 2.1 can therefore be seen as a generalization of the results of Pesaran (2006) (Theorem 4) to the case when $T/N \rightarrow 0$ is violated, making it more relevant for applied work. The presence of bias is, however, not unexpected given the results of Bai (2009) (Theorem 3) and Westerlund and Urbain (2013a) (Theorem 2.1).

Remark 2.6. The C³E estimator considered here is based on “within” pooling, whereby the data are summed over the cross-section before taking the ratio. Another approach is to use “between” pooling, in which case the ratio is taken prior to summing over the cross-section. Pesaran (2006) considers both types of pooling. However, since in his Monte Carlo study within pooling generally leads to the best performing estimator, here we only consider this type.

The remainder of this subsection is devoted to a discussion of the necessity of the rank condition in (2.5). While under Assumptions 2.1 and 2.2 original CCE requires $m + 1 \geq r$, as is made clear in Pesaran (2006), by strengthening the assumptions placed on $\boldsymbol{\lambda}_i$ and $\boldsymbol{\Lambda}_i$ it is actually possible to allow $m + 1 < r$ (without the need for additional combinations). To illustrate this, suppose that instead of (2.1) we have $\mathbf{y}_i = \mathbf{X}_i \boldsymbol{\beta}_i + \boldsymbol{\epsilon}_i$, where $\boldsymbol{\beta}_i = \boldsymbol{\beta} + \mathbf{v}_i$ and \mathbf{v}_i is i.i.d. with mean zero and positive definite covariance matrix $\boldsymbol{\Sigma}_v$. Also, \mathbf{v}_i is independent of all other random elements of the DGP. Suppose first that \mathbf{F} is known and that $m + 1 = r$, in which case

$$\begin{aligned} \sqrt{N}(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}) &= \left(\frac{1}{NT} \sum_{i=1}^N \mathbf{X}'_i \mathbf{M}_F \mathbf{X}_i \right)^{-1} \frac{1}{\sqrt{NT}} \sum_{i=1}^N \mathbf{X}'_i \mathbf{M}_F (\mathbf{X}_i \mathbf{v}_i + \mathbf{F} \boldsymbol{\lambda}_i + \boldsymbol{\epsilon}_i) \\ &= \left(\frac{1}{NT} \sum_{i=1}^N \boldsymbol{\eta}'_i \mathbf{M}_F \boldsymbol{\eta}_i \right)^{-1} \frac{1}{\sqrt{NT}} \sum_{i=1}^N \boldsymbol{\eta}'_i \mathbf{M}_F \boldsymbol{\eta}_i \mathbf{v}_i + o_p(1) \\ &= \overline{\boldsymbol{\Sigma}}_{\boldsymbol{\eta}}^{-1} \frac{1}{\sqrt{N}} \sum_{i=1}^N \boldsymbol{\Sigma}_{\boldsymbol{\eta},i} \mathbf{v}_i + o_p(1) \end{aligned}$$

$$\sim \left(\bar{\Sigma}_\eta^{-1} \frac{1}{N} \sum_{i=1}^N \Sigma_{\eta,i} \Sigma_{\mathbf{v}} \Sigma_{\eta,i} \bar{\Sigma}_\eta^{-1} \right)^{1/2} N(0, \mathbf{I}_m),$$

where the last result follows from applying a central limit law to $N^{-1/2} \sum_{i=1}^N \Sigma_{\eta,i} \mathbf{v}_i$. Thus, in this case normality is a direct consequence of the assumed randomness of β_i . A similar result applies in case $m+1 < r$, but then normality requires that λ_i and Λ_i have similar random coefficient representations as β_i . In fact, as pointed out by Westerlund and Urbain (2013b), in this case one also has to assume that λ_i and Λ_i are mutually independent, which seems like a rather restrictive assumption. For example, when regressing investments on savings, as is commonly done in the literature on the so-called “Feldstein–Horioka puzzle”, a common shock that increases savings is going to push interest rates down and investments up, suggesting that λ_i and Λ_i should be negatively correlated. Thus, while the requirement that $m+1 \geq r$ can be relaxed also within the original CCE framework, this does not come free of charge. Then there is also the dependence of the asymptotic distribution of $\sqrt{N}(\hat{\beta} - \beta)$ on $\Sigma_{\mathbf{v}}$, which makes for relatively complicated covariance estimation. Finally, if β_i is allowed to vary, the rate of convergence is reduced, from \sqrt{NT} to \sqrt{N} , which is true also in case of a violation of $m+1 \geq r$.

One may think that the above mentioned reduction in the rate of consistency is due to the fact that β_i is allowed to vary, and that imposing $\beta_1 = \dots = \beta_N = \beta$ would prevent this from happening, regardless of whether $m+1 \geq r$ or $m+1 < r$.⁴ However, this is not the case. The reason for this is easily appreciated by replacing \mathbf{F} in the above expansion of $\sqrt{N}(\hat{\beta} - \beta)$ with $\hat{\mathbf{F}} = \bar{\mathbf{W}}$. By imposing $\mathbf{v}_1 = \dots = \mathbf{v}_N = 0$ and noting that $(NT)^{-1/2} \sum_{i=1}^N \mathbf{X}'_i \mathbf{M}_{\bar{\mathbf{W}}} \epsilon_i = O_p(1)$, the numerator $\sqrt{NT}(\hat{\beta} - \beta)$ reduces to

$$\frac{1}{\sqrt{NT}} \sum_{i=1}^N \mathbf{X}'_i \mathbf{M}_{\bar{\mathbf{W}}} (\mathbf{X}_i \mathbf{v}_i + \mathbf{F} \lambda_i + \epsilon_i) = \frac{1}{\sqrt{NT}} \sum_{i=1}^N \mathbf{X}'_i \mathbf{M}_{\bar{\mathbf{W}}} \mathbf{F} \lambda_i + O_p(1).$$

On the one hand, if $m+1 \geq r$, then, since in such cases $\hat{\mathbf{F}} = \mathbf{F}\bar{\mathbf{C}} + o_p(1)$, $\mathbf{X}'_i \mathbf{M}_{\bar{\mathbf{W}}} \mathbf{F} = \mathbf{X}'_i \mathbf{M}_{\mathbf{F}\bar{\mathbf{C}}} \mathbf{F} = 0$, and so we obtain $\sqrt{NT}(\hat{\beta} - \beta) = O_p(1)$. Hence, provided that $m+1 \geq r$, imposing $\beta_1 = \dots = \beta_N = \beta$ restores \sqrt{NT} -consistency.

⁴It is not clear from Pesaran (2006) whether one can have $\beta_1 = \dots = \beta_N = \beta$, while at the same time permitting $m+1 < r$.

On the other hand, if $m + 1 < r$, then $\mathbf{X}'_i \mathbf{M}_{\overline{\mathbf{W}}} \mathbf{F} = O_p(T)$, and therefore

$$\begin{aligned} \sqrt{NT}(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}) &= \sqrt{T} \left(\frac{1}{NT} \sum_{i=1}^N \mathbf{X}'_i \mathbf{M}_{\overline{\mathbf{W}}} \mathbf{X}_i \right)^{-1} \frac{1}{\sqrt{N}} \sum_{i=1}^N T^{-1} \mathbf{X}'_i \mathbf{M}_{\overline{\mathbf{W}}} \mathbf{F} \boldsymbol{\lambda}_i \\ &\quad + O_p(1), \end{aligned}$$

whose order is determined by the order of the first term on the right, which in turn depends on $\boldsymbol{\lambda}_i$ and $\boldsymbol{\Lambda}_i$. If $\boldsymbol{\lambda}_i$ is i.i.d. independently of $\boldsymbol{\Lambda}_i$, then the first term is $O_p(\sqrt{T})$, whereas if $\boldsymbol{\lambda}_i$ is non-i.i.d. and/or correlated with $\boldsymbol{\Lambda}_i$, then the same term is $O_p(\sqrt{NT})$. Thus, the rate of consistency is \sqrt{N} , at best, and if $\boldsymbol{\lambda}_i$ is non-i.i.d. and/or correlated with $\boldsymbol{\Lambda}_i$, then $\widehat{\boldsymbol{\beta}}$ is even inconsistent. The proposed estimator is not only very simple, but is also \sqrt{NT} -consistent regardless of the specification of $\boldsymbol{\lambda}_i$ and $\boldsymbol{\Lambda}_i$, provided that Assumptions 2.1 and 2.2 are satisfied.

2.3.2 Bias Correction

As pointed out by Bai (2009), an obvious solution to the problem with bias in the case when $T/N \rightarrow \tau > 0$ is to use bias correction. Let us therefore define the following bias-adjusted version of $\widehat{\boldsymbol{\beta}}$:

$$\widehat{\boldsymbol{\beta}}_{BA} = \widehat{\boldsymbol{\beta}} - N^{-1} \widehat{\boldsymbol{\Sigma}}_{\eta}^{-1} \widehat{\mathbf{B}}.$$

Here $\widehat{\mathbf{B}}$ is \mathbf{B} that is defined in Theorem 2.1, with $\boldsymbol{\beta}$, $\boldsymbol{\Sigma}_{\eta,i}$ and $\sigma_{\epsilon,i}^2$ replaced by $\widehat{\boldsymbol{\beta}}$, $\widehat{\boldsymbol{\Sigma}}_{\eta,i}$ and $\widehat{\sigma}_{\epsilon,i}^2$, respectively, where $\widehat{\boldsymbol{\Sigma}}_{\eta,i}$ and $\widehat{\sigma}_{\epsilon,i}^2$ are the usual heteroskedasticity and autocorrelation consistent (HAC) covariance matrix estimators based on $\mathbf{M}_{\widehat{\mathbf{F}}} \mathbf{X}_i$ and $(\mathbf{y}_i - \mathbf{X}_i \widehat{\boldsymbol{\beta}} - \widehat{\mathbf{F}} \widehat{\boldsymbol{\lambda}}_i)$, respectively, and $\widehat{\boldsymbol{\Sigma}}_{\eta} = N^{-1} \sum_{i=1}^N \widehat{\boldsymbol{\Sigma}}_{\eta,i}$. The estimators $\widehat{\boldsymbol{\lambda}}_i$ and $\widehat{\boldsymbol{\Lambda}}_i$ of $\boldsymbol{\lambda}_i$ and $\boldsymbol{\Lambda}_i$, respectively, are obtained by simply picking the appropriate elements in $\widehat{\mathbf{C}}_i$, which can be obtained from a LS regression of \mathbf{w}_{it} onto $\widehat{\mathbf{f}}_t$.

Proposition 2.3.1. *Under Assumptions 2.1 and 2.2, as $(N, T) \rightarrow \infty$ with $\sqrt{T}/N \rightarrow 0$ and $\sqrt{N}/T \rightarrow 0$,*

$$\sqrt{NT}(\widehat{\boldsymbol{\beta}}_{BA} - \boldsymbol{\beta}) \sim N(0, \overline{\boldsymbol{\Sigma}}_{\eta}^{-1} \overline{\mathbf{W}} \overline{\boldsymbol{\Sigma}}_{\eta}^{-1}).$$

According to Proposition 2.3.1, asymptotically the bias correction is successful. Moreover, the correction does not contribute to the limiting variance, which is the same as in Theorem 2.1.

2.3.3 Selecting the Combinations

A problem in applications is how to pick the combinations, \mathbf{z}_i . The most obvious approach is to exploit if there are natural candidates. \mathbf{z}_i is ideally chosen to be uncorrelated with ϵ_{it} , but highly correlated with \mathbf{C} , measuring the exposure of each individual unit to the common shocks. As such, natural candidates for \mathbf{z}_i are actually not difficult to find. For example, if \mathbf{w}_{it} includes GDP and/or inflation, then \mathbf{z}_i might be a vector of trade shares, or, if \mathbf{w}_{it} includes stock prices, then \mathbf{z}_i might be the book-to-market ratio and/or the earnings-price ratio (aggregated over the whole time period). In fact, \mathbf{z}_i can even include preliminary consistent estimates of (the space spanned by) \mathbf{C} . The only requirement is that the rate of consistency must be at least \sqrt{N} , which is sufficiently relaxed to enable estimation by, for example, principal components (see Bai (2003), Theorem 2). Deterministic instruments are even simpler to come by. Again, if \mathbf{w}_{it} includes GDP, then \mathbf{z}_i might be a vector of distance rankings, or, if \mathbf{w}_{it} includes stock prices, then \mathbf{z}_i might be industrial classification. Also, as Chudik et al. (2011) show, the cross-sectional average can be quite effective in mopping up cross-section dependence. Thus, a vector of ones is always a good starting point when constructing \mathbf{z}_i .

Even if candidates for \mathbf{z}_i are easy to find, in practice there will most likely be uncertainty regarding which candidates to use. If this is the case, then \mathbf{z}_i can be estimated by an IC of the form

$$IC(s) = V(s, \widehat{\mathbf{F}}^s) + sg(N), \quad (2.6)$$

where $\widehat{\mathbf{F}}^s$ is $\widehat{\mathbf{F}}$ based on s combinations,

$$V(s, \widehat{\mathbf{F}}^s) = (NT)^{-1} \sum_{i=1}^N (\mathbf{y}'_i \mathbf{M}_{\mathbf{X}_i} \mathbf{M}_{\widehat{\mathbf{F}}^s} \mathbf{M}_{\mathbf{X}_i} \mathbf{y}_i)$$

is the sum of squared residuals, and $g(N)$ is a penalty function. To fix ideas, suppose for sake of argument that m is known and that the columns of \mathbf{z} are ordered according to relevance, as measured by the (cross-section) correlation with (the unobserved) columns of \mathbf{C} . The problem of determining the columns \mathbf{z}_i to be included in the final version of \mathbf{z} therefore reduces to the problem of consistent estimation of k , the dimension of \mathbf{z} . The IC estimator \widehat{k} of k is given by

$$\widehat{k} = \arg \min_{s=0, \dots, \mathbf{K}_{max}} IC(s),$$

where $\mathbf{K}_{max} \geq k$. By Assumption 2, $k \geq \lfloor r/(m+1) \rfloor$, where $\lfloor x \rfloor$ is the largest integer less than x . Hence, we do not propose to use \widehat{k} as an estimator of the number of factors, r , but as an estimator of the minimal number of columns of \mathbf{z} required to approximate the underlying factor structure.

Proposition 2.3.2. *Under Assumptions 2.1 and 2.2, if $g(N) \rightarrow 0$ and $Ng(N) \rightarrow \infty$, as $N \rightarrow \infty$,*

$$P(\widehat{k} = k) \rightarrow 1.$$

Remark 2.7. Bai and Ng (2002) proposes several IC that are appropriate in the context of principal components estimation approximate common factor models. The class of IC considered here is very similar. The only difference is the penalty, which is here assumed not to depend on T . The reason for this difference is that while the rate of consistency of the principal components estimator of \mathbf{F} depends on both N and T , the rate of consistency of $\widehat{\mathbf{F}}$ only depends on N .

Remark 2.8. In order to appreciate the implications of Proposition 2.3.2 for the asymptotic distribution of $\sqrt{NT}(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta})$ it is useful to treat $\widehat{\boldsymbol{\beta}}$ as a function k . Let us therefore write $\widehat{\boldsymbol{\beta}}^k$ for $\widehat{\boldsymbol{\beta}}$. Clearly,

$$\begin{aligned} P[\sqrt{NT}|\widehat{\boldsymbol{\beta}}^k - \boldsymbol{\beta}| \leq \delta] &= P[\sqrt{NT}|\widehat{\boldsymbol{\beta}}^k - \boldsymbol{\beta}| \leq \delta | \widehat{k} = k]P(\widehat{k} = k) \\ &+ P[\sqrt{NT}|\widehat{\boldsymbol{\beta}}^k - \boldsymbol{\beta}| \leq \delta | \widehat{k} \neq k]P(\widehat{k} \neq k), \end{aligned}$$

where $\delta > 0$. Because $P(\widehat{k} = k) \rightarrow 1$ and $P(\widehat{k} \neq k) \rightarrow 0$, the second term on the right-hand side converges to zero, and $P[\sqrt{NT}|\widehat{\boldsymbol{\beta}}^k - \boldsymbol{\beta}| \leq \delta] = 1 + o(1)$. Moreover, conditional on $\widehat{k} = k$, $\widehat{\boldsymbol{\beta}}^k = \widehat{\boldsymbol{\beta}}$. Thus,

$$|P[\sqrt{NT}|\widehat{\boldsymbol{\beta}}^k - \boldsymbol{\beta}| \leq \delta] - P[\sqrt{NT}|\widehat{\boldsymbol{\beta}}^k - \boldsymbol{\beta}| \leq \delta]| \rightarrow 0,$$

suggesting that Theorem 2.1 is unaffected by the estimation of k .

The minimization can be done via grid search over all possible combinations of columns of \mathbf{z} ; however, given that the large number of potential candidates, a computationally less demanding procedure seems preferable. To implement the minimization, the following two-step approach may be used:

1. We begin by ordering the candidates. This can be done by considering the cross-section correlation between each element of \mathbf{z} and each element

of \mathbf{C}_i . But since the latter is not observed, we propose to use $\overline{\mathbf{W}}_i = T^{-1} \sum_{t=1}^T \mathbf{w}_{it} = \mathbf{C}'_i \overline{\mathbf{F}} + \overline{\mathbf{u}}_i = \mathbf{C}'_i \overline{\mathbf{F}} + o_p(1)$ to estimate the space spanned by \mathbf{C}_i . This practice leads to a $(m+1) \times k$ matrix of correlations, and hence $(m+1)$ potentially different orderings. As a way of summarizing these we propose to take the average of the correlations for each candidate over the elements of $\overline{\mathbf{W}}_i$.

2. Once an ordering of the candidates is obtained, the problem of selecting which combinations to use reduces to the relatively simple problem of selecting the number of combinations, which can be done by using an IC. The specific IC used in this chapter can be seen as a version of the IC_{p1} criterion of Bai and Ng (2002) and is given by

$$IC(s) = \ln(V(s, \widehat{\mathbf{F}}^s)) + sN^{-1} \ln(N).$$

Proposition 2.3.2 implies that \widehat{k} based on this criterion is consistent for k .

2.4 Monte Carlo Simulations

In this section we evaluate briefly the small sample properties of the C³E estimator. The DGP used for this purpose can be seen as a restricted version of (2.1)–(2.3), and sets $m = 1$, $\beta = -2$ and $(F'_t, \eta_{it}, \epsilon_{it}) \sim N(0, \mathbf{I}_{r+2})$. The difference between the experiments considered lies in how we generate $\boldsymbol{\lambda}_i$, $\boldsymbol{\Lambda}_i$ and \mathbf{z}_i . Seven experiments, denoted E1–E7, are considered. In E1–E3, the rank condition in (2.5) is satisfied, whereas in E4–E7, the condition is violated. In E4 and E5, $\boldsymbol{\lambda}_i$ and $\boldsymbol{\Lambda}_i$ are i.i.d. and independent of each other, as required in original CCE, whereas in E6 and E7, $\boldsymbol{\lambda}_i$ and $\boldsymbol{\Lambda}_i$ are non-i.i.d.. Exactly how $\boldsymbol{\lambda}_i$, $\boldsymbol{\Lambda}_i$ and \mathbf{z}_i are generated is described in Table 2.1 below.

For each experiment, 19 (N, T) pairs are considered. In the first 16, $N, T = 30, 50, 100, 200$, whereas in last three, $N = \lfloor T^{4/3} \rfloor$. The motivation behind the last three pairs is to assess the performance when the conditions of Theorem 2.1 and Proposition 2.3.1 (that is, $\sqrt{N}/T \rightarrow 0$ and $\sqrt{T}/N \rightarrow 0$) are met. We report the bias and the size of a nominal 5% level t -test. The reported results are based on 5,000 replications and the parameters of the models are kept constant across replications.

The performance of C³E is compared with that of the naive LS estimator where the cross-sectional dependence is ignored, original CCE, and the principal

Table 2.1: Description of experiments.

Exp.	r	Observables	Factor loadings	Combinations
E1	2	$y_{it} = \beta x_{it} + \lambda_{1i} f_{1t} + \lambda_{2i} f_{2t} + \epsilon_{it}$ $x_{it} = \Lambda_{1i} f_{1t} + \Lambda_{2i} f_{2t} + \eta_{it}$	$\lambda_{1i} = z_{1i} + z_{2i} + \tau_{\lambda 1i}$ $\lambda_{2i} = z_{1i} + z_{2i} + \tau_{\lambda 2i}$ $\Lambda_{1i} = z_{1i} + \tau_{\Lambda 1i}$ $\Lambda_{2i} = z_{2i} + \tau_{\Lambda 2i}$	$z_{1i} \sim N(1.0, 0.5)$ $z_{2i} \sim N(0.2, 0.5)$
E2	2	$y_{it} = \beta x_{it} + \lambda_{1i} f_{1t} + \lambda_{2i} f_{2t} + \epsilon_{it}$ $x_{it} = \Lambda_{1i} f_{1t} + \Lambda_{2i} f_{2t} + \eta_{it}$	$\lambda_{1i} = z_{1i} + z_{2i} + \tau_{\lambda 1i}$ $\lambda_{2i} = z_{1i} + z_{3i} + \tau_{\lambda 2i}$ $\Lambda_{1i} = z_{1i} + z_{3i} \tau_{\Lambda 1i}$ $\Lambda_{2i} = z_{4i} + \tau_{\Lambda 2i}$	$z_{1i} \sim N(1.0, 0.5)$ $z_{2i} \sim N(0.2, 0.5)$ $z_{3i} \sim N(0.5, 0.5)$ $z_{4i} \sim N(0.5, 0.5)$
E3	2	$y_{it} = \beta x_{it} + \lambda_{1i} f_{1t} + \lambda_{2i} f_{2t} + \epsilon_{it}$ $x_{it} = \Lambda_{1i} f_{1t} + \Lambda_{2i} f_{2t} + \eta_{it}$	$\lambda_{1i} \sim N(1.0, 0.2)$ $\lambda_{2i} \sim N(0.5, 0.5)$ $\Lambda_{1i} \sim N(0.5, 0.5)$ $\Lambda_{2i} = 0$	$z_{1i} \sim N(1.0, 0.5)$ $z_{2i} \sim N(0.2, 0.5)$ $z_{3i} \sim N(0.5, 0.5)$ $z_{4i} \sim N(0.5, 0.5)$
E4	2	$y_{it} = \beta x_{it} + \lambda_{1i} f_{1t} + \lambda_{2i} f_{2t} + \epsilon_{it}$ $x_{it} = \Lambda_{1i} f_{1t} + \Lambda_{2i} f_{2t} + \eta_{it}$	$\lambda_{1i} = z_{1i} + 0.5\tau_{\lambda 1i}$ $\lambda_{2i} = z_{2i} + 0.5\tau_{\lambda 2i}$ $\Lambda_{1i} = z_{3i} + 0.5\tau_{\Lambda 1i}$ $\Lambda_{2i} = z_{4i} + 0.5\tau_{\Lambda 2i}$	$z_{1i} \sim N(1.0, 0.5)$ $z_{2i} \sim N(0, 0.5)$ $z_{3i} \sim N(0.5, 0.5)$ $z_{4i} \sim N(0, 0.5)$
E5	3	$y_{it} = \beta x_{it} + \lambda_{1i} f_{1t} + \lambda_{2i} f_{2t}$ $\quad + \lambda_{3i} f_{3t} + \epsilon_{it}$ $x_{it} = \Lambda_{1i} f_{1t} + \Lambda_{2i} f_{2t}$ $\quad + \Lambda_{3i} f_{3t} + \eta_{it}$	$\lambda_{1i} \sim N(1.0, 0.2)$ $\lambda_{2i} \sim N(0.5, 0.5)$ $\lambda_{3i} \sim N(0, 0.5)$ $\Lambda_{1i} \sim N(0.5, 0.5)$ $\Lambda_{2i} \sim N(0.2, 0.5)$ $\Lambda_{3i} = 0$	$z_{1i} \sim N(1.0, 0.5)$ $z_{2i} \sim N(0.2, 0.5)$ $z_{3i} \sim N(0.5, 0.5)$ $z_{4i} \sim N(0.5, 0.5)$
E6	2	$y_{it} = \beta x_{it} + \lambda_{1i} f_{1t} + \lambda_{2i} f_{2t} + \epsilon_{it}$ $x_{it} = \Lambda_{1i} f_{1t} + \Lambda_{2i} f_{2t} + \eta_{it}$	$\lambda_{1i} = z_{1i} + z_{4i} + \tau_{\lambda 1i}$ $\lambda_{2i} = z_{1i} - 5z_{3i} + \tau_{\lambda 2i}$ $\Lambda_{1i} = z_{3i} + \tau_{\Lambda 1i}$ $\Lambda_{2i} = z_{3i} - z_{4i} + \tau_{\Lambda 2i}$	$z_{1i} \sim N(1.0, 0.5)$ $z_{2i} \sim N(0.5, 0.5)$ $z_{3i} \sim N(0.2, 0.5)$ $z_{4i} \sim N(0.2, 0.5)$
E7	2	$y_{it} = \beta x_{it} + \lambda_{1i} f_{1t} + \lambda_{2i} f_{2t} + \epsilon_{it}$ $x_{it} = \Lambda_{1i} f_{1t} + \Lambda_{2i} f_{2t} + \eta_{it}$	$\lambda_{1i} = z_{1i} - 5z_{2i} + \tau_{\lambda 1i}$ $\lambda_{2i} = z_{1i} - 2z_{3i} + \tau_{\lambda 2i}$ $\Lambda_{1i} = z_{1i} - 2z_{3i} + \tau_{\Lambda 1i}$ $\Lambda_{2i} = z_{4i} + \tau_{\Lambda 2i}$	$z_{1i} \sim N(1.0, 0.5)$ $z_{2i} \sim N(0.2, 0.5)$ $z_{3i} \sim N(0.5, 0.5)$ $z_{4i} \sim N(0, 0.5)$

Notes: $(\tau_{\lambda 1i}, \tau_{\lambda 2i}, \tau_{\Lambda 1i}, \tau_{\Lambda 2i})' \sim N(\mathbf{1}_4, I_4)$, where $\mathbf{1}_4 = (1, 1, 1, 1)'$.

components (PC) estimator of Bai (2009). Four versions of C³E is considered; (i) C³E based on the combinations reported in Table 2.1, (ii) C³E|1 based on the Table 2.1-combinations plus the cross-sectional average of the observeables, (iii) C³ES, which is C³E based on IC selection of the Table 2.1-combinations and, (iv) C³ES|1, which is C³E based on IC selection among the combinations reported in Table 2.1 plus the cross-sectional averages. The reason for the special treatment of the cross-section averages is that we would like to be able to study separately a violation of the CCE rank condition in (2.5) and the one in Assumption 2.2. As explained in Section 2.3 C³E|1 and C³ES|1 are obtained by simply including in \mathbf{z} a vector of ones. There estimators therefore nest original CCE. The selection of the combinations (in C³ES and C³ES|1) is done as explained in Section 2.3.3. We begin by ordering the combinations according to their cross-section correlation

with the time averages of the observables. Once an ordering is obtained, starting with the full set of combinations, the proposed IC is used to select the appropriate combinations to use.

2.4.1 Simulation Results

The results for each experiment are reported in Tables 2.E1–2.E7. The conclusions that can be drawn from these tables are summarized as follows.

- E1.** The aim of this experiment is to evaluate the accuracy of the theoretically predicted bias (as given in Theorem 2.1) and also to evaluate the performance of the bias-adjusted C^3E estimator.⁵ The results reported in Table 2.E1 suggest that if $N < T$, then the bias of the C^3E estimator is slightly smaller than expected, whereas if $N \geq T$, then the empirical bias is generally very close to the one predicted by theory. As expected, bias correction leads to a considerable improvement for all estimators. The overall best performance is obtained by using bias-adjusted C^3E , followed by CCE and PC.
- E2.** The aim of this experiment is to compare the performance of original CCE and C^3E when all the conditions required for both methods to be valid, including (2.5), are met. All the combinations are highly correlated with the loadings. Under these conditions both CCE and (the versions of) C^3E should perform equally well. This is reflected in Table 2.E2. CCE tend to perform slightly better, but the difference is not very large, and it gets smaller as N and T increases. As expected given Proposition 2.3.2, in large samples C^3ES and $C^3E|1$ do just as well as C^3E and $C^3E|1$, respectively.
- E3.** In this experiment the combinations are generated independently of the loadings, such that Assumption 2.2 is violated. The expectation is that this should cause C^3E to break down. However, (2.5) is still satisfied, and therefore CCE is still expected to perform well. Interestingly, this is not what we see in Table 2.E3. Specifically, while the performance of the C^3E estimators is generally worse than in E2, the other estimators also perform worse than in E2. In fact, in most of the cases C^3E outperforms the competition, including CCE. A comparison of C^3E and $C^3E|1$ shows that when the loadings are generated independently of the combinations, including a vector of

⁵The theoretical bias of the CCE and C^3E estimators are obtained from Theorem 2.1. In case of the PC estimator the theoretical prediction is taken from Bai (2009).

ones to the combination matrix decreases the bias for most of the N and T combinations. We also experimented with a subset of invalid combinations and obtained very similar results.

- E4.** This experiment is conducted to compare the performance of CCE and the versions of C^3E when (2.5) is not satisfied, as in this case

$$E(\mathbf{C}_i) = \begin{bmatrix} 1.5 & 0.5 \\ 0 & 0 \end{bmatrix}$$

(see Table 2.1). But the factor loadings are still i.i.d. The theory suggests that in this case the rate of consistency of CCE should be reduced from \sqrt{NT} to \sqrt{N} . The results reported in Table 2.E4 are quite suggestive of this; while decreasing in N , the bias is roughly constant in T . We also see that the size of the bias of the CCE is substantially higher when compared to E3. But while the bias is constant, the size distortions of the CCE are actually increasing in T . Moreover, although the bias is generally decreasing in N , the size distortions remain or even increase with increases in N . This is unexpected, because inference based on the CCE t -statistic should be asymptotically valid regardless of the relative expansion rate of N and T (see Pesaran, 2006, Theorem 3). By contrast, the C^3E t -statistic maintains the nominal level well and the accuracy increases with increases in N and/or T . The \sqrt{NT} -consistency of the C^3E estimator is also clearly visible; increasing N and/or T leads a reduction in bias. Furthermore including a vector of ones to the combination matrix (so that C^3E nests original CCE) leads to a decrease in the bias.

- E5.** In this experiment (2.5) is again violated; however, this time the violation is due to the presence of too many factors; $m + 1 = 2 < r = 3$. Also, the loadings are independent of the combinations. The results reported in Table 2.E5 show that while the conclusions regarding the CCE bias are roughly the same as in E4, the size distortions of this estimator are now markedly smaller. However, there is still a tendency for the distortions to increase with T . As expected, since there are now even more invalid combinations than in E3, the C^3E results reported in Table 2.E5 are worse than those reported in Table 2.E3. However, in most of the cases, performance is still better than for CCE.

E6. In E6,

$$E(\mathbf{C}_i) = \begin{bmatrix} 1.7 & 0.2 \\ 0 & 0 \end{bmatrix}$$

and therefore (2.5) is violated. Also, λ_{2i} and λ_{1i} are correlated with each other, so the factor loadings are non-i.i.d.. As expected, CCE breaks down completely. However, since the combinations are still correlated with the loadings, C³E still performs well.

E7. All loadings are mean zero and non-i.i.d. in this experiment. The CCE estimator is now even more biased than in E7. The C³E estimator still performs well, as does C³ES.

2.5 Conclusion

This chapter considers the problem of consistent estimation of a factor-augmented panel regression model in which the number of factors, r , is potentially larger than the number of observables, $m+1$. The estimator that we propose can be viewed as an extension of the CCE estimator of Pesaran (2006), which is based on using the cross-section averages of the observables as proxies for the latent factors. While CCE does allow $r > m+1$, it does so at a cost. In particular, it is required that the factor loadings are i.i.d., which in most cases of practical relevance is likely to be violated. Also, even if the assumption is met, the rate of consistency is only \sqrt{N} , as opposed to the usual \sqrt{NT} -rate. In this chapter we take this feature of CCE as our starting point. The purpose is to provide a simple extension that preserves \sqrt{NT} -consistency without for that matter requiring i.i.d. loadings.

The idea behind the proposed C³E estimator is to use not only the cross-section average but also other (cross-section) combinations of the observables. By taking $k \geq 1$ such combinations we can allow $k(m+1) \geq m+1$ common factors. The new estimator is shown to be \sqrt{NT} -consistent and asymptotically normal under the condition that $T/N \rightarrow \tau < \infty$. This condition is more general than the $T/N \rightarrow 0$ condition of Pesaran (2006), whose relaxation is shown to have important consequences. In particular, it is shown that the estimator is biased whenever $\tau > 0$. As a response to this, a bias-adjusted C³E estimator is proposed, which is shown to support asymptotically normal inference under $\sqrt{T}/N \rightarrow 0$ and $\sqrt{N}/T \rightarrow 0$. This is true if the combinations are known. If there

is uncertainty over which combinations to use an IC can be used to select the appropriate combinations.

The small sample performance of the C^3E estimator is examined through a series of Monte Carlo experiments. The results suggest that whenever the assumptions of Pesaran (2006) are satisfied, the performance of the CCE and C^3E estimators are comparable. If, however, the assumptions are not met, then the C^3E estimator continues to work well while the CCE estimator breaks down. We also find that the proposed bias-adjustment and IC-based combination selection procedures seem to work well, leading to estimators with good small-sample properties.

2.A Tables

Table 2.E1: Bias and bias correction.

N	T	Bias $\times \sqrt{NT}$								
		PC	Theory	BA	CCE	Theory	BA	C ³ E	Theory	BA
30	30	2.37	2.20	0.34	0.64	0.71	0.01	0.24	0.31	-0.01
50	30	1.80	1.70	0.16	0.24	0.26	-0.01	0.13	0.20	-0.06
100	30	1.35	1.32	0.07	0.23	0.27	-0.02	0.10	0.10	-0.06
200	30	1.01	0.97	0.06	0.11	0.12	-0.00	0.02	0.03	-0.00
30	50	2.98	2.84	0.36	0.82	0.92	0.02	0.31	0.40	-0.02
50	50	2.28	2.19	0.16	0.31	0.34	-0.02	0.17	0.25	-0.07
100	50	1.74	1.71	0.08	0.32	0.35	-0.01	0.15	0.13	-0.05
200	50	1.27	1.26	0.03	0.13	0.15	-0.02	0.01	0.03	-0.02
30	100	4.18	4.01	0.46	1.16	1.30	0.02	0.45	0.56	-0.02
50	100	3.20	3.10	0.20	0.46	0.48	-0.01	0.26	0.36	-0.09
100	100	2.45	2.41	0.09	0.47	0.49	-0.00	0.22	0.18	-0.07
200	100	1.81	1.78	0.05	0.22	0.22	0.00	0.05	0.05	-0.00
30	200	5.89	5.68	0.62	1.64	1.84	0.03	0.62	0.80	-0.04
50	200	4.47	4.39	0.23	0.64	0.68	-0.02	0.36	0.51	-0.13
100	200	3.48	3.42	0.14	0.69	0.70	0.01	0.33	0.26	-0.08
200	200	2.53	2.52	0.05	0.31	0.31	-0.00	0.06	0.07	-0.01
93	30	1.45	1.37	0.12	0.27	0.27	0.00	0.13	0.08	-0.08
184	50	1.34	1.32	0.04	0.20	0.21	-0.00	0.05	0.04	-0.03
464	100	1.21	1.20	0.01	0.10	0.12	-0.01	-0.01	-0.01	-0.01

Table 2.E1 cont'd: Bias and bias correction.

N	T	Size					
		PC	BA	CCE	BA	C ³ E	BA
30	30	57.9	10.4	11.3	9.5	8.0	12.2
50	30	40.8	8.5	6.7	8.4	7.7	12.0
100	30	27.4	7.1	6.1	6.6	7.1	11.1
200	30	18.2	6.2	6.0	6.1	6.5	10.2
30	50	78.5	9.7	13.9	9.2	8.3	9.9
50	50	59.2	7.3	6.7	7.3	6.3	8.9
100	50	39.6	6.3	6.6	6.1	6.3	8.5
200	50	24.8	5.9	6.0	6.3	6.7	8.2
30	100	96.4	10.0	23.4	8.9	8.4	8.6
50	100	86.1	7.7	8.3	7.4	7.0	7.5
100	100	66.7	6.5	8.2	6.5	6.5	7.0
200	100	43.5	6.1	5.6	5.6	5.7	6.3
30	200	100.0	12.1	39.0	9.4	11.1	8.1
50	200	98.9	6.7	9.6	6.3	6.8	7.0
100	200	92.1	6.5	11.5	6.2	6.9	6.0
200	200	69.9	5.7	6.4	5.6	5.6	5.9
93	30	30.8	7.5	7.1	7.4	7.4	11.7
184	50	26.8	6.0	6.0	6.0	5.8	7.7
464	100	22.8	5.7	5.7	5.8	5.6	6.7

Notes: “PC” and “CCE” refer to the principal components estimator of Bai (2009), and the common correlated effects estimator of Pesaran (2006), respectively. “C³E” is the modified CCE estimator proposed in this chapter. “Theory” and “BA” refer to the theoretical prediction and bias-adjusted version of each estimator. The theoretical bias of the CCE and C³E estimators are obtained from Theorem 2.1. In case of the PC estimator the theoretical prediction is taken from Bai (2009).

2 CCE ESTIMATION WITH MORE FACTORS THAN OBSERVABLES

Table 2.E2: All conditions of PC, CCE and C³E are satisfied.

		Bias \times 100						
N	T	LS	PC	CCE	C ³ E	C ³ E 1	C ³ ES	C ³ ES 1
30	30	44.61	7.75	-0.73	0.27	0.29	0.28	0.20
50	30	41.34	4.89	0.08	0.17	0.14	0.17	0.14
100	30	42.77	2.31	0.10	0.16	0.16	0.16	0.15
200	30	39.14	1.06	-0.07	0.04	0.04	0.04	0.04
30	50	44.87	7.62	-0.71	0.26	0.30	0.53	-0.53
50	50	41.66	4.78	0.07	0.23	0.20	0.42	0.10
100	50	42.52	2.32	0.06	0.14	0.13	0.14	0.12
200	50	39.51	1.04	-0.09	0.04	0.03	0.04	0.02
30	100	44.89	7.57	-0.68	0.31	0.35	0.59	-0.69
50	100	41.73	4.77	0.10	0.25	0.22	0.62	0.10
100	100	42.64	2.32	0.08	0.15	0.15	0.27	0.07
200	100	39.51	1.05	-0.07	0.04	0.04	0.04	0.03
30	200	44.97	7.54	-0.67	0.31	0.36	0.58	-0.67
50	200	41.71	4.72	0.06	0.22	0.20	0.58	0.06
100	200	42.70	2.29	0.04	0.14	0.14	0.23	0.07
200	200	39.53	1.04	-0.08	0.04	0.04	0.05	-0.08
93	30	43.19	2.56	0.05	0.17	0.18	0.17	0.14
184	50	39.66	1.15	-0.07	0.04	0.04	0.04	0.02
464	100	42.35	0.44	-0.11	0.00	0.00	0.00	-0.00
		Size						
30	30	100.0	57.7	6.8	12.7	15.6	12.8	11.7
50	30	100.0	44.2	6.4	10.7	12.6	10.7	10.4
100	30	100.0	26.5	5.7	10.5	12.0	10.5	9.8
200	30	100.0	14.1	6.6	10.1	11.5	10.1	9.7
30	50	100.0	76.6	7.2	10.1	12.0	9.2	7.3
50	50	100.0	61.6	5.8	8.7	9.8	8.3	7.0
100	50	100.0	36.1	5.5	7.9	8.1	7.9	7.6
200	50	100.0	18.5	5.4	7.3	8.2	7.3	7.2
30	100	100.0	96.2	7.0	8.5	9.5	10.1	7.1
50	100	100.0	88.6	5.6	7.6	7.9	11.0	5.6
100	100	100.0	61.9	5.7	7.0	7.8	7.3	5.7
200	100	100.0	31.8	5.3	6.1	6.5	6.1	6.4
30	200	100.0	99.9	8.9	8.3	8.8	13.1	8.9
50	200	100.0	99.1	5.8	7.4	7.8	16.4	5.8
100	200	100.0	87.9	5.6	6.5	6.9	8.5	5.6
200	200	100.0	53.7	5.2	5.6	5.7	6.3	5.2
93	100.0	100.0	27.0	6.5	10.5	12.1	10.5	10.4
184	100.0	100.0	20.5	6.0	7.8	8.5	7.8	7.9
464	100.0	100.0	16.1	5.4	5.6	5.9	5.6	5.7

Notes: “LS” and “C³ES” refers to the least squares and C³E estimator based on IC selection of the combinations. See Table 2.E1 for an explanation of the rest.

Table 2.E3: The combinations are uncorrelated with the loadings.

		Bias $\times 100$						
N	T	LS	PC	CCE	C ³ E	C ³ E 1	C ³ ES	C ³ ES 1
30	30	25.64	10.51	-0.97	-0.58	-0.14	-0.49	-0.29
50	30	29.16	6.25	0.39	0.00	0.29	0.00	0.32
100	30	30.07	3.34	0.17	0.12	0.17	0.12	0.18
200	30	31.31	1.75	0.04	0.05	0.03	0.05	0.04
30	50	25.74	10.05	-0.96	-0.53	-0.16	-0.35	-0.75
50	50	29.48	6.02	0.36	0.00	0.32	0.11	0.35
100	50	30.12	3.19	0.15	0.10	0.15	0.10	0.16
200	50	31.56	1.66	0.02	0.05	0.03	0.05	0.02
30	100	25.89	9.84	-0.94	-0.50	-0.12	-0.16	-0.90
50	100	29.61	5.93	0.39	0.05	0.35	0.54	0.39
100	100	30.28	3.13	0.16	0.10	0.15	0.14	0.16
200	100	31.71	1.63	0.03	0.06	0.04	0.06	0.04
30	200	25.94	9.72	-0.92	-0.48	-0.11	-0.09	-0.90
50	200	29.62	5.84	0.36	0.01	0.32	0.54	0.36
100	200	30.36	3.08	0.15	0.10	0.15	0.19	0.15
200	200	31.69	1.61	0.03	0.05	0.04	0.07	0.03
93	30	30.09	3.61	0.13	0.05	0.13	0.05	0.14
184	50	31.52	1.81	0.04	0.07	0.06	0.07	0.05
464	100	30.75	0.70	0.00	0.01	0.00	0.01	0.00
		Size						
30	30	99.9	79.1	6.7	12.6	15.5	12.4	11.8
50	30	100.0	61.1	6.3	10.4	12.2	10.4	10.8
100	30	100.0	43.2	6.1	9.7	11.8	9.7	9.7
200	30	100.0	27.3	6.4	9.8	11.5	9.8	9.5
30	50	100.0	92.6	7.9	10.3	11.5	8.7	8.1
50	50	100.0	79.7	6.5	8.5	9.7	8.0	7.0
100	50	100.0	58.5	5.6	7.8	8.7	7.8	7.6
200	50	100.0	38.6	5.5	7.5	8.1	7.5	7.6
30	100	100.0	99.8	8.5	8.9	9.1	8.1	8.4
50	100	100.0	97.4	6.4	7.1	8.5	12.0	6.4
100	100	100.0	85.3	6.0	7.0	8.0	6.9	6.0
200	100	100.0	62.6	5.3	6.3	6.9	6.3	6.3
30	200	100.0	100.0	11.0	8.8	7.8	9.3	11.0
50	200	100.0	100.0	7.0	6.9	8.6	17.3	7.0
100	200	100.0	98.4	5.9	6.3	7.0	7.4	5.9
200	200	100.0	88.2	5.0	5.9	6.0	5.7	5.0
93	30	100.0	45.7	6.1	10.0	12.3	10.0	10.3
184	50	100.0	41.0	5.9	7.9	8.3	7.9	7.7
464	100	100.0	32.3	4.8	5.9	5.9	5.9	5.5

Notes: “LS” and “C³ES” refers to the least squares and C³E estimator based on IC selection of the combinations. See Table 2.E1 for an explanation of the rest.

2 CCE ESTIMATION WITH MORE FACTORS THAN OBSERVABLES

Table 2.E4: Condition (2.5) is not satisfied but loadings are i.i.d..

		Bias $\times 100$						
N	T	LS	PC	CCE	C ³ E	C ³ E 1	C ³ ES	C ³ ES 1
30	30	15.02	5.99	-3.01	0.24	0.23	0.25	0.11
50	30	15.05	3.53	-2.81	0.10	0.06	0.10	0.08
100	30	20.47	1.84	1.36	0.10	0.10	0.10	0.11
200	30	15.47	0.87	-1.10	0.01	0.01	0.01	0.01
30	50	15.01	5.86	-3.02	0.18	0.16	0.15	-1.74
50	50	15.19	3.40	-2.88	0.14	0.11	0.10	-0.91
100	50	20.54	1.76	1.31	0.07	0.07	0.07	0.08
200	50	15.49	0.83	-1.15	0.01	0.01	0.01	0.00
30	100	15.03	5.77	-3.08	0.23	0.22	0.20	-2.22
50	100	15.27	3.38	-2.84	0.16	0.13	0.16	-2.62
100	100	20.62	1.74	1.31	0.08	0.08	0.11	1.10
200	100	15.59	0.84	-1.13	0.02	0.02	0.02	-0.00
30	200	15.07	5.74	-3.06	0.24	0.23	0.20	-2.23
50	200	15.26	3.34	-2.87	0.14	0.12	0.14	-2.65
100	200	20.67	1.72	1.34	0.07	0.07	0.12	1.22
200	200	15.53	0.82	-1.13	0.02	0.02	0.01	-1.09
93	30	20.34	2.01	0.98	0.13	0.13	0.13	0.13
184	50	14.60	0.90	-1.34	0.01	0.01	0.01	0.00
464	100	14.05	0.35	-1.94	-0.00	-0.00	-0.00	-0.00
		Size						
30	30	96.8	40.1	14.0	12.4	15.2	12.3	12.2
50	30	99.1	27.1	18.9	10.4	12.9	10.4	10.9
100	30	100.0	18.0	12.2	10.0	11.8	10.0	9.5
200	30	100.0	11.6	15.5	10.3	11.8	10.3	9.8
30	50	99.8	56.9	19.5	10.4	11.7	8.9	14.4
50	50	100.0	38.1	30.3	8.4	9.3	7.8	15.1
100	50	100.0	23.8	15.2	7.7	8.1	7.7	7.4
200	50	100.0	13.2	22.3	7.1	7.9	7.1	7.3
30	100	100.0	83.8	38.9	8.2	9.0	6.7	27.1
50	100	100.0	63.3	52.1	7.6	7.7	6.5	46.7
100	100	100.0	39.7	25.5	7.1	7.4	6.2	21.0
200	100	100.0	22.3	37.4	6.2	6.6	6.1	6.9
30	200	100.0	98.8	68.1	7.8	8.5	7.0	45.2
50	200	100.0	89.1	82.4	7.5	7.9	6.7	75.0
100	200	100.0	66.6	47.7	6.3	6.4	6.0	41.3
200	200	100.0	38.0	63.4	5.2	5.8	5.0	60.1
93	30	100.0	19.6	9.3	10.6	12.1	10.6	10.7
184	50	100.0	14.8	26.8	7.9	8.6	7.9	8.0
464	100	100.0	12.0	98.4	5.7	6.0	5.7	5.6

Notes: “LS” and “C³ES” refers to the least squares and C³E estimator based on IC selection of the combinations. See Table 2.E1 for an explanation of the rest.

Table 2.E5: $m + 1 < r$ but loadings are i.i.d. and independent of the combinations.

		Bias $\times 100$						
N	T	LS	PC	CCE	C ³ E	C ³ E 1	C ³ ES	C ³ ES 1
30	30	42.21	14.16	1.87	1.72	-0.12	1.73	0.11
50	30	44.69	8.81	0.59	0.77	-0.15	0.77	-0.10
100	30	42.52	4.67	0.50	0.36	0.17	0.36	0.29
200	30	42.12	2.35	0.64	0.34	0.28	0.34	0.36
30	50	42.35	13.42	1.89	1.67	-0.15	1.77	0.28
50	50	44.96	8.38	0.63	0.81	-0.13	0.82	-0.03
100	50	42.81	4.41	0.48	0.38	0.19	0.38	0.30
200	50	42.21	2.25	0.66	0.37	0.31	0.37	0.39
30	100	42.55	13.08	1.90	1.75	-0.07	2.02	0.66
50	100	45.19	8.08	0.62	0.80	-0.14	0.82	0.11
100	100	42.95	4.25	0.48	0.39	0.20	0.33	0.34
200	100	42.62	2.16	0.64	0.36	0.29	0.36	0.37
30	200	42.64	12.87	1.90	1.73	-0.10	2.11	0.82
50	200	45.27	7.92	0.63	0.78	-0.16	0.81	0.24
100	200	42.96	4.15	0.48	0.38	0.19	0.39	0.41
200	200	42.70	2.12	0.67	0.36	0.30	0.36	0.42
93	30	42.51	5.04	0.61	0.64	0.40	0.64	0.47
184	50	43.77	2.45	0.65	0.42	0.34	0.42	0.43
464	100	41.44	0.91	0.01	0.05	0.02	0.05	0.02
		Size						
30	30	100.0	94.0	7.1	15.3	14.6	15.3	12.0
50	30	100.0	84.2	5.3	11.9	12.9	11.9	10.3
100	30	100.0	66.2	6.3	10.8	12.4	10.8	10.4
200	30	100.0	43.3	7.8	11.0	12.5	11.0	10.8
30	50	100.0	99.4	8.0	14.2	11.3	14.5	8.4
50	50	100.0	96.3	5.6	10.0	9.1	9.9	7.7
100	50	100.0	84.1	6.4	8.1	8.7	8.1	7.9
200	50	100.0	58.7	9.1	9.0	9.6	9.0	9.4
30	100	100.0	100.0	13.6	18.0	9.4	21.3	9.0
50	100	100.0	99.9	6.4	10.8	8.1	10.4	7.2
100	100	100.0	97.9	7.1	8.6	7.7	8.6	8.0
200	100	100.0	83.4	12.8	9.8	9.0	9.8	9.5
30	200	100.0	100.0	22.9	27.0	8.6	35.2	12.9
50	200	100.0	100.0	7.9	12.4	7.6	12.8	7.5
100	200	100.0	100.0	9.4	9.5	7.2	9.7	10.8
200	200	100.0	98.7	20.7	11.5	9.2	11.5	12.9
93	30	100.0	69.0	6.4	11.0	12.4	11.0	10.8
184	50	100.0	62.5	8.9	9.1	9.4	9.1	9.3
464	100	100.0	49.1	5.3	6.3	6.6	6.3	5.9

Notes: "LS" and "C³ES" refers to the least squares and C³E estimator based on IC selection of the combinations. See Table 2.E1 for an explanation of the rest.

2 CCE ESTIMATION WITH MORE FACTORS THAN OBSERVABLES

Table 2.E6: Condition (2.5) is not satisfied and loadings are non-i.i.d..

		Bias $\times 100$						
N	T	LS	PC	CCE	C ³ E	C ³ E 1	C ³ ES	C ³ ES 1
30	30	-46.52	7.29	-35.30	0.12	0.09	0.11	-3.20
50	30	-29.76	4.34	-28.93	0.10	0.05	0.10	-1.20
100	30	-27.59	2.25	-5.20	0.06	0.05	0.06	-0.49
200	30	-31.54	1.22	-14.18	0.02	0.01	0.02	-0.42
30	50	-47.00	7.22	-35.36	0.07	0.07	0.10	-9.99
50	50	-29.93	4.27	-29.02	0.13	0.13	0.17	-3.31
100	50	-27.32	2.20	-5.29	0.03	0.03	0.03	-0.56
200	50	-32.06	1.20	-14.29	0.02	0.02	0.02	-0.44
30	100	-47.08	7.24	-35.49	0.12	0.13	0.12	-17.76
50	100	-29.95	4.30	-29.17	0.16	0.15	0.29	-13.79
100	100	-27.40	2.22	-5.27	0.04	0.04	0.09	-3.00
200	100	-31.93	1.21	-14.25	0.03	0.03	0.03	-0.49
30	200	-47.23	7.20	-35.59	0.12	0.11	0.11	-20.34
50	200	-29.97	4.27	-29.28	0.13	0.14	0.27	-14.59
100	200	-27.45	2.20	-5.28	0.04	0.03	0.12	-3.23
200	200	-32.16	1.20	-14.27	0.03	0.01	0.06	-6.95
93	30	-28.46	2.40	-7.17	0.09	0.10	0.09	-0.62
184	50	-34.19	1.31	-13.40	0.02	0.01	0.02	-0.41
464	100	-45.95	0.50	-65.32	0.00	0.00	0.00	-0.18
		Size						
30	30	99.9	53.0	100.0	12.3	15.3	12.2	27.7
50	30	95.4	37.3	99.9	10.4	11.9	10.4	16.0
100	30	95.3	23.8	68.9	9.9	11.3	9.9	12.7
200	30	97.2	16.7	100.0	10.5	11.3	10.5	12.9
30	50	100.0	73.5	100.0	10.0	11.5	9.2	50.9
50	50	99.2	53.9	100.0	8.6	10.1	8.1	25.2
100	50	98.8	33.5	87.4	7.7	8.2	7.7	12.0
200	50	99.5	22.5	100.0	7.0	7.7	7.0	13.2
30	100	100.0	95.3	100.0	8.2	9.2	7.5	85.5
50	100	100.0	82.7	100.0	7.3	7.8	7.1	68.8
100	100	100.0	58.3	99.0	7.0	7.4	6.4	58.1
200	100	100.0	39.5	100.0	6.4	6.7	6.4	17.7
30	200	100.0	100.0	100.0	7.1	8.1	7.2	99.1
50	200	100.0	97.6	100.0	7.5	7.9	8.3	85.9
100	200	100.0	86.0	100.0	6.1	6.4	6.3	70.3
200	200	100.0	65.1	100.0	5.7	5.9	5.6	75.7
93	30	96.6	24.6	84.3	10.5	12.2	10.5	13.7
184	50	99.7	24.9	100.0	7.0	8.3	7.0	12.5
464	100	100.0	18.8	100.0	5.6	5.8	5.6	10.2

Notes: “LS” and “C³ES” refers to the least squares and C³E estimator based on IC selection of the combinations. See Table 2.E1 for an explanation of the rest.

Table 2.E7: Loadings are non-i.i.d. and mean zero.

		Bias $\times 100$						
N	T	LS	PC	CCE	C ³ E	C ³ E 1	C ³ ES	C ³ ES 1
30	30	7.36	8.01	-12.07	0.51	0.48	0.53	0.77
50	30	3.96	4.85	11.64	0.18	0.10	0.18	0.45
100	30	21.59	2.58	-3.26	0.14	0.14	0.14	0.31
200	30	19.54	1.21	-16.14	0.02	0.02	0.02	0.07
30	50	7.41	7.89	-12.16	0.46	0.437	0.58	0.86
50	50	4.06	4.78	11.88	0.22	0.173	0.24	1.72
100	50	21.28	2.52	-3.46	0.11	0.104	0.11	0.28
200	50	19.81	1.19	-16.37	0.02	0.022	0.03	0.06
30	100	7.35	7.87	-12.26	0.50	0.486	0.64	0.75
50	100	3.98	4.78	11.86	0.22	0.169	0.37	2.67
100	100	21.31	2.53	-3.42	0.11	0.112	0.18	1.09
200	100	19.74	1.20	-16.48	0.02	0.028	0.02	0.05
30	200	7.29	7.87	-12.33	0.50	0.484	0.65	0.24
50	200	3.95	4.76	11.92	0.20	0.159	0.36	3.42
100	200	21.32	2.51	-3.51	0.11	0.109	0.17	1.02
200	200	19.68	1.19	-16.55	0.02	0.024	0.03	-1.33
93	30	20.13	2.68	-6.14	0.15	0.15	0.15	0.32
184	50	21.60	1.35	-13.82	0.01	0.01	0.01	0.06
464	100	15.23	0.50	-4.65	-0.00	-0.00	-0.00	0.00
		Size						
30	30	29.7	60.1	72.6	12.5	15.3	12.5	12.6
50	30	33.0	44.0	70.2	10.8	12.6	10.8	11.6
100	30	98.3	29.4	28.4	10.3	11.6	10.3	10.5
200	30	95.3	17.0	100.0	10.1	11.2	10.1	9.6
30	50	39.2	80.1	89.1	9.8	11.7	9.1	9.7
50	50	37.9	62.4	84.3	8.9	9.5	8.9	21.1
100	50	99.8	41.7	40.9	7.9	8.4	7.9	7.9
200	50	99.0	22.3	100.0	7.2	7.7	7.2	6.6
30	100	57.9	97.4	99.5	9.2	10.2	9.6	11.3
50	100	44.5	89.1	96.7	7.6	8.1	7.6	35.9
100	100	100.0	69.4	62.1	6.9	7.7	7.0	26.8
200	100	100.0	39.5	100.0	6.4	6.7	6.4	6.3
30	200	81.0	100.0	100.0	9.1	9.5	10.2	17.4
50	200	57.6	99.1	99.7	7.6	7.9	8.8	45.1
100	200	100.0	93.3	87.5	6.3	6.5	6.8	37.8
200	200	100.0	65.5	100.0	5.0	5.4	5.0	21.1
93	30	96.8	29.2	56.7	10.1	11.9	10.1	10.3
184	50	99.0	25.6	100.0	7.5	8.1	7.5	7.5
464	100	99.8	19.5	97.6	5.5	5.8	5.5	5.5

Notes: “LS” and “C³ES” refers to the least squares and C³E estimator based on IC selection of the combinations. See Table 2.E1 for an explanation of the rest.

2.B Appendix

2.B.1 Preliminary Results

We start with some notation. The model for $\mathbf{w}_{it} = (y_{it}, \mathbf{x}'_{it})'$ can be written in matrix notation as

$$\mathbf{W}_i = \mathbf{F}\mathbf{C}_i + \mathbf{U}_i, \quad (2.B.1)$$

where $\mathbf{W}_i = (\mathbf{y}_i, \mathbf{X}_i) = (\mathbf{w}_{i1}, \dots, \mathbf{w}_{iT})'$ is $T \times (m+1)$, $\mathbf{F} = (\mathbf{f}_1, \dots, \mathbf{f}_T)'$ is $T \times r$, $\mathbf{C}_i = (\boldsymbol{\lambda}'_i \boldsymbol{\beta} + \boldsymbol{\lambda}_i, \boldsymbol{\lambda}'_i)$ is $r \times (m+1)$ and $\mathbf{U}_i = (\mathbf{u}_{i1}, \dots, \mathbf{u}_{iT})' = (\boldsymbol{\eta}_i \boldsymbol{\beta} + \boldsymbol{\epsilon}_i, \boldsymbol{\eta}_i)$ is $T \times (m+1)$. Alternatively, the model for \mathbf{w}_{it} can be written as the following N -dimensional system:

$$\mathbf{w}_t = \mathbf{C}\mathbf{f}_t + \mathbf{u}_t, \quad (2.B.2)$$

where $\mathbf{w}_t = (\mathbf{w}'_{1t}, \dots, \mathbf{w}'_{Nt})'$ and $\mathbf{u}_t = (\mathbf{u}'_{1t}, \dots, \mathbf{u}'_{Nt})'$ are $N(m+1) \times 1$, and $\mathbf{C} = (\mathbf{C}_1, \dots, \mathbf{C}_N)'$ is $N(m+1) \times r$. The matrix notation

$$\mathbf{W} = \mathbf{F}\mathbf{C}' + \mathbf{U} \quad (2.B.3)$$

will also be used, where $\mathbf{W} = (\mathbf{W}_1, \dots, \mathbf{W}_N)$ and $\mathbf{U} = (\mathbf{U}_1, \dots, \mathbf{U}_N)$ are $T \times N(m+1)$. In what follows the representations in (2.B.1)–(2.B.3) will be used interchangeably.

Many of the results can be expressed in terms of $(\widehat{\mathbf{F}} - \mathbf{F}\overline{\mathbf{H}}')$. It is therefore going to be convenient to introduce some special notation to simplify such expressions. It is therefore useful to define

$$\mathbf{D} = \widehat{\mathbf{F}} - \mathbf{F}\overline{\mathbf{H}}' = \frac{1}{N} \sum_{i=1}^N \mathbf{U}_i \mathbf{Z}_i,$$

whose dimension is given by $T \times (m+1)k$. It is further convenient to write $\mathbf{D} = (\mathbf{d}_1, \dots, \mathbf{d}_T)'$, where

$$\mathbf{d}_t = \widehat{\mathbf{f}}_t - \overline{\mathbf{H}}\mathbf{f}_t = \frac{1}{N} \sum_{i=1}^N \mathbf{Z}'_i \mathbf{u}_{it}$$

is $(m+1)k \times 1$.

Before we come to the proof of Theorem 2.1 we state some useful lemmas.

Lemma 2.B.1. *Under Assumptions 2.1 and 2.2,*

$$\frac{1}{T} \sum_{t=1}^T \|\mathbf{d}_t\|^2 = O_p(N^{-1}).$$

Proof of Lemma 2.B.1: The proof of Lemma 2.B.1 is a simple consequence of the fact that $\|N^{-1/2} \sum_{i=1}^N \mathbf{Z}'_i \mathbf{u}_{it}\| = O_p(1)$, as seen by writing

$$\frac{N}{T} \sum_{t=1}^T \|\mathbf{d}_t\|^2 \leq \frac{1}{T} \sum_{t=1}^T \left\| \frac{1}{\sqrt{N}} \sum_{i=1}^N \mathbf{Z}'_i \mathbf{u}_{it} \right\|^2 = O_p(1).$$

■

Lemma 2.B.2. *Under the conditions of Lemma 2.B.1,*

$$\|\sqrt{NT}^{-1/2} \mathbf{F}' \mathbf{D}\| = O_p(1).$$

Proof of Lemma 2.B.2: The proof is completed by noting that,

$$\sqrt{NT}^{-1/2} \mathbf{F}' \mathbf{D} = \frac{\sqrt{N}}{\sqrt{T}} \sum_{t=1}^T \mathbf{f}_t \mathbf{d}'_t = \frac{1}{\sqrt{T}} \sum_{t=1}^T \mathbf{f}_t \frac{1}{\sqrt{N}} \sum_{i=1}^N \mathbf{u}'_{it} \mathbf{Z}_i = O_p(1).$$

■

Lemma 2.B.3. *Under the conditions of Lemma 2.B.1,*

$$NT^{-1} \mathbf{D}' \mathbf{D} = \Theta + O_p(T^{-1/2})$$

where $\Theta = \lim_{N \rightarrow \infty} N^{-1} \mathbf{Z}' \Sigma_u \mathbf{Z}$.

Proof of Lemma 2.B.3: A direct calculation reveals that

$$\begin{aligned} NT^{-1} \mathbf{D}' \mathbf{D} &= \frac{N}{T} \sum_{t=1}^T \mathbf{d}_t \mathbf{d}'_t = \frac{1}{NT} \sum_{t=1}^T \sum_{i=1}^N \sum_{j=1}^N \mathbf{Z}'_i \mathbf{u}_{it} \mathbf{u}'_{jt} \mathbf{Z}_j \\ &= \frac{1}{NT} \sum_{t=1}^T \sum_{i=1}^N \mathbf{Z}'_i \mathbf{u}_{it} \mathbf{u}'_{it} \mathbf{Z}_i + \frac{1}{\sqrt{T}} \frac{1}{N\sqrt{T}} \sum_{t=1}^T \sum_{i=1}^N \sum_{j \neq i}^N \mathbf{Z}'_i \mathbf{u}_{it} \mathbf{u}'_{jt} \mathbf{Z}_j \\ &= \frac{1}{NT} \sum_{t=1}^T \sum_{i=1}^N \mathbf{Z}'_i \mathbf{u}_{it} \mathbf{u}'_{it} \mathbf{Z}_i + O_p(T^{-1/2}) \\ &= \Theta + \frac{1}{NT} \sum_{t=1}^T \sum_{i=1}^N (\mathbf{Z}'_i \mathbf{u}_{it} \mathbf{u}'_{it} \mathbf{Z}_i - \Theta) + O_p(T^{-1/2}) \end{aligned}$$

$$= \Theta + O_p(T^{-1/2}),$$

as was to be shown. ■

Lemma 2.B.4. *Under the conditions of Lemma 2.B.1,*

$$\frac{1}{T} \sum_{i=1}^N \boldsymbol{\eta}'_i \mathbf{D} = \frac{1}{N} \sum_{i=1}^N \boldsymbol{\Sigma}_{\eta,i}(\boldsymbol{\beta}, \mathbf{I}_m) \mathbf{Z}_i + O_p(T^{-1/2}).$$

Proof of Lemma 2.B.4: By using the fact that

$$E(\boldsymbol{\eta}_{it} \mathbf{u}'_{it}) = E[\boldsymbol{\eta}_{it}(\epsilon_{it} + \boldsymbol{\eta}'_{it} \boldsymbol{\beta}, \boldsymbol{\eta}'_{it})] = (\boldsymbol{\Sigma}_{\eta,i} \boldsymbol{\beta}, \boldsymbol{\Sigma}_{\eta,i}) = \boldsymbol{\Sigma}_{\eta,i}(\boldsymbol{\beta}, \mathbf{I}_m),$$

we obtain

$$\begin{aligned} \frac{1}{T} \sum_{i=1}^N \boldsymbol{\eta}'_i \mathbf{D} &= \frac{1}{T} \sum_{i=1}^N \sum_{t=1}^T \boldsymbol{\eta}_{it} \mathbf{d}'_t = \frac{1}{T} \sum_{i=1}^N \sum_{t=1}^T \boldsymbol{\eta}_{it} \frac{1}{N} \sum_{j=1}^N \mathbf{u}'_{jt} \mathbf{Z}_j \\ &= \frac{1}{NT} \sum_{i=1}^N \sum_{t=1}^T \sum_{j=1}^N \boldsymbol{\eta}_{it} \mathbf{u}'_{jt} \mathbf{Z}_j \\ &= \frac{1}{NT} \sum_{i=1}^N \sum_{t=1}^T \boldsymbol{\eta}_{it} \mathbf{u}'_{it} \mathbf{Z}_i + \frac{1}{NT} \sum_{i=1}^N \sum_{t=1}^T \sum_{j \neq i}^N \boldsymbol{\eta}_{it} \mathbf{u}'_{jt} \mathbf{Z}_j \\ &= \frac{1}{N} \sum_{i=1}^N \boldsymbol{\Sigma}_{\eta,i}(\boldsymbol{\beta}, \mathbf{I}_m) \mathbf{Z}_i + \frac{1}{\sqrt{NT}} \frac{1}{\sqrt{NT}} \sum_{i=1}^N \sum_{t=1}^T [\boldsymbol{\eta}_{it} \mathbf{u}'_{it} \mathbf{Z}_i - \boldsymbol{\Sigma}_{\eta,i}(\boldsymbol{\beta}, \mathbf{I}_m) \mathbf{Z}_i] \\ &\quad + \frac{1}{\sqrt{T}} \frac{1}{N\sqrt{T}} \sum_{i=1}^N \sum_{t=1}^T \sum_{j \neq i}^N \boldsymbol{\eta}_{it} \mathbf{u}'_{jt} \mathbf{Z}_j \\ &= \frac{1}{N} \sum_{i=1}^N \boldsymbol{\Sigma}_{\eta,i}(\boldsymbol{\beta}, \mathbf{I}_m) \mathbf{Z}_i + O_p(T^{-1/2}), \end{aligned}$$

as required for the proof. ■

Lemma 2.B.5. *Under the conditions of Lemma 2.B.1,*

$$NT^{-1} \boldsymbol{\epsilon}'_i \mathbf{D} = \sigma_{\epsilon,i}^2(1, 0_m) \mathbf{Z}_i + O_p(T^{-1/2}).$$

Proof of Lemma 2.B.5: The proof is implied by

$$E(\boldsymbol{\epsilon}_{it} \mathbf{u}'_{it}) = E[\boldsymbol{\epsilon}_{it}(\epsilon_{it} + \boldsymbol{\eta}'_{it} \boldsymbol{\beta}, \boldsymbol{\eta}'_{it})] = (\sigma_{\epsilon,i}^2, 0)$$

and

$$\begin{aligned}
& E \left[\left(\frac{1}{\sqrt{T}} \sum_{t=1}^T \sum_{j \neq i}^N \epsilon_{it} \mathbf{u}'_{jt} \mathbf{Z}_j \right)' \left(\frac{1}{\sqrt{T}} \sum_{t=1}^T \sum_{j \neq i}^N \epsilon_{it} \mathbf{u}'_{jt} \mathbf{Z}_j \right) \right] \\
&= \frac{1}{T} \sum_{t=1}^T \sum_{s=1}^T \sum_{k \neq i}^N \sum_{j \neq i}^N E(\mathbf{Z}'_j \mathbf{u}_{jt} \epsilon_{it} \epsilon_{is} \mathbf{u}'_{ks} \mathbf{Z}_k) \\
&= \frac{1}{T} \sum_{t=1}^T \sum_{k \neq i}^N \sum_{j \neq i}^N E(\mathbf{Z}'_j \mathbf{u}_{jt} \epsilon_{it} \epsilon_{it} \mathbf{u}'_{kt} \mathbf{Z}_k) + \frac{1}{T} \sum_{t \neq s}^T \sum_{s=1}^T \sum_{k \neq i}^N \sum_{j \neq i}^N E(\mathbf{Z}'_j \mathbf{u}_{jt} \epsilon_{it} \epsilon_{is} \mathbf{u}'_{ks} \mathbf{Z}_k) \\
&= \frac{1}{T} \sum_{t=1}^T \sigma_{\epsilon,i}^2 E(\mathbf{Z}'_j \mathbf{u}_{jt} \mathbf{u}'_{jt} \mathbf{Z}_j) + \frac{1}{T} \sum_{t=1}^T \sum_{k \neq i}^N \sum_{j \neq i, j \neq k}^N \sigma_{\epsilon,i}^2 E(\mathbf{Z}'_j \mathbf{u}_{jt} \mathbf{u}'_{kt} \mathbf{Z}_k) = \sigma_{\epsilon,i}^2 \boldsymbol{\Theta}_j,
\end{aligned}$$

where $\boldsymbol{\Theta}_j = E(\mathbf{Z}'_j \boldsymbol{\Sigma}_{u,j} \mathbf{Z}_j)$.

Since, $\|\boldsymbol{\Theta}_j\| < \infty$, we have $\|T^{-1/2} \sum_{t=1}^T \sum_{j \neq i}^N \epsilon_{it} \mathbf{u}'_{jt} \mathbf{Z}_j\| = O_p(1)$, as is clear from

$$\begin{aligned}
NT^{-1} \boldsymbol{\epsilon}'_i \mathbf{D} &= \frac{1}{T} \sum_{t=1}^T \epsilon_{it} \frac{1}{N} \sum_{j=1}^N \mathbf{u}'_{jt} \mathbf{Z}_j \\
&= \frac{1}{T} \sum_{t=1}^T \sum_{j=1}^N \epsilon_{it} \mathbf{u}'_{jt} \mathbf{Z}_j = \frac{1}{T} \sum_{t=1}^T \epsilon_{it} \mathbf{u}'_{it} \mathbf{Z}_i + \frac{1}{T} \sum_{t=1}^T \sum_{j \neq i}^N \epsilon_{it} \mathbf{u}'_{jt} \mathbf{Z}_j \\
&= \sigma_{\epsilon,i}^2(1,0) \mathbf{Z}_i + \frac{1}{\sqrt{T}} \frac{1}{\sqrt{T}} \sum_{t=1}^T [\epsilon_{it} \mathbf{u}'_{it} \mathbf{Z}_i - \sigma_{\epsilon,i}^2(1,0) \mathbf{Z}_i] \\
&\quad + \frac{1}{\sqrt{T}} \frac{1}{\sqrt{T}} \sum_{t=1}^T \sum_{j \neq i}^N \epsilon_{it} \mathbf{u}'_{jt} \mathbf{Z}_j \\
&= \sigma_{\epsilon,i}^2(1,0_m) \mathbf{Z}_i + O_p(T^{-1/2}).
\end{aligned}$$

■

2.B.2 Proof of Theorem 2.1

Since $\bar{\mathbf{H}}$ is $k(m+1) \times r$ with $rk(\bar{\mathbf{H}}) = r \leq k(m+1)$, the $r \times r$ matrix $\bar{\mathbf{H}}' \bar{\mathbf{H}}$ is nonsingular. We may therefore define $\bar{\mathbf{H}}^- = \bar{\mathbf{H}}(\bar{\mathbf{H}}' \bar{\mathbf{H}})^{-1}$, such that $\bar{\mathbf{H}}' \bar{\mathbf{H}}^- = I_r$. The equation for \mathbf{y}_i can now be written as

$$\mathbf{y}_i = \mathbf{X}_i \boldsymbol{\beta} + \widehat{\mathbf{F}} \bar{\mathbf{H}}^- \boldsymbol{\lambda}_i - \mathbf{D} \bar{\mathbf{H}}^- \boldsymbol{\lambda}_i + \boldsymbol{\epsilon}_i, \quad (2.B.4)$$

where $\mathbf{D} = \widehat{\mathbf{F}} - \mathbf{F}\overline{\mathbf{H}}'$ is as before. The estimator of $\boldsymbol{\beta}$ is given by

$$\widehat{\boldsymbol{\beta}} = \left(\sum_{i=1}^N \mathbf{X}'_i \mathbf{M}_{\widehat{\mathbf{F}}} \mathbf{X}_i \right)^{-1} \sum_{i=1}^N \mathbf{X}'_i \mathbf{M}_{\widehat{\mathbf{F}}} \mathbf{y}_i,$$

suggesting that

$$\sqrt{NT}(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}) = \left(\frac{1}{NT} \sum_{i=1}^N \mathbf{X}'_i \mathbf{M}_{\widehat{\mathbf{F}}} \mathbf{X}_i \right)^{-1} \frac{1}{\sqrt{NT}} \sum_{i=1}^N \mathbf{X}'_i \mathbf{M}_{\widehat{\mathbf{F}}} (\boldsymbol{\epsilon}_i - \mathbf{D}\overline{\mathbf{H}}^- \boldsymbol{\lambda}_i).$$

We begin by considering the second term in the numerator. Clearly,

$$\mathbf{M}_{\widehat{\mathbf{F}}} \mathbf{D}\overline{\mathbf{H}}^- = \mathbf{M}_{\widehat{\mathbf{F}}} (\widehat{\mathbf{F}} - \mathbf{F}\overline{\mathbf{H}}') \overline{\mathbf{H}}^- = -\mathbf{M}_{\widehat{\mathbf{F}}} \mathbf{F},$$

and therefore

$$\begin{aligned} -\frac{1}{\sqrt{NT}} \sum_{i=1}^N \mathbf{X}'_i \mathbf{M}_{\widehat{\mathbf{F}}} \mathbf{D}\overline{\mathbf{H}}^- \boldsymbol{\lambda}_i &= \frac{1}{\sqrt{NT}} \sum_{i=1}^N \mathbf{X}'_i \mathbf{M}_{\widehat{\mathbf{F}}} \mathbf{F} \boldsymbol{\lambda}_i \\ &= \frac{1}{\sqrt{NT}} \sum_{i=1}^N \boldsymbol{\Lambda}_i \mathbf{F}' \mathbf{M}_{\widehat{\mathbf{F}}} \mathbf{F} \boldsymbol{\lambda}_i + \frac{1}{\sqrt{NT}} \sum_{i=1}^N \boldsymbol{\eta}'_i \mathbf{M}_{\widehat{\mathbf{F}}} \mathbf{F} \boldsymbol{\lambda}_i \\ &= \mathbf{K}_1 + \mathbf{K}_2. \end{aligned}$$

Consider \mathbf{K}_1 . Using A^+ to denote the Moore-Penrose inverse A , we define $\mathbf{M}_{\mathbf{F}\overline{\mathbf{H}}'} = \mathbf{I}_T - \mathbf{F}\overline{\mathbf{H}}' (\overline{\mathbf{H}}\mathbf{F}'\mathbf{F}\overline{\mathbf{H}}')^+ \overline{\mathbf{H}}\mathbf{F}'$. It is necessary to use the Moore-Penrose inverse in the definition, since $(\overline{\mathbf{H}}\mathbf{F}'\mathbf{F}\overline{\mathbf{H}}')$ is $k(m+1) \times k(m+1)$ whereas its rank is r . From

$$\overline{\mathbf{H}}\mathbf{F}' \mathbf{M}_{\widehat{\mathbf{F}}} \mathbf{F}\overline{\mathbf{H}}' = \mathbf{D}' \mathbf{M}_{\widehat{\mathbf{F}}} \mathbf{D} = \mathbf{D}' \mathbf{M}_{\mathbf{F}\overline{\mathbf{H}}'} \mathbf{D} - \mathbf{D}' (\mathbf{M}_{\mathbf{F}\overline{\mathbf{H}}'} - \mathbf{M}_{\widehat{\mathbf{F}}}) \mathbf{D},$$

we obtain

$$\begin{aligned} \mathbf{K}_1 &= \frac{1}{\sqrt{NT}} \sum_{i=1}^N \boldsymbol{\Lambda}_i \mathbf{F}' \mathbf{M}_{\widehat{\mathbf{F}}} \mathbf{F} \boldsymbol{\lambda}_i \\ &= \frac{1}{\sqrt{NT}} \sum_{i=1}^N \boldsymbol{\Lambda}_i (\overline{\mathbf{H}}^-)' \overline{\mathbf{H}}\mathbf{F}' \mathbf{M}_{\widehat{\mathbf{F}}} \mathbf{F}\overline{\mathbf{H}}' \overline{\mathbf{H}}^- \boldsymbol{\lambda}_i \\ &= \frac{1}{\sqrt{NT}} \sum_{i=1}^N \boldsymbol{\Lambda}_i (\overline{\mathbf{H}}^-)' \mathbf{D}' \mathbf{M}_{\mathbf{F}\overline{\mathbf{H}}'} \mathbf{D}\overline{\mathbf{H}}^- \boldsymbol{\lambda}_i \\ &\quad + \frac{1}{\sqrt{NT}} \sum_{i=1}^N \boldsymbol{\Lambda}_i (\overline{\mathbf{H}}^-)' \mathbf{D}' (\mathbf{M}_{\mathbf{F}\overline{\mathbf{H}}'} - \mathbf{M}_{\widehat{\mathbf{F}}}) \mathbf{D}\overline{\mathbf{H}}^- \boldsymbol{\lambda}_i \\ &= \mathbf{K}_{11} + \mathbf{K}_{12}. \end{aligned}$$

Consider \mathbf{K}_{12} , from the definitions of $\mathbf{M}_{F\bar{\mathbf{H}}'}$ and $\mathbf{M}_{\hat{\mathbf{F}}}$ he have

$$\begin{aligned}\mathbf{M}_{F\bar{\mathbf{H}}'} - \mathbf{M}_{\hat{\mathbf{F}}} &= \mathbf{D}(\hat{\mathbf{F}}'\hat{\mathbf{F}})^{-1}\mathbf{D}' + \mathbf{D}(\hat{\mathbf{F}}'\hat{\mathbf{F}})^{-1}\bar{\mathbf{H}}\mathbf{F}' \\ &+ \mathbf{F}\bar{\mathbf{H}}'(\hat{\mathbf{F}}'\hat{\mathbf{F}})^{-1}\mathbf{D}' + \mathbf{F}\bar{\mathbf{H}}'[(\hat{\mathbf{F}}'\hat{\mathbf{F}})^{-1} - (\bar{\mathbf{H}}\mathbf{F}'\bar{\mathbf{H}}')^+]\bar{\mathbf{H}}\mathbf{F}',\end{aligned}$$

suggesting that

$$\begin{aligned}\mathbf{D}'(\mathbf{M}_{F\bar{\mathbf{H}}'} - \mathbf{M}_{\hat{\mathbf{F}}})\mathbf{D} &= \mathbf{D}'\mathbf{D}(\hat{\mathbf{F}}'\hat{\mathbf{F}})^{-1}\mathbf{D}'\mathbf{D} + \mathbf{D}'\mathbf{D}(\hat{\mathbf{F}}'\hat{\mathbf{F}})^{-1}\bar{\mathbf{H}}\mathbf{F}'\mathbf{D} \\ &+ \mathbf{D}'\mathbf{F}\bar{\mathbf{H}}'(\hat{\mathbf{F}}'\hat{\mathbf{F}})^{-1}\mathbf{D}'\mathbf{D} \\ &+ \mathbf{D}'\mathbf{F}\bar{\mathbf{H}}'[(\hat{\mathbf{F}}'\hat{\mathbf{F}})^{-1} - (\bar{\mathbf{H}}\mathbf{F}'\bar{\mathbf{H}}')^+]\bar{\mathbf{H}}\mathbf{F}'\mathbf{D}.\end{aligned}\quad (2.B.5)$$

Consider the fourth term. Using the properties of generalized inverses,

$$(\bar{\mathbf{H}}\mathbf{F}'\bar{\mathbf{H}}')(\bar{\mathbf{H}}\mathbf{F}'\bar{\mathbf{H}}')^+(\bar{\mathbf{H}}\mathbf{F}'\bar{\mathbf{H}}') = (\bar{\mathbf{H}}\mathbf{F}'\bar{\mathbf{H}}').$$

Multiplying the equation above by $(\mathbf{F}'\mathbf{F})^{-1}(\bar{\mathbf{H}}'\bar{\mathbf{H}})^{-1}\bar{\mathbf{H}}'$ from the left and $\bar{\mathbf{H}}(\bar{\mathbf{H}}'\bar{\mathbf{H}})^{-1}(\mathbf{F}'\mathbf{F})^{-1}$ from the right,

$$\bar{\mathbf{H}}'(\bar{\mathbf{H}}\mathbf{F}'\bar{\mathbf{H}}')^+\bar{\mathbf{H}} = (\mathbf{F}'\mathbf{F})^{-1}.\quad (2.B.6)$$

Insertion into $\mathbf{F}\bar{\mathbf{H}}'[(\hat{\mathbf{F}}'\hat{\mathbf{F}})^{-1} - (\bar{\mathbf{H}}\mathbf{F}'\bar{\mathbf{H}}')^+]\bar{\mathbf{H}}\mathbf{F}'$ yields

$$\begin{aligned}\mathbf{F}\bar{\mathbf{H}}'[(\hat{\mathbf{F}}'\hat{\mathbf{F}})^{-1} - (\bar{\mathbf{H}}\mathbf{F}'\bar{\mathbf{H}}')^+]\bar{\mathbf{H}}\mathbf{F}' &= \mathbf{F}[\bar{\mathbf{H}}'(\hat{\mathbf{F}}'\hat{\mathbf{F}})^{-1}\bar{\mathbf{H}} - (\mathbf{F}'\mathbf{F})^{-1}]\mathbf{F}' \\ &= \mathbf{F}\bar{\mathbf{H}}'[(\hat{\mathbf{F}}'\hat{\mathbf{F}})^{-1} - \bar{\mathbf{H}}^-(\mathbf{F}'\mathbf{F})^{-1}\bar{\mathbf{H}}'^-]\bar{\mathbf{H}}\mathbf{F}' \\ &= \mathbf{F}\bar{\mathbf{H}}'(\hat{\mathbf{F}}'\hat{\mathbf{F}})^{-1}[\bar{\mathbf{H}}\mathbf{F}'\bar{\mathbf{H}}' - \hat{\mathbf{F}}'\hat{\mathbf{F}}] \\ &\quad \times \bar{\mathbf{H}}^-(\mathbf{F}'\mathbf{F})^{-1}\bar{\mathbf{H}}'^-\bar{\mathbf{H}}\mathbf{F}' \\ &= \mathbf{F}\bar{\mathbf{H}}'(\hat{\mathbf{F}}'\hat{\mathbf{F}})^{-1}[\mathbf{D}'\mathbf{F}\bar{\mathbf{H}}' - \hat{\mathbf{F}}'\mathbf{D}] \\ &\quad \times \bar{\mathbf{H}}^-(\mathbf{F}'\mathbf{F})^{-1}\bar{\mathbf{H}}'^-\bar{\mathbf{H}}\mathbf{F}'.\end{aligned}$$

By Lemmas 2.B.1 and 2.B.2,

$$\begin{aligned}\|\sqrt{N}T^{-1/2}\mathbf{D}'\hat{\mathbf{F}}\| &\leq \sqrt{T}N^{-1/2}\|NT^{-1}\mathbf{D}'\mathbf{D}\| + \|\sqrt{N}T^{-1/2}\mathbf{D}'\mathbf{F}\|\|\bar{\mathbf{H}}\| \\ &= O_p(\sqrt{T}N^{-1/2}) + O_p(1),\end{aligned}$$

giving

$$\begin{aligned}T\|(\hat{\mathbf{F}}'\hat{\mathbf{F}})^{-1} - (\bar{\mathbf{H}}\mathbf{F}'\bar{\mathbf{H}}')^+\| \\ \leq \|(T^{-1}\hat{\mathbf{F}}'\hat{\mathbf{F}})^{-1}\|T^{-1}\|\mathbf{D}'\hat{\mathbf{F}} + \bar{\mathbf{H}}\mathbf{F}'\mathbf{D}\|\|\bar{\mathbf{H}}^-\|\|(T^{-1}\mathbf{F}'\mathbf{F})^{-1}\|\|\bar{\mathbf{H}}'^-\|\end{aligned}$$

$$= O_p(N^{-1}) + O_p((NT)^{-1/2}),$$

and therefore

$$\begin{aligned} & \|T^{-1}\mathbf{D}'(\mathbf{M}_{F\bar{H}'} - \mathbf{M}_{\hat{F}})\mathbf{D}\| \\ & \leq \|T^{-1}\mathbf{D}'\mathbf{D}\|^2 \|(T^{-1}\hat{\mathbf{F}}'\hat{\mathbf{F}})^{-1}\| + 2\|\bar{\mathbf{H}}\| \|T^{-1}\mathbf{D}'\mathbf{D}\| \|(T^{-1}\hat{\mathbf{F}}'\hat{\mathbf{F}})^{-1}\| \|T^{-1}\mathbf{F}'\mathbf{D}\| \\ & \quad + \|\bar{\mathbf{H}}\|^2 \|T^{-1}\mathbf{D}'\mathbf{F}\|^2 T \|(\hat{\mathbf{F}}'\hat{\mathbf{F}})^{-1} - (\bar{\mathbf{H}}\mathbf{F}'\mathbf{F}\bar{\mathbf{H}})'\| \\ & = O_p(N^{-2}) + O_p(N^{-1})O_p((NT)^{-1/2}) \\ & \quad + [O_p(N^{-1}) + O_p((NT)^{-1/2})]O_p((NT)^{-1}) \\ & = O_p(N^{-2}) + O_p(N^{-3/2}T^{-1/2}). \end{aligned}$$

Hence,

$$\begin{aligned} & \left\| \frac{1}{\sqrt{NT}} \sum_{i=1}^N \boldsymbol{\Lambda}_i (\bar{\mathbf{H}}^-)' \mathbf{D}' (\mathbf{M}_{F\bar{H}'} - \mathbf{M}_{\hat{F}}) \mathbf{D} \bar{\mathbf{H}}^- \boldsymbol{\lambda}_i \right\| \\ & \leq \sqrt{NT} \|\bar{\mathbf{H}}^-\|^2 \|T^{-1}\mathbf{D}'(\mathbf{M}_{F\bar{H}'} - \mathbf{M}_{\hat{F}})\mathbf{D}\| \frac{1}{N} \sum_{i=1}^N \|\boldsymbol{\Lambda}_i\| \|\boldsymbol{\lambda}_i\| \\ & = O_p(\sqrt{T}N^{-3/2}) + O_p(N^{-1}). \end{aligned}$$

Consider \mathbf{K}_{12} . Since $\mathbf{D}'\mathbf{M}_{F\bar{H}'}\mathbf{D} = \mathbf{D}'\mathbf{D} - \mathbf{D}'\mathbf{F}\bar{\mathbf{H}}(\bar{\mathbf{H}}\mathbf{F}'\mathbf{F}\bar{\mathbf{H}})'\mathbf{D} + \bar{\mathbf{H}}\mathbf{F}'\mathbf{D}$, and using the result in (2.B.6), we have $\mathbf{D}'\mathbf{M}_{F\bar{H}'}\mathbf{D} = \mathbf{D}'\mathbf{D} - \mathbf{D}'\mathbf{F}(\mathbf{F}'\mathbf{F})^{-1}\mathbf{F}'\mathbf{D}$ with

$$\begin{aligned} & \left\| \frac{1}{\sqrt{NT}} \sum_{i=1}^N \boldsymbol{\Lambda}_i (\bar{\mathbf{H}}^-)' \mathbf{D}' \mathbf{F} (\mathbf{F}'\mathbf{F})^{-1} \mathbf{F}' \mathbf{D} \bar{\mathbf{H}}^- \boldsymbol{\lambda}_i \right\| \\ & \leq \frac{1}{\sqrt{NT}} \|\bar{\mathbf{H}}^-\|^2 \|\bar{\mathbf{H}}\|^2 \|\sqrt{NT}T^{-1/2}\mathbf{D}'\mathbf{F}\|^2 \|(T^{-1}\mathbf{F}'\mathbf{F})^{-1}\| \frac{1}{N} \sum_{i=1}^N \|\boldsymbol{\Lambda}_i\| \|\boldsymbol{\lambda}_i\| \\ & = O_p((NT)^{-1/2}), \end{aligned}$$

we obtain

$$\begin{aligned} \mathbf{K}_1 &= \frac{\sqrt{T}}{\sqrt{N}} \frac{1}{N} \sum_{i=1}^N \boldsymbol{\Lambda}_i (\bar{\mathbf{H}}^-)' NT^{-1} \mathbf{D}' \mathbf{D} \bar{\mathbf{H}}^- \boldsymbol{\lambda}_i + O_p(\sqrt{T}N^{-3/2}) \\ & \quad + O_p(N^{-1}) + O_p((NT)^{-1/2}). \end{aligned}$$

Application of Lemma 2.B.3 now yields

$$\mathbf{K}_1 = \sqrt{T}N^{-1/2}\mathbf{B}_1 + O_p(\sqrt{T}N^{-3/2}) + O_p(N^{-1/2}),$$

where, via $\bar{\mathbf{H}}^- = \bar{\mathbf{H}}(\bar{\mathbf{H}}'\bar{\mathbf{H}})^{-1}$,

$$\mathbf{B}_1 = \frac{1}{N} \sum_{i=1}^N \boldsymbol{\Lambda}_i (\bar{\mathbf{H}}^-)' \boldsymbol{\Theta} \bar{\mathbf{H}}^- \boldsymbol{\lambda}_i = \frac{1}{N} \sum_{i=1}^N \boldsymbol{\Lambda}_i (\bar{\mathbf{H}}'\bar{\mathbf{H}})^{-1} \bar{\mathbf{H}}' \boldsymbol{\Theta} \bar{\mathbf{H}} (\bar{\mathbf{H}}'\bar{\mathbf{H}})^{-1} \boldsymbol{\lambda}_i.$$

Next, consider \mathbf{K}_2 . By using $\mathbf{M}_{F\bar{\mathbf{H}}'} \mathbf{F} \boldsymbol{\lambda}_i = \mathbf{M}_{F\bar{\mathbf{H}}'} \mathbf{F} \bar{\mathbf{H}}' \bar{\mathbf{H}}^- \boldsymbol{\lambda}_i = 0$, and then substitution for $(\mathbf{M}_{F\bar{\mathbf{H}}'} - \mathbf{M}_{\hat{\mathbf{F}}})$,

$$\begin{aligned} \mathbf{K}_2 &= \frac{1}{\sqrt{NT}} \sum_{i=1}^N \boldsymbol{\eta}'_i \mathbf{M}_{\hat{\mathbf{F}}} \mathbf{F} \boldsymbol{\lambda}_i \\ &= -\frac{1}{\sqrt{NT}} \sum_{i=1}^N \boldsymbol{\eta}'_i (\mathbf{M}_{F\bar{\mathbf{H}}'} - \mathbf{M}_{\hat{\mathbf{F}}}) \mathbf{F} \boldsymbol{\lambda}_i \\ &= -\frac{1}{\sqrt{NT}} \sum_{i=1}^N \boldsymbol{\eta}'_i \mathbf{D} (\hat{\mathbf{F}}'\hat{\mathbf{F}})^{-1} \mathbf{D}' \mathbf{F} \boldsymbol{\lambda}_i - \frac{1}{\sqrt{NT}} \sum_{i=1}^N \boldsymbol{\eta}'_i \mathbf{D} (\hat{\mathbf{F}}'\hat{\mathbf{F}})^{-1} \bar{\mathbf{H}} \mathbf{F}' \mathbf{F} \boldsymbol{\lambda}_i \\ &\quad - \frac{1}{\sqrt{NT}} \sum_{i=1}^N \boldsymbol{\eta}'_i \mathbf{F} \bar{\mathbf{H}}' (\hat{\mathbf{F}}'\hat{\mathbf{F}})^{-1} \mathbf{D}' \mathbf{F} \boldsymbol{\lambda}_i - \frac{1}{\sqrt{NT}} \sum_{i=1}^N \boldsymbol{\eta}'_i \mathbf{F} \bar{\mathbf{H}}' [(\hat{\mathbf{F}}'\hat{\mathbf{F}})^{-1} \\ &\quad - (\bar{\mathbf{H}} \mathbf{F}' \mathbf{F} \bar{\mathbf{H}}')^+] \bar{\mathbf{H}} \mathbf{F}' \mathbf{F} \boldsymbol{\lambda}_i \\ &= -\mathbf{K}_{21} - \dots - \mathbf{K}_{24}. \end{aligned}$$

Since $\mathbf{d}'_t (\hat{\mathbf{F}}'\hat{\mathbf{F}})^{-1} \mathbf{d}_s$ and $\mathbf{F}'_t (\hat{\mathbf{F}}'\hat{\mathbf{F}})^{-1} \mathbf{D}' \mathbf{F} \boldsymbol{\lambda}_i$ are just scalars, the orders of \mathbf{K}_{21} and \mathbf{K}_{23} can be inferred as follows:

$$\begin{aligned} \|\mathbf{K}_{21}\| &= \left\| \frac{1}{\sqrt{NT}} \sum_{i=1}^N \boldsymbol{\eta}'_i \mathbf{D} (\hat{\mathbf{F}}'\hat{\mathbf{F}})^{-1} \mathbf{D}' \mathbf{F} \boldsymbol{\lambda}_i \right\| \\ &= \left\| \frac{1}{\sqrt{NT}} \sum_{i=1}^N \sum_{t=1}^T \boldsymbol{\eta}_{it} \mathbf{d}'_t (\hat{\mathbf{F}}'\hat{\mathbf{F}})^{-1} \sum_{s=1}^T \mathbf{d}_s \mathbf{f}'_s \boldsymbol{\lambda}_i \right\| \\ &= \left\| \frac{1}{\sqrt{NT}} \sum_{t=1}^T \sum_{s=1}^T \mathbf{d}'_t (\hat{\mathbf{F}}'\hat{\mathbf{F}})^{-1} \mathbf{d}_s \sum_{i=1}^N \boldsymbol{\eta}_{it} \mathbf{f}'_s \boldsymbol{\lambda}_i \right\| \\ &\leq \sqrt{T} \|(T^{-1} \hat{\mathbf{F}}'\hat{\mathbf{F}})^{-1}\| \frac{1}{T} \sum_{t=1}^T \|\mathbf{d}_t\|^2 \left(\frac{1}{T^2} \sum_{t=1}^T \sum_{s=1}^T \left\| \frac{1}{\sqrt{N}} \sum_{i=1}^N \boldsymbol{\eta}_{it} \mathbf{f}'_s \boldsymbol{\lambda}_i \right\|^2 \right)^{1/2} \\ &= O_p(\sqrt{T} N^{-1}), \end{aligned}$$

and

$$\begin{aligned}
 \|\mathbf{K}_{23}\| &= \left\| \frac{1}{\sqrt{NT}} \sum_{i=1}^N \boldsymbol{\eta}'_i \mathbf{F} \bar{\mathbf{H}}' (\hat{\mathbf{F}}' \hat{\mathbf{F}})^{-1} \mathbf{D}' \mathbf{F} \boldsymbol{\lambda}_i \right\| \\
 &= \left\| \frac{1}{\sqrt{NT}} \sum_{i=1}^N \sum_{t=1}^T \boldsymbol{\eta}_{it} \mathbf{F}'_t \bar{\mathbf{H}}' (\hat{\mathbf{F}}' \hat{\mathbf{F}})^{-1} \mathbf{D}' \mathbf{F} \boldsymbol{\lambda}_i \right\| \\
 &= \left\| \frac{1}{\sqrt{NT}} \sum_{t=1}^T \mathbf{F}'_t \bar{\mathbf{H}}' (\hat{\mathbf{F}}' \hat{\mathbf{F}})^{-1} \mathbf{D}' \mathbf{F} \sum_{i=1}^N \boldsymbol{\lambda}_i \boldsymbol{\eta}_{it} \right\| \\
 &\leq \frac{1}{\sqrt{N}} \|\bar{\mathbf{H}}\| \left(\frac{1}{T} \sum_{t=1}^T \|\mathbf{F}_t\|^2 \right)^{1/2} \|(T^{-1} \hat{\mathbf{F}}' \hat{\mathbf{F}})^{-1}\| \|\sqrt{NT} T^{-1/2} \mathbf{D}' \mathbf{F}\| \\
 &\quad \times \left(\frac{1}{T} \sum_{t=1}^T \left\| \frac{1}{\sqrt{N}} \sum_{i=1}^N \boldsymbol{\lambda}_i \boldsymbol{\eta}_{it} \right\|^2 \right)^{1/2} \\
 &= O_p(N^{-1/2}).
 \end{aligned}$$

Similarly, since $\mathbf{F}'_t \bar{\mathbf{H}}' [(\hat{\mathbf{F}}' \hat{\mathbf{F}})^{-1} - (\bar{\mathbf{H}} \mathbf{F}' \mathbf{F} \bar{\mathbf{H}}')^+] \bar{\mathbf{H}} \mathbf{F}' \boldsymbol{\lambda}_i$ is a scalar,

$$\begin{aligned}
 \|\mathbf{K}_{24}\| &= \left\| \frac{1}{\sqrt{NT}} \sum_{i=1}^N \boldsymbol{\eta}'_i \mathbf{F} \bar{\mathbf{H}}' T [(\hat{\mathbf{F}}' \hat{\mathbf{F}})^{-1} - (\bar{\mathbf{H}} \mathbf{F}' \mathbf{F} \bar{\mathbf{H}}')^+] T^{-1} \bar{\mathbf{H}} \mathbf{F}' \boldsymbol{\lambda}_i \right\| \\
 &= \left\| \frac{1}{\sqrt{NT}} \sum_{t=1}^T \mathbf{F}'_t \bar{\mathbf{H}}' [(\hat{\mathbf{F}}' \hat{\mathbf{F}})^{-1} - (\bar{\mathbf{H}} \mathbf{F}' \mathbf{F} \bar{\mathbf{H}}')^+] \bar{\mathbf{H}} \mathbf{F}' \mathbf{F} \sum_{i=1}^N \boldsymbol{\lambda}_i \boldsymbol{\eta}_{it} \right\| \\
 &\leq \sqrt{T} \|\bar{\mathbf{H}}\|^2 \left(\frac{1}{T} \sum_{t=1}^T \|\mathbf{f}_t\|^2 \right)^{1/2} T \|(\hat{\mathbf{F}}' \hat{\mathbf{F}})^{-1} - (\bar{\mathbf{H}} \mathbf{F}' \mathbf{F} \bar{\mathbf{H}}')^+\| \|T^{-1} \mathbf{F}' \mathbf{F}\| \\
 &\quad \times \left(\frac{1}{T} \sum_{t=1}^T \left\| \frac{1}{\sqrt{N}} \sum_{i=1}^N \boldsymbol{\lambda}_i \boldsymbol{\eta}_{it} \right\|^2 \right)^{1/2} \\
 &= O_p(\sqrt{T} N^{-1}) + O_p(N^{-1/2}).
 \end{aligned}$$

\mathbf{K}_{22} can be expanded as follows, by adding and subtracting $\bar{\mathbf{H}}^- (T^{-1} \mathbf{F}' \mathbf{F})^{-1} \bar{\mathbf{H}}^{-\prime}$:

$$\begin{aligned}
 \mathbf{K}_{22} &= \frac{1}{\sqrt{NT}} \sum_{i=1}^N \boldsymbol{\eta}'_i \mathbf{D} (\hat{\mathbf{F}}' \hat{\mathbf{F}})^{-1} \bar{\mathbf{H}} \mathbf{F}' \boldsymbol{\lambda}_i \\
 &= \frac{1}{\sqrt{NT}} \sum_{i=1}^N \boldsymbol{\eta}'_i \mathbf{D} (T^{-1} \hat{\mathbf{F}}' \hat{\mathbf{F}})^{-1} T^{-1} \bar{\mathbf{H}} \mathbf{F}' \mathbf{F} \bar{\mathbf{H}}^- \boldsymbol{\lambda}_i
 \end{aligned}$$

$$\begin{aligned}
&= \frac{1}{\sqrt{NT}} \sum_{i=1}^N \boldsymbol{\eta}'_i \mathbf{D} \bar{\mathbf{H}}^{-1} \boldsymbol{\lambda}_i \\
&\quad + \sqrt{NT} \frac{1}{N} \sum_{i=1}^N T^{-1} \boldsymbol{\eta}'_i \mathbf{D} [(T^{-1} \hat{\mathbf{F}}' \hat{\mathbf{F}})^{-1} - \bar{\mathbf{H}}^{-1} (T^{-1} \mathbf{F}' \mathbf{F})^{-1} \bar{\mathbf{H}}^{-1}] T^{-1} \bar{\mathbf{H}} \mathbf{F}' \mathbf{F} \boldsymbol{\lambda}_i,
\end{aligned}$$

where the summand in last term on the right is

$$\begin{aligned}
&T^{-1} \|\boldsymbol{\eta}'_i \mathbf{D} [(T^{-1} \hat{\mathbf{F}}' \hat{\mathbf{F}})^{-1} - \bar{\mathbf{H}}^{-1} (T^{-1} \mathbf{F}' \mathbf{F})^{-1} \bar{\mathbf{H}}^{-1}] T^{-1} \bar{\mathbf{H}} \mathbf{F}' \mathbf{F} \bar{\mathbf{H}}^{-1} \boldsymbol{\lambda}_i\| \\
&\leq \left(\frac{1}{T} \sum_{t=1}^T \|\boldsymbol{\eta}_{it}\|^2 \right)^{1/2} \left(\frac{1}{T} \sum_{t=1}^T \|\mathbf{d}_t\|^2 \right)^{1/2} \|(T^{-1} \hat{\mathbf{F}}' \hat{\mathbf{F}})^{-1} - \bar{\mathbf{H}}^{-1} (T^{-1} \mathbf{F}' \mathbf{F})^{-1} \bar{\mathbf{H}}^{-1}\| \\
&\quad \times \|T^{-1} \bar{\mathbf{H}} \mathbf{F}' \mathbf{F} \bar{\mathbf{H}}^{-1}\| \|\boldsymbol{\lambda}_i\| \\
&= O_p(N^{-1/2}) [O_p(N^{-1}) + O_p((NT)^{-1/2})] \\
&= O_p(T^{-1/2} N^{-1}) + O_p(N^{-3/2}).
\end{aligned}$$

The order of the second term in \mathbf{K}_{22} is \sqrt{NT} times this, which is $O_p(N^{-1/2}) + O_p(\sqrt{T} N^{-1})$. The first term is, via Lemma 2.B.4,

$$\begin{aligned}
\frac{1}{\sqrt{NT}} \sum_{i=1}^N \boldsymbol{\eta}'_i \mathbf{D} \bar{\mathbf{H}}^{-1} \boldsymbol{\lambda}_i &= \frac{\sqrt{T}}{\sqrt{N}} \frac{1}{T} \sum_{i=1}^N \boldsymbol{\eta}'_i \mathbf{D} \bar{\mathbf{H}}^{-1} \boldsymbol{\lambda}_i \\
&= \frac{\sqrt{T}}{\sqrt{N}} \frac{1}{N} \sum_{i=1}^N \boldsymbol{\Sigma}_{\eta,i}(\boldsymbol{\beta}, \mathbf{I}_m) \mathbf{Z}_i \bar{\mathbf{H}}^{-1} \boldsymbol{\lambda}_i + O_p(N^{-1/2})
\end{aligned}$$

By using this we obtain

$$\mathbf{K}_{22} = \sqrt{T} N^{-1/2} \mathbf{B}_2 + O_p(\sqrt{T} N^{-1}) + O_p(N^{-1/2}),$$

where $\mathbf{B}_2 = N^{-1} \sum_{i=1}^N \boldsymbol{\Sigma}_{\eta,i}(\boldsymbol{\beta}, \mathbf{I}_m) \mathbf{Z}_i \bar{\mathbf{H}} (\bar{\mathbf{H}}' \bar{\mathbf{H}})^{-1} \boldsymbol{\lambda}_i$. Thus, by adding the results,

$$\begin{aligned}
-\frac{1}{\sqrt{NT}} \sum_{i=1}^N \mathbf{X}'_i \mathbf{M}_{\hat{\mathbf{F}}} \mathbf{D} \bar{\mathbf{H}}^{-1} \boldsymbol{\lambda}_i &= \mathbf{K}_1 + \mathbf{K}_2 \\
&= \sqrt{T} N^{-1/2} (\mathbf{B}_1 - \mathbf{B}_2) + O_p(N^{-1/2}) \\
&\quad + O_p(\sqrt{T} N^{-1}).
\end{aligned}$$

Next, consider $(NT)^{-1/2} \sum_{i=1}^N \mathbf{X}'_i \mathbf{M}_{\hat{\mathbf{F}}} \boldsymbol{\epsilon}_i$, the first term in the numerator of $\sqrt{NT}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})$. Clearly,

$$\frac{1}{\sqrt{NT}} \sum_{i=1}^N \mathbf{X}'_i \mathbf{M}_{\hat{\mathbf{F}}} \boldsymbol{\epsilon}_i = \frac{1}{\sqrt{NT}} \sum_{i=1}^N \mathbf{X}'_i \mathbf{M}_{F\bar{H}'} \boldsymbol{\epsilon}_i - \frac{1}{\sqrt{NT}} \sum_{i=1}^N \mathbf{X}'_i (\mathbf{M}_{F\bar{H}'} - \mathbf{M}_{\hat{\mathbf{F}}}) \boldsymbol{\epsilon}_i,$$

where

$$\begin{aligned} & \frac{1}{\sqrt{NT}} \sum_{i=1}^N \mathbf{X}'_i (\mathbf{M}_{F\bar{H}'} - \mathbf{M}_{\hat{\mathbf{F}}}) \boldsymbol{\epsilon}_i \\ &= \frac{1}{\sqrt{NT}} \sum_{i=1}^N \mathbf{X}'_i \mathbf{D} (\hat{\mathbf{F}}' \hat{\mathbf{F}})^{-1} \mathbf{D}' \boldsymbol{\epsilon}_i + \frac{1}{\sqrt{NT}} \sum_{i=1}^N \mathbf{X}'_i \mathbf{D} (\hat{\mathbf{F}}' \hat{\mathbf{F}})^{-1} \bar{\mathbf{H}} \mathbf{F}' \boldsymbol{\epsilon}_i \\ & \quad + \frac{1}{\sqrt{NT}} \sum_{i=1}^N \mathbf{X}'_i \mathbf{F} \bar{\mathbf{H}}' (\hat{\mathbf{F}}' \hat{\mathbf{F}})^{-1} \mathbf{D}' \boldsymbol{\epsilon}_i \\ & \quad + \frac{1}{\sqrt{NT}} \sum_{i=1}^N \mathbf{X}'_i \mathbf{F} \bar{\mathbf{H}}' [(\hat{\mathbf{F}}' \hat{\mathbf{F}})^{-1} - (\bar{\mathbf{H}} \mathbf{F}' \bar{\mathbf{H}}')^+] \bar{\mathbf{H}} \mathbf{F}' \boldsymbol{\epsilon}_i \\ &= \mathbf{L}_1 + \dots + \mathbf{L}_4. \end{aligned}$$

The order of $\mathbf{L}_1, \dots, \mathbf{L}_4$ can be obtained by using the same steps as when analyzing \mathbf{K}_2 . For \mathbf{L}_1 , we use the fact that $\mathbf{x}_{it} = \boldsymbol{\Lambda}_i \mathbf{f}_t + \boldsymbol{\eta}_{it}$, giving

$$\left\| \frac{1}{\sqrt{N}} \sum_{i=1}^N \mathbf{x}_{it} \boldsymbol{\epsilon}_{is} \right\| \leq \left\| \frac{1}{\sqrt{N}} \sum_{i=1}^N \boldsymbol{\Lambda}_i \boldsymbol{\epsilon}_{is} \right\| \|\mathbf{F}_t\| + \left\| \frac{1}{\sqrt{N}} \sum_{i=1}^N \boldsymbol{\eta}_{it} \boldsymbol{\epsilon}_{is} \right\| = O_p(1),$$

which in turn implies

$$\begin{aligned} \|\mathbf{L}_1\| &= \left\| \frac{1}{\sqrt{NT}} \sum_{i=1}^N \mathbf{X}'_i \mathbf{D} (\hat{\mathbf{F}}' \hat{\mathbf{F}})^{-1} \mathbf{D}' \boldsymbol{\epsilon}_i \right\| \\ &= \left\| \frac{1}{\sqrt{NT}} \sum_{t=1}^T \sum_{s=1}^T \mathbf{d}'_t (\hat{\mathbf{F}}' \hat{\mathbf{F}})^{-1} \mathbf{d}_s \sum_{i=1}^N \mathbf{x}_{it} \boldsymbol{\epsilon}_{is} \right\| \\ &\leq \sqrt{T} \frac{1}{T} \sum_{t=1}^T \|\mathbf{d}_t\|^2 \|(T^{-1} \hat{\mathbf{F}}' \hat{\mathbf{F}})^{-1}\| \left(\frac{1}{T^2} \sum_{t=1}^T \sum_{s=1}^T \left\| \frac{1}{\sqrt{N}} \sum_{i=1}^N \mathbf{x}_{it} \boldsymbol{\epsilon}_{is} \right\|^2 \right)^{1/2} \\ &= O_p(\sqrt{T} N^{-1}). \end{aligned}$$

Similarly,

$$\begin{aligned}
\|\mathbf{L}_4\| &= \left\| \frac{1}{\sqrt{NT}} \sum_{i=1}^N \mathbf{X}'_i \mathbf{F} \bar{\mathbf{H}}' [(\hat{\mathbf{F}}' \hat{\mathbf{F}})^{-1} - (\bar{\mathbf{H}} \mathbf{F}' \bar{\mathbf{F}} \mathbf{H}')^+] \bar{\mathbf{H}} \mathbf{F}' \boldsymbol{\epsilon}_i \right\| \\
&\leq \sqrt{N} \|\bar{\mathbf{H}}\|^2 \left(\frac{1}{N} \sum_{i=1}^N \|T^{-1} \mathbf{X}'_i \mathbf{F}\|^2 \right)^{1/2} T \|(\hat{\mathbf{F}}' \hat{\mathbf{F}})^{-1} - (\bar{\mathbf{H}} \mathbf{F}' \bar{\mathbf{F}} \mathbf{H}')^+\| \\
&\quad \times \left(\frac{1}{N} \sum_{i=1}^N \|T^{-1/2} \mathbf{F}' \boldsymbol{\epsilon}_i\|^2 \right)^{1/2} \\
&= O_p(N^{-1/2}) + O_p(T^{-1/2}).
\end{aligned}$$

Consider \mathbf{L}_2 . Adding and subtracting $\bar{\mathbf{H}}^- (\mathbf{F}' \mathbf{F})^{-1} \bar{\mathbf{H}}^{-\prime}$ give

$$\begin{aligned}
\mathbf{L}_2 &= \frac{1}{\sqrt{NT}} \sum_{i=1}^N \mathbf{X}'_i \mathbf{D} (\hat{\mathbf{F}}' \hat{\mathbf{F}})^{-1} \bar{\mathbf{H}} \mathbf{F}' \boldsymbol{\epsilon}_i \\
&= \frac{1}{\sqrt{NT}} \sum_{i=1}^N \mathbf{X}'_i \mathbf{D} \bar{\mathbf{H}}^- (\mathbf{F}' \mathbf{F})^{-1} \mathbf{F}' \boldsymbol{\epsilon}_i \\
&\quad + \frac{1}{\sqrt{NT}} \sum_{i=1}^N \mathbf{X}'_i \mathbf{D} [(\hat{\mathbf{F}}' \hat{\mathbf{F}})^{-1} - \bar{\mathbf{H}}^- (\mathbf{F}' \mathbf{F})^{-1} \bar{\mathbf{H}}^{-\prime}] \bar{\mathbf{H}} \mathbf{F}' \boldsymbol{\epsilon}_i \\
&= \mathbf{L}_{21} + \mathbf{L}_{22},
\end{aligned}$$

where

$$\begin{aligned}
\|\mathbf{L}_{22}\| &= \left\| \frac{1}{\sqrt{NT}} \sum_{i=1}^N \mathbf{X}'_i \mathbf{D} [(\hat{\mathbf{F}}' \hat{\mathbf{F}})^{-1} - \bar{\mathbf{H}}^- (\mathbf{F}' \mathbf{F})^{-1} \bar{\mathbf{H}}^{-\prime}] \bar{\mathbf{H}} \mathbf{F}' \boldsymbol{\epsilon}_i \right\| \\
&= \left\| \frac{1}{\sqrt{NT}} \sum_{i=1}^N \sum_{t=1}^T \sum_{s=1}^T \mathbf{x}_{it} \mathbf{d}'_t [(\hat{\mathbf{F}}' \hat{\mathbf{F}})^{-1} - \bar{\mathbf{H}}^- (\mathbf{F}' \mathbf{F})^{-1} \bar{\mathbf{H}}^{-\prime}] \bar{\mathbf{H}} \mathbf{f}_s \boldsymbol{\epsilon}_{is} \right\| \\
&= \sqrt{T} \|\bar{\mathbf{H}}\| \left(\frac{1}{T} \sum_{t=1}^T \|\mathbf{d}_t\|^2 \right)^{1/2} T \|(\hat{\mathbf{F}}' \hat{\mathbf{F}})^{-1} - \bar{\mathbf{H}}^- (\mathbf{F}' \mathbf{F})^{-1} \bar{\mathbf{H}}^{-\prime}\| \\
&\quad \times \left(\frac{1}{T} \sum_{s=1}^T \|\mathbf{f}_s\|^2 \right)^{1/2} \left(\frac{1}{T^2} \sum_{t=1}^T \sum_{s=1}^T \left\| \frac{1}{\sqrt{N}} \sum_{i=1}^N \boldsymbol{\epsilon}_{is} \mathbf{x}_{it} \right\|^2 \right)^{1/2} \\
&= \sqrt{T} O_p(N^{-1/2}) [O_p(N^{-1}) + O_p((NT)^{-1/2})] \\
&= O_p(N^{-1}) + O_p(\sqrt{T} N^{-3/2}).
\end{aligned}$$

Also, from $\mathbf{X}_i = \mathbf{F}\boldsymbol{\Lambda}'_i + \boldsymbol{\eta}_i$,

$$\begin{aligned}\mathbf{L}_{21} &= \frac{1}{\sqrt{NT}} \sum_{i=1}^N \mathbf{X}'_i \mathbf{D} \bar{\mathbf{H}}^{-1} (\mathbf{F}'\mathbf{F})^{-1} \mathbf{F}' \boldsymbol{\epsilon}_i \\ &= \frac{1}{\sqrt{NT}} \sum_{i=1}^N \boldsymbol{\Lambda}_i \mathbf{F}' \mathbf{D} \bar{\mathbf{H}}^{-1} (\mathbf{F}'\mathbf{F})^{-1} \mathbf{F}' \boldsymbol{\epsilon}_i + \frac{1}{\sqrt{NT}} \sum_{i=1}^N \boldsymbol{\eta}'_i \mathbf{D} \bar{\mathbf{H}}^{-1} (\mathbf{F}'\mathbf{F})^{-1} \mathbf{F}' \boldsymbol{\epsilon}_i.\end{aligned}$$

Lemma 2.B.4 implies $\|NT^{-1}\boldsymbol{\eta}'_i \mathbf{D}\| = O_p(1)$, from which it follows that

$$\begin{aligned}&\left\| \frac{1}{\sqrt{NT}} \sum_{i=1}^N \boldsymbol{\eta}'_i \mathbf{D} \bar{\mathbf{H}}^{-1} (\mathbf{F}'\mathbf{F})^{-1} \mathbf{F}' \boldsymbol{\epsilon}_i \right\| \\ &\leq \frac{1}{\sqrt{N}} \frac{1}{N} \sum_{i=1}^N \|NT^{-1}\boldsymbol{\eta}'_i \mathbf{D}\| \|\bar{\mathbf{H}}^{-1}\| \|(T^{-1}\mathbf{F}'\mathbf{F})^{-1}\| \|T^{-1/2}\mathbf{F}' \boldsymbol{\epsilon}_i\| = O_p(N^{-1/2}),\end{aligned}$$

and by further use of Lemma 2.B.2,

$$\begin{aligned}&\left\| \frac{1}{\sqrt{NT}} \sum_{i=1}^N \boldsymbol{\Lambda}_i \mathbf{F}' \mathbf{D} \bar{\mathbf{H}}^{-1} (\mathbf{F}'\mathbf{F})^{-1} \mathbf{F}' \boldsymbol{\epsilon}_i \right\| \\ &\leq \frac{1}{\sqrt{T}} \frac{1}{N} \sum_{i=1}^N \|\boldsymbol{\Lambda}_i\| \|\sqrt{NT}^{-1/2} \mathbf{F}' \mathbf{D}\| \|\bar{\mathbf{H}}^{-1}\| \|(T^{-1}\mathbf{F}'\mathbf{F})^{-1}\| \|T^{-1/2}\mathbf{F}' \boldsymbol{\epsilon}_i\| \\ &= O_p(T^{-1/2}).\end{aligned}$$

Consequently,

$$\begin{aligned}\|\mathbf{L}_{21}\| &\leq \left\| \frac{1}{\sqrt{NT}} \sum_{i=1}^N \boldsymbol{\Lambda}_i \mathbf{F}' \mathbf{D} \bar{\mathbf{H}}^{-1} (\mathbf{F}'\mathbf{F})^{-1} \mathbf{F}' \boldsymbol{\epsilon}_i \right\| \\ &\quad + \left\| \frac{1}{\sqrt{NT}} \sum_{i=1}^N \boldsymbol{\eta}'_i \mathbf{D} \bar{\mathbf{H}}^{-1} (\mathbf{F}'\mathbf{F})^{-1} \mathbf{F}' \boldsymbol{\epsilon}_i \right\| \\ &= O_p(N^{-1/2}) + O_p(T^{-1/2}),\end{aligned}$$

leading to the following result for $\|\mathbf{L}_2\|$:

$$\|\mathbf{L}_2\| \leq \|\mathbf{L}_{21}\| + \|\mathbf{L}_{22}\| = O_p(N^{-1/2}) + O_p(T^{-1/2}) + O_p(\sqrt{T}N^{-3/2}). \quad (2.B.7)$$

Consider \mathbf{L}_3 . As when evaluating \mathbf{L}_2 , we begin by adding and subtracting $\bar{\mathbf{H}}^{-}(\mathbf{F}'\mathbf{F})^{-1}\bar{\mathbf{H}}^{-\prime}$;

$$\begin{aligned}\mathbf{L}_3 &= \frac{1}{\sqrt{NT}} \sum_{i=1}^N \mathbf{X}'_i \mathbf{F} \bar{\mathbf{H}}' (\hat{\mathbf{F}}' \hat{\mathbf{F}})^{-1} \mathbf{D}' \epsilon_i \\ &= \frac{1}{\sqrt{NT}} \sum_{i=1}^N \mathbf{X}'_i \mathbf{F} (\mathbf{F}' \mathbf{F})^{-1} \bar{\mathbf{H}}^{-\prime} \mathbf{D}' \epsilon_i \\ &\quad + \frac{1}{\sqrt{NT}} \sum_{i=1}^N \mathbf{X}'_i \mathbf{F} \bar{\mathbf{H}}' [(\hat{\mathbf{F}}' \hat{\mathbf{F}})^{-1} - \bar{\mathbf{H}}^{-}(\mathbf{F}' \mathbf{F})^{-1} \bar{\mathbf{H}}^{-\prime}] \mathbf{D}' \epsilon_i,\end{aligned}$$

where, in analogy to $\|\mathbf{L}_{22}\|$,

$$\begin{aligned}&\left\| \frac{1}{\sqrt{NT}} \sum_{i=1}^N \mathbf{X}'_i \mathbf{F} \bar{\mathbf{H}}' [(\hat{\mathbf{F}}' \hat{\mathbf{F}})^{-1} - \bar{\mathbf{H}}^{-}(\mathbf{F}' \mathbf{F})^{-1} \bar{\mathbf{H}}^{-\prime}] \mathbf{D}' \epsilon_i \right\| \\ &= \left\| \frac{1}{\sqrt{NT}} \sum_{i=1}^N \sum_{t=1}^T \sum_{s=1}^T \mathbf{x}_{it} \mathbf{F}'_t \bar{\mathbf{H}}' [(\hat{\mathbf{F}}' \hat{\mathbf{F}})^{-1} - \bar{\mathbf{H}}^{-}(\mathbf{F}' \mathbf{F})^{-1} \bar{\mathbf{H}}^{-\prime}] \mathbf{d}_s \epsilon_{is} \right\| \\ &\leq \sqrt{T} \left(\frac{1}{T} \sum_{t=1}^T \|\mathbf{f}_t\|^2 \right)^{1/2} \|\bar{\mathbf{H}}\| T \|(\hat{\mathbf{F}}' \hat{\mathbf{F}})^{-1} - \bar{\mathbf{H}}^{-}(\mathbf{F}' \mathbf{F})^{-1} \bar{\mathbf{H}}^{-\prime}\| \left(\frac{1}{T} \sum_{s=1}^T \|\mathbf{d}_s\|^2 \right)^{1/2} \\ &\quad \times \left(\frac{1}{T^2} \sum_{t=1}^T \sum_{s=1}^T \left\| \frac{1}{\sqrt{N}} \sum_{i=1}^N \mathbf{x}_{it} \epsilon_{is} \right\|^2 \right)^{1/2} \\ &= \sqrt{T} [O_p(N^{-1}) + O_p((NT)^{-1/2})] O_p(N^{-1/2}) \\ &= O_p(N^{-1}) + O_p(\sqrt{T} N^{-3/2}).\end{aligned}$$

and, by substituting for \mathbf{X}_i , via Lemma 2.B.5 the first term in \mathbf{L}_3 can be written

$$\begin{aligned}&\frac{1}{\sqrt{NT}} \sum_{i=1}^N \mathbf{X}'_i \mathbf{F} (\mathbf{F}' \mathbf{F})^{-1} \bar{\mathbf{H}}^{-\prime} \mathbf{D}' \epsilon_i \\ &= \frac{1}{\sqrt{NT}} \sum_{i=1}^N \boldsymbol{\Lambda}_i \mathbf{F}' \mathbf{F} (\mathbf{F}' \mathbf{F})^{-1} \bar{\mathbf{H}}^{-\prime} \mathbf{D}' \epsilon_i + \frac{1}{\sqrt{NT}} \sum_{i=1}^N \boldsymbol{\eta}'_i \mathbf{F} (\mathbf{F}' \mathbf{F})^{-1} \bar{\mathbf{H}}^{-\prime} \mathbf{D}' \epsilon_i \\ &= \frac{\sqrt{T}}{\sqrt{N}} \frac{1}{N} \sum_{i=1}^N \boldsymbol{\Lambda}_i (\bar{\mathbf{H}}^{-})' N T^{-1} \mathbf{D}' \epsilon_i + \frac{1}{\sqrt{N}} \frac{1}{N} \sum_{i=1}^N T^{-1/2} \boldsymbol{\eta}'_i \mathbf{F} (\mathbf{F}' \mathbf{F})^{-1} \bar{\mathbf{H}}^{-\prime} \mathbf{D}' \epsilon_i \\ &= \frac{\sqrt{T}}{\sqrt{N}} \frac{1}{N} \sum_{i=1}^N \sigma_{\epsilon_i}^2 \boldsymbol{\lambda}_i (\bar{\mathbf{H}}^{-})' \mathbf{Z}'_i (1, 0_m)' + O_p(N^{-1/2}).\end{aligned}$$

Hence, letting $\mathbf{B}_3 = N^{-1} \sum_{i=1}^N \sigma_{\epsilon_i}^2 \boldsymbol{\Lambda}_i (\overline{\mathbf{H}}' \overline{\mathbf{H}})^{-1} \overline{\mathbf{H}}' \mathbf{Z}'_i (1, 0_m)'$, we obtain

$$\mathbf{L}_3 = \sqrt{T} N^{-1/2} \mathbf{B}_3 + O_p(N^{-1/2}) + O_p(\sqrt{T} N^{-3/2}).$$

The results for $\mathbf{L}_1, \dots, \mathbf{L}_4$ imply

$$\begin{aligned} & \frac{1}{\sqrt{NT}} \sum_{i=1}^N \mathbf{X}'_i (\mathbf{M}_{F\overline{\mathbf{H}}'} - \mathbf{M}_{\widehat{F}}) \boldsymbol{\epsilon}_i \\ &= \mathbf{L}_1 + \dots + \mathbf{L}_4 \\ &= \sqrt{T} N^{-1/2} \mathbf{B}_3 + O_p(\sqrt{T} N^{-1}) + O_p(T^{-1/2}) + O_p(N^{-1/2}). \end{aligned}$$

which in turn implies

$$\begin{aligned} \frac{1}{\sqrt{NT}} \sum_{i=1}^N \mathbf{X}'_i \mathbf{M}_{\widehat{F}} \boldsymbol{\epsilon}_i &= \frac{1}{\sqrt{NT}} \sum_{i=1}^N \mathbf{X}'_i \mathbf{M}_{F\overline{\mathbf{H}}'} \boldsymbol{\epsilon}_i - \sqrt{T} N^{-1/2} \mathbf{B}_3 + O_p(\sqrt{T} N^{-1}) \\ &\quad + O_p(T^{-1/2}) + O_p(N^{-1/2}). \end{aligned} \tag{2.B.8}$$

Consider the first term in of the equation above. The variance of

$$(NT)^{-1/2} \sum_{i=1}^N \boldsymbol{\eta}'_i \mathbf{F} \overline{\mathbf{H}}' (\overline{\mathbf{H}} \mathbf{F}' \mathbf{F} \overline{\mathbf{H}}')^+ \overline{\mathbf{H}} \mathbf{F}' \boldsymbol{\epsilon}_i$$

is $O_p(T^{-1})$; hence,

$$\left\| \frac{1}{\sqrt{NT}} \sum_{i=1}^N \boldsymbol{\eta}'_i \mathbf{F} \overline{\mathbf{H}}' (\overline{\mathbf{H}} \mathbf{F}' \mathbf{F} \overline{\mathbf{H}}')^+ \overline{\mathbf{H}} \mathbf{F}' \boldsymbol{\epsilon}_i \right\| = O_p(T^{-1/2}).$$

This result, together with the fact that $\mathbf{M}_{F\overline{\mathbf{H}}'} \mathbf{X}_i = \mathbf{M}_{F\overline{\mathbf{H}}'} \boldsymbol{\eta}_i$, implies

$$\begin{aligned} \frac{1}{\sqrt{NT}} \sum_{i=1}^N \mathbf{X}'_i \mathbf{M}_{F\overline{\mathbf{H}}'} \boldsymbol{\epsilon}_i &= \frac{1}{\sqrt{NT}} \sum_{i=1}^N \boldsymbol{\eta}'_i \mathbf{M}_{F\overline{\mathbf{H}}'} \boldsymbol{\epsilon}_i \\ &= \frac{1}{\sqrt{NT}} \sum_{i=1}^N \boldsymbol{\eta}'_i \boldsymbol{\epsilon}_i - \frac{1}{\sqrt{NT}} \sum_{i=1}^N \boldsymbol{\eta}'_i \mathbf{F} \overline{\mathbf{H}}' (\overline{\mathbf{H}} \mathbf{F}' \mathbf{F} \overline{\mathbf{H}}')^+ \overline{\mathbf{H}} \mathbf{F}' \boldsymbol{\epsilon}_i \\ &= \frac{1}{\sqrt{NT}} \sum_{i=1}^N \boldsymbol{\eta}'_i \boldsymbol{\epsilon}_i + O_p(T^{-1/2}). \end{aligned}$$

Consider the variance of $\frac{1}{\sqrt{NT}} \sum_{i=1}^N \boldsymbol{\eta}'_i \boldsymbol{\epsilon}_i$. We have

$$\begin{aligned}
 E(\boldsymbol{\epsilon}_{it} \boldsymbol{\epsilon}_{i,t-s}) &= E[\phi_i(L) \boldsymbol{\epsilon}_{it} \phi_i(L) \boldsymbol{\epsilon}_{i,t-s}] = \sum_{j=0}^{\infty} \sum_{k=0}^{\infty} \phi_{j,i} \phi_{k,i} E(\boldsymbol{\epsilon}_{i,t-j} \boldsymbol{\epsilon}_{i,t-s-k}) \\
 &= \sum_{k=0}^{\infty} \phi_{s+k,i} \phi_{k,i} E(\boldsymbol{\epsilon}_{i,t-s-k}^2) = \sum_{k=0}^{\infty} \phi_{s+k,i} \phi_{k,i} = \gamma_{s,i}, \\
 E(\boldsymbol{\eta}_{it} \boldsymbol{\eta}'_{i,t-s}) &= E[\Phi_i(L) \mathbf{v}_{it} (\phi_i(L) \mathbf{v}_{i,t-s})'] = \sum_{j=0}^{\infty} \sum_{k=0}^{\infty} \Phi_{j,i} E(\mathbf{v}_{i,t-j} \mathbf{v}'_{i,t-s-k}) \Phi'_{k,i} \\
 &= \sum_{k=0}^{\infty} \Phi_{s+k,i} E(\mathbf{v}_{i,t-s-k} \mathbf{v}'_{i,t-s-k}) \Phi'_{k,i} = \sum_{k=0}^{\infty} \Phi_{s+k,i} \Phi'_{k,i} = \Gamma_{s,i}
 \end{aligned}$$

for any $s = -(T-t), \dots, t-1$. Hence, letting $\mathbf{W}_i = T^{-1} \sum_{t=1}^T \sum_{s=-(T-t)}^{t-1} \gamma_{s,i} \boldsymbol{\Gamma}_{s,i}$ and $\overline{\mathbf{W}} = N^{-1} \sum_{i=1}^N \mathbf{W}_i$,

$$\begin{aligned}
 &E \left[\left(\frac{1}{\sqrt{NT}} \sum_{i=1}^N \boldsymbol{\eta}'_i \boldsymbol{\epsilon}_i \right) \left(\frac{1}{\sqrt{NT}} \sum_{i=1}^N \boldsymbol{\eta}'_i \boldsymbol{\epsilon}_i \right)' \right] \\
 &= \frac{1}{NT} \sum_{i=1}^N \sum_{j=1}^N E[\boldsymbol{\eta}'_i E(\boldsymbol{\epsilon}_i \boldsymbol{\epsilon}'_j | \boldsymbol{\eta}_i, \boldsymbol{\eta}_j) \boldsymbol{\eta}_j] = \frac{1}{NT} \sum_{i=1}^N \sum_{t=1}^T \sum_{s=1}^T E(\boldsymbol{\epsilon}_{it} \boldsymbol{\epsilon}_{is}) E(\boldsymbol{\eta}_{it} \boldsymbol{\eta}'_{is}) \\
 &= \frac{1}{NT} \sum_{i=1}^N \sum_{t=1}^T E(\boldsymbol{\epsilon}_{it}^2) E(\boldsymbol{\eta}_{it} \boldsymbol{\eta}'_{it}) + \frac{1}{NT} \sum_{i=1}^N \sum_{t=1}^T \sum_{s=1}^{t-1} E(\boldsymbol{\epsilon}_{it} \boldsymbol{\epsilon}_{i,t-s}) E(\boldsymbol{\eta}_{it} \boldsymbol{\eta}'_{i,t-s}) \\
 &\quad + \frac{1}{NT} \sum_{i=1}^N \sum_{t=1}^T \sum_{s=1}^{T-t} E(\boldsymbol{\epsilon}_{it} \boldsymbol{\epsilon}_{i,t+s}) E(\boldsymbol{\eta}_{it} \boldsymbol{\eta}'_{i,t+s}) \\
 &= \frac{1}{NT} \sum_{i=1}^N \sum_{t=1}^T \gamma_{0,i} \boldsymbol{\Gamma}_{0,i} + \frac{1}{NT} \sum_{i=1}^N \sum_{t=1}^T \sum_{s=1}^{t-1} \gamma_{s,i} \boldsymbol{\Gamma}_{s,i} + \frac{1}{NT} \sum_{i=1}^N \sum_{t=1}^T \sum_{s=1}^{T-t} \gamma_{-s,i} \boldsymbol{\Gamma}_{-s,i} \\
 &= \frac{1}{NT} \sum_{i=1}^N \sum_{t=1}^T \sum_{s=-(T-t)}^{t-1} \gamma_{s,i} \boldsymbol{\Gamma}_{s,i} = \overline{\mathbf{W}}.
 \end{aligned}$$

Moreover, since the fourth-order moments of $\boldsymbol{\eta}_{it}$ and $\boldsymbol{\epsilon}_{it}$ are bounded by assumption, by a central limit law for i.i.d. variates,

$$\frac{1}{\sqrt{NT}} \sum_{i=1}^N \boldsymbol{\eta}'_i \boldsymbol{\epsilon}_i \sim N(0, \overline{\mathbf{W}})$$

as $N, T \rightarrow \infty$, where \sim signifies asymptotic equivalence. Thus, provided that $\sqrt{TN}^{-1} = o(1)$,

$$\begin{aligned} \frac{1}{\sqrt{NT}} \sum_{i=1}^N \mathbf{X}'_i \mathbf{M}_{\hat{\mathbf{F}}} \boldsymbol{\epsilon}_i &= \frac{1}{\sqrt{NT}} \sum_{i=1}^N \boldsymbol{\eta}'_i \boldsymbol{\epsilon}_i - \sqrt{T} N^{-1/2} \mathbf{B}_3 + O_p(\sqrt{T} N^{-1}) \\ &\quad + O_p(T^{-1/2}) + O_p(N^{-1/2}) \\ &\sim N(0, \overline{\mathbf{W}}) - \sqrt{T} N^{-1/2} \mathbf{B}_3. \end{aligned}$$

Let $\mathbf{B} = \mathbf{B}_1 - \mathbf{B}_2 - \mathbf{B}_3$. The above results suggest the following limit for the numerator of $\sqrt{NT}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})$:

$$\begin{aligned} &\frac{1}{\sqrt{NT}} \sum_{i=1}^N (\mathbf{X}'_i \mathbf{M}_{\hat{\mathbf{F}}} \boldsymbol{\epsilon}_i - \mathbf{X}'_i \mathbf{M}_{\hat{\mathbf{F}}} \mathbf{D} \overline{\mathbf{H}}^{-1} \boldsymbol{\lambda}_i) \\ &= \frac{1}{\sqrt{NT}} \sum_{i=1}^N \boldsymbol{\eta}'_i \boldsymbol{\epsilon}_i + \sqrt{T} N^{-1/2} \mathbf{B} + O_p(\sqrt{T} N^{-1}) + O_p(N^{-1/2}) + O_p(T^{-1/2}) \\ &\sim N(0, \overline{\mathbf{W}}) + \sqrt{T} N^{-1/2} \mathbf{B}, \end{aligned} \tag{2.B.9}$$

which holds as $N, T \rightarrow \infty$ with $\sqrt{TN}^{-1} \rightarrow 0$.

Next, consider the denominator of $\sqrt{NT}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})$, which we expand as

$$\frac{1}{NT} \sum_{i=1}^N \mathbf{X}'_i \mathbf{M}_{\hat{\mathbf{F}}} \mathbf{X}_i = \frac{1}{NT} \sum_{i=1}^N \mathbf{X}'_i \mathbf{M}_{F\overline{\mathbf{H}}} \mathbf{X}_i - \frac{1}{NT} \sum_{i=1}^N \mathbf{X}'_i (\mathbf{M}_{F\overline{\mathbf{H}}} - \mathbf{M}_{\hat{\mathbf{F}}}) \mathbf{X}_i.$$

By Lemma 2.B.3, $NT^{-1} \|\mathbf{D}\|^2 = O_p(1)$, implying $\sqrt{NT} T^{-1/2} \|\mathbf{D}\| = O_p(1)$. Similarly, since $T^{-1} \|\mathbf{F}\|^2 = T^{-1} \sum_{t=1}^T \|\mathbf{f}_t\|^2 = O_p(1)$, we have $T^{-1/2} \|\mathbf{F}\| = O_p(1)$.

Hence,

$$\begin{aligned} \|\mathbf{M}_{F\overline{\mathbf{H}}} - \mathbf{M}_{\hat{\mathbf{F}}}\| &\leq T^{-1} \|\mathbf{D}\|^2 \|(T^{-1} \hat{\mathbf{F}}' \hat{\mathbf{F}})^{-1}\| + 2T^{-1} \|\overline{\mathbf{H}}\| \|\mathbf{D}\| \|\mathbf{F}\| \|(T^{-1} \hat{\mathbf{F}}' \hat{\mathbf{F}})^{-1}\| \\ &\quad + T^{-1} \|\overline{\mathbf{H}}\|^2 \|\mathbf{F}\|^2 T \|\hat{\mathbf{F}}' \hat{\mathbf{F}}\|^{-1} - (\overline{\mathbf{H}} \mathbf{F}' \mathbf{F} \overline{\mathbf{H}}')^+ \\ &= O_p(N^{-1/2}). \end{aligned}$$

implying

$$\left\| \frac{1}{NT} \sum_{i=1}^N \mathbf{X}'_i (\mathbf{M}_{F\overline{\mathbf{H}}} - \mathbf{M}_{\hat{\mathbf{F}}}) \mathbf{X}_i \right\| \leq \|\mathbf{M}_{F\overline{\mathbf{H}}} - \mathbf{M}_{\hat{\mathbf{F}}}\| \frac{1}{NT} \sum_{i=1}^N \|\mathbf{X}_i\|^2 = O_p(N^{-1/2}).$$

By using this and

$$\mathbf{X}'_i \mathbf{M}_{F\overline{\mathbf{H}}} \mathbf{X}_i = \boldsymbol{\eta}'_i \mathbf{M}_{F\overline{\mathbf{H}}} \boldsymbol{\eta}_i = \boldsymbol{\eta}'_i \boldsymbol{\eta}_i - T^{-1/2} \boldsymbol{\eta}'_i \mathbf{F} (T^{-1} \mathbf{F}' \mathbf{F})^{-1} T^{-1/2} \mathbf{F}' \boldsymbol{\eta}_i$$

$$= \boldsymbol{\eta}'_i \boldsymbol{\eta}_i + O_p(1),$$

we obtain

$$\begin{aligned} \frac{1}{NT} \sum_{i=1}^N \mathbf{X}'_i \mathbf{M}_{\widehat{\mathbf{F}}} \mathbf{X}_i &= \frac{1}{NT} \sum_{i=1}^N \mathbf{X}'_i \mathbf{M}_{F\overline{H}} \mathbf{X}_i + O_p(N^{-1/2}) \\ &= \overline{\boldsymbol{\Sigma}}_\eta + \frac{1}{N} \sum_{i=1}^N (T^{-1} \boldsymbol{\eta}'_i \boldsymbol{\eta}_i - \boldsymbol{\Sigma}_{\eta,i}) + O_p(N^{-1/2}) \\ &= \overline{\boldsymbol{\Sigma}}_\eta + O_p(N^{-1/2}). \end{aligned} \quad (2.B.10)$$

By adding all the results, as $N, T \rightarrow \infty$ with $\sqrt{TN}^{-1} \rightarrow 0$,

$$\begin{aligned} &\sqrt{NT}(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}) \\ &= \left(\frac{1}{NT} \sum_{i=1}^N \mathbf{X}'_i \mathbf{M}_{\widehat{\mathbf{F}}} \mathbf{X}_i \right)^{-1} \frac{1}{\sqrt{NT}} \sum_{i=1}^N (\mathbf{X}'_i \mathbf{M}_{\widehat{\mathbf{F}}} \boldsymbol{\epsilon}_i - \mathbf{X}'_i \mathbf{M}_{\widehat{\mathbf{F}}} \mathbf{D}\overline{\mathbf{H}}^{-1} \boldsymbol{\lambda}_i) \\ &= \overline{\boldsymbol{\Sigma}}_\eta^{-1} \left(\frac{1}{\sqrt{NT}} \sum_{i=1}^N \boldsymbol{\eta}'_i \boldsymbol{\epsilon}_i + \sqrt{TN}^{-1/2} \mathbf{B} \right) + O_p(\sqrt{TN}^{-1}) + O_p(N^{-1/2}) \\ &\quad + O_p(T^{-1/2}) \\ &\sim N(0, \overline{\boldsymbol{\Sigma}}_\eta^{-1} \overline{\mathbf{W}} \overline{\boldsymbol{\Sigma}}_\eta^{-1}) + \overline{\boldsymbol{\Sigma}}_\eta^{-1} \sqrt{TN}^{-1/2} \mathbf{B}. \end{aligned}$$

This completes the proof. ■

2.B.3 Proof of Proposition 2.3.1

Write

$$\begin{aligned} \sqrt{NT}(\widehat{\boldsymbol{\beta}}_{BA} - \boldsymbol{\beta}) &= \sqrt{NT}(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}) - \sqrt{TN}^{-1/2} \widehat{\boldsymbol{\Sigma}}_\eta^{-1} \widehat{\mathbf{B}} \\ &= \sqrt{NT}(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}) - \sqrt{TN}^{-1/2} \boldsymbol{\Sigma}_\eta^{-1} \mathbf{B} - \sqrt{TN}^{-1/2} \widehat{\boldsymbol{\Sigma}}_\eta^{-1} (\widehat{\mathbf{B}} - \mathbf{B}) \\ &\quad - \sqrt{TN}^{-1/2} (\widehat{\boldsymbol{\Sigma}}_\eta^{-1} - \boldsymbol{\Sigma}_\eta^{-1}) \mathbf{B}. \end{aligned}$$

Consider $\sqrt{TN}^{-1/2} \widehat{\boldsymbol{\Sigma}}_\eta^{-1} (\widehat{\mathbf{B}} - \mathbf{B})$. From Theorem 2.1, we have that $(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}) = O_p((NT)^{-1/2}) + O_p(N^{-1})$. $\widehat{\mathbf{C}}_i$ is given by

$$\widehat{\mathbf{C}}_i = (T^{-1} \widehat{\mathbf{F}}' \widehat{\mathbf{F}})^{-1} T^{-1} \widehat{\mathbf{F}}' \mathbf{W}_i,$$

where

$$T^{-1} \widehat{\mathbf{F}}' \mathbf{W}_i = T^{-1} \widehat{\mathbf{F}}' \mathbf{F} \mathbf{C}_i + T^{-1} \widehat{\mathbf{F}}' \mathbf{U}_i$$

$$\begin{aligned}
 &= T^{-1}\overline{\mathbf{H}}\mathbf{F}'\mathbf{F}\mathbf{C}_i + T^{-1}\mathbf{D}'\mathbf{F}\mathbf{C}_i + T^{-1}\widehat{\mathbf{F}}'\mathbf{U}_i \\
 &= T^{-1}\overline{\mathbf{H}}\mathbf{F}'\mathbf{F}\mathbf{C}_i + T^{-1}\overline{\mathbf{H}}\mathbf{F}'\mathbf{U}_i + T^{-1}\mathbf{D}'\mathbf{F}\mathbf{C}_i + T^{-1}\mathbf{D}'\mathbf{U}_i.
 \end{aligned}$$

Clearly, $\|T^{-1}\mathbf{F}'\mathbf{U}_i\| = O_p(T^{-1/2})$, and by Lemmas 2.B.2, 2.B.4 and 2.B.5, $\|T^{-1}\mathbf{D}'\mathbf{F}\|$ and $\|T^{-1}\mathbf{D}'\mathbf{u}_i\|$ are $O_p((NT)^{-1/2})$. Moreover, from the proof of Theorem 2.1,

$$T\|(\widehat{\mathbf{F}}'\widehat{\mathbf{F}})^{-1} - \overline{\mathbf{H}}^-(\mathbf{F}'\mathbf{F})^{-1}\overline{\mathbf{H}}^{-'}\| = O_p(N^{-1}) + O_p((NT)^{-1/2}).$$

It follows that, with $\overline{\mathbf{H}}^- = \overline{\mathbf{H}}(\overline{\mathbf{H}}'\overline{\mathbf{H}})^{-1}$,

$$\begin{aligned}
 \widehat{\mathbf{C}}_i &= (T^{-1}\widehat{\mathbf{F}}'\widehat{\mathbf{F}})^{-1}T^{-1}\widehat{\mathbf{F}}'\mathbf{W}_i \\
 &= \overline{\mathbf{H}}^-(T^{-1}\mathbf{F}'\mathbf{F})^{-1}\overline{\mathbf{H}}^{-'}(T^{-1}\overline{\mathbf{H}}\mathbf{F}'\mathbf{F}\mathbf{C}_i + T^{-1}\overline{\mathbf{H}}\mathbf{F}'\mathbf{U}_i \\
 &\quad + T^{-1}\mathbf{D}'\mathbf{F}\mathbf{C}_i + T^{-1}\mathbf{D}'\mathbf{U}_i) + O_p(N^{-1}) + O_p((NT)^{-1/2}) \\
 &= \overline{\mathbf{H}}^-(T^{-1}\mathbf{F}'\mathbf{F})^{-1}\overline{\mathbf{H}}^{-'}T^{-1}\overline{\mathbf{H}}\mathbf{F}'\mathbf{F}\mathbf{C}_i + O_p(N^{-1}) + O_p(T^{-1/2}) \\
 &= \overline{\mathbf{H}}^-\mathbf{C}_i + O_p(N^{-1}) + O_p(T^{-1/2})
 \end{aligned}$$

$(\widehat{\boldsymbol{\Sigma}}_\eta - \boldsymbol{\Sigma}_\eta)$, $(\widehat{\boldsymbol{\Sigma}}_{\eta,i} - \boldsymbol{\Sigma}_{\eta,i})$ and $(\widehat{\sigma}_{\epsilon,i}^2 - \sigma_{\epsilon,i}^2)$ are all $O_p(T^{-1/2})$. This implies that

$$\|\widehat{\mathbf{B}} - \mathbf{B}\| = O_p(T^{-1/2}) + O_p(N^{-1}),$$

and therefore, with $\|\widehat{\boldsymbol{\Sigma}}_\eta^{-1}\| = O_p(1)$,

$$\begin{aligned}
 \|\sqrt{T}N^{-1/2}\widehat{\boldsymbol{\Sigma}}_\eta^{-1}(\widehat{\mathbf{B}} - \mathbf{B})\| &\leq \sqrt{T}N^{-1/2}\|\widehat{\boldsymbol{\Sigma}}_\eta^{-1}\|\|\widehat{\mathbf{B}} - \mathbf{B}\| \\
 &= O_p(N^{-1/2}) + O_p(\sqrt{T}N^{-3/2}),
 \end{aligned}$$

which is $o_p(1)$ under our assumption that $\sqrt{T}N^{-1} = o(1)$. Similarly, since $\|\mathbf{B}\| = O_p(1)$ and, by Taylor expansion, $\|\widehat{\boldsymbol{\Sigma}}_\eta^{-1} - \boldsymbol{\Sigma}_\eta^{-1}\| = O_p(T^{-1/2})$,

$$\|\sqrt{T}N^{-1/2}(\widehat{\boldsymbol{\Sigma}}_\eta^{-1} - \boldsymbol{\Sigma}_\eta^{-1})\mathbf{B}\| \leq \sqrt{T}N^{-1/2}\|\widehat{\boldsymbol{\Sigma}}_\eta^{-1} - \boldsymbol{\Sigma}_\eta^{-1}\|\|\mathbf{B}\| = O_p(N^{-1/2}).$$

Together with Theorem 2.1 these results imply

$$\sqrt{NT}(\widehat{\boldsymbol{\beta}}_{BA} - \boldsymbol{\beta}) = \sqrt{NT}(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}) - \sqrt{\tau}\boldsymbol{\Sigma}_\eta^{-1}\mathbf{B} + o_p(1) \rightarrow_d N(0, \overline{\boldsymbol{\Sigma}}_\eta^{-1}\overline{\mathbf{W}}\overline{\boldsymbol{\Sigma}}_\eta^{-1})$$

as $N, T \rightarrow \infty$ with $\sqrt{T}N^{-1} \rightarrow 0$ and $\sqrt{NT}^{-1} \rightarrow 0$. ■

2.B.4 Proof of Proposition 2.3.2

Theorem 2 of Bai and Ng (2002) states the consistency of the IC when applied to the PC estimator. According to Corollary 2 of Bai and Ng (2002) this result holds also when the method of estimation for the factors differs from PC, provided that the rate of consistency of the factor estimator is not greater than the rate that applies to PC, which is $\min\{\sqrt{N}, \sqrt{T}\}$. In our case, $\|\mathbf{d}_t\| = O_p(N^{-1/2})$, which means that the condition of Corollary 2 is not satisfied when $N < T$. Hence, since we cannot rely on Corollary 2 of Bai and Ng (2002), we adapt the proof of their Theorem 2 to the current setup.

We need to prove that

$$P[IC(s) < IC(k)] = P[V(s, \hat{\mathbf{F}}^s) - V(k, \hat{\mathbf{F}}^k) < (k-s)g(N)] \rightarrow 0 \quad (2.B.11)$$

as $N, T \rightarrow \infty$. Consider first the case when $s < k$. Under the condition that $g(N) \rightarrow 0$ as $N \rightarrow \infty$, it is sufficient to show that $V(s, \hat{\mathbf{F}}^s) - V(k, \hat{\mathbf{F}}^k)$ has a positive limit for each $1 \leq s < k$. For this purpose, we use the identity

$$\begin{aligned} V(s, \hat{\mathbf{F}}^s) - V(k, \hat{\mathbf{F}}^k) &= [V(s, \hat{\mathbf{F}}^s) - V(s, \mathbf{F}\bar{\mathbf{H}}^s)] + [V(s, \mathbf{F}\bar{\mathbf{H}}^s) - V(k, \mathbf{F}\bar{\mathbf{H}}^k)] \\ &\quad + [V(k, \mathbf{F}\bar{\mathbf{H}}^k) - V(k, \hat{\mathbf{F}}^k)]. \end{aligned} \quad (2.B.12)$$

It can be established that

$$V(s, \hat{\mathbf{F}}^s) - V(s, \mathbf{F}\bar{\mathbf{H}}^s) = O_p(N^{-1/2}).$$

The first and the third terms of (2.B.12) are of this form, so they are $O_p(N^{-1/2})$. The proof of this result is completely analogous to the proof of Lemma 2 in Bai and Ng (2002). For the second term of (2.B.12), we can show that, for any s with $1 \leq s \leq k$, there exists a $\kappa_s > 0$, such that

$$V(s, \mathbf{F}\bar{\mathbf{H}}^s) - V(k, \mathbf{F}\bar{\mathbf{H}}^k) \rightarrow \kappa_s$$

as $N, T \rightarrow \infty$. The proof of this result is completely analogous to the proof of Lemma 3 in Bai and Ng (2002). It follows that, if $s < k$, then $V(s, \hat{\mathbf{F}}^s) - V(k, \hat{\mathbf{F}}^k) = \kappa_s + o_p(1)$, as required.

Next, consider the case when $s \geq k$. The rank condition in Assumption 2.2 is satisfied in this case. Using this we now show that

$$|V(k, \hat{\mathbf{F}}^k) - V(s, \hat{\mathbf{F}}^s)| = O_p(N^{-1}).$$

We begin by noting that when $s \geq k$,

$$|V(k, \widehat{\mathbf{F}}^k) - V(s, \widehat{\mathbf{F}}^s)| \leq 2 \max_{k \leq s \leq \mathbf{K}_{max}} |V(k, \mathbf{F}) - V(s, \widehat{\mathbf{F}}^s)|.$$

Hence, if we can show that $|V(k, \mathbf{F}) - V(s, \widehat{\mathbf{F}}^s)| = O_p(N^{-1})$, then we are done.

Considering (2.B.4) and assuming that $\boldsymbol{\beta} = 0$ to simplify things, we have

$$\begin{aligned} V(s, \widehat{\mathbf{F}}^s) - V(k, \mathbf{F}) &= \frac{1}{NT} \sum_{i=1}^N (\boldsymbol{\epsilon}_i - d\overline{\mathbf{H}}^- \boldsymbol{\lambda}_i)' \mathbf{M}_{\widehat{\mathbf{F}}^s} (\boldsymbol{\epsilon}_i - d\overline{\mathbf{H}}^- \boldsymbol{\lambda}_i) \\ &\quad - \frac{1}{NT} \sum_{i=1}^N \boldsymbol{\epsilon}_i' \mathbf{M}_F \boldsymbol{\epsilon}_i \\ &= \frac{1}{NT} \sum_{i=1}^N \boldsymbol{\epsilon}_i' (\mathbf{M}_{\widehat{\mathbf{F}}^s} - \mathbf{M}_F) \boldsymbol{\epsilon}_i - \frac{1}{NT} \sum_{i=1}^N \boldsymbol{\lambda}_i' \overline{\mathbf{H}}^{-'} \mathbf{D}' \mathbf{M}_{\widehat{\mathbf{F}}^s} \boldsymbol{\epsilon}_i \\ &\quad - \frac{1}{NT} \sum_{i=1}^N \boldsymbol{\epsilon}_i' \mathbf{M}_{\widehat{\mathbf{F}}^s} d\overline{\mathbf{H}}^- \boldsymbol{\lambda}_i + \frac{1}{NT} \sum_{i=1}^N \boldsymbol{\lambda}_i' \overline{\mathbf{H}}^{-'} \mathbf{D}' \mathbf{M}_{\widehat{\mathbf{F}}^s} d\overline{\mathbf{H}}^- \boldsymbol{\lambda}_i. \end{aligned}$$

Since the rank condition is satisfied, $\mathbf{M}_F = \mathbf{M}_{\mathbf{F}\overline{\mathbf{H}}^{s'}}$. The fourth term on the right-hand side of the above expansion is $(NT)^{-1/2}$ times the order of \mathbf{K}_1 in the proof of Theorem 2.1. Hence, this term is $O_p(N^{-1})$, and so are second and the third terms. To find the order of the first term we write

$$\begin{aligned} \boldsymbol{\epsilon}_i' (\mathbf{M}_{\widehat{\mathbf{F}}^s} - \mathbf{M}_F) \boldsymbol{\epsilon}_i &= \boldsymbol{\epsilon}_i' d(\widehat{\mathbf{F}}^{s'} \widehat{\mathbf{F}}^s)^{-1} \mathbf{D}' \boldsymbol{\epsilon}_i + \boldsymbol{\epsilon}_i' d(\widehat{\mathbf{F}}^{s'} \widehat{\mathbf{F}}^s)^{-1} \overline{\mathbf{H}}^s \mathbf{F}' \boldsymbol{\epsilon}_i \\ &\quad + \boldsymbol{\epsilon}_i' \mathbf{F} \overline{\mathbf{H}}^{s'} (\widehat{\mathbf{F}}^{s'} \widehat{\mathbf{F}}^s)^{-1} \mathbf{D}' \boldsymbol{\epsilon}_i \\ &\quad + \boldsymbol{\epsilon}_i' \mathbf{F} \overline{\mathbf{H}}^{s'} [(\widehat{\mathbf{F}}^{s'} \widehat{\mathbf{F}}^s)^{-1} - (\overline{\mathbf{H}}^s \mathbf{F}' \mathbf{F} \overline{\mathbf{H}}^{s'})^{-1}] \overline{\mathbf{H}}^s \mathbf{F}' \boldsymbol{\epsilon}_i. \end{aligned}$$

Using Lemma 2.B.5, we know that $\boldsymbol{\epsilon}_i' \mathbf{D} = O_p(TN^{-1})$, and it is straightforward to show that $\mathbf{F}' \boldsymbol{\epsilon}_i = O_p(\sqrt{T})$. These results imply

$$\frac{1}{NT} \sum_{i=1}^N \boldsymbol{\epsilon}_i' (\mathbf{M}_{\widehat{\mathbf{F}}^s} - \mathbf{M}_F) \boldsymbol{\epsilon}_i = O_p(N^{-1}),$$

Using this result and the condition that $Ng(N) \rightarrow \infty$ it is obvious that (2.B.11) holds for the case with $s \geq k$. This completes the proof. \blacksquare

Chapter 3

A CCE Estimator for Dynamic Panel ECM's with Factors

3.1 Introduction

In the recent literature on panel data there has been an increased interest in the analysis of panel data models in which the assumption of cross-sectional independence is violated. An overview of the existing literature is provided by Sarafidis and Wansbeek (2012) and Chudik and Pesaran (2013d). The assumption of cross-sectional independence is very restrictive and difficult to justify for panels of long economic time series. In the case of cross-sectional dependence, the standard estimation methods do usually not produce consistent estimates of the coefficients of interest, and much effort has therefore gone recently into the development of methods that account for cross-sectional dependence. In particular, the use of a framework allowing for a reduced number of unobserved common factors causing the cross-sectional dependence has recently become very popular. The focus has mainly been on unobserved factors since observed factors can in principle easily be included in the model in question to overcome the problem of cross-section dependence.

One of the approaches which deals with the estimation and inference problems in cross-sectionally dependent panels with unobserved common factors is the Prin-

Principal Components approach, see Stock and Watson (2002) and Bai (2003). In a stationary world, they show that the PC estimates are consistent for the space spanned by the true factors instead of the factors themselves. An alternative is proposed by Pesaran (2006), in which the cross-sectional averages of the dependent variables and the observed regressors can be used to approximate the factor space, e.g. linear combinations of unobserved common factors. One of the advantages of this method is that, the number of unobserved factors needs not to be known beforehand. The estimators obtained using the cross-sectional averages as an approximation to the unobserved factors are called the CCE (Common Correlated Effects) estimators.

In the case of nonstationary panels with $I(1)$ variables, the literature on the estimation of non-stationary cointegrated dynamic panel models with unobserved common factors is still scarce. Notable exceptions are Kapetanios et al. (2011), Bai et al. (2009) and Bai and Kao (2006) that allow for $I(1)$ factors. As shown in Bai et al. (2009), the (pooled) least squares estimation of the (homogenous) cointegration parameters in a cross-sectionally dependent static framework requires specific corrections and modifications to ensure consistent estimation and nuisance parameter free inference. To solve this, they adopt a static triangular system framework and propose a continuously updated fully-modified estimator (CUP-FM), very much in the spirit of the standard fully modified estimator. In their set-up, the unobserved factors are estimated by principal components.

Alternatively, in the framework of Kapetanios et al. (2011), the $I(1)$ factors are approximated by cross-sectional averages and are used as additional regressors. This Common Correlated Effect (CCE) estimator extends the work of Pesaran (2006) by allowing the factors to be $I(1)$. In this chapter, we build on this branch of the literature and propose the estimation of long run parameters in dynamic, cointegrated heterogeneous panels with cross-sectional dependence. We consider an error correction model (ECM) representation for each cross-sectional unit in the panel. The error correction model allows us to model the short run dynamics explicitly. As pointed out in Gengenbach et al. (2013), this model representation for cointegrated panels can be used as an alternative to the triangular representation of Bai et al. (2009). It is also used in Gengenbach et al. (2014). Similarly to the latter paper, we adopt the CCE framework and use cross-sectional averages to approximate the unobserved, and potentially $I(1)$ common factors. This approach is also related to the recent work of Banerjee and Carrion-i Silvestre (2012), however, our set-up is dynamic and in this sense is much related to the work of

Pesaran et al. (1999) that we extend by allowing for cross-sectional dependence. Under suitable assumptions, most notably the weak exogeneity assumption, a single equation augmented heterogenous panel ECM provides us with a simple and attractive framework to estimate both the long run and short run parameters. A similar analysis is provided for the dynamic stationary case with weakly exogenous variables by Chudik and Pesaran (2013b).

The structure of the chapter is as follows. In Section 2, the model is described, and the necessary assumptions are presented. In the third section, we propose the estimators and their theoretical properties. In Section 4, some Monte Carlo results are presented to provide evidence on the small sample properties of the estimators. A final section concludes.

3.2 Model and Assumptions

We consider a panel where we observe a $(1 + m)$ -dimensional $I(1)$ vector, $\mathbf{z}_{it} = (y_{it}, \mathbf{x}'_{it})'$, which consists of a scalar variable y_{it} and an m -vector \mathbf{x}_{it} of idiosyncratic variables, for individuals $i = 1, \dots, N$ over time periods $t = -p, \dots, T$. y_{it} is the observation on the i^{th} cross-section unit at time t . Furthermore, we assume that we have a k -vector of $I(1)$ unobserved common factors \mathbf{f}_t . Denote the $(1 + m + k)$ vector $\mathbf{z}_{it}^* = (\mathbf{z}'_{it}, \mathbf{f}'_t)'$. For notational simplification we ignore deterministic trends (See Remark 3.7). For each unit i , the DGP is given by the following vector error correction model

$$\Delta \mathbf{z}_{it}^* = \mathbf{A}_i^* \mathbf{z}_{i,t-1}^* + \sum_{\ell=1}^{p-1} \mathbf{A}_{i\ell} \Delta \mathbf{z}_{i,t-\ell}^* + \boldsymbol{\varepsilon}_{it}^*. \quad (3.1)$$

We next introduce the assumptions imposed on the model above. We present these in three main groups. Assumption 3.1 is related to the dynamics and the cointegration structure in the model, Assumption 3.2 contains the restrictions related to the exogeneity properties of the model. Assumption 3.3 describes the nature of heterogeneity in the long-run and in the short-run.

Assumption 3.1.

- (i) The vector of errors, $\boldsymbol{\varepsilon}_{it}^* = (\varepsilon_{yit}, \boldsymbol{\varepsilon}'_{xit}, \boldsymbol{\varepsilon}'_{ft})'$ is i.i.d across t with zero mean and finite variance,

$$\boldsymbol{\Sigma}_i_{(1+m+k) \times (1+m+k)} = \begin{pmatrix} \sigma_{yyi} & \boldsymbol{\Sigma}_{yxi} & 0 \\ \boldsymbol{\Sigma}_{xyi} & \boldsymbol{\Sigma}_{xxi} & 0 \\ 0 & 0 & \boldsymbol{\Sigma}_{ff} \end{pmatrix}, \quad (3.2)$$

and has finite fourth order moment. Furthermore ε_{yit} and $\boldsymbol{\varepsilon}_{xit}$ are i.i.d across i .

- (ii) The roots of the determinant of the characteristic function,

$$\mathbf{F}(z) = (1 - z)\mathbf{I}_{(1+m+k)} - \mathbf{A}_i^* z - \sum_{\ell=1}^{p-1} \mathbf{A}_{i\ell}(1 - z)z^\ell$$

fall outside the unit circle for all $i = 1, \dots, N$.

- (iii) We assume that cointegration holds for each individual subsystem with common cointegration rank, $r_i = r = 1$ for $i = 1, \dots, N$; such that for all $i = 1, \dots, N$; \mathbf{A}_i^* can be expressed in terms of two vectors of rank 1:

$$\mathbf{A}_i^* = \boldsymbol{\alpha}_i \boldsymbol{\beta}'_i,$$

where $\boldsymbol{\alpha}_i$ and $\boldsymbol{\beta}_i$ are $(1 + m + k)$ -vectors.

- (iv) We normalize $\boldsymbol{\beta}_i$ such that the first element in $\boldsymbol{\beta}_i$ is equal to 1, which implies

$$\mathbf{A}_i^* = \boldsymbol{\alpha}_i \boldsymbol{\beta}'_i = \begin{pmatrix} \alpha_{yi} \\ \boldsymbol{\alpha}_{xi} \\ \boldsymbol{\alpha}_f \end{pmatrix} \begin{pmatrix} 1 - \boldsymbol{\beta}'_{xi} - \boldsymbol{\lambda}'_i \end{pmatrix}.$$

Assumption 3.2.

- (i) The idiosyncratic variables, \mathbf{x}_{it} are assumed to be weakly exogenous for the long-run parameters, so that $\boldsymbol{\alpha}_{xi} = 0$ for $i = 1, 2, \dots, N$.

(ii) The coefficient of the short-run dynamics for the ℓ^{th} lag order, $\mathbf{A}_{i\ell}$, can be written in a block specific form as

$$\mathbf{A}_{i\ell}^{(1+m+k) \times (1+m+k)} = \begin{pmatrix} A_{yyi\ell} & \mathbf{A}_{yxi\ell} & \mathbf{A}_{yfi\ell} \\ 1 \times 1 & 1 \times m & 1 \times k \\ \mathbf{A}_{xyi\ell} & \mathbf{A}_{xxi\ell} & \mathbf{A}_{xfi\ell} \\ m \times 1 & m \times m & m \times k \\ 0_{k \times 1} & 0_{k \times m} & \mathbf{A}_{ffi\ell} \\ & & k \times k \end{pmatrix}.$$

(iii) The latent factors are strongly exogenous for the long-run parameters, so that, in addition to (ii) we assume that $\boldsymbol{\alpha}_f = 0$.

Assumption 3.3.

(i) We assume that the error-correction coefficients α_{yi} are random, i.i.d. across i , independent of all other model parameters for all i and characterized by

$$\alpha_{yi} = \alpha_y + \eta_{yi}, \quad \eta_{yi} \sim IID(0, \omega_{\eta y}), \quad \omega_{\eta y} > 0 \quad \text{for } i = 1, 2, \dots, N,$$

where $\omega_{\eta y}$ is the variance of the random deviations and $\|\omega_{\eta y}\| < K$, for some constant K . The random deviations are distributed independently of the model parameters and variables. For each i , the support of the distribution of α_{yi} lies in the stable range $(-2, 0)$.

(ii) We assume that the long-run parameters β_{xi} for the idiosyncratic variables are random, i.i.d. across i , independent of all other model parameters for all i and characterized by

$$\beta_{xi} = \beta_x + \boldsymbol{\eta}_{xi}, \quad \boldsymbol{\eta}_{xi} \sim IID(0, \boldsymbol{\Omega}_{\eta x}), \quad \boldsymbol{\Omega}_{\eta x} > 0 \quad \text{for } i = 1, 2, \dots, N,$$

where $\boldsymbol{\Omega}_{\eta x}$ is the covariance matrix of the random deviations and the random deviations are distributed independently of the model parameters and variables.

(iii) The factor loadings in the long-run are random, i.i.d. across i and characterized by

$$\boldsymbol{\lambda}_i = \boldsymbol{\lambda} + \boldsymbol{\eta}_{fi}, \quad \boldsymbol{\eta}_{fi} \sim IID(0, \boldsymbol{\Omega}_{\eta f}), \quad \boldsymbol{\Omega}_{\eta f} > 0 \quad \text{for } i = 1, 2, \dots, N;$$

where $\mathbf{\Omega}_{\eta f}$ is the covariance matrix of the random deviations and the random deviations are distributed independently of the model parameters and variables.

- (iv) For all $i = 1, 2, \dots, N$; $\mathbf{A}_{zzi\ell}$ and $\mathbf{A}_{zfi\ell}$ are random coefficients, distributed uniformly bounded over i and have bounded variances for all $\ell = 1, \dots, p$; where

$$\mathbf{A}_{zzi\ell} = \begin{pmatrix} A_{yyi\ell} & \mathbf{A}_{yxi\ell} \\ \mathbf{A}_{xyi\ell} & \mathbf{A}_{xxi\ell} \end{pmatrix},$$

and where $\mathbf{A}_{zfi\ell} = \begin{pmatrix} \mathbf{A}_{yfi\ell} \\ \mathbf{A}_{xfi\ell} \end{pmatrix}$.

- (v) All factors are supposed to be strong - see Chudik et al. (2011). In particular, we assume that, for each cross-section unit i ,

$$\lim_{N \rightarrow \infty} N^{-1} \sum_{i=1}^N |\lambda_i| = K > 0, \quad \lim_{N \rightarrow \infty} N^{-1} \sum_{i=1}^N |\mathbf{A}_{zfi\ell}| = \mathbf{K} > 0.$$

for $\ell = 1, \dots, p-1$.

Under the assumptions stated in this section we can rewrite (3.1) as

$$\begin{aligned} \Delta y_{it} &= \alpha_{yi}(y_{i,t-1} - \beta'_{xi}\mathbf{x}_{i,t-1} - \lambda'_i\mathbf{f}_{t-1}) + \sum_{\ell=1}^{p-1} A_{yyi\ell} \Delta y_{i,t-\ell} + \sum_{\ell=1}^{p-1} \mathbf{A}_{yxi\ell} \Delta \mathbf{x}_{i,t-\ell} \\ &\quad + \sum_{\ell=1}^{p-1} \mathbf{A}_{yfi\ell} \Delta \mathbf{f}_{t-\ell} + \varepsilon_{yit}, \end{aligned} \quad (3.3)$$

$$\Delta \mathbf{x}_{it} = \sum_{\ell=1}^{p-1} \mathbf{A}_{xyi\ell} \Delta y_{i,t-\ell} + \sum_{\ell=1}^{p-1} \mathbf{A}_{xxi\ell} \Delta \mathbf{x}_{i,t-\ell} + \sum_{\ell=1}^{p-1} \mathbf{A}_{xfi\ell} \Delta \mathbf{f}_{t-\ell} + \varepsilon_{xit}, \quad (3.4)$$

$$\Delta \mathbf{f}_t = \sum_{\ell=1}^{p-1} \mathbf{A}_{ff\ell} \Delta \mathbf{f}_{t-\ell} + \varepsilon_{ft}. \quad (3.5)$$

A few remarks are in order to correctly interpret these assumptions.

Remark 3.1. Assumption 3.1 implies that the errors are restricted to be serially uncorrelated. Given the dynamic nature of our modelling approach this is easily achieved by increasing the lag order in the models presented above if necessary.

Also, without loss of generality we assume $p_i = p$, e.g. the same lag order for all unit in the panel.

Remark 3.2. Another implication of Assumption 3.1 is that a multivariate invariance principle holds for vector ε_{it}^* , such that

$$\frac{1}{\sqrt{T}} \sum_{t=1}^{[Tr]} \varepsilon_{it}^* \xrightarrow{w} \mathbf{B}_i(r) \quad \text{as } T \rightarrow \infty \quad \text{for all } i, \quad (3.6)$$

where $\mathbf{B}_i(r) = (B_{yi}(r), \mathbf{B}'_{xi}(r), \mathbf{B}'_f(r))'$ is a $1 + m + k$ -vector Brownian motion process with a covariance matrix Σ_i that is defined in (3.2). We should keep in mind the presence of $\mathbf{B}_f(r)$ implies that \mathbf{B}_i are correlated across i .

Remark 3.3. The assumed cross-sectional independence of ε_{it}^* and the form of Σ_i implies that \mathbf{f}_t is the only source of cross-sectional dependence. While this might seem restrictive, the model considered here is actually more general than almost all those considered in the existing literature. In our model \mathbf{f}_t is allowed to enter in both levels and first-differences, which means that the cross-sectional dependence can be both long-run and short-run. Also, in case where $\boldsymbol{\lambda}_i = 0$ so that the common factors do not enter the cointegrating relationships, our specification still allows for $I(0)$ factors in the form of $\Delta \mathbf{f}_{t-j}$. Note also that a similar structure to the one we impose by Assumptions 3.1(i) and 3.2(iii) is used by Bai et al. (2009).

Remark 3.4. All the results reported in this chapter are based on the important assumption that all the factors are of the strong form. One reason for this is the conceptual difficulty involved in interpreting a cointegrated relationship with weak factors. See also the discussion in Remark 1 of the supplement of Gengenbach et al. (2014).

Remark 3.5. As discussed in Gengenbach et al. (2013), the representation used in this chapter can actually be derived as an alternative to the triangular systems representation adopted by Bai et al. (2009). The main difference is that the short-run dynamics are explicitly treated parametrically. Similarly to their framework, the cointegrating rank is assumed to be one for all members of the panel. Furthermore, as in Bai et al. (2009), Banerjee and Carrion-i Silvestre (2012) and Bai and Carrion-i Silvestre (2009), we assume that the factors may be $I(1)$ and potentially enter the cointegrating relations so that the variables do not cointegrate alone without the factors.

Remark 3.6. The model for $\Delta \mathbf{x}_{it}$ does not include an error correction term. As a result of Johansen (1992) or Urbain (1992), this is a necessary and a sufficient condition for \mathbf{x}_{it} to be weakly exogenous for $\boldsymbol{\alpha}_i$ and $\boldsymbol{\beta}_i$. This implies that there is no information for the long-run relation in the model for \mathbf{x}_{it} . This in turn makes it sufficient to use the conditional model for Δy_{it} , to estimate the long-run relation in the system. Assuming that the conditioning variables are weakly exogeneous is a restriction compared to Bai et al. (2009). A similar assumption (while not made explicitly) is made by Kapetanios et al. (2011). Note that this assumption can also be tested using Lagrange multiplier type tests. Furthermore, the strong exogeneity assumption for the factors is a standard natural assumption.

Remark 3.7. We will ignore any kinds of deterministic components in the model for notational simplicity. While this may seem restrictive, it is possible and rather straightforward to assume the presence of deterministic components in, for example, the long-run relation and consider the various cases of interest such as in Boswijk (1995) or Gengenbach et al. (2014) for example. Besides notational complications, the implications are mainly at the level of the asymptotic results where one will have to replace the Brownian motions by suitable demeaned or detrended Brownian motions.

3.2.1 Conditional Model

Given Assumptions 3.1 and 3.2, it is natural to consider a conditional model. Using (3.3)-(3.5) and given the block specific elements of $\boldsymbol{\Sigma}_i$ by defining $\boldsymbol{\kappa}'_{xi} = \boldsymbol{\Sigma}_{yxi} \boldsymbol{\Sigma}_{xxi}^{-1}$, the conditional parameters can be written as

$$\begin{aligned} \mathbf{A}_{y.xi0} &= \boldsymbol{\kappa}'_{xi}, \\ A_{y.yi\ell} &= A_{yyi\ell} - \boldsymbol{\kappa}'_{xi} \mathbf{A}_{xyi\ell} \quad \text{for } \ell = 1, \dots, p-1, \\ \mathbf{A}_{y.xi\ell} &= \mathbf{A}_{yxi\ell} - \boldsymbol{\kappa}'_{xi} \mathbf{A}_{xxi\ell} \quad \text{for } \ell = 1, \dots, p-1, \\ \mathbf{A}_{y.fi\ell} &= \mathbf{A}_{yfi\ell} - \boldsymbol{\kappa}'_{xi} \mathbf{A}_{xfi\ell} \quad \text{for } \ell = 1, \dots, p-1. \end{aligned}$$

The conditional model for Δy_{it} given the past values of Δy_{it} current and past $\Delta \mathbf{x}_{it}$ s, past values of $\Delta \mathbf{f}_t$ until their $(p-1)^{th}$ lag and given $\mathbf{y}_{i,t-1}$ $\mathbf{x}_{i,t-1}$ and $\mathbf{f}_{i,t-1}$ can then be written as

$$\Delta y_{it} = \alpha_{yi}(y_{i,t-1} - \boldsymbol{\beta}'_{xi} \mathbf{x}_{i,t-1} - \boldsymbol{\lambda}'_i \mathbf{f}_{i,t-1}) + \sum_{\ell=1}^{p-1} \mathbf{A}_{y.yi\ell} \Delta y_{i,t-\ell} + \sum_{\ell=0}^{p-1} \mathbf{A}_{y.xi\ell} \Delta \mathbf{x}_{i,t-\ell}$$

$$+ \sum_{\ell=1}^{p-1} \mathbf{A}_{y,fi\ell} \Delta \mathbf{f}_{t-\ell} + \boldsymbol{\varepsilon}_{y,xit}, \quad (3.7)$$

where $\varepsilon_{y,xit} = \varepsilon_{yit} - \boldsymbol{\kappa}'_{xi} \boldsymbol{\varepsilon}_{xit}$.

The previously assumed properties of the disturbances are such that the following invariance principle applies:

$$\frac{1}{\sqrt{T}} \sum_{t=1}^{[Tr]} \begin{pmatrix} \varepsilon_{y,xit} \\ \boldsymbol{\varepsilon}_{xit} \\ \boldsymbol{\varepsilon}_{ft} \end{pmatrix} \xrightarrow{w} \mathbf{B}_i^*(r) = \begin{pmatrix} B_{y.xi}(r) \\ \mathbf{B}_{xi}(r) \\ \mathbf{B}_f(r) \end{pmatrix} \quad \text{as } T \rightarrow \infty \text{ for all } i, \quad (3.8)$$

where the Brownian motion, $\mathbf{B}_i^*(r)$, has a covariance matrix given by

$$\boldsymbol{\Sigma}_i^* = \begin{pmatrix} \sigma_{y.xi} & 0_{1 \times m} & 0_{1 \times k} \\ 0_{m \times 1} & \boldsymbol{\Sigma}_{xxi} & 0_{m \times k} \\ 0_{k \times 1} & 0_{k \times m} & \boldsymbol{\Sigma}_{ff} \end{pmatrix}, \quad (3.9)$$

where $\sigma_{y.xi} = \sigma_{yyi} - \boldsymbol{\Sigma}_{yxi} \boldsymbol{\Sigma}_{xxi}^{-1} \boldsymbol{\Sigma}_{xyi}$. Note the difference in the covariance structure between (3.2) and (3.9). However, $\mathbf{B}_i^*(r)$ is still correlated across i given the presence of $\mathbf{B}_f(r)$.

3.3 Estimation

In this section we discuss the estimation by distinguishing between two cases. In the first case we assume that the factors are observed. Under the assumption of weak exogeneity of \mathbf{x}_{it} and the strong exogeneity of the common factors, the long-run parameters can be estimated by using the conditional model (3.7). Using the conditional model without additional augmentation we obtain the infeasible estimators and give the asymptotic properties of the estimators. We provide the asymptotic properties of the estimators for the long-run and the short-run coefficients including the coefficients of the error correction term, for the individual specific parameters. In the second case we go back to the assumption of unobserved factors. First we provide a method to approximate the unobserved factors. Then the approximations are used to write the augmented conditional model. Finally, we discuss the asymptotic properties of the short-run and the long-run coefficient estimators for the individual specific parameters and asymptotic properties of the mean-group estimator for the long-run coefficient. The proofs of the results presented in this section can be found in the Appendix.

By stacking the observations over time, we write the conditional model (3.7) in matrix notation as

$$\Delta \mathbf{y}_i = (\mathbf{y}_{i,-1} - \mathbf{X}_{i,-1} \boldsymbol{\beta}_{xi} - \mathbf{F}_{-1} \boldsymbol{\lambda}_i) \alpha_{yi} + \mathbf{G}_{yi} \boldsymbol{\Gamma}_{y.yi} + \mathbf{G}_{xi} \boldsymbol{\Gamma}_{y.xi} + \mathbf{G}_f \boldsymbol{\Gamma}_{y.fi} + \boldsymbol{\varepsilon}_{y.xi}, \quad (3.10)$$

where

$$\begin{aligned} \Delta \mathbf{y}_i &= \begin{pmatrix} \Delta y_{i1} \\ \vdots \\ \Delta y_{iT} \end{pmatrix}_{T \times 1}, \quad \mathbf{y}_{i,-1} = \begin{pmatrix} y_{i0} \\ \vdots \\ y_{i,T-1} \end{pmatrix}_{T \times 1}, \quad \mathbf{X}_{i,-1} = \begin{pmatrix} \mathbf{x}'_{i0} \\ \vdots \\ \mathbf{x}'_{i,T-1} \end{pmatrix}_{T \times m}, \\ \mathbf{F}_{-1} &= \begin{pmatrix} \mathbf{f}'_0 \\ \vdots \\ \mathbf{f}'_{T-1} \end{pmatrix}_{T \times k}, \quad \boldsymbol{\varepsilon}_{y.xi} = \begin{pmatrix} \varepsilon_{y.xi1} \\ \vdots \\ \varepsilon_{y.xiT} \end{pmatrix}_{T \times 1}. \end{aligned}$$

For the current and lagged values of the first differences we have

$$\mathbf{G}_{yi} = \begin{pmatrix} \mathbf{g}'_{yi1} \\ \vdots \\ \mathbf{g}'_{yiT} \end{pmatrix}_{T \times (p-1)}, \quad \mathbf{G}_{xi} = \begin{pmatrix} \mathbf{g}'_{xi1} \\ \vdots \\ \mathbf{g}'_{xiT} \end{pmatrix}_{T \times pm}, \quad \mathbf{G}_f = \begin{pmatrix} \mathbf{g}'_{f1} \\ \vdots \\ \mathbf{g}'_{fT} \end{pmatrix}_{T \times (p-1)k},$$

with

$$\mathbf{g}_{yit} = \begin{pmatrix} \Delta y_{i,t-1} \\ \vdots \\ \Delta y_{i,t-p+1} \end{pmatrix}_{(p-1) \times 1}, \quad \mathbf{g}_{xit} = \begin{pmatrix} \Delta \mathbf{x}_{it} \\ \vdots \\ \Delta \mathbf{x}_{i,t-p+1} \end{pmatrix}_{mp \times 1}, \quad \mathbf{g}_{ft} = \begin{pmatrix} \Delta \mathbf{f}_{t-1} \\ \vdots \\ \Delta \mathbf{f}_{t-p+1} \end{pmatrix}_{k(p-1) \times 1}.$$

Note that \mathbf{g}_{xit} includes the current values $\Delta \mathbf{x}_{it}$. Furthermore the coefficients in the model are defined as follows: $\boldsymbol{\Gamma}_{y.yi} = (A_{y.yi1}, \dots, A_{y.yi,p-1})'$ is a $(p-1) \times 1$ vector, $\boldsymbol{\Gamma}_{y.xi} = (\mathbf{A}_{y.xi0}, \dots, \mathbf{A}_{y.xi,p-1})'$ is a $pm \times 1$ vector and $\boldsymbol{\Gamma}_{y.fi} = (\mathbf{A}_{y.fi1}, \dots, \mathbf{A}_{y.fi,p-1})'$ is a $(p-1)k \times 1$ vector.

Defining the lagged levels of the non-stationary variables entering into the error correction term as $\mathbf{Z}_{i,-1}^* = (\mathbf{y}_{i,-1}, \mathbf{X}_{i,-1}, \mathbf{F}_{-1})$, which is a $T \times (1 + m + k)$ matrix, the unrestricted conditional error correction model can be rewritten as

$$\Delta \mathbf{y}_i = \mathbf{Z}_{i,-1}^* \boldsymbol{\Pi}_i + \mathbf{G}_i \boldsymbol{\Gamma}_i + \boldsymbol{\varepsilon}_{y.xi}, \quad (3.11)$$

where

$$\mathbf{\Pi}_i = \begin{pmatrix} \alpha_{yi} \\ \mathbf{\Pi}_{xi} \\ \mathbf{\Pi}_{fi} \end{pmatrix} = \begin{pmatrix} \alpha_{yi} \\ -\boldsymbol{\beta}_{xi}\alpha_{yi} \\ -\lambda_i\alpha_{yi} \end{pmatrix},$$

and $\mathbf{G}_i = (\mathbf{G}_{yi}, \mathbf{G}_{xi}, \mathbf{G}_f)$ is a $T \times q$ matrix where $q = (p-1)(1+k) + mp$ and $\boldsymbol{\Gamma}_i = (\boldsymbol{\Gamma}'_{y.yi}, \boldsymbol{\Gamma}'_{y.xi}, \boldsymbol{\Gamma}'_{y.fi})'$ is a $q \times 1$ vector.

3.3.1 Observed Factors

In the presence of observed factors the problem boils down to the standard time series case for which the estimation of the short-run and the long-run parameters of the error correction models and/or autoregressive distributed lag models is a standard problem. For example, Pesaran and Shin (1995) provides a cointegration analysis for ADL models whereas Boswijk (1995) provides an estimation and inference analysis for structural error correction models. In this section we follow the latter and adapt the existing literature to our model and notation. We propose to estimate the short-run parameters by OLS and long-run coefficients by indirect least squares method as proposed by Boswijk (1995). The proofs of the theorems provided in this section can be found in Appendix 3.B.1.

Assuming that the factors explaining the cross-sectional dependence in the model are observed, the unrestricted conditional model in (3.11) can be estimated using the OLS method. Denoting the OLS estimators of $\mathbf{\Pi}_i$ and $\boldsymbol{\Gamma}_i$ by $\widehat{\mathbf{\Pi}}_i$ and $\widehat{\boldsymbol{\Gamma}}_i$ respectively, the OLS estimator for the coefficients of the model in (3.11) can simply be written as

$$\begin{pmatrix} \widehat{\mathbf{\Pi}}_i \\ \widehat{\boldsymbol{\Gamma}}_i \end{pmatrix} = \begin{pmatrix} \mathbf{Z}'_{i,-1}\mathbf{Z}^*_{i,-1} & \mathbf{Z}'_{i,-1}\mathbf{G}_i \\ \mathbf{G}'_i\mathbf{Z}^*_{i,-1} & \mathbf{G}'_i\mathbf{G}_i \end{pmatrix}^{-1} \begin{pmatrix} \mathbf{Z}'_{i,-1}\Delta\mathbf{y}_i \\ \mathbf{G}'_i\Delta\mathbf{y}_i \end{pmatrix}. \quad (3.12)$$

The long-run coefficients of the model in (3.10) can be estimated using indirect least squares method, by

$$\begin{pmatrix} \widehat{\boldsymbol{\beta}}_{xi} \\ \widehat{\lambda}_i \end{pmatrix} = -\frac{1}{\widehat{\alpha}_{yi}} \left(\mathbf{R}'_{i,-1}\mathbf{M}_i\mathbf{R}_{i,-1} \right)^{-1} \left(\mathbf{R}'_{i,-1}\mathbf{M}_i(\Delta\mathbf{y}_i - \mathbf{y}_{i,-1}\widehat{\alpha}_{yi}) \right), \quad (3.13)$$

where $\mathbf{R}_i = (\mathbf{X}_i, \mathbf{F})$, $\mathbf{M}_i = \mathbf{I} - \mathbf{G}_i(\mathbf{G}'_i\mathbf{G}_i)^{-1}\mathbf{G}'_i$. The asymptotic properties of the estimators stated in (3.12) and (3.13) are summarized in the theorems below. The

theorems provided below follow from standard results in the relevant literature, see for example Chapter 7 of Lütkepohl (2007).

Theorem 3.1. *Suppose that \mathbf{z}_{it}^* for $i = 1, \dots, N$ and $t = 1, \dots, T$ are generated by (3.3)-(3.5), that Assumptions 3.1, 3.2 and 3.3(i), (iv), (v) hold and consider the model form in (3.11). Let $T \rightarrow \infty$, then $\widehat{\boldsymbol{\Pi}}_i$ and $\widehat{\boldsymbol{\Gamma}}_i$ defined in (3.12) are consistent estimators of $\boldsymbol{\Pi}_i$ and $\boldsymbol{\Gamma}_i$ respectively and their asymptotic distribution is given by*

$$\sqrt{T} \begin{pmatrix} \widehat{\boldsymbol{\Pi}}_i - \boldsymbol{\Pi}_i \\ \widehat{\boldsymbol{\Gamma}}_i - \boldsymbol{\Gamma}_i \end{pmatrix} \xrightarrow{d} N \left(0, \begin{pmatrix} \boldsymbol{\beta}_i & 0 \\ 0 & \mathbf{I}_q \end{pmatrix} \boldsymbol{\Psi}_i^{*-1} \begin{pmatrix} \boldsymbol{\beta}_i & 0 \\ 0 & \mathbf{I}_q \end{pmatrix}' \sigma_{y.xi} \right),$$

where

$$\begin{pmatrix} T^{-1} \mathbf{U}'_{i,-1} \mathbf{U}_{i,-1} & T^{-1} \mathbf{U}'_{i,-1} \mathbf{G}_i \\ T^{-1} \mathbf{G}'_i \mathbf{U}_{i,-1} & T^{-1} \mathbf{G}'_i \mathbf{G}_i \end{pmatrix} \xrightarrow{p} \boldsymbol{\Psi}_i^*$$

with $\mathbf{U}_{i,-1} = \mathbf{Z}_{i,-1}^* \boldsymbol{\beta}_i$ and $\sigma_{y.xi}$ is defined subsequent to (3.9). Furthermore,

$$\begin{pmatrix} \boldsymbol{\beta}_i & 0 \\ 0 & \mathbf{I}_q \end{pmatrix} \boldsymbol{\Psi}_i^{*-1} \begin{pmatrix} \boldsymbol{\beta}_i & 0 \\ 0 & \mathbf{I}_q \end{pmatrix}',$$

can consistently be estimated by

$$T^{-1} \begin{pmatrix} \mathbf{Z}'_{i,-1} \mathbf{Z}_{i,-1}^* & \mathbf{Z}'_{i,-1} \mathbf{G}_i \\ \mathbf{G}'_i \mathbf{Z}_{i,-1}^* & \mathbf{G}'_i \mathbf{G}_i \end{pmatrix}^{-1}.$$

And to consistently estimate the residual variance one can use

$$\hat{\sigma}_{y.xi} = \frac{1}{T-1-m-k-q} (\Delta \mathbf{y}_i - \mathbf{Z}_{i,-1}^* \widehat{\boldsymbol{\Pi}}_i - \mathbf{G}_i \widehat{\boldsymbol{\Gamma}}_i)' (\Delta \mathbf{y}_i - \mathbf{Z}_{i,-1}^* \widehat{\boldsymbol{\Pi}}_i - \mathbf{G}_i \widehat{\boldsymbol{\Gamma}}_i).$$

Theorem 3.2. *Suppose that \mathbf{z}_{it}^* for $i = 1, \dots, N$ and $t = 1, \dots, T$ are generated by (3.3)-(3.5), that Assumptions 3.1, 3.2 and 3.3(i), (iv), (v) hold and consider the conditional error correction model in (3.10). Let $T \rightarrow \infty$, then the estimators $\widehat{\boldsymbol{\beta}}_{xi}$ and $\widehat{\boldsymbol{\lambda}}_i$ defined by (3.13) are consistent estimators of $\boldsymbol{\beta}_{xi}$ and $\boldsymbol{\lambda}_i$ respectively and their asymptotic distribution is given by*

$$T \begin{pmatrix} \widehat{\boldsymbol{\beta}}_{xi} - \boldsymbol{\beta}_{xi} \\ \widehat{\boldsymbol{\lambda}}_i - \boldsymbol{\lambda}_i \end{pmatrix}$$

$$\xrightarrow{d} -\frac{1}{\alpha_{yi}} \begin{bmatrix} \int_0^1 \mathbf{C}_{xzi} \mathbf{B}_{bi}(r) \mathbf{B}'_{bi}(r) dr \mathbf{C}'_{xzi} & \int_0^1 \mathbf{C}_{xzi} \mathbf{B}_{bi}(r) \mathbf{B}'_f(r) dr \mathbf{C}'_{ff} \\ \int_0^1 \mathbf{C}_{ff} \mathbf{B}_f(r) \mathbf{B}'_{bi}(r) dr \mathbf{C}'_{xzi} & \int_0^1 \mathbf{C}_{ff} \mathbf{B}_f(r) \mathbf{B}'_f(r) dr \mathbf{C}'_{ff} \end{bmatrix}^{-1} \\ \times \begin{bmatrix} \int_0^1 \mathbf{C}_{xzi} \mathbf{B}_{bi}(r) dB'_{y.xi}(r) \\ \int_0^1 \mathbf{C}_{ff} \mathbf{B}_f(r) dB'_{y.xi}(r) \end{bmatrix}.$$

where $\mathbf{B}_{bi}(r) = (\mathbf{B}'_{xi}(r), \mathbf{B}'_f(r))'$ and $B_{y.xi}(r)$, $\mathbf{B}_{xi}(r)$, $\mathbf{B}_f(r)$ are Brownian motions defined in (3.8), where $\mathbf{B}_i^*(r) = (B_{y.xi}(r), \mathbf{B}_{xi}(r)', \mathbf{B}_f(r)')$ has a covariance matrix defined in (3.9) and \mathbf{C}_{xzi} , \mathbf{C}_{ff} are defined in Appendix 3.B.1, in (3.B.22) and (3.B.4) respectively. Furthermore, following Bårdsen (1989), the variance of the long-run coefficients can consistently be estimated by

$$\widehat{\mathbf{V}} \begin{bmatrix} \widehat{\boldsymbol{\beta}}_i \\ \widehat{\boldsymbol{\lambda}}_i \end{bmatrix} = \frac{1}{\widehat{\alpha}_{yi}^2} \begin{pmatrix} \widehat{\boldsymbol{\beta}}_i & \mathbf{I}_m & 0 \\ \widehat{\boldsymbol{\lambda}}_i & 0 & \mathbf{I}_k \end{pmatrix} \widehat{\mathbf{V}}[\widehat{\boldsymbol{\Pi}}_i] \begin{pmatrix} \widehat{\boldsymbol{\beta}}_i' \\ \widehat{\boldsymbol{\lambda}}_i' \\ 0 & \mathbf{I}_k \end{pmatrix},$$

where $\widehat{\mathbf{V}}[\widehat{\boldsymbol{\Pi}}_i]$ is the estimated variance of the estimates of the unrestricted coefficients, $\boldsymbol{\Pi}_i$.

Note that to obtain the results in Theorems 3.1 and 3.2 the cross-section dimension is not used.

Remark 3.8. Given the weak (and strong for the factors) exogeneity of the conditioning variables, $\mathbf{B}_{bi}(r)$ and $B_{y.xi}(r)$ are independent so that the limiting distribution above is mixed normal, see for example Boswijk (1995) or Davidson (2000), Chapter 8. The asymptotic distribution of the estimator can consequently be stated as

$$T \begin{pmatrix} \mathbf{X}'_{i,-1} \mathbf{X}_{i,-1} & \mathbf{X}'_{i,-1} \mathbf{F}_{-1} \\ \mathbf{F}'_{-1} \mathbf{X}_{i,-1} & \mathbf{F}'_{-1} \mathbf{F}_{-1} \end{pmatrix}^{1/2} \begin{pmatrix} \widehat{\boldsymbol{\beta}}_i - \boldsymbol{\beta}_i \\ \widehat{\boldsymbol{\lambda}}_i - \boldsymbol{\lambda}_i \end{pmatrix} \xrightarrow{d} N(0, \alpha_{yi}^{-2} \sigma_{y.xi} \otimes \mathbf{I}_{m+k}).$$

Remark 3.9. The approximation for the estimation of the variance of the long-run coefficients that is proposed in Bårdsen (1989) and is proposed in Boswijk (1995) in a very similar set-up where $N = 1$. The usefulness of the formula comes from the fact that the OLS estimation provides all the necessary estimators of the parameters to be used in the estimation of the variance.

3.3.2 Unobserved Factors

The consistent estimation of the cointegrating vector in the presence of unobserved factors is possible only if the unobserved factors are approximated. There are two

widely used methods in the literature to approximate the unobserved common factors, namely the principal components and common correlated effects approaches. The principal components method (PC) is proposed by Bai et al. (2009) for the cointegration case where the latent factors are unit root non-stationary. They propose to use an iterative procedure to jointly estimate the coefficients of the idiosyncratic variables and an approximation for the unobserved factors. They employ a static cointegrated panel model in their paper. The alternative is the Common Correlated Effects method proposed by Pesaran (2006) for stationary data and extended to the non-stationary case by Kapetanios et al. (2011). In a stationary world, the cross-sectional averages can be used to approximate the unobserved common factors and as $N \rightarrow \infty$ the approximation errors become negligible. We opt for the CCE method to approximate the unobserved factors. Following Pesaran (2006), we first obtain an approximation for the levels and for the first differences of the unobserved factors by using the cross-sectional averages of the observed variables and of their first differences and show that the approximation errors are asymptotically negligible as $N \rightarrow \infty$ and under the additional condition that $\frac{\sqrt{T}}{\sqrt{N}} \rightarrow 0$ as $T, N \rightarrow \infty$. Note that we will express the relation between the unobserved factors and the approximations in terms of the model parameters and variables. As it is well-known, in this set-up only the factor space can be identified so that the unobserved factors can only be estimated up to a non-singular (rotation) matrix. We therefore also provide the detailed expressions for the rotation matrices. The approach used in this section shows that a proper approximation for the levels is possible only if one uses the current and lagged values of the first differences besides the levels of the cross-sectional averages of the observables, for a similar result in the dynamic stationary case see Chudik and Pesaran (2013b). As for the approximation for the first differences of the factors, using current and lagged values of the cross-sectional averages of the first differences of the observables as an approximation suffices. We also provide the resulting rates of convergence of the approximation errors.

Approximating the unobserved factors

We use the common trends representation to obtain an approximation for the unobserved factors. The motivation behind this is that the common trends representation allows us to track the rank reduction in the model and invertibility of the matrices used in the analysis. The common trends representation is derived in Section 3.B.1, in Appendix 3.B.1 in detail. For our purposes in this section we

start with writing the common trends representation for the observed variables and the unobserved common factors separately. For a given cross-section i , by ignoring the initial values the common trends representation for the DGP given by equations (3.3)-(3.5), can be written as

$$\mathbf{z}_{it} = \mathbf{C}_{zzi} \sum_{s=1}^t \boldsymbol{\varepsilon}_{zis} + \mathbf{C}_{zfi} \sum_{s=1}^t \boldsymbol{\varepsilon}_{fs} + \mathbf{C}_{zzi}^*(L) \boldsymbol{\varepsilon}_{zit} + \mathbf{C}_{zfi}^*(L) \boldsymbol{\varepsilon}_{ft}, \quad (3.14)$$

$$\mathbf{f}_t = \mathbf{C}_{ff} \sum_{s=1}^t \boldsymbol{\varepsilon}_{fs} + \mathbf{C}_{ff}^*(L) \boldsymbol{\varepsilon}_{ft}, \quad (3.15)$$

where $\mathbf{z}_{it} = (y_{it}, \mathbf{x}'_{it})'$, $\boldsymbol{\varepsilon}_{zit} = (\varepsilon_{yit}, \boldsymbol{\varepsilon}'_{xit})'$ and where $\mathbf{C}_{zzi} = \begin{pmatrix} C_{yyi} & C_{yxi} \\ C_{xyi} & C_{xxi} \end{pmatrix}$ and $\mathbf{C}_{zfi} = (\mathbf{C}'_{yfi}, \mathbf{C}'_{xfi})'$. The specific expressions for the blocks of \mathbf{C}_{zzi} and \mathbf{C}_{zfi} are provided in Section 3.B.1, in Appendix 3.B.1, which shows that $(C_{yyi}, \mathbf{C}'_{xyi})' = 0$, which is an important implication of the weak exogeneity of the idiosyncratic variables for the long-run coefficients. The infinite order matrix lag-polynomials $\mathbf{C}_{zzi}^*(L) = \begin{pmatrix} C_{yyi}^*(L) & C_{yxi}^*(L) \\ C_{xyi}^*(L) & C_{xxi}^*(L) \end{pmatrix}$ and $\mathbf{C}_{zfi}^*(L) = (\mathbf{C}_{yfi}^{*'}(L), \mathbf{C}_{xfi}^{*'}(L))'$, $\mathbf{C}_{ff}^*(L)$ are also determined by the model parameters. They are the coefficients matrices of the stationary component of \mathbf{z}_{it}^* .

Before presenting the results regarding the approximation, we state the necessary assumptions we impose on specific functions of the model parameters. For this purpose we first define the following matrix and polynomial matrix

$$\mathbf{H}'_1 = \overline{\mathbf{C}}_{zf} \mathbf{C}_{ff}^{-1}, \quad (3.16)$$

$$\mathbf{H}'_4(L) = (\overline{\mathbf{C}}_{zf} + \overline{\mathbf{C}}_{zf}^*(L)(1-L))(\mathbf{C}_{ff} + \mathbf{C}_{ff}^*(L)(1-L))^{-1}, \quad (3.17)$$

where $\overline{\mathbf{C}}_{zf} = \frac{1}{N} \sum_{i=1}^N \mathbf{C}_{zfi}$, $\overline{\mathbf{C}}_{zf}^*(L) = \frac{1}{N} \sum_{i=1}^N \mathbf{C}_{zfi}^*(L)$. As it is apparent from (3.17), $\mathbf{H}'_4(L)$ is an infinite order lag-polynomial.

We make the following assumption on the rank of these matrices.¹

Assumption 3.4.

(i) $\mathbf{H}'_1 : (m+1) \times k$ has a full column rank.

(ii) The infinite order lag-polynomial $\mathbf{H}'_4(L) : (m+1) \times k$ has a full column rank.

¹ For a definition of the rank of a polynomial matrix see Lütkepohl (1996), Section 11.3.

Remark 3.10. Notice that Assumption 3.4 implies that $k \leq m + 1$ which is the standard rank assumption that is usually made in CCE type of approaches, namely that the number of factors is at most equal to the number of observable variables.²

Let us first consider the approximation for $\Delta \mathbf{f}_t$. Using (3.14) and (3.15) and taking the first differences of both sides of the equations yields

$$\Delta \mathbf{z}_{it} = (\mathbf{C}_{zzi} + \mathbf{C}_{zzi}^*(L)(1-L))\boldsymbol{\varepsilon}_{zit} + (\mathbf{C}_{zfi} + \mathbf{C}_{zfi}^*(L)(1-L))\boldsymbol{\varepsilon}_{ft}, \quad (3.18)$$

$$\Delta \mathbf{f}_t = (\mathbf{C}_{ff} + \mathbf{C}_{ff}^*(L)(1-L))\boldsymbol{\varepsilon}_{ft}. \quad (3.19)$$

Taking the cross-sectional average of (3.18), we obtain

$$\overline{\Delta \mathbf{z}}_t = (\overline{\mathbf{C}}_{zf} + \overline{\mathbf{C}}_{zf}^*(L)(1-L))\boldsymbol{\varepsilon}_{ft} + \frac{1}{N} \sum_{i=1}^N (\mathbf{C}_{zzi} + \mathbf{C}_{zzi}^*(L)(1-L))\boldsymbol{\varepsilon}_{zit}. \quad (3.20)$$

By substituting the expression for $\boldsymbol{\varepsilon}_{ft}$, which can be obtained by rearranging (3.19), we get

$$\begin{aligned} \overline{\Delta \mathbf{z}}_t &= (\overline{\mathbf{C}}_{zf} + \overline{\mathbf{C}}_{zf}^*(L)(1-L))(\mathbf{C}_{ff} + \mathbf{C}_{ff}^*(L)(1-L))^{-1} \Delta \mathbf{f}_t \\ &\quad + \frac{1}{N} \sum_{i=1}^N (\mathbf{C}_{zzi} + \mathbf{C}_{zzi}^*(L)(1-L))\boldsymbol{\varepsilon}_{zit}, \end{aligned} \quad (3.21)$$

where the last equation is obtained by noting that \mathbf{C}_{ff} has full rank k (See Assumption 3.2(iii)) since, by assumption, the latent factors are not cointegrated alone. We can rewrite the approximation error as

$$\overline{\Delta \mathbf{z}}_t - \mathbf{H}'_4(L)\Delta \mathbf{f}_t = \frac{1}{N} \sum_{i=1}^N (\mathbf{C}_{zzi} + \mathbf{C}_{zzi}^*(L)(1-L))\boldsymbol{\varepsilon}_{zit} = \mathbf{v}_{Bt}, \quad (3.22)$$

where $\mathbf{H}'_4(L)$ is defined in (3.17). Thus, \mathbf{v}_{Bt} is the approximation error for the first differences of the latent factors when they are approximated by the current and lagged values of the first differences of the cross-sectional averages of the observables. The nature of the rotation matrix implies that for a proper approximation one should use the current and lagged values of the approximations and requires the imposition of further conditions on the rotation matrix. This is discussed

²In the static, stationary case, this assumption may be violated provided some additional assumptions are made, for more details see Karabiyik et al. (2014).

later in the chapter. Note that similar "polynomial rotation matrices" are used in Chudik and Pesaran (2013b).

Moreover, we have

$$\mathbf{v}_{Bt} = O_p(N^{-1/2}),$$

which is a result stated and proved in Lemma 3.B.4 in Section 3.B.2 of Appendix 3.B.2.

In order to obtain an approximation for the levels of the unobserved factors we start with taking the cross-sectional average³ of both sides of (3.14), thus we have

$$\bar{\mathbf{z}}_t = \frac{1}{N} \sum_{i=1}^N \mathbf{C}_{zzi} \sum_{s=1}^t \boldsymbol{\varepsilon}_{zis} + \bar{\mathbf{C}}_{zf} \sum_{s=1}^t \boldsymbol{\varepsilon}_{fs} + \frac{1}{N} \sum_{i=1}^N \mathbf{C}_{zzi}^*(L) \boldsymbol{\varepsilon}_{zit} + \bar{\mathbf{C}}_{zf}^*(L) \boldsymbol{\varepsilon}_{ft}. \quad (3.23)$$

Rearranging the equation and making necessary substitutions (See Section 3.B.2 of Appendix 3.B.2 for details), we obtain

$$\bar{\mathbf{z}}_t - \mathbf{H}'_1 \mathbf{f}_t = \mathbf{H}_3(L) \bar{\Delta} \bar{\mathbf{z}}_t + \mathbf{v}_{At}, \quad (3.24)$$

where \mathbf{H}'_1 is defined in (3.16), $\mathbf{H}'_3(L)$ is an infinite order lag-polynomial matrix defined in (3.B.39) and

$$\begin{aligned} \mathbf{v}_{At} &= \frac{1}{N} \sum_{i=1}^N \mathbf{C}_{zzi} \sum_{s=1}^t \boldsymbol{\varepsilon}_{zis} \\ &\quad + \frac{1}{N} \sum_{i=1}^N (\mathbf{C}_{zzi}^*(L) + \mathbf{H}_3(L) (\mathbf{C}_{zzi} + \mathbf{C}_{zzi}^*(L)(1-L))) \boldsymbol{\varepsilon}_{zit}. \end{aligned} \quad (3.25)$$

\mathbf{v}_{At} is the approximation error for the levels of the latent factors when they are approximated by the cross-sectional averages of the levels of the observables and corrected for the short run effects by including current and lagged values of the cross-sectional averages of the first differences of the observables. The order of the approximation error can be stated formally as

$$\mathbf{v}_{At} = O_p(T^{1/2} N^{-1/2}),$$

³An alternative would be to simply compute the partial sum of approximation for $\Delta \mathbf{f}_t$. Both lead to the same outcome, but the derivation below clarifies the originale of each term.

which is a result proved in Lemma 3.B.5. Furthermore, it can be shown that the approximation error becomes asymptotically $I(0)$ as $N \rightarrow \infty$ under the condition that $\frac{\sqrt{T}}{\sqrt{N}} \rightarrow 0$.

Remark 3.11. The Granger representation theorem suggests that for any $i \in \{1, \dots, N\}$,

$$\mathbf{C}_i_{(1+m+k) \times (1+m+k)} = \begin{pmatrix} C_{yyi} & C_{yxi} & C_{yfi} \\ \mathbf{C}_{xyi} & \mathbf{C}_{xxi} & \mathbf{C}_{xfi} \\ 0 & 0 & \mathbf{C}_{ffi} \end{pmatrix},$$

has a reduced rank, $m+k$ in our set-up. Under our assumptions, the rank reduction comes from the first column of \mathbf{C}_i , i.e. C_{yyi} and \mathbf{C}_{xyi} are 0. This implies that \mathbf{C}_{zfi} has a full column rank. Assumption 3.4(i) is still required since \mathbf{C}_{zfi} being full column rank for all i does not imply that the average over cross-sectional units $\overline{\mathbf{C}}_{zf}$ is full column rank.

Remark 3.12. In order for the approximation method proposed in this section to work with standard matrix inverses, certain matrices and lag polynomials and functions of lag polynomial matrices, namely \mathbf{H}_1 , $\mathbf{H}_4(L)$, which are $k \times (1+m)$ matrices, and $(\overline{\mathbf{C}}_{zf} + \overline{\mathbf{C}}_{zf}^*(L)(1-L))$, with size $(1+m) \times k$, must be square and non-singular. However, they are non-square matrices unless $m+1 = k$, which is a very restrictive assumption since it implies that the number of observables should be equal to the number of latent factors. Nevertheless, a particular form of generalized inverses, namely the right inverse for \mathbf{H}_1 , $\mathbf{H}_4(L)$ and the left inverse for $(\overline{\mathbf{C}}_{zf} + \overline{\mathbf{C}}_{zf}^*(L)(1-L))$ can be used as long as certain matrices involved in the the expressions have full column ranks, where the left inverse and the right inverse are defined in the paragraph devoted to the notation, in Section 1.5. This condition is satisfied under Assumption (3.4).

Remark 3.13. $\mathbf{H}_3(L)$ and $\mathbf{H}_4(L)$, which are defined in (3.B.39) and (3.17) respectively, are infinite order lag polynomials. This infiniteness stems from the inversion of lag polynomial $(\overline{\mathbf{C}}_{zf} + \overline{\mathbf{C}}_{zf}^*(L)(1-L))$ for $\mathbf{H}_3(L)$ and $(\mathbf{C}_{ff} + \mathbf{C}_{ff}^*(L)(1-L))$ for $\mathbf{H}_4(L)$. In practice, the infinite lag length needs to be truncated to make the estimation feasible while augmenting the regression by the first differences of the cross-sectional averages. As shown later, the truncation does not affect the asymptotic properties of the proposed estimator provided certain conditions on the growth rate of the truncation lag are met. We discuss this issue further in Section 3.3.2.

Augmented model

In order to augment the model with the factor approximations obtained in the previous section, we consider the model with matrix polynomials notation. The model with true factors is given by

$$\begin{aligned}\Delta \mathbf{y}_i &= (\mathbf{y}_{i,-1} - \mathbf{X}_{i,-1}\boldsymbol{\beta}_{xi} - \mathbf{F}_{-1}\boldsymbol{\lambda}_i)\alpha_{yi} + \Delta \mathbf{y}_{i,-1}\mathbf{A}'_{y,yi}(L) + \Delta \mathbf{X}_i\mathbf{A}'_{y,xi}(L) \\ &\quad + \Delta \mathbf{F}_{-1}\mathbf{A}'_{y,fi}(L) + \boldsymbol{\varepsilon}_{y,xi},\end{aligned}$$

where the definitions of the lag-polynomials are implicit from (3.7). We next replace the true factors in the model above with the approximations obtained in the previous section. Rearranging (3.22) and (3.24), stacking the t dimension and substituting the resulting expressions into the model above yields

$$\begin{aligned}\Delta \mathbf{y}_i &= (\mathbf{y}_{i,-1} - \mathbf{X}_{i,-1}\boldsymbol{\beta}_{xi} - \bar{\mathbf{Z}}_{-1}\boldsymbol{\lambda}_i^*)\alpha_{yi} + \Delta \mathbf{y}_{i,-1}\mathbf{A}'_{y,yi}(L) + \Delta \mathbf{X}_{i,-1}\mathbf{A}'_{y,xi}(L) \\ &\quad + \bar{\Delta \mathbf{Z}}_{-1}\mathbf{A}'_{y,fi}(L) + \boldsymbol{\varepsilon}_{y,xi} + \mathbf{V}_{A,-1}\boldsymbol{\lambda}_i\alpha_{yi} \\ &\quad - \mathbf{V}_{B,-1}\mathbf{H}_4^-(L)\mathbf{A}'_{y,fi}(L),\end{aligned}\tag{3.26}$$

with $\boldsymbol{\lambda}_i^* = \mathbf{H}_1^- \boldsymbol{\lambda}_i$ with $\mathbf{H}_1^- = \mathbf{H}'_1(\mathbf{H}_1\mathbf{H}'_1)^{-1}$ and $\mathbf{A}_{y,fi}^{*'}(L) = \mathbf{H}_4^-(L)\mathbf{A}'_{y,fi}(L) + \mathbf{H}'_3(L)\mathbf{H}_1^- \boldsymbol{\lambda}_i\alpha_{yi}$, with $\mathbf{H}_4^-(L) = \mathbf{H}'_4[\mathbf{H}_4(L)\mathbf{H}'_4(L)]^{-1}$.

The estimation of the augmented model is feasible only when the infinite order matrix lag-polynomial $\mathbf{A}_{y,fi}^{*'}(L)$ is approximated by a finite matrix lag-polynomial. This issue is discussed in the next section.

Truncation

This section discusses the truncation issue in some depth. As mentioned before, one needs to truncate the infinite lag polynomials to make the estimation feasible. The truncation is possible as long as the coefficients of the infinite order lag-polynomial are exponentially decaying as the lag order increases. The infinite order lag-polynomial matrix we consider is the one that multiplies the first differences of the cross-sectional averages, that is $\mathbf{A}_{y,fi}^{*'}(L) = \sum_{\ell=0}^{\infty} \mathbf{A}_{y,fi\ell}^* L^\ell$. Now, consider

$$\mathbf{A}_{y,fi}^{*'}(L) = \mathbf{H}_4^-(L)\mathbf{A}'_{y,fi}(L) + \mathbf{H}'_3(L)\mathbf{H}_1^- \boldsymbol{\lambda}_i\alpha_{yi}.\tag{3.27}$$

To show that $\mathbf{A}_{y,fi}^{*'}(L)$ has exponentially decaying coefficients we analyze the two terms on the right-hand side separately. First, consider the first term, and note

that

$$\mathbf{H}'_4(L) = (\bar{\mathbf{C}}_{zf} + \bar{\mathbf{C}}^*_{zf}(L)(1-L))(\mathbf{C}_{ff} + \mathbf{C}^*_{ff}(L)(1-L))^{-1}.$$

Consider $\mathbf{C}_{ff} + \mathbf{C}^*_{ff}(L)(1-L)$. It is the moving average coefficients of the stationary process $\Delta \mathbf{f}_t$. The MA representation can be found in (3.19). By the stability assumption, 3.1(ii), it is invertible and has exponentially decaying coefficients. Then by using Lemma A.1 of Chudik and Pesaran (2013c), it can be established that the inverse, $(\mathbf{C}_{ff} + \mathbf{C}^*_{ff}(L)(1-L))^{-1}$ exists and has exponentially decaying coefficients. Now, consider $\bar{\mathbf{C}}^*_{zf}(L)(1-L)$, which is the cross-sectional average of the coefficients in the moving average of a stationary process, that can be found in (3.18). By using the same arguments as before it can be established that it has exponentially decaying coefficients, see for example Chudik and Pesaran (2013b). Since $\mathbf{H}'_4(L)$ is obtained by a linear combination of the aforementioned lag-polynomials and their products, it also has exponentially decaying coefficients. Having established that $\mathbf{H}'_4(L)$ has exponentially decaying coefficients, similar arguments can be used to obtain the same result for the whole first term and the second term of (3.27).

By choosing a proper lag order for the first differences of the observables one can show that the truncation error disappears at an exponential decay rate. In the theorems provided in this section we refer to this exponential decay rate. So at this point it is necessary to define an exponential decay rate. Following Chudik and Pesaran (2013c) we assume that there exist real positive constants $K < \infty$ and $\rho_A < 1$ such that, $|\mathbf{A}^*_{y,fi\ell}| < K\rho_A^\ell$, for any $\ell \in \{0, 1, 2, \dots\}$, uniformly in i .

We use p_T for the lag order to highlight that the truncation order should increase with the number of time series observations T .

Estimation with unobserved factors

Going back to the notation with coefficients as vectors instead of lag polynomials, the model in equation (3.26) can be rewritten as

$$\begin{aligned} \Delta \mathbf{y}_i &= (\mathbf{y}_{i,-1} - \mathbf{X}_{i,-1}\boldsymbol{\beta}_{xi} - \bar{\mathbf{Z}}_{-1}\boldsymbol{\lambda}_i^*)\alpha_{yi} + \mathbf{G}_{yi}\boldsymbol{\Gamma}_{y,yi} + \mathbf{G}_{xi}\boldsymbol{\Gamma}_{y,xi} \\ &\quad + \bar{\mathbf{G}}_z\boldsymbol{\Gamma}^*_{y,fi} + \tilde{\boldsymbol{\varepsilon}}_{y,xi}, \end{aligned} \tag{3.28}$$

where

$$\tilde{\boldsymbol{\varepsilon}}_{y,xi} = \boldsymbol{\varepsilon}_{y,xi} + \mathbf{V}_{A,-1}\mathbf{H}_1^-\boldsymbol{\lambda}_i\alpha_{yi} - \mathbf{V}_{B,-1}\mathbf{H}_4^-(L)\mathbf{A}'_{y,fi}(L)$$

$$+\overline{\Delta\mathbf{Z}}_{-p_T-1}\mathbf{A}_{y,fi}^{**'}(L), \quad (3.29)$$

and where

$$\begin{aligned} \Delta\mathbf{y}_i &= \begin{pmatrix} \Delta y_{i,p_T+1} \\ \vdots \\ \Delta y_{iT} \end{pmatrix}, & \mathbf{y}_{i,-1} &= \begin{pmatrix} y_{i,p_T} \\ \vdots \\ y_{i,T-1} \end{pmatrix}, \\ \mathbf{X}_{i,-1} &= \begin{pmatrix} \mathbf{x}'_{i,p_T} \\ \vdots \\ \mathbf{x}'_{i,T-1} \end{pmatrix}, & \overline{\mathbf{Z}}_{-1} &= \begin{pmatrix} \overline{\mathbf{z}}'_{p_T} \\ \vdots \\ \overline{\mathbf{z}}'_{T-1} \end{pmatrix}, \\ \tilde{\mathbf{e}}_{y,xi} &= \begin{pmatrix} \tilde{e}_{y,xi,p_T+1} \\ \vdots \\ \tilde{e}_{y,xiT} \end{pmatrix}, & \overline{\Delta\mathbf{Z}}_{-p_T-1} &= \begin{pmatrix} \overline{\Delta\mathbf{z}}'_0 \\ \vdots \\ \overline{\Delta\mathbf{z}}'_{T-p_T-1} \end{pmatrix}. \end{aligned}$$

For the current and lagged values of the first differences we have

$$\begin{aligned} \mathbf{G}_{yi} &= \begin{pmatrix} \mathbf{g}'_{yi,p_T+1} \\ \vdots \\ \mathbf{g}'_{yiT} \end{pmatrix}, & \mathbf{G}_{xi} &= \begin{pmatrix} \mathbf{g}'_{xi,p_T+1} \\ \vdots \\ \mathbf{g}'_{xiT} \end{pmatrix}, \\ \overline{\mathbf{G}}_z &= \begin{pmatrix} \overline{\mathbf{g}}'_{z,p_T+1} \\ \vdots \\ \overline{\mathbf{g}}'_{zT} \end{pmatrix}, \end{aligned}$$

with

$$\begin{aligned} \mathbf{g}_{yit} &= \begin{pmatrix} \Delta y_{i,t-1} \\ \vdots \\ \Delta y_{i,t-p+1} \end{pmatrix}, & \mathbf{g}_{xit} &= \begin{pmatrix} \Delta \mathbf{x}_{it} \\ \vdots \\ \Delta \mathbf{x}_{i,t-p+1} \end{pmatrix}, \\ \overline{\mathbf{g}}_{zt} &= \begin{pmatrix} \overline{\Delta \mathbf{z}}_{t-1} \\ \vdots \\ \overline{\Delta \mathbf{z}}_{t-p_T} \end{pmatrix}, \end{aligned}$$

and where $\mathbf{\Gamma}_{y.yi}$ and $\mathbf{\Gamma}_{y.xi}$ are defined subsequent to (3.10),

$$\mathbf{\Gamma}_{y.fi}^* = (\mathbf{A}_{y.fi1}^*, \dots, \mathbf{A}_{y.fi,p_T}^*)'$$

is a $p_T(1+m) \times 1$ vector and $\mathbf{A}_{y.fi}^{**}(L) = \sum_{\ell=0}^{\infty} \mathbf{A}_{y.fi,\ell+p_T}^{**} L^\ell$. Furthermore, for the approximation errors we have

$$\mathbf{V}_{A,-1}^{(T-p_T) \times (1+m)} = \begin{pmatrix} \mathbf{v}_{A,p_T} \\ \vdots \\ \mathbf{v}_{A,T-1} \end{pmatrix}, \quad \mathbf{V}_{B,-1}^{(T-p_T) \times (1+m)} = \begin{pmatrix} \Delta \mathbf{v}_{B,p_T} \\ \vdots \\ \mathbf{v}_{B,T-1} \end{pmatrix}.$$

Note the difference between the definitions of the vectors and matrices for the unobserved factors and the observed factors cases.

The augmented conditional model can then be rewritten as

$$\Delta \mathbf{y}_i = \tilde{\mathbf{Z}}_{i,-1}^* \mathbf{\Pi}_i^* + \mathbf{G}_{zi} \mathbf{\Gamma}_{zi} + \bar{\mathbf{G}}_z \mathbf{\Gamma}_{y.fi}^* + \tilde{\mathbf{e}}_{y.xi}, \quad (3.30)$$

where

$$\mathbf{\Pi}_i^*_{(2+2m) \times 1} = \begin{pmatrix} \alpha_{yi} \\ \mathbf{\Pi}_{xi} \\ \mathbf{H}_1^- \mathbf{\Pi}_{fi} \end{pmatrix}, \quad \mathbf{\Gamma}_{zi}_{(p-1+pm) \times 1} = \begin{pmatrix} \mathbf{\Gamma}_{y.yi} \\ \mathbf{\Gamma}_{y.xi} \end{pmatrix}, \quad (3.31)$$

and $\tilde{\mathbf{Z}}_{i,-1}^* = (\mathbf{Y}_{i,-1}, \mathbf{X}_{i,-1}, \bar{\mathbf{Z}}_{-1})$, $\mathbf{G}_{zi} = (\mathbf{G}_{yi}, \mathbf{G}_{xi})$. To find the asymptotic properties of the estimators for the long-run parameters, we first have to show that α_{yi} can be estimated consistently. For this purpose it is useful to present the asymptotic properties of the short-run parameters for the case in which the factors are approximated by the cross-sectional averages of the observables. We use the notation $\hat{\mathbf{\Pi}}_i$ and $\hat{\mathbf{\Gamma}}_{zi}$ to denote the OLS estimators of $\mathbf{\Pi}_i^*$ and $\mathbf{\Gamma}_{zi}$ for the case with unobserved factors. $\mathbf{\Pi}_i^*$ and $\mathbf{\Gamma}_{zi}$ are defined in (3.31). The estimator can be written as

$$\begin{pmatrix} \hat{\mathbf{\Pi}}_i \\ \hat{\mathbf{\Gamma}}_{zi} \end{pmatrix} = \begin{pmatrix} \tilde{\mathbf{Z}}_{i,-1}^{*'} \bar{\mathbf{M}}_Z \tilde{\mathbf{Z}}_{i,-1}^* & \tilde{\mathbf{Z}}_{i,-1}^{*'} \bar{\mathbf{M}}_Z \mathbf{G}_{zi} \\ \mathbf{G}_{zi}' \bar{\mathbf{M}}_Z \tilde{\mathbf{Z}}_{i,-1}^* & \mathbf{G}_{zi}' \bar{\mathbf{M}}_Z \mathbf{G}_{zi} \end{pmatrix}^{-1} \begin{pmatrix} \tilde{\mathbf{Z}}_{i,-1}^{*'} \bar{\mathbf{M}}_Z \Delta \mathbf{y}_i \\ \mathbf{G}_{zi}' \bar{\mathbf{M}}_Z \Delta \mathbf{y}_i \end{pmatrix}, \quad (3.32)$$

where

$$\bar{\mathbf{M}}_Z = \mathbf{I} - \bar{\mathbf{G}}_z (\bar{\mathbf{G}}_z' \bar{\mathbf{G}}_z)^{-1} \bar{\mathbf{G}}_z'. \quad (3.33)$$

This estimator is the unobserved factors counterpart of the estimator defined in (3.12). In addition to the orthogonal projection matrix defined above, we define

$$\ddot{\mathbf{M}}_F = \mathbf{I} - \ddot{\mathbf{G}}_F(\ddot{\mathbf{G}}_F' \ddot{\mathbf{G}}_F)^+ \ddot{\mathbf{G}}_F', \quad (3.34)$$

where $\ddot{\mathbf{G}}_F = (\ddot{\mathbf{g}}_{f,p_T+1}, \dots, \ddot{\mathbf{g}}_{fT})'$, with

$$\ddot{\mathbf{g}}_{ft} = \begin{pmatrix} \mathbf{H}'_4(L) \Delta \mathbf{f}_{t-1} \\ \mathbf{H}'_4(L) \Delta \mathbf{f}_{t-2} \\ \vdots \\ \mathbf{H}'_4(L) \Delta \mathbf{f}_{t-p_T} \end{pmatrix}.$$

$\ddot{\mathbf{g}}_{ft}$ is defined considering the approximation result for the first differences of the factors which is given in (3.22). We make use of the result that for all $t = 1, 2, \dots, T$ the approximation error is $O_p(N^{-1/2})$ to show that having $\bar{\mathbf{M}}_Z$ or $\ddot{\mathbf{M}}_F$ in the estimator is asymptotically equivalent under certain conditions placed on the relative growth rates of N , T and p_T . We establish this result in the proof of the following theorem.

Now consider the model

$$\Delta \mathbf{y}_i = (\mathbf{y}_{i,-1} - \mathbf{X}_{i,-1} \boldsymbol{\beta}_{xi} - \bar{\mathbf{Z}}_{-1} \boldsymbol{\lambda}_i^*) \alpha_{yi} + \tilde{\mathbf{G}}_i \boldsymbol{\Gamma}_i^* + \tilde{\boldsymbol{\varepsilon}}_{y.xi}, \quad (3.35)$$

where $\tilde{\mathbf{G}}_i = (\mathbf{G}_{yi}, \mathbf{G}_{xi}, \bar{\mathbf{G}}_z)$, $\boldsymbol{\Gamma}_i^* = (\boldsymbol{\Gamma}'_{y.yi}, \boldsymbol{\Gamma}'_{y.xi}, \boldsymbol{\Gamma}'_{y.fi})'$. We use $\hat{\boldsymbol{\beta}}_{xi}$ and $\hat{\boldsymbol{\lambda}}_i$ to denote the indirect least squares estimators for $\boldsymbol{\beta}_{xi}$ and $\boldsymbol{\lambda}_i^*$ respectively. The indirect least squares estimator for the long-run coefficients can be obtained as

$$\begin{pmatrix} \hat{\boldsymbol{\beta}}_{xi} \\ \hat{\boldsymbol{\lambda}}_i \end{pmatrix} = -\frac{1}{\hat{\alpha}_{yi}} \left(\tilde{\mathbf{R}}'_{i,-1} \tilde{\mathbf{M}}_i \tilde{\mathbf{R}}_{i,-1} \right)^{-1} \left(\tilde{\mathbf{R}}'_{i,-1} \tilde{\mathbf{M}}_i (\Delta \mathbf{y}_i - \mathbf{y}_{i,-1} \hat{\alpha}_{yi}) \right), \quad (3.36)$$

where $\tilde{\mathbf{R}}_i = (\mathbf{X}_i, \bar{\mathbf{Z}})$ and

$$\tilde{\mathbf{M}}_i = \mathbf{I} - \tilde{\mathbf{G}}_i (\tilde{\mathbf{G}}_i' \tilde{\mathbf{G}}_i)^{-1} \tilde{\mathbf{G}}_i', \quad (3.37)$$

and $\tilde{\mathbf{G}}_i$ is defined subsequent to (3.35) and $\hat{\alpha}_{yi}$ is the estimator for the coefficient of the error correction term, obtained by (3.32). The asymptotic properties of the estimators (3.32) and (3.36) are stated in the following theorems.

Theorem 3.3. *Suppose that \mathbf{z}_{it}^* for $i = 1, \dots, N$ and $t = 1, \dots, T$ are generated by (3.3)-(3.5), that Assumptions 3.1, 3.2, 3.3 and 3.4 hold and consider the model in (3.30). Let $\frac{T}{N} \rightarrow 0$, $\frac{p_T^2}{T} \rightarrow 0$ and $\sqrt{T} \rho_A^{p_T} \rightarrow 0$ as $(N, T, p_T) \rightarrow \infty$, then $\hat{\boldsymbol{\Pi}}_i$ and*

$\widehat{\Gamma}_{zi}$, defined by (3.32) are consistent estimators of Π_i^* and Γ_{zi}^* respectively and

$$\sqrt{T} \begin{pmatrix} \widehat{\Pi}_i - \Pi_i^* \\ \widehat{\Gamma}_{zi} - \Gamma_{zi}^* \end{pmatrix} \xrightarrow{d} N \left(0, \begin{pmatrix} \beta_i^* & 0 \\ 0 & \mathbf{I}_{q^*} \end{pmatrix} \Psi_i^{-1} \begin{pmatrix} \beta_i^* & 0 \\ 0 & \mathbf{I}_{q^*} \end{pmatrix}' \sigma_{y.xi} \right),$$

where $\beta_i^* = (1, -\beta'_{xi}, -\lambda_i^*)'$ and

$$\begin{pmatrix} T^{-1} \mathbf{U}'_{i,-1} \ddot{\mathbf{M}}_F \mathbf{U}_{i,-1} & T^{-1} \mathbf{U}'_{i,-1} \ddot{\mathbf{M}}_F \mathbf{G}_{zi} \\ T^{-1} \mathbf{G}'_{zi} \ddot{\mathbf{M}}_F \mathbf{U}_{i,-1} & T^{-1} \mathbf{G}'_{zi} \ddot{\mathbf{M}}_F \mathbf{G}_{zi} \end{pmatrix} \xrightarrow{p} \Psi_i \quad (3.38)$$

with $\ddot{\mathbf{M}}_F$ defined in (3.34). $\begin{pmatrix} \beta_i^* & 0 \\ 0 & \mathbf{I}_{q^*} \end{pmatrix} \Psi_i^{-1} \begin{pmatrix} \beta_i^* & 0 \\ 0 & \mathbf{I}_{q^*} \end{pmatrix}'$ can be consistently estimated by

$$T^{-1} \begin{pmatrix} \widetilde{\mathbf{Z}}'_{i,-1} \overline{\mathbf{M}}_Z \widetilde{\mathbf{Z}}^*_{i,-1} & \widetilde{\mathbf{Z}}'_{i,-1} \overline{\mathbf{M}}_Z \mathbf{G}_{zi} \\ \mathbf{G}'_{zi} \overline{\mathbf{M}}_Z \widetilde{\mathbf{Z}}^*_{i,-1} & \mathbf{G}'_{zi} \overline{\mathbf{M}}_Z \mathbf{G}_{zi} \end{pmatrix}^{-1}. \quad (3.39)$$

Furthermore, to consistently estimate the residual variance we can use

$$\hat{\sigma}_{y.xi} = \frac{1}{d_k} (\Delta \mathbf{y}_i - \widetilde{\mathbf{Z}}^*_{i,-1} \widehat{\Pi}_i - \widetilde{\mathbf{G}}_i \widehat{\Gamma}_i)' (\Delta \mathbf{y}_i - \widetilde{\mathbf{Z}}^*_{i,-1} \widehat{\Pi}_i - \widetilde{\mathbf{G}}_i \widehat{\Gamma}_i), \quad (3.40)$$

where $d_k = T - 3 - 2m - (m-1)p - (m+1)p_T$.

Theorem 3.4. Suppose that \mathbf{z}_{it}^* for $i = 1, \dots, N$ and $t = 1, \dots, T$ are generated by (3.3)-(3.5) and Assumptions 3.1, 3.2, 3.3 and 3.4 hold, consider the conditional error correction model in (3.35). Let $\frac{T}{\sqrt{N}} \rightarrow 0$, $\frac{p_T^2}{T} \rightarrow 0$ and $\sqrt{T} \rho_A^{p_T} \rightarrow 0$ as $(N, T, p_T) \rightarrow \infty$. $\widehat{\beta}_{xi}$ and $\widehat{\lambda}_i$ defined by (3.36) are consistent estimators of β_{xi} and λ_i^* respectively and their asymptotic distribution is given as

$$T \begin{pmatrix} \widehat{\beta}_{xi} - \beta_{xi} \\ \widehat{\lambda}_i - \lambda_i^* \end{pmatrix} \xrightarrow{d} -\frac{1}{\alpha_{yi}} \left[\mathbf{H}_R \begin{pmatrix} \int_0^1 \mathbf{C}_{xzi} \mathbf{B}_{bi}(r) \mathbf{B}'_{bi}(r) dr \mathbf{C}'_{xzi} & \int_0^1 \mathbf{C}_{xzi} \mathbf{B}_{bi}(r) \mathbf{B}'_f(r) dr \mathbf{C}'_{ff} \\ \int_0^1 \mathbf{C}_{ff} \mathbf{B}_f(r) \mathbf{B}'_{bi}(r) dr \mathbf{C}'_{xzi} & \int_0^1 \mathbf{C}_{ff} \mathbf{B}_f(r) \mathbf{B}'_f(r) dr \mathbf{C}'_{ff} \end{pmatrix} \mathbf{H}'_R \right]^{-1} \times \left[\mathbf{H}_R \begin{pmatrix} \int_0^1 \mathbf{C}_{xzi} \mathbf{B}_{bi}(r) dB'_{y.xi}(r) \\ \int_0^1 \mathbf{C}_{ff} \mathbf{B}_f(r) dB'_{y.xi}(r) \end{pmatrix} \right],$$

where $\mathbf{H}_R = \begin{pmatrix} \mathbf{I} & 0 \\ 0 & \mathbf{H}_1 \end{pmatrix}$, $\mathbf{B}_{bi}(r) = (\mathbf{B}'_{xi}(r), \mathbf{B}'_f(r))'$ and $B_{y.xi}(r)$, $\mathbf{B}_{xi}(r)$, $\mathbf{B}_f(r)$ are Brownian motions defined in (3.8), where $\mathbf{B}_i^*(r) = (B_{y.xi}(r), \mathbf{B}'_{xi}(r), \mathbf{B}'_f(r))'$ has a covariance matrix defined in (3.9) and \mathbf{C}_{xzi} , \mathbf{C}_{ff} defined in Appendix 3.B.1, in (3.B.22) and (3.B.4) respectively and \mathbf{H}_1 is defined in (3.16).

The proofs of the theorems provided in this section can be found in Appendix 3.B.2. The Δ -method suggested by Bardsen (1989) can be used in computing the estimator for the variance of the long-run coefficients, the details can be found in Theorem 3.2 of the current chapter.

Remark 3.14. As pointed out with observed factors, the weak and strong exogeneity assumptions ensure that the estimators for the long-run coefficients are asymptotically mixed-normal. See Remark 3.8 for a detailed explanation.

Remark 3.15. The asymptotic representation above suggests that the short-run and the long-run coefficients of the unobserved factors can only be estimated up to a rotation matrix. λ_i^* is the rotation of the long-run factor loadings, namely $\lambda_i^* = \mathbf{H}_1^- \lambda_i$.

Remark 3.16. In Theorem 3.3 we present the estimator in terms of Ψ_i , which is defined as the limit of the left-hand side of (3.38) as $(N, T, p_T) \rightarrow \infty$. Then in the second part of the theorem we propose an estimator for that limit. Since it is established that the limit can be estimated directly by (3.39), we do not provide an explicit expression for the limit result in the theorem.

Remark 3.17. In Theorem 3.4, the required condition regarding the relative growth rates of N and T is $\frac{T}{\sqrt{N}} \rightarrow 0$. This is required to eliminate the effects of approximating the levels of the unobserved factors. Since in our set-up, contrary to other papers on this branch of the literature, (See, for example Bai et al. (2009) or Kapetanios et al. (2011)), the factor and the idiosyncratic variables are not cointegrated among themselves without the dependent variable and the factors are not cointegrated among themselves. Under these assumptions, the approximation error for the levels includes the cross-sectional average of a partial sum for finite N , whereas for example in Kapetanios et al. (2011) the approximation error is a covariance stationary process. The presence of the partial sum in the approximation error necessitates a more restrictive condition that is $\frac{T}{\sqrt{N}} \rightarrow 0$ than the ones in the existing literature, i.e., Kapetanios et al. (2011) requires $\frac{\sqrt{T}}{N} \rightarrow 0$ for the consistent estimation of the individual coefficients.

Remark 3.18. The last two conditions stated in Theorems 3.3 and 3.4 are $\frac{p_T^2}{T} \rightarrow 0$ and $\sqrt{T}\rho_A^{p_T} \rightarrow 0$, where p_T is the truncation lag order for the first differences of the cross-sectional averages and ρ_A^ℓ is the exponential decay rate of the coefficients of the lag-polynomial matrix $\mathbf{A}_{y.fi}^{**}(L) = \sum_{\ell=0}^{\infty} \mathbf{A}_{y.fi}^{**} L^\ell$ that appears in (3.29). These conditions are required to make the truncation work. $\frac{p_T^2}{T} \rightarrow 0$ imposes an upper bound for the growth rate of the truncation lag, which allows for enough degrees of freedom for consistent estimation. The second condition regarding the choice of the truncation lag order is $\sqrt{T}\rho_A^{p_T} \rightarrow 0$. This imposes a lower bound for the truncation lag order that would make the bias that arises because of the truncation of the infinite lag order sufficiently small. For a further discussion see, for example, Berk (1974), Said and Dickey (1984), Chudik and Pesaran (2013c).

3.3.3 Mean Group Estimator

In this section we consider a panel estimator where the parameter of interest is the mean of the individual long-run parameters, that is $\beta_x = E(\beta_{xi})$. A natural estimator is the mean group estimator proposed in Pesaran and Smith (1995) and also considered in Kapetanios et al. (2011). It is simply obtained by averaging individual coefficients over cross-section units as

$$\widehat{\beta}_{xMG} = \frac{1}{N} \sum_{i=1}^N \widehat{\beta}_{xi}, \quad (3.41)$$

where $\widehat{\beta}_{xi}$ is given by (3.36). We are only interested in the mean group estimator of the long-run coefficients of the idiosyncratic variables, since the λ_i can in any case only be estimated up to a rotation matrix and hence is of no direct interest. For this reason we define the orthogonal projection matrix

$$\overline{\mathbf{M}}_Q = \mathbf{I} - \overline{\mathbf{Z}}_{-1}(\overline{\mathbf{Z}}'_{-1}\overline{\mathbf{Z}}_{-1})^{-1}\overline{\mathbf{Z}}'_{-1}.$$

Using this and the orthogonal projection matrix $\widetilde{\mathbf{M}}_i$ defined in (3.37), which projects onto the orthogonal space spanned by the first differences in the augmented model and using Assumption 3.3(ii), we can rewrite the estimator as

$$\begin{aligned} \widehat{\beta}_{xMG} - \beta_x &= \frac{1}{N} \sum_{i=1}^N \eta_{xi} \\ &= \frac{1}{N} \sum_{i=1}^N \frac{1}{\widehat{\alpha}_{yi}} \left(\mathbf{X}'_{i,-1} \overline{\mathbf{M}}_Q \widetilde{\mathbf{M}}_i \overline{\mathbf{M}}_Q \mathbf{X}_{i,-1} \right)^{-1} \left(\mathbf{X}'_{i,-1} \overline{\mathbf{M}}_Q \widetilde{\mathbf{M}}_i \overline{\mathbf{M}}_Q \varepsilon^*_{y.xi} \right), \end{aligned} \quad (3.42)$$

where $\tilde{\varepsilon}_{y.xi}^*$ is defined analogously to (3.29) as

$$\begin{aligned}\tilde{\varepsilon}_{y.xi}^* &= \varepsilon_{y.xi} + (\varepsilon_{y.xi}^* - \varepsilon_{y.xi}) + \mathbf{V}_{A,-1} \mathbf{H}_1^- \boldsymbol{\lambda}_i \alpha_{yi} + \mathbf{V}_{A,-1} \mathbf{H}_1^- \boldsymbol{\lambda}_i (\hat{\alpha}_{yi} - \alpha_{yi}) \\ &\quad - \mathbf{V}_{B,-1} \mathbf{H}_4^- (L) \mathbf{A}'_{y.fi}(L) + \overline{\Delta \mathbf{Z}}_{-p_T-1} \mathbf{A}^{**'}_{y.fi}(L),\end{aligned}\quad (3.43)$$

where the superscript ‘*’ of $\varepsilon_{y.xi}^*$ indicates that it is the error term of the model where the true value of α_{yi} is replaced by the estimated value $\hat{\alpha}_{yi}$ that is estimated by (3.32). The result regarding the consistency of this estimator can be stated formally as follows:

Theorem 3.5. *Suppose that \mathbf{z}_{it}^* for $i = 1, \dots, N$ and $t = 1, \dots, T$ are generated by (3.3)-(3.5) and assumptions 3.1-3.3 and 3.4 hold, consider the model in (3.30). Let $(N, T, p_T) \rightarrow \infty$, under the condition that $\frac{p_T^2}{T} \rightarrow 0$, we have*

$$\hat{\boldsymbol{\beta}}_{xMG} - \boldsymbol{\beta}_x \xrightarrow{P} 0,$$

where $\hat{\boldsymbol{\beta}}_{xMG}$ is defined by (3.41).

The proof of this theorem can be found in Appendix 3.B.3. The consistency result does not require any conditions on the relative expansion rates of T , N and p_T , except an upper bound for the truncated lag order that is $\frac{p_T^2}{T} \rightarrow 0$, which is a necessary condition to ensure that we have enough degrees of freedom for the consistent estimation.

In order to obtain the asymptotic distribution of the estimator in question we first multiply both sides of (3.42) by \sqrt{N} ,

$$\begin{aligned}\sqrt{N}(\hat{\boldsymbol{\beta}}_{xMG} - \boldsymbol{\beta}_x) &= \frac{1}{\sqrt{N}} \sum_{i=1}^N \eta_{xi} \\ &\quad - \frac{1}{\sqrt{N}} \sum_{i=1}^N \frac{1}{\hat{\alpha}_{yi}} \left(\mathbf{X}'_{i,-1} \overline{\mathbf{M}}_Q \widetilde{\mathbf{M}}_i \overline{\mathbf{M}}_Q \mathbf{X}_{i,-1} \right)^{-1} \left(\mathbf{X}'_{i,-1} \overline{\mathbf{M}}_Q \widetilde{\mathbf{M}}_i \overline{\mathbf{M}}_Q \tilde{\varepsilon}_{y.xi}^* \right).\end{aligned}$$

The first term on the right-hand side has a zero mean and a finite variance $\boldsymbol{\Omega}_{x\eta}$ and emerges from the random coefficients assumption on $\boldsymbol{\beta}_{xi}$, by Assumption 3.3(ii). The result regarding the asymptotic distribution of $\hat{\boldsymbol{\beta}}_{xMG}$ can now be stated in a formal way as follows:

Theorem 3.6. *Suppose that \mathbf{z}_{it}^* for $i = 1, \dots, N$ and $t = 1, \dots, T$ are generated by (3.3)-(3.5) and Assumptions 3.1-3.4 hold, consider the conditional error correction model in (3.28). Let $\frac{p_T^2}{T} \rightarrow 0$ and $\frac{\rho_A^{p_T} \sqrt{N}}{T} \rightarrow 0$ as $(N, T, p_T) \rightarrow \infty$. Then $\hat{\boldsymbol{\beta}}_{xMG}$*

defined by (3.41) has an asymptotic distribution given by

$$\sqrt{N}(\widehat{\beta}_{xMG} - \beta_x) \xrightarrow{d} N(0, \Sigma_{MG}),$$

with

$$\Sigma_{MG} = \Omega_{\eta x} + \overline{\Lambda}_q,$$

where

$$\overline{\Lambda}_q = \lim_{N, T, p_T \rightarrow \infty} \frac{1}{N} \sum_{i=1}^N \text{Var} \left[\left(\frac{(\mathbf{S}_{xi,-1}^C)' \mathbf{S}_{xi,-1}^C}{T^2} \right)^{-1} \left(\frac{(\mathbf{S}_{xi,-1}^C)' \overline{\mathbf{S}}_{x,-i,-1}^C}{T^2} \right) \mathbf{H}_1^- \boldsymbol{\eta}_{fi} \right],$$

where $\mathbf{S}_{xi,-1}^C = \overline{\mathbf{S}}_{x,-i,-1} \overline{\mathbf{C}}'_{zx,-i}$, $\overline{\mathbf{S}}_{x,-i,-1} \overline{\mathbf{C}}'_{zx,-i} = \overline{\mathbf{S}}_{x,-i,-1}^C$ and $\mathbf{S}_{xi,-1}$ is defined in (3.B.5) in Appendix 3.B.1. $\overline{\mathbf{S}}_{x,-i,-1}$ and $\overline{\mathbf{C}}_{zx,-i}$ are defined in (3.B.150) and (3.B.151) respectively, in Appendix 3.B.3. $\mathbf{H}_1^- = \mathbf{H}'_1 (\mathbf{H}_1 \mathbf{H}'_1)^{-1}$ and \mathbf{H}'_1 is defined in (3.16). Furthermore, following Pesaran (2006), the variance can consistently be estimated by a non-parametric method by using

$$\widehat{\Sigma}_{MG} = \frac{1}{N-1} \sum_{i=1}^N (\widehat{\beta}_i - \widehat{\beta}_{MG})(\widehat{\beta}_i - \widehat{\beta}_{MG})'. \quad (3.44)$$

The proof of the theorem can be found in Appendix 3.B.3.

Remark 3.19. We exploit the heterogeneity and the random coefficients structure of the long-run coefficients to obtain the \sqrt{N} consistency in the theorem above.

Remark 3.20. The conditions required for Theorem 3.6 to work are $\frac{p_T^2}{T} \rightarrow 0$ and $\frac{\rho_A^{p_T} \sqrt{N}}{T} \rightarrow 0$. Similar to the previous theorems, the first one is necessary to have a sufficient number of degrees of freedom for consistent estimation. The second one imposes a restriction on the relative growth rates of N , T and p_T , where p_T is the truncation lag order. This condition is necessary to make the bias arising due to truncation sufficiently small.

Remark 3.21. The second term in the variance $\overline{\Lambda}_q$ of the estimator arises due to the heterogeneity of the factor loadings in the long-run.

3.4 Monte Carlo Simulations

In this section we evaluate the finite sample performance of the estimators provided in the previous section. We consider the individual specific statistics and the mean-group estimator of the long-run coefficients of the idiosyncratic variables. The estimators are defined by (3.36) and (3.41) respectively. In addition to the estimators proposed in this chapter, which will be denoted as the DCCE estimators in this section, we consider some alternative estimators. Namely, we consider the infeasible estimator as a benchmark to see the effect of approximation on the performance of the estimator. The infeasible estimator is obtained by using the realization of the true factors. The second estimator we consider is the estimator obtained from the static model of Kapetanios et al. (2011) and denoted as the KCCE in this section. They consider the estimator for the long-run parameters by estimating the static long-run model without parametrically modeling the short-run dynamics. Kapetanios et al. (2011) show that the KCCE estimator outperforms the principal components estimator proposed by Bai et al. (2009). Hence, we do not consider the latter in our simulation experiments. The theoretical findings of the current chapter imply that the lag order of the cross-sectional averages should be truncated. To be able to observe the effect of using different lag selection methods we consider three different methods, (i) using the lag order used in the DGP for the observed variables and the truncation lag order which is computed using $p_T = \lfloor T^{1/3} \rfloor$ for the first differences of the cross-sectional averages, (ii) using lag order used in the DGP for the observed variables and the lag order used in the DGP, which is equal to 1, (iii) using the Akaike-Information Criteria to select the lag order for the first differences.

Furthermore we consider three different types of long-run relations. First we restrict the long-run parameters to be homogeneous across cross-section units, for the second case we let the factor loadings in the long-run relation be heterogeneous up to a certain degree, lastly we let the parameter of interest β_{xi} be heterogeneous across cross-section units as well as the factor loadings. We also consider three different setups regarding the weak exogeneity. In the first case, the DGP does not violate the assumption of weak exogeneity, in the second case the individual cross-section units violate the assumption of weak exogeneity whereas on average the assumption holds and in the third case the weak exogeneity assumption is violated also on average.

3.4.1 Data Generating Process

The data generating process for the simulations is analogous to the one we used in the theoretical part of this chapter. We generate the data according to the following model:

$$\begin{aligned}\Delta y_{it} &= \alpha_{yi}(y_{i,t-1} - \beta_{xi}x_{i,t-1} - \lambda_{1i}f_{1,t-1} - \lambda_{2i}f_{2,t-1}) + A_{yyi}\Delta y_{i,t-1} \\ &\quad + A_{yxi}\Delta x_{i,t-1} + A_{yf1i}\Delta f_{1,t-1} + A_{yf2i}\Delta f_{2,t-1} + \varepsilon_{yit} + b_i\varepsilon_{xit}, \\ \Delta x_{it} &= \alpha_{xi}(y_{i,t-1} - \beta_{xi}x_{i,t-1} - \lambda_{1i}f_{1,t-1} - \lambda_{2i}f_{2,t-1}) + A_{xyi}\Delta y_{i,t-1} \\ &\quad + A_{xxi}\Delta x_{i,t-1} + A_{xf1i}\Delta f_{1,t-1} + A_{xf2i}\Delta f_{2,t-1} + \varepsilon_{xit},\end{aligned}$$

for $i = 1, \dots, N$ and $t = 1, \dots, T$. The number of idiosyncratic variables, $m = 1$. The number of common factors, $k = 2$ and they are generated according to

$$\begin{aligned}\Delta f_{1t} &= A_{ff11}\Delta f_{1,t-1} + A_{ff12}\Delta f_{2,t-1} + \varepsilon_{f1t}, \\ \Delta f_{2t} &= A_{ff22}\Delta f_{2,t-1} + \varepsilon_{f2t},\end{aligned}$$

satisfying the strong exogeneity assumption. The true lag order, $p = 2$ in terms of the notation of the model in (3.7). The parameters of the first differences of $\Delta y_{i,t-1}$, $\Delta x_{i,t-1}$, $\Delta f_{1,t-1}$ and $\Delta f_{2,t-1}$ in the models for Δy_{it} and Δx_{it} are generated as

$$\begin{aligned}\begin{pmatrix} A_{yyi} & A_{xyi} \\ A_{xyi} & A_{xxi} \end{pmatrix} &\sim IID \begin{pmatrix} N(-0.50, 0.25) & N(0.25, 0.25) \\ N(0.25, 0.25) & N(-0.50, 0.25) \end{pmatrix}, \\ \begin{pmatrix} A_{yf1i} & A_{yf2i} \\ A_{xf1i} & A_{xf2i} \end{pmatrix} &\sim IID \begin{pmatrix} N(-0.25, 0.25) & 0 \\ 0 & N(-0.25, 0.25) \end{pmatrix}.\end{aligned}$$

The parameters of the lagged first differences of the common factors in the models for Δf_{1t} and Δf_{2t} are equal to

$$\begin{pmatrix} A_{ff11} & A_{ff12} \\ 0 & A_{ff22} \end{pmatrix} = \begin{pmatrix} 0.50 & -0.25 \\ 0 & -0.50 \end{pmatrix}.$$

We draw the error terms as

$$\begin{aligned}\varepsilon_{yit} &\sim IIDN(0, 1), & \varepsilon_{xit} &\sim IIDN(0, 1), \\ \varepsilon_{f1t} &\sim IIDN(0, 1), & \varepsilon_{f2t} &\sim IIDN(0, 1),\end{aligned}$$

and $b_i = 0.2$. The coefficient of the error correction term is generated as

$$\alpha_{yi} \sim U[-1.25, -0.75].$$

The parameters of the long-run relation are generated as

$$\beta_{xi} = -1 + \eta_{xi}, \quad \lambda_{1i} = -2 + \eta_{f1i}, \quad \lambda_{2i} = -2 + \eta_{f2i},$$

where η_{xi} , η_{f1i} and η_{f2i} depend on the experiment. We report the results for experiments with 5 different DGP's. The heterogeneity structure of the long-run relation and the set-up for the weak exogeneity for different experiments are shown in the table below.

Table 3.1: Description of Experiments

Exp	Table # 3.	Long-run heterogeneity		Weak exogeneity
E1	E1.A, E1.B	$\beta_{xi} = -1$	$\lambda_{ij} = -2$	$\alpha_{xi} = 0$
E2	E2.A, E2.B	$\beta_{xi} = -1$	$\lambda_{ij} \sim U[-2.1, -1.9]$	$\alpha_{xi} = 0$
E3	E3	$\beta_{xi} \sim U[-1.1, -0.9]$	$\lambda_{ij} \sim U[-2.1, -1.9]$	$\alpha_{xi} = 0$
E4	E4	$\beta_{xi} = -1$	$\lambda_{ij} \sim U[-2.1, -1.9]$	$\alpha_{xi} \sim U[-0.15, 0.15]$
E5	E5	$\beta_{xi} = -1$	$\lambda_{ij} = -2$	$\alpha_{xi} \sim U[-0.45, -0.15]$

We report the bias, root mean squared error, size and power estimates of the estimators. The results reported in the tables are based on 5000 replications. For each experiment 20 (N, T) pairs are considered with $N, T \in \{40, 60, 100, 150, 200\}$. The size and the power results are based on t-test statistics that are obtained by using the variance estimators proposed in the chapter, i.e. for the individual estimator that we use (3.39) and (3.40). For the mean group estimator we use (3.44). To assess the power we set $H_0 : \beta_{xi} = -1.1$.

3.4.2 Simulation Results

The results for each experiment are reported in Tables 3.E1.A-3.E5. The conclusion that can be drawn from the tables are summarized as follows.

- E1.** The aim is to investigate the small sample performance of the DCCE estimator under weak exogeneity and with homogeneous long-run. The results are reported in Tables 3.E1.A for the first individual unit and in 3.E1.B for the mean-group estimator. From the second column of Table 3.E1.A we see that the bias of the DCCE estimator of the individual long-run coefficient, β_{x1} , is

close to the bias of the infeasible estimator. As T and N increase jointly, we observe a decrease in the bias for all different lag order specifications, which supports our theoretical finding that suggests that under certain conditions on the relative growth rates of N and T , the DCCE estimator of the long-run parameters are T consistent. Also as expected by the standard theory, AIC works better as T increases. RMSE and size decrease with T , but are not affected by an increase in N . This is also an evidence in line with the T consistency of the individual estimator. We observe low power for small values of T , whereas an increase in power as T increases. The poor performance of the KCCE estimator is observed in almost all the experiments. From Table 3.E1.B, we see that the mean group estimator has a smaller bias than the estimator for the i^{th} unit. For small values of T , AIC method leads to a higher bias. A comparison of the DCCE estimator and the infeasible estimator shows that the effect of the approximation disappears as T and N increase, as expected by the theory. The results regarding the size and power suggest that the T-tests based on the proposed estimator are correctly sized and have a power increasing with N and T .

- E2.** The aim of this experiment is to investigate the performance of the estimator for the case with heterogeneous factor loadings in the long-run. The results are reported in Tables 3.E2.A for the first individual unit and in 3.E2.B for the mean-group estimator. As expected by the theory, regarding the bias of the estimators of β_{xi} and β_{xMG} the difference between the DCCE estimators and the infeasible estimators disappears as N and T increase. For small values of T the DCCE estimator with $p_f = p_T$, the DCCE and the infeasible estimators with p_f selected by AIC are relatively biased compared to the DCCE and the infeasible estimator with $p_f = 1$ (lag order used in the DGP). But as T increases the difference disappears. The RMSE is decreasing by the increase in the time-series dimension. The size and power properties of the DCCE estimators are very similar to those of the infeasible estimator which is correctly sized and has a high power.
- E3.** The long-run parameters are generated heterogeneously in this experiment. Only the results for the mean-group estimator are reported and can be found in Table 3.E3. The bias and the RMSE decrease substantially as N and T increase jointly as expected by the theoretical findings of the current chapter. The size results suggest that as T increases the t-test becomes substantially

undersized also for the infeasible estimator. This finding requires further investigation. Furthermore, we observe that as N and T increase power increases.

- E4.** In this experiment we allow for the violation of the weak exogeneity assumption on an individual basis with a heterogeneous long-run, on average over the cross-section units the weak exogeneity assumption still holds. Therefore we draw the coefficient of the error correction term of the marginal model as $\alpha_{xi} \sim U[-0.15, 0.15]$ where $E(\alpha_{xi}) = 0$. We only report the results for the mean-group estimator, which can be found in Table 3.E4. We observe that approximation methods work better for large values of N and T , but the bias of the DCCE estimator remains slightly higher compared to the infeasible estimator, whereas the reported values for the RMSE for the two estimators are similar. As for the size and the power one should expect a well-performing t-test.
- E5.** The aim of this experiment to observe the effect of the violation of the weak exogeneity assumption. We generate the data by setting $E(\alpha_{xi}) = -0.3$ while remaining in the long-run homogeneity case. We report the results only for the estimator for the individual cross-section unit estimator. The violation of the weak exogeneity assumption effects the infeasible estimator as well as the DCCE estimator. Overall as expected, we observe a worsening in the performances of the estimators compared to the case in which the weak exogeneity assumption holds. One can still observe the decrease in the bias and the RMSE as T increases, which is promising for theoretical counterpart. The size of the t-test stays at a reasonable level and converges to 5% as T increases.

3.5 Conclusion

In this chapter we develop an asymptotic theory for a cointegrated heterogeneous panel where the assumption of cross-sectionally independent errors is violated. In the literature, one way to model the cross-sectional dependence is using latent common factor models. In this work we assume a factor structure with unobserved and possibly $I(1)$ factors. The approximation method we use for the unobserved factors is the one introduced in Pesaran (2006). We adopt the CCE framework and use cross-sectional averages to approximate the unobserved, possibly $I(1)$ com-

mon factors. The idiosyncratic variables are weakly exogenous in our model. A Lagrange Multiplier test for weak exogeneity of the idiosyncratic variables is proposed in Chapter 4. We estimate the conditional error correction model where we assume a cointegrating relation between the observables and the factors.

We present the asymptotic properties of the OLS estimators of the coefficients in the ECM model and the indirect least squares estimators for the long-run coefficients for the case in which the factors are assumed to be known and for the case in which the factors are approximated by the cross-sectional averages of the observable variables. Furthermore, we present the asymptotic properties of the mean group estimator for the long-run coefficients of the idiosyncratic variables. The approximation of the factors in our case differs from the ones existing in the related literature. In our set-up, consistent estimation of the model parameters requires that the levels of the factors appear in the cointegrating relation to be approximated using the levels and also the lagged first differences of the observables.

The asymptotic results obtained show that the OLS estimation of the model when the factors are assumed to be observed yields \sqrt{T} -consistent group specific estimators and the indirect least squares estimation yields T -consistent estimates for the long-run coefficients in the cointegrating vector which has a mixed normal distribution as $T \rightarrow \infty$. These results are adaptations of the existing results in the literature to our set-up. Our contribution comes into being for the case where the factors are not observed and the model is augmented by the cross-sectional averages. The method we use requires a truncation for the lag order for the first differences of the of the cross-sectional averages that are required to obtain an asymptotically negligible approximation error in the estimation. The OLS estimator of the short run coefficients and the indirect least squares estimator for the long-run coefficients yield \sqrt{T} and T -consistent estimates respectively and have normal distributions as $(N, T) \rightarrow \infty$ and under the conditions $\frac{T}{\sqrt{N}} \rightarrow 0$, $\frac{v_T^2}{T} \rightarrow 0$ and $\sqrt{T}\rho_A^{p_T} \rightarrow 0$, where p_T is the aforementioned truncation lag order and ρ_A^ℓ , for all $\ell = 1, \dots$, is the exponential decay rate of the lag-polynomial that is in front of the first differences of the cross-sectional averages in the augmented model.

The mean group estimator is found to be consistent and has a normal distribution with a variance that depends on the heterogeneity of the long-run relation, as $\frac{v_T^2}{T} \rightarrow 0$ and $\frac{\rho_A^{p_T}\sqrt{N}}{T} \rightarrow 0$.

Monte Carlo results are consistent with our theoretical findings. We conduct experiments to observe the effects of heterogeneity in the long-run and of violation of weak exogeneity on the small sample performance of the proposed estimator. In all different simulation the performance of the proposed estimator is similar to that of the infeasible estimator, which we present as a benchmark.

3.A Tables

Table 3.E1.A: β_{xi} , weak exogeneity assumption is satisfied, homogeneous long-run

N/T	Bias $\times 100$					RMSE				
	40	60	100	150	200	40	60	100	150	200
DCCE with $p_f = p_T$										
40	0.60	0.09	0.10	0.14	-0.01	23.79	13.33	7.35	4.70	3.36
60	-0.03	0.48	0.11	0.02	0.04	25.47	13.16	7.29	4.63	3.43
100	0.05	0.32	0.20	0.10	0.05	24.41	13.35	7.45	4.76	3.35
150	-0.39	0.07	0.27	0.09	0.06	26.51	14.37	7.41	4.84	3.56
200	-0.52	0.15	0.09	0.26	0.09	25.51	13.92	7.75	4.93	3.40
DCCE with $p_f = 1$										
40	1.15	0.64	0.43	0.38	0.17	19.70	11.88	6.75	4.38	3.30
60	-0.08	0.33	0.16	0.07	0.10	19.17	11.68	6.55	4.32	3.27
100	-0.06	0.42	0.20	0.06	0.05	19.58	11.73	6.74	4.31	3.15
150	-0.36	-0.13	0.02	-0.09	-0.04	20.51	12.56	6.69	4.48	3.37
200	-0.97	-0.18	-0.25	-0.07	-0.12	20.56	12.01	6.98	4.48	3.18
DCCE with p_f selected by AIC										
40	1.46	0.49	0.13	0.15	-0.03	29.25	13.56	7.15	4.50	3.29
60	0.87	0.73	0.10	0.00	0.02	26.81	13.38	7.13	4.51	3.36
100	1.50	0.70	0.12	-0.03	-0.02	24.42	13.11	7.21	4.51	3.24
150	0.79	0.40	0.17	-0.03	0.01	39.53	13.89	7.20	4.63	3.46
200	1.38	0.57	-0.06	0.09	-0.03	26.71	13.44	7.44	4.74	3.27
Infeasible with $p_f = 1$										
40	-0.10	0.01	0.12	0.06	0.00	19.47	12.13	6.52	4.16	3.15
60	-0.71	-0.20	-0.07	0.00	0.06	19.08	11.63	6.46	4.25	3.19
100	-0.76	-0.03	0.07	-0.04	0.03	19.58	11.45	6.54	4.20	3.03
150	-0.43	-0.12	0.00	0.02	0.02	19.90	11.86	6.37	4.34	3.17
200	-0.25	0.08	-0.02	0.07	-0.02	19.20	11.62	6.51	4.27	3.04
Infeasible with p_f selected by AIC										
40	0.37	0.26	0.13	0.09	0.00	27.77	13.53	6.84	4.28	3.22
60	-0.86	0.00	-0.03	0.02	0.07	46.82	13.06	6.74	4.40	3.26
100	-0.13	0.12	0.12	-0.03	0.03	25.32	12.62	6.96	4.31	3.09
150	0.59	0.05	0.03	0.05	0.01	24.69	13.33	6.73	4.49	3.23
200	0.39	0.27	0.04	0.08	0.00	29.03	12.98	6.88	4.37	3.10
KCCE (Static)										
40	71.18	60.27	48.40	39.45	33.79	82.90	70.54	57.00	47.08	40.51
60	57.32	47.44	36.97	30.22	26.02	68.86	57.45	45.01	37.27	32.04
100	61.76	54.08	42.34	34.96	29.36	72.23	63.12	50.36	41.80	35.41
150	61.82	52.89	42.20	34.70	29.42	71.70	61.38	49.74	41.28	35.10
200	61.18	52.80	41.77	34.18	28.92	71.34	61.93	49.51	40.78	34.87

Table 3.E1.A cont'd: β_{xi} , weak exogeneity assumption is satisfied, homogeneous long-run

N/T	Size					Power				
	40	60	100	150	200	40	60	100	150	200
DCCE with $p_f = p_T$										
40	7.76	6.18	6.48	5.88	5.40	17.26	26.16	46.48	69.62	83.46
60	7.06	6.46	6.18	6.02	5.22	17.00	26.94	46.90	67.22	83.32
100	6.20	6.12	5.86	5.50	5.10	16.24	25.76	46.08	66.54	82.62
150	7.56	6.56	5.20	4.96	5.10	16.46	24.22	43.58	65.72	81.84
200	6.70	6.28	6.06	6.10	5.18	14.30	23.20	44.18	67.98	83.26
DCCE with $p_f = 1$										
40	8.64	7.06	7.76	7.38	6.74	22.52	32.40	54.46	75.70	87.88
60	6.54	6.52	6.36	6.28	6.26	20.10	30.88	53.10	72.84	86.46
100	6.24	6.38	6.24	5.62	5.22	18.56	29.32	51.52	71.08	85.56
150	6.64	5.64	5.16	5.30	5.60	17.90	26.10	47.26	69.52	84.20
200	5.96	5.34	5.24	5.52	5.14	15.42	25.14	47.78	70.08	84.54
DCCE with p_f selected by AIC										
40	13.50	9.44	7.70	6.60	6.32	26.46	32.08	51.70	72.32	84.92
60	12.50	9.84	8.10	6.90	6.50	24.70	32.60	52.56	71.44	84.80
100	12.10	9.48	8.16	6.86	6.02	24.42	31.42	51.54	70.66	85.10
150	13.46	9.28	7.50	6.86	6.80	24.68	30.10	49.42	71.14	84.26
200	12.76	9.32	7.82	7.76	6.76	22.84	29.72	49.98	71.88	85.26
Infeasible with $p_f = 1$										
40	6.22	6.70	6.02	5.56	5.08	18.70	29.42	52.34	74.08	86.56
60	6.42	6.24	5.38	5.92	5.68	19.00	29.54	51.62	73.40	87.10
100	6.78	6.06	5.80	5.58	4.92	18.90	29.24	52.34	71.62	86.44
150	7.44	6.14	5.46	5.50	5.58	19.62	29.36	52.28	73.78	86.48
200	6.76	5.64	6.04	5.44	4.48	18.80	29.54	51.48	74.42	87.06
Infeasible with p_f selected by AIC										
40	13.98	9.96	7.30	6.32	5.54	27.08	33.02	52.88	73.90	86.44
60	14.08	9.80	6.40	7.08	5.94	26.76	32.50	52.20	73.46	87.02
100	13.84	9.56	7.14	6.24	5.54	26.08	32.36	53.14	71.60	86.34
150	15.40	9.26	7.22	6.36	6.08	27.52	32.56	52.62	73.72	86.46
200	14.64	8.92	7.34	6.14	5.04	26.32	33.12	51.80	74.30	87.14
KCCE (Static)										
40	91.04	92.54	93.30	94.62	94.86	94.98	96.22	98.20	98.88	99.28
60	88.88	89.82	90.94	91.54	93.30	93.52	95.66	97.56	98.44	99.08
100	90.70	92.92	92.96	93.56	94.60	94.62	96.82	97.74	98.72	99.32
150	91.64	93.06	93.54	93.96	93.62	94.86	96.56	97.48	98.54	98.90
200	90.62	91.46	92.00	92.92	92.02	94.08	95.30	96.54	97.62	98.26

Notes: “DCCE” and “Infeasible” refer to the estimators proposed in (3.36) which is based on the model (3.28) and (3.13) which is based on the model (3.10). p_f is the lag order for the first differences of the factors (cross-sectional averages for DCCE and true factors for the infeasible estimator). p_T is the truncation lag order and is computed by using the formula $\lfloor T^{1/3} \rfloor$ i.e. p_T is equal to 3, 3, 4, 5, 5 respectively for T values 40, 60, 100, 150, 200. KCCE (Static) refers to the estimator proposed in Kapetanios et al. (2011).

3 A CCE ESTIMATOR FOR DYNAMIC PANEL ECM'S

Table 3.E1.B : β_{xMG} , weak exogeneity assumption is satisfied, homogeneous long-run

N/T	Bias $\times 100$					RMSE				
	40	60	100	150	200	40	60	100	150	200
DCCE with $p_f = p_T$										
40	0.04	0.23	0.08	0.04	0.03	5.90	1.85	1.00	0.65	0.47
60	0.04	0.17	0.10	0.04	0.03	4.74	1.63	0.89	0.56	0.41
100	-0.12	0.19	0.10	0.06	0.04	9.96	1.25	0.67	0.43	0.30
150	1.94	0.24	0.10	0.07	0.04	138.39	1.09	0.56	0.36	0.26
200	-0.06	0.23	0.10	0.07	0.04	21.11	0.94	0.51	0.32	0.22
DCCE with $p_f = 1$										
40	0.11	0.19	0.10	0.05	0.04	2.84	1.65	0.92	0.61	0.45
60	-0.10	0.02	0.02	0.01	0.00	2.42	1.43	0.80	0.51	0.39
100	-0.04	0.02	0.02	0.01	0.01	1.85	1.08	0.61	0.40	0.29
150	-0.06	0.10	0.07	0.05	0.03	1.69	0.92	0.51	0.33	0.25
200	-0.05	0.08	0.07	0.04	0.03	1.84	0.80	0.45	0.29	0.21
DCCE with p_f selected by AIC										
40	0.34	0.22	0.06	0.02	0.03	12.26	1.88	1.00	0.64	0.46
60	-0.28	0.14	0.04	0.02	0.01	31.40	1.59	0.86	0.53	0.40
100	-0.32	0.02	0.01	0.00	0.01	19.52	1.21	0.65	0.42	0.30
150	-0.35	-0.03	0.02	0.02	0.02	12.13	1.05	0.54	0.35	0.26
200	-0.28	-0.03	0.01	0.01	0.01	5.75	0.89	0.48	0.30	0.21
Infeasible with $p_f = 1$										
40	-0.26	-0.07	0.00	-0.02	-0.01	2.89	1.66	0.90	0.59	0.44
60	-0.50	-0.07	-0.01	-0.01	0.00	16.27	1.44	0.79	0.50	0.38
100	-0.26	-0.08	-0.01	0.00	0.00	1.87	1.08	0.60	0.39	0.28
150	-0.40	-0.10	-0.02	-0.01	0.00	1.69	0.88	0.49	0.31	0.23
200	-0.37	-0.10	-0.02	-0.01	0.00	1.79	0.77	0.43	0.27	0.20
Infeasible with p_f selected by AIC										
40	1.90	0.03	0.03	0.00	0.00	35.18	1.82	0.95	0.61	0.45
60	-0.18	0.02	0.02	0.00	0.01	15.46	1.62	0.84	0.52	0.39
100	-0.37	-0.05	0.00	0.00	0.00	5.38	1.17	0.62	0.41	0.29
150	0.16	-0.05	0.00	0.00	0.00	3.04	0.99	0.51	0.33	0.24
200	-0.44	-0.05	0.01	0.00	0.00	4.78	0.84	0.45	0.28	0.20
KCCE (Static)										
40	42.47	37.87	32.55	28.52	25.94	44.02	39.10	33.55	29.40	26.75
60	44.99	40.65	35.51	31.76	29.24	46.11	41.64	36.27	32.42	29.82
100	39.99	36.26	31.24	27.62	24.97	40.95	37.04	31.89	28.17	25.46
150	38.37	34.26	29.26	25.72	23.34	39.21	35.00	29.89	26.25	23.80
200	36.61	32.36	27.47	23.88	21.42	37.51	33.11	28.09	24.37	21.87

Notes: See Table 3.E1.A for an explanation.

Table 3.E1.B cont'd: β_{xMG} , weak exogeneity assumption is satisfied, homogeneous long-run

N/T	Size					Power				
	40	60	100	150	200	40	60	100	150	200
	DCCE with $p_f = p_T$									
40	5.84	6.22	5.50	6.12	5.46	81.76	99.28	100.0	100.0	100.0
60	5.54	5.42	5.88	5.30	5.72	87.20	99.78	100.0	100.0	100.0
100	6.02	5.22	5.14	5.50	4.84	95.40	99.98	100.0	100.0	100.0
150	5.42	6.32	5.32	5.76	5.04	96.52	100.0	100.0	100.0	100.0
200	5.58	6.76	6.82	6.08	5.44	97.84	100.0	100.0	100.0	100.0
	DCCE with $p_f = 1$									
40	5.80	5.94	5.86	5.52	5.82	92.10	99.70	100.0	100.0	100.0
60	5.42	5.40	5.20	5.20	5.44	95.24	99.86	100.0	100.0	100.0
100	5.48	4.92	4.82	5.22	5.04	98.80	100.0	100.0	100.0	100.0
150	5.14	5.46	4.82	5.04	5.24	99.30	100.0	100.0	100.0	100.0
200	5.28	5.36	6.18	5.66	5.58	99.70	100.0	100.0	100.0	100.0
	DCCE with p_f selected by AIC									
40	5.78	6.20	5.74	6.38	5.78	81.12	99.18	100.0	100.0	100.0
60	5.20	5.10	5.32	4.94	5.52	85.78	99.72	99.96	100.0	100.0
100	5.38	5.18	5.02	5.48	4.56	94.04	99.98	100.0	100.0	100.0
150	4.94	5.24	4.40	4.94	5.00	94.90	99.94	100.0	100.0	100.0
200	5.26	5.26	5.56	5.54	4.88	96.04	99.98	100.0	100.0	100.0
	Infeasible with $p_f = 1$									
40	5.66	6.06	5.62	5.48	5.56	89.58	99.64	100.0	100.0	100.0
60	5.24	6.02	5.08	4.86	5.54	93.66	99.88	100.0	100.0	100.0
100	5.06	5.44	4.58	5.22	4.76	98.76	100.0	100.0	100.0	100.0
150	5.20	4.72	4.62	4.80	4.88	99.42	100.0	100.0	100.0	100.0
200	5.48	5.34	5.20	5.16	4.74	99.68	100.0	100.0	100.0	100.0
	Infeasible with p_f selected by AIC									
40	5.32	5.84	5.96	5.44	5.66	80.80	99.22	100.0	100.0	100.0
60	4.18	5.28	5.38	4.70	5.56	85.88	99.74	100.0	100.0	100.0
100	4.80	4.90	4.36	5.64	5.20	94.28	99.92	100.0	100.0	100.0
150	4.44	4.52	4.22	4.96	5.02	95.16	99.94	100.0	100.0	100.0
200	5.06	4.88	5.48	5.26	4.82	96.76	99.94	100.0	100.0	100.0
	KCCE (Static)									
40	98.06	98.28	97.28	96.48	95.30	99.82	99.94	99.92	99.94	99.96
60	99.92	99.92	99.90	99.96	99.82	100.0	100.0	100.0	100.0	100.0
100	100.0	100.0	100.0	100.0	99.98	100.0	100.0	100.0	100.0	100.0
150	99.98	100.0	100.0	99.98	100.0	100.0	100.0	100.0	100.0	100.0
200	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0

Notes: See Table 3.E1.A for an explanation.

3 A CCE ESTIMATOR FOR DYNAMIC PANEL ECM'S

Table 3.E2.A: β_{xi} , weak exogeneity assumption is satisfied, $\beta_{xi} = -1$, $\lambda_{ij} \sim U[-2.1, -1.9]$ for all i and $j = 1, 2$

N/T	Bias $\times 100$					RMSE				
	40	60	100	150	200	40	60	100	150	200
DCCE with $p_f = p_T$										
40	0.42	0.12	0.11	-0.02	-0.03	4.07	1.86	0.89	0.52	0.35
60	0.25	0.03	0.13	-0.02	-0.01	3.49	1.62	0.82	0.46	0.32
100	0.29	0.07	0.12	0.02	0.01	3.19	1.41	0.73	0.40	0.28
150	0.27	0.07	0.06	0.00	0.00	2.92	1.30	0.64	0.38	0.26
200	0.22	0.10	0.05	0.00	0.00	2.88	1.34	0.62	0.37	0.24
DCCE with $p_f = 1$										
40	-0.28	-0.27	-0.16	-0.11	-0.09	2.73	1.49	0.76	0.46	0.33
60	-0.02	-0.07	-0.05	-0.01	-0.01	2.44	1.27	0.67	0.40	0.29
100	0.01	-0.02	0.00	0.01	0.01	2.19	1.11	0.58	0.35	0.26
150	0.02	-0.03	-0.02	-0.02	-0.01	1.96	1.04	0.53	0.33	0.23
200	-0.06	-0.03	-0.03	-0.03	-0.02	1.99	1.04	0.52	0.32	0.23
DCCE with p_f selected by AIC										
40	0.25	0.00	0.00	0.01	-0.01	4.38	1.73	0.82	0.49	0.34
60	0.29	0.06	0.02	0.03	0.03	3.39	1.50	0.76	0.44	0.31
100	0.32	0.07	0.04	0.03	0.02	3.08	1.30	0.66	0.38	0.27
150	0.32	0.07	0.02	0.01	0.01	2.86	1.20	0.58	0.35	0.24
200	0.17	0.09	0.03	0.01	0.01	5.72	1.23	0.57	0.35	0.23
Infeasible with $p_f = 1$										
40	0.15	0.05	0.02	0.01	0.00	1.68	0.90	0.43	0.26	0.19
60	0.16	0.05	0.02	0.01	0.00	1.66	0.88	0.43	0.26	0.19
100	0.17	0.05	0.02	0.00	0.01	1.69	0.87	0.44	0.26	0.18
150	0.20	0.05	0.02	0.01	0.00	1.63	0.90	0.43	0.26	0.19
200	0.13	0.07	0.02	0.01	0.00	1.67	0.87	0.42	0.26	0.18
Infeasible with p_f selected by AIC										
40	0.23	0.11	0.02	0.01	0.00	2.14	1.04	0.46	0.27	0.19
60	0.34	0.08	0.02	0.01	0.01	3.26	1.00	0.46	0.27	0.19
100	0.34	0.07	0.03	0.01	0.01	4.80	0.98	0.48	0.27	0.19
150	0.33	0.08	0.02	0.01	0.00	2.35	1.03	0.46	0.27	0.19
200	0.22	0.10	0.03	0.01	0.00	2.23	0.99	0.45	0.28	0.18
KCCE (Static)										
40	3.78	2.83	1.71	1.09	0.73	7.35	6.06	4.88	4.13	3.78
60	2.44	1.82	1.12	0.58	0.36	6.39	5.21	4.13	3.57	3.31
100	2.03	1.12	0.58	0.20	0.12	5.81	4.42	3.66	3.28	2.90
150	0.98	0.44	-0.09	-0.45	-0.54	5.39	4.47	3.57	3.10	2.82
200	1.78	0.76	-0.02	-0.41	-0.64	6.28	4.67	3.71	3.20	2.92

Notes: See Table 3.E1.A for an explanation.

Table 3.E2.A cont'd: β_{xi} , weak exogeneity assumption is satisfied, $\beta_{xi} = -1$, $\lambda_{ij} \sim U[-2.1, -1.9]$ for all i and $j = 1, 2$

N/T	Size					Power				
	40	60	100	150	200	40	60	100	150	200
DCCE with $p_f = p_T$										
40	4.66	4.00	4.82	5.46	5.30	82.88	98.38	100.0	100.0	100.0
60	4.76	3.90	4.58	4.80	4.96	86.56	98.62	100.0	100.0	100.0
100	5.16	4.78	5.28	5.32	5.72	89.78	99.38	100.0	100.0	100.0
150	5.78	4.26	5.26	5.40	5.40	92.98	99.54	99.96	100.0	100.0
200	5.72	4.76	4.92	6.20	5.60	91.88	99.60	100.0	100.0	100.0
DCCE with $p_f = 1$										
40	5.30	5.46	4.80	4.88	5.14	91.46	99.10	100.0	100.0	100.0
60	5.00	3.62	3.68	4.28	3.66	94.68	99.46	100.0	100.0	100.0
100	4.94	4.88	5.26	5.48	6.24	96.16	99.82	100.0	100.0	100.0
150	5.86	5.12	5.48	6.24	6.00	97.54	99.86	100.0	100.0	100.0
200	6.18	6.06	6.14	6.74	6.50	96.78	99.88	100.0	100.0	100.0
DCCE with p_f selected by AIC										
40	11.70	9.02	8.90	8.18	7.50	92.00	99.22	100.0	100.0	100.0
60	12.88	8.74	8.12	7.88	7.58	94.64	99.48	100.0	100.0	100.0
100	12.84	9.20	8.00	7.46	7.88	95.76	99.66	100.0	100.0	100.0
150	12.86	9.20	7.86	7.46	7.00	97.18	99.84	100.0	100.0	100.0
200	13.26	9.40	7.70	8.02	7.22	96.36	99.82	100.0	100.0	100.0
Infeasible with $p_f = 1$										
40	6.36	5.74	5.06	5.10	5.12	99.28	99.98	100.0	100.0	100.0
60	5.98	5.56	5.02	5.24	5.12	99.38	100.0	100.0	100.0	100.0
100	6.22	5.30	5.68	5.16	4.28	99.14	99.96	100.0	100.0	100.0
150	5.60	5.50	5.22	5.38	5.52	99.46	99.98	100.0	100.0	100.0
200	5.76	5.34	4.98	5.14	4.52	99.04	99.92	100.0	100.0	100.0
Infeasible with p_f selected by AIC										
40	12.74	9.42	6.92	5.84	5.68	98.90	99.98	100.0	100.0	100.0
60	12.48	8.90	6.96	6.24	5.76	98.94	99.98	100.0	100.0	100.0
100	12.22	9.14	7.90	6.08	5.10	98.66	99.96	100.0	100.0	100.0
150	12.68	9.16	7.58	6.44	6.14	99.10	99.94	100.0	100.0	100.0
200	12.12	9.16	7.16	6.38	5.20	99.06	99.90	100.0	100.0	100.0
KCCE (Static)										
40	51.94	56.36	61.58	68.04	70.66	80.34	83.74	85.14	87.02	89.06
60	50.54	56.42	62.30	68.20	70.46	79.58	83.04	86.26	88.74	90.24
100	48.24	53.20	61.44	68.50	70.44	80.26	83.14	87.08	89.18	92.54
150	47.48	55.72	63.56	69.88	72.96	75.66	79.44	84.78	87.70	90.36
200	50.12	55.50	63.66	69.66	73.74	77.30	79.36	83.54	87.36	89.96

Notes: See Table 3.E1.A for an explanation.

3 A CCE ESTIMATOR FOR DYNAMIC PANEL ECM'S

Table 3.E2.B: β_{xMG} , weak exogeneity assumption is satisfied, $\beta_{xi} = -1$, $\lambda_{ij} \sim U[-2.1, -1.9]$ for all i and $j = 1, 2$

N/T	Bias $\times 100$					RMSE				
	40	60	100	150	200	40	60	100	150	200
DCCE with $p_f = p_T$										
40	-0.86	-0.07	-0.05	-0.13	-0.14	9.68	2.66	1.43	1.13	1.15
60	-0.04	0.22	0.08	0.08	0.07	8.74	1.80	1.17	0.90	0.78
100	0.02	0.05	0.01	0.05	0.04	10.67	1.43	0.91	0.65	0.56
150	-0.32	-0.01	0.00	-0.01	-0.01	6.06	1.05	0.66	0.55	0.43
200	-0.32	0.03	0.02	0.02	0.01	2.58	0.97	0.58	0.43	0.44
DCCE with $p_f = 1$										
40	-0.22	0.01	-0.03	-0.07	-0.09	5.98	1.99	1.37	1.11	1.03
60	-0.09	0.19	0.13	0.09	0.09	5.98	1.61	1.15	0.90	0.78
100	-0.06	0.05	0.06	0.08	0.06	4.87	2.53	1.18	0.66	0.58
150	-0.06	0.07	0.05	0.04	0.04	1.71	0.96	0.65	0.52	0.44
200	-0.13	0.10	0.09	0.08	0.07	7.41	0.84	0.57	0.44	0.39
DCCE with p_f selected by AIC										
40	-1.65	-0.45	-0.12	-0.13	-0.12	20.56	8.79	1.58	1.16	2.16
60	-0.55	-0.05	0.04	0.00	0.00	8.29	2.14	2.94	1.01	0.85
100	-0.50	-0.12	-0.01	0.01	0.01	12.76	1.62	0.89	0.69	0.59
150	0.33	-0.14	-0.05	-0.03	-0.02	61.71	1.16	0.68	0.50	0.44
200	-0.44	-0.14	-0.05	0.00	0.00	7.57	0.95	1.31	0.44	0.37
Infeasible with $p_f = 1$										
40	-0.63	-0.22	-0.05	-0.04	-0.02	3.06	1.68	0.91	0.59	0.44
60	-0.27	-0.13	-0.04	-0.02	0.00	15.63	1.37	0.75	0.48	0.36
100	-0.48	-0.16	-0.04	-0.01	0.00	2.46	1.06	0.59	0.38	0.28
150	-0.50	-0.11	-0.02	-0.02	-0.01	4.70	0.84	0.47	0.30	0.22
200	-0.38	-0.12	-0.03	-0.01	0.00	1.41	0.73	0.40	0.26	0.19
Infeasible with p_f selected by AIC										
40	-0.64	-0.21	-0.04	-0.03	-0.02	9.40	1.87	0.96	0.60	0.45
60	-0.40	-0.07	-0.01	-0.01	0.01	14.32	1.51	0.80	0.49	0.37
100	-0.49	-0.08	-0.01	0.01	0.00	9.11	1.15	0.62	0.39	0.28
150	-0.47	-0.02	0.01	0.00	0.00	10.79	0.93	0.49	0.31	0.22
200	-0.13	-0.06	0.00	0.00	0.00	30.49	0.80	0.43	0.26	0.19
KCCE (Static)										
40	35.45	29.99	23.67	19.47	16.82	36.71	31.04	24.50	20.12	17.40
60	40.65	35.53	29.59	25.31	22.44	41.87	36.56	30.42	26.02	23.06
100	36.46	31.87	26.67	22.81	20.28	37.52	32.75	27.36	23.40	20.79
150	35.06	30.67	25.43	21.84	19.53	36.04	31.47	26.10	22.40	20.01
200	34.11	29.80	24.74	21.07	18.63	34.97	30.57	25.36	21.57	19.08

Notes: See Table 3.E1.A for an explanation.

Table 3.E2.B cont'd: β_{xMG} , weak exogeneity assumption is satisfied, $\beta_{xi} = -1$, $\lambda_{ij} \sim U[-2.1, -1.9]$ for all i and $j = 1, 2$

N/T	Size					Power				
	40	60	100	150	200	40	60	100	150	200
DCCE with $p_f = p_T$										
40	4.58	5.44	5.16	4.90	4.98	73.02	97.36	99.46	99.58	99.58
60	5.60	5.74	5.18	5.06	5.48	84.90	99.16	99.74	99.80	99.88
100	5.48	5.26	5.28	5.48	5.02	93.70	99.82	99.86	99.94	100.0
150	5.24	4.82	4.62	4.90	4.80	97.08	99.94	99.92	99.96	100.0
200	5.48	5.36	5.72	5.14	4.98	97.96	99.92	100.0	99.96	99.98
DCCE with $p_f = 1$										
40	4.90	5.84	5.38	4.62	4.76	85.96	98.48	99.26	99.42	99.60
60	5.62	5.78	5.64	4.94	5.04	93.38	99.42	99.62	99.76	99.88
100	5.56	5.30	5.40	5.74	5.14	98.12	99.80	99.82	99.96	99.96
150	5.64	5.00	5.42	4.74	5.26	99.36	99.94	99.92	99.96	100.0
200	5.40	5.76	6.28	6.02	5.72	99.52	99.96	99.96	99.98	99.98
DCCE with p_f selected by AIC										
40	4.66	5.44	4.76	4.74	4.82	69.52	96.36	99.18	99.34	99.44
60	4.92	5.26	5.04	4.66	4.66	80.32	98.76	99.60	99.56	99.86
100	4.98	5.20	5.32	5.18	4.40	91.28	99.58	99.80	99.82	99.96
150	5.00	5.08	4.72	4.28	4.44	95.00	99.74	99.90	99.96	100.0
200	5.14	5.24	5.42	5.10	4.46	95.20	99.88	99.86	99.98	100.0
Infeasible with $p_f = 1$										
40	5.34	5.32	5.50	5.58	5.76	87.16	99.42	100.0	100.0	100.0
60	5.28	5.46	5.12	4.74	5.60	94.40	99.90	100.0	100.0	100.0
100	5.34	5.30	4.86	5.12	4.92	98.52	100.0	100.0	100.0	100.0
150	4.98	4.44	4.98	4.82	5.00	99.42	100.0	100.0	100.0	100.0
200	5.90	5.02	5.24	4.54	4.98	99.54	100.0	100.0	100.0	100.0
Infeasible with p_f selected by AIC										
40	5.28	5.48	5.80	5.72	6.00	76.30	98.72	99.98	100.0	100.0
60	4.62	5.46	5.16	5.10	5.80	86.68	99.70	100.0	100.0	100.0
100	5.24	4.74	4.62	5.56	4.86	93.70	99.96	100.0	100.0	100.0
150	4.50	4.80	4.62	5.06	5.04	95.90	99.94	100.0	100.0	100.0
200	5.00	4.70	5.26	4.92	4.98	97.06	99.98	100.0	100.0	100.0
KCCE (Static)										
40	99.40	99.60	99.44	99.16	99.00	100.0	100.0	100.0	100.0	100.0
60	99.96	99.98	99.92	99.80	99.80	100.0	100.0	100.0	100.0	100.0
100	100.0	100.0	100.0	100.0	99.96	100.0	100.0	100.0	100.0	100.0
150	100.0	100.0	99.98	99.98	100.0	100.0	100.0	100.0	100.0	100.0
200	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0

Notes: See Table 3.E1.A for an explanation.

3 A CCE ESTIMATOR FOR DYNAMIC PANEL ECM'S

Table 3.E3 : β_{xMG} , weak exogeneity assumption is satisfied, $\beta_{xi} \sim U[-1.1, -1.0]$, $\lambda_{ij} \sim U[-2.1, -1.9]$ for all i and $j = 1, 2$

N/T	Bias $\times 100$					RMSE				
	40	60	100	150	200	40	60	100	150	200
DCCE with $p_f = p_T$										
40	-0.33	-0.10	-0.22	-0.21	-0.21	6.12	2.01	1.34	1.08	0.94
60	-0.23	0.04	-0.02	-0.01	-0.04	3.28	1.63	1.01	0.77	0.69
100	-0.08	0.05	-0.05	-0.03	-0.03	2.52	1.24	0.75	0.54	0.46
150	-0.33	-0.07	-0.10	-0.06	-0.07	2.74	1.01	0.59	0.43	0.36
200	0.30	-0.08	-0.12	-0.07	-0.08	46.92	0.87	0.53	0.38	0.32
DCCE with $p_f = 1$										
40	-0.50	-0.28	-0.26	-0.26	-0.24	3.04	1.84	1.26	1.04	0.91
60	-0.33	-0.10	-0.04	-0.02	-0.03	2.96	1.45	0.94	0.74	0.66
100	-0.12	-0.07	-0.06	-0.05	-0.04	1.82	1.11	0.69	0.52	0.42
150	-0.17	-0.12	-0.09	-0.10	-0.09	5.77	0.93	0.54	0.41	0.35
200	-0.30	-0.14	-0.12	-0.11	-0.11	1.28	0.78	0.49	0.37	0.32
DCCE with p_f selected by AIC										
40	1.91	-0.37	-0.30	-0.29	-0.27	70.52	3.32	1.37	1.12	0.96
60	-0.05	-0.19	-0.08	-0.05	-0.06	51.79	1.75	1.04	0.79	0.70
100	-0.43	-0.12	-0.07	-0.07	-0.05	7.08	1.23	1.15	0.55	0.44
150	-0.91	-0.16	-0.12	-0.12	-0.11	31.33	1.19	0.58	0.43	0.37
200	-0.27	-0.16	-0.14	-0.12	-0.11	7.21	0.96	0.53	0.39	0.33
Infeasible with $p_f = 1$										
40	-0.27	-0.08	-0.01	-0.02	-0.01	2.86	1.59	0.89	0.57	0.42
60	-0.38	-0.09	-0.02	-0.01	0.00	2.28	1.31	0.71	0.45	0.34
100	-0.26	-0.08	-0.02	0.00	0.00	1.77	1.01	0.56	0.36	0.26
150	-0.27	-0.06	0.00	-0.01	0.00	1.45	0.81	0.46	0.30	0.22
200	-0.29	-0.08	-0.01	-0.01	0.00	1.38	0.71	0.40	0.25	0.19
Infeasible with p_f selected by AIC										
40	-0.27	0.02	0.03	0.00	0.00	6.96	1.75	0.93	0.59	0.43
60	-0.46	-0.02	0.01	0.00	0.01	4.88	1.41	0.75	0.47	0.35
100	-0.31	-0.01	0.01	0.01	0.01	11.43	1.09	0.59	0.38	0.27
150	-0.24	0.01	0.03	0.01	0.01	2.97	0.89	0.48	0.30	0.22
200	-0.19	-0.01	0.02	0.01	0.00	4.54	0.76	0.42	0.26	0.19
KCCE (Static)										
40	29.11	25.49	21.50	18.27	16.51	30.18	26.36	22.19	18.86	17.02
60	30.29	26.54	22.16	18.95	16.83	31.11	27.25	22.76	19.45	17.27
100	33.20	29.21	24.18	20.37	17.95	34.12	30.02	24.86	20.94	18.45
150	32.48	28.12	22.95	19.46	17.11	33.38	28.93	23.62	19.99	17.57
200	32.98	28.59	23.57	19.98	17.60	33.88	29.35	24.20	20.49	18.05

Notes: See Table 3.E1.A for an explanation.

Table 3.E3 cont'd: β_{xMG} , weak exogeneity assumption is satisfied, $\beta_{xi} \sim U[-1.1, -1.0]$, $\lambda_{ij} \sim U[-2.1, -1.9]$ for all i and $j = 1, 2$

N/T	Size					Power				
	40	60	100	150	200	40	60	100	150	200
DCCE with $p_f = p_T$										
40	4.18	2.70	1.34	0.68	0.50	76.42	97.82	99.48	99.58	99.78
60	4.36	2.50	1.26	0.60	0.30	87.14	99.56	99.90	99.92	99.88
100	4.10	2.74	1.34	0.38	0.16	95.74	99.96	100.0	100.0	99.98
150	4.38	2.64	1.26	0.52	0.16	98.20	99.94	100.0	100.0	100.0
200	4.36	3.06	1.18	0.36	0.34	98.50	99.98	100.0	100.0	100.0
DCCE with $p_f = 1$										
40	3.88	2.40	1.16	0.58	0.44	86.52	98.60	99.56	99.56	99.90
60	3.92	2.58	1.02	0.48	0.28	94.96	99.86	99.96	99.92	99.94
100	4.02	2.86	0.82	0.24	0.10	99.18	100.0	100.0	100.0	100.0
150	4.14	2.20	0.90	0.30	0.12	99.60	99.96	100.0	100.0	100.0
200	4.26	2.26	1.30	0.42	0.36	99.90	99.96	100.0	100.0	100.0
DCCE with p_f selected by AIC										
40	4.38	2.92	1.22	0.64	0.44	73.70	96.58	99.34	99.54	99.70
60	3.72	2.56	1.24	0.56	0.40	83.80	99.20	99.86	99.92	99.86
100	4.00	2.82	1.10	0.42	0.12	93.16	99.90	99.98	99.96	99.96
150	4.24	2.66	1.16	0.56	0.12	95.68	99.92	100.0	100.0	100.0
200	4.08	2.80	1.54	0.32	0.34	96.54	99.90	100.0	100.0	100.0
Infeasible with $p_f = 1$										
40	3.86	2.18	0.50	0.04	0.00	89.14	99.66	100.0	100.0	100.0
60	3.42	2.22	0.24	0.02	0.00	95.42	99.96	100.0	100.0	100.0
100	4.04	2.44	0.40	0.08	0.00	99.16	100.0	100.0	100.0	100.0
150	3.50	1.78	0.68	0.02	0.00	99.68	100.0	100.0	100.0	100.0
200	3.90	1.90	0.70	0.00	0.00	99.84	100.0	100.0	100.0	100.0
Infeasible with p_f selected by AIC										
40	3.88	2.78	0.70	0.06	0.00	80.82	99.22	100.0	100.0	100.0
60	3.08	2.62	0.46	0.02	0.00	88.04	99.82	100.0	100.0	100.0
100	4.16	2.18	0.64	0.06	0.00	95.18	100.0	100.0	100.0	100.0
150	4.04	2.28	0.62	0.06	0.00	97.14	99.98	100.0	100.0	100.0
200	4.22	2.18	0.86	0.00	0.00	97.96	100.0	100.0	100.0	100.0
KCCE (Static)										
40	95.42	95.54	95.90	94.38	93.76	99.68	99.78	99.96	99.98	100.0
60	99.58	99.48	99.50	99.32	99.38	100.0	100.0	100.0	100.0	100.0
100	99.60	99.36	99.04	98.84	97.82	99.98	99.94	100.0	100.0	100.0
150	99.68	99.46	99.30	99.08	98.76	100.0	100.0	100.0	100.0	100.0
200	99.96	99.94	99.84	99.88	99.86	100.0	100.0	100.0	100.0	100.0

Notes: See Table 3.E1.A for an explanation.

3 A CCE ESTIMATOR FOR DYNAMIC PANEL ECM'S

Table 3.E4 : $\beta_{xMG}, \alpha_{xi} \sim U[-0.15, 0.15]$, homogeneous long-run

N/T	Bias $\times 100$					RMSE				
	40	60	100	150	200	40	60	100	150	200
DCCE with $p_f = p_T$										
40	-0.17	0.14	0.16	0.09	0.08	4.33	2.03	1.10	0.69	0.50
60	-0.22	0.06	0.08	0.04	0.03	6.07	1.58	0.85	0.55	0.39
100	-0.38	-0.03	-0.02	0.00	0.00	3.67	1.23	0.68	0.43	0.30
150	-0.44	0.03	0.02	0.01	0.01	3.56	0.97	0.53	0.34	0.25
200	-0.43	-0.04	-0.03	0.00	-0.01	2.32	0.88	0.47	0.30	0.21
DCCE with $p_f = 1$										
40	0.19	0.22	0.18	0.10	0.09	3.00	1.74	1.00	0.63	0.48
60	-0.12	0.00	0.01	0.01	0.00	2.42	1.37	0.76	0.50	0.37
100	-0.19	-0.03	0.00	0.00	0.00	1.85	1.07	0.61	0.39	0.29
150	-0.14	0.01	0.03	0.00	0.01	1.50	0.86	0.48	0.31	0.23
200	-0.30	-0.09	-0.06	-0.05	-0.04	2.57	0.78	0.43	0.29	0.21
DCCE with p_f selected by AIC										
40	0.39	0.04	0.14	0.07	0.08	48.30	2.03	1.09	0.67	0.50
60	-0.27	-0.17	-0.04	-0.02	-0.01	10.33	1.57	0.83	0.53	0.39
100	-0.45	-0.19	-0.08	-0.04	-0.03	7.81	1.22	0.67	0.42	0.30
150	-0.31	-0.11	-0.04	-0.03	-0.01	11.76	0.98	0.53	0.33	0.24
200	-0.67	-0.14	-0.07	-0.04	-0.03	13.31	0.89	0.47	0.30	0.21
Infeasible with $p_f = 1$										
40	-0.13	0.06	0.10	0.06	0.06	2.99	1.76	0.95	0.62	0.47
60	-0.34	-0.10	-0.03	-0.01	0.00	2.30	1.37	0.75	0.48	0.36
100	-0.49	-0.20	-0.08	-0.04	-0.03	2.19	1.07	0.58	0.38	0.28
150	-0.47	-0.12	-0.03	-0.02	-0.01	5.45	0.84	0.48	0.30	0.23
200	-0.45	-0.18	-0.07	-0.04	-0.02	1.45	0.77	0.42	0.27	0.20
Infeasible with p_f selected by AIC										
40	0.01	0.13	0.13	0.08	0.06	13.69	2.22	1.00	0.64	0.47
60	-2.30	0.01	0.02	0.01	0.01	129.16	1.50	0.80	0.50	0.37
100	-0.54	-0.11	-0.05	-0.02	-0.02	7.18	1.14	0.61	0.39	0.28
150	-0.30	-0.01	0.01	0.00	0.00	7.00	0.92	0.50	0.31	0.23
200	-1.67	-0.09	-0.03	-0.02	-0.02	98.70	0.84	0.44	0.28	0.20
KCCE (Static)										
40	40.32	35.72	30.55	26.39	23.83	41.58	36.82	31.42	27.16	24.51
60	34.27	30.19	25.31	21.86	19.46	35.34	31.10	26.03	22.50	20.01
100	37.05	33.39	28.44	24.81	22.35	37.92	34.10	29.04	25.34	22.82
150	36.14	32.06	26.93	23.51	21.19	37.05	32.81	27.56	24.03	21.64
200	36.41	32.29	27.34	23.83	21.46	37.30	33.05	27.98	24.35	21.91

Notes: See Table 3.E1.A for an explanation.

Table 3.E4 cont'd: $\beta_{xMG}, \alpha_{xi} \sim U[-0.15, 0.15]$, homogeneous long-run

N/T	Size					Power				
	40	60	100	150	200	40	60	100	150	200
	DCCE with $p_f = p_T$									
40	5.76	5.24	5.26	5.62	5.00	75.96	98.28	100.0	100.0	100.0
60	4.42	5.18	5.00	5.44	5.46	85.90	99.72	100.0	100.0	100.0
100	4.92	4.82	5.18	4.86	4.58	93.30	99.96	100.0	100.0	100.0
150	5.46	4.86	4.60	5.26	4.90	97.46	100.0	100.0	100.0	100.0
200	4.90	5.50	5.06	4.96	4.72	98.12	100.0	100.0	100.0	100.0
	DCCE with $p_f = 1$									
40	5.66	5.56	5.52	5.26	5.66	89.60	99.40	100.0	100.0	100.0
60	5.26	5.16	4.64	4.94	5.16	95.84	99.94	100.0	100.0	100.0
100	4.74	4.70	5.12	4.60	4.58	99.04	100.0	100.0	100.0	100.0
150	4.92	4.92	4.72	5.54	5.16	99.72	100.0	100.0	100.0	100.0
200	5.34	5.58	5.14	5.08	4.80	99.82	100.0	100.0	100.0	100.0
	DCCE with p_f selected by AIC									
40	4.84	5.24	5.50	5.20	5.10	75.32	97.98	100.0	100.0	100.0
60	4.76	4.98	4.72	5.16	5.46	85.12	99.54	100.0	100.0	100.0
100	5.18	4.84	5.16	4.48	4.82	91.92	99.96	100.0	100.0	100.0
150	4.98	5.28	4.94	5.46	5.40	95.04	100.0	100.0	100.0	100.0
200	4.80	5.40	5.34	5.56	5.12	96.86	99.98	100.0	100.0	100.0
	Infeasible with $p_f = 1$									
40	4.92	5.92	5.06	5.04	5.68	88.82	99.42	99.98	100.0	100.0
60	5.10	5.38	4.82	4.66	5.84	95.56	99.94	100.0	100.0	100.0
100	5.46	4.88	4.48	5.10	4.76	98.74	100.0	100.0	100.0	100.0
150	4.98	4.80	4.76	4.88	4.90	99.72	100.0	100.0	100.0	100.0
200	5.92	5.40	4.96	4.90	4.42	99.86	100.0	100.0	100.0	100.0
	Infeasible with p_f selected by AIC									
40	4.66	6.00	5.14	5.32	5.48	78.52	98.72	99.98	100.0	100.0
60	5.16	5.28	5.00	4.88	5.66	87.78	99.76	100.0	100.0	100.0
100	4.64	4.28	4.42	5.06	4.42	94.08	99.98	100.0	100.0	100.0
150	4.82	4.52	4.88	5.06	4.76	96.52	99.96	100.0	100.0	100.0
200	4.82	5.00	4.84	4.98	4.88	97.36	99.98	100.0	100.0	100.0
	KCCE (Static)									
40	99.88	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
60	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
100	99.98	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
150	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
200	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0

Notes: See Table 3.E1.A for an explanation.

3 A CCE ESTIMATOR FOR DYNAMIC PANEL ECM'S

Table 3.E5 : $\beta_{xi}, \alpha_{xi} \sim U[-0.45, -0.15], \beta_{xi} = -1, \lambda_{ij} = -2$ for all i and $j = 1, 2$

N/T	Bias $\times 100$					RMSE				
	40	60	100	150	200	40	60	100	150	200
	DCCE with $p_f = p_T$									
40	4.53	3.08	1.90	1.47	1.00	24.83	14.32	8.15	5.49	3.91
60	3.72	3.19	1.88	1.32	1.01	24.71	14.49	8.31	5.41	4.00
100	3.20	3.08	1.88	1.28	0.93	25.99	14.79	8.22	5.41	3.91
150	3.19	3.20	1.86	1.32	0.94	27.20	15.16	8.20	5.54	4.03
200	3.34	2.95	1.79	1.29	0.94	25.97	14.36	8.32	5.46	3.92
	DCCE with $p_f = 1$									
40	2.97	1.86	1.40	0.98	0.61	19.61	12.18	7.28	4.91	3.68
60	2.89	2.72	1.58	1.06	0.84	19.06	12.03	7.11	4.77	3.70
100	2.76	2.47	1.65	1.06	0.76	19.34	12.47	7.20	4.91	3.60
150	3.18	2.75	1.73	1.21	0.85	19.48	12.63	7.23	4.92	3.70
200	3.40	2.37	1.48	1.03	0.69	19.61	12.33	7.41	4.83	3.58
	DCCE with p_f selected by AIC									
40	4.17	2.43	1.57	1.16	0.78	30.93	13.93	7.85	5.12	3.80
60	3.68	2.64	1.58	1.06	0.86	25.45	15.87	7.82	5.02	3.88
100	-4.19	2.97	1.78	1.07	0.78	13.32	14.40	7.90	5.14	3.79
150	-3.35	2.97	1.78	1.21	0.82	27.20	14.37	8.00	5.23	3.85
200	3.56	2.45	1.56	1.04	0.70	24.92	13.90	8.06	5.10	3.72
	Infeasible with $p_f = 1$									
40	3.90	2.91	1.94	1.27	0.96	22.93	14.17	8.01	5.07	3.87
60	3.27	2.77	1.69	1.25	0.97	21.72	13.85	7.77	5.13	3.91
100	3.11	2.87	1.86	1.22	0.98	22.67	13.79	7.90	5.14	3.79
150	3.30	2.84	1.81	1.27	0.95	22.63	14.15	7.72	5.29	3.85
200	3.91	2.91	1.81	1.27	0.89	22.82	13.66	7.82	5.19	3.69
	Infeasible with p_f selected by AIC									
40	3.50	3.64	2.24	1.51	1.08	37.15	15.40	8.37	5.28	4.03
60	4.13	3.22	2.05	1.44	1.11	26.36	14.71	8.13	5.36	4.04
100	3.81	3.42	2.28	1.42	1.12	26.29	14.68	8.49	5.34	3.95
150	3.93	3.45	2.15	1.48	1.06	27.05	15.12	8.25	5.50	3.96
200	4.26	3.51	2.13	1.45	1.03	28.24	15.05	8.24	5.38	3.85
	KCCE (Static)									
40	46.90	40.93	34.61	29.91	26.94	53.35	46.77	39.81	34.78	31.53
60	45.40	39.85	33.55	28.75	25.90	52.08	45.82	38.63	33.74	30.38
100	43.10	38.45	31.45	27.22	23.64	50.00	44.41	36.79	31.90	27.96
150	40.05	35.26	29.05	25.11	22.03	46.54	40.65	33.95	29.51	26.06
200	38.58	33.99	28.15	24.20	21.41	44.87	39.22	32.74	28.16	25.28

Notes: See Table 3.E1.A for an explanation.

Table 3.E5 cont'd: $\beta_{xi}, \alpha_{xi} \sim U[-0.45, -0.15]$, $\beta_{xi} = -1$, $\lambda_{ij} = -2$ for all i and $j = 1, 2$

N/T	Size					Power				
	40	60	100	150	200	40	60	100	150	200
DCCE with $p_f = p_T$										
40	8.34	7.24	6.78	6.44	6.14	19.70	29.08	48.10	71.78	85.08
60	8.90	8.32	6.90	7.00	5.90	19.44	29.92	48.52	68.48	84.80
100	7.80	7.44	6.54	6.06	5.70	18.36	27.48	48.16	68.00	83.16
150	8.04	7.94	6.96	6.00	6.02	18.96	28.64	46.82	68.42	82.12
200	8.04	7.48	6.82	6.06	5.96	18.58	28.76	47.28	69.86	83.82
DCCE with $p_f = 1$										
40	7.88	7.18	7.12	7.54	7.40	23.40	32.44	54.14	74.38	86.60
60	7.72	8.06	6.76	6.64	6.74	23.48	35.60	55.76	74.72	88.18
100	6.76	6.40	5.60	5.18	5.34	21.50	31.28	52.50	71.96	84.92
150	7.68	7.34	6.28	5.92	5.72	22.16	32.82	53.42	73.46	85.74
200	7.82	6.96	6.34	6.34	6.06	22.72	32.82	53.00	73.80	86.12
DCCE with p_f selected by AIC										
40	13.80	9.72	8.36	7.92	7.44	28.14	35.32	54.32	74.16	86.46
60	12.96	11.30	8.54	7.52	6.92	28.42	36.28	54.22	71.86	86.10
100	12.42	10.52	8.50	7.16	6.82	27.36	34.76	54.38	72.18	84.76
150	12.48	10.74	8.36	7.10	6.68	26.56	35.34	54.12	73.44	85.16
200	11.92	10.06	8.34	7.94	7.00	26.50	34.18	53.44	72.98	85.74
Infeasible with $p_f = 1$										
40	7.86	8.12	7.42	6.40	5.52	20.86	30.86	51.58	72.66	85.76
60	7.62	7.32	6.52	5.94	6.56	21.06	30.54	51.52	71.72	85.96
100	7.84	7.88	7.16	6.32	6.16	20.90	30.08	51.60	70.96	85.68
150	8.56	7.48	6.32	6.38	5.92	21.22	30.86	51.92	72.40	85.32
200	8.44	7.46	6.78	6.20	5.68	20.94	30.70	50.78	73.42	86.02
Infeasible with p_f selected by AIC										
40	12.76	11.02	9.00	7.56	6.46	26.68	34.76	53.54	74.22	86.16
60	13.12	10.56	8.50	7.52	7.52	27.14	33.90	54.12	73.58	87.24
100	12.54	10.00	8.86	7.70	7.06	26.40	34.22	55.26	72.86	86.44
150	13.72	10.86	8.66	7.68	7.08	27.38	34.26	54.38	73.66	86.16
200	13.62	10.04	8.78	7.28	6.54	26.86	35.28	53.50	74.90	86.90
KCCE (Static)										
40	94.70	95.32	95.28	95.68	95.60	96.98	97.68	98.02	98.52	98.96
60	94.30	94.98	95.36	94.94	95.28	96.76	97.64	98.38	98.20	98.58
100	94.56	95.66	95.16	95.52	95.80	97.34	98.24	98.68	99.28	99.48
150	94.10	95.32	95.38	95.84	95.74	97.04	98.04	98.12	98.70	98.92
200	93.90	94.92	95.48	96.02	95.80	96.78	97.90	98.06	98.90	98.98

Notes: See Table E1.A for an explanation.

3.B Appendix

3.B.1 The Case with Observed Factors

Lemmas and proofs

In this section we provide some preliminary results to be used in the proofs. First, we rewrite the model given in (3.3)-(3.5) in a common trends representation form. We use Theorem 4.1 of Johansen (1991) for this purpose. We choose the orthogonal complements of α_i and β_i denoted by $\alpha_{i\perp}$ and $\beta_{i\perp}$ such that $\alpha_i' \alpha_{i\perp} = 0$ and $\beta_i' \beta_{i\perp} = 0$ as

$$\begin{aligned} \alpha_{i\perp} &= \begin{pmatrix} \alpha_{yi} \\ 0 \\ 0 \end{pmatrix}_{\perp} = \begin{pmatrix} 0 & 0 \\ I_m & 0 \\ 0 & I_k \end{pmatrix}, \\ \beta_{i\perp} &= \begin{pmatrix} \mathbf{I} \\ -\beta_{xi} \\ -\lambda_i \end{pmatrix}_{\perp} = \begin{pmatrix} \beta'_{xi} & \lambda'_i \\ \mathbf{I}_m & 0 \\ 0 & \mathbf{I}_k \end{pmatrix}. \end{aligned} \quad (3.B.1)$$

Letting

$$\mathbf{C}_i = \beta_{i\perp} \left[\alpha'_{i\perp} \left(\mathbf{I}_{(1+m+k)} - \sum_{\ell=1}^{p-1} \mathbf{A}_{i\ell} \right) \beta_{i\perp} \right]^{-1} \alpha'_{i\perp}, \quad (3.B.2)$$

we can write the common trends representation for \mathbf{z}_{it}^* as

$$\mathbf{z}_{it}^* = \mathbf{C}_i \sum_{s=1}^t \varepsilon_{is}^* + \mathbf{C}_i^*(L) \varepsilon_{it}^* + \mathbf{z}_{i0}^*, \quad (3.B.3)$$

where $\mathbf{C}_i^*(L)$ is a $(1+m+k) \times (1+m+k)$ infinite lag polynomial matrix that is a function of the model parameters and \mathbf{z}_{i0}^* contains the initial values.

In the proofs of the theorems and lemmas, we frequently use block specific representations for the variables in the $\mathbf{z}_{it}^* = (y_{it}, \mathbf{x}'_{it}, \mathbf{f}'_t)'$ vector. To write the common trends representations for y_{it} , \mathbf{x}_{it} and \mathbf{f}_t separately, we need the block

specific expressions for the blocks of \mathbf{C}_i . Let

$$\mathbf{C}_i = \begin{pmatrix} \mathbf{C}_{yyi} & \mathbf{C}_{yxi} & \mathbf{C}_{yfi} \\ \mathbf{C}_{xyi} & \mathbf{C}_{xxi} & \mathbf{C}_{xfi} \\ 0_{k \times 1} & 0_{k \times m} & \mathbf{C}_{ffi} \end{pmatrix},$$

where the lower left block of size $k \times (1+m)$ is a zero matrix thanks to the strong exogeneity assumption. By using the definition for the orthogonal complement matrices $\boldsymbol{\alpha}_{i\perp}$ and $\boldsymbol{\beta}_{i\perp}$ in (3.B.2), we obtain the following expressions for the specific blocks of \mathbf{C}_i ,

$$\begin{aligned} \mathbf{C}_{yyi} &= 0, \\ \mathbf{C}_{yxi} &= \boldsymbol{\beta}'_{xi} [\mathbf{A}_{xxi}(1) + \mathbf{A}_{xyi}(1)\boldsymbol{\beta}'_{xi}]^{-1}, \\ \mathbf{C}_{yfi} &= \boldsymbol{\lambda}'_i \mathbf{A}_{ff}(1)^{-1} \\ &\quad - \boldsymbol{\beta}'_{xi} [\mathbf{A}_{xxi}(1) + \mathbf{A}_{xyi}(1)\boldsymbol{\beta}'_{xi}]^{-1} (\mathbf{A}_{xfi}(1) + \mathbf{A}_{xyi}(1)\boldsymbol{\lambda}'_i) \mathbf{A}_{ff}(1)^{-1}, \\ \mathbf{C}_{xyi} &= 0, \\ \mathbf{C}_{xxi} &= (\mathbf{A}_{xxi}(1) + \mathbf{A}_{xyi}(1)\boldsymbol{\beta}'_{xi})^{-1}, \\ \mathbf{C}_{xfi} &= -(\mathbf{A}_{xxi}(1) + \mathbf{A}_{xyi}(1)\boldsymbol{\beta}'_{xi})^{-1} (\mathbf{A}_{xfi}(1) + \mathbf{A}_{xyi}(1)\boldsymbol{\lambda}'_i) \mathbf{A}_{ff}(1)^{-1}, \\ \mathbf{C}_{ffi} &= \mathbf{A}_{ff}(1)^{-1}, \end{aligned} \tag{3.B.4}$$

where

$$\begin{aligned} \mathbf{A}_{xyi}(1) &= -\sum_{\ell=1}^{p-1} \mathbf{A}_{xyi\ell}, \quad \mathbf{A}_{xxi}(1) = \mathbf{I}_m - \sum_{\ell=1}^{p-1} \mathbf{A}_{xxi\ell}, \\ \mathbf{A}_{yfi}(1) &= -\sum_{\ell=1}^{p-1} \mathbf{A}_{yfi\ell}, \quad \mathbf{A}_{xfi}(1) = -\sum_{\ell=1}^{p-1} \mathbf{A}_{xfi\ell}, \\ \mathbf{A}_{ffi}(1) &= \mathbf{I}_k - \sum_{\ell=1}^{p-1} \mathbf{A}_{ffi\ell}. \end{aligned}$$

Also by defining the partial sum processes, such that

$$\mathbf{s}_{xit} = \sum_{s=1}^t \boldsymbol{\varepsilon}_{xis}, \quad \mathbf{s}_{ft} = \sum_{s=1}^t \boldsymbol{\varepsilon}_{fs}, \tag{3.B.5}$$

we can write

$$y_{it} = \mathbf{C}_{yxi}\mathbf{s}_{xit} + \mathbf{C}_{yfi}\mathbf{s}_{ft} + C_{yyi}^*(L)\varepsilon_{yit} + \mathbf{C}_{yxi}^*(L)\varepsilon_{xit} + \mathbf{C}_{yfi}^*(L)\varepsilon_{ft} + y_{i0}^*, \quad (3.B.6)$$

$$\mathbf{x}_{it} = \mathbf{C}_{xxi}\mathbf{s}_{xit} + \mathbf{C}_{xfi}\mathbf{s}_{ft} + \mathbf{C}_{xyi}^*(L)\varepsilon_{yit} + \mathbf{C}_{xxi}^*(L)\varepsilon_{xit} + \mathbf{C}_{xfi}^*(L)\varepsilon_{ft} + \mathbf{x}_{i0}^*, \quad (3.B.7)$$

$$\mathbf{f}_t = \mathbf{C}_{ff}\mathbf{s}_{ft} + \mathbf{C}_{ff}^*(L)\varepsilon_{ft} + \mathbf{f}_0^*, \quad (3.B.8)$$

where y_{i0}^* , \mathbf{x}_{i0}^* and \mathbf{f}_0^* contain initial values and will be ignored in the sequel. The lag polynomials are obtained by the partition of $\mathbf{C}_i^*(L)$ such that

$$\mathbf{C}_i^*(L) = \begin{pmatrix} C_{yyi}^*(L) & \mathbf{C}_{yxi}^*(L) & \mathbf{C}_{yfi}^*(L) \\ \mathbf{C}_{xyi}^*(L) & \mathbf{C}_{xxi}^*(L) & \mathbf{C}_{xfi}^*(L) \\ 0 & 0 & \mathbf{C}_{ff}^*(L) \end{pmatrix}. \quad (3.B.9)$$

Thanks to the assumptions placed on the error process ε_{it}^* , a functional central limit theorem can be applied to the processes, \mathbf{x}_{it} and \mathbf{f}_t . In the lemma below we present the limits of the processes.

Lemma 3.B.1. *Let $\varepsilon_{y.xit}$ be the conditional error term that is i.i.d. across t , \mathbf{x}_{it} and \mathbf{f}_t have the common trends representation shown in (3.B.6)-(3.B.8). As $T \rightarrow \infty$, for all $i = 1, \dots, N$, we have*

$$\frac{1}{\sqrt{T}} \sum_{s=1}^t \varepsilon_{y.xis} \xrightarrow{w} B_{y.xi}(r), \quad (3.B.10)$$

$$\frac{1}{\sqrt{T}} \mathbf{x}_{it} \xrightarrow{w} \mathbf{C}_{xxi} \mathbf{B}_{xi}(r) + \mathbf{C}_{xfi} \mathbf{B}_f(r), \quad (3.B.11)$$

$$\frac{1}{\sqrt{T}} \mathbf{f}_t \xrightarrow{w} \mathbf{C}_{ff} \mathbf{B}_f(r), \quad (3.B.12)$$

where $B_{y.xi}(r)$, $\mathbf{B}_{xi}(r)$, $\mathbf{B}_f(r)$ are Brownian motion processes and the covariance matrix of $\mathbf{B}_i^*(r) = (B_{y.xi}(r), \mathbf{B}_{xi}(r)', \mathbf{B}_f(r)')'$ given in (3.9).

Proof of Lemma 3.B.1: The results are established by a simple adaptation of the existing results in the literature on non-stationary processes. For details, for example, see Park and Phillips (1988). ■

At this stage it is useful to introduce some further notation. As also defined in Theorem 3.1, the error correction term at time t for the i^{th} cross-section unit

can be written as

$$\begin{aligned} u_{it} &= \boldsymbol{\beta}'_i \mathbf{z}_{it}^* \\ &= y_{it} - \boldsymbol{\beta}'_{xi} \mathbf{x}_{it} - \boldsymbol{\lambda}'_i \mathbf{f}_t. \end{aligned} \quad (3.B.13)$$

Additionally, using the common trends representation for \mathbf{z}_{it}^* and premultiplying it by $\boldsymbol{\beta}_i$, the error correction term can also be expressed as

$$\mathbf{u}_{it} = \boldsymbol{\beta}'_i \mathbf{C}_i^*(L) \boldsymbol{\varepsilon}_{it}^*. \quad (3.B.14)$$

We provide the lemmas below regarding the limits of the cross products of the processes appear in the estimators.

Lemma 3.B.2. *Let the t^{th} rows of \mathbf{X}_i , \mathbf{F} be generated by (3.B.7) and (3.B.8) respectively, the t^{th} row of \mathbf{U}_i be defined in (3.B.13) and $\boldsymbol{\varepsilon}_{y.xi}$ be the vector of conditional error term. Under the assumptions that are stated in Section 3.2, as $T \rightarrow \infty$ for all $i = 1, \dots, N$, we have*

$$\frac{\mathbf{F}'_{-1} \mathbf{F}_{-1}}{T^2} \xrightarrow{d} \mathbf{C}_{ff} \int_0^1 \mathbf{B}_f(r) \mathbf{B}_f(r)' dr \mathbf{C}'_{ff}, \quad (3.B.15)$$

$$\frac{\mathbf{X}'_{i,-1} \mathbf{X}_{i,-1}}{T^2} \xrightarrow{d} \mathbf{C}_{xzi} \int_0^1 \mathbf{B}_{bi}(r) \mathbf{B}_{bi}(r)' dr \mathbf{C}'_{xzi}, \quad (3.B.16)$$

$$\frac{\mathbf{X}'_{i,-1} \mathbf{F}_{-1}}{T^2} \xrightarrow{d} \mathbf{C}_{xzi} \int_0^1 \mathbf{B}_{bi}(r) \mathbf{B}_f(r)' dr \mathbf{C}'_{ff}, \quad (3.B.17)$$

$$\frac{\mathbf{X}'_{i,-1} \boldsymbol{\varepsilon}_{y.xi}}{T} \xrightarrow{d} \mathbf{C}_{xzi} \int_0^1 \mathbf{B}_{bi}(r) dB_{y.xi}(r), \quad (3.B.18)$$

$$\frac{\mathbf{F}'_{-1} \boldsymbol{\varepsilon}_{y.xi}}{T} \xrightarrow{d} \mathbf{C}_{ff} \int_0^1 \mathbf{B}_f(r) dB_{y.xi}(r), \quad (3.B.19)$$

$$\frac{\mathbf{U}'_{i,-1} \mathbf{U}_{i,-1}}{T} \xrightarrow{p} \boldsymbol{\beta}'_i \sum_{j=0}^{\infty} \mathbf{C}_{ij}^* \boldsymbol{\Sigma}_i \mathbf{C}_{ij}^{*'} \boldsymbol{\beta}_i, \quad (3.B.20)$$

$$\frac{\mathbf{U}'_{i,-1} \boldsymbol{\varepsilon}_{y.xi}}{\sqrt{T}} \xrightarrow{d} N(0, \boldsymbol{\beta}'_i \sum_{j=0}^{\infty} \mathbf{C}_{ij}^* \boldsymbol{\Sigma}_i \mathbf{C}_{ij}^{*'} \boldsymbol{\beta}_i \sigma_{y.xi}). \quad (3.B.21)$$

where $\mathbf{B}_{bi}(r) = (\mathbf{B}_{xi}(r)', \mathbf{B}_f(r)')'$ and

$$\mathbf{C}_{xzi} = (\mathbf{C}_{xvi}, \mathbf{C}_{xfi}), \quad (3.B.22)$$

where $\mathbf{C}_{ij}^{*'}$ are the coefficients of the lag-polynomial, $\mathbf{C}_i^*(L) = \sum_{j=0}^{\infty} \mathbf{C}_{ij}^* L^j$ and $\boldsymbol{\Sigma}_i$ is defined in (3.2).

Proof of Lemma 3.B.2: Using the results that are stated in Lemma 3.B.1 and applying the continuous mapping theorem is sufficient to obtain the results in (3.B.15)-(3.B.19). To prove (3.B.20) one can use the standard weak law of large numbers because $\mathbf{U}_{i,-1}$ is a stationary process under cointegration. We have

$$\begin{aligned}
 T^{-1}E(\mathbf{U}'_{i,-1}\mathbf{U}_{i,-1}) &= T^{-1}\sum_{t=1}^TE(u_{t-1}^2) \\
 &= T^{-1}\boldsymbol{\beta}'_i\sum_{t=1}^T\sum_{j=0}^{\infty}\sum_{\ell=0}^{\infty}\mathbf{C}_{ij}^*E(\boldsymbol{\varepsilon}_{i,t-j}\boldsymbol{\varepsilon}'_{i,t-\ell})\mathbf{C}_{i\ell}^*\boldsymbol{\beta}'_i \\
 &= \boldsymbol{\beta}'_i\sum_{j=0}^{\infty}\mathbf{C}_{ij}^*\boldsymbol{\Sigma}_i\mathbf{C}_{ij}^*\boldsymbol{\beta}'_i.
 \end{aligned} \tag{3.B.23}$$

Since the left-hand side of (3.B.21) involves a cross product of two independent covariance stationary processes the limit can be also be obtained using standard results in the literature. \blacksquare

Lemma 3.B.3. *Under the assumptions that are stated in Section 3.2, we have*

$$\begin{aligned}
 \frac{\mathbf{G}'_i\boldsymbol{\varepsilon}_{y.xi}}{\sqrt{T}} &= O_p(1), & \frac{\mathbf{U}'_{i,-1}\mathbf{G}_i}{T} &= O_p(1), & \frac{\mathbf{G}'_i\mathbf{G}_i}{T} &= O_p(1), \\
 \frac{\mathbf{X}'_{i,-1}\mathbf{U}_{i,-1}}{T^{3/2}} &= o_p(1), & \frac{\mathbf{X}'_{i,-1}\mathbf{G}_i}{T^{3/2}} &= o_p(1), \\
 \frac{\mathbf{F}'_{-1}\mathbf{U}_{i,-1}}{T^{3/2}} &= o_p(1), & \frac{\mathbf{F}'_{-1}\mathbf{G}_i}{T^{3/2}} &= o_p(1),
 \end{aligned}$$

where $\mathbf{G}_i = (\mathbf{G}_{yi}, \mathbf{G}_{xi}, \mathbf{G}_f)$ and \mathbf{G}_{yi} , \mathbf{G}_{xi} and \mathbf{G}_f are defined subsequent to (3.10).

Proof of Lemma 3.B.3: This lemma is an adaptation of the standard results. For example, see Lütkepohl (2007), Lemma 7.2 for details. \blacksquare

Proof of Theorem 3.1

We use a similar analogy as in Lütkepohl (2007) and adapt the proofs to our case where we estimate a single equation, conditional model. First, we define

$$\mathbf{Q}_i^* = \begin{pmatrix} 1 & 0 & 0 & 0 \\ -\boldsymbol{\beta}_{xi} & 0 & \mathbf{I}_m & 0 \\ -\boldsymbol{\lambda}_i & 0 & 0 & \mathbf{I}_k \\ 0 & \mathbf{I}_q & 0 & 0 \end{pmatrix}, \quad \mathbf{Q}_i^{*-1} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & \mathbf{I}_q \\ \boldsymbol{\beta}_{xi} & \mathbf{I}_m & 0 & 0 \\ \boldsymbol{\lambda}_i & 0 & \mathbf{I}_k & 0 \end{pmatrix}, \tag{3.B.24}$$

where $q = (p-1)(1+k) + mp$ is the number of columns of \mathbf{G}_i in (3.11). Considering (3.11), we perform the following operations:

$$\begin{aligned}\Delta \mathbf{y}_i &= \begin{pmatrix} \mathbf{Z}_{i,-1}^* & \mathbf{G}_i \end{pmatrix} \mathbf{Q}_i^* \mathbf{Q}_i^{*-1} \begin{pmatrix} \mathbf{\Pi}_i \\ \mathbf{\Gamma}_i \end{pmatrix} + \boldsymbol{\varepsilon}_{y.xi} \\ &= \begin{pmatrix} \mathbf{U}_{i,-1} & \mathbf{G}_i & \mathbf{X}_{i,-1} & \mathbf{F}_{-1} \end{pmatrix} \mathbf{Q}_i^{*-1} \begin{pmatrix} \mathbf{\Pi}_i \\ \mathbf{\Gamma}_i \end{pmatrix} + \boldsymbol{\varepsilon}_{y.xi},\end{aligned}$$

where $\mathbf{U}_{i,-1}$ is the $T \times 1$ vector of error correction, where the t^{th} row is defined in (3.B.13). Then we define a diagonal standardization matrix $\mathbf{D} = T^{1/2} \mathbf{I}_{(1+q+m+k)}$. In order to derive the asymptotic properties of the OLS estimator stated in (3.12), premultiplying $\mathbf{Q}_i^{*-1} \begin{pmatrix} \mathbf{\Pi}_i \\ \mathbf{\Gamma}_i \end{pmatrix}$ by \mathbf{D} yields

$$\begin{aligned}\mathbf{D} \mathbf{Q}_i^{*-1} \begin{pmatrix} \widehat{\mathbf{\Pi}}_i - \mathbf{\Pi}_i \\ \widehat{\mathbf{\Gamma}}_i - \mathbf{\Gamma}_i \end{pmatrix} &= \begin{pmatrix} \frac{\mathbf{U}'_{i,-1} \mathbf{U}_{i,-1}}{T} & \frac{\mathbf{U}'_{i,-1} \mathbf{G}_i}{T} & \frac{\mathbf{U}'_{i,-1} \mathbf{X}_{i,-1}}{T^{3/2}} & \frac{\mathbf{U}'_{i,-1} \mathbf{F}_{-1}}{T^{3/2}} \\ \frac{\mathbf{G}'_i \mathbf{U}_{i,-1}}{T} & \frac{\mathbf{G}'_i \mathbf{G}_i}{T} & \frac{\mathbf{G}'_i \mathbf{X}_{i,-1}}{T^{3/2}} & \frac{\mathbf{G}'_i \mathbf{F}_{-1}}{T^{3/2}} \\ \frac{\mathbf{X}'_{i,-1} \mathbf{U}_{i,-1}}{T^{3/2}} & \frac{\mathbf{X}'_{i,-1} \mathbf{G}_i}{T^{3/2}} & \frac{\mathbf{X}'_{i,-1} \mathbf{X}_{i,-1}}{T^2} & \frac{\mathbf{X}'_{i,-1} \mathbf{F}_{-1}}{T^2} \\ \frac{\mathbf{F}'_{-1} \mathbf{U}_{i,-1}}{T^{3/2}} & \frac{\mathbf{F}'_{-1} \mathbf{G}_i}{T^{3/2}} & \frac{\mathbf{F}'_{-1} \mathbf{X}_{i,-1}}{T^2} & \frac{\mathbf{F}'_{-1} \mathbf{F}_{-1}}{T^2} \end{pmatrix}^{-1} \\ &\quad \times \begin{pmatrix} \frac{\mathbf{U}'_{i,-1} \boldsymbol{\varepsilon}_{y.xi}}{T^{1/2}} \\ \frac{\mathbf{G}'_i \boldsymbol{\varepsilon}_{y.xi}}{T^{1/2}} \\ \frac{\mathbf{X}'_{i,-1} \boldsymbol{\varepsilon}_{y.xi}}{T^{3/2}} \\ \frac{\mathbf{F}'_{-1} \boldsymbol{\varepsilon}_{y.xi}}{T^{3/2}} \end{pmatrix}. \tag{3.B.25}\end{aligned}$$

Using the results in Lemma 3.B.2 and Lemma 3.B.3 in the equation above, it is straightforward to show that the off-block-diagonal elements of the first part of the right-hand side are $o_p(1)$. This result allows us to write

$$\begin{aligned}\mathbf{D} \mathbf{Q}_i^{*-1} \begin{pmatrix} \widehat{\mathbf{\Pi}}_i - \mathbf{\Pi}_i \\ \widehat{\mathbf{\Gamma}}_i - \mathbf{\Gamma}_i \end{pmatrix} &= \begin{pmatrix} \begin{pmatrix} \frac{\mathbf{U}'_{i,-1} \mathbf{U}_{i,-1}}{T} & \frac{\mathbf{U}'_{i,-1} \mathbf{G}_i}{T} \end{pmatrix}^{-1} \begin{pmatrix} \frac{\mathbf{U}'_{i,-1} \boldsymbol{\varepsilon}_{y.xi}}{T^{1/2}} \\ \frac{\mathbf{G}'_i \boldsymbol{\varepsilon}_{y.xi}}{T^{1/2}} \end{pmatrix} \\ \begin{pmatrix} \frac{\mathbf{X}'_{i,-1} \mathbf{X}_{i,-1}}{T^2} & \frac{\mathbf{X}'_{i,-1} \mathbf{F}_{-1}}{T^2} \\ \frac{\mathbf{F}'_{-1} \mathbf{X}_{i,-1}}{T^2} & \frac{\mathbf{F}'_{-1} \mathbf{F}_{-1}}{T^2} \end{pmatrix}^{-1} \begin{pmatrix} \frac{\mathbf{X}'_{i,-1} \boldsymbol{\varepsilon}_{y.xi}}{T^{3/2}} \\ \frac{\mathbf{F}'_{-1} \boldsymbol{\varepsilon}_{y.xi}}{T^{3/2}} \end{pmatrix} \end{pmatrix} \\ &\quad + o_p(1) \\ &= \begin{pmatrix} \begin{pmatrix} \frac{\mathbf{U}'_{i,-1} \mathbf{U}_{i,-1}}{T} & \frac{\mathbf{U}'_{i,-1} \mathbf{G}_i}{T} \end{pmatrix}^{-1} \begin{pmatrix} \frac{\mathbf{U}'_{i,-1} \boldsymbol{\varepsilon}_{y.xi}}{T^{1/2}} \\ \frac{\mathbf{G}'_i \boldsymbol{\varepsilon}_{y.xi}}{T^{1/2}} \end{pmatrix} \\ o_p(1) \end{pmatrix} \\ &\quad + o_p(1),\end{aligned}$$

where we used the results stated in Lemma 3.B.3 to obtain the second equality in the above equation. As a next step, we premultiply both sides of the equation above by \mathbf{Q}_i^* . Note that using the definition of $\boldsymbol{\alpha}_{i\perp}$ that can be found in (3.B.1), we can also write $\mathbf{Q}_i^* = \begin{pmatrix} \boldsymbol{\beta}_i & 0 & \boldsymbol{\alpha}_{i\perp} \\ 0 & \mathbf{I}_q & 0 \end{pmatrix}$, then the premultiplication which yields

$$\sqrt{T} \begin{pmatrix} \widehat{\boldsymbol{\Pi}}_i - \boldsymbol{\Pi}_i \\ \widehat{\boldsymbol{\Gamma}}_i - \boldsymbol{\Gamma}_i \end{pmatrix} = \begin{pmatrix} \frac{\mathbf{U}'_{i,-1}\mathbf{U}_{i,-1}}{T} & \frac{\mathbf{U}'_{i,-1}\mathbf{G}_i}{T} \\ \frac{\mathbf{G}'_i\mathbf{U}_{i,-1}}{T} & \frac{\mathbf{G}'_i\mathbf{G}_i}{T} \end{pmatrix}^{-1} \begin{pmatrix} \frac{\mathbf{U}'_{i,-1}\boldsymbol{\varepsilon}_{y.xi}}{T^{1/2}} \\ \frac{\mathbf{G}'_i\boldsymbol{\varepsilon}_{y.xi}}{T^{1/2}} \end{pmatrix} + o_p(1).$$

By letting

$$\boldsymbol{\Psi}_i^* = \lim_{T \rightarrow \infty} \begin{pmatrix} \frac{\mathbf{U}'_{i,-1}\mathbf{U}_{i,-1}}{T} & \frac{\mathbf{U}'_{i,-1}\mathbf{G}_i}{T} \\ \frac{\mathbf{G}'_i\mathbf{U}_{i,-1}}{T} & \frac{\mathbf{G}'_i\mathbf{G}_i}{T} \end{pmatrix}, \quad (3.B.26)$$

we can write the asymptotic distribution for the estimator defined in (3.12), as

$$\sqrt{T} \begin{pmatrix} \widehat{\boldsymbol{\Pi}}_i - \boldsymbol{\Pi}_i \\ \widehat{\boldsymbol{\Gamma}}_i - \boldsymbol{\Gamma}_i \end{pmatrix} \xrightarrow{d} N \left(0, \begin{pmatrix} \boldsymbol{\beta}_i & 0 \\ 0 & \mathbf{I}_p \end{pmatrix} \boldsymbol{\Psi}_i^{*-1} \begin{pmatrix} \boldsymbol{\beta}_i & 0 \\ 0 & \mathbf{I}_p \end{pmatrix}' \sigma_{y.xi} \right),$$

which gives us the desired result.

To prove the second part of the theorem it is useful to start with premultiplying the expression

$$T^{-1} \begin{pmatrix} \mathbf{Z}'_{i,-1}\mathbf{Z}_{i,-1}^* & \mathbf{Z}'_{i,-1}\mathbf{G}'_i \\ \mathbf{G}'_i\mathbf{Z}_{i,-1}^* & \mathbf{G}'_i\mathbf{G}_i \end{pmatrix}^{-1} \text{ by } \mathbf{Q}_i^* \mathbf{Q}_i^{*-1} \text{ and postmultiplying by } \mathbf{Q}_i^{*-1'} \mathbf{Q}_i^{*'},$$

this operation yields

$$\begin{aligned} & T^{-1} \begin{pmatrix} \mathbf{Z}'_{i,-1}\mathbf{Z}_{i,-1}^* & \mathbf{Z}'_{i,-1}\mathbf{G}'_i \\ \mathbf{G}'_i\mathbf{Z}_{i,-1}^* & \mathbf{G}'_i\mathbf{G}_i \end{pmatrix}^{-1} \\ &= \mathbf{Q}_i^* \begin{pmatrix} \frac{\mathbf{U}'_{i,-1}\mathbf{U}_{i,-1}}{T} & \frac{\mathbf{U}'_{i,-1}\mathbf{G}_i}{T} & \frac{\mathbf{U}'_{i,-1}\mathbf{X}_{i,-1}}{T^{3/2}} & \frac{\mathbf{U}'_{i,-1}\mathbf{F}_{-1}}{T^{3/2}} \\ \frac{\mathbf{G}'_i\mathbf{U}_{i,-1}}{T} & \frac{\mathbf{G}'_i\mathbf{G}_i}{T} & \frac{\mathbf{G}'_i\mathbf{X}_{i,-1}}{T^{3/2}} & \frac{\mathbf{G}'_i\mathbf{F}_{-1}}{T^{3/2}} \\ \frac{\mathbf{X}'_{i,-1}\mathbf{U}_{i,-1}}{T^{3/2}} & \frac{\mathbf{X}'_{i,-1}\mathbf{G}_i}{T^{3/2}} & \frac{\mathbf{X}'_{i,-1}\mathbf{X}_{i,-1}}{T^2} & \frac{\mathbf{X}'_{i,-1}\mathbf{F}_{-1}}{T^2} \\ \frac{\mathbf{F}'_{-1}\mathbf{U}_{i,-1}}{T^{3/2}} & \frac{\mathbf{F}'_{-1}\mathbf{G}_i}{T^{3/2}} & \frac{\mathbf{F}'_{-1}\mathbf{X}_{i,-1}}{T^2} & \frac{\mathbf{F}'_{-1}\mathbf{F}_{-1}}{T^2} \end{pmatrix}^{-1} \mathbf{Q}_i^{*'} \end{aligned}$$

The remaining part of the proof is provided in detail for the case in which the factors are not observed (See the proof of Theorem 3.4), hence it will be omitted for the current case. Using the rules for partitioned inverses and applying the

appropriate limit results it is straightforward to show that

$$T^{-1} \begin{pmatrix} \mathbf{Z}_{i,-1}^{*'} \mathbf{Z}_{i,-1}^* & \mathbf{Z}_{i,-1}^{*'} \mathbf{G}_i \\ \mathbf{G}_i' \mathbf{Z}_{i,-1}^* & \mathbf{G}_i' \mathbf{G}_i \end{pmatrix}^{-1} \xrightarrow{p} \mathbf{Q}_i^* \begin{pmatrix} \Psi_i^{*-1} + o_p(1) & o_p(1) \\ o_p(1) & o_p(1) \end{pmatrix} \mathbf{Q}_i^{*'}.$$

Using the definition of \mathbf{Q}_i^* we get

$$T^{-1} \begin{pmatrix} \mathbf{Z}_{i,-1}^{*'} \mathbf{Z}_{i,-1}^* & \mathbf{Z}_{i,-1}^{*'} \mathbf{G}_i \\ \mathbf{G}_i' \mathbf{Z}_{i,-1}^* & \mathbf{G}_i' \mathbf{G}_i \end{pmatrix}^{-1} \xrightarrow{p} \begin{pmatrix} \beta_i & 0 \\ 0 & \mathbf{I}_q \end{pmatrix} \Psi_i^{*-1} \begin{pmatrix} \beta_i' & 0 \\ 0 & \mathbf{I}_q \end{pmatrix}.$$

This completes the proof of the second part of the theorem. \blacksquare

Proof of Theorem 3.2

To establish the result provided in the theorem, we start with the case where we assume that α_{yi} is known. We denote the estimators in this case by $\widehat{\beta}_{xi}$ and $\widehat{\lambda}_i$ for β_{xi} and λ_i respectively. Using the definition $\mathbf{R}_i = (\mathbf{X}_i, \mathbf{F})$, we start with rewriting the estimator stated in (3.13) as

$$\begin{pmatrix} \widehat{\beta}_{xi} \\ \widehat{\lambda}_i \end{pmatrix} = -\frac{1}{\alpha_{yi}} \begin{pmatrix} \mathbf{X}_{i,-1}' \mathbf{M}_i \mathbf{X}_{i,-1} & \mathbf{X}_{i,-1}' \mathbf{M}_i \mathbf{F}_{-1} \\ \mathbf{F}_{-1}' \mathbf{M}_i \mathbf{X}_{i,-1} & \mathbf{F}_{-1}' \mathbf{M}_i \mathbf{F}_{-1} \end{pmatrix}^{-1} \times \begin{pmatrix} \mathbf{X}_{i,-1}' \mathbf{M}_i (\Delta \mathbf{y}_i - \mathbf{y}_{i,-1} \alpha_{yi}) \\ \mathbf{F}_{-1}' \mathbf{M}_i (\Delta \mathbf{y}_i - \mathbf{y}_{i,-1} \alpha_{yi}) \end{pmatrix}.$$

Then considering the model

$$\Delta \mathbf{y}_i - \mathbf{y}_{i,-1} \alpha_{yi} = -\mathbf{X}_{i,-1} \beta_{xi} \alpha_{yi} - \mathbf{F}_{-1} \lambda_i \alpha_{yi} + \mathbf{G}_i \Gamma_i + \varepsilon_{y.xi},$$

we replace $(\Delta \mathbf{y}_i - \mathbf{y}_{i,-1} \alpha_{yi})$ in the estimator above with the right-hand side of the equation above and premultiply with T , which yields

$$T \begin{pmatrix} \widehat{\beta}_{xi} - \beta_{xi} \\ \widehat{\lambda}_i - \lambda_i \end{pmatrix} = -\frac{1}{\alpha_{yi}} \begin{pmatrix} \frac{\mathbf{X}_{i,-1}' \mathbf{M}_i \mathbf{X}_{i,-1}}{T^2} & \frac{\mathbf{X}_{i,-1}' \mathbf{M}_i \mathbf{F}_{-1}}{T^2} \\ \frac{\mathbf{F}_{-1}' \mathbf{M}_i \mathbf{X}_{i,-1}}{T^2} & \frac{\mathbf{F}_{-1}' \mathbf{M}_i \mathbf{F}_{-1}}{T^2} \end{pmatrix}^{-1} \times \begin{pmatrix} \frac{\mathbf{X}_{i,-1}' \mathbf{M}_i \varepsilon_{y.xi}}{T} \\ \frac{\mathbf{F}_{-1}' \mathbf{M}_i \varepsilon_{y.xi}}{T} \end{pmatrix}. \quad (3.B.27)$$

Thanks to the results provided in Lemma 3.B.3 we have

$$\frac{\mathbf{X}_{i,-1}' \mathbf{M}_i \mathbf{X}_{i,-1}}{T^2} = \frac{\mathbf{X}_{i,-1}' \mathbf{X}_{i,-1}}{T^2} + o_p(1), \quad \frac{\mathbf{F}_{-1}' \mathbf{M}_i \mathbf{F}_{-1}}{T^2} = \frac{\mathbf{F}_{-1}' \mathbf{F}_{-1}}{T^2} + o_p(1),$$

$$\frac{\mathbf{X}'_{i,-1}\mathbf{M}_i\mathbf{F}_{-1}}{T^2} = \frac{\mathbf{X}'_{i,-1}\mathbf{F}_{-1}}{T^2} + o_p(1).$$

Then by an application of the standard methods in the literature and considering the results and definitions provided in Lemma 3.B.2, we obtain

$$\begin{aligned} T \begin{pmatrix} \widehat{\boldsymbol{\beta}}_{xi} - \boldsymbol{\beta}_{xi} \\ \widehat{\boldsymbol{\lambda}}_i - \boldsymbol{\lambda}_i \end{pmatrix} &\xrightarrow{d} \\ & - \frac{1}{\alpha_{yi}} \begin{bmatrix} \int_0^1 \mathbf{C}_{xzi}\mathbf{B}_{bi}(r)\mathbf{B}'_{bi}(r)dr\mathbf{C}'_{xzi} & \int_0^1 \mathbf{C}_{xzi}\mathbf{B}_{bi}(r)\mathbf{B}'_f(r)dr\mathbf{C}'_{ff} \\ \int_0^1 \mathbf{C}_{ff}\mathbf{B}_f(r)\mathbf{B}'_{bi}(r)dr\mathbf{C}'_{xzi} & \int_0^1 \mathbf{C}_{ff}\mathbf{B}_f(r)\mathbf{B}'_f(r)dr\mathbf{C}'_{ff} \end{bmatrix}^{-1} \\ & \times \begin{bmatrix} \int_0^1 \mathbf{C}_{xzi}\mathbf{B}_{bi}(r)d\mathbf{B}'_{y.xi}(r) \\ \int_0^1 \mathbf{C}_{ff}\mathbf{B}_f(r)d\mathbf{B}'_{y.xi}(r) \end{bmatrix}. \end{aligned}$$

Note that the asymptotic representation above is obtained under the strong assumption of known α_{yi} . We need to show that the feasible estimator has the same asymptotic properties. We do this by showing that

$$T \begin{pmatrix} \widehat{\boldsymbol{\beta}}_{xi} - \widehat{\boldsymbol{\beta}}_{xi} \\ \widehat{\boldsymbol{\lambda}}_i - \widehat{\boldsymbol{\lambda}}_i \end{pmatrix} = o_p(1), \quad (3.B.28)$$

where

$$\begin{aligned} \begin{pmatrix} \widehat{\boldsymbol{\beta}}_{xi} \\ \widehat{\boldsymbol{\lambda}}_i \end{pmatrix} &= -\frac{1}{\widehat{\alpha}_{yi}} \begin{pmatrix} \mathbf{X}'_{i,-1}\mathbf{M}_i\mathbf{X}_{i,-1} & \mathbf{X}'_{i,-1}\mathbf{M}_i\mathbf{F}_{-1} \\ \mathbf{F}'_{-1}\mathbf{M}_i\mathbf{X}_{i,-1} & \mathbf{F}'_{-1}\mathbf{M}_i\mathbf{F}_{-1} \end{pmatrix}^{-1} \\ & \times \begin{pmatrix} \mathbf{X}'_{i,-1}\mathbf{M}_i(\Delta\mathbf{y}_i - \mathbf{y}_{i,-1}\widehat{\alpha}_{yi}) \\ \mathbf{F}'_{-1}\mathbf{M}_i(\Delta\mathbf{y}_i - \mathbf{y}_{i,-1}\widehat{\alpha}_{yi}) \end{pmatrix}. \end{aligned} \quad (3.B.29)$$

Consider the model where the unknown α_{yi} is replaced by its OLS estimator that can be written as

$$\Delta\mathbf{y}_i - \mathbf{y}_{i,-1}\widehat{\alpha}_{yi} = -\mathbf{X}_{i,-1}\boldsymbol{\beta}_{xi}\widehat{\alpha}_{yi} - \mathbf{F}_{-1}\boldsymbol{\lambda}_i\widehat{\alpha}_{yi} + \mathbf{G}_i\boldsymbol{\Gamma}_i + \boldsymbol{\varepsilon}_{y.xi}^*,$$

where $\boldsymbol{\varepsilon}_{y.xi}^*$ is the error term of the model in which the unknown α_{yi} is replaced by its OLS estimator. Using this equation in the estimator (3.B.29) and premultiplying with T yields

$$T \begin{pmatrix} \widehat{\boldsymbol{\beta}}_{xi} - \boldsymbol{\beta}_{xi} \\ \widehat{\boldsymbol{\lambda}}_i - \boldsymbol{\lambda}_i \end{pmatrix} = -\frac{1}{\widehat{\alpha}_{yi}} \begin{pmatrix} \frac{\mathbf{X}'_{i,-1}\mathbf{M}_i\mathbf{X}_{i,-1}}{T^2} & \frac{\mathbf{X}'_{i,-1}\mathbf{M}_i\mathbf{F}_{-1}}{T^2} \\ \frac{\mathbf{F}'_{-1}\mathbf{M}_i\mathbf{X}_{i,-1}}{T^2} & \frac{\mathbf{F}'_{-1}\mathbf{M}_i\mathbf{F}_{-1}}{T^2} \end{pmatrix}^{-1}$$

$$\begin{pmatrix} \frac{\mathbf{X}'_{i,-1}\mathbf{M}_i\boldsymbol{\varepsilon}_{y,xi}^*}{T} \\ \frac{\mathbf{F}'_{-1}\mathbf{M}_i\boldsymbol{\varepsilon}_{y,xi}^*}{T} \end{pmatrix}. \quad (3.B.30)$$

The difference between (3.B.30) and (3.B.27) can be written as

$$\begin{aligned} T \begin{pmatrix} (\widehat{\boldsymbol{\beta}}_{xi} - \widehat{\boldsymbol{\beta}}_{xi}) \\ (\widehat{\boldsymbol{\lambda}}_i - \widehat{\boldsymbol{\lambda}}_i) \end{pmatrix} &= \frac{1}{\widehat{\alpha}_{yi}} \begin{pmatrix} \frac{\mathbf{X}'_{i,-1}\mathbf{M}_i\mathbf{X}_{i,-1}}{T^2} & \frac{\mathbf{X}'_{i,-1}\mathbf{M}_i\mathbf{F}_{-1}}{T^2} \\ \frac{\mathbf{F}'_{-1}\mathbf{M}_i\mathbf{X}_{i,-1}}{T^2} & \frac{\mathbf{F}'_{-1}\mathbf{M}_i\mathbf{F}_{-1}}{T^2} \end{pmatrix}^{-1} \\ &+ \begin{pmatrix} \frac{\mathbf{X}'_{i,-1}\mathbf{M}_i(\boldsymbol{\varepsilon}_{y,xi}^* - \boldsymbol{\varepsilon}_{y,xi})}{T} \\ \frac{\mathbf{F}'_{-1}\mathbf{M}_i(\boldsymbol{\varepsilon}_{y,xi}^* - \boldsymbol{\varepsilon}_{y,xi})}{T} \end{pmatrix} \\ &- \left(\frac{1}{\widehat{\alpha}_{yi}} - \frac{1}{\alpha_{yi}} \right) \begin{pmatrix} \frac{\mathbf{X}'_{i,-1}\mathbf{M}_i\mathbf{X}_{i,-1}}{T^2} & \frac{\mathbf{X}'_{i,-1}\mathbf{M}_i\mathbf{F}_{-1}}{T^2} \\ \frac{\mathbf{F}'_{-1}\mathbf{M}_i\mathbf{X}_{i,-1}}{T^2} & \frac{\mathbf{F}'_{-1}\mathbf{M}_i\mathbf{F}_{-1}}{T^2} \end{pmatrix}^{-1} \\ &\times \begin{pmatrix} \frac{\mathbf{X}'_{i,-1}\mathbf{M}_i\boldsymbol{\varepsilon}_{y,xi}^*}{T} \\ \frac{\mathbf{F}'_{-1}\mathbf{M}_i\boldsymbol{\varepsilon}_{y,xi}^*}{T} \end{pmatrix}. \end{aligned} \quad (3.B.31)$$

To establish the order of the first term on the right-hand side above we write

$$\begin{aligned} \boldsymbol{\varepsilon}_{y,xi}^* - \boldsymbol{\varepsilon}_{y,xi} &= \mathbf{y}_{i,-1}(\widehat{\alpha}_{yi} - \alpha_{yi}) - \mathbf{X}_{i,-1}\boldsymbol{\beta}_{xi}(\widehat{\alpha}_{yi} - \alpha_{yi}) - \mathbf{F}_{-1}\boldsymbol{\lambda}_i(\widehat{\alpha}_{yi} - \alpha_{yi}) \\ &= \mathbf{U}_{i,-1}(\widehat{\alpha}_{yi} - \alpha_{yi}). \end{aligned}$$

Substituting this into the last part of the first term on the right-hand side of (3.B.31) yields

$$\begin{pmatrix} T^{-1}\mathbf{X}'_{i,-1}\mathbf{M}_i(\boldsymbol{\varepsilon}_{y,xi}^* - \boldsymbol{\varepsilon}_{y,xi}) \\ T^{-1}\mathbf{F}'_{-1}\mathbf{M}_i(\boldsymbol{\varepsilon}_{y,xi}^* - \boldsymbol{\varepsilon}_{y,xi}) \end{pmatrix} = \begin{pmatrix} T^{-1}\mathbf{X}'_{i,-1}\mathbf{M}_i\mathbf{U}_{i,-1}(\widehat{\alpha}_{yi} - \alpha_{yi}) \\ T^{-1}\mathbf{F}'_{-1}\mathbf{M}_i\mathbf{U}_{i,-1}(\widehat{\alpha}_{yi} - \alpha_{yi}) \end{pmatrix}.$$

Then by using the results in Lemma 3.B.3 and the consistency property of $\widehat{\alpha}_{yi}$, which implies $(\widehat{\alpha}_{yi} - \alpha_{yi}) = O_p(T^{-1/2})$, we can show that the first term of (3.B.31) is $o_p(1)$. The order of the second term can be found equal to $o_p(1)$ by again using the consistency of $\widehat{\alpha}_{yi}$. This completes the proof of Theorem 3.2. \blacksquare

3.B.2 The Case with Unobserved Factors

Approximation for the unobserved factors

This section is a supplementary to Section 3.3.2. We provide the details regarding the steps we use to obtain an approximation for the unobserved factors and establish the orders of the approximation errors. As it is also made clear in Section 3.3.2, we obtain approximation for the first differences and for the levels of the unobserved factors separately. In the first part of the section we find approximations for the first differences of the factors. In (3.22) we define \mathbf{v}_{Bt} as the approximation error for the first differences of the factors. We have

$$\mathbf{v}_{Bt} = \frac{1}{N} \sum_{i=1}^N (\mathbf{C}_{zzi} + \mathbf{C}_{zzi}^*(L)(1-L)) \boldsymbol{\varepsilon}_{zit}. \quad (3.B.32)$$

Below we provide the lemma regarding the order of this process.

Lemma 3.B.4. *Suppose Assumption 3.1(i) holds. For the order of the random process \mathbf{v}_{Bt} , defined by (3.B.32), we have*

$$\mathbf{v}_{Bt} = O_p(N^{-1/2}). \quad (3.B.33)$$

Proof of Lemma 3.B.4: To prove this result we first define

$$\mathbf{C}_{vbi}^*(L) = (\mathbf{C}_{zzi} + \mathbf{C}_{zzi}^*(L)(1-L)), \quad (3.B.34)$$

then (3.B.32) can be written as

$$\mathbf{v}_{Bt} = \frac{1}{N} \sum_{i=1}^N \mathbf{C}_{vbi}^*(L) \boldsymbol{\varepsilon}_{zit}. \quad (3.B.35)$$

We show that $E(\mathbf{v}_{Bt}) = 0$ and $E\|\mathbf{v}_{Bt}\|^2 = O(N^{-1})$. First note that by Assumption 3.1(i), $E(\boldsymbol{\varepsilon}_{zit}) = 0$, then we have

$$E(\mathbf{v}_{Bt}) = E\left(\frac{1}{N} \sum_{i=1}^N \mathbf{C}_{vbi}^*(L) \boldsymbol{\varepsilon}_{zit}\right) \leq \frac{1}{N} \sum_{i=1}^N \mathbf{C}_{vbi}^*(L) E(\boldsymbol{\varepsilon}_{zit}) = 0, \quad (3.B.36)$$

where we use the triangle inequality, the independence of the common trends representation parameters from the error terms and the result that the lag polynomial $\mathbf{C}_{vbi}^*(L)$ having absolute summable coefficients to obtain the second inequality.

Also, again by Assumption 3.1(i), $E(\boldsymbol{\varepsilon}_{zit}\boldsymbol{\varepsilon}'_{zit}) = \boldsymbol{\Sigma}_{zzi} < K$. We have

$$E\|\mathbf{v}_{Bt}\|^2 = \frac{1}{N^2} \sum_{i=1}^N \mathbf{C}_{vbi}^*(L) \boldsymbol{\Sigma}_{zzi} \mathbf{C}_{vbi}^{*'}(L) + o(1) = O(N^{-1}),$$

where the last equation is obtained by using Assumption 3.3, which implies that the common trends representation parameters are independent of the error terms across cross-section units and the result that the lag polynomial $\mathbf{C}_{vbi}^*(L)$ having absolute summable coefficients. Then,

$$E\|\mathbf{v}_{Bt}\| \leq [E\|\mathbf{v}_{Bt}\|^2]^{1/2} = O(N^{-1/2}) \quad (3.B.37)$$

yields the required result. \blacksquare

Next we obtain an approximation for the levels of the unobserved factors. We start with taking the cross-sectional averages of both sides of (3.14), the resulting expression is provided in (3.23). Consider the common trends representation for \mathbf{f}_t , which is provided in (3.15). We solve that equation for the partial sum process $\sum_{s=1}^t \boldsymbol{\varepsilon}_{fs}$ using the invertibility of \mathbf{C}_{ff} , then we substitute the resulting expression for $\sum_{s=1}^t \boldsymbol{\varepsilon}_{fs}$ in (3.23). This operation and a rearrangement of the terms yield

$$\begin{aligned} \bar{\mathbf{z}}_t &= \bar{\mathbf{C}}_{zf} \mathbf{C}_{ff}^{-1} \mathbf{f}_t + (\bar{\mathbf{C}}_{zf} \mathbf{C}_{ff}^{-1} \mathbf{C}_{ff}^*(L) - \bar{\mathbf{C}}_{zf}^*(L)) \boldsymbol{\varepsilon}_{ft} + \frac{1}{N} \sum_{i=1}^N \mathbf{C}_{zzi} \sum_{s=1}^t \boldsymbol{\varepsilon}_{zis} \\ &\quad + \frac{1}{N} \sum_{i=1}^N \mathbf{C}_{zzi}^*(L) \boldsymbol{\varepsilon}_{zit}. \end{aligned} \quad (3.B.38)$$

One way to eliminate the presence of the cyclical effects caused by the short-run dynamics of the unobserved factors is to correct the levels of the cross-sectional averages for the short run dynamics of the cross-sectional averages. To do that we make use of Assumption 3.4. Solving (3.20) for $\boldsymbol{\varepsilon}_{ft}$ and substituting the resulting expression to (3.B.38) we obtain

$$\begin{aligned} \bar{\mathbf{z}}_t &= \mathbf{H}'_1 \mathbf{f}_t + \mathbf{H}_3(L) \bar{\Delta} \bar{\mathbf{z}}_t + \frac{1}{N} \sum_{i=1}^N \mathbf{C}_{zzi} \sum_{s=1}^t \boldsymbol{\varepsilon}_{zis} + \frac{1}{N} \sum_{i=1}^N \mathbf{C}_{zzi}^*(L) \boldsymbol{\varepsilon}_{zit} \\ &\quad + \mathbf{H}_3(L) \frac{1}{N} \sum_{i=1}^N (\mathbf{C}_{zzi} + \mathbf{C}_{zzi}^*(L)(1-L)) \boldsymbol{\varepsilon}_{zit}, \end{aligned}$$

where $\mathbf{H}'_1 = \bar{\mathbf{C}}_{zf} \mathbf{C}_{ff}^{-1}$ and where

$$\mathbf{H}_3(L) = (\bar{\mathbf{C}}_{zf} \mathbf{C}_{ff}^{-1} \mathbf{C}_{ff}^*(L) - \bar{\mathbf{C}}_{zf}^*(L)) (\bar{\mathbf{C}}_{zf} + \bar{\mathbf{C}}_{zf}^*(L)(1-L))^{-1}, \quad (3.B.39)$$

is a lag polynomial that is a function of model parameters and need not be known for estimation purposes (See Remark 3.12), and where

$$\begin{aligned} (\overline{\mathbf{C}}_{zf} + \overline{\mathbf{C}}_{zf}^*(L)(1-L))^{-1}_{k \times (1+m)} &= [(\overline{\mathbf{C}}_{zf} + \overline{\mathbf{C}}_{zf}^*(L)(1-L))'(\overline{\mathbf{C}}_{zf} + \overline{\mathbf{C}}_{zf}^*(L)(1-L))]^{-1} \\ &\quad \times (\overline{\mathbf{C}}_{zf} + \overline{\mathbf{C}}_{zf}^*(L)(1-L))' \end{aligned}$$

is the left inverse of $(\overline{\mathbf{C}}_{zf} + \overline{\mathbf{C}}_{zf}^*(L)(1-L))$. Finally, rearranging the above equation yields

$$\begin{aligned} \bar{\mathbf{z}}_t - \mathbf{H}_3(L)\overline{\Delta\mathbf{z}}_t - \mathbf{H}'_1\mathbf{f}_t &= \frac{1}{N} \sum_{i=1}^N \mathbf{C}_{zzi} \sum_{s=1}^t \boldsymbol{\varepsilon}_{zis} + \frac{1}{N} \sum_{i=1}^N \mathbf{C}_{zzi}^*(L)\boldsymbol{\varepsilon}_{zit} \\ &\quad + \mathbf{H}_3(L) \frac{1}{N} \sum_{i=1}^N (\mathbf{C}_{zzi} + \mathbf{C}_{zzi}^*(L)(1-L))\boldsymbol{\varepsilon}_{zit}. \end{aligned} \tag{3.B.40}$$

For further use we write

$$\bar{\mathbf{z}}_t - \mathbf{H}'_1\mathbf{f}_t = \mathbf{H}_3(L)\overline{\Delta\mathbf{z}}_t + \mathbf{v}_{At} \tag{3.B.41}$$

by defining

$$\mathbf{v}_{At} = \frac{1}{N} \sum_{i=1}^N \mathbf{C}_{zzi} \sum_{s=1}^t \boldsymbol{\varepsilon}_{zis} + \frac{1}{N} \sum_{i=1}^N (\mathbf{C}_{zzi}^*(L) + \mathbf{H}_3(L)(\mathbf{C}_{zzi} + \mathbf{C}_{zzi}^*(L)(1-L))) \boldsymbol{\varepsilon}_{zit}.$$

Lemma 3.B.5. *Consider \mathbf{v}_{At} defined above, we have*

$$\mathbf{v}_{At} = O_p(T^{1/2}N^{-1/2}). \tag{3.B.42}$$

Proof of Lemma 3.B.5: The result can be obtained by using the result in (3.B.49) Lemma 3.B.6, which is provided in the next section.

Lemmas and proofs

In this section we provide some preliminary results to be used in the proofs of the theorems. We group the preliminary results based on the components involved in the cross products. For instance in Lemma 3.B.6-3.B.7 we present the probability orders of the cross products involving the approximation errors.

Throughout the proofs we use some of the definitions and representations stated in Appendix 3.B.1, such as the partial sum \mathbf{s}_{ft} defined in (3.B.5) and the common

trends representation for \mathbf{x}_{it} and \mathbf{f}_t which are stated in (3.B.7) and (3.B.8) respectively. We introduce some further notation regarding the partial sum processes, such as

$$\mathbf{s}_{zit} = \sum_{s=1}^t \boldsymbol{\varepsilon}_{zis}, \quad (3.B.43)$$

where $\boldsymbol{\varepsilon}_{zit} = (\varepsilon_{yit}, \boldsymbol{\varepsilon}'_{xit})'$.

At this stage it is useful to state some of the standard results that will be used in the proofs. The partial sum processes defined above in (3.B.5) and (3.B.43) are obtained by cumulating i.i.d. increments that implies the order results stated below.

$$\frac{1}{T^2} \sum_{t=1}^T \mathbf{s}_{zit} \mathbf{s}'_{zjt} = O_p(1), \quad \frac{1}{T^2} \sum_{t=1}^T \mathbf{s}_{ft} \mathbf{s}'_{ft} = O_p(1), \quad (3.B.44)$$

$$\frac{1}{T^2} \sum_{t=1}^T \mathbf{s}_{zit} \mathbf{s}'_{ft} = O_p(1), \quad (3.B.45)$$

$$\frac{1}{T} \sum_{t=1}^T \mathbf{s}_{zit} \boldsymbol{\varepsilon}'_{zjt} = O_p(1), \quad \frac{1}{T} \sum_{t=1}^T \mathbf{s}_{ft} \boldsymbol{\varepsilon}'_{ft} = O_p(1), \quad (3.B.46)$$

$$\frac{1}{T} \sum_{t=1}^T \mathbf{s}_{zit} \boldsymbol{\varepsilon}'_{ft} = O_p(1), \quad (3.B.47)$$

for all $i, j = 1, \dots, N$. Furthermore, considering the error correction term u_{it} that is defined in (3.B.13) in vector notation, which is a covariance stationary process under cointegration, we can state the following standard results.

$$\frac{1}{T} \sum_{t=1}^T \mathbf{s}_{zit} u_{jt} = O_p(1), \quad \frac{1}{T} \sum_{t=1}^T \mathbf{s}_{ft} u_{it} = O_p(1), \quad (3.B.48)$$

for all $i, j = 1, \dots, N$.

Remark 3.B.1. The augmentation of the lagged first differences of the cross-sectional averages until their p_T^{th} lag order causes the number of time series observations to drop to $T - p_T$ (See the definitions of the matrices below (3.28)). This implies that we should use $T - p_T$ to scale the cross products. But we work under the condition that $T - p_T \rightarrow \infty$, which implies that using T or $T - p_T$ would be asymptotically equivalent. In order to keep the notation simple we use T as a

scaling factor for the cross products. Note that T and $T - p_T$ are interchangeable from an asymptotic point of view.

The results provided in the lemma below involve the cross products with the approximation error, $\mathbf{V}_{A,-1}$ that is a $(T - p_T) \times (m + 1)$ matrix. The t^{th} row of \mathbf{V}_A is defined in (3.25) and includes the cross-sectional average of a partial-sum process, this gives occasion to use the results from the literature on non-stationary processes, especially the results stated in (3.B.44)-(3.B.48).

Lemma 3.B.6. *Let $\mathbf{V}_{A,-1} = (\mathbf{v}_{A,p_T+1}, \dots, \mathbf{v}_{A,T-1})'$, where $\mathbf{v}_{A,t}$ is defined in (3.25). Under the assumptions stated in the chapter, the probability orders of the cross products with $\mathbf{V}_{A,-1}$ can be stated as*

$$T^{-1} \mathbf{V}'_{A,-1} \mathbf{V}_{A,-1} = O_p(TN^{-1}), \quad (3.B.49)$$

$$T^{-1} \mathbf{U}'_{i,-1} \mathbf{V}_{A,-1} = O_p(N^{-1/2}), \quad (3.B.50)$$

$$T^{-1} \mathbf{X}'_{i,-1} \mathbf{V}_{A,-1} = O_p(TN^{-1/2}), \quad (3.B.51)$$

$$T^{-1} \mathbf{F}'_{-1} \mathbf{V}_{A,-1} = O_p(TN^{-1/2}), \quad (3.B.52)$$

$$T^{-1/2} \mathbf{V}'_{A,-1} \boldsymbol{\varepsilon}_{y.xi} = O_p(T^{1/2} N^{-1/2}), \quad (3.B.53)$$

uniformly in i , where $i = 1, \dots, N$.

Proof of Lemma 3.B.6: We start with the definition of $\mathbf{v}_{A,t}$ stated in (3.25) and let $\mathbf{C}_{vai}^*(L) = \mathbf{C}_{zzi}^*(L) + \mathbf{H}_3(L)(\mathbf{C}_{zzi} + \mathbf{C}_{zzi}^*(L)(1 - L))$ to simplify the notation, which allows us to write $\mathbf{v}_{A,t-1}$ as

$$\mathbf{v}_{A,t-1} = \frac{1}{N} \sum_{i=1}^N \mathbf{C}_{zzi} \mathbf{s}_{zi,t-1} + \frac{1}{N} \sum_{i=1}^N \mathbf{C}_{vai}^*(L) \boldsymbol{\varepsilon}_{zi,t-1}, \quad (3.B.54)$$

where the first term is the cross-sectional average of a partial sum process and the second term is the cross-sectional average of a stationary process. Throughout the proofs of this lemma we use some of the facts induced by our assumptions. Before starting to prove the results we list some of the properties of the parameters and the error processes that appear in the proofs:

- (P1) The entries in \mathbf{C}_{zzi} are: (i) bounded and (ii) independently distributed across cross-section units for all $i = 1, \dots, N$.
- (P2) Each lag-polynomial elements in $\mathbf{C}_{zzi}^*(L)$: (i) have absolutely summable and (ii) cross-sectionally independently distributed coefficients.

- (P3) Each lag polynomial elements of $\mathbf{C}_i^*(L)$: (i) have absolutely summable coefficients.
- (P4) The coefficients of the lag-polynomial $\mathbf{C}_{vai}^*(L)$ are absolutely summable.
- (P5) The elements of ε_{zit} are: (i) serially uncorrelated and (ii) independently distributed across cross-section units for all $t = 1, \dots, T$ and $i = 1, \dots, N$.

Now, to prove (3.B.49) we consider (3.B.54) and write,

$$\begin{aligned} \frac{\sum_{t=1}^{T-1} \mathbf{v}_{A,t-1} \mathbf{v}'_{A,t-1}}{T^2} &= \frac{1}{T^2 N^2} \sum_{t=1}^T \sum_{i=1}^N \sum_{j=1}^N \mathbf{C}_{zzi} \mathbf{s}_{zi,t-1} \mathbf{s}'_{zj,t-1} \mathbf{C}'_{zzj} \\ &+ \frac{1}{T^2 N^2} \sum_{t=1}^T \sum_{i=1}^N \sum_{j=1}^N \mathbf{C}_{zzi} \mathbf{s}_{zi,t-1} \varepsilon'_{zj,t-1} \mathbf{C}_{vaj}^*(L) \\ &+ \frac{1}{T^2 N^2} \sum_{t=1}^T \sum_{i=1}^N \sum_{j=1}^N \mathbf{C}_{vai}^*(L) \varepsilon_{zi,t-1} \mathbf{s}'_{zj,t-1} \mathbf{C}'_{zzj} \\ &+ \frac{1}{T^2 N^2} \sum_{t=1}^T \sum_{i=1}^N \sum_{j=1}^N \mathbf{C}_{vai}^*(L) \varepsilon_{zi,t-1} \varepsilon'_{zj,t-1} \mathbf{C}_{vaj}^*(L). \end{aligned}$$

The order is determined by the first term on the right-hand side, since that is the dominating term in terms of the order. It is straightforward to show that the remaining three terms are of smaller order by using the results in (3.B.44) and P1(i), P4 and P5. So we consider the first term on the right-hand side and write

$$\begin{aligned} &\frac{1}{T^2 N^2} \sum_{t=1}^{T-1} \sum_{i=1}^N \sum_{j=1}^N \mathbf{C}_{zzi} \mathbf{s}_{zi,t-1} \mathbf{s}'_{zj,t-1} \mathbf{C}'_{zzj} \\ &= \frac{1}{N^2} \sum_{i=1}^N \mathbf{C}_{zzi} \frac{\sum_{t=1}^T \mathbf{s}_{zi,t-1} \mathbf{s}'_{zi,t-1}}{T^2} \mathbf{C}'_{zzi} \\ &\quad + \frac{1}{N^2} \sum_{i=1}^N \sum_{j \neq i}^N \mathbf{C}_{zzi} \frac{\sum_{t=1}^T \mathbf{s}_{zi,t-1} \mathbf{s}'_{zj,t-1}}{T^2} \mathbf{C}'_{zzj}. \end{aligned}$$

By using the result stated in the first part of (3.B.44) and P1(i), it is straightforward to show that the first term on the right-hand side has an order of $O_p(N^{-1})$. As for the second term P1(ii) and P4(ii) imply that the order is $O_p(N^{-1/2})$. Then by multiplying the order of the last term with T yields the result stated in (3.B.49).

To prove (3.B.50), note that u_{it} is the error correction term so that under cointegration it is a covariance stationary process. Using (3.B.54) we obtain

$$\begin{aligned}
 \frac{\sum_{t=1}^T u_{i,t-1} \mathbf{v}'_{A,t-1}}{T} &= \frac{1}{N} \frac{1}{T} \sum_{t=1}^T u_{i,t-1} \mathbf{s}'_{zi,t-1} \mathbf{C}'_{zzi} \\
 &+ \frac{1}{\sqrt{N}} \frac{1}{T\sqrt{N}} \sum_{t=1}^T \sum_{j \neq i}^N u_{i,t-1} \mathbf{s}'_{zj,t-1} \mathbf{C}'_{zzj} \\
 &+ \frac{1}{N} \frac{1}{T} \sum_{t=1}^{T-1} u_{i,t-1} \boldsymbol{\varepsilon}'_{zi,t-1} \mathbf{C}^{*'}_{vai}(L) \\
 &+ \frac{1}{\sqrt{NT}} \frac{1}{\sqrt{NT}} \sum_{t=1}^T \sum_{j \neq i}^N u_{i,t-1} \boldsymbol{\varepsilon}'_{zj,t-1} \mathbf{C}^{*'}_{vaj}(L) \\
 &= O_p(N^{-1/2}),
 \end{aligned}$$

where the order is obtained by using the standard result stated in (3.B.48) and P1(i), P3 and P4.

To prove (3.B.51) we use the common trends representation for \mathbf{x}_{it} stated in (3.B.7) and the results stated in (3.B.44) and (3.B.46). We write

$$\begin{aligned}
 \frac{\sum_{t=1}^T \mathbf{x}_{i,t-1} \mathbf{v}'_{A,t-1}}{T} &= \frac{T}{N} \frac{1}{T^2} \sum_{t=1}^T \mathbf{C}_{xzi} \mathbf{s}_{zi,t-1} \mathbf{s}'_{zi,t-1} \mathbf{C}'_{zzi} \\
 &+ \frac{T}{\sqrt{N}} \frac{1}{T^2 \sqrt{N}} \sum_{t=1}^T \sum_{j \neq i}^N \mathbf{C}_{xzi} \mathbf{s}_{zi,t-1} \mathbf{s}'_{zj,t-1} \mathbf{C}'_{zzj} \\
 &+ \frac{T}{N} \frac{1}{T^2} \sum_{t=1}^T \mathbf{C}_{xfi} \mathbf{s}_{f,t-1} \mathbf{s}'_{zi,t-1} \mathbf{C}'_{zzi} \\
 &+ \frac{T}{\sqrt{N}} \frac{1}{T^2 \sqrt{N}} \sum_{t=1}^T \sum_{j \neq i}^N \mathbf{C}_{xfi} \mathbf{s}_{f,t-1} \mathbf{s}'_{zj,t-1} \mathbf{C}'_{zzj} \\
 &+ o_p(TN^{-1}) \\
 &= O_p(TN^{-1/2}).
 \end{aligned}$$

The terms which has a smaller order than $o_p(TN^{-1})$ are left out in the equation above to save space. The first and the third terms are of order $O_p(TN^{-1})$, which can be established by using the results in (3.B.44) and (3.B.45) and P1. The second and the fourth terms determine the order of the expression in question.

By using the results in (3.B.44) and (3.B.45) and P1, P2, P4, the result can be obtained.

(3.B.52) can be established using similar arguments hence the proof is omitted. This time one needs to use (3.B.8), the common trends representation for \mathbf{f}_t and the results provided in (3.B.44)-(3.B.46).

To prove (3.B.53), first note that $\varepsilon_{y.xit}$ is the scalar conditional error term. Using (3.B.54) we write

$$\begin{aligned}
 \frac{\sum_{t=1}^T \mathbf{v}_{A,t-1} \varepsilon_{y.xit}}{\sqrt{T}} &= \frac{\sqrt{T}}{N} \frac{1}{T} \sum_{t=1}^T \mathbf{C}_{zzi} \mathbf{s}_{zi,t-1} \varepsilon_{y.xit} \\
 &+ \frac{\sqrt{T}}{\sqrt{N}} \frac{1}{T\sqrt{N}} \sum_{j \neq i}^N \sum_{t=1}^T \mathbf{C}_{zzi} \mathbf{s}_{zi,t-1} \varepsilon_{y.xjt} \\
 &+ \frac{1}{N} \frac{1}{\sqrt{T}} \sum_{t=1}^T \mathbf{C}_{vai}^*(L) \varepsilon_{zi,t-1} \varepsilon_{y.xit} \\
 &+ \frac{1}{\sqrt{N}} \frac{1}{\sqrt{TN}} \sum_{j \neq i}^N \sum_{t=1}^T \mathbf{C}_{vai}^*(L) \varepsilon_{zi,t-1} \varepsilon_{y.xjt} \\
 &= O_p(T^{1/2} N^{-1/2}),
 \end{aligned}$$

where the order is determined by the second term, by using P1-P4. \blacksquare

Lemma 3.B.7. *Let $\mathbf{V}_B = (\mathbf{v}_{B,p_T+1}, \dots, \mathbf{v}_{B,T})'$, where \mathbf{v}_{Bt} is defined in (3.22). Under the assumption stated in the chapter, we have*

$$T^{-1/2} \mathbf{U}'_{i,-1} \mathbf{V}_{B,-1} = O_p(T^{1/2} N^{-1}) + O_p(N^{-1/2}), \quad (3.B.55)$$

$$T^{-1/2} \mathbf{X}'_{i,-1} \mathbf{V}_{B,-1} = O_p(T^{1/2} N^{-1/2}), \quad (3.B.56)$$

$$T^{-1/2} \mathbf{F}'_{-1} \mathbf{V}_{B,-1} = O_p(T^{1/2} N^{-1/2}), \quad (3.B.57)$$

$$T^{-1/2} \mathbf{V}'_{A,-1} \mathbf{V}_{B,-1} = O_p(T^{1/2} N^{-1}), \quad (3.B.58)$$

uniformly in i , where $i = 1, \dots, N$.

Proof of Lemma 3.B.7: To simplify the notation we use the definition $\mathbf{C}_{vbi}^*(L) = (\mathbf{C}_{zzi} + \mathbf{C}_{zzi}^*(L)(1-L))$, which was first used in (3.B.34), then for the t^{th} row of \mathbf{V}_B we can use the expression on the right-hand side of (3.B.35). Moreover we list some of the properties of the parameters that appear in the proofs used to establish the results.

(P6) The coefficients of the lag polynomial $\mathbf{C}_{vbi}^*(L)$ are (i) absolutely summable and (ii) cross-sectionally independent of each other for all $i = 1, \dots, N$.

To prove (3.B.55) using (3.B.14) we write

$$\begin{aligned} \frac{\sum_{t=1}^T u_{i,t-1} \mathbf{v}'_{B,t-1}}{\sqrt{T}} &= \frac{\sqrt{T}}{N} \frac{1}{T} \sum_{t=1}^T \beta_i \mathbf{C}_i^*(L) \boldsymbol{\varepsilon}_{i,t-1}^* \boldsymbol{\varepsilon}'_{zi,t-1} \mathbf{C}_{vbj}^{*'}(L) \\ &\quad + \frac{1}{\sqrt{N}} \frac{1}{\sqrt{TN}} \sum_{j \neq i}^N \sum_{t=1}^T \beta_i \mathbf{C}_i^*(L) \boldsymbol{\varepsilon}_{i,t-1}^* \boldsymbol{\varepsilon}'_{zj,t-1} \mathbf{C}_{vbj}^{*'}(L) \\ &= O_p(T^{1/2} N^{-1}) + O_p(N^{-1/2}). \end{aligned}$$

The probability order is obtained by using P3, P5, P6 and using Assumption 3.1(i), which assumes that $\boldsymbol{\varepsilon}_{zit}$ is a cross-sectionally independent covariance stationary processes for all $i, j = 1, \dots, N$ with finite fourth order moments.

The result in (3.B.56) can be proved by using (3.B.7). We write

$$\begin{aligned} \frac{\sum_{t=1}^T \mathbf{x}_{i,t-1} \mathbf{v}'_{B,t-1}}{\sqrt{T}} &= \frac{\sqrt{T}}{\sqrt{N}} \frac{1}{T\sqrt{N}} \sum_{j=1}^N \sum_{t=1}^T \mathbf{C}_{xzi} \mathbf{s}_{zi,t-1} \boldsymbol{\varepsilon}'_{zj,t-1} \mathbf{C}_{vbj}^{*'}(L) \\ &\quad + \frac{\sqrt{T}}{\sqrt{N}} \frac{1}{T\sqrt{N}} \sum_{j=1}^N \sum_{t=1}^T \mathbf{C}_{xfi} \mathbf{s}_{f,t-1} \boldsymbol{\varepsilon}'_{zj,t-1} \mathbf{C}_{vbj}^{*'}(L) \\ &\quad + \frac{\sqrt{T}}{\sqrt{N}} \frac{1}{T\sqrt{N}} \sum_{j=1}^N \sum_{t=1}^T \mathbf{C}_{xzi}^*(L) \boldsymbol{\varepsilon}_{zi,t-1} \boldsymbol{\varepsilon}'_{zj,t-1} \mathbf{C}_{vbj}^{*'}(L) \\ &\quad + \frac{\sqrt{T}}{\sqrt{N}} \frac{1}{T\sqrt{N}} \sum_{j=1}^N \sum_{t=1}^T \mathbf{C}_{xfi}^*(L) \boldsymbol{\varepsilon}_{f,t-1} \boldsymbol{\varepsilon}'_{zj,t-1} \mathbf{C}_{vbj}^{*'}(L) \\ &= O_p(T^{1/2} N^{-1/2}). \end{aligned}$$

The order is obtained by using the results in (3.B.46) and P1, P2, P5, P6 and Assumption 3.1(i).

(3.B.57)-(3.B.58) can be proved using similar arguments as we used to prove (3.B.56). ■

Next, we provide some further useful notation. Firstly, let

$$\tilde{\mathbf{U}}_{i,-1} = \mathbf{y}_{i,-1} - \mathbf{X}_{i,-1} \boldsymbol{\beta}_{xi} - \bar{\mathbf{Z}}_{-1} \mathbf{H}_1^- \boldsymbol{\lambda}_i, \quad (3.B.59)$$

where $\tilde{\mathbf{U}}_{i,-1}$ is the error correction term of the i^{th} group in which the true factors are replaced by the cross-sectional averages and \mathbf{H}_1^- is the right inverse of \mathbf{H}_1 such

that

$$\mathbf{H}_1^- = \mathbf{H}_1'(\mathbf{H}_1\mathbf{H}_1')^{-1}.$$

(3.B.59) can be rewritten in terms of the error correction term with true factors and approximation errors as

$$\tilde{\mathbf{U}}_{i,-1} = \mathbf{U}_{i,-1} - (\bar{\mathbf{Z}}_{-1} - \mathbf{F}_{-1}\mathbf{H}_1)\mathbf{H}_1^- \boldsymbol{\lambda}_i,$$

where $\mathbf{U}_{i,-1}$ is defined in (3.B.13). Secondly, stacking (3.24) over T yields

$$\bar{\mathbf{Z}}_{-1} - \mathbf{F}_{-1}\mathbf{H}_1 = \bar{\Delta\mathbf{Z}}_{-1}\mathbf{H}_3'(L) + \mathbf{V}_{A,-1}.$$

Now we premultiply $\tilde{\mathbf{U}}_{i,-1}$ and $\bar{\mathbf{Z}}_{-1}$ by the orthogonal projection matrix $\bar{\mathbf{M}}_Z$ defined in (3.33) we obtain

$$\bar{\mathbf{M}}_Z\tilde{\mathbf{U}}_{i,-1} = \bar{\mathbf{M}}_Z\mathbf{U}_{i,-1} - \bar{\mathbf{M}}_Z\mathbf{V}_{A,-1}\mathbf{H}_1^- \boldsymbol{\lambda}_i - \bar{\mathbf{M}}_Z\mathbf{V}_{C,-p_T}\mathbf{H}_1^- \boldsymbol{\lambda}_i, \quad (3.B.60)$$

$$\bar{\mathbf{M}}_Z\bar{\mathbf{Z}}_{-1} = \bar{\mathbf{M}}_Z\mathbf{F}_{-1}\mathbf{H}_1 + \bar{\mathbf{M}}_Z\mathbf{V}_{A,-1} + \bar{\mathbf{M}}_Z\mathbf{V}_{C,-p_T}, \quad (3.B.61)$$

where $\mathbf{V}_{C,-p_T}$ is the part which is not orthogonal to the space spanned by $\bar{\mathbf{G}}_z$ and can be defined as

$$\mathbf{V}_{C,-p_T} = \begin{pmatrix} (\sum_{\ell=p_T+1}^{\infty} \mathbf{H}_{3\ell}\bar{\Delta\mathbf{z}}_{p_T+1-\ell})' \\ (\sum_{\ell=p_T+1}^{\infty} \mathbf{H}_{3\ell}\bar{\Delta\mathbf{z}}_{p_T+2-\ell})' \\ \vdots \\ (\sum_{\ell=p_T+1}^{\infty} \mathbf{H}_{3\ell}\bar{\Delta\mathbf{z}}_{T-\ell})' \end{pmatrix}, \quad (3.B.62)$$

where $\mathbf{H}_3(L) = \sum_{\ell=0}^{\infty} \mathbf{H}_{3\ell}L^\ell$ is defined in (3.B.39). In the following lemma we present the results for the cross products involving $\mathbf{V}_{C,-p_T}$. But before that we need to show that the coefficients of $\mathbf{H}_3(L)$ has exponentially decaying coefficients. This result can be established by considering that it is obtained by a linear combination of lag-polynomials with exponentially decaying coefficients. For details see Section 3.3.2. Then again by following Chudik and Pesaran (2013c) we assume that there exist real positive constants $K < \infty$ and $\rho_H < 1$ such that

$$|\mathbf{H}_{3\ell}| < K\rho_H^\ell, \quad \text{for any } \ell \in \{0, 1, 2, \dots\}. \quad (3.B.63)$$

Lemma 3.B.8. *Let $\mathbf{V}_{C,-p_T}$ be as defined by (3.B.62) and ρ_H is the exponential decay rate of the coefficients of the lag polynomial $\mathbf{H}_3(L)$, such that $|\mathbf{H}_{3\ell}| < K\rho_H^\ell$, for any $\ell \in \{0, 1, 2, \dots\}$ where $K < \infty$ and $\rho_H < 1$ are real positive constants. And*

let p_T be the truncation lag order for the first differences of the approximations for the unobserved factors in the augmented model (3.28) and let $T - p_T \rightarrow \infty$. Under the assumptions made in the chapter, we have

$$T^{-1} \mathbf{X}'_{i,-1} \mathbf{V}_{C,-p_T} = O_p(\rho_H^{p_T}), \quad (3.B.64)$$

$$T^{-1} \mathbf{F}'_{-1} \mathbf{V}_{C,-p_T} = O_p(\rho_H^{p_T}), \quad (3.B.65)$$

$$T^{-1} \mathbf{U}'_i \mathbf{V}_{C,-p_T} = O_p(\rho_H^{p_T}), \quad (3.B.66)$$

$$T^{-1/2} \mathbf{V}'_{C,-p_T} \boldsymbol{\varepsilon}_{y.xi} = O_p(\rho_H^{p_T}), \quad (3.B.67)$$

$$T^{-1} \mathbf{V}'_{A,-1} \mathbf{V}_{C,-p_T} = O_p(\rho_H^{p_T} N^{-1/2}), \quad (3.B.68)$$

$$T^{-1/2} \mathbf{V}'_{C,-p_T} \mathbf{V}_{B,-1} = O_p(\rho_H^{p_T} T^{1/2} N^{-1/2}), \quad (3.B.69)$$

$$T^{-1} \mathbf{V}'_{C,-p_T} \mathbf{V}_{C,-p_T} = O_p(\rho_H^{2p_T}), \quad (3.B.70)$$

uniformly in i for all $i = 1, \dots, N$.

Proof of Lemma 3.B.8: To prove (3.B.64) we write the right-hand side as

$$\begin{aligned} \frac{\mathbf{X}'_{i,-1} \mathbf{V}_{C,-p_T}}{T} &= T^{-1} \sum_{t=1}^T \mathbf{x}_{i,t-1} \sum_{\ell=p_T+1}^{\infty} \overline{\Delta \mathbf{z}'_{t-\ell}} \mathbf{H}'_{3\ell} \\ &= \sum_{\ell=p_T+1}^{\infty} \left(\frac{\sum_{t=1}^T \mathbf{x}_{i,t-1} \overline{\Delta \mathbf{z}'_{t-\ell}}}{T} \right) \mathbf{H}'_{3\ell}. \end{aligned}$$

Then we can write

$$E \left| \sum_{\ell=p_T+1}^{\infty} \left(\frac{\sum_{t=1}^T \mathbf{x}_{i,t-1} \overline{\Delta \mathbf{z}'_{t-\ell}}}{T} \right) \mathbf{H}'_{3\ell} \right| \leq \sum_{\ell=p_T+1}^{\infty} E \left| \frac{\sum_{t=1}^T \mathbf{x}_{i,t-1} \overline{\Delta \mathbf{z}'_{t-\ell}}}{T} \right| |\mathbf{H}'_{3\ell}|$$

From standard unit root asymptotics we know that $E \left| \frac{\sum_{t=1}^T \mathbf{x}_{i,t-1} \overline{\Delta \mathbf{z}'_{t-\ell}}}{T} \right| = O(1)$ and the corresponding limit is independently distributed of the coefficients of the lag-polynomial $\mathbf{H}_3(L)$. Now consider the coefficients for the lags, we have

$$\sum_{\ell=p_T+1}^{\infty} |\mathbf{H}_{3\ell}| \leq K \frac{\rho_H^{p_T}}{1 - \rho_H}, \quad (3.B.71)$$

since $\mathbf{H}_{3\ell}$ are exponentially decaying as $\ell \rightarrow \infty$. This implies

$$\sum_{\ell=p_T+1}^{\infty} \left(\frac{\sum_{t=1}^T \mathbf{x}_{i,t-1} \overline{\Delta \mathbf{z}'_{t-\ell}}}{T} \right) \mathbf{H}'_{3\ell} = O_p(\rho_H^{p_T}).$$

The second result in the lemma, (3.B.65), can be proved in a similar way by noting that $E \left| \frac{\sum_{t=1}^T \mathbf{f}_{t-1} \overline{\Delta \mathbf{z}'_{t-\ell}}}{T} \right| < K$ by standard unit root asymptotics.

Now consider (3.B.66), by taking the expectation of the expression on the left-hand side we obtain

$$\begin{aligned} E \left| \frac{1}{T} \sum_{t=p_T+1}^T u_{i,t-1} \mathbf{v}'_{C,t-p_T} \right| &\leq \frac{1}{T} \sum_{t=p_T+1}^T E |u_{i,t-1} \mathbf{v}'_{C,t-p_T}| \\ &\leq \frac{1}{T} \sum_{t=p_T+1}^T [E(u_{i,t-1})^2 E(\mathbf{v}_{C,t-p_T} \mathbf{v}'_{C,t-p_T})]^{1/2} \\ &\leq \frac{1}{T} \sum_{t=p_T+1}^T E [u_{i,t-1}^2]^{1/2} \\ &\quad \times \sum_{\ell=p_T+1}^{\infty} |\mathbf{H}_{3\ell}| \left[E(\overline{\Delta \mathbf{z}_{t-\ell}} \overline{\Delta \mathbf{z}'_{t-\ell}}) \right]^{1/2}, \end{aligned}$$

where we used the triangle inequality to obtain the first and the third inequalities and Cauchy-Schwarz inequality to obtain the second inequality. Under the assumptions we made, $E [u_{it}^2]^{1/2}$ and $\left[E(\overline{\Delta \mathbf{z}_{t-\ell}} \overline{\Delta \mathbf{z}'_{t-\ell}}) \right]^{1/2}$ are $O(1)$. By using these results and (3.B.71) we obtain

$$\frac{1}{T} \sum_{t=p_T+1}^T u_{it} \mathbf{v}'_{C,t-p_T} = O_p(\rho_H^{p_T}).$$

The result in (3.B.67) can be established using similar arguments by noting that $\varepsilon_{y,xi}$ is the serially uncorrelated conditional error term which is independent of the process driving $\mathbf{V}_{C,-p_T}$.

To prove (3.B.68) we start with writing the expectation of an individual element of the matrix on the left-hand side as

$$\begin{aligned} E \left| \frac{1}{T} \sum_{t=1}^T \mathbf{v}_{A,t-1} \mathbf{v}'_{C,t-p_T} \right| &\leq \frac{1}{T} \sum_{t=p_T+1}^T E |\mathbf{v}_{A,t-1} \mathbf{v}'_{C,t-p_T}| \\ &\leq \frac{1}{\sqrt{N}} \sum_{\ell=p_T+1}^{\infty} E \left| \frac{\sqrt{N} \sum_{t=p_T+1}^T \mathbf{v}_{A,t-1} \overline{\Delta \mathbf{z}'_{t-\ell}}}{T} \right| |\mathbf{H}'_{3\ell}|, \end{aligned}$$

where we used triangle inequality to obtain the inequalities. Analogous to the results we obtain regarding the cross products involving \mathbf{V}_A , we can show that

$E \left| \frac{\sqrt{N} \sum_{t=p_T+1}^T \mathbf{v}_{A,t-1} \overline{\Delta \mathbf{z}'_{t-\ell}}}{T} \right| = O(1)$. By using this result and (3.B.71) yields

$$\frac{1}{T} \sum_{t=1}^T \mathbf{v}_{A,t-1} \mathbf{v}'_{C,t-p_T} = O_p \left(\frac{\rho_H^{p_T}}{\sqrt{N}} \right).$$

(3.B.69) can be established similarly to (3.B.68).

To prove (3.B.70) we write the left-hand side as

$$\begin{aligned} E \left| \frac{1}{T} \sum_{t=1}^T \mathbf{v}_{C,t-p_T} \mathbf{v}'_{C,t-p_T} \right| &\leq \frac{1}{T} \sum_{t=p_T+1}^T E |\mathbf{v}_{C,t-p_T} \mathbf{v}'_{C,t-p_T}| \\ &\leq \frac{1}{T} \sum_{t=p_T+1}^T [E(\mathbf{v}_{C,t-p_T})^2 E(\mathbf{v}_{C,t-p_T})^2]^{1/2} \\ &\leq \frac{1}{T} \sum_{t=p_T+1}^T \sum_{\ell=p_T+1}^{\infty} |\mathbf{H}_{3\ell}| \left[E(\overline{\Delta \mathbf{z}_{t-p_T}} \overline{\Delta \mathbf{z}'_{t-p_T}}) \right]^{1/2} \\ &\quad \times \sum_{\ell=p_T+1}^{\infty} |\mathbf{H}_{3\ell}| \left[E(\overline{\Delta \mathbf{z}_{t-p_T}} \overline{\Delta \mathbf{z}'_{t-p_T}}) \right]^{1/2}, \end{aligned}$$

where we used the triangle inequality to obtain the first and the third inequalities and the Cauchy-Schwarz inequality for the second inequality. As mentioned before $\left[E(\overline{\Delta \mathbf{z}_{t-\ell}} \overline{\Delta \mathbf{z}'_{t-\ell}}) \right]^{1/2}$ is $O(1)$. This together with (3.B.71) imply the resulting order, which can be stated formally as

$$\frac{1}{T} \sum_{t=1}^T \mathbf{v}_{C,t-p_T} \mathbf{v}'_{C,t-p_T} = O_p(\rho_H^{2p_T}).$$

■

Next, we consider the error term of the augmented model, $\tilde{\varepsilon}_{y.xi}$, that is defined in (3.29) as

$$\begin{aligned} \tilde{\varepsilon}_{y.xit} &= \varepsilon_{y.xit} + \mathbf{V}_{A,-1} \mathbf{H}_1^{-1} \boldsymbol{\lambda}_i \alpha_{yi} - \mathbf{V}_{B,-1} \mathbf{H}_4^-(L) \mathbf{A}'_{y.fi}(L) \\ &\quad + \overline{\Delta \mathbf{Z}}_{p_T-1} \mathbf{A}_{y.fi}^{**'}(L). \end{aligned} \tag{3.B.72}$$

The fourth term represents the truncation error. We show that, under certain conditions, this term does not effect the consistency and also the asymptotic dis-

tribution of our estimators. For this purpose we define

$$\mathbf{V}_{D_i, -p_T} = \begin{pmatrix} (\sum_{\ell=p_T+1}^{\infty} \mathbf{A}_{y.fi, \ell+p_T}^* \overline{\Delta \mathbf{z}_{p_T+1-\ell}})' \\ (\sum_{\ell=p_T+1}^{\infty} \mathbf{A}_{y.fi, \ell+p_T}^* \overline{\Delta \mathbf{z}_{p_T+2-\ell}})' \\ \vdots \\ (\sum_{\ell=p_T+1}^{\infty} \mathbf{A}_{y.fi, \ell+p_T}^* \overline{\Delta \mathbf{z}_{T-\ell}})' \end{pmatrix}, \quad (3.B.73)$$

similarly to $\mathbf{V}_{C, -p_T}$. Now we can rewrite (3.B.72) as

$$\begin{aligned} \tilde{\varepsilon}_{y.xfit} &= \varepsilon_{y.xi} + \mathbf{V}_{A, -1} \mathbf{H}_1^{-1} \lambda_i \alpha_{yi} - \mathbf{V}_{B, -1} \mathbf{H}_4^{-1}(L) \mathbf{A}'_{y.fi}(L) \\ &\quad + \mathbf{V}_{D_i, -p_T}. \end{aligned} \quad (3.B.74)$$

In the lemma below, we provide the probability order for the cross products that involve $\mathbf{V}_{D_i, -p_T}$.

Lemma 3.B.9. *Let $\mathbf{V}_{D, -p_T}$ be as defined by (3.B.73) and ρ_A is the exponential decay rate of the coefficients of the lag polynomial $\mathbf{A}_{y.fi}^*(L) = \sum_{\ell=0}^{\infty} \mathbf{A}_{y.fi}^* L^\ell$, such that $|\mathbf{A}_{y.fi, \ell}^*| < K \rho_A^\ell$ for any $\ell \in \{0, 1, 2, \dots\}$ where $K < \infty$ and $\rho_A < 1$ are real positive constants. And let p_T be the truncation lag order for the first differences of the approximations for the unobserved factors in the augmented model and let $T - p_T \rightarrow \infty$. Under the assumptions made in the chapter, we have*

$$T^{-1} \mathbf{U}'_i \mathbf{V}_{D_i, -p_T} = O_p(\rho_A^{p_T}), \quad (3.B.75)$$

$$T^{-1} \mathbf{X}'_{i, -1} \mathbf{V}_{D_i, -p_T} = O_p(\rho_A^{p_T}), \quad (3.B.76)$$

$$T^{-1} \mathbf{F}'_{-1} \mathbf{V}_{D_i, -p_T} = O_p(\rho_A^{p_T}), \quad (3.B.77)$$

$$T^{-1} \mathbf{V}'_{A, -1} \mathbf{V}_{D_i, -p_T} = O_p(\rho_A^{p_T} N^{-1/2}), \quad (3.B.78)$$

$$T^{-1} \mathbf{V}'_{C, -p_T} \mathbf{V}_{D_i, -p_T} = O_p(\rho_A^{2p_T}), \quad (3.B.79)$$

uniformly in i for all $i = 1, \dots, N$.

Proof of Lemma 3.B.9: The proof of the lemma follows from the proof of Lemma 3.B.8. One can make use of the method used in the previous lemma, noting that $\mathbf{V}_{D_i, -p_T}$ defined similarly to $\mathbf{V}_{C, -p_T}$ and the lag polynomial $\mathbf{A}_{y.fi}^{**'}(L)$ has similar properties, i.e. the coefficients of $\mathbf{A}_{y.fi}^{**'}(L)$ are exponentially decaying with a rate of $\rho_A^{p_T}$ as $p_T \rightarrow \infty$. ■

In Section 3.3.2 we define two different orthogonal projection matrices in (3.37) and (3.33) as

$$\begin{aligned}\widetilde{\mathbf{M}}_i &= \mathbf{I} - \widetilde{\mathbf{G}}_i(\widetilde{\mathbf{G}}_i' \widetilde{\mathbf{G}}_i)^{-1} \widetilde{\mathbf{G}}_i', \\ \widetilde{\mathbf{M}}_z &= \mathbf{I} - \overline{\mathbf{G}}_z(\overline{\mathbf{G}}_z' \overline{\mathbf{G}}_z)^{-1} \overline{\mathbf{G}}_z',\end{aligned}\quad (3.B.80)$$

where $\widetilde{\mathbf{G}}_i = (\mathbf{G}_{yi}, \mathbf{G}_{xi}, \overline{\mathbf{G}}_z)$ and $\overline{\mathbf{G}}_z$ are $(T - p_T) \times q$ and $(T - p_T) \times p_T(1 + m)$ matrices respectively, where $q = (p - 1) + pm + p_T(1 + m)$. \mathbf{G}_{yi} , \mathbf{G}_{xi} and $\overline{\mathbf{G}}_z$ are defined below (3.28). As it can be seen from the definition of $\widetilde{\mathbf{M}}_i$, it projects on the orthogonal complement of the vector space spanned by the lagged values of $\Delta \mathbf{y}_i$, current and lagged values of $\Delta \mathbf{X}_i$'s until the p^{th} lag and lagged values of the cross-sectional averages, $\Delta \mathbf{Z}$ until the p_T^{th} lag. In addition to these, we define two more orthogonal projection matrices to be used in the proofs of the theorems. Let

$$\ddot{\mathbf{M}}_i = \mathbf{I} - \ddot{\mathbf{G}}_i(\ddot{\mathbf{G}}_i' \ddot{\mathbf{G}}_i)^+ \ddot{\mathbf{G}}_i', \quad (3.B.81)$$

$$\ddot{\mathbf{M}}_F = \mathbf{I} - \ddot{\mathbf{G}}_F(\ddot{\mathbf{G}}_F' \ddot{\mathbf{G}}_F)^+ \ddot{\mathbf{G}}_F', \quad (3.B.82)$$

with $\ddot{\mathbf{G}}_i = (\mathbf{G}_{yi}, \mathbf{G}_{xi}, \ddot{\mathbf{G}}_F)$, which is a $(T - p_T) \times q$ matrix with $\ddot{\mathbf{G}}_F = (\ddot{\mathbf{g}}_{f,p_T+1}, \dots, \ddot{\mathbf{g}}_{fT})'$, where

$$\ddot{\mathbf{g}}_{ft} = \begin{pmatrix} \mathbf{H}'_4(L) \Delta \mathbf{f}_{t-1} \\ \mathbf{H}'_4(L) \Delta \mathbf{f}_{t-2} \\ \vdots \\ \mathbf{H}'_4(L) \Delta \mathbf{f}_{t-p_T} \end{pmatrix}, \quad (3.B.83)$$

where the definition of $\mathbf{H}'_4(L)$ can be found in (3.17). From (3.22) in Section 3.3.2 we know that $\mathbf{H}_4(L) = \sum_{\ell=0}^{\infty} H_{4\ell} L^\ell$, then we can also write

$$\ddot{\mathbf{g}}_{ft} = \left(\sum_{\ell=0}^{\infty} \Delta \mathbf{f}'_{t-\ell} H_{4\ell}, \sum_{\ell=0}^{\infty} \Delta \mathbf{f}'_{t-1-\ell} H_{4\ell}, \dots, \sum_{\ell=0}^{\infty} \Delta \mathbf{f}'_{t-p_T-\ell} H_{4\ell} \right)'$$

Remark 3.B.2. Following Chudik and Pesaran (2013c), $\ddot{\mathbf{g}}_{ft}$, can be considered as the large N counterpart of $\overline{\mathbf{g}}_{zt}$, in the sense that

$$\overline{\mathbf{g}}_{zt} = \ddot{\mathbf{g}}_{ft} + O_p(N^{-1/2}). \quad (3.B.84)$$

Note that the orthogonal projection matrices are defined using Moore-Penrose inverses instead of standard inverses. This is due to the rank assumption, As-

sumption 3.4(ii), we make for $\mathbf{H}_4(L)$, which implies that $\mathbf{H}_4(L)$ has a rank of k , which in turn implies that $(\ddot{\mathbf{G}}_F' \ddot{\mathbf{G}}_F)$ has a reduced rank that is equal to kp_T .

Remark 3.B.3. The orthogonal projection matrices $\ddot{\mathbf{M}}_i$ and $\ddot{\mathbf{M}}_F$ are defined analogously to \mathbf{M}_q in Kapetanios et al. (2011). In our set-up $\ddot{\mathbf{M}}_i$ projects on the orthogonal complement of the space of the first differences of the true factors multiplied by the polynomial rotation matrix besides the lagged values of Δy_{it} and current and lagged values of $\Delta \mathbf{x}_{it}$.

In the following lemma we present the results for the matrices which includes $\ddot{\mathbf{G}}_i$ and $\ddot{\mathbf{G}}_F$.

Lemma 3.B.10. *Let $\ddot{\mathbf{G}}_i$ be as defined below (3.B.82) and $\ddot{\mathbf{G}}_F$ be as defined in (3.B.83). As $(N, T, p_T) \rightarrow \infty$, where $\frac{p_T^2}{T} \rightarrow 0$, then we have*

$$T^{-1} \mathbf{U}'_{i,-1} \ddot{\mathbf{G}}_i = O_p(1), \quad T^{-1} \mathbf{U}'_{i,-1} \ddot{\mathbf{G}}_F = O_p(1), \quad (3.B.85)$$

$$T^{-1} \mathbf{X}'_{i,-1} \ddot{\mathbf{G}}_i = O_p(1), \quad T^{-1} \mathbf{X}'_{i,-1} \ddot{\mathbf{G}}_F = O_p(1), \quad (3.B.86)$$

$$T^{-1} \mathbf{F}'_{-1} \ddot{\mathbf{G}}_i = O_p(1), \quad T^{-1} \mathbf{F}'_{-1} \ddot{\mathbf{G}}_F = O_p(1), \quad (3.B.87)$$

$$T^{-1} \ddot{\mathbf{G}}_i' \ddot{\mathbf{G}}_i = O_p(1), \quad T^{-1} \ddot{\mathbf{G}}_F' \ddot{\mathbf{G}}_F = O_p(1), \quad (3.B.88)$$

$$T^{-1/2} \ddot{\mathbf{G}}_i' \boldsymbol{\varepsilon}_{y.xi} = O_p(1), \quad T^{-1/2} \ddot{\mathbf{G}}_F' \boldsymbol{\varepsilon}_{y.xi} = O_p(1), \quad (3.B.89)$$

$$T^{-1} \mathbf{V}'_{A,-1} \ddot{\mathbf{G}}_i = O_p(N^{-1/2}), \quad T^{-1} \mathbf{V}'_{A,-1} \ddot{\mathbf{G}}_F = O_p(N^{-1/2}), \quad (3.B.90)$$

$$T^{-1/2} \mathbf{V}'_B \ddot{\mathbf{G}}_i = O_p(T^{1/2} N^{-1/2}), \quad T^{-1/2} \mathbf{V}'_B \ddot{\mathbf{G}}_F = O_p(N^{-1/2}), \quad (3.B.91)$$

$$T^{-1} \bar{\mathbf{Z}}'_{-1} \ddot{\mathbf{G}}_i = O_p(1), \quad T^{-1} \bar{\mathbf{Z}}'_{-1} \ddot{\mathbf{G}}_F = O_p(1), \quad (3.B.92)$$

$$T^{-1} \ddot{\mathbf{G}}_i' \mathbf{V}_{C,-p_T} = O_p(\rho_H^{p_T}), \quad T^{-1} \ddot{\mathbf{G}}_F' \mathbf{V}_{C,-p_T} = O_p(\rho_H^{p_T}), \quad (3.B.93)$$

$$T^{-1} \ddot{\mathbf{G}}_i' \mathbf{V}_{Di,-p_T} = O_p(\rho_A^{p_T}), \quad T^{-1} \ddot{\mathbf{G}}_F' \mathbf{V}_{Di,-p_T} = O_p(\rho_A^{p_T}). \quad (3.B.94)$$

Proof of Lemma 3.B.10: Noting that the processes $\ddot{\mathbf{g}}_{it}$, $\ddot{\mathbf{g}}_{ft}$ and \mathbf{u}_{it} are stationary processes with absolute summable autocovariances and $\boldsymbol{\varepsilon}_{y.xi}$ is the conditional error term, which is serially uncorrelated, whereas \mathbf{x}_{it} , and \mathbf{f}_t are non-stationary processes with a common trends representation provided in (3.B.7) and (3.B.8) respectively, the results in (3.B.85)-(3.B.89) can be established straightforwardly by using the result in Berk (1974).

The results in (3.B.90)-(3.B.91) can be established by noting that $\mathbf{V}_{A,-1} = O_p(TN^{-1/2})$ and $\mathbf{V}_{B,-1} = O_p(\sqrt{T}N^{-1/2})$.

For (3.B.92), we use

$$\bar{\mathbf{Z}}_{-1} = \mathbf{F}_{-1} \mathbf{H}_1 + \mathbf{V}_{A,-1} + \bar{\Delta} \bar{\mathbf{Z}}_{-1} \mathbf{H}'_3(L).$$

Then by combining the results in (3.B.87), (3.B.90) and (3.B.91) and using the boundedness of \mathbf{H}_1 , we obtain the result.

To prove (3.B.93), let an individual element of the $(T - p_T) \times [(p - 1) + pm + p_T(1 + m)]$ matrix $\ddot{\mathbf{G}}_i$ be denoted as \ddot{g}_{its} and an individual element of $(T - p_T) \times (m + 1)$ matrix $\mathbf{V}_{C, -p_T}$ be denoted as $v_{C, t-p_T, j}$ where $j = 1, \dots, m + 1$. Then for an individual element of the cross product $T^{-1}\ddot{\mathbf{G}}_i'\mathbf{V}_{C, -p_T}$ we can write

$$\begin{aligned} E \left| \frac{1}{T} \sum_{t=1}^T \ddot{g}_{its} v_{C, t-p_T, j} \right| &\leq \frac{1}{T} \sum_{t=p_T+1}^T E |\ddot{g}_{its} v_{C, t-p_T, j}| \\ &\leq \frac{1}{T} \sum_{t=p_T+1}^T [E(\ddot{g}_{its})^2 E(v_{C, t-p_T, j})^2]^{1/2} \\ &\leq \frac{1}{T} \sum_{t=p_T+1}^T [E(\ddot{g}_{its})^2]^{1/2} \\ &\quad \times \sum_{\ell=p_T+1}^{\infty} |\mathbf{H}_{3m\ell}| \left[E(\overline{\Delta \mathbf{z}_{t-\ell}} \overline{\Delta \mathbf{z}'_{t-\ell}}) \right]^{1/2}, \end{aligned}$$

where $\mathbf{H}_{3m\ell}$ is the m^{th} column of $\mathbf{H}_{3\ell}$ where $m = 1, \dots, m + 1$. From the standard results we know that $[E(\ddot{g}_{its})^2]^{1/2} = O(1)$ and $[E(\overline{\Delta \mathbf{z}_{t-\ell}} \overline{\Delta \mathbf{z}'_{t-\ell}})]^{1/2} = O(1)$. Then using (3.B.71) we have

$$E \left| \frac{1}{T} \sum_{t=1}^T \ddot{g}_{its} v_{C, t-p_T, j} \right| = O(\rho_H^{p_T}),$$

which is the required result. (3.B.94) can be established using similar arguments. ■

In the following lemmas we present some asymptotic results regarding the difference between $\overline{\mathbf{M}}_z$ (defined in (3.33)) and $\ddot{\mathbf{M}}_F$ (defined in (3.B.82)), which is required to prove Theorem 3.3. These results can be generalized to the case with $\widetilde{\mathbf{M}}_i$ (defined in (3.37)) and $\ddot{\mathbf{M}}_i$ (defined in (3.B.81)) straightforwardly, which will be used in the proof of Theorem 3.4.

Lemma 3.B.11. *Note that $\overline{\mathbf{G}}_z$ and $\ddot{\mathbf{G}}_F$ are $(T - p_T) \times (m + 1)p_T$ random matrices, where in our set-up T and p_T are deterministic sequences nondecreasing in N . Let $\delta_{\ddot{\mathbf{G}}_F, i1}$ and $\delta_{\ddot{\mathbf{G}}_F, i2}$ represent any vectors belonging to the column space of $\ddot{\mathbf{G}}_F$. Similarly let $\delta_{\overline{\mathbf{G}}_z, i1}$ and $\delta_{\overline{\mathbf{G}}_z, i2}$ represent any vectors belonging to the column space of $\overline{\mathbf{G}}_z$. Suppose there exist vectors \mathbf{c}_{F1} , \mathbf{c}_{F2} , \mathbf{c}_{z1} and \mathbf{c}_{z2} which satisfy*

$\delta_{\ddot{G}_F, i1} = \ddot{\mathbf{G}}_F \mathbf{c}_{F1}$, $\delta_{\ddot{G}_F, i2} = \ddot{\mathbf{G}}_F \mathbf{c}_{F2}$, $\delta_{\overline{\mathbf{G}}_z, i1} = \overline{\mathbf{G}}_z \mathbf{c}_{G1}$ and $\delta_{\overline{\mathbf{G}}_z, i2} = \overline{\mathbf{G}}_z \mathbf{c}_{G2}$ such that $\|\mathbf{c}_{F1}\| < K$, $\|\mathbf{c}_{F2}\| < K$, $\|\mathbf{c}_{G1}\| < K$ and $\|\mathbf{c}_{G2}\| < K$, where $K < \infty$ is a constant which does not depend on N . Under these conditions we have the following results:

$$\begin{aligned} \|\overline{\mathbf{M}}_Z \delta_{\overline{\mathbf{G}}_z, i1}\| &= O_p\left(\frac{p_T \sqrt{T - p_T}}{\sqrt{N}}\right), \\ \|\ddot{\mathbf{M}}_F \delta_{\ddot{G}_F, i1}\| &= O_p\left(\frac{p_T \sqrt{T - p_T}}{\sqrt{N}}\right). \end{aligned}$$

Furthermore, we obtain the following results regarding the inner products of the vectors:

$$\begin{aligned} \langle \overline{\mathbf{M}}_Z \delta_{\overline{\mathbf{G}}_z, i1}, \overline{\mathbf{M}}_Z \delta_{\overline{\mathbf{G}}_z, i2} \rangle &= O_p\left(\frac{p_T^2 (T - p_T)}{N}\right), \\ \langle \ddot{\mathbf{M}}_F \delta_{\ddot{G}_F, i1}, \ddot{\mathbf{M}}_F \delta_{\ddot{G}_F, i2} \rangle &= O_p\left(\frac{p_T^2 (T - p_T)}{N}\right), \end{aligned}$$

where $\overline{\mathbf{M}}_Z$ and $\ddot{\mathbf{M}}_F$ are the orthogonal complement matrices project onto the orthogonal complements of the column spaces of $\overline{\mathbf{G}}_z$ and $\ddot{\mathbf{G}}_F$ respectively.

Proof of Lemma 3.B.11: This lemma is an adaptation of Lemma A.1 provided in Chudik and Pesaran (2013b) to our set-up. To adapt the set-up in the mentioned chapter to our set-up, we define

$$\mathbf{V}_{B, (p_T)} = (\mathbf{V}_{B, -1}, \mathbf{V}_{B, -2}, \dots, \mathbf{V}_{B, -p_T}). \quad (3.B.95)$$

By considering the definitions of $\overline{\mathbf{G}}_z$ and $\ddot{\mathbf{G}}_F$, we can write

$$\overline{\mathbf{G}}_z = \ddot{\mathbf{G}}_F + \mathbf{V}_{B, (p_T)}. \quad (3.B.96)$$

Note that $\|\overline{\mathbf{g}}_{Z\ell}\| = O_p(\sqrt{T - p_T})$, $\|\ddot{\mathbf{g}}_{F\ell}\| = O_p(\sqrt{T - p_T})$ and $\|\mathbf{v}_{B, (p_T)\ell}\| = O_p(\sqrt{T - p_T} N^{-1/2})$ for all $\ell = 1, 2, \dots, (m + 1)p_T$, where $\overline{\mathbf{g}}_{Z\ell}$, $\ddot{\mathbf{g}}_{F\ell}$ and $\mathbf{v}_{B, (p_T)\ell}$ are the ℓ^{th} columns of $\overline{\mathbf{G}}_z$, $\ddot{\mathbf{G}}_F$, $\mathbf{V}_{B, (p_T)}$ respectively. By an application of the Hilbert projection theorem (see Rudin (1987)), the proof of the result follows from the proof of Lemma A.1 Chudik and Pesaran (2013). \blacksquare

Lemma 3.B.12. Suppose $(N, T, p_T) \rightarrow \infty$ and $\frac{p_T^2}{T} \rightarrow 0$. Let \mathbf{A}_i and \mathbf{B}_i be any two matrices with random elements of size $(T - p_T) \times q_A$ and $(T - p_T) \times q_B$ respectively. Denoting the individual columns of \mathbf{A}_i and \mathbf{B}_i by \mathbf{a}_{ij} and \mathbf{b}_{is} , where $j = 1, \dots, q_A$

and $s = 1, \dots, q_B$, let \mathbf{A}_i and \mathbf{B}_i to satisfy

$$\begin{aligned} (\sqrt{T - p_T})^{-1} \|\ddot{\mathbf{M}}_F \mathbf{a}_{ij}\| &= O_p(R_A), \\ (\sqrt{T - p_T})^{-1} \|\ddot{\mathbf{M}}_F \mathbf{b}_{is}\| &= O_p(R_B), \\ (\sqrt{T - p_T})^{-1} \|\overline{\mathbf{M}}_Z \mathbf{a}_{ij}\| &= O_p(R_A), \\ (\sqrt{T - p_T})^{-1} \|\overline{\mathbf{M}}_Z \mathbf{b}_{is}\| &= O_p(R_B), \end{aligned}$$

for all individual columns, where R_A and R_B are deterministic functions of T , N , ρ_A and ρ_H . Then we have

$$\frac{\mathbf{A}'_i \overline{\mathbf{M}}_Z \mathbf{B}_i}{T} - \frac{\mathbf{A}'_i \ddot{\mathbf{M}}_F \mathbf{B}_i}{T} = O_p\left(\frac{p_T \times \max\{R_A, R_B\}}{\sqrt{N}}\right) + O_p\left(\frac{p_T^2}{N}\right), \quad (3.B.97)$$

uniformly in i .

Proof of Lemma 3.B.12: This lemma is an adaptation of Lemma A.6 of Chudik and Pesaran (2013b). Hence the proof follows similarly. We first use the Cauchy-Schwarz inequality to write

$$\mathbf{a}'_{ij} \overline{\mathbf{M}}_Z \mathbf{b}_{is} - \mathbf{a}'_{ij} \ddot{\mathbf{M}}_F \mathbf{b}_{is} \leq \|\overline{\mathbf{M}}_Z \mathbf{a}_{ij}\| \|\overline{\mathbf{M}}_Z \mathbf{b}_{is}\| - \|\ddot{\mathbf{M}}_F \mathbf{a}_{ij}\| \|\ddot{\mathbf{M}}_F \mathbf{b}_{is}\|. \quad (3.B.98)$$

Then by using the Hilbert projection theorem we can write

$$\|\overline{\mathbf{M}}_Z \mathbf{a}_{ij}\| \|\overline{\mathbf{M}}_Z \mathbf{b}_{is}\| \leq \|\mathbf{a}_{ij} - \boldsymbol{\delta}_{zai}\| \|\mathbf{b}_{is} - \boldsymbol{\delta}_{zbi}\|,$$

for any vectors $\boldsymbol{\delta}_{zai}$ and $\boldsymbol{\delta}_{zbi}$ belonging to the column space of $\overline{\mathbf{G}}_z$. For our purposes we choose $\boldsymbol{\delta}_{zai}$ and $\boldsymbol{\delta}_{zbi}$ as

$$\begin{aligned} \boldsymbol{\delta}_{zai} &= \ddot{\mathbf{P}}_F \mathbf{a}_{ij} - \overline{\mathbf{M}}_Z \ddot{\mathbf{P}}_F \mathbf{a}_{ij}, \\ \boldsymbol{\delta}_{zbi} &= \ddot{\mathbf{P}}_F \mathbf{b}_{is} - \overline{\mathbf{M}}_Z \ddot{\mathbf{P}}_F \mathbf{b}_{is}, \end{aligned}$$

where $\ddot{\mathbf{P}}_F = \mathbf{I} - \ddot{\mathbf{M}}_F$. Note that $\boldsymbol{\delta}_{zai} = (\mathbf{I}_{T-p_T} - \overline{\mathbf{M}}_Z) \ddot{\mathbf{P}}_F \mathbf{a}_{ij}$ and $\boldsymbol{\delta}_{zbi} = (\mathbf{I}_{T-p_T} - \overline{\mathbf{M}}_Z) \ddot{\mathbf{P}}_F \mathbf{b}_{is}$ indeed belong to the column space of $\overline{\mathbf{G}}_z$. Then we can write

$$\begin{aligned} \|\overline{\mathbf{M}}_Z \mathbf{a}_{ij}\| \|\overline{\mathbf{M}}_Z \mathbf{b}_{is}\| &\leq \|\mathbf{a}_{ij} - \ddot{\mathbf{P}}_F \mathbf{a}_{ij} + \overline{\mathbf{M}}_Z \ddot{\mathbf{P}}_F \mathbf{a}_{ij}\| \|\mathbf{b}_{is} - \ddot{\mathbf{P}}_F \mathbf{b}_{is} + \overline{\mathbf{M}}_Z \ddot{\mathbf{P}}_F \mathbf{b}_{is}\| \\ &\leq \|\ddot{\mathbf{M}}_F \mathbf{a}_{ij} + \overline{\mathbf{M}}_Z \ddot{\mathbf{P}}_F \mathbf{a}_{ij}\| \|\ddot{\mathbf{M}}_F \mathbf{b}_{is} + \overline{\mathbf{M}}_Z \ddot{\mathbf{P}}_F \mathbf{b}_{is}\| \\ &\leq \|\ddot{\mathbf{M}}_F \mathbf{a}_{ij}\| \|\ddot{\mathbf{M}}_F \mathbf{b}_{is}\| + \|\ddot{\mathbf{M}}_F \mathbf{a}_{ij}\| \|\overline{\mathbf{M}}_Z \ddot{\mathbf{P}}_F \mathbf{b}_{is}\| \\ &\quad + \|\overline{\mathbf{M}}_Z \ddot{\mathbf{P}}_F \mathbf{a}_{ij}\| \|\ddot{\mathbf{M}}_F \mathbf{b}_{is}\| + \|\overline{\mathbf{M}}_Z \ddot{\mathbf{P}}_F \mathbf{a}_{ij}\| \|\overline{\mathbf{M}}_Z \ddot{\mathbf{P}}_F \mathbf{b}_{is}\| \\ &= \|\ddot{\mathbf{M}}_F \mathbf{a}_{ij}\| \|\ddot{\mathbf{M}}_F \mathbf{b}_{is}\| + \epsilon_{iU,NT}. \end{aligned}$$

The third inequality yields an upper bound for $\|\overline{\mathbf{M}}_Z \mathbf{a}_{ij}\| \|\overline{\mathbf{M}}_Z \mathbf{b}_{is}\|$. An upper bound for $\|\ddot{\mathbf{M}}_F \mathbf{a}_{ij}\| \|\ddot{\mathbf{M}}_F \mathbf{b}_{is}\|$ can be obtained in a similar way as

$$\begin{aligned} \|\ddot{\mathbf{M}}_F \mathbf{a}_{ij}\| \|\ddot{\mathbf{M}}_F \mathbf{b}_{is}\| &\leq \|\overline{\mathbf{M}}_Z \mathbf{a}_{ij}\| \|\overline{\mathbf{M}}_Z \mathbf{b}_{is}\| + \|\overline{\mathbf{M}}_Z \mathbf{a}_{ij}\| \|\ddot{\mathbf{M}}_F \overline{\mathbf{P}}_Z \mathbf{b}_{is}\| \\ &\quad + \|\ddot{\mathbf{M}}_F \overline{\mathbf{P}}_Z \mathbf{a}_{ij}\| \|\overline{\mathbf{M}}_Z \mathbf{b}_{is}\| + \|\ddot{\mathbf{M}}_F \overline{\mathbf{P}}_Z \mathbf{a}_{ij}\| \|\ddot{\mathbf{M}}_F \overline{\mathbf{P}}_Z \mathbf{b}_{is}\| \\ &= \|\overline{\mathbf{M}}_Z \mathbf{a}_{ij}\| \|\overline{\mathbf{M}}_Z \mathbf{b}_{is}\| + \epsilon_{iL,NT}, \end{aligned}$$

where $\overline{\mathbf{P}}_Z = \mathbf{I} - \overline{\mathbf{M}}_Z$. The last two inequalities provide a lower and an upper bound for the expression on the right-hand side of (3.B.98), such that

$$\epsilon_{iL_B,NT} \leq \|\overline{\mathbf{M}}_Z \mathbf{a}_{ij}\| \|\overline{\mathbf{M}}_Z \mathbf{b}_{is}\| - \|\ddot{\mathbf{M}}_F \mathbf{a}_{ij}\| \|\ddot{\mathbf{M}}_F \mathbf{b}_{is}\| \leq \epsilon_{iU_B,NT}. \quad (3.B.99)$$

Evaluating the orders of the bounds and showing that the orders are equal to each other will give us the order of the term we are interested in. By setting $\delta_{\overline{G}_Z, i1} = \overline{\mathbf{P}}_Z \mathbf{a}_{ij}$, $\delta_{\overline{G}_Z, i2} = \overline{\mathbf{P}}_Z \mathbf{b}_{is}$, $\delta_{\ddot{G}_F, i1} = \ddot{\mathbf{P}}_F \mathbf{a}_{ij}$ and $\delta_{\ddot{G}_F, i2} = \ddot{\mathbf{P}}_F \mathbf{b}_{is}$, we can make use of Lemma 3.B.11 to establish the order results below.

$$\begin{aligned} \|\ddot{\mathbf{M}}_F \overline{\mathbf{P}}_Z \mathbf{a}_{ij}\| &= O_p \left(\frac{p_T \sqrt{T - p_T}}{\sqrt{N}} \right), \\ \|\ddot{\mathbf{M}}_F \overline{\mathbf{P}}_Z \mathbf{b}_{is}\| &= O_p \left(\frac{p_T \sqrt{T - p_T}}{\sqrt{N}} \right), \\ \|\overline{\mathbf{M}}_Z \ddot{\mathbf{P}}_F \mathbf{a}_{ij}\| &= O_p \left(\frac{p_T \sqrt{T - p_T}}{\sqrt{N}} \right), \\ \|\overline{\mathbf{M}}_Z \ddot{\mathbf{P}}_F \mathbf{b}_{is}\| &= O_p \left(\frac{p_T \sqrt{T - p_T}}{\sqrt{N}} \right). \end{aligned}$$

The orders of $\|\overline{\mathbf{M}}_Z \mathbf{a}_{ij}\|$, $\|\overline{\mathbf{M}}_Z \mathbf{b}_{is}\|$, $\|\ddot{\mathbf{M}}_F \mathbf{a}_{ij}\|$ and $\|\ddot{\mathbf{M}}_F \mathbf{b}_{is}\|$ are stated in the lemma. By combining those conditions with the order results above, we obtain

$$\epsilon_{i\ell,NT} = O_p \left(\frac{p_T(T - p_T)R_A}{\sqrt{N}} \right) + O_p \left(\frac{p_T(T - p_T)R_B}{\sqrt{N}} \right) + O_p \left(\frac{p_T^2(T - p_T)}{N} \right),$$

for $\ell = L_B, U_B$. Using this result in (3.B.99) yields

$$\begin{aligned} &\left\| \frac{\overline{\mathbf{M}}_Z \mathbf{a}_{ij}}{\sqrt{T}} \right\| \left\| \frac{\overline{\mathbf{M}}_Z \mathbf{b}_{is}}{\sqrt{T}} \right\| - \left\| \frac{\ddot{\mathbf{M}}_F \mathbf{a}_{ij}}{\sqrt{T}} \right\| \left\| \frac{\ddot{\mathbf{M}}_F \mathbf{b}_{is}}{\sqrt{T}} \right\| \\ &= O_p \left(\frac{p_T(T - p_T)R_A}{T\sqrt{N}} \right) + O_p \left(\frac{p_T(T - p_T)R_B}{T\sqrt{N}} \right) + O_p \left(\frac{p_T^2(T - p_T)}{TN} \right). \end{aligned}$$

This result gives us the orders of each element in $\frac{\mathbf{A}'_i \bar{\mathbf{M}}_Z \mathbf{B}_i}{T} - \frac{\mathbf{A}'_i \ddot{\mathbf{M}}_F \mathbf{B}_i}{T}$. We further simplify the orders by noting that $T - p_T \rightarrow T$, which yields the result stated in the lemma. \blacksquare

Lemma 3.B.13. *We have the following results regarding the processes that appear in the estimators:*

$$\begin{aligned}
 (\sqrt{T - p_T})^{-1} \|\ddot{\mathbf{M}}_F u_{ij_1}\| &= O_p(1), \\
 (\sqrt{T - p_T})^{-1} \|\ddot{\mathbf{M}}_F \mathbf{g}_{zij_2}\| &= O_p(1), \\
 (\sqrt{T - p_T})^{-1} \|\ddot{\mathbf{M}}_F \varepsilon_{ij_3}\| &= O_p(1), \\
 (\sqrt{T - p_T})^{-1} \|\ddot{\mathbf{M}}_F \mathbf{x}_{ij_4}\| &= O_p(\sqrt{T - p_T}), \\
 (\sqrt{T - p_T})^{-1} \|\ddot{\mathbf{M}}_F \mathbf{f}_{j_5}\| &= O_p(\sqrt{T - p_T}), \\
 (\sqrt{T - p_T})^{-1} \|\ddot{\mathbf{M}}_F \mathbf{v}_{Aj_6}\| &= O_p(\sqrt{T - p_T} N^{-1/2}), \\
 (\sqrt{T - p_T})^{-1} \|\ddot{\mathbf{M}}_F \mathbf{v}_{Bj_7}\| &= O_p(N^{-1/2}), \\
 (\sqrt{T - p_T})^{-1} \|\ddot{\mathbf{M}}_F \mathbf{v}_{Cj_8}\| &= O_p(\rho_H^{p_T}), \\
 (\sqrt{T - p_T})^{-1} \|\ddot{\mathbf{M}}_F \mathbf{v}_{Dj_8}\| &= O_p(\rho_A^{p_T}),
 \end{aligned}$$

uniformly in $i = 1, \dots, N$, where $j_1, j_2, j_3, j_4, j_5, j_6, j_7, j_8$ are used to denote the individual columns of the corresponding matrices.

Proof of Lemma 3.B.13: The results are established by using the inequality: $\|\ddot{\mathbf{M}}_F \mathbf{a}_{ij}\| \leq \|\mathbf{a}_{ij}\|$. \blacksquare

The results in the following lemmas are obtained by using the result in the lemmas above and will be used in the proofs of the theorems.

Lemma 3.B.14. $\bar{\mathbf{M}}_Z$, $\ddot{\mathbf{M}}_F$ and $\tilde{\mathbf{U}}_{i,-1}$ are defined in (3.B.80), (3.B.82) and (3.B.59) respectively. As $(N, T, p_T) \rightarrow \infty$ with $\frac{p_T^2}{T} \rightarrow 0$ and under the assumptions stated in the chapter, the following results hold uniformly in i :

$$\begin{aligned}
 \frac{\tilde{\mathbf{U}}'_{i,-1} \bar{\mathbf{M}}_Z \tilde{\mathbf{U}}_{i,-1}}{T} - \frac{\mathbf{U}'_{i,-1} \ddot{\mathbf{M}}_F \mathbf{U}_{i,-1}}{T} &= O_p\left(\frac{T}{N}\right) + O_p\left(\frac{p_T^2}{N}\right) + O_p\left(\frac{p_T}{\sqrt{N}}\right) \\
 &\quad + O_p\left(\frac{p_T \sqrt{T}}{N}\right), \tag{3.B.100}
 \end{aligned}$$

$$\begin{aligned}
 \frac{\tilde{\mathbf{U}}'_{i,-1} \bar{\mathbf{M}}_Z \mathbf{G}_{zi}}{T} - \frac{\mathbf{U}'_{i,-1} \ddot{\mathbf{M}}_F \mathbf{G}_{zi}}{T} &= O_p\left(\frac{p_T^2}{N}\right) + O_p\left(\frac{p_T}{\sqrt{N}}\right) + O_p\left(\frac{p_T \sqrt{T}}{N}\right), \tag{3.B.101}
 \end{aligned}$$

$$\frac{\mathbf{G}'_{zi} \bar{\mathbf{M}}_Z \mathbf{G}_{zi}}{T} - \frac{\mathbf{G}'_{zi} \ddot{\mathbf{M}}_F \mathbf{G}_{zi}}{T} = O_p\left(\frac{p_T^2}{N}\right) + O_p\left(\frac{p_T}{\sqrt{N}}\right), \quad (3.B.102)$$

$$\frac{\mathbf{X}'_{i,-1} \bar{\mathbf{M}}_Z \mathbf{X}_{i,-1}}{T^2} - \frac{\mathbf{X}'_{i,-1} \ddot{\mathbf{M}}_F \mathbf{X}_{i,-1}}{T^2} = o_p(1), \quad (3.B.103)$$

$$\frac{\bar{\mathbf{Z}}'_{-1} \bar{\mathbf{M}}_Z \mathbf{X}_{i,-1}}{T^2} - \frac{\mathbf{H}'_1 \mathbf{F}'_{-1} \ddot{\mathbf{M}}_F \mathbf{X}_{i,-1}}{T^2} = o_p(1), \quad (3.B.104)$$

$$\frac{\bar{\mathbf{Z}}'_{-1} \bar{\mathbf{M}}_Z \bar{\mathbf{Z}}_{-1}}{T^2} - \frac{\mathbf{H}'_1 \mathbf{F}'_{-1} \ddot{\mathbf{M}}_F \mathbf{F}_{-1} \mathbf{H}_1}{T^2} = o_p(1). \quad (3.B.105)$$

Lemma 3.B.15. $\bar{\mathbf{M}}_Z$, $\ddot{\mathbf{M}}_F$ and $\tilde{\mathbf{U}}_{i,-1}$ are defined in (3.B.80), (3.B.82) and (3.B.59) respectively. As $(N, T, p_T) \rightarrow \infty$ with $\frac{p_T^2}{T} \rightarrow 0$ and under the assumptions stated in the chapter, the following results holds uniformly in i :

$$\frac{\tilde{\mathbf{U}}'_{i,-1} \bar{\mathbf{M}}_Z \mathbf{X}_{i,-1}}{T^{3/2}} = O_p\left(\frac{\sqrt{T}}{\sqrt{N}}\right) + O_p\left(\frac{p_T}{\sqrt{N}}\right) + O_p\left(\frac{p_T^2}{N}\right), \quad (3.B.106)$$

$$\frac{\tilde{\mathbf{U}}'_{i,-1} \bar{\mathbf{M}}_Z \bar{\mathbf{Z}}_{-1}}{T^{3/2}} = O_p\left(\frac{\sqrt{T}}{\sqrt{N}}\right) + O_p\left(\frac{p_T}{\sqrt{N}}\right), \quad (3.B.107)$$

$$\frac{\mathbf{G}'_{zi} \bar{\mathbf{M}}_Z \mathbf{X}_{i,-1}}{T^{3/2}} = O_p\left(\frac{p_T}{\sqrt{N}}\right), \quad (3.B.108)$$

$$\frac{\mathbf{G}'_{zi} \bar{\mathbf{M}}_Z \bar{\mathbf{Z}}_{-1}}{T^{3/2}} = O_p\left(\frac{p_T}{\sqrt{N}}\right). \quad (3.B.109)$$

Lemma 3.B.16. $\bar{\mathbf{M}}_Z$, $\ddot{\mathbf{M}}_F$ and $\tilde{\mathbf{U}}_{i,-1}$ are defined in (3.B.80), (3.B.82) and (3.B.59) respectively. As $(N, T, p_T) \rightarrow \infty$ with $\frac{p_T^2}{T} \rightarrow 0$ and under the assumptions stated in the chapter, the following results holds uniformly in i :

$$\begin{aligned} \frac{\tilde{\mathbf{U}}'_{i,-1} \bar{\mathbf{M}}_Z \tilde{\boldsymbol{\varepsilon}}_{y.xi}}{\sqrt{T}} - \frac{\mathbf{U}'_{i,-1} \ddot{\mathbf{M}}_F \boldsymbol{\varepsilon}_{y.xi}}{\sqrt{T}} &= O_p(\sqrt{T} \rho_A^{p_T}) + O_p\left(\frac{p_T \sqrt{T}}{\sqrt{N}}\right) + O_p\left(\frac{p_T T}{N}\right) \\ &\quad + O_p\left(\frac{p_T^2 \sqrt{T}}{N}\right), \end{aligned} \quad (3.B.110)$$

$$\begin{aligned} \frac{\mathbf{G}'_{zi} \bar{\mathbf{M}}_Z \tilde{\boldsymbol{\varepsilon}}_{y.xi}}{\sqrt{T}} - \frac{\mathbf{G}'_{zi} \ddot{\mathbf{M}}_F \boldsymbol{\varepsilon}_{y.xi}}{\sqrt{T}} &= O_p(\sqrt{T} \rho_A^{p_T}) + O_p\left(\frac{p_T \sqrt{T}}{\sqrt{N}}\right) + O_p\left(\frac{p_T T}{N}\right) \\ &\quad + O_p\left(\frac{p_T^2 \sqrt{T}}{N}\right), \end{aligned} \quad (3.B.111)$$

$$\frac{\mathbf{X}'_{i,-1}\bar{\mathbf{M}}_Z\tilde{\varepsilon}_{y.xi}}{T} - \frac{\mathbf{X}'_{i,-1}\ddot{\mathbf{M}}_F\varepsilon_{y.xi}}{T} = O_p\left(\frac{T}{\sqrt{N}}\right) + O_p\left(\frac{p_T\sqrt{T}}{\sqrt{N}}\right) + O_p\left(\frac{p_T^2}{N}\right), \quad (3.B.112)$$

$$\frac{\bar{\mathbf{Z}}'_{-1}\bar{\mathbf{M}}_Z\tilde{\varepsilon}_{y.xi}}{T} - \frac{\mathbf{H}'_1\mathbf{F}'_{-1}\ddot{\mathbf{M}}_F\varepsilon_{y.xi}}{T} = O_p\left(\frac{T}{\sqrt{N}}\right) + O_p\left(\frac{p_T\sqrt{T}}{\sqrt{N}}\right) + O_p\left(\frac{p_T^2}{N}\right). \quad (3.B.113)$$

Proof of Lemma 3.B.14: Throughout the proofs of this lemma we use the equality

$$\bar{\mathbf{M}}_Z = \ddot{\mathbf{M}}_F + (\bar{\mathbf{M}}_Z - \ddot{\mathbf{M}}_F). \quad (3.B.114)$$

To prove (3.B.100) we first make use of (3.B.60) and (3.B.114) to write

$$\begin{aligned} T^{-1}\tilde{\mathbf{U}}'_{i,-1}\bar{\mathbf{M}}_Z\tilde{\mathbf{U}}_{i,-1} &= T^{-1}\mathbf{U}'_{i,-1}\ddot{\mathbf{M}}_F\mathbf{U}_{i,-1} + T^{-1}\mathbf{U}'_{i,-1}(\bar{\mathbf{M}}_Z - \ddot{\mathbf{M}}_F)\mathbf{U}_{i,-1} \\ &\quad - T^{-1}\mathbf{U}'_{i,-1}\bar{\mathbf{M}}_Z\mathbf{V}_{A,-1}\mathbf{H}_1^-\boldsymbol{\lambda}_i \\ &\quad - T^{-1}\mathbf{U}'_{i,-1}\bar{\mathbf{M}}_Z\mathbf{V}_{C,-p_T}\mathbf{H}_1^-\boldsymbol{\lambda}_i \\ &\quad - T^{-1}\boldsymbol{\lambda}'_i\mathbf{H}_1^-\mathbf{V}'_{A,-1}\bar{\mathbf{M}}_Z\mathbf{U}_{i,-1} \\ &\quad + T^{-1}\boldsymbol{\lambda}'_i\mathbf{H}_1^-\mathbf{V}'_{A,-1}\bar{\mathbf{M}}_Z\mathbf{V}_{A,-1}\mathbf{H}_1^-\boldsymbol{\lambda}_i \\ &\quad + T^{-1}\boldsymbol{\lambda}'_i\mathbf{H}_1^-\mathbf{V}'_{A,-1}\bar{\mathbf{M}}_Z\mathbf{V}_{C,-p_T}\mathbf{H}_1^-\boldsymbol{\lambda}_i \\ &\quad - T^{-1}\boldsymbol{\lambda}'_i\mathbf{H}_1^-\mathbf{V}'_{C,-p_T}\bar{\mathbf{M}}_Z\mathbf{U}_{i,-1} \\ &\quad + T^{-1}\boldsymbol{\lambda}'_i\mathbf{H}_1^-\mathbf{V}'_{C,-p_T}\bar{\mathbf{M}}_Z\mathbf{V}_{A,-1}\mathbf{H}_1^-\boldsymbol{\lambda}_i \\ &\quad + T^{-1}\boldsymbol{\lambda}'_i\mathbf{H}_1^-\mathbf{V}'_{C,-p_T}\bar{\mathbf{M}}_Z\mathbf{V}_{C,-p_T}\mathbf{H}_1^-\boldsymbol{\lambda}_i \\ &= \mathbf{K}_{11} + \mathbf{K}_{12} - \mathbf{K}_{13} - \mathbf{K}_{14} - \mathbf{K}_{15} + \mathbf{K}_{16} + \mathbf{K}_{17} - \mathbf{K}_{18} \\ &\quad + \mathbf{K}_{19} + \mathbf{K}_{110} \end{aligned}$$

The order of \mathbf{K}_{11} can be obtained as $O_p(1)$ using the preliminary results (3.B.20), (3.B.85) and (3.B.88) and the boundedness of \mathbf{H}_1^- and $\boldsymbol{\lambda}_i$. The order of \mathbf{K}_{12} can be found using Lemma 3.B.12, noting that $(\sqrt{T} - p_T)^{-1}\|\ddot{\mathbf{M}}_F\mathbf{u}_{i,-1,j_1}\| = O_p(1)$, which is a result stated in Lemma 3.B.13 and where the index j_1 signifies the j_1^{st} column of $\mathbf{U}_{i,-1}$. We have

$$\|T^{-1}\mathbf{U}'_i(\bar{\mathbf{M}}_Z - \ddot{\mathbf{M}}_F)\mathbf{U}_i\| = O_p(p_T N^{-1/2}) + O_p(p_T^2 N^{-1}).$$

\mathbf{K}_{13} can be written as a sum of two terms using (3.B.114) as

$$\begin{aligned} T^{-1}\mathbf{U}'_{i,-1}\bar{\mathbf{M}}_Z\mathbf{V}_{A,-1}\mathbf{H}_1^-\boldsymbol{\lambda}_i &= T^{-1}\mathbf{U}'_{i,-1}\ddot{\mathbf{M}}_F\mathbf{V}_{A,-1}\mathbf{H}_1^-\boldsymbol{\lambda}_i \\ &\quad + T^{-1}\mathbf{U}'_{i,-1}(\bar{\mathbf{M}}_Z - \ddot{\mathbf{M}}_F)\mathbf{V}_{A,-1}\mathbf{H}_1^-\boldsymbol{\lambda}_i. \end{aligned} \quad (3.B.115)$$

We first find the order of second term on the right-hand side, which can be established using Lemma 3.B.11 and the results $(\sqrt{T-p_T})^{-1}\|\ddot{\mathbf{M}}_F\mathbf{v}_{A,-1,j_6}\| = O_p(T^{1/2}N^{-1/2})$ and $(\sqrt{T-p_T})^{-1}\|\ddot{\mathbf{M}}_F u_{i,-1,j_1}\| = O_p(1)$ that are stated in 3.B.13 and where j_6 and j_1 signifies the j_6^{th} and the j_1^{st} columns of $\mathbf{V}_{A,-1}$ and $\mathbf{U}_{i,-1}$ respectively. Note that as mentioned before from an asymptotic point of view $T-p_T$ and T are replaceable. Then the order result can be formally written as

$$\begin{aligned} \|T^{-1}\mathbf{U}'_{i,-1}(\bar{\mathbf{M}}_Z - \ddot{\mathbf{M}}_F)\mathbf{V}_{A,-1}\mathbf{H}_1^-\boldsymbol{\lambda}_i\| &= O_p\left(\frac{p_T\sqrt{T}}{N}\right) + O_p\left(\frac{p_T}{\sqrt{N}}\right) \\ &\quad + O_p\left(\frac{p_T^2}{N}\right). \end{aligned}$$

And for the first term of (3.B.115), under the condition that $\frac{p_T^2}{T} \rightarrow 0$, we have

$$\begin{aligned} \|T^{-1}\mathbf{U}'_{i,-1}\ddot{\mathbf{M}}_F\mathbf{V}_{A,-1}\mathbf{H}_1^-\boldsymbol{\lambda}_i\| &\leq \|T^{-1}\mathbf{U}'_{i,-1}\mathbf{V}_{A,-1}\|\|\mathbf{H}_1^-\|\|\boldsymbol{\lambda}_i\| \\ &= O_p(N^{-1/2}). \end{aligned}$$

The order is obtained using Cauchy-Schwarz inequality and (3.B.50). As for \mathbf{K}_{14} we again use (3.B.114) to write the expression into two terms as

$$\begin{aligned} T^{-1}\mathbf{U}'_{i,-1}\bar{\mathbf{M}}_Z\mathbf{V}_{C,-p_T}\mathbf{H}_1^-\boldsymbol{\lambda}_i &= T^{-1}\mathbf{U}'_{i,-1}\ddot{\mathbf{M}}_F\mathbf{V}_{C,-p_T}\mathbf{H}_1^-\boldsymbol{\lambda}_i \\ &\quad + T^{-1}\mathbf{U}'_{i,-1}(\bar{\mathbf{M}}_Z - \ddot{\mathbf{M}}_F)\mathbf{V}_{C,-p_T}\mathbf{H}_1^-\boldsymbol{\lambda}_i. \end{aligned}$$

For the second term, using Lemmas 3.B.12 and 3.B.13 we have

$$\begin{aligned} \|T^{-1}\mathbf{U}'_{i,-1}(\bar{\mathbf{M}}_Z - \ddot{\mathbf{M}}_F)\mathbf{V}_{C,-p_T}\mathbf{H}_1^-\boldsymbol{\lambda}_i\| \\ \leq \|T^{-1}\mathbf{U}'_{i,-1}(\bar{\mathbf{M}}_Z - \ddot{\mathbf{M}}_F)\mathbf{V}_{C,-p_T}\|\|\mathbf{H}_1^-\|\|\boldsymbol{\lambda}_i\| &= O_p(p_T^2N^{-1}). \end{aligned}$$

For the first term we have

$$\begin{aligned} \|T^{-1}\mathbf{U}'_{i,-1}\ddot{\mathbf{M}}_F\mathbf{V}_{C,-p_T}\mathbf{H}_1^-\boldsymbol{\lambda}_i\| &\leq \|T^{-1}\mathbf{U}'_{i,-1}\mathbf{V}_{C,-p_T}\|\|\mathbf{H}_1^-\|\|\boldsymbol{\lambda}_i\| \\ &= O_p(\rho_H^{p_T}). \end{aligned}$$

The order is obtained using (3.B.66). \mathbf{K}_{15} is the transpose of \mathbf{K}_{13} , hence they have the same order. To find the order of \mathbf{K}_{16} we follow the same analogy as before, we first rewrite the expression by using (3.B.114) and obtain two terms, the order of the second term can be established as

$$\begin{aligned} & \|T^{-1}\boldsymbol{\lambda}'_i\mathbf{H}_1^{-'}\mathbf{V}'_{A,-1}(\overline{\mathbf{M}}_Z - \ddot{\mathbf{M}}_F)\mathbf{V}_{A,-1}\mathbf{H}_1^{-}\boldsymbol{\lambda}_i\| \\ & \leq \|T^{-1}\boldsymbol{\lambda}'_i\mathbf{H}_1^{-'}\mathbf{V}'_{A,-1}(\overline{\mathbf{M}}_Z - \ddot{\mathbf{M}}_F)\mathbf{V}_{A,-1}\|\|\mathbf{H}_1^{-}\|\|\boldsymbol{\lambda}_i\| \\ & = O_p\left(\frac{p_T\sqrt{T}}{N}\right) + O_p\left(\frac{p_T^2}{N}\right), \end{aligned}$$

by using the Cauchy-Schwarz inequality, Lemmas 3.B.12 and 3.B.13 and for the first term we have

$$\begin{aligned} \|T^{-1}\boldsymbol{\lambda}'_i\mathbf{H}_1^{-'}\mathbf{V}'_{A,-1}\ddot{\mathbf{M}}_F\mathbf{V}_{A,-1}\mathbf{H}_1^{-}\boldsymbol{\lambda}_i\| & \leq \|\boldsymbol{\lambda}'_i\|\|\mathbf{H}_1^{-'}\|\|T^{-1}\mathbf{V}'_{A,-1}\mathbf{V}_{A,-1}\|\|\mathbf{H}_1^{-}\|\|\boldsymbol{\lambda}_i\| \\ & = O_p(TN^{-1}), \end{aligned}$$

by using the preliminary result in (3.B.49). To find the order of \mathbf{K}_{17} we establish the following results:

$$\begin{aligned} & \|T^{-1}\boldsymbol{\lambda}'_i\mathbf{H}_1^{-'}\mathbf{V}'_{A,-1}(\overline{\mathbf{M}}_Z - \ddot{\mathbf{M}}_F)\mathbf{V}_{C,-p_T}\mathbf{H}_1^{-}\boldsymbol{\lambda}_i\| \\ & \leq \|\boldsymbol{\lambda}'_i\|\|\mathbf{H}_1^{-'}\|\|T^{-1}\mathbf{V}'_{A,-1}(\overline{\mathbf{M}}_Z - \ddot{\mathbf{M}}_F)\mathbf{V}_{C,-p_T}\|\|\mathbf{H}_1^{-}\|\|\boldsymbol{\lambda}_i\| \\ & = O_p\left(\frac{p_T\sqrt{T}}{N}\right) + O_p\left(\frac{p_T^2}{N}\right), \end{aligned}$$

by using Lemmas 3.B.12 and 3.B.13, and

$$\begin{aligned} & \|T^{-1}\boldsymbol{\lambda}'_i\mathbf{H}_1^{-'}\mathbf{V}'_{A,-1}\overline{\mathbf{M}}_Z\mathbf{V}_{C,-p_T}\mathbf{H}_1^{-}\boldsymbol{\lambda}_i\| \\ & \leq \|\boldsymbol{\lambda}'_i\|\|\mathbf{H}_1^{-'}\|\|T^{-1}\mathbf{V}'_{A,-1}\mathbf{V}_{C,-p_T}\|\|\mathbf{H}_1^{-}\|\|\boldsymbol{\lambda}_i\| \\ & = O_p(\rho_H^{p_T}N^{-1/2}), \end{aligned}$$

by using (3.B.68). \mathbf{K}_{18} and \mathbf{K}_{19} are the transposes of \mathbf{K}_{14} and \mathbf{K}_{17} hence have the same orders of convergence. Finally for the order of \mathbf{K}_{110} we establish the following results. Firstly,

$$\begin{aligned} & \|T^{-1}\boldsymbol{\lambda}'_i\mathbf{H}_1^{-'}\mathbf{V}'_{C,-p_T}(\overline{\mathbf{M}}_Z - \ddot{\mathbf{M}}_F)\mathbf{V}_{C,-p_T}\mathbf{H}_1^{-}\boldsymbol{\lambda}_i\| \\ & \leq \|\boldsymbol{\lambda}'_i\|\|\mathbf{H}_1^{-'}\|\|T^{-1}\mathbf{V}'_{C,-p_T}(\overline{\mathbf{M}}_Z - \ddot{\mathbf{M}}_F)\mathbf{V}_{C,-p_T}\|\|\mathbf{H}_1^{-}\|\|\boldsymbol{\lambda}_i\| \\ & = O_p\left(\frac{p_T\rho_A^{p_T}}{\sqrt{N}}\right) + O_p\left(\frac{p_T^2}{N}\right). \end{aligned}$$

Secondly,

$$\begin{aligned}
& \|T^{-1}\boldsymbol{\lambda}'_i\mathbf{H}_1^{-'}\mathbf{V}'_{C,-p_T}\bar{\mathbf{M}}_Z\mathbf{V}_{C,-p_T}\mathbf{H}_1^-\boldsymbol{\lambda}_i\| \\
& \leq \|\boldsymbol{\lambda}'_i\|\|\mathbf{H}_1^{-'}\|\|T^{-1}\mathbf{V}'_{C,-p_T}\mathbf{V}_{C,-p_T}\|\|\mathbf{H}_1^-\|\|\boldsymbol{\lambda}_i\| \\
& = O_p(\rho_H^{2p_T}),
\end{aligned}$$

where (3.B.70) is used to obtain the order. Combining the orders of \mathbf{K}_1 - \mathbf{K}_{110} and considering the conditions of the lemma yield the required result.

To prove (3.B.101) we write the left-hand side as

$$\begin{aligned}
T^{-1}\tilde{\mathbf{U}}'_{i,-1}\bar{\mathbf{M}}_Z\mathbf{G}_{zi} &= T^{-1}\mathbf{U}'_{i,-1}\ddot{\mathbf{M}}_F\mathbf{G}_{zi} + T^{-1}\mathbf{U}'_{i,-1}(\bar{\mathbf{M}}_Z - \ddot{\mathbf{M}}_F)\mathbf{G}_{zi} \\
&\quad - T^{-1}\boldsymbol{\lambda}'_i\mathbf{H}_1^{-'}\mathbf{V}'_{A,-1}\bar{\mathbf{M}}_Z\mathbf{G}_{zi} - T^{-1}\boldsymbol{\lambda}'_i\mathbf{H}_1^{-'}\mathbf{V}_{C,-p_T}\bar{\mathbf{M}}_Z\mathbf{G}_{zi} \\
&= \mathbf{K}_{21} + \mathbf{K}_{22} - \mathbf{K}_{23} - \mathbf{K}_{24}.
\end{aligned} \tag{3.B.116}$$

The order of \mathbf{K}_{22} can be established by using Lemmas 3.B.12 and 3.B.13 as

$$\|T^{-1}\mathbf{U}'_{i,-1}(\bar{\mathbf{M}}_Z - \ddot{\mathbf{M}}_F)\mathbf{G}_{zi}\| = O_p(p_T N^{-1/2}) + O_p(p_T^2 N^{-1}).$$

The orders of the rest of the terms are can be found by first using (3.B.114) and by applying the triangle inequality and the Cauchy-Schwarz inequality. We have

$$\begin{aligned}
\|\mathbf{K}_{23}\| &\leq \|\boldsymbol{\lambda}'_i\|\|\mathbf{H}_1^{-'}\|\|T^{-1}\mathbf{V}'_{A,-1}\ddot{\mathbf{M}}_F\mathbf{G}_{zi}\| \\
&\quad + \|\boldsymbol{\lambda}'_i\|\|\mathbf{H}_1^{-'}\|\|T^{-1}\mathbf{V}'_{A,-1}(\bar{\mathbf{M}}_Z - \ddot{\mathbf{M}}_F)\mathbf{G}_{zi}\| \\
&= O_p\left(\frac{p_T\sqrt{T}}{N}\right) + O_p\left(\frac{p_T^2}{N}\right),
\end{aligned}$$

by using (3.B.88), (3.B.90) and Lemmas 3.B.12 and 3.B.13. For \mathbf{K}_{24} we have

$$\begin{aligned}
\|\mathbf{K}_{24}\| &\leq \|\boldsymbol{\lambda}'_i\|\|\mathbf{H}_1^{-'}T^{-1}\|\|\mathbf{V}'_{C,-p_T}\ddot{\mathbf{M}}_F\mathbf{G}_{zi}\| \\
&\quad + \|\boldsymbol{\lambda}'_i\|\|\mathbf{H}_1^{-'}\|\|T^{-1}\mathbf{V}'_{C,-p_T}(\bar{\mathbf{M}}_Z - \ddot{\mathbf{M}}_F)\mathbf{G}_{zi}\| \\
&= O_p\left(\frac{p_T^2}{N}\right) + O_p\left(\frac{p_T}{\sqrt{N}}\right) + O_p(\rho_H^{p_T}),
\end{aligned}$$

where the last equality is obtained using (3.B.88), (3.B.93) and Lemmas 3.B.12 and 3.B.13. Then by combining the orders of \mathbf{K}_{22} , \mathbf{K}_{23} and \mathbf{K}_{24} gives the result in the lemma as $(N, T, p_T) \rightarrow \infty$.

To prove (3.B.102) we write the left-hand side as

$$\begin{aligned} T^{-1}\mathbf{G}'_{zi}\bar{\mathbf{M}}_Z\mathbf{G}_{zi} &= T^{-1}\mathbf{G}'_{zi}\ddot{\mathbf{M}}_F\mathbf{G}_{zi} + T^{-1}\mathbf{G}'_{zi}(\bar{\mathbf{M}}_Z - \ddot{\mathbf{M}}_F)\mathbf{G}_{zi} \\ &= \mathbf{K}_{31} + \mathbf{K}_{32}. \end{aligned} \quad (3.B.117)$$

The order of \mathbf{K}_{31} is $O_p(1)$ from the preliminary result (3.B.90). The order of \mathbf{K}_{32} can be found as

$$\|T^{-1}\mathbf{G}'_{zi}(\bar{\mathbf{M}}_Z - \ddot{\mathbf{M}}_F)\mathbf{G}_{zi}\| = O_p\left(\frac{p_T}{\sqrt{N}}\right) + O_p\left(\frac{p_T^2}{N}\right),$$

by using Lemmas 3.B.12 and 3.B.13. Combining the result for \mathbf{K}_{31} and \mathbf{K}_{32} gives the required result.

To prove (3.B.103) we write the left-hand side of the equation as

$$\begin{aligned} T^{-2}\mathbf{X}'_{i,-1}\bar{\mathbf{M}}_Z\mathbf{X}_{i,-1} &= T^{-2}\mathbf{X}'_{i,-1}\ddot{\mathbf{M}}_F\mathbf{X}_{i,-1} + T^{-2}\mathbf{X}'_{i,-1}(\bar{\mathbf{M}}_Z - \ddot{\mathbf{M}}_F)\mathbf{X}_{i,-1} \\ &= \mathbf{K}_{41} + \mathbf{K}_{42}. \end{aligned}$$

From the preliminary results (3.B.16), (3.B.86) and (3.B.88), it can be shown that \mathbf{K}_{41} is $O_p(1)$. As for the order of \mathbf{K}_{42} we write

$$\|T^{-2}\mathbf{X}'_{i,-1}(\bar{\mathbf{M}}_Z - \ddot{\mathbf{M}}_F)\mathbf{X}_{i,-1}\| = O_p\left(\frac{p_T}{\sqrt{NT}}\right) + O_p\left(\frac{p_T^2}{NT}\right),$$

where the order is obtained by using Lemmas 3.B.12 and 3.B.13.

To prove (3.B.104) we write the left-hand side of the equation as

$$\begin{aligned} T^{-2}\bar{\mathbf{Z}}'_{-1}\bar{\mathbf{M}}_Z\mathbf{X}_{i,-1} &= T^{-2}\mathbf{H}'_1\mathbf{F}'_{-1}\ddot{\mathbf{M}}_F\mathbf{X}_{i,-1} + T^{-2}\mathbf{H}'_1\mathbf{F}'_{-1}(\bar{\mathbf{M}}_Z - \ddot{\mathbf{M}}_F)\mathbf{X}_{i,-1} \\ &\quad + T^{-2}\mathbf{V}'_{A,-1}\bar{\mathbf{M}}_Z\mathbf{X}_{i,-1} + T^{-2}\mathbf{V}'_{C,-p_T}\bar{\mathbf{M}}_Z\mathbf{X}_{i,-1} \\ &= \mathbf{K}_{51} + \mathbf{K}_{52} + \mathbf{K}_{53} + \mathbf{K}_{54}. \end{aligned} \quad (3.B.118)$$

The order of \mathbf{K}_{51} is $O_p(1)$, which can be established by using (3.B.17), (3.B.86), (3.B.87) and (3.B.88). By using Lemmas 3.B.12 and 3.B.13, the order of \mathbf{K}_{52} can be obtained as

$$\|T^{-2}\mathbf{H}'_1\mathbf{F}'_{-1}(\bar{\mathbf{M}}_Z - \ddot{\mathbf{M}}_F)\mathbf{X}_{i,-1}\| = O_p\left(\frac{p_T}{\sqrt{NT}}\right) + O_p\left(\frac{p_T^2}{NT}\right).$$

The orders of the rest of the terms are

$$\|\mathbf{K}_{53}\| \leq \|T^{-2}\mathbf{V}'_{A,-1}\ddot{\mathbf{M}}_F\mathbf{X}_{i,-1}\| + \|T^{-2}\mathbf{V}'_{A,-1}(\bar{\mathbf{M}}_Z - \ddot{\mathbf{M}}_F)\mathbf{X}_{i,-1}\|$$

$$= O_p\left(\frac{1}{\sqrt{N}}\right) + O_p\left(\frac{p_T}{\sqrt{NT}}\right) + O_p\left(\frac{p_T^2}{NT}\right),$$

by (3.B.51), (3.B.86), (3.B.88) and (3.B.90) and by Lemmas 3.B.12 and 3.B.13,

$$\begin{aligned} \|\mathbf{K}_{54}\| &\leq \|T^{-2}\mathbf{V}'_{C,-p_T}\ddot{\mathbf{M}}_F\mathbf{X}_{i,-1}\| + \|T^{-2}\mathbf{V}'_{C,-p_T}(\overline{\mathbf{M}}_Z - \ddot{\mathbf{M}}_F)\mathbf{X}_{i,-1}\| \\ &= O_p\left(\frac{\rho_H^{p_T}}{T}\right) + O_p\left(\frac{p_T\rho_H^{p_T}}{T\sqrt{N}}\right) + O_p\left(\frac{p_T^2}{NT}\right) + O_p\left(\frac{p_T}{\sqrt{NT}}\right), \end{aligned}$$

which can be obtained using (3.B.64), (3.B.86), (3.B.88), (3.B.90) and using Lemmas 3.B.12 and 3.B.13.

To prove (3.B.105) we write

$$\begin{aligned} T^{-2}\overline{\mathbf{Z}}'_- \overline{\mathbf{M}}_Z \overline{\mathbf{Z}}_- &= T^{-2}\mathbf{H}'_1\mathbf{F}'_{-1}\ddot{\mathbf{M}}_F\mathbf{F}_{-1}\mathbf{H}_1 + T^{-2}\mathbf{H}'_1\mathbf{F}'_{-1}(\overline{\mathbf{M}}_Z - \ddot{\mathbf{M}}_F)\mathbf{F}_{-1}\mathbf{H}_1 \\ &\quad + T^{-2}\mathbf{H}'_1\mathbf{F}'_{-1}\overline{\mathbf{M}}_Z\mathbf{V}_{A,-1} + T^{-2}\mathbf{H}'_1\mathbf{F}'_{-1}\overline{\mathbf{M}}_Z\mathbf{V}_{C,-p_T} \\ &\quad + T^{-2}\mathbf{V}'_{A,-1}\overline{\mathbf{M}}_Z\mathbf{F}_{-1}\mathbf{H}_1 + T^{-2}\mathbf{V}'_{A,-1}\overline{\mathbf{M}}_Z\mathbf{V}_{A,-1} \\ &\quad + T^{-2}\mathbf{V}'_{A,-1}\overline{\mathbf{M}}_Z\mathbf{V}_{C,-p_T} + T^{-2}\mathbf{V}'_{C,-p_T}\overline{\mathbf{M}}_Z\mathbf{F}_{-1}\mathbf{H}_1 \\ &\quad + T^{-2}\mathbf{V}'_{C,-p_T}\overline{\mathbf{M}}_Z\mathbf{V}_{A,-1} + T^{-2}\mathbf{V}'_{C,-p_T}\overline{\mathbf{M}}_Z\mathbf{V}_{C,-p_T} \\ &= \mathbf{K}_{61} + \mathbf{K}_{62} + \mathbf{K}_{63} + \mathbf{K}_{64} + \mathbf{K}_{65} + \mathbf{K}_{66} \\ &\quad + \mathbf{K}_{67} + \mathbf{K}_{68} + \mathbf{K}_{69} + \mathbf{K}_{610}. \end{aligned}$$

We analyze each term on the right-hand side separately. For the terms \mathbf{K}_{63} - \mathbf{K}_{610} , first we use (3.B.114) to write it in two terms, then we apply the triangle inequality and the Cauchy-Schwarz inequality (if necessary). Finally by using the relevant preliminary results and the boundedness of \mathbf{H}_1 we find the orders of the terms. \mathbf{K}_{61} is $O_p(1)$. For the order of \mathbf{K}_{62} we have

$$\|T^{-2}\mathbf{H}'_1\mathbf{F}'_{-1}(\overline{\mathbf{M}}_Z - \ddot{\mathbf{M}}_F)\mathbf{F}_{-1}\mathbf{H}'_1\| = O_p\left(\frac{p_T}{\sqrt{NT}}\right) + O_p\left(\frac{p_T^2}{NT}\right),$$

by using Lemmas (3.B.12) and (3.B.13). The order of \mathbf{K}_{63} can be found as

$$\begin{aligned} \|T^{-2}\mathbf{H}'_1\mathbf{F}'_{-1}\overline{\mathbf{M}}_Z\mathbf{V}_{A,-1}\| &\leq \|\mathbf{H}'_1\| \|T^{-2}\mathbf{F}'_{-1}\ddot{\mathbf{M}}_F\mathbf{V}_{A,-1}\| \\ &\quad + \|\mathbf{H}'_1\| \|T^{-2}\mathbf{F}'_{-1}(\overline{\mathbf{M}}_Z - \ddot{\mathbf{M}}_F)\mathbf{V}_{A,-1}\| \\ &= O_p\left(\frac{1}{\sqrt{N}}\right) + O_p\left(\frac{p_T}{\sqrt{NT}}\right) + O_p\left(\frac{p_T^2}{NT}\right), \end{aligned}$$

by using (3.B.52), (3.B.87), (3.B.88), (3.B.90) and Lemmas 3.B.12 and 3.B.13. The order of \mathbf{K}_{64} can be obtained as

$$\begin{aligned} \|T^{-2}\mathbf{H}'_1\mathbf{F}'_{-1}\bar{\mathbf{M}}_Z\mathbf{V}_{C,-p_T}\| &\leq \|\mathbf{H}'_1\| \|T^{-2}\mathbf{F}'_{-1}\ddot{\mathbf{M}}_F\mathbf{V}_{C,-p_T}\| \\ &\quad + \|\mathbf{H}'_1\| \|T^{-2}\mathbf{F}'_{-1}(\bar{\mathbf{M}}_Z - \ddot{\mathbf{M}}_F)\mathbf{V}_{C,-p_T}\| \\ &= O_p\left(\frac{\rho_H^{p_T}}{T}\right) + O_p\left(\frac{p_T}{\sqrt{NT}}\right) \\ &\quad + O_p\left(\frac{p_T^2}{NT}\right) + O_p\left(\frac{p_T\rho_H^{p_T}}{T\sqrt{N}}\right), \end{aligned}$$

by using (3.B.65), (3.B.87), (3.B.88), (3.B.93) and Lemmas 3.B.12 and 3.B.13. \mathbf{K}_{65} is the transpose of \mathbf{K}_{63} , so they have the same order. \mathbf{K}_{66} and \mathbf{K}_{67} have the same orders as $T^{-1}\mathbf{K}_{16}$ and $T^{-1}\mathbf{K}_{17}$ and the orders can be written formally as

$$\begin{aligned} \|\mathbf{K}_{66}\| &= O_p\left(\frac{1}{N}\right) + O_p\left(\frac{p_T}{N\sqrt{T}}\right) + O_p\left(\frac{p_T^2}{TN}\right), \\ \|\mathbf{K}_{67}\| &= O_p\left(\frac{\rho_H^{p_T}}{T\sqrt{N}}\right) + O_p\left(\frac{p_T}{N\sqrt{T}}\right) + O_p\left(\frac{p_T^2}{TN}\right). \end{aligned}$$

\mathbf{K}_{68} and \mathbf{K}_{69} are the transposes of \mathbf{K}_{64} and \mathbf{K}_{67} , so their orders are the same as their transposes. \mathbf{K}_{610} has the same order as $T^{-1}\mathbf{K}_{110}$, which is

$$\|\mathbf{K}_{610}\| = O_p\left(\frac{\rho_H^{2p_T}}{T}\right) + O_p\left(\frac{p_T\rho_A^{2p_T}}{T\sqrt{N}}\right) + O_p\left(\frac{p_T^2}{TN}\right).$$

Combining the orders of \mathbf{K}_{61} - \mathbf{K}_{610} we obtain the required result.

Proof of Lemma 3.B.15: As in the proof of Lemma 3.B.14 we first use (3.B.60) and (3.B.61) to replace $\bar{\mathbf{M}}_Z\tilde{\mathbf{U}}_{i,-1}$, $\bar{\mathbf{M}}_Z\bar{\mathbf{Z}}_{-1}$ respectively. Then we analyze the resulting terms by rewriting them by using (3.B.114). To prove (3.B.106) we write

$$\begin{aligned} T^{-3/2}\tilde{\mathbf{U}}'_{i,-1}\bar{\mathbf{M}}_Z\mathbf{X}_{i,-1} &= T^{-3/2}\mathbf{U}'_{i,-1}\bar{\mathbf{M}}_Z\mathbf{X}_{i,-1} - T^{-3/2}\boldsymbol{\lambda}'_i\mathbf{H}'_1\mathbf{V}'_{A,-1}\bar{\mathbf{M}}_Z\mathbf{X}_{i,-1} \\ &\quad - T^{-3/2}\boldsymbol{\lambda}'_i\mathbf{H}'_1\mathbf{V}'_{C,-p_T}\bar{\mathbf{M}}_Z\mathbf{X}_{i,-1} \\ &= \mathbf{K}_{71} - \mathbf{K}_{72} - \mathbf{K}_{73}. \end{aligned} \tag{3.B.119}$$

By using (3.B.114), the orders of the terms on the right-hand side of the last equation are can be obtained as follows:

$$\|\mathbf{K}_{71}\| \leq \|T^{-3/2}\mathbf{U}'_{i,-1}\ddot{\mathbf{M}}_F\mathbf{X}_{i,-1}\| + \|T^{-3/2}\mathbf{U}'_{i,-1}(\bar{\mathbf{M}}_Z - \ddot{\mathbf{M}}_F)\mathbf{X}_{i,-1}\|$$

$$= O_p\left(\frac{1}{\sqrt{T}}\right) + O_p\left(\frac{p_T}{\sqrt{N}}\right) + O_p\left(\frac{p_T^2}{N}\right),$$

by Lemma 3.B.3, (3.B.85), (3.B.86), (3.B.88) and Lemmas 3.B.12 and 3.B.13.

$$\begin{aligned} \|\mathbf{K}_{72}\| &\leq \|\boldsymbol{\lambda}'_i\| \|\mathbf{H}_1^{-1}\| T^{-3/2} \mathbf{V}'_{A,-1} \ddot{\mathbf{M}}_F \mathbf{X}_{i,-1} \| \\ &\quad + \|\boldsymbol{\lambda}'_i\| \|\mathbf{H}_1^{-1}\| T^{-3/2} \mathbf{V}'_{A,-1} (\bar{\mathbf{M}}_Z - \ddot{\mathbf{M}}_F) \mathbf{X}_{i,-1} \| \\ &= O_p\left(\frac{\sqrt{T}}{\sqrt{N}}\right) + O_p\left(\frac{p_T}{\sqrt{N}}\right) + O_p\left(\frac{p_T^2}{N}\right), \end{aligned}$$

by (3.B.51), (3.B.86), (3.B.88), (3.B.90) and Lemmas 3.B.12 and 3.B.13.

$$\begin{aligned} \|\mathbf{K}_{73}\| &\leq \|\boldsymbol{\lambda}'_i\| \|\mathbf{H}_1^{-1}\| T^{-3/2} \mathbf{V}'_{C,-p_T} \ddot{\mathbf{M}}_F \mathbf{X}_{i,-1} \| \\ &\quad + \|\boldsymbol{\lambda}'_i\| \|\mathbf{H}_1^{-1}\| T^{-3/2} \mathbf{V}'_{C,-p_T} (\bar{\mathbf{M}}_Z - \ddot{\mathbf{M}}_F) \mathbf{X}_{i,-1} \| \\ &= O_p\left(\frac{\rho_H^{p_T}}{\sqrt{T}}\right) + O_p\left(\frac{p_T}{\sqrt{N}}\right) + O_p\left(\frac{p_T^2}{N}\right), \end{aligned}$$

by (3.B.64), (3.B.86), (3.B.88), (3.B.93) and Lemmas 3.B.12 and 3.B.13. Then by combining these order results we obtain the result stated in (3.B.106).

To prove (3.B.107) we write,

$$\begin{aligned} T^{-3/2} \tilde{\mathbf{U}}'_{i,-1} \bar{\mathbf{M}}_Z \bar{\mathbf{Z}}_{-1} &= T^{-3/2} \mathbf{U}'_{i,-1} \bar{\mathbf{M}}_Z \mathbf{F}_{-1} \mathbf{H}_1 \\ &\quad - T^{-3/2} \boldsymbol{\lambda}'_i \mathbf{H}_1^{-1} \mathbf{V}'_{A,-1} \bar{\mathbf{M}}_Z \mathbf{F}_{-1} \mathbf{H}_1 \\ &\quad - T^{-3/2} \boldsymbol{\lambda}'_i \mathbf{H}_1^{-1} \mathbf{V}'_{C,-p_T} \bar{\mathbf{M}}_Z \mathbf{F}_{-1} \mathbf{H}_1 \\ &\quad + T^{-3/2} \mathbf{U}'_{i,-1} \bar{\mathbf{M}}_Z \mathbf{V}_{A,-1} \\ &\quad - T^{-3/2} \boldsymbol{\lambda}'_i \mathbf{H}_1^{-1} \mathbf{V}'_{A,-1} \bar{\mathbf{M}}_Z \mathbf{V}_{A,-1} \\ &\quad - T^{-3/2} \boldsymbol{\lambda}'_i \mathbf{H}_1^{-1} \mathbf{V}'_{C,-p_T} \bar{\mathbf{M}}_Z \mathbf{V}_{A,-1} \\ &\quad + T^{-3/2} \mathbf{U}'_{i,-1} \bar{\mathbf{M}}_Z \mathbf{V}_{C,-p_T} \\ &\quad - T^{-3/2} \boldsymbol{\lambda}'_i \mathbf{H}_1^{-1} \mathbf{V}'_{A,-1} \bar{\mathbf{M}}_Z \mathbf{V}_{C,-p_T} \\ &\quad - T^{-3/2} \boldsymbol{\lambda}'_i \mathbf{H}_1^{-1} \mathbf{V}'_{C,-p_T} \bar{\mathbf{M}}_Z \mathbf{V}_{C,-p_T} \\ &= \mathbf{K}_{81} - \mathbf{K}_{82} - \mathbf{K}_{83} + \mathbf{K}_{84} - \mathbf{K}_{85} - \mathbf{K}_{86} + \mathbf{K}_{87} \\ &\quad - \mathbf{K}_{88} - \mathbf{K}_{89}. \end{aligned}$$

We do not provide the detailed steps we use to obtain the orders of the each term on the right-hand side of the above equation. The orders can be obtained by following similar steps as in the proof of (3.B.106) and they can be stated formally

as

$$\begin{aligned}
 \|\mathbf{K}_{81}\| &= O_p\left(\frac{1}{\sqrt{T}}\right) + O_p\left(\frac{p_T}{\sqrt{N}}\right) + O_p\left(\frac{p_T^2}{N\sqrt{T}}\right), \\
 \|\mathbf{K}_{82}\| &= O_p\left(\frac{\sqrt{T}}{\sqrt{N}}\right) + O_p\left(\frac{p_T}{\sqrt{N}}\right) + O_p\left(\frac{p_T^2}{N\sqrt{T}}\right), \\
 \|\mathbf{K}_{83}\| &= O_p\left(\frac{\rho_H^{p_T}}{\sqrt{T}}\right) + O_p\left(\frac{p_T}{\sqrt{N}}\right) + O_p\left(\frac{p_T^2}{N\sqrt{T}}\right), \\
 \|\mathbf{K}_{84}\| &= O_p\left(\frac{p_T}{\sqrt{NT}}\right) + O_p\left(\frac{p_T}{N}\right) + O_p\left(\frac{p_T^2}{N\sqrt{T}}\right), \\
 \|\mathbf{K}_{85}\| &= O_p\left(\frac{\sqrt{T}}{N}\right) + O_p\left(\frac{p_T}{N}\right) + O_p\left(\frac{p_T^2}{N\sqrt{T}}\right), \\
 \|\mathbf{K}_{86}\| &= O_p\left(\frac{p_T \rho_H^{p_T}}{\sqrt{NT}}\right) + O_p\left(\frac{p_T}{N}\right) + O_p\left(\frac{p_T^2}{N\sqrt{T}}\right), \\
 \|\mathbf{K}_{87}\| &= O_p\left(\frac{\rho_H^{p_T}}{\sqrt{T}}\right) + O_p\left(\frac{p_T}{\sqrt{NT}}\right) + O_p\left(\frac{p_T^2}{N\sqrt{T}}\right), \\
 \|\mathbf{K}_{88}\| &= O_p\left(\frac{\rho_H^{p_T}}{\sqrt{NT}}\right) + O_p\left(\frac{p_T}{N}\right) + O_p\left(\frac{p_T^2}{N\sqrt{T}}\right), \\
 \|\mathbf{K}_{89}\| &= O_p\left(\frac{\rho_H^{2p_T}}{\sqrt{T}}\right) + O_p\left(\frac{p_T \rho_H^{p_T}}{\sqrt{TN}}\right) + O_p\left(\frac{p_T^2}{N\sqrt{T}}\right),
 \end{aligned}$$

which can be found using the relevant preliminary results. Combining these order results yields the required result.

To prove (3.B.108), using (3.B.114), we directly write the left-hand side in a normal form as

$$\begin{aligned}
 \|T^{3/2}\mathbf{G}'_{zi}\bar{\mathbf{M}}_Z\mathbf{X}_{i,-1}\| &\leq \|T^{3/2}\mathbf{G}'_{zi}\ddot{\mathbf{M}}_F\mathbf{X}_{i,-1}\| + \|T^{3/2}\mathbf{G}'_{zi}(\bar{\mathbf{M}}_Z - \ddot{\mathbf{M}}_F)\mathbf{X}_{i,-1}\| \\
 &= \left(\frac{1}{\sqrt{T}}\right) + O_p\left(\frac{p_T}{\sqrt{N}}\right) + O_p\left(\frac{p_T^2}{N\sqrt{T}}\right),
 \end{aligned}$$

where the order result is obtained by using Lemma 3.B.3, (3.B.86), (3.B.88) and Lemmas 3.B.12 and 3.B.13.

Finally to prove (3.B.109) we write

$$\begin{aligned}
 T^{-3/2}\mathbf{G}'_{zi}\bar{\mathbf{M}}_Z\bar{\mathbf{Z}}_{-1} &= T^{-3/2}\mathbf{G}'_{zi}\bar{\mathbf{M}}_Z\mathbf{F}_{-1}\mathbf{H}_1 + T^{-3/2}\mathbf{G}'_{zi}\bar{\mathbf{M}}_Z\mathbf{V}_{A,-1} \\
 &\quad + T^{-3/2}\mathbf{G}'_{zi}\bar{\mathbf{M}}_Z\mathbf{V}_{C,-p_T}
 \end{aligned}$$

$$= \mathbf{K}_{91} + \mathbf{K}_{92} + \mathbf{K}_{93}. \quad (3.B.120)$$

The orders of the terms on the right-hand side of the last equation are

$$\begin{aligned} \|\mathbf{K}_{91}\| &= \left(\frac{1}{\sqrt{T}}\right) + O_p\left(\frac{p_T}{\sqrt{N}}\right) + O_p\left(\frac{p_T^2}{N\sqrt{T}}\right), \\ \|\mathbf{K}_{92}\| &= \left(\frac{1}{\sqrt{NT}}\right) + O_p\left(\frac{p_T}{\sqrt{NT}}\right) + O_p\left(\frac{p_T}{N}\right) + O_p\left(\frac{p_T^2}{N\sqrt{T}}\right), \\ \|\mathbf{K}_{93}\| &= \left(\frac{\rho_H^{p_T}}{\sqrt{T}}\right) + O_p\left(\frac{p_T}{\sqrt{NT}}\right) + O_p\left(\frac{p_T^2}{N\sqrt{T}}\right), \end{aligned}$$

which can be found using the relevant preliminary results. Combining these order results yields the required result. \blacksquare

Proof of Lemma 3.B.16: To prove (3.B.110) we write

$$\begin{aligned} T^{-1/2}\tilde{\mathbf{U}}'_{i,-1}\bar{\mathbf{M}}_Z\tilde{\boldsymbol{\varepsilon}}_{y.xi} &= T^{-1/2}\mathbf{U}'_{i,-1}\ddot{\mathbf{M}}_F\boldsymbol{\varepsilon}_{y.xi} + T^{-1/2}\mathbf{U}'_{i,-1}(\bar{\mathbf{M}}_Z - \ddot{\mathbf{M}}_F)\boldsymbol{\varepsilon}_{y.xi} \\ &\quad + T^{-1/2}\mathbf{U}'_{i,-1}\bar{\mathbf{M}}_Z\mathbf{V}_{A,-1}\mathbf{H}_1^-\boldsymbol{\lambda}_i\alpha_{yi} \\ &\quad + T^{-1/2}\mathbf{U}'_{i,-1}\bar{\mathbf{M}}_Z\mathbf{V}_B\mathbf{H}_4^-(L)\mathbf{A}'_{y.fi}(L) \\ &\quad + T^{-1/2}\mathbf{U}'_{i,-1}\bar{\mathbf{M}}_Z\mathbf{V}_{Di,-p_T} \\ &\quad + T^{-1/2}\boldsymbol{\lambda}'_i\mathbf{H}_1^-\mathbf{V}'_{A,-1}\bar{\mathbf{M}}_Z\boldsymbol{\varepsilon}_{y.xi} \\ &\quad + T^{-1/2}\boldsymbol{\lambda}'_i\mathbf{H}_1^-\mathbf{V}'_{A,-1}\bar{\mathbf{M}}_Z\mathbf{V}_{A,-1}\mathbf{H}_1^-\boldsymbol{\lambda}_i\alpha_{yi} \\ &\quad + T^{-1/2}\boldsymbol{\lambda}'_i\mathbf{H}_1^-\mathbf{V}'_{A,-1}\bar{\mathbf{M}}_Z\mathbf{V}_B\mathbf{H}_4^-(L)\mathbf{A}'_{y.fi}(L) \\ &\quad + T^{-1/2}\boldsymbol{\lambda}'_i\mathbf{H}_1^-\mathbf{V}'_{A,-1}\bar{\mathbf{M}}_Z\mathbf{V}_{Di,-p_T} \\ &\quad + T^{-1/2}\boldsymbol{\lambda}'_i\mathbf{H}_1^-\mathbf{V}'_{C,-p_T}\bar{\mathbf{M}}_Z\boldsymbol{\varepsilon}_{y.xi} \\ &\quad + T^{-1/2}\boldsymbol{\lambda}'_i\mathbf{H}_1^-\mathbf{V}'_{C,-p_T}\bar{\mathbf{M}}_Z\mathbf{V}_{A,-1}\mathbf{H}_1^-\boldsymbol{\lambda}_i\alpha_{yi} \\ &\quad + T^{-1/2}\boldsymbol{\lambda}'_i\mathbf{H}_1^-\mathbf{V}'_{C,-p_T}\bar{\mathbf{M}}_Z\mathbf{V}_B\mathbf{H}_4^-(L)\mathbf{A}'_{y.fi}(L) \\ &\quad + T^{-1/2}\boldsymbol{\lambda}'_i\mathbf{H}_1^-\mathbf{V}'_{C,-p_T}\bar{\mathbf{M}}_Z\mathbf{V}_{Di,-p_T} \\ &= \mathbf{L}_{11} + \mathbf{L}_{12} + \mathbf{L}_{13} + \mathbf{L}_{14} + \mathbf{L}_{15} + \mathbf{L}_{16} + \mathbf{L}_{17} + \mathbf{L}_{18} + \mathbf{L}_{19} \\ &\quad + \mathbf{L}_{110} + \mathbf{L}_{111} + \mathbf{L}_{112} + \mathbf{L}_{113}. \end{aligned}$$

The order of \mathbf{L}_{11} can be obtained as $O_p(1)$ by using (3.B.21), (3.B.85), (3.B.88) and (3.B.89). The orders of the terms on the right-hand side of the last equation

can be stated formally as

$$\begin{aligned}
 \|\mathbf{L}_{12}\| &= O_p\left(\frac{p_T\sqrt{T}}{\sqrt{N}}\right) + O_p\left(\frac{p_T^2\sqrt{T}}{N}\right), \\
 \|\mathbf{L}_{13}\| &= O_p\left(\frac{p_T T}{N}\right) + O_p\left(\frac{p_T\sqrt{T}}{\sqrt{N}}\right) + O_p\left(\frac{p_T^2\sqrt{T}}{N}\right), \\
 \|\mathbf{L}_{14}\| &= O_p\left(\frac{p_T\sqrt{T}}{\sqrt{N}}\right) + O_p\left(\frac{p_T^2\sqrt{T}}{N}\right), \\
 \|\mathbf{L}_{15}\| &= O_p(\sqrt{T}\rho_A^{p_T}) + O_p\left(\frac{p_T\sqrt{T}}{\sqrt{N}}\right) + O_p\left(\frac{p_T^2\sqrt{T}}{N}\right), \\
 \|\mathbf{L}_{16}\| &= O_p\left(\frac{p_T\sqrt{T}}{\sqrt{N}}\right) + O_p\left(\frac{p_T T}{N}\right) + O_p\left(\frac{p_T^2\sqrt{T}}{N}\right), \\
 \|\mathbf{L}_{17}\| &= O_p\left(\frac{p_T T}{N}\right) + O_p\left(\frac{p_T\rho_A^{p_T}\sqrt{T}}{\sqrt{N}}\right) + O_p\left(\frac{p_T^2\sqrt{T}}{N}\right), \\
 \|\mathbf{L}_{18}\| &= O_p\left(\frac{p_T T}{N}\right) + O_p\left(\frac{p_T^2\sqrt{T}}{N}\right), \\
 \|\mathbf{L}_{19}\| &= O_p\left(\frac{\sqrt{T}\rho_A^{p_T} p_T}{\sqrt{N}}\right) + O_p\left(\frac{p_T T}{N}\right) + O_p\left(\frac{p_T^2\sqrt{T}}{N}\right), \\
 \|\mathbf{L}_{110}\| &= O_p(\rho_H^{p_T}) + O_p\left(\frac{p_T\sqrt{T}}{\sqrt{N}}\right) + O_p\left(\frac{p_T^2\sqrt{T}}{N}\right), \\
 \|\mathbf{L}_{111}\| &= O_p\left(\frac{\sqrt{T}\rho_H^{p_T} p_T}{\sqrt{N}}\right) + O_p\left(\frac{p_T T}{N}\right) + O_p\left(\frac{p_T^2\sqrt{T}}{N}\right), \\
 \|\mathbf{L}_{112}\| &= O_p\left(\frac{\sqrt{T}\rho_H^{p_T} p_T}{\sqrt{N}}\right) + O_p\left(\frac{p_T^2\sqrt{T}}{N}\right), \\
 \|\mathbf{L}_{113}\| &= O_p\left(\frac{\sqrt{T}\rho_H^{p_T} p_T}{\sqrt{N}}\right) + O_p\left(\frac{\sqrt{T}\rho_A^{p_T} p_T}{\sqrt{N}}\right) + O_p\left(\frac{p_T^2\sqrt{T}}{N}\right).
 \end{aligned}$$

The results stated above can be obtained by following the same method as in the proofs above, hence the detailed steps are not provided here.

To prove (3.B.111) we write

$$\begin{aligned}
 T^{1/2}\mathbf{G}'_{zi}\bar{\mathbf{M}}_Z\tilde{\varepsilon}_{y.xi} &= T^{-1/2}\mathbf{G}'_{zi}\ddot{\mathbf{M}}_F\varepsilon_{y.xi} + T^{-1/2}\mathbf{G}'_{zi}(\bar{\mathbf{M}}_Z - \ddot{\mathbf{M}}_F)\varepsilon_{y.xi} \\
 &\quad + T^{-1}\mathbf{G}'_{zi}\bar{\mathbf{M}}_Z\mathbf{V}_{A,-1}\mathbf{H}_1^-\lambda_i\alpha_{yi} \\
 &\quad + T^{-1}\mathbf{G}'_{zi}\bar{\mathbf{M}}_Z\mathbf{V}_B\mathbf{H}_4^-(L)\mathbf{A}'_{y.fi}(L) \\
 &\quad + T^{-1}\mathbf{G}'_{zi}\bar{\mathbf{M}}_Z\mathbf{V}_{Di,-p_T}
 \end{aligned}$$

$$= \mathbf{L}_{21} + \mathbf{L}_{22} + \mathbf{L}_{23} + \mathbf{L}_{24} + \mathbf{L}_{25}. \quad (3.B.121)$$

The order of \mathbf{L}_{21} is $O_p(1)$, that can be obtained by using Lemma 3.B.3, (3.B.88) and (3.B.89). The orders of the remainder terms can be established by using the relevant preliminary results as

$$\begin{aligned} \|\mathbf{L}_{22}\| &= O_p\left(\frac{p_T\sqrt{T}}{\sqrt{N}}\right) + O_p\left(\frac{p_T^2\sqrt{T}}{N}\right), \\ \|\mathbf{L}_{23}\| &= O_p\left(\frac{p_T T}{N}\right) + O_p\left(\frac{p_T\sqrt{T}}{\sqrt{N}}\right) + O_p\left(\frac{p_T^2\sqrt{T}}{N}\right), \\ \|\mathbf{L}_{24}\| &= O_p\left(\frac{p_T\sqrt{T}}{\sqrt{N}}\right) + O_p\left(\frac{p_T^2\sqrt{T}}{N}\right), \\ \|\mathbf{L}_{25}\| &= O_p(\sqrt{T}\rho_A^{p_T}) + O_p\left(\frac{p_T\sqrt{T}}{\sqrt{N}}\right) + O_p\left(\frac{p_T^2\sqrt{T}}{N}\right). \end{aligned}$$

To prove (3.B.112) we write

$$\begin{aligned} T^{-1}\mathbf{X}'_{i,-1}\bar{\mathbf{M}}_Z\tilde{\varepsilon}_{y.xi} &= T^{-1}\mathbf{X}'_{i,-1}\ddot{\mathbf{M}}_F\varepsilon_{y.xi} + T^{-1}\mathbf{X}'_{i,-1}(\bar{\mathbf{M}}_Z - \ddot{\mathbf{M}}_F)\varepsilon_{y.xi} \\ &\quad + T^{-1}\mathbf{X}'_{i,-1}\bar{\mathbf{M}}_Z\mathbf{V}_{A,-1}\mathbf{H}_1^-\lambda_i\alpha_{yi} \\ &\quad + T^{-1}\mathbf{X}'_{i,-1}\bar{\mathbf{M}}_Z\mathbf{V}_B\mathbf{H}_4^-(L)\mathbf{A}_{y.fi}^*(L) \quad (3.B.122) \\ &\quad + T^{-1}\mathbf{X}'_{i,-1}\bar{\mathbf{M}}_Z\mathbf{V}_{D_i,-p_T} \\ &= \mathbf{L}_{31} + \mathbf{L}_{32} + \mathbf{L}_{33} + \mathbf{L}_{34} + \mathbf{L}_{35}. \quad (3.B.123) \end{aligned}$$

The order of \mathbf{L}_{31} is $O_p(1)$, that can be obtained by using (3.B.18), (3.B.86), (3.B.88) and (3.B.89). The orders of the rest of the terms are

$$\begin{aligned} \|\mathbf{L}_{32}\| &= O_p\left(\frac{p_T\sqrt{T}}{\sqrt{N}}\right) + O_p\left(\frac{p_T^2}{N}\right), \quad \|\mathbf{L}_{33}\| = O_p\left(\frac{T}{\sqrt{N}}\right) + O_p\left(\frac{p_T\sqrt{T}}{\sqrt{N}}\right) + O_p\left(\frac{p_T^2}{N}\right), \\ \|\mathbf{L}_{34}\| &= O_p\left(\frac{p_T\sqrt{T}}{\sqrt{N}}\right) + O_p\left(\frac{p_T^2}{N}\right), \quad \|\mathbf{L}_{35}\| = O_p(\rho_A^{p_T}) + O_p\left(\frac{p_T\sqrt{T}}{\sqrt{N}}\right) + O_p\left(\frac{p_T^2}{N}\right), \end{aligned}$$

which can be established straightforwardly by following the similar steps that are used in the proofs of Lemma 3.B.14 and using the relevant preliminary results. To prove (3.B.113) we write

$$\begin{aligned} T^{-1}\bar{\mathbf{Z}}'_{-1}\bar{\mathbf{M}}_Z\tilde{\varepsilon}_{y.xi} &= T^{-1}\mathbf{H}'_1\mathbf{F}'_{-1}\ddot{\mathbf{M}}_F\varepsilon_{y.xi} + T^{-1}\mathbf{H}'_1\mathbf{F}'_{-1}(\bar{\mathbf{M}}_Z - \ddot{\mathbf{M}}_F)\varepsilon_{y.xi} \\ &\quad + T^{-1}\mathbf{H}'_1\mathbf{F}'_{-1}\bar{\mathbf{M}}_Z\mathbf{V}_{A,-1}\mathbf{H}_1^-\lambda_i\alpha_{yi} \end{aligned}$$

$$\begin{aligned}
 & +T^{-1}\mathbf{H}'_1\mathbf{F}'_{-1}\bar{\mathbf{M}}_Z\mathbf{V}_B\mathbf{H}_4^-(L)\mathbf{A}'_{y.fi}(L) \\
 & +T^{-1}\mathbf{H}'_1\mathbf{F}'_{-1}\bar{\mathbf{M}}_Z\mathbf{V}_{Di,-p_T} + T^{-1}\mathbf{V}'_{A,-1}\bar{\mathbf{M}}_Z\varepsilon_{y.xi} \\
 & +T^{-1}\mathbf{V}'_{A,-1}\bar{\mathbf{M}}_Z\mathbf{V}_{A,-1}\mathbf{H}_1^-\lambda_i\alpha_{yi} \\
 & +T^{-1}\mathbf{V}'_{A,-1}\bar{\mathbf{M}}_Z\mathbf{V}_B\mathbf{H}_4^-(L)\mathbf{A}'_{y.fi}(L) \\
 & +T^{-1}\mathbf{V}'_{A,-1}\bar{\mathbf{M}}_Z\mathbf{V}_{Di,-p_T} + T^{-1}\mathbf{V}'_{C,-p_T}\bar{\mathbf{M}}_Z\varepsilon_{y.xi} \\
 & +T^{-1}\mathbf{V}'_{C,-p_T}\bar{\mathbf{M}}_Z\mathbf{V}_{A,-1}\mathbf{H}_1^-\lambda_i\alpha_{yi} \\
 & +T^{-1}\mathbf{V}'_{C,-p_T}\bar{\mathbf{M}}_Z\mathbf{V}_B\mathbf{H}_4^-(L)\mathbf{A}'_{y.fi}(L) \\
 & +T^{-1}\mathbf{V}'_{C,-p_T}\bar{\mathbf{M}}_Z\mathbf{V}_{Di,-p_T} \\
 = & \mathbf{L}_{41} + \mathbf{L}_{42} + \mathbf{L}_{43} + \mathbf{L}_{44} + \mathbf{L}_{45} + \mathbf{L}_{46} + \mathbf{L}_{47} + \mathbf{L}_{48} + \mathbf{L}_{49} \\
 & + \mathbf{L}_{410} + \mathbf{L}_{411} + \mathbf{L}_{412} + \mathbf{L}_{413}.
 \end{aligned}$$

\mathbf{L}_{41} is $O_p(1)$ by (3.B.19), (3.B.87), (3.B.88) and (3.B.89). The orders of remainder of the terms can be stated as

$$\begin{aligned}
 \|\mathbf{L}_{42}\| &= O_p\left(\frac{p_T\sqrt{T}}{\sqrt{N}}\right) + O_p\left(\frac{p_T^2}{N}\right), \\
 \|\mathbf{L}_{43}\| &= O_p\left(\frac{T}{\sqrt{N}}\right) + O_p\left(\frac{p_T\sqrt{T}}{\sqrt{N}}\right) + O_p\left(\frac{p_T^2}{N}\right), \\
 \|\mathbf{L}_{44}\| &= O_p\left(\frac{p_T\sqrt{T}}{\sqrt{N}}\right) + O_p\left(\frac{p_T^2}{N}\right), \\
 \|\mathbf{L}_{45}\| &= O_p(\rho_A^{p_T}) + O_p\left(\frac{p_T\sqrt{T}}{\sqrt{N}}\right) + O_p\left(\frac{p_T^2}{N}\right), \\
 \|\mathbf{L}_{46}\| &= O_p\left(\frac{p_T\sqrt{T}}{N}\right) + O_p\left(\frac{p_T}{\sqrt{N}}\right) + O_p\left(\frac{p_T^2}{N}\right), \\
 \|\mathbf{L}_{47}\| &= O_p\left(\frac{T}{N}\right) + O_p\left(\frac{p_T\sqrt{T}}{N}\right) + O_p\left(\frac{p_T^2}{N}\right), \\
 \|\mathbf{L}_{48}\| &= O_p\left(\frac{T}{N}\right) + O_p\left(\frac{p_T\sqrt{T}}{N}\right) + O_p\left(\frac{p_T^2}{N}\right), \\
 \|\mathbf{L}_{49}\| &= O_p\left(\frac{p_T\sqrt{T}}{N}\right) + O_p\left(\frac{p_T\rho_A^{p_T}}{\sqrt{N}}\right) + O_p\left(\frac{p_T^2}{N}\right), \\
 \|\mathbf{L}_{410}\| &= O_p\left(\frac{\rho_H^{p_T}}{\sqrt{T}}\right) + O_p\left(\frac{p_T}{\sqrt{N}}\right) + O_p\left(\frac{p_T^2}{N}\right), \\
 \|\mathbf{L}_{411}\| &= O_p\left(\frac{p_T\rho_H^{p_T}}{\sqrt{N}}\right) + O_p\left(\frac{p_T\sqrt{T}}{\sqrt{N}}\right) + O_p\left(\frac{p_T^2}{N}\right),
 \end{aligned}$$

$$\|\mathbf{L}_{412}\| = O_p\left(\frac{p_T \rho_H^{p_T}}{\sqrt{N}}\right) + O_p\left(\frac{p_T^2}{N}\right),$$

$$\|\mathbf{L}_{413}\| = O_p\left(\frac{p_T \rho_H^{p_T}}{\sqrt{N}}\right) + O_p\left(\frac{p_T \rho_A^{p_T}}{\sqrt{N}}\right) + O_p\left(\frac{p_T^2}{N}\right).$$

The order in (3.B.113) can be obtained by combining the relevant results above.

■

Proof of Theorem 3.3

To prove the theorem we use a technique similar to the one used in the proof of Theorem 3.1. We start with rewriting the estimator by substituting the augmented model in the estimator in (3.32), which yields

$$\begin{pmatrix} \widehat{\Pi}_i - \Pi_i^* \\ \widehat{\Gamma}_{zi} - \Gamma_{zi} \end{pmatrix} = \begin{pmatrix} \widetilde{\mathbf{Z}}_{i,-1}^{*'} \overline{\mathbf{M}}_Z \widetilde{\mathbf{Z}}_{i,-1}^* & \widetilde{\mathbf{Z}}_{i,-1}^{*'} \overline{\mathbf{M}}_Z \mathbf{G}_{zi} \\ \mathbf{G}'_{zi} \overline{\mathbf{M}}_Z \widetilde{\mathbf{Z}}_{i,-1}^* & \mathbf{G}'_{zi} \overline{\mathbf{M}}_Z \mathbf{G}_{zi} \end{pmatrix}^{-1} \begin{pmatrix} \widetilde{\mathbf{Z}}_{i,-1}^{*'} \overline{\mathbf{M}}_Z \widetilde{\boldsymbol{\varepsilon}}_{y.xi} \\ \mathbf{G}'_{zi} \overline{\mathbf{M}}_Z \widetilde{\boldsymbol{\varepsilon}}_{y.xi} \end{pmatrix}. \quad (3.B.124)$$

It is useful to remind the notation at this stage. Note that

$\widetilde{\mathbf{Z}}_{i,-1}^* = (\mathbf{Y}_{i,-1}, \mathbf{X}_{i,-1}, \overline{\mathbf{Z}}_{-1})$ and $\mathbf{G}_{zi} = (\mathbf{G}_{yi}, \mathbf{G}_{xi})$, $\overline{\mathbf{M}}_Z = \mathbf{I} - \overline{\mathbf{G}}_z (\overline{\mathbf{G}}_z' \overline{\mathbf{G}}_z)^{-1} \overline{\mathbf{G}}_z'$ and

$$\begin{aligned} \mathbf{G}_{yi} &= (\Delta \mathbf{y}_{i,-1}, \dots, \Delta \mathbf{y}_{i,-p+1}), \\ \mathbf{G}_{xi} &= (\Delta \mathbf{X}_i, \dots, \Delta \mathbf{X}_{i,-p+1}), \\ \overline{\mathbf{G}}_z &= (\overline{\Delta \mathbf{Z}}_{-1}, \dots, \overline{\Delta \mathbf{Z}}_{-p_T}), \end{aligned}$$

where $\Delta \mathbf{y}_i = (\Delta y_{i,p_T+1}, \dots, \Delta y_{i,T})'$, $\Delta \mathbf{X}_i = (\Delta \mathbf{x}_{i,p_T+1}, \dots, \Delta \mathbf{x}_{i,T})'$ and $\overline{\Delta \mathbf{Z}} = (\overline{\Delta \mathbf{z}}_{p_T+1}, \dots, \overline{\Delta \mathbf{z}}_T)'$.

We define \mathbf{Q}_i and \mathbf{Q}_i^{-1} analogously to \mathbf{Q}_i^* and \mathbf{Q}_i^{*-1} defined in (3.B.24) as

$$\mathbf{Q}_i = \begin{pmatrix} 1 & 0 & 0 & 0 \\ -\boldsymbol{\beta}_{xi} & 0 & \mathbf{I}_m & 0 \\ -\mathbf{H}_1^- \boldsymbol{\lambda}_i & 0 & 0 & \mathbf{I}_k \\ 0 & \mathbf{I}_{q^*} & 0 & 0 \end{pmatrix}, \quad \mathbf{Q}_i^{-1} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & \mathbf{I}_{q^*} \\ \boldsymbol{\beta}_{xi} & \mathbf{I}_m & 0 & 0 \\ \mathbf{H}_1^- \boldsymbol{\lambda}_i & 0 & \mathbf{I}_k & 0 \end{pmatrix}, \quad (3.B.125)$$

where $q^* = (1 + m)p - 1$ is the number of columns of \mathbf{G}_{zi} . Using the diagonal standardization matrix $\mathbf{D} = \sqrt{T} \times \mathbf{I}_{1+q^*+m+k}$ we write

$$\begin{aligned} & \mathbf{DQ}_i^{-1} \begin{pmatrix} \widehat{\Pi}_i - \Pi_i^* \\ \widehat{\Gamma}_{zi} - \Gamma_{zi} \end{pmatrix} \\ &= \begin{pmatrix} \frac{\widetilde{\mathbf{U}}'_{i,-1} \overline{\mathbf{M}}_Z \widetilde{\mathbf{U}}_{i,-1}}{T} & \frac{\widetilde{\mathbf{U}}'_{i,-1} \overline{\mathbf{M}}_Z \mathbf{G}_{zi}}{T} & \frac{\widetilde{\mathbf{U}}'_{i,-1} \overline{\mathbf{M}}_Z \mathbf{X}_{i,-1}}{T^{3/2}} & \frac{\widetilde{\mathbf{U}}'_{i,-1} \overline{\mathbf{M}}_Z \overline{\mathbf{Z}}_{-1}}{T^{3/2}} \\ \frac{\mathbf{G}'_{zi} \overline{\mathbf{M}}_Z \widetilde{\mathbf{U}}_{i,-1}}{T} & \frac{\mathbf{G}'_{zi} \overline{\mathbf{M}}_Z \mathbf{G}_{zi}}{T} & \frac{\mathbf{G}'_{zi} \overline{\mathbf{M}}_Z \mathbf{X}_{i,-1}}{T^{3/2}} & \frac{\mathbf{G}'_{zi} \overline{\mathbf{M}}_Z \overline{\mathbf{Z}}_{-1}}{T^{3/2}} \\ \frac{\mathbf{X}'_{i,-1} \overline{\mathbf{M}}_Z \widetilde{\mathbf{U}}_{i,-1}}{T^{3/2}} & \frac{\mathbf{X}'_{i,-1} \overline{\mathbf{M}}_Z \mathbf{G}_{zi}}{T^{3/2}} & \frac{\mathbf{X}'_{i,-1} \overline{\mathbf{M}}_Z \mathbf{X}_{i,-1}}{T^2} & \frac{\mathbf{X}'_{i,-1} \overline{\mathbf{M}}_Z \overline{\mathbf{Z}}_{-1}}{T^2} \\ \frac{\overline{\mathbf{Z}}'_{-1} \overline{\mathbf{M}}_Z \widetilde{\mathbf{U}}_{i,-1}}{T^{3/2}} & \frac{\overline{\mathbf{Z}}'_{-1} \overline{\mathbf{M}}_Z \mathbf{G}_{zi}}{T^{3/2}} & \frac{\overline{\mathbf{Z}}'_{-1} \overline{\mathbf{M}}_Z \mathbf{X}_{i,-1}}{T^2} & \frac{\overline{\mathbf{Z}}'_{-1} \overline{\mathbf{M}}_Z \overline{\mathbf{Z}}_{-1}}{T^2} \end{pmatrix}^{-1} \\ & \quad \times \begin{pmatrix} \frac{\widetilde{\mathbf{U}}'_{i,-1} \overline{\mathbf{M}}_Z \tilde{\epsilon}_{y.xi}}{T^{1/2}} \\ \frac{\mathbf{G}'_{zi} \overline{\mathbf{M}}_Z \tilde{\epsilon}_{y.xi}}{T^{1/2}} \\ \frac{\mathbf{X}'_{i,-1} \overline{\mathbf{M}}_Z \tilde{\epsilon}_{y.xi}}{T^{3/2}} \\ \frac{\overline{\mathbf{Z}}'_{-1} \overline{\mathbf{M}}_Z \tilde{\epsilon}_{y.xi}}{T^{3/2}} \end{pmatrix}, \end{aligned} \quad (3.B.126)$$

where $\widetilde{\mathbf{U}}_{i,-1}$ is defined in (3.B.59). By using Lemma 3.B.15, it can be shown that the off-block-diagonal blocks of the first part of the right-hand side are $o_p(1)$ under the conditions $\frac{T}{N} \rightarrow 0$, $\frac{p_T^2}{T} \rightarrow 0$ as $(N, T, p_T) \rightarrow \infty$. Then by using the familiar rules on partitioned matrix inverses we can write

$$\begin{aligned} & \mathbf{DQ}_i^{-1} \begin{pmatrix} \widehat{\Pi}_i - \Pi_i^* \\ \widehat{\Gamma}_{zi} - \Gamma_{zi} \end{pmatrix} \\ &= \begin{pmatrix} \left(\begin{pmatrix} \frac{\widetilde{\mathbf{U}}'_{i,-1} \overline{\mathbf{M}}_Z \widetilde{\mathbf{U}}_{i,-1}}{T} & \frac{\widetilde{\mathbf{U}}'_{i,-1} \overline{\mathbf{M}}_Z \mathbf{G}_{zi}}{T} \\ \frac{\mathbf{G}'_{zi} \overline{\mathbf{M}}_Z \widetilde{\mathbf{U}}_{i,-1}}{T} & \frac{\mathbf{G}'_{zi} \overline{\mathbf{M}}_Z \mathbf{G}_{zi}}{T} \end{pmatrix}^{-1} \begin{pmatrix} \frac{\widetilde{\mathbf{U}}'_{i,-1} \overline{\mathbf{M}}_Z \tilde{\epsilon}_{y.xi}}{T^{1/2}} \\ \frac{\mathbf{G}'_{zi} \overline{\mathbf{M}}_Z \tilde{\epsilon}_{y.xi}}{T^{1/2}} \end{pmatrix} \right) \\ \left(\begin{pmatrix} \frac{\mathbf{X}'_{i,-1} \overline{\mathbf{M}}_Z \mathbf{X}_{i,-1}}{T^2} & \frac{\mathbf{X}'_{i,-1} \overline{\mathbf{M}}_Z \overline{\mathbf{Z}}_{-1}}{T^2} \\ \frac{\overline{\mathbf{Z}}'_{-1} \overline{\mathbf{M}}_Z \mathbf{X}_{i,-1}}{T^2} & \frac{\overline{\mathbf{Z}}'_{-1} \overline{\mathbf{M}}_Z \overline{\mathbf{Z}}_{-1}}{T^2} \end{pmatrix}^{-1} \begin{pmatrix} \frac{\mathbf{X}'_{i,-1} \overline{\mathbf{M}}_Z \tilde{\epsilon}_{y.xi}}{T^{3/2}} \\ \frac{\overline{\mathbf{Z}}'_{-1} \overline{\mathbf{M}}_Z \tilde{\epsilon}_{y.xi}}{T^{3/2}} \end{pmatrix} \right) \end{pmatrix} \\ & \quad + o_p(1). \end{aligned}$$

Now using the results in Lemmas 3.B.14-3.B.16, we analyze the expressions separately. We have

$$\begin{pmatrix} \frac{\widetilde{\mathbf{U}}'_{i,-1} \overline{\mathbf{M}}_Z \widetilde{\mathbf{U}}_{i,-1}}{T} & \frac{\widetilde{\mathbf{U}}'_{i,-1} \overline{\mathbf{M}}_Z \mathbf{G}_{zi}}{T} \\ \frac{\mathbf{G}'_{zi} \overline{\mathbf{M}}_Z \widetilde{\mathbf{U}}_{i,-1}}{T} & \frac{\mathbf{G}'_{zi} \overline{\mathbf{M}}_Z \mathbf{G}_{zi}}{T} \end{pmatrix} - \begin{pmatrix} \frac{\mathbf{U}'_{i,-1} \check{\mathbf{M}}_F \mathbf{U}_{i,-1}}{T} & \frac{\mathbf{U}'_{i,-1} \check{\mathbf{M}}_F \mathbf{G}_{zi}}{T} \\ \frac{\mathbf{G}'_{zi} \check{\mathbf{M}}_F \mathbf{U}_{i,-1}}{T} & \frac{\mathbf{G}'_{zi} \check{\mathbf{M}}_F \mathbf{G}_{zi}}{T} \end{pmatrix}$$

$$= O_p\left(\frac{T}{N}\right) + O_p\left(\frac{p_T^2}{N}\right) + O_p\left(\frac{p_T\sqrt{T}}{N}\right),$$

by the order results stated in (3.B.100)-(3.B.102). Similarly we have

$$\begin{aligned} & \begin{pmatrix} \frac{\mathbf{X}'_{i,-1}\bar{\mathbf{M}}_Z\mathbf{X}_{i,-1}}{T^2} & \frac{\mathbf{X}'_{i,-1}\bar{\mathbf{M}}_Z\bar{\mathbf{Z}}_{-1}}{T^2} \\ \frac{\bar{\mathbf{Z}}'_{-1}\bar{\mathbf{M}}_Z\mathbf{X}_{i,-1}}{T^2} & \frac{\bar{\mathbf{Z}}'_{-1}\bar{\mathbf{M}}_Z\bar{\mathbf{Z}}_{-1}}{T^2} \end{pmatrix} - \begin{pmatrix} \frac{\mathbf{X}'_{i,-1}\ddot{\mathbf{M}}_F\mathbf{X}_{i,-1}}{T^2} & \frac{\mathbf{X}'_{i,-1}\ddot{\mathbf{M}}_F\mathbf{F}_{-1}\mathbf{H}_1}{T^2} \\ \frac{\mathbf{H}'_1\mathbf{F}'_{-1}\ddot{\mathbf{M}}_F\mathbf{X}_{i,-1}}{T^2} & \frac{\mathbf{H}'_1\mathbf{F}'_{-1}\ddot{\mathbf{M}}_F\mathbf{F}_{-1}\mathbf{H}_1}{T^2} \end{pmatrix} \\ &= O_p\left(\frac{p_T^2}{NT}\right) + O_p\left(\frac{p_T}{\sqrt{NT}}\right), \end{aligned}$$

by (3.B.103)-(3.B.105). And

$$\begin{aligned} & \begin{pmatrix} \frac{\tilde{\mathbf{U}}'_{i,-1}\bar{\mathbf{M}}_Z\tilde{\boldsymbol{\varepsilon}}_{y,xi}}{T^{1/2}} \\ \frac{\mathbf{G}'_{zi}\bar{\mathbf{M}}_Z\tilde{\boldsymbol{\varepsilon}}_{y,xi}}{T^{1/2}} \end{pmatrix} - \begin{pmatrix} \frac{\mathbf{U}'_{i,-1}\ddot{\mathbf{M}}_F\boldsymbol{\varepsilon}_{y,xi}}{T^{1/2}} \\ \frac{\mathbf{G}'_{zi}\ddot{\mathbf{M}}_F\boldsymbol{\varepsilon}_{y,xi}}{T^{1/2}} \end{pmatrix} \\ &= O_p(\sqrt{T}\rho_A^{p_T}) + O_p(\rho_H^{p_T}) + O_p\left(\frac{p_T\sqrt{T}}{\sqrt{N}}\right) + O_p\left(\frac{p_T T}{N}\right) + O_p\left(\frac{p_T^2\sqrt{T}}{N}\right), \end{aligned}$$

by (3.B.110) and (3.B.111). And finally we have

$$\begin{pmatrix} \frac{\mathbf{X}'_{i,-1}\bar{\mathbf{M}}_Z\tilde{\boldsymbol{\varepsilon}}_{y,xi}}{T^{3/2}} \\ \frac{\bar{\mathbf{Z}}'_{-1}\bar{\mathbf{M}}_Z\tilde{\boldsymbol{\varepsilon}}_{y,xi}}{T^{3/2}} \end{pmatrix} = O_p\left(\frac{1}{\sqrt{T}}\right) + O_p\left(\frac{\sqrt{T}}{\sqrt{N}}\right) + O_p\left(\frac{p_T}{\sqrt{N}}\right) + O_p\left(\frac{p_T^2}{N\sqrt{T}}\right),$$

by using (3.B.112), (3.B.113). Then by combining the results above, under the conditions of the theorem, we have

$$\begin{aligned} & \mathbf{DQ}_i^{-1} \begin{pmatrix} \widehat{\boldsymbol{\Pi}}_i - \boldsymbol{\Pi}_i^* \\ \widehat{\boldsymbol{\Gamma}}_{zi} - \boldsymbol{\Gamma}_{zi} \end{pmatrix} \\ &= \begin{pmatrix} \begin{pmatrix} \frac{\mathbf{U}'_{i,-1}\ddot{\mathbf{M}}_F\mathbf{U}_{i,-1}}{T} & \frac{\mathbf{U}'_{i,-1}\ddot{\mathbf{M}}_F\mathbf{G}_{zi}}{T} \\ \frac{\mathbf{G}'_{zi}\ddot{\mathbf{M}}_F\mathbf{U}_{i,-1}}{T} & \frac{\mathbf{G}'_{zi}\ddot{\mathbf{M}}_F\mathbf{G}_{zi}}{T} \end{pmatrix}^{-1} \begin{pmatrix} \frac{\mathbf{U}'_{i,-1}\ddot{\mathbf{M}}_F\boldsymbol{\varepsilon}_{y,xi}}{T^{1/2}} \\ \frac{\mathbf{G}'_{zi}\ddot{\mathbf{M}}_F\boldsymbol{\varepsilon}_{y,xi}}{T^{1/2}} \end{pmatrix} \\ o_p(1) \end{pmatrix} \\ &+ o_p(1). \end{aligned}$$

A final step is to premultiply both sides of the equation above by \mathbf{Q}_i . Note that using the definition for $\boldsymbol{\alpha}_{i\perp}$ given in (3.B.1) and by noting the definition $\boldsymbol{\beta}_i^* = (1, -\boldsymbol{\beta}'_{xi}, -\mathbf{H}_1^{-1}\boldsymbol{\lambda}'_i)'$, we can write $\mathbf{Q}_i = \begin{pmatrix} \boldsymbol{\beta}_i^* & 0 & \boldsymbol{\alpha}_{i\perp} \\ 0 & \mathbf{I}_{q^*} & 0 \end{pmatrix}$. Then the pre-

multiplication yields

$$\begin{aligned} & \sqrt{T} \begin{pmatrix} \widehat{\Pi}_i - \Pi_i^* \\ \widehat{\Gamma}_{zi} - \Gamma_{zi} \end{pmatrix} \\ &= \begin{pmatrix} \beta_i^* & 0 \\ 0 & \mathbf{I}_{q^*} \end{pmatrix} \begin{pmatrix} T^{-1} \mathbf{U}'_{i,-1} \ddot{\mathbf{M}}_F \mathbf{U}_{i,-1} & T^{-1} \mathbf{U}'_{i,-1} \ddot{\mathbf{M}}_F \mathbf{G}_{zi} \\ T^{-1} \mathbf{G}'_{zi} \ddot{\mathbf{M}}_F \mathbf{U}_{i,-1} & T^{-1} \mathbf{G}'_{zi} \ddot{\mathbf{M}}_F \mathbf{G}_{zi} \end{pmatrix}^{-1} \\ & \quad \times \begin{pmatrix} T^{-1/2} \mathbf{U}'_{i,-1} \ddot{\mathbf{M}}_F \varepsilon_{y.xi} \\ T^{-1/2} \mathbf{G}'_{zi} \ddot{\mathbf{M}}_F \varepsilon_{y.xi} \end{pmatrix} + o_p(1). \end{aligned}$$

Using the assumption that the conditional error term $\varepsilon_{y.xit}$ are i.i.d. for all i and t , the asymptotic distribution of the estimators can be written as

$$\sqrt{T} \begin{pmatrix} \widehat{\Pi}_i - \Pi_i^* \\ \widehat{\Gamma}_i - \Gamma_i^* \end{pmatrix} \xrightarrow{d} N \left(0, \begin{pmatrix} \beta_i^* & 0 \\ 0 & \mathbf{I}_{q^*} \end{pmatrix} \Psi_i^{-1} \begin{pmatrix} \beta_i^* & 0 \\ 0 & \mathbf{I}_{q^*} \end{pmatrix}' \sigma_{y.xfi} \right),$$

where

$$\Psi_i = \lim_{N,T,p_T \rightarrow \infty} \begin{pmatrix} T^{-1} \mathbf{U}'_{i,-1} \ddot{\mathbf{M}}_F \mathbf{U}_{i,-1} & T^{-1} \mathbf{U}'_{i,-1} \ddot{\mathbf{M}}_F \mathbf{G}_{zi} \\ T^{-1} \mathbf{G}'_{zi} \ddot{\mathbf{M}}_F \mathbf{U}_{i,-1} & T^{-1} \mathbf{G}'_{zi} \ddot{\mathbf{M}}_F \mathbf{G}_{zi} \end{pmatrix}.$$

In the second part of the theorem we propose an estimator for Ψ_i . The remainder of the proof is devoted to the proof of the proposition that the proposed estimator consistently estimates Ψ_i . For this purpose it is useful to start with writing

$$T^{-1} \begin{pmatrix} \widetilde{\mathbf{Z}}'_{i,-1} \overline{\mathbf{M}}_Z \widetilde{\mathbf{Z}}^*_{i,-1} & \widetilde{\mathbf{Z}}'_{i,-1} \overline{\mathbf{M}}_Z \mathbf{G}_{zi} \\ \mathbf{G}'_{zi} \overline{\mathbf{M}}_Z \widetilde{\mathbf{Z}}^*_{i,-1} & \mathbf{G}'_{zi} \overline{\mathbf{M}}_Z \mathbf{G}_{zi} \end{pmatrix}^{-1} = \mathbf{Q}_i \begin{pmatrix} \widehat{\Phi}_{11i} & \widehat{\Phi}_{12i} \\ \widehat{\Phi}_{21i} & \widehat{\Phi}_{22i} \end{pmatrix}^{-1} \mathbf{Q}'_i,$$

where

$$\begin{aligned} \widehat{\Phi}_{11i} &= T^{-1} \begin{pmatrix} \widetilde{\mathbf{U}}'_{i,-1} \overline{\mathbf{M}}_Z \widetilde{\mathbf{U}}_{i,-1} & \widetilde{\mathbf{U}}'_{i,-1} \overline{\mathbf{M}}_Z \mathbf{G}_{zi} \\ \mathbf{G}'_{zi} \overline{\mathbf{M}}_Z \widetilde{\mathbf{U}}_{i,-1} & \mathbf{G}'_{zi} \overline{\mathbf{M}}_Z \mathbf{G}_{zi} \end{pmatrix}, \\ \widehat{\Phi}_{12i} &= T^{-1} \begin{pmatrix} \widetilde{\mathbf{U}}'_{i,-1} \overline{\mathbf{M}}_Z \mathbf{X}_{i,-1} & \widetilde{\mathbf{U}}'_{i,-1} \overline{\mathbf{M}}_Z \overline{\mathbf{Z}}_{-1} \\ \mathbf{G}'_{zi} \overline{\mathbf{M}}_Z \mathbf{X}_{i,-1} & \mathbf{G}'_{zi} \overline{\mathbf{M}}_Z \overline{\mathbf{Z}}_{-1} \end{pmatrix}, \\ \widehat{\Phi}_{22i} &= T^{-1} \begin{pmatrix} \mathbf{X}'_{i,-1} \overline{\mathbf{M}}_Z \mathbf{X}_{i,-1} & \mathbf{X}'_{i,-1} \overline{\mathbf{M}}_Z \overline{\mathbf{Z}}_{-1} \\ \overline{\mathbf{Z}}'_{-1} \overline{\mathbf{M}}_Z \mathbf{X}_{i,-1} & \overline{\mathbf{Z}}'_{-1} \overline{\mathbf{M}}_Z \overline{\mathbf{Z}}_{-1} \end{pmatrix}, \\ \widehat{\Phi}_{21i} &= \widehat{\Phi}'_{12i}. \end{aligned}$$

We need the orders of the stochastic matrices above, which can be obtained by using the results in Lemmas 3.B.14 and 3.B.15. We have

$$\begin{aligned}\widehat{\Phi}_{11i} &= O_p(1) + O_p(TN^{-1}) + O_p(p_T^2 N^{-1}) + O_p(p_T T^{-1/2} N^{-1}), \\ \frac{\widehat{\Phi}_{12i}}{\sqrt{T}} &= O_p(T^{1/2} N^{-1/2}) + O_p(p_T N^{-1/2}) + O_p(p_T^2 N^{-1}), \\ \frac{\widehat{\Phi}_{22i}}{T} &= O_p(1) + O_p(p_T T^{-1/2} N^{-1/2}) + O_p(p_T^2 T^{-1} N^{-1}), \\ \frac{\widehat{\Phi}_{21i}}{\sqrt{T}} &= O_p(T^{1/2} N^{-1/2}) + O_p(p_T N^{-1/2}) + O_p(p_T^2 N^{-1}).\end{aligned}$$

We write

$$\begin{pmatrix} \widehat{\Phi}_{11i} & \widehat{\Phi}_{12i} \\ \widehat{\Phi}_{21i} & \widehat{\Phi}_{22i} \end{pmatrix}^{-1} = \begin{pmatrix} \widehat{\Phi}_{11i}^I & \widehat{\Phi}_{12i}^I \\ \widehat{\Phi}_{21i}^I & \widehat{\Phi}_{22i}^I \end{pmatrix},$$

then by using the familiar rules for partitioned inverses we have

$$\begin{aligned}\widehat{\Phi}_{11i}^I &= \widehat{\Phi}_{11i}^{-1} + \widehat{\Phi}_{11i}^{-1} \left(\frac{\widehat{\Phi}_{12i}}{\sqrt{T}} \right) \left[\left(\frac{\widehat{\Phi}_{22i}}{T} \right)^{-1} - \left(\frac{\widehat{\Phi}_{21i}}{\sqrt{T}} \right) \widehat{\Phi}_{11i}^{-1} \left(\frac{\widehat{\Phi}_{12i}}{\sqrt{T}} \right) \right]^{-1} \\ &\quad \times \frac{\widehat{\Phi}_{21i}}{\sqrt{T}} \widehat{\Phi}_{11i}^{-1}, \\ \widehat{\Phi}_{12i}^I &= -\frac{1}{\sqrt{T}} \widehat{\Phi}_{11i}^{-1} \frac{\widehat{\Phi}_{12i}}{\sqrt{T}} \left[\left(\frac{\widehat{\Phi}_{22i}}{T} \right)^{-1} - \left(\frac{\widehat{\Phi}_{21i}}{\sqrt{T}} \right) \widehat{\Phi}_{11i}^{-1} \left(\frac{\widehat{\Phi}_{12i}}{\sqrt{T}} \right) \right]^{-1}, \\ \widehat{\Phi}_{21i}^I &= -\frac{1}{\sqrt{T}} \left[\left(\frac{\widehat{\Phi}_{22i}}{T} \right)^{-1} - \left(\frac{\widehat{\Phi}_{21i}}{\sqrt{T}} \right) \widehat{\Phi}_{11i}^{-1} \left(\frac{\widehat{\Phi}_{12i}}{\sqrt{T}} \right) \right]^{-1} \frac{\widehat{\Phi}_{21i}}{\sqrt{T}} \widehat{\Phi}_{11i}^{-1}, \\ \widehat{\Phi}_{22i}^I &= \frac{1}{T} \left[\left(\frac{\widehat{\Phi}_{22i}}{T} \right)^{-1} - \left(\frac{\widehat{\Phi}_{21i}}{\sqrt{T}} \right) \widehat{\Phi}_{11i}^{-1} \left(\frac{\widehat{\Phi}_{12i}}{\sqrt{T}} \right) \right]^{-1}.\end{aligned}$$

Using the results provided in lemmas 3.B.14 and 3.B.15 and by considering the definition of $\widehat{\Phi}_{11i}$, under the conditions of the theorem it can be shown that

$$T^{-1} \begin{pmatrix} \widetilde{\mathbf{Z}}_{i,-1}^{*'} \overline{\mathbf{M}}_Z \widetilde{\mathbf{Z}}_{i,-1}^* & \widetilde{\mathbf{Z}}_{i,-1}^{*'} \overline{\mathbf{M}}_Z \mathbf{G}_{zi} \\ \mathbf{G}'_{zi} \overline{\mathbf{M}}_Z \widetilde{\mathbf{Z}}_{i,-1}^* & \mathbf{G}'_{zi} \overline{\mathbf{M}}_Z \mathbf{G}_{zi} \end{pmatrix}^{-1}$$

$$= \mathbf{Q}_i \left(T^{-1} \begin{pmatrix} \tilde{\mathbf{U}}'_{i,-1} \bar{\mathbf{M}}_Z \tilde{\mathbf{U}}_{i,-1} & \tilde{\mathbf{U}}'_{i,-1} \bar{\mathbf{M}}_Z \mathbf{G}_{zi} \\ \mathbf{G}'_{zi} \bar{\mathbf{M}}_Z \tilde{\mathbf{U}}_{i,-1} & \mathbf{G}'_{zi} \bar{\mathbf{M}}_Z \mathbf{G}_{zi} \end{pmatrix}^{-1} + o_p(1) \quad o_p(1) \right) \mathbf{Q}'_i.$$

Premultiplying the right-hand side expression with \mathbf{Q}_i^{-1} and post-multiplying with $\mathbf{Q}_i^{-1'}$ yields the following result:

$$T^{-1} \begin{pmatrix} \tilde{\mathbf{Z}}'^*_{i,-1} \bar{\mathbf{M}}_Z \tilde{\mathbf{Z}}^*_{i,-1} & \tilde{\mathbf{Z}}'^*_{i,-1} \bar{\mathbf{M}}_Z \mathbf{G}_{zi} \\ \mathbf{G}'_{zi} \bar{\mathbf{M}}_Z \tilde{\mathbf{Z}}^*_{i,-1} & \mathbf{G}'_{zi} \bar{\mathbf{M}}_Z \mathbf{G}_{zi} \end{pmatrix}^{-1} \\ \xrightarrow{p} \begin{pmatrix} \boldsymbol{\beta}^*_i & 0 \\ 0 & \mathbf{I}_{q^*} \end{pmatrix} \boldsymbol{\Psi}_i^{-1} \begin{pmatrix} \boldsymbol{\beta}^{*'}_i & 0 \\ 0 & \mathbf{I}_{q^*} \end{pmatrix}.$$

where $\boldsymbol{\Psi}_i$ is the expectation of $\boldsymbol{\Phi}_{11i}$ as $T, N, p_T \rightarrow \infty$. This completes the proof of the second part of the theorem. \blacksquare

Proof of Theorem 3.4

To prove the theorem we use a technique that is similar to the one used in the the proof of Theorem 3.2. First, we assume that α_{yi} is known and rewrite the model in (3.28) as

$$\Delta \mathbf{y}_i - \mathbf{y}_{i,-1} \alpha_{yi} = -\mathbf{X}_{i,-1} \boldsymbol{\beta}_{xi} \alpha_{yi} - \bar{\mathbf{Z}}_{-1} \boldsymbol{\lambda}_i^* \alpha_{yi} + \tilde{\mathbf{G}}_i \boldsymbol{\Gamma}_i^* + \tilde{\boldsymbol{\varepsilon}}_{y.xi}. \quad (3.B.127)$$

We denote the indirect least squares estimator for $\boldsymbol{\beta}_{xi}$ and $\boldsymbol{\lambda}_i^*$ for the case where α_{yi} is observed by $\hat{\hat{\boldsymbol{\beta}}}_{xi}$ and $\hat{\hat{\boldsymbol{\lambda}}}_i$ respectively. The estimator can be written as

$$\begin{pmatrix} \hat{\hat{\boldsymbol{\beta}}}_{xi} \\ \hat{\hat{\boldsymbol{\lambda}}}_i \end{pmatrix} = -\frac{1}{\alpha_{yi}} \begin{pmatrix} \mathbf{X}'_{i,-1} \tilde{\mathbf{M}}_i \mathbf{X}_{i,-1} & \mathbf{X}'_{i,-1} \tilde{\mathbf{M}}_i \bar{\mathbf{Z}}_{-1} \\ \bar{\mathbf{Z}}'_{-1} \tilde{\mathbf{M}}_i \mathbf{X}_{i,-1} & \bar{\mathbf{Z}}'_{-1} \tilde{\mathbf{M}}_i \bar{\mathbf{Z}}_{-1} \end{pmatrix}^{-1} \\ \times \begin{pmatrix} \mathbf{X}'_{i,-1} \tilde{\mathbf{M}}_i \\ \bar{\mathbf{Z}}'_{-1} \tilde{\mathbf{M}}_i \end{pmatrix} (\Delta \mathbf{y}_i - \mathbf{y}_{i,-1} \alpha_{yi}).$$

Replacing $(\Delta \mathbf{y}_i - \mathbf{y}_{i,-1} \alpha_{yi})$ by the right-hand side of (3.B.127) yields

$$T \begin{pmatrix} \hat{\hat{\boldsymbol{\beta}}}_{xi} - \boldsymbol{\beta}_{xi} \\ \hat{\hat{\boldsymbol{\lambda}}}_i - \boldsymbol{\lambda}_i^* \end{pmatrix} = -\frac{1}{\alpha_{yi}} \begin{pmatrix} \frac{\mathbf{X}'_{i,-1} \tilde{\mathbf{M}}_i \mathbf{X}_{i,-1}}{T^2} & \frac{\mathbf{X}'_{i,-1} \tilde{\mathbf{M}}_i \bar{\mathbf{Z}}_{-1}}{T^2} \\ \frac{\bar{\mathbf{Z}}'_{-1} \tilde{\mathbf{M}}_i \mathbf{X}_{i,-1}}{T^2} & \frac{\bar{\mathbf{Z}}'_{-1} \tilde{\mathbf{M}}_i \bar{\mathbf{Z}}_{-1}}{T^2} \end{pmatrix}^{-1}$$

$$\times \begin{pmatrix} \frac{\mathbf{X}'_{i,-1} \widetilde{\mathbf{M}}_i \bar{\boldsymbol{\varepsilon}}_{y.xi}}{T} \\ \frac{\bar{\mathbf{Z}}'_{-1} \widetilde{\mathbf{M}}_i \bar{\boldsymbol{\varepsilon}}_{y.xi}}{T} \end{pmatrix}. \quad (3.B.128)$$

Notice that the lemmas 3.B.11, 3.B.12 and 3.B.13 are given in terms of $\bar{\mathbf{M}}_Z$ and its large N counterpart $\ddot{\mathbf{M}}_F$, whereas in the current proof we need the results on the processes with $\widetilde{\mathbf{M}}_i$, which has a large N counterpart $\ddot{\mathbf{M}}_i$ defined in (3.B.81). Since the non-zero elements of the difference $(\widetilde{\mathbf{M}}_i - \ddot{\mathbf{M}}_i)$ are the same as the elements of the difference $(\bar{\mathbf{M}}_Z - \ddot{\mathbf{M}}_F)$, it can be shown that the results presented in the above mentioned lemmas are valid for the orthogonal projection matrices $\widetilde{\mathbf{M}}_i$ and $\ddot{\mathbf{M}}_i$. By using this fact it is straightforward to show that the results (3.B.103), (3.B.104) and (3.B.105) in Lemma 3.B.14 and (3.B.112) and (3.B.113) in Lemma 3.B.16 hold if we replace the orthogonal projection matrices $\bar{\mathbf{M}}_Z$ and $\ddot{\mathbf{M}}_F$ with $\widetilde{\mathbf{M}}_i$ and $\ddot{\mathbf{M}}_i$ respectively. This is shown in Chapter 4. These results imply that under the conditions $\frac{p_T^2}{T} \rightarrow 0$ and $\frac{T}{\sqrt{N}} \rightarrow 0$ as $N, T, p_T \rightarrow \infty$ the left-hand side of (3.B.128) can be written as

$$\begin{aligned} & T \begin{pmatrix} \widehat{\widehat{\boldsymbol{\beta}}}_{xi} - \boldsymbol{\beta}_{xi} \\ \widehat{\widehat{\boldsymbol{\lambda}}}_i - \boldsymbol{\lambda}_i^* \end{pmatrix} \\ &= -\frac{1}{\alpha_{yi}} \begin{pmatrix} \frac{\mathbf{X}'_{i,-1} \ddot{\mathbf{M}}_i \mathbf{X}_{i,-1}}{T^2} + o_p(1) & \frac{\mathbf{X}'_{i,-1} \ddot{\mathbf{M}}_i \mathbf{F}_{-1} \mathbf{H}_1}{T^2} + o_p(1) \\ \frac{\mathbf{H}'_1 \mathbf{F}'_{-1} \ddot{\mathbf{M}}_i \mathbf{X}_{i,-1}}{T^2} + o_p(1) & \frac{\mathbf{H}'_1 \mathbf{F}'_{-1} \ddot{\mathbf{M}}_i \mathbf{F}_{-1} \mathbf{H}_1}{T^2} + o_p(1) \end{pmatrix}^{-1} \\ & \quad \times \begin{pmatrix} \frac{\mathbf{X}'_{i,-1} \ddot{\mathbf{M}}_i \boldsymbol{\varepsilon}_{y.xi}}{T} + o_p(1) \\ \frac{\mathbf{H}'_1 \mathbf{F}'_{-1} \ddot{\mathbf{M}}_i \boldsymbol{\varepsilon}_{y.xi}}{T} + o_p(1) \end{pmatrix}. \end{aligned} \quad (3.B.129)$$

Furthermore, under the conditions that $\frac{p_T^2}{T} \rightarrow 0$ as $T, p_T \rightarrow \infty$ and thanks to the preliminary results (3.B.86), (3.B.87), (3.B.88), (3.B.89) we have

$$\begin{aligned} \frac{\mathbf{X}'_{i,-1} \ddot{\mathbf{M}}_i \mathbf{X}_{i,-1}}{T^2} &= \frac{\mathbf{X}'_{i,-1} \mathbf{X}_{i,-1}}{T^2} + o_p(1), \\ \frac{\mathbf{X}'_{i,-1} \ddot{\mathbf{M}}_i \mathbf{F}_{-1} \mathbf{H}'_1}{T^2} &= \frac{\mathbf{X}'_{i,-1} \mathbf{F}_{-1} \mathbf{H}_1}{T^2} + o_p(1), \\ \frac{\mathbf{H}'_1 \mathbf{F}'_{-1} \ddot{\mathbf{M}}_i \mathbf{F}_{-1} \mathbf{H}_1}{T^2} &= \frac{\mathbf{H}'_1 \mathbf{F}'_{-1} \mathbf{F}_{-1} \mathbf{H}_1}{T^2} + o_p(1), \\ \frac{\mathbf{X}'_{i,-1} \ddot{\mathbf{M}}_i \boldsymbol{\varepsilon}_{y.xi}}{T} &= \frac{\mathbf{X}'_{i,-1} \boldsymbol{\varepsilon}_{y.xi}}{T} + o_p(1), \\ \frac{\mathbf{H}'_1 \mathbf{F}'_{-1} \ddot{\mathbf{M}}_i \boldsymbol{\varepsilon}_{y.xi}}{T} &= \frac{\mathbf{H}'_1 \mathbf{F}'_{-1} \boldsymbol{\varepsilon}_{y.xi}}{T} + o_p(1). \end{aligned}$$

Using these results we can rewrite (3.B.129) as

$$T \begin{pmatrix} \widehat{\widehat{\beta}}_{xi} - \beta_{xi} \\ \widehat{\widehat{\lambda}}_i - \lambda_i^* \end{pmatrix} = \frac{1}{\alpha_{yi}} \begin{pmatrix} \frac{\mathbf{X}'_{i,-1}\mathbf{X}_{i,-1}}{T^2} + o_p(1) & \frac{\mathbf{X}'_{i,-1}\mathbf{F}_{-1}\mathbf{H}_1}{T^2} + o_p(1) \\ \frac{\mathbf{H}'_1\mathbf{F}'_{-1}\mathbf{X}_{i,-1}}{T^2} + o_p(1) & \frac{\mathbf{H}'_1\mathbf{F}'_{-1}\mathbf{F}_{-1}\mathbf{H}_1}{T^2} + o_p(1) \end{pmatrix}^{-1} \times \begin{pmatrix} \frac{\mathbf{X}'_{i,-1}\varepsilon_{y,xi}}{T} + o_p(1) \\ \frac{\mathbf{H}'_1\mathbf{F}'_{-1}\varepsilon_{y,xi}}{T} + o_p(1) \end{pmatrix}. \quad (3.B.130)$$

After the simplification above, we reintroduce the estimator for the case in which α_{yi} is unobserved, which is denoted by $\widehat{\widehat{\beta}}_{xi}$ and $\widehat{\widehat{\lambda}}_i$. We have

$$\begin{aligned} & \begin{pmatrix} \widehat{\widehat{\beta}}_{xi} - \widehat{\beta}_{xi} \\ \widehat{\widehat{\lambda}}_i - \widehat{\lambda}_i \end{pmatrix} \\ &= \left(\frac{1}{\widehat{\widehat{\alpha}}_{yi}} - \frac{1}{\alpha_{yi}} \right) \begin{pmatrix} \frac{\mathbf{X}'_{i,-1}\mathbf{X}_{i,-1}}{T^2} + o_p(1) & \frac{\mathbf{X}'_{i,-1}\mathbf{F}_{-1}\mathbf{H}_1}{T^2} + o_p(1) \\ \frac{\mathbf{H}'_1\mathbf{F}'_{-1}\mathbf{X}_{i,-1}}{T^2} + o_p(1) & \frac{\mathbf{H}'_1\mathbf{F}'_{-1}\mathbf{F}_{-1}\mathbf{H}_1}{T^2} + o_p(1) \end{pmatrix}^{-1} \\ & \quad \times \begin{pmatrix} \frac{\mathbf{X}'_{i,-1}\varepsilon_{y,xi}}{T} + o_p(1) \\ \frac{\mathbf{H}'_1\mathbf{F}'_{-1}\varepsilon_{y,xi}}{T} + o_p(1) \end{pmatrix} \\ & \quad + \frac{1}{\widehat{\widehat{\alpha}}_{yi}} \begin{pmatrix} \frac{\mathbf{X}'_{i,-1}\mathbf{X}_{i,-1}}{T^2} + o_p(1) & \frac{\mathbf{X}'_{i,-1}\mathbf{F}_{-1}\mathbf{H}_1}{T^2} + o_p(1) \\ \frac{\mathbf{H}'_1\mathbf{F}'_{-1}\mathbf{X}_{i,-1}}{T^2} + o_p(1) & \frac{\mathbf{H}'_1\mathbf{F}'_{-1}\mathbf{F}_{-1}\mathbf{H}_1}{T^2} + o_p(1) \end{pmatrix}^{-1} \\ & \quad \times \begin{pmatrix} \frac{\mathbf{X}'_{i,-1}(\varepsilon_{y,xi}^* - \varepsilon_{y,xi})}{T} + o_p(1) \\ \frac{\mathbf{H}'_1\mathbf{F}'_{-1}(\varepsilon_{y,xi}^* - \varepsilon_{y,xi})}{T} + o_p(1) \end{pmatrix} \\ &= o_p(1), \end{aligned} \quad (3.B.131)$$

where the order result is obtained by using the consistency property of $\widehat{\widehat{\alpha}}_{yi}$, which is established in Theorem 3.3, where $\widehat{\widehat{\alpha}}_{yi}$ is the first element of the estimator $\widehat{\widehat{\boldsymbol{\Pi}}}_i$ (For details see Proof of Theorem 3.2.).

The result in (3.B.131) and the preliminary results (3.B.15), (3.B.16), (3.B.17), (3.B.18), (3.B.19) yield the below result on the convergence of the distribution of the estimator

$$T \begin{pmatrix} \widehat{\widehat{\beta}}_{xi} - \beta_{xi} \\ \widehat{\widehat{\lambda}}_i - \lambda_i^* \end{pmatrix} \xrightarrow{d} -\frac{1}{\alpha_{yi}} \left[\mathbf{H}_R \begin{pmatrix} \int_0^1 \mathbf{C}_{xzi}\mathbf{B}_{bi}(r)\mathbf{B}'_{bi}(r)dr\mathbf{C}'_{xzi} & \int_0^1 \mathbf{C}_{xzi}\mathbf{B}_{bi}(r)\mathbf{B}'_f(r)dr\mathbf{C}'_{ff} \\ \int_0^1 \mathbf{C}_{ff}\mathbf{B}_f(r)\mathbf{B}'_{bi}(r)dr\mathbf{C}'_{xzi} & \int_0^1 \mathbf{C}_{ff}\mathbf{B}_f(r)\mathbf{B}'_f(r)dr\mathbf{C}'_{ff} \end{pmatrix} \mathbf{H}'_R \right]^{-1}$$

$$\times \left[\mathbf{H}_R \begin{array}{c} \int_0^1 \mathbf{C}_{xzi} \mathbf{B}_{bi}(r) dB'_{y.xi}(r) \\ \int_0^1 \mathbf{C}_{ff} \mathbf{B}_f(r) dB'_{y.xi}(r) \end{array} \right], \quad (3.B.132)$$

where $\mathbf{H}_R = \begin{pmatrix} \mathbf{I} & 0 \\ 0 & \mathbf{H}_1 \end{pmatrix}$. ■

3.B.3 The Mean Group Estimator

Lemmas and proofs

To prove Theorems 3.5 and 3.6, we start with rewriting the estimator as

$$\begin{aligned}
 \widehat{\beta}_{xMG} - \beta_x &= \frac{1}{N} \sum_{i=1}^N \eta_{xi} - \frac{1}{N} \sum_{i=1}^N \frac{1}{\alpha_{yi}} \left(\frac{\mathbf{X}'_{i,-1} \overline{\mathbf{M}}_Q \widetilde{\mathbf{M}}_i \overline{\mathbf{M}}_Q \mathbf{X}_{i,-1}}{T^2} \right)^{-1} \\
 &\quad \times \left(\frac{\mathbf{X}'_{i,-1} \overline{\mathbf{M}}_Q \widetilde{\mathbf{M}}_i \overline{\mathbf{M}}_Q \tilde{\varepsilon}_{y.xi}^*}{T^2} \right) \\
 &\quad - \frac{1}{N} \sum_{i=1}^N \left(\frac{1}{\widehat{\alpha}_{yi}} - \frac{1}{\alpha_{yi}} \right) \left(\frac{\mathbf{X}'_{i,-1} \overline{\mathbf{M}}_Q \widetilde{\mathbf{M}}_i \overline{\mathbf{M}}_Q \mathbf{X}_{i,-1}}{T^2} \right)^{-1} \\
 &\quad \times \left(\frac{\mathbf{X}'_{i,-1} \overline{\mathbf{M}}_Q \widetilde{\mathbf{M}}_i \overline{\mathbf{M}}_Q \tilde{\varepsilon}_{y.xi}^*}{T^2} \right), \tag{3.B.133}
 \end{aligned}$$

where

$$\begin{aligned}
 \tilde{\varepsilon}_{y.xi}^* &= \varepsilon_{y.xi} + (\varepsilon_{y.xi}^* - \varepsilon_{y.xi}) + \mathbf{V}_{A,-1} \mathbf{H}_1^- \boldsymbol{\lambda}_i \alpha_{yi} + \mathbf{V}_{A,-1} \mathbf{H}_1^- \boldsymbol{\lambda}_i (\widehat{\alpha}_{yi} - \alpha_{yi}) \\
 &\quad - \mathbf{V}_{B,-1} \mathbf{H}_4^- (L) \mathbf{A}'_{y.fi} (L) + \overline{\Delta \mathbf{Z}}_{-p_T-1} \mathbf{A}^{**'}_{y.fi} (L). \tag{3.B.134}
 \end{aligned}$$

We provide the following lemmas regarding the orders of the processes appear in the equation above.

Lemma 3.B.17. *Let $\frac{p_T^2}{T} \rightarrow 0$, as $(N, T, p_T) \rightarrow \infty$, we have*

$$\frac{\mathbf{X}'_{i,-1} \overline{\mathbf{M}}_Q \widetilde{\mathbf{M}}_i \overline{\mathbf{M}}_Q \mathbf{X}_{i,-1}}{T^2} = \frac{\mathbf{C}_{xxi} \mathbf{S}'_{xi,-1} \mathbf{S}_{xi,-1} \mathbf{C}'_{xxi}}{T^2} + o_p(1). \tag{3.B.135}$$

Furthermore, we have

$$\frac{\mathbf{C}_{xxi} \mathbf{S}'_{xi,-1} \mathbf{S}_{xi,-1} \mathbf{C}'_{xxi}}{T^2} = O_p(1). \tag{3.B.136}$$

Proof of Lemma 3.B.17: To prove the lemma, we first consider $\widetilde{\mathbf{M}}_i = \mathbf{I} - \widetilde{\mathbf{G}}_i (\widetilde{\mathbf{G}}'_i \widetilde{\mathbf{G}}_i)^{-1} \widetilde{\mathbf{G}}'_i$, where $\widetilde{\mathbf{G}}_i = (\mathbf{G}_{yi}, \mathbf{G}_{xi}, \overline{\mathbf{G}}_z)$. We write $\widetilde{\mathbf{G}}_i = \ddot{\mathbf{G}}_i + (\widetilde{\mathbf{G}}_i - \ddot{\mathbf{G}}_i)$, where $\ddot{\mathbf{G}}_i = (\mathbf{G}_{yi}, \mathbf{G}_{xi}, \ddot{\mathbf{G}}_F)$ and the t^{th} row of $\ddot{\mathbf{G}}_F$ is defined in (3.B.83). The non-zero elements of $(\widetilde{\mathbf{G}}_i - \ddot{\mathbf{G}}_i)$ forms the $(T - p_T) \times ((m + 1)p_T)$ matrix $\mathbf{V}_{B,(p_T)}$ defined in (3.B.95). Furthermore we can write

$$\widetilde{\mathbf{G}}'_i \widetilde{\mathbf{G}}_i = \ddot{\mathbf{G}}'_i \ddot{\mathbf{G}}_i + \ddot{\mathbf{G}}'_i (\widetilde{\mathbf{G}}_i - \ddot{\mathbf{G}}_i) + (\widetilde{\mathbf{G}}_i - \ddot{\mathbf{G}}_i)' \ddot{\mathbf{G}}_i + (\widetilde{\mathbf{G}}_i - \ddot{\mathbf{G}}_i)' (\widetilde{\mathbf{G}}_i - \ddot{\mathbf{G}}_i).$$

Then by using the results in (3.B.88) and (3.B.91) in Lemma 3.B.10 we have $\frac{\tilde{\mathbf{G}}'_i \tilde{\mathbf{G}}_i}{T} = O_p(1) + O_p\left(\frac{1}{\sqrt{N}}\right)$. Using similar steps one can also show that $\frac{\tilde{\mathbf{G}}'_i \mathbf{X}_{i,-1}}{T} = O_p(1) + O_p\left(\frac{1}{\sqrt{N}}\right)$ by using (3.B.56) and (3.B.86) and $\frac{\tilde{\mathbf{G}}'_i \bar{\mathbf{Z}}_{-1}}{T} = O_p(1) + O_p\left(\frac{1}{\sqrt{N}}\right)$ by using (3.B.57) and (3.B.58). Then by writing

$$\begin{aligned} \frac{\mathbf{X}'_{i,-1} \bar{\mathbf{M}}_Q \tilde{\mathbf{M}}_i \bar{\mathbf{M}}_Q \mathbf{X}_{i,-1}}{T^2} &= \frac{\mathbf{X}'_{i,-1} \bar{\mathbf{M}}_Q \mathbf{X}_{i,-1}}{T^2} \\ &\quad - \frac{\mathbf{X}'_{i,-1} \bar{\mathbf{M}}_Q \tilde{\mathbf{G}}_i}{T^2} \left(\frac{\tilde{\mathbf{G}}'_i \tilde{\mathbf{G}}_i}{T} \right)^{-1} \frac{\tilde{\mathbf{G}}'_i \bar{\mathbf{M}}_Q \mathbf{X}_{i,-1}}{T^2}, \end{aligned}$$

and using the order results obtained above, we can easily show that the second term is $O_p(T^{-1})$. Now we analyze the second term on the right-hand side above, which can be rewritten as

$$\frac{\mathbf{X}'_{i,-1} \bar{\mathbf{M}}_Q \mathbf{X}_{i,-1}}{T^2} = \frac{\mathbf{X}'_{i,-1} \mathbf{M}_H \mathbf{X}_{i,-1}}{T^2} + \frac{\mathbf{X}'_{i,-1} (\bar{\mathbf{M}}_Q - \mathbf{M}_H) \mathbf{X}_{i,-1}}{T^2}. \quad (3.B.137)$$

For the second term above, by considering the definitions,

$$\bar{\mathbf{M}}_Q = \mathbf{I} - \bar{\mathbf{Z}}_{-1} (\bar{\mathbf{Z}}'_{-1} \bar{\mathbf{Z}}_{-1})^{-1} \bar{\mathbf{Z}}'_{-1}$$

and

$$\mathbf{M}_H = \mathbf{I} - \mathbf{F}_{-1} \mathbf{H}_1 (\mathbf{H}'_1 \mathbf{F}'_{-1} \mathbf{F}_{-1} \mathbf{H}_1)^+ \mathbf{H}'_1 \mathbf{F}'_{-1},$$
 we write

$$\begin{aligned} &\left\| \frac{\mathbf{X}'_{i,-1} (\bar{\mathbf{M}}_Q - \mathbf{M}_H) \mathbf{X}_{i,-1}}{T^2} \right\| \\ &\leq \left\| \frac{\mathbf{X}'_{i,-1} (\bar{\mathbf{Z}}_{-1} - \mathbf{F}_{-1} \mathbf{H}_1)}{T^2} \right\| \left\| \left(\frac{\bar{\mathbf{Z}}'_{-1} \bar{\mathbf{Z}}_{-1}}{T^2} \right)^{-1} \right\| \left\| \frac{(\bar{\mathbf{Z}}_{-1} - \mathbf{F}_{-1} \mathbf{H}_1)' \mathbf{X}_{i,-1}}{T^2} \right\| \\ &\quad + \left\| \frac{\mathbf{X}'_{i,-1} (\bar{\mathbf{Z}}_{-1} - \mathbf{F}_{-1} \mathbf{H}_1)}{T^2} \right\| \left\| \left(\frac{\bar{\mathbf{Z}}'_{-1} \bar{\mathbf{Z}}_{-1}}{T^2} \right)^{-1} \right\| \left\| \frac{\mathbf{H}'_1 \mathbf{F}'_{-1} \mathbf{X}_{i,-1}}{T^2} \right\| \\ &\quad + \left\| \frac{\mathbf{X}'_{i,-1} \mathbf{F}_{-1} \mathbf{H}_1}{T^2} \right\| \left\| \left(\frac{\bar{\mathbf{Z}}'_{-1} \bar{\mathbf{Z}}_{-1}}{T^2} \right)^{-1} \right\| \left\| \frac{(\bar{\mathbf{Z}}_{-1} - \mathbf{F}_{-1} \mathbf{H}_1)' \mathbf{X}_{i,-1}}{T^2} \right\| \\ &\quad + \left\| \frac{\mathbf{X}'_{i,-1} \mathbf{F}_{-1} \mathbf{H}_1}{T^2} \right\| \left\| \left[\left(\frac{\bar{\mathbf{Z}}'_{-1} \bar{\mathbf{Z}}_{-1}}{T^2} \right)^{-1} - \left(\frac{\mathbf{H}'_1 \mathbf{F}'_{-1} \mathbf{F}_{-1} \mathbf{H}_1}{T^2} \right)^+ \right] \right\| \\ &\quad \times \left\| \frac{\mathbf{H}'_1 \mathbf{F}'_{-1} \mathbf{X}_{i,-1}}{T^2} \right\| = O_p\left(\frac{1}{\sqrt{N}}\right), \end{aligned} \quad (3.B.138)$$

where the inequality is obtained using the Cauchy-Schwarz inequality and the triangle inequality and the order result is obtained by using (3.B.15), (3.B.17), (3.B.49), (3.B.51) and (3.B.52) considering the equation (3.B.41). This result implies that

$$\frac{\mathbf{X}'_{i,-1}\bar{\mathbf{M}}_Q\mathbf{X}_{i,-1}}{T^2} = \frac{\mathbf{X}'_{i,-1}\mathbf{M}_F\mathbf{X}_{i,-1}}{T^2} + O_p\left(\frac{1}{\sqrt{N}}\right),$$

where

$$\mathbf{M}_F = \mathbf{I} - \mathbf{F}_{-1}(\mathbf{F}'_{-1}\mathbf{F}_{-1})^{-1}\mathbf{F}'_{-1}, \quad (3.B.139)$$

where we note that under the rank condition imposed by Assumption 3.4(i) using familiar results on generalized inverses we have $\mathbf{M}_F = \mathbf{M}_H$.

As a next step, using the common trends representation for \mathbf{X}_i and \mathbf{F}

$$\mathbf{X}_{i,-1} = \mathbf{S}_{xi,-1}\mathbf{C}'_{xxi} + \mathbf{F}_{-1}\mathbf{C}_{ff}^{-1}\mathbf{C}'_{xfi} + \boldsymbol{\varepsilon}_{i,-1}^*\mathbf{C}_{xz^*i}^*(L), \quad (3.B.140)$$

where $\mathbf{C}_{xz^*i}^*(L) = (\mathbf{C}_{xyi}^*(L), \mathbf{C}_{xxi}^*(L), \mathbf{C}_{xfi}^*(L))$. We use this representation to write

$$\frac{\mathbf{X}'_{i,-1}\mathbf{M}_F\mathbf{X}_{i,-1}}{T^2} = \frac{\mathbf{C}_{xxi}\mathbf{S}'_{xi,-1}\mathbf{S}_{xi,-1}\mathbf{C}'_{xxi}}{T^2} + o_p(1). \quad (3.B.141)$$

The result obtained by noting that the last term in (3.B.140) is a covariance-stationary process and noting the orthogonality of $\mathbf{S}_{xi,-1}$ to \mathbf{F}_{-1} . This gives the required result. The second result in the lemma can be established using the standard results in the literature on non-stationary processes by noting that $\mathbf{S}_{xi,-1}$ is the partial sum of i.i.d. increments and \mathbf{C}_{xxi} is bounded. ■

Lemma 3.B.18. *Defining*

$$\mathbf{Q}_i^\# = \frac{\mathbf{C}_{xxi}\mathbf{S}'_{xi,-1}\mathbf{S}_{xi,-1}\mathbf{C}'_{xxi}}{T^2},$$

as $(N, T, p_T) \rightarrow \infty$ such that $\frac{p_T^2}{T} \rightarrow 0$, we have

$$\frac{1}{\sqrt{N}}\frac{1}{T}\sum_{i=1}^N\frac{1}{\alpha_{yi}}(\mathbf{Q}_i^\#)^{-1}\frac{\mathbf{X}'_{i,-1}\bar{\mathbf{M}}_Q\widetilde{\mathbf{M}}_i\bar{\mathbf{M}}_Q\boldsymbol{\varepsilon}_{y.xi}}{T} = O_p\left(\frac{1}{T}\right) \quad (3.B.142)$$

$$\frac{1}{\sqrt{N}}\frac{1}{T}\sum_{i=1}^N\frac{1}{\alpha_{yi}}(\mathbf{Q}_i^\#)^{-1}\frac{\mathbf{X}'_{i,-1}\bar{\mathbf{M}}_Q\widetilde{\mathbf{M}}_i\bar{\mathbf{M}}_Q(\boldsymbol{\varepsilon}_{y.xi}^* - \boldsymbol{\varepsilon}_{y.xi})}{T} = O_p\left(\frac{1}{T^{3/2}}\right), \quad (3.B.143)$$

$$\frac{1}{N} \sum_{i=1}^N (\mathbf{Q}_i^\#)^{-1} \frac{\sqrt{N} \mathbf{X}'_{i,-1} \bar{\mathbf{M}}_Q \tilde{\mathbf{M}}_i \bar{\mathbf{M}}_Q \mathbf{V}_{A,-1} \mathbf{H}_1^- \boldsymbol{\lambda}_i}{T^2} = O_p(1), \quad (3.B.144)$$

$$\frac{1}{N} \sum_{i=1}^N \frac{(\tilde{\alpha}_{yi} - \alpha_{yi})}{\alpha_{yi}} (\mathbf{Q}_i^\#)^{-1} \frac{\sqrt{N} \mathbf{X}'_{i,-1} \bar{\mathbf{M}}_Q \tilde{\mathbf{M}}_i \bar{\mathbf{M}}_Q \mathbf{V}_{A,-1} \mathbf{H}_1^- \boldsymbol{\lambda}_i}{T^2} = O_p\left(\frac{1}{\sqrt{T}}\right), \quad (3.B.145)$$

$$\frac{1}{N} \sum_{i=1}^N \frac{1}{\alpha_{yi}} (\mathbf{Q}_i^\#)^{-1} \frac{\sqrt{N} \mathbf{X}'_{i,-1} \bar{\mathbf{M}}_Q \tilde{\mathbf{M}}_i \bar{\mathbf{M}}_Q \mathbf{V}_{B,-1} \mathbf{H}_4^-(L) \mathbf{A}'_{y.fi}(L)}{T^2} = O_p\left(\frac{1}{T}\right), \quad (3.B.146)$$

$$\frac{1}{N} \sum_{i=1}^N \frac{1}{\alpha_{yi}} (\mathbf{Q}_i^\#)^{-1} \frac{\sqrt{N} \mathbf{X}'_{i,-1} \bar{\mathbf{M}}_Q \tilde{\mathbf{M}}_i \bar{\mathbf{M}}_Q \bar{\Delta} \bar{\mathbf{Z}}_{-p_T-1} \mathbf{A}_{y.fi}^{**'}(L)}{T^2} = O_p\left(\frac{\rho_A^{p_T} \sqrt{N}}{T}\right). \quad (3.B.147)$$

Proof of Lemma 3.B.18: To prove the first part of the lemma, we first write

$$\frac{\mathbf{X}'_{i,-1} \bar{\mathbf{M}}_Q \tilde{\mathbf{M}}_i \bar{\mathbf{M}}_Q \boldsymbol{\varepsilon}_{y.xi}}{T} = \frac{\mathbf{X}'_{i,-1} \bar{\mathbf{M}}_Q \boldsymbol{\varepsilon}_{y.xi}}{T} - \frac{\mathbf{X}'_{i,-1} \bar{\mathbf{M}}_Q \tilde{\mathbf{G}}_i}{T} \left(\frac{\tilde{\mathbf{G}}_i' \tilde{\mathbf{G}}_i}{T} \right)^{-1} \frac{\tilde{\mathbf{G}}_i' \bar{\mathbf{M}}_Q \boldsymbol{\varepsilon}_{y.xi}}{T}.$$

The first and the second parts of the second term can be established as $O_p(1)$ by using (3.B.15), (3.B.17), (3.B.56), (3.B.57), (3.B.86), (3.B.87), (3.B.88) and (3.B.91). The third part of the second term can be written as

$$\frac{\tilde{\mathbf{G}}_i' \bar{\mathbf{M}}_Q \boldsymbol{\varepsilon}_{y.xi}}{T} = \frac{\tilde{\mathbf{G}}_i' \boldsymbol{\varepsilon}_{y.xi}}{T} - \frac{\tilde{\mathbf{G}}_i' \bar{\mathbf{Z}}_{-1}}{T} \left(\frac{\bar{\mathbf{Z}}_{-1}' \bar{\mathbf{Z}}_{-1}}{T^2} \right)^{-1} \frac{\bar{\mathbf{Z}}_{-1}' \boldsymbol{\varepsilon}_{y.xi}}{T^2},$$

where the first term is $O_p(T^{-1/2})$ since $\boldsymbol{\varepsilon}_{y.xi}$ is the conditional error term, which is assumed to be i.i.d. across t . The first and second parts of the second term are $O_p(1)$, which can be established by using (3.B.15), (3.B.49), (3.B.58), (3.B.57), (3.B.87) and (3.B.90). The third part of the second term is $O_p(T^{-1})$, which can be established by using (3.B.19) and (3.B.53). Then we have

$$\frac{\mathbf{X}'_{i,-1} \bar{\mathbf{M}}_Q \tilde{\mathbf{M}}_i \bar{\mathbf{M}}_Q \boldsymbol{\varepsilon}_{y.xi}}{T} = \frac{\mathbf{X}'_{i,-1} \bar{\mathbf{M}}_Q \boldsymbol{\varepsilon}_{y.xi}}{T} + O_p\left(\frac{1}{\sqrt{T}}\right).$$

Furthermore we can write

$$\frac{\mathbf{X}'_{i,-1} \bar{\mathbf{M}}_Q \boldsymbol{\varepsilon}_{y.xi}}{T} = \frac{\mathbf{X}'_{i,-1} \mathbf{M}_H \boldsymbol{\varepsilon}_{y.xi}}{T} + \frac{\mathbf{X}'_{i,-1} (\bar{\mathbf{M}}_Q - \mathbf{M}_H) \boldsymbol{\varepsilon}_{y.xi}}{T}$$

$$= \frac{\mathbf{X}'_{i,-1} \mathbf{M}_H \boldsymbol{\varepsilon}_{y.xi}}{T} + O_p\left(\frac{1}{\sqrt{N}}\right),$$

where the second inequality can be obtained by using a similar method to the one in (3.B.138). Using these results allows us to write the left-hand side of (3.B.142) as

$$\frac{1}{\sqrt{N}} \frac{1}{T} \sum_{i=1}^N \frac{1}{\alpha_{yi}} \left(\frac{\mathbf{C}_{xxi} \mathbf{S}'_{xi,-1} \mathbf{S}_{xi,-1} \mathbf{C}'_{xxi}}{T^2} \right)^{-1} \frac{\mathbf{C}_{xxi} \mathbf{S}'_{xi,-1} \boldsymbol{\varepsilon}_{y.xi}}{T}.$$

To show that the expression above is $O_p\left(\frac{1}{\sqrt{T}}\right)$, we show that a CLT holds for

$$\frac{1}{\sqrt{N}} \sum_{i=1}^N \frac{1}{\alpha_{yi}} \left(\frac{\mathbf{C}_{xxi} \mathbf{S}'_{xi,-1} \mathbf{S}_{xi,-1} \mathbf{C}'_{xxi}}{T^2} \right)^{-1} \frac{\mathbf{C}_{xxi} \mathbf{S}'_{xi,-1} \boldsymbol{\varepsilon}_{y.xi}}{T}.$$

Thanks to the results derived in Appendix 3.B.1, we know that the expression has finite second order moments. Then the only remaining thing to show is that $\left(\frac{\mathbf{C}_{xxi} \mathbf{S}'_{xi,-1} \mathbf{S}_{xi,-1} \mathbf{C}'_{xxi}}{T^2} \right)^{-1} \frac{\mathbf{C}_{xxi} \mathbf{S}'_{xi,-1} \boldsymbol{\varepsilon}_{y.xi}}{T}$ is a martingale difference sequence, which is true such that

$$E \left[(\mathbf{Q}_i^\#)^{-1} \mathbf{C}_{xxi} \mathbf{s}_{xi,t-1} \boldsymbol{\varepsilon}_{y.xit} \mid (\mathbf{Q}_i^\#)^{-1} \mathbf{C}_{xxi} \mathbf{s}_{x,i-1,t^*-1} \boldsymbol{\varepsilon}_{y.x,i-1,t^*} \right] = 0,$$

for all t, t^* and for any ordering of i . (3.B.143) can be proved using similar steps and by noting that $\widehat{\alpha}_{yi}$ is a \sqrt{T} -consistent estimator of α_{yi} . The left-hand side of (3.B.144) is analyzed in detail in the next section so we omit the proof here. (3.B.145) can be proved by noting that $(\widehat{\alpha}_{yi} - \alpha_{yi})$ is $O_p(T^{-1/2})$, by the \sqrt{T} -consistency of the estimator $\widehat{\alpha}_{yi}$. (3.B.146) can be established in a similar way as the result in (3.B.142) by noting that $\mathbf{V}_{B,-1}$ is the approximation error and with $E\|\mathbf{v}_{B,t-1}\| = O_p\left(\frac{1}{\sqrt{N}}\right)$, which is a result established in Lemma 3.B.4. To establish the result in (3.B.147) we use similar steps as in the proof of (3.B.142) and also use the result in (3.B.76) in Lemma 3.B.9. ■

Proof of Theorem 3.5

To show that $\widehat{\beta}_{MG}$ is a consistent estimator for β defined in Assumption 3.3(ii) we consider (3.B.133). The estimator in (3.B.133) has three terms on the right-hand side. We know that by Assumption 3.3(ii) the first term converges to zero in probability. We analyze the second term in detail. For this reason we substitute the right-hand side of (3.B.134) for $\boldsymbol{\varepsilon}_{y.xi}^*$ in the estimator. This operation yields six terms. The results provided in Lemma 3.B.17 and 3.B.18 can be used to establish

the orders of the six terms. By multiplying the order results by $\frac{1}{\sqrt{N}}$ we show that

$$\widehat{\boldsymbol{\beta}}_{xMG} - \boldsymbol{\beta}_x = O_p\left(\frac{\rho_A^{pT}}{T}\right)$$

which yields the consistency result as $(N, T, p_T) \rightarrow \infty$. \blacksquare

Proof of Theorem 3.6

To find the asymptotic distribution of the mean group estimator we write

$$\begin{aligned} \sqrt{N}(\widehat{\boldsymbol{\beta}}_{xMG} - \boldsymbol{\beta}_x) &= \frac{1}{\sqrt{N}} \sum_{i=1}^N \eta_{xi} \\ &\quad - \frac{1}{\sqrt{N}} \sum_{i=1}^N \frac{1}{\alpha_{yi}} \left(\frac{\mathbf{X}'_{i,-1} \overline{\mathbf{M}}_Q \widetilde{\mathbf{M}}_i \overline{\mathbf{M}}_Q \mathbf{X}_{i,-1}}{T^2} \right)^{-1} \\ &\quad \times \left(\frac{\mathbf{X}'_{i,-1} \overline{\mathbf{M}}_Q \widetilde{\mathbf{M}}_i \overline{\mathbf{M}}_Q \tilde{\boldsymbol{\varepsilon}}_{y.xi}^*}{T^2} \right) \\ &\quad - \frac{1}{\sqrt{N}} \sum_{i=1}^N \left(\frac{1}{\widehat{\alpha}_{yi}} - \frac{1}{\alpha_{yi}} \right) \left(\frac{\mathbf{X}'_{i,-1} \overline{\mathbf{M}}_Q \widetilde{\mathbf{M}}_i \overline{\mathbf{M}}_Q \mathbf{X}_{i,-1}}{T^2} \right)^{-1} \\ &\quad \times \left(\frac{\mathbf{X}'_{i,-1} \overline{\mathbf{M}}_Q \widetilde{\mathbf{M}}_i \overline{\mathbf{M}}_Q \tilde{\boldsymbol{\varepsilon}}_{y.xi}^*}{T^2} \right), \end{aligned}$$

where

$$\begin{aligned} \tilde{\boldsymbol{\varepsilon}}_{y.xi}^* &= \boldsymbol{\varepsilon}_{y.xi} + (\boldsymbol{\varepsilon}_{y.xi}^* - \boldsymbol{\varepsilon}_{y.xi}) + \mathbf{V}_{A,-1} \mathbf{H}_1^- \boldsymbol{\lambda}_i \alpha_{yi} + \mathbf{V}_{A,-1} \mathbf{H}_1^- \boldsymbol{\lambda}_i (\widehat{\alpha}_{yi} - \alpha_{yi}) \\ &\quad - \mathbf{V}_{B,-1} \mathbf{H}_4^-(L) \mathbf{A}'_{y.fi}(L) + \overline{\Delta \mathbf{Z}}_{-p_T-1} \mathbf{A}^{**'}_{y.fi}(L). \end{aligned}$$

The third term is $o_p(1)$ given the consistency of $\widehat{\alpha}_{yi}$. The results in Lemma C.2 show that under the conditions of the theorem none of the terms, except (3.B.144), has an effect on the asymptotic distribution. Now, we analyse (3.B.144) in detail to see its effect on the asymptotic distribution of the mean group estimator. First we write

$$\begin{aligned} &\frac{1}{N} \sum_{i=1}^N (\mathbf{Q}_i^\#)^{-1} \frac{\sqrt{N} \mathbf{X}'_{i,-1} \overline{\mathbf{M}}_Q \mathbf{V}_{A,-1} \mathbf{H}_1^- \boldsymbol{\lambda}_i}{T^2} \\ &= \frac{1}{N} \sum_{i=1}^N (\mathbf{Q}_i^\#)^{-1} \frac{\sqrt{N} \mathbf{X}'_{i,-1} \mathbf{M}_H \mathbf{V}_{A,-1} \mathbf{H}_1^- \boldsymbol{\lambda}_i}{T^2} \end{aligned}$$

$$\begin{aligned}
 & + \frac{1}{N} \sum_{i=1}^N (\mathbf{Q}_i^\#)^{-1} \frac{\sqrt{N} \mathbf{X}'_{i,-1} (\overline{\mathbf{M}}_Q - \mathbf{M}_H) \mathbf{V}_{A,-1} \mathbf{H}_1^- \boldsymbol{\lambda}_i}{T^2}. \\
 \end{aligned} \tag{3.B.148}$$

We can show that the second term is $o_p(1)$. Now we consider the first term of the equation above. Using (3.B.140) and Assumption 3.3(iv), we have

$$\begin{aligned}
 & \frac{1}{N} \sum_{i=1}^N (\mathbf{Q}_i^\#)^{-1} \frac{\sqrt{N} \mathbf{X}'_{i,-1} \mathbf{M}_H \mathbf{V}_{A,-1} \mathbf{H}_1^- \boldsymbol{\lambda}_i}{T^2} \\
 & = \frac{1}{N} \sum_{i=1}^N (\mathbf{Q}_i^\#)^{-1} \frac{\sqrt{N} \mathbf{C}_{xxi} \mathbf{S}'_{xi,-1} \mathbf{V}_{A,-1} \mathbf{H}_1^- \bar{\boldsymbol{\lambda}}}{T^2} \\
 & \quad + \frac{1}{N} \sum_{i=1}^N (\mathbf{Q}_i^\#)^{-1} \frac{\sqrt{N} \mathbf{C}_{xxi} \mathbf{S}'_{xi,-1} \mathbf{V}_{A,-1} \mathbf{H}_1^- (\boldsymbol{\eta}_{fi} - \bar{\boldsymbol{\eta}}_f)}{T^2}.
 \end{aligned}$$

We can write the first term as

$$\begin{aligned}
 & \frac{1}{N^2} \sum_{i=1}^N \sum_{j=1}^N (\mathbf{Q}_i^\#)^{-1} \frac{\sqrt{N} \mathbf{C}_{xxi} \mathbf{S}'_{xi,-1} \mathbf{V}_{A,-1} \mathbf{H}_1^- \boldsymbol{\lambda}_j}{T^2} \\
 & = \frac{1}{N^2} \sum_{i=1}^N (\mathbf{Q}_i^\#)^{-1} \frac{\sqrt{N} \mathbf{C}_{xxi} \mathbf{S}'_{xi,-1} \mathbf{V}_{A,-1} \mathbf{H}_1^- \boldsymbol{\lambda}_i}{T^2} \\
 & \quad + \frac{1}{N^2} \sum_{i=1}^N \sum_{j \neq i}^N (\mathbf{Q}_i^\#)^{-1} \frac{\sqrt{N} \mathbf{C}_{xxi} \mathbf{S}'_{xi,-1} \mathbf{V}_{A,-1} \mathbf{H}_1^- \boldsymbol{\lambda}_j}{T^2}.
 \end{aligned}$$

Noting that, from the results of the previous section, $\frac{\mathbf{X}'_{i,-1} \mathbf{M}_H \mathbf{V}_{A,-1}}{T^2} = O_p\left(\frac{1}{\sqrt{N}}\right)$, then the first term is $O_p(N^{-1})$ and the second term is also $O_p(N^{-1})$. We need to analyze the first term on the right-hand side of (3.B.148). We write

$$\frac{1}{\sqrt{N}} \sum_{i=1}^N (\mathbf{Q}_i^\#)^{-1} \frac{\mathbf{C}_{xxi} \mathbf{S}'_{xi,-1} \mathbf{V}_{A,-1} \mathbf{H}_1^- (\boldsymbol{\eta}_{fi} - \bar{\boldsymbol{\eta}}_f)}{T^2}. \tag{3.B.149}$$

Note that the definition of $\mathbf{V}_{A,-1}$ is as follows:

$$\mathbf{V}_{A,-1} = \frac{1}{N} \sum_{i=1}^N \mathbf{S}_{xi,-1} \mathbf{C}'_{zxi} + \frac{1}{N} \sum_{i=1}^N \boldsymbol{\varepsilon}_{zi,-1} \mathbf{C}_{vai}^{*'}(L).$$

Using this in (3.B.149) we obtain

$$\begin{aligned}
 & \frac{1}{\sqrt{N}} \sum_{i=1}^N (\mathbf{Q}_i^\#)^{-1} \frac{\mathbf{C}_{xxi} \mathbf{S}'_{xi,-1} \mathbf{V}_{A,-1} \mathbf{H}_1^- (\boldsymbol{\eta}_{fi} - \bar{\boldsymbol{\eta}}_f)}{T^2} \\
 = & \frac{1}{N^{3/2}} \sum_{i=1}^N \sum_{j=1}^N (\mathbf{Q}_i^\#)^{-1} \frac{\mathbf{C}_{xxi} \mathbf{S}'_{xi,-1} \mathbf{S}_{xj,-1} \mathbf{C}'_{zzj} \mathbf{H}_1^- (\boldsymbol{\eta}_{fi} - \bar{\boldsymbol{\eta}}_f)}{T^2} + o_p(1) \\
 = & \frac{1}{\sqrt{N}} \sum_{i=1}^N \sum_{j \neq i}^N (\mathbf{Q}_i^\#)^{-1} \frac{\mathbf{C}_{xxi} \mathbf{S}'_{xi,-1} \mathbf{S}_{xj,-1} \mathbf{C}'_{zzj} \mathbf{H}_1^- (\boldsymbol{\eta}_{fi} - \bar{\boldsymbol{\eta}}_f)}{T^2} \\
 & + \frac{1}{\sqrt{N}} \sum_{i=1}^N (\mathbf{Q}_i^\#)^{-1} \frac{\mathbf{C}_{xxi} \mathbf{S}'_{xi,-1} \mathbf{S}_{xi,-1} \mathbf{C}'_{zxi} \mathbf{H}_1^- (\boldsymbol{\eta}_{fi} - \bar{\boldsymbol{\eta}}_f)}{T^2} + o_p(1),
 \end{aligned}$$

where the second term is clearly $o_p(1)$. For the first term we define

$$\bar{\mathbf{S}}_{x,-i,-1} = \frac{1}{N} \sum_{j \neq i}^N \mathbf{S}_{xj,-1}, \quad (3.B.150)$$

$$\bar{\mathbf{C}}_{zx,-i} = \frac{1}{N} \sum_{j \neq i}^N \mathbf{C}_{zxj}. \quad (3.B.151)$$

These definitions allows us to write

$$\frac{1}{N} \sum_{j \neq i}^N \mathbf{C}_{zxj} \mathbf{S}_{xj} = \bar{\mathbf{C}}_{zx,-i} \bar{\mathbf{S}}_{x,-i,-1} + \frac{1}{N} \sum_{j \neq i}^N (\mathbf{C}_{zxj} - \bar{\mathbf{C}}_{zx,-i}) \mathbf{S}_{xj}.$$

Then we have

$$\begin{aligned}
 & \frac{1}{\sqrt{N}} \sum_{i=1}^N (\mathbf{Q}_i^\#)^{-1} \frac{\mathbf{C}_{xxi} \mathbf{S}'_{xi,-1} \mathbf{V}_{A,-1} \mathbf{H}_1^- (\boldsymbol{\eta}_{fi} - \bar{\boldsymbol{\eta}}_f)}{T^2} \\
 = & \frac{1}{\sqrt{N}} \sum_{i=1}^N (\mathbf{Q}_i^\#)^{-1} \frac{\mathbf{C}_{xxi} \mathbf{S}'_{xi,-1} \bar{\mathbf{S}}_{x,-i,-1} \bar{\mathbf{C}}'_{zx,-i} \mathbf{H}_1^- (\boldsymbol{\eta}_{fi} - \bar{\boldsymbol{\eta}}_f)}{T^2} + o_p(1).
 \end{aligned}$$

Now we define

$$\mathbf{Q}_i^{\#\#} = (\mathbf{Q}_i^\#)^{-1} \frac{\mathbf{C}_{xxi} \mathbf{S}'_{xi,-1} \bar{\mathbf{S}}_{x,-i,-1} \bar{\mathbf{C}}'_{zx,-i} \mathbf{H}_1^-}{T^2}.$$

The following lines follows from the proof of Theorem 1 in Kapetanios et al. (2011). We write

$$\frac{1}{\sqrt{N}} \sum_{i=1}^N \mathbf{Q}_i^{\#\#} (\boldsymbol{\eta}_{fi} - \bar{\boldsymbol{\eta}}_f) = \frac{1}{\sqrt{N}} \sum_{i=1}^N (\mathbf{Q}_i^{\#\#} - \bar{\mathbf{Q}}_i^{\#\#}) \boldsymbol{\eta}_{fi}. \quad (3.B.152)$$

Noting that $\boldsymbol{\eta}_{fi}$ is i.i.d. with zero mean and finite variance Ω_{η_f} and independent of all other stochastic processes in $\mathbf{Q}_i^{\#\#}$, by Assumption 3.3(iv) and $(\mathbf{Q}_i^{\#\#} - \bar{\mathbf{Q}}_i^{\#\#})$ is a triangular array of random variables, we can use a central limit theorem for martingale difference triangular arrays (see for instance Theorem 24.3 of Davidson (1994)), since $(\mathbf{Q}_i^{\#\#} - \bar{\mathbf{Q}}_i^{\#\#}) \boldsymbol{\eta}_{fi}$ is a martingale difference triangular array. Then the right-hand side of (3.B.152) converges to a normal distribution, provided that the variance of $(\mathbf{Q}_i^{\#\#} - \bar{\mathbf{Q}}_i^{\#\#}) \boldsymbol{\eta}_{fi}$ is finite. For this it is sufficient to show that $\|(\mathbf{Q}_i^{\#\#} - \bar{\mathbf{Q}}_i^{\#\#})\|$ has finite second order moments. It is straightforward to see that the second order moments of the norm in question exists if we write

$$\mathbf{Q}_i^{\#\#} = \left(\frac{\mathbf{C}_{xxi} \mathbf{S}'_{xi,-1} \mathbf{S}_{xi,-1} \mathbf{C}'_{xxi}}{T^2} \right)^{-1} \left(\frac{\mathbf{C}_{xxi} \mathbf{S}'_{xi,-1} \bar{\mathbf{S}}_{x,-i,-1} \bar{\mathbf{C}}_{zx,-i}' \mathbf{H}_1^-}{T^2} \right),$$

since $\mathbf{S}_{xi,-1}$ is the partial sum of i.i.d. increments over n and t and \mathbf{C}_{xxi} is a full rank matrix. Then, for the variance of the mean group estimator we have that

$$\boldsymbol{\Sigma}_{MG} = \boldsymbol{\Omega}_{\eta_x} + \bar{\boldsymbol{\Lambda}}_q, \quad (3.B.153)$$

where

$$\bar{\boldsymbol{\Lambda}}_q = \lim_{N,T,p_T \rightarrow \infty} \frac{1}{N} \sum_{i=1}^N \text{Var} \left[\mathbf{Q}_i^{\#\#} \boldsymbol{\eta}_{fi} \right].$$

■

Chapter 4

Testing Weak Exogeneity in Cross-Sectionally Dependent Cointegrated Panels

4.1 Introduction

Consider a panel data regression involving two sets of non-stationary variables, \mathbf{y}_{it} and \mathbf{x}_{it} , where $i = 1, \dots, N$ and $t = 1, \dots, T$ index the cross-sectional and time series dimensions, respectively. The analysis of such regressions has been a growing field of econometric research in recent years, see for example Breitung and Pesaran (2008) for a recent overview. In particular, in many economic applications it is an important question whether the relationship is cointegrated or whether it is spurious.

Kao (1999) and Pedroni (1999, 2004) were among the first to propose residual-based tests for the null hypothesis of no cointegration in cross-sectionally independent panels. However, it has since then become clear that these tests do not work in general, as cross-section dependence is likely to be the rule rather than the exception. In fact, as Gengenbach et al. (2006) show the presence of unattended cross-section dependence in the form of non-stationary common factors can actually cause the test statistics to diverge as N and T grows. As a response to this, they propose to estimate separately the common and idiosyncratic components of \mathbf{x}_{it}

and y_{it} using the principal components method of Bai and Ng (2004), 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 principal components method to \mathbf{x}_{it} and y_{it} directly, they apply it to the residuals of a first-stage regression of y_{it} onto \mathbf{x}_{it} . Cointegration requires that both the common and idiosyncratic components of the residuals are stationary. However, while 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 so called ‘common factor’ critique of Kremers et al. (1992), leading to tests with potentially 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 (2009).

Larsson et al. (2001), and Larsson and Lyhagen (2007) alleviate these problems by developing rank tests based on a full-blown panel vector error correction model (ECM). However, for this approach to work properly, one usually requires T being substantially larger than N , a condition that is rarely fulfilled in practice, except in some macro-panels with few entities but a long observation period. Being based on numerical maximization of the likelihood it is also computationally demanding, which is likely to lead to poor small-sample performance, see Wagner and Hlouskova (2010) for a discussion and for some confirmatory simulation results.

Gengenbach et al. (2014) propose two tests for the null hypothesis of no cointegration by inferring whether the error correction term in a conditional ECM is equal to zero. They show that under the null of no error correction the asymptotic distributions of the tests are free of nuisance parameters, depending only on the number of unit roots in the system, a result that holds regardless of whether the factors are treated as known or estimated using the cross-sectional average of the observed data. However, the assumption of weakly exogenous regressors is crucial for the correct interpretation of the test outcome. In a pure time series context, for example Johansen (1992) and Urbain (1992) address the conditions for weak exogeneity in cointegrated system. The crucial condition is that the cointegrating relations do not contain any information for the conditioning variables. According to this condition the absence of the cointegrating relations in the models of the conditioning variables is sufficient to claim that the conditioning variables are weakly exogenous for the long-run parameters.

In this chapter, we consider the problem of testing the validity of the weak exogeneity assumption using panel extensions of the Lagrange multiplier tests proposed by Boswijk and Urbain (1997). Extending Boswijk and Urbain (1997) to a panel case is possible if the possible cross-sectional dependence is properly accounted for. In our setting, the cross-sectional dependence existing in the panel can be characterized by a common factor model. We show that using the common correlated effects method that is developed by Pesaran (2006) for stationary panels and extended to a non-stationary set-up by Kapetanios et al. (2011) is sufficient to account for the cross-sectional dependence in the panel properly.

It is important to mention the related, albeit different, work of Trapani (2014) that proposes a test for the null hypothesis that the long run covariance between the regressors and the error term are different from zero in a static cointegrated panel set-up. In contrast to this work, we address the issue of exogeneity in a parametric dynamic panel ECM model. Related independent work to ours is also that of Moral-Benito and Servén (2013), but their results do not easily carry over to cross-sectionally dependent panels.

The remainder of this chapter is organized as follows. Section 4.2 presents the model of interest. Sections 4.3 and 4.4 then present the tests for weak exogeneity and its asymptotic properties. Section 4.5 presents a Monte Carlo study of the small-sample properties of the tests, while Section 4.6 concludes.

4.2 Model and Assumptions

We consider a panel where we observe a $(1 + m)$ -dimensional $I(1)$ vector $\mathbf{z}_{it} = (y_{it}, \mathbf{x}'_{it})'$, which is comprised of a scalar variable y_{it} and an m -vector \mathbf{x}_{it} of idiosyncratic variables, for individuals $i = 1, \dots, N$, over time periods $t = -p, \dots, T$. Furthermore, we have a k -vector of $I(1)$ unobserved common factors \mathbf{f}_t . We assume that the data generating process of $\mathbf{z}_{it}^* = ((y_{it}, \mathbf{x}'_{it}, \mathbf{f}'_t)'$ can be written as a VECM for each cross-section unit i as

$$\Delta \mathbf{z}_{it}^* = \mathbf{A}_i^* \mathbf{z}_{i,t-1}^* + \sum_{\ell=1}^{p-1} \mathbf{A}_{i\ell} \Delta \mathbf{z}_{i,t-\ell}^* + \boldsymbol{\varepsilon}_{it}^*. \quad (4.1)$$

This DGP is similar to the one assumed in Chapter 3. In the following lines we present the assumptions of our set-up.

Assumption 4.1.

- (i) The vector of errors, $\boldsymbol{\varepsilon}_{it}^* = (\varepsilon_{yit}, \boldsymbol{\varepsilon}'_{xit}, \boldsymbol{\varepsilon}'_{ft})'$ is i.i.d across t with zero mean and finite variance,

$$\boldsymbol{\Sigma}_i_{(1+m+k) \times (1+m+k)} = \begin{pmatrix} \sigma_{yyi} & \boldsymbol{\Sigma}_{yxi} & 0 \\ \boldsymbol{\Sigma}_{xyi} & \boldsymbol{\Sigma}_{xxi} & 0 \\ 0 & 0 & \boldsymbol{\Sigma}_{ff} \end{pmatrix}, \quad (4.2)$$

and has finite fourth order moment. Furthermore ε_{yit} and $\boldsymbol{\varepsilon}_{xit}$ are i.i.d across i .

- (ii) The roots of the determinant of characteristic function,

$$\mathbf{F}(z) = (1 - z)\mathbf{I}_{(1+m+k)} - \mathbf{A}_i^* z - \sum_{\ell=1}^{p-1} \mathbf{A}_{i\ell}(1 - z)z^\ell$$

fall outside or on the unit circle for all $i = 1, \dots, N$.

- (iii) We assume that cointegration holds for each individual subsystem with common cointegration rank, $r_i = r = 1$ for $i = 1, \dots, N$; such that for all $i = 1, \dots, N$, \mathbf{A}_i^* can be expressed in terms of two vectors of rank 1:

$$\mathbf{A}_i^* = \boldsymbol{\alpha}_i \boldsymbol{\beta}'_i,$$

where $\boldsymbol{\alpha}_i$ and $\boldsymbol{\beta}_i$ are $(1 + m + k)$ -vectors.

- (iv) We normalize $\boldsymbol{\beta}_i$ such that the first element in $\boldsymbol{\beta}_i$ is equal to 1.

Assumption 4.2.

- (i) The coefficient of the short-run dynamics for the ℓ^{th} lag order, $\mathbf{A}_{i\ell}$, can be written in a block specific form as

$$\mathbf{A}_{i\ell} = \begin{pmatrix} A_{yyi\ell} & \mathbf{A}_{yxil} & \mathbf{A}_{yfi\ell} \\ \mathbf{A}_{xyi\ell} & \mathbf{A}_{xxi\ell} & \mathbf{A}_{xfi\ell} \\ 0 & 0 & \mathbf{A}_{ffi\ell} \end{pmatrix}.$$

- (ii) The latent factors are strongly exogenous for the long-run parameters, so that, in addition to (i) we assume that $\boldsymbol{\alpha}_f = 0$.

Assumption 4.3.

- (i) We assume that the error-correction coefficient α_{zi} are random, i.i.d. across i , independent of all other model parameters for all i and characterized by

$$\alpha_{zi} = \alpha_z + \eta_{zi}, \quad \eta_{zi} \sim IID(0, \omega_{\eta z}), \quad \omega_{\eta z} > 0 \quad \text{for } i = 1, 2, \dots, N,$$

where $\omega_{\eta z}$ is the variance of the random deviations and $\|\omega_{\eta z}\| < K$, for some constant K . The random deviations are distributed independently of the model parameters and variables. For each i , the support of the distribution of α_{zi} lies in the stable range.

- (ii) We assume that the long-run parameters for the idiosyncratic variables β_{xi} are random, i.i.d. across i , independent of all other model parameters for all i and characterized by

$$\beta_{xi} = \beta_x + \eta_{xi}, \quad \eta_{xi} \sim IID(0, \Omega_{\eta x}), \quad \Omega_{\eta x} > 0 \quad \text{for } i = 1, 2, \dots, N,$$

where $\Omega_{\eta x}$ is the variance of the random deviations and the random deviations are distributed independently of the model parameters and variables.

- (iii) The factor loadings in the long-run are random, i.i.d. across i and characterized by

$$\lambda_i = \lambda + \eta_{fi}, \quad \eta_{fi} \sim IID(0, \Omega_{\eta f}), \quad \Omega_{\eta f} > 0 \quad \text{for } i = 1, 2, \dots, N;$$

where $\lambda \neq 0$, $\Omega_{\eta f}$ is the variance of the random deviations and the random deviations are distributed independently of the model parameters and variables.

- (iv) For all $i = 1, 2, \dots, N$; $\mathbf{A}_{zzi\ell}$ and \mathbf{A}_{zfil} are random coefficients, distributed uniformly bounded over i and have bounded variances for all $\ell = 1, \dots, p$; where

$$\mathbf{A}_{zzi\ell} = \begin{pmatrix} A_{yyi\ell} & \mathbf{A}_{yxil} \\ \mathbf{A}_{xyi\ell} & \mathbf{A}_{xxi\ell} \end{pmatrix},$$

$$\text{and where } \mathbf{A}_{zfil} = \begin{pmatrix} \mathbf{A}_{yfil} \\ \mathbf{A}_{xfil} \end{pmatrix}.$$

(v) All factors are supposed to be strong - see Chudik et al. (2011). In particular, we assume that, such that for each cross-section unit i ,

$$\lim_{N \rightarrow \infty} N^{-1} \sum_{i=1}^N |\lambda_i| = K > 0, \quad \lim_{N \rightarrow \infty} N^{-1} \sum_{i=1}^N |\mathbf{A}_{zfil}| = \mathbf{K} > 0.$$

for $\ell = 1, \dots, p-1$ and where $\mathbf{A}_{zfil} = \begin{pmatrix} \mathbf{A}_{yfil} \\ \mathbf{A}_{xfil} \end{pmatrix}$.

Remark 4.2.1. In Assumption 4.1(i) the errors are assumed to be i.i.d. across t . This assumption restricts the errors to be serially uncorrelated. Given the dynamic nature of our modelling approach, serially uncorrelated errors can be obtained by increasing the lag order for the first differences.

Remark 4.2.2. Assumption 4.1(i) restricts the correlation between ε_{zit} and ε_{ft} to zero. This implies that the long-run and the short-run effects of cross-sectional dependence are fully parameterized by the inclusion of the latent factors both in levels and in first-differences. For a further discussion on this issue, see Remark 3.3 in Chapter 3. Bai et al. (2009) suggests when ε_{zit} is cross-sectionally dependent, the problem can be solved by including $I(0)$ common factors besides the $I(1)$ factors. In our set-up the first differences of the factors are already present in the model, which makes it legitimate to assume cross-sectional independence for ε_{zit} .

Remark 4.2.3. The factors are assumed to be strong based on the definitions provided in Chudik et al. (2011), see Assumption 4.3(v). As it is also mentioned in Gengenbach et al. (2014), this is due to the conceptual difficulty arises in the interpretation of the non-stationary weak factors, that enter into the cointegrating relation. Chudik et al. (2011) suggests that in the $I(0)$ factors case, under certain conditions, the weak factors can be omitted in the estimation. But in the $I(1)$ factors case, the weak factors will have an accumulated effect on y_{it} and on \mathbf{x}_{it} and the order of the approximation error will depend on the exponent of the cross-sectional dependence. See Bailey et al. (2012) for details, which assume that N^α of the N factor loadings are individually important, where α is defined as the exponent of the cross-sectional dependence.

Remark 4.2.4. Similarly to Chapter 3 we ignore any kinds of deterministic trends and observed common effect. See Remark 3.7 of Chapter 3 for details.

Remark 4.2.5. Assumptions 4.1(iv) and 4.2(ii) allows us to write

$$\mathbf{A}_i^* = \begin{pmatrix} \alpha_{yi} \\ \boldsymbol{\alpha}_{xi} \\ 0 \end{pmatrix} \begin{pmatrix} 1 & -\boldsymbol{\beta}'_{xi} & -\boldsymbol{\lambda}'_i \end{pmatrix},$$

where $\boldsymbol{\beta}_{xi}$ and $\boldsymbol{\lambda}_i$ are the coefficient vectors in the single long-run relation and α_{yi} and $\boldsymbol{\alpha}_{xi}$ are the coefficients of the error correction term in the models for Δy_{it} and $\Delta \mathbf{x}_{it}$ respectively.

Remark 4.2.6. Notice that the random coefficient specification of the heterogeneity made in Assumption 4.3 (i)-(iii) are strictly speaking not necessary. These could be replaced by assuming alternatively heterogeneous but non-random coefficient.

Using the assumptions we stated in this section, we can write the model in (4.1) separately for the scalar Δy_{it} , for the $m \times 1$ vector of idiosyncratic variables and for $k \times 1$ vector of unobserved factors, as

$$\begin{aligned} \Delta y_{it} &= \alpha_{yi}(y_{i,t-1} - \boldsymbol{\beta}'_{xi}\mathbf{x}_{i,t-1} - \boldsymbol{\lambda}'_i\mathbf{f}_{t-1}) + \sum_{\ell=1}^{p-1} A_{yyi\ell}\Delta y_{i,t-\ell} + \sum_{\ell=1}^{p-1} \mathbf{A}_{yxi\ell}\Delta \mathbf{x}_{i,t-\ell} \\ &\quad + \sum_{\ell=1}^{p-1} \mathbf{A}_{yfi\ell}\Delta \mathbf{f}_{t-\ell} + \varepsilon_{yit}, \end{aligned} \quad (4.3)$$

$$\begin{aligned} \Delta \mathbf{x}_{it} &= \boldsymbol{\alpha}_{xi}(y_{i,t-1} - \boldsymbol{\beta}'_{xi}\mathbf{x}_{i,t-1} - \boldsymbol{\lambda}'_i\mathbf{f}_{t-1}) + \sum_{\ell=1}^{p-1} \mathbf{A}_{xyi\ell}\Delta y_{i,t-\ell} + \sum_{\ell=1}^{p-1} \mathbf{A}_{xxi\ell}\Delta \mathbf{x}_{i,t-\ell} \\ &\quad + \sum_{\ell=1}^{p-1} \mathbf{A}_{xfi\ell}\Delta \mathbf{f}_{t-\ell} + \boldsymbol{\varepsilon}_{xit}, \end{aligned} \quad (4.4)$$

$$\Delta \mathbf{f}_t = \sum_{\ell=1}^{p-1} \mathbf{A}_{ff\ell}\Delta \mathbf{f}_{t-\ell} + \boldsymbol{\varepsilon}_{ft}. \quad (4.5)$$

4.2.1 Conditional Model

Given our purposes, we consider the conditional model. In this section we write the conditional model without imposing any further restrictions to the set-up we present in the previous section. Using (4.3)-(4.4) and given the block specific elements of $\boldsymbol{\Sigma}_i$, by defining $\boldsymbol{\kappa}'_{xi} = \boldsymbol{\Sigma}_{yxi}\boldsymbol{\Sigma}_{xxi}^{-1}$, the conditional parameters can be

written as

$$\begin{aligned}
 \alpha_{y.xi} &= \alpha_{yi} - \boldsymbol{\kappa}'_{xi} \boldsymbol{\alpha}_{xi}, \\
 \mathbf{A}_{y.xi0} &= \boldsymbol{\kappa}'_{xi}, \\
 A_{y.yi\ell} &= A_{yyi\ell} - \boldsymbol{\kappa}'_{xi} \mathbf{A}_{xyi\ell}, \quad \text{for } \ell = 1, \dots, p-1, \\
 \mathbf{A}_{y.xi\ell} &= \mathbf{A}_{yxi\ell} - \boldsymbol{\kappa}'_{xi} \mathbf{A}_{xxi\ell}, \quad \text{for } \ell = 1, \dots, p-1, \\
 \mathbf{A}_{y.fil} &= \mathbf{A}_{yfil} - \boldsymbol{\kappa}'_{xi} \mathbf{A}_{xfil}, \quad \text{for } \ell = 1, \dots, p-1.
 \end{aligned}$$

Using these definitions, the conditional model for $\Delta \mathbf{y}_{it}$ given current and past $\Delta \mathbf{x}_{it}$ s can be written as

$$\begin{aligned}
 \Delta y_{it} &= \alpha_{y.xi}(y_{i,t-1} - \boldsymbol{\beta}'_{xi} \mathbf{x}_{i,t-1} - \boldsymbol{\lambda}'_i \mathbf{f}_{t-1}) + \sum_{\ell=1}^{p-1} A_{y.yi\ell} \Delta y_{i,t-\ell} \\
 &\quad + \sum_{\ell=0}^{p-1} \mathbf{A}_{y.xi\ell} \Delta \mathbf{x}_{i,t-\ell} + \sum_{\ell=1}^{p-1} \mathbf{A}_{y.fil} \Delta \mathbf{f}_{t-\ell} + \varepsilon_{y.xit}, \quad (4.6)
 \end{aligned}$$

where $\varepsilon_{y.xit} = \varepsilon_{yit} - \boldsymbol{\kappa}'_{xi} \boldsymbol{\varepsilon}_{xit}$. Then,

$$\boldsymbol{\varepsilon}_{z.xit} = \begin{pmatrix} \varepsilon_{y.xit} \\ \boldsymbol{\varepsilon}_{xit} \\ \boldsymbol{\varepsilon}_{ft} \end{pmatrix} \quad (4.7)$$

has a covariance matrix given by

$$\boldsymbol{\Sigma}_{(1+m+k) \times (1+m+k)}^* = \begin{pmatrix} \sigma_{y.xi} & 0_{1 \times m} & 0_{1 \times k} \\ 0_{m \times 1} & \boldsymbol{\Sigma}_{xxi} & 0_{m \times k} \\ 0_{k \times 1} & 0_{k \times m} & \boldsymbol{\Sigma}_{ff} \end{pmatrix}, \quad (4.8)$$

where $\sigma_{y.xi} = \sigma_{yyi} - \boldsymbol{\Sigma}_{yxi} \boldsymbol{\Sigma}_{xxi}^{-1} \boldsymbol{\Sigma}_{xyi}$. Note the difference in the covariance structure between (4.2) and (4.8).

Remark 4.2.7. The conditioning in (4.6) is only on $\Delta \mathbf{x}_{it}$ and its past values, as the common factors do not enter contemporaneously in (4.3) and ε_{yit} and $\boldsymbol{\varepsilon}_{ft}$ are independent by assumption. In practice we would condition on $\Delta \mathbf{f}_t$ as well, and it is straightforward to extend the model in such a way, without qualitatively altering the results of the subsequent sections.

4.3 Weak Exogeneity

In the previous section we condition the scalar endogenous variable Δy_{it} on the m -vector of idiosyncratic variables $\Delta \mathbf{x}_{it}$ and their past values. This yields a single equation conditional error correction model for Δy_{it} that is stated in (4.6). Conducting valid inference without loss of information from conditioning requires weak exogeneity of the conditioning variables, see Engle et al. (1983). Conditions for weak exogeneity in cointegrated systems is analyzed in several studies and are by now well-known, see Boswijk (1992), Urbain (1992), Johansen (1992), Boswijk (1995) and Boswijk and Urbain (1997) for time series models. A set of conditioning variables are said to be weakly exogenous for a set of parameters belonging to the conditional model, if (i) the parameters of interest that are of the conditional model are only a function of the parameters of the conditional model, (ii) the parameters of the conditional model and the parameters of the marginal models for the conditioning variables are variation free.

In an error correction model set-up where the parameters of interest are the long-run parameters of the conditional model, a sufficient condition for the weak exogeneity of the conditioning variables for the parameters of interest is the absence of the error correcting behaviour of the conditioning variables as shown in Johansen (1992) and Urbain (1992). In our set-up, this is satisfied when $\alpha_{xi} = 0, \forall i = 1, \dots, N$, where α_{xi} is the adjustment coefficient in (4.4).

The assumption of weak exogeneity allows the researcher to conduct analysis efficiently within partial systems, by treating some variables as endogenous and conditioning on the rest of the variables. Intuitively, weak exogeneity implies that there is no loss of information (hence efficiency) in estimating the conditional model relative to the estimation of the joint model, i.e. the model for $\Delta \mathbf{z}_{it}$. If the assumption of weak exogeneity is violated this may lead to inefficient inference and, more importantly, the asymptotic distribution of the estimators for the long-run parameters obtained from the conditional model are no longer mixed normal. The serious consequences of the violation of weak exogeneity, hence the validity of the results in Chapter 3 necessitates a testing procedure for weak exogeneity. The testing procedure is discussed in the next section.

4.4 A Lagrange Multiplier Test for Weak Exogeneity

Following Boswijk (1992) and Boswijk and Urbain (1997), we use Lagrange multiplier type tests to test for the hypothesis that the coefficient of the error correction term in the marginal models are zero. In a time series set-up, Boswijk (1995) shows that this testing procedure has the same interpretation as the F-test proposed by Johansen (1992) to test for weak exogeneity, hence the Lagrange multiplier test can be considered as a variable addition test. The advantages of using Lagrange multiplier test for testing weak exogeneity is discussed in Boswijk (1992). One of the main advantages comes from the fact that we work under the null hypothesis of weak exogeneity and this allows us to analyze the conditional model and the marginal model separately without loss of efficiency. Another advantage is that while testing for the restriction $\alpha_{xi} = 0$, we can interpret and analyze the individual elements of α_{xi} separately.

A two step approach is used to obtain the test statistic. In a first step, we use the conditional model (4.6) to obtain super-consistent estimates of the long-parameters β_{xi} and λ_i by using an indirect least squares estimator. The estimators that are denoted by $\hat{\beta}_{xi}$ and $\hat{\lambda}_i$ are known to be T -(super-)consistent estimates of the long run parameters which are obtained as a result of the normalization restriction stated in Assumption 4.1(iv), see Chapter 3 for the proof. These estimates are used to construct an estimated the error correction term, $\hat{u}_{it} = y_{it} - \hat{\beta}_{xi}x_{it} - \hat{\lambda}_i f_t$, which is then used to obtain estimates of α_{xi} from the marginal model (4.4) and the associated LM statistics for weak exogeneity for cross-section unit $i = 1, \dots, N$.

We will consider the asymptotic properties of the Lagrange multiplier test statistic for individual units for the case in which the factors are assumed to be observed and for the case in which the factors are not observed. In the latter case we rely on the approximation method for the unobserved factors considered in Chapter 3. The approximation method is an adaptation of the method developed in Pesaran (2006), which proposed to use the cross-sectional averages of the observed variables as an approximation for the unobserved common factors. The details regarding the approximation for the unobserved factors is provided later in the chapter. After providing the results regarding the asymptotic properties of the individual Lagrange multiplier test statistics, we discuss the possible methods for obtaining a pooled test statistics to test the joint hypothesis of the presence of weak exogeneity for all cross section units.

4.4.1 Observed Factors

Let us first focus on the case of observed common factors first as a benchmark case. We rewrite the models in (4.4) and (4.6) in matrix notation by stacking the observations over time, for all $i = 1, \dots, N$; as

$$\begin{aligned} \Delta \mathbf{y}_i &= (\mathbf{y}_{i,-1} + \mathbf{X}_{i,-1} \boldsymbol{\beta}_{xi} + \mathbf{F}_{-1} \boldsymbol{\lambda}_i) \alpha_{yi} + \mathbf{G}_{yi} \boldsymbol{\Gamma}_{y.yi} + \mathbf{G}_{xi}^* \boldsymbol{\Gamma}_{y.xi} \\ &\quad + \mathbf{G}_f \boldsymbol{\Gamma}_{y.fi} + \boldsymbol{\varepsilon}_{y.xi}, \end{aligned} \quad (4.9)$$

$$\begin{aligned} \Delta \mathbf{X}_i &= (\mathbf{y}_{i,-1} + \mathbf{X}_{i,-1} \boldsymbol{\beta}_{xi} + \mathbf{F}_{-1} \boldsymbol{\lambda}_i) \alpha'_{xi} + \mathbf{G}_{yi} \boldsymbol{\Gamma}_{xyi} + \mathbf{G}_{xi} \boldsymbol{\Gamma}_{xxi} \\ &\quad + \mathbf{G}_f \boldsymbol{\Gamma}_{xfi} + \boldsymbol{\varepsilon}_{xi}, \end{aligned} \quad (4.10)$$

where

$$\begin{aligned} \Delta \mathbf{y}_i &= \begin{pmatrix} \Delta y_{i1} \\ \vdots \\ \Delta y_{iT} \end{pmatrix}_{T \times 1}, \quad \mathbf{y}_{i,-1} = \begin{pmatrix} y_{i0} \\ \vdots \\ y_{i,T-1} \end{pmatrix}_{T \times 1}, \quad \mathbf{X}_{i,-1} = \begin{pmatrix} \mathbf{x}'_{i0} \\ \vdots \\ \mathbf{x}'_{i,T-1} \end{pmatrix}_{T \times m}, \\ \mathbf{F}_{-1} &= \begin{pmatrix} \mathbf{f}'_0 \\ \vdots \\ \mathbf{f}'_{T-1} \end{pmatrix}_{T \times k}, \quad \boldsymbol{\varepsilon}_{y.xi} = \begin{pmatrix} \varepsilon_{y.xi1} \\ \vdots \\ \varepsilon_{y.xiT} \end{pmatrix}_{T \times 1}, \quad \boldsymbol{\varepsilon}_{xi} = \begin{pmatrix} \varepsilon'_{xi1} \\ \vdots \\ \varepsilon'_{xiT} \end{pmatrix}_{T \times 1}. \end{aligned}$$

For the current and lagged values of the first differences we have

$$\begin{aligned} \mathbf{G}_{yi} &= \begin{pmatrix} \mathbf{g}'_{yi1} \\ \vdots \\ \mathbf{g}'_{yiT} \end{pmatrix}_{T \times (p-1)}, \quad \mathbf{G}_{xi}^* = \begin{pmatrix} \mathbf{g}'_{xi1} \\ \vdots \\ \mathbf{g}'_{xiT} \end{pmatrix}_{T \times pm}, \quad \mathbf{G}_{xi} = \begin{pmatrix} \mathbf{g}'_{xi1} \\ \vdots \\ \mathbf{g}'_{xiT} \end{pmatrix}_{T \times (p-1)m}, \\ \mathbf{G}_f &= \begin{pmatrix} \mathbf{g}'_{f1} \\ \vdots \\ \mathbf{g}'_{fT} \end{pmatrix}_{T \times pk}, \end{aligned}$$

with

$$\begin{aligned} \mathbf{g}_{yit} &= \begin{pmatrix} \Delta y_{i,t-1} \\ \vdots \\ \Delta y_{i,t-p+1} \end{pmatrix}, & \mathbf{g}_{xit}^* &= \begin{pmatrix} \Delta \mathbf{x}_{it} \\ \vdots \\ \Delta \mathbf{x}_{i,t-p+1} \end{pmatrix}, \\ \mathbf{g}_{xit} &= \begin{pmatrix} \Delta \mathbf{x}_{i,t-1} \\ \vdots \\ \Delta \mathbf{x}_{i,t-p+1} \end{pmatrix}, & \mathbf{g}_{ft} &= \begin{pmatrix} \Delta \mathbf{f}_{t-1} \\ \vdots \\ \Delta \mathbf{f}_{t-p+1} \end{pmatrix}. \end{aligned}$$

Note that \mathbf{g}_{xit}^* includes the current value of $\Delta \mathbf{x}_{it}$. The coefficients of the conditional model are defined as follows: $\mathbf{\Gamma}_{y.yi} = (A_{y.yi1}, \dots, A_{y.yi,p-1})'$ is a $(p-1) \times 1$ vector, $\mathbf{\Gamma}_{y.xi} = (\mathbf{A}_{y.xi0}, \dots, \mathbf{A}_{y.xi,p-1})'$ is a $pm \times 1$ vector and $\mathbf{\Gamma}_{y.fi} = (\mathbf{A}_{y.fi1}, \dots, \mathbf{A}_{y.fi,p-1})'$ is a $(p-1)k \times 1$ vector. The coefficients of the marginal model are defined as $\mathbf{\Gamma}_{xyi} = (\mathbf{A}_{xyi1}, \dots, \mathbf{A}_{xyi,p-1})'$ is a $(p-1) \times m$ matrix, $\mathbf{\Gamma}_{xxi} = (\mathbf{A}_{xxi1}, \dots, \mathbf{A}_{xxi,p-1})'$ is a $(p-1)m \times m$ matrix and $\mathbf{\Gamma}_{xfi} = (\mathbf{A}_{xfi1}, \dots, \mathbf{A}_{xfi,p-1})'$ is a $(p-1)k \times m$ matrix.

As it is stated before, the Lagrange multiplier test statistic is obtained under the null hypothesis of weak exogeneity. The null hypothesis can be stated as

$$H_0 : \boldsymbol{\alpha}_{xi} = 0.$$

As also shown in Chapter 3 under weak exogeneity of the conditioning variables for the parameters of the conditional model, the indirect least squares estimators of the long-run coefficients of the conditional model are super-consistent and have a mixed normal distribution. Given the estimates $\hat{\boldsymbol{\beta}}_{xi}$ and $\hat{\boldsymbol{\lambda}}_i$, we can construct the estimated error correction term as $\hat{\mathbf{U}}_{i,-1} = \mathbf{Y}_{i,-1} - \mathbf{X}_{i,-1}\hat{\boldsymbol{\beta}}_{xi} - \mathbf{F}_{-1}\hat{\boldsymbol{\lambda}}_i$. Then the marginal model for $\Delta \mathbf{X}_i$ can be written by substituting the estimated error correction term $\hat{\mathbf{U}}_{i,-1}$ to (4.10) as

$$\Delta \mathbf{X}_i = \hat{\mathbf{U}}_{i,-1}\boldsymbol{\alpha}'_{xi} + \mathbf{G}_{yi}\mathbf{\Gamma}_{xyi} + \mathbf{G}_{xi}\mathbf{\Gamma}_{xxi} + \mathbf{G}_f\mathbf{\Gamma}_{xfi} + \mathbf{e}_{xi}, \quad (4.11)$$

where the error term is $\mathbf{e}_{xi} = (\mathbf{e}_{xi1}, \dots, \mathbf{e}_{xiT})'$. Note that the error term is no longer $\boldsymbol{\varepsilon}_{xi}$ as it was in the original marginal model for $\Delta \mathbf{x}_{it}$ namely (4.4), since we replaced the true error correction term with the estimated error correction term. It can be written as $\mathbf{e}_{xi} = \boldsymbol{\varepsilon}_{xi} + (\mathbf{U}_{i,-1} - \hat{\mathbf{U}}_{i,-1})\boldsymbol{\alpha}'_{xi}$. Note that under the null hypothesis of weak exogeneity \mathbf{e}_{xi} boils down to $\boldsymbol{\varepsilon}_{xi}$.

Defining the orthogonal projection matrix that projects onto the orthogonal space spanned by the columns of the first differences in the marginal model as $\mathbf{M}_i = \mathbf{I} - \mathbf{G}_i(\mathbf{G}'_i\mathbf{G}_i)^{-1}\mathbf{G}'_i$, where $\mathbf{G}_i = (\mathbf{G}_{yi}, \mathbf{G}_{xi}, \mathbf{G}_f)$ is a $T \times (p-1)(1+m+k)$ matrix, the least squares estimator of $\boldsymbol{\alpha}'_{xi}$ can be written as

$$\hat{\boldsymbol{\alpha}}'_{xi} = \left(\widehat{\mathbf{U}}'_{i,-1}\mathbf{M}_i\widehat{\mathbf{U}}_{i,-1} \right)^{-1} \left(\widehat{\mathbf{U}}'_{i,-1}\mathbf{M}_i\Delta\mathbf{X}_i \right). \quad (4.12)$$

The associated LM statistic for the i^{th} individual now can be written as

$$LM_i = \hat{\boldsymbol{\alpha}}'_{xi} \left[\widehat{\mathbf{V}}(\hat{\boldsymbol{\alpha}}_{xi}) \right]^{-1} \hat{\boldsymbol{\alpha}}_{xi}, \quad (4.13)$$

where

$$\widehat{\mathbf{V}}(\hat{\boldsymbol{\alpha}}_{xi}) = \widehat{\boldsymbol{\Sigma}}_{xxi}(\widehat{\mathbf{U}}'_{i,-1}\mathbf{M}_i\widehat{\mathbf{U}}_{i,-1})^{-1}$$

and $\widehat{\boldsymbol{\Sigma}}_{xxi} = T^{-1}\widehat{\mathbf{e}}^{0'}_{xi}\widehat{\mathbf{e}}^0_{xi}$ where $\widehat{\mathbf{e}}^0_{xi}$ are the residuals from estimating (4.11) imposing $\boldsymbol{\alpha}_{xi} = 0$.

Remark 4.4.8. One can also consider a Wald type test, denoted by w_i , if we replace the restricted estimator of $\boldsymbol{\Sigma}_{xxi}$ by the unrestricted one from (4.11).

As shown in Boswijk and Urbain (1997), LM_i has an asymptotic $\chi^2(m)$ distribution under the null, where m is the number of conditioning variables.

4.4.2 Unobserved Factors

Approximating F

There are two widely used methods to account for the cross-sectional dependence in the presence of unobserved common factors, namely the principal components approach and the common correlated effects approach. The principal components approach is proposed by Coakley et al. (2002), which has the shortcoming of yielding inconsistent estimates when the explanatory variables and the unobserved factors are correlated. This shortcoming is overcome by an iterative approach proposed by Bai (2009). This method is extended to a non-stationary case in Bai et al. (2009). The main disadvantage of the principal components method is that the estimation of the number of unobserved factors is necessary and this introduces additional uncertainty to the model. Pesaran (2006) proposes a computationally easier method which is called the Common Correlated Effects method, suggesting to use the cross sectional averages of the observed variables to augment the model

before performing a least squares regression. This method is extended to a static cointegrated panel case by Kapetanios et al. (2011). As for an error correction set-up with parametrically modelled dynamics including the dynamic behaviour of the unobserved common factors, Chapter 3 show that augmenting by lagged levels and lagged first differences of the cross-sectional averages of the observables $\bar{\mathbf{z}}_t$ captures the effect of the common factor when estimating the parameters of the long-run relation in the conditional model (4.6). For a detailed recent survey on the literature on cross-sectional dependence in panel data models see Chudik and Pesaran (2013d).

The approximation method we use in this chapter is proposed in Chapter 3 under the assumption of weak exogeneity. The presence of cointegration and the presence of parametrically modelled short-run dynamics in both the conditional and the marginal model requires a particular method for the approximation of the factors.

As discussed in the previous section we compute the Lagrange multiplier statistic in two steps. First we use the single equation conditional error correction model to obtain an estimate for the cointegrating vectors and we use this estimate in the marginal model to obtain the Lagrange multiplier statistic. This implies that we need to augment both the conditional model for Δy_{it} and the marginal model for $\Delta \mathbf{x}_{it}$ with the approximations. The method proposed in Chapter 3 uses the common trends representation of the model as a starting point and obtains the approximations for the levels of the unobserved factors that enter into the cointegrating relation and for the first differences of the unobserved factors separately. We do not provide the details of the method here. The reader is referred to Chapter 3 for the details. But for the sake of completeness, we present the model in common trends representation form then we state the approximation results using the notation of this form.

Using Theorem 4.1 of Johansen (1991), under Assumptions 4.1-4.3 and under the null hypothesis of weak exogeneity of the conditioning variables from the conditional model parameters, the common trends representation of the model (4.1) can be written as

$$\begin{aligned} \mathbf{z}_{it} &= \mathbf{C}_{zzi} \sum_{s=1}^t \boldsymbol{\varepsilon}_{zis} + \mathbf{C}_{zfi} \sum_{s=1}^t \boldsymbol{\varepsilon}_{fs} + \mathbf{C}_{zzi}^*(L) \boldsymbol{\varepsilon}_{zit} + \mathbf{C}_{zfi}^*(L) \boldsymbol{\varepsilon}_{ft} \\ &\quad + \mathbf{z}_{i0}^*, \end{aligned} \tag{4.14}$$

$$\mathbf{f}_t = \mathbf{C}_{ff} \sum_{s=1}^t \boldsymbol{\varepsilon}_{fs} + \mathbf{C}_{ff}^*(L) \boldsymbol{\varepsilon}_{ft} + \mathbf{f}_0^*, \quad (4.15)$$

where $\mathbf{z}_{it} = (y_{it}, \mathbf{x}'_{it})'$, $\boldsymbol{\varepsilon}_{zit} = (\varepsilon_{yit}, \boldsymbol{\varepsilon}'_{xit})'$, $\mathbf{C}_{zzi} = \begin{pmatrix} C_{yyi} & C_{yxi} \\ C_{xyi} & C_{xxi} \end{pmatrix}$, $\mathbf{C}_{zfi} = \begin{pmatrix} C_{yfi} \\ C_{xfi} \end{pmatrix}$ and \mathbf{z}_{i0}^* and \mathbf{f}_0^* contains the initial values. For notational simplicity we ignore the initial values. Starting with the representation (4.14)-(4.15) and applying the same steps as in Chapter 3, which includes taking the cross-sectional averages of the first differences of the observables at some point, we obtain

$$\overline{\Delta \mathbf{z}_t} - \mathbf{H}'_4(L) \Delta \mathbf{f}_t = \mathbf{v}_{Bt}, \quad (4.16)$$

where

$$\mathbf{v}_{Bt} = \frac{1}{N} \sum_{i=1}^N (\mathbf{C}_{zzi} + \mathbf{C}_{zzi}^*(L)(1-L)) \boldsymbol{\varepsilon}_{zit} \quad (4.17)$$

and $\mathbf{H}'_4(L) = (\overline{\mathbf{C}}_{zf} + \overline{\mathbf{C}}_{zf}^*(L)(1-L))(\mathbf{C}_{ff} + \mathbf{C}_{ff}^*(L)(1-L))^{-1}$. \mathbf{v}_{Bt} is the resulting approximation error when the cross-sectional averages of the first differences of the observables are used as an approximation for the unobserved factors. The rotation matrix polynomial $\mathbf{H}'_4(L)$ is an infinite order lag-polynomial and this implies that one needs to augment the model with the approximations with a lag order of infinity. This issue is addressed in Remark 4.4.9 below. Furthermore $\mathbf{H}'_4(L)$ has to be invertible. To be able to use the right inverse suffices to obtain the results provided in the current chapter. One of the conditions for the existence of the right inverse of the lag-polynomial $\mathbf{H}'_4(L)$ of size $(m+1) \times k$ is imposed by the assumption below.

Assumption 4.4. *The infinite order lag-polynomial $\mathbf{H}'_4(L) : (m+1) \times k$ has full column rank.*

The second condition required for invertibility is that the coefficients of $\mathbf{H}'_4(L)$ should have exponentially decaying coefficients. This is implied by the stability of the condition imposed in Assumption 4.1(ii). This issue is discussed in detail in Chapter 3. The assumption above together with the second condition ensures that a right inverse of the lag-polynomial $\mathbf{H}_4(L)$ exists. We can write the left inverse as

$$\mathbf{H}_4(L)^- = \mathbf{H}'_4(L) [\mathbf{H}_4(L) \mathbf{H}'_4(L)]^{-1}, \quad (4.18)$$

which has a size of $(m + 1) \times k$.

Now we provide the approximation result for the levels of the unobserved factors that enter into the cointegrating relation. The approximation error can be expressed in terms of the common trends representation parameters as

$$\bar{\mathbf{z}}_t - \mathbf{H}'_1 \mathbf{f}_t = \mathbf{H}_3(L) \overline{\Delta \mathbf{z}}_t + \mathbf{v}_{At}, \quad (4.19)$$

where

$$\begin{aligned} \mathbf{v}_{At} &= \frac{1}{N} \sum_{i=1}^N \mathbf{C}_{zzi} \sum_{s=1}^t \boldsymbol{\varepsilon}_{zis} \\ &\quad + \frac{1}{N} \sum_{i=1}^N (\mathbf{C}_{zzi}^*(L) + \mathbf{H}_3(L)(\mathbf{C}_{zzi} + \mathbf{C}_{zzi}^*(L)(1 - L))) \boldsymbol{\varepsilon}_{zit}, \end{aligned} \quad (4.20)$$

and $\mathbf{H}'_1 = \overline{\mathbf{C}}_{zf} \mathbf{C}_{ff}^{-1}$ and $\mathbf{H}_3(L) = (\overline{\mathbf{C}}_{zf} \mathbf{C}_{ff}^{-1} \mathbf{C}_{ff}^*(L) - \overline{\mathbf{C}}_{zf}^*(L)) (\overline{\mathbf{C}}_{zf} + \overline{\mathbf{C}}_{zf}^*(L)(1 - L))^{-1}$. \mathbf{v}_{At} involves the cross-sectional average of a partial sum process. At this stage we need to make a further assumption. The rotation matrix \mathbf{H}'_1 has to be invertible and again the existence of a right inverse would suffice to obtain the results. We make the following assumption for this purpose.

Assumption 4.5. $\mathbf{H}'_1 : (m + 1) \times k$ has a full column rank.

This assumption allows us to define the right inverse of the rotation matrix \mathbf{H}_1 as

$$\mathbf{H}_1^- = \mathbf{H}'_1 (\mathbf{H}_1 \mathbf{H}'_1)^{-1}, \quad (4.21)$$

where \mathbf{H}_1^- has a size of $(m + 1) \times k$.

Remark 4.4.9. $\mathbf{H}_3(L)$ and $\mathbf{H}_4(L)$ are lag polynomials of infinite orders. This infiniteness stems from the right-inversion of finite order lag polynomials. According to this result infinite lag length should be used while augmenting the regression by the first differences of the cross-sectional averages. However this is not a very practical result. In practice one can use a truncation lag order if certain conditions for the coefficients of the infinite order lag polynomial are satisfied. We discuss the necessary conditions which make truncation possible in the next section.

The augmented marginal model

In this section we rewrite the marginal model for the conditioning variables by replacing the unobserved common factors with their approximations obtained in

the previous section. Consider the marginal model for $\Delta \mathbf{x}_{it}$ stated in (4.4). By stacking the observations over t and by defining the lag-polynomials $\mathbf{A}_{xyi}(L) = \sum_{\ell=0}^{p-1} \mathbf{A}_{xyi\ell} L^\ell$, $\mathbf{A}_{xxi}(L) = \sum_{\ell=0}^{p-1} \mathbf{A}_{xxi\ell} L^\ell$ and $\mathbf{A}_{xfi}(L) = \sum_{\ell=0}^{p-1} \mathbf{A}_{xfi\ell} L^\ell$ we can write (4.4) as

$$\begin{aligned} \Delta \mathbf{X}_i &= (\mathbf{y}_{i,-1} - \mathbf{X}_{i,-1} \boldsymbol{\beta}_{xi} - \mathbf{F}_{-1} \boldsymbol{\lambda}_i) \boldsymbol{\alpha}'_{xi} + \Delta \mathbf{y}_{i,-1} \mathbf{A}'_{xyi}(L) + \Delta \mathbf{X}_{i,-1} \mathbf{A}'_{xxi}(L) \\ &\quad + \Delta \mathbf{F}_{-1} \mathbf{A}'_{xfi}(L) + \boldsymbol{\varepsilon}_{xi}. \end{aligned}$$

Considering the approximation results stated in (4.16) and (4.19) and the definitions for the approximation errors stated in (4.17) and (4.20) we can rewrite the marginal model above as

$$\begin{aligned} \Delta \mathbf{X}_i &= (\mathbf{y}_{i,-1} - \mathbf{X}_{i,-1} \boldsymbol{\beta}_{xi} - \bar{\mathbf{Z}}_{-1} \boldsymbol{\lambda}_i^*) \boldsymbol{\alpha}'_{xi} + \Delta \mathbf{y}_{i,-1} \mathbf{A}'_{xyi}(L) + \Delta \mathbf{X}_{i,-1} \mathbf{A}'_{xxi}(L) \\ &\quad + \bar{\Delta} \bar{\mathbf{Z}}_{-1} \mathbf{A}'_{xfi}(L) + \mathbf{V}_{A,-1} \mathbf{H}_1^- \boldsymbol{\lambda}_i \boldsymbol{\alpha}'_{xi} + \mathbf{V}_{B,-1} \mathbf{H}_4^-(L) \mathbf{A}'_{xfi}(L) \\ &\quad + \boldsymbol{\varepsilon}_{xi}, \end{aligned} \tag{4.22}$$

where $\boldsymbol{\lambda}_i^* = \mathbf{H}_1^- \boldsymbol{\lambda}_i$ and $\mathbf{A}'_{xfi}(L) = \mathbf{H}_4^-(L) \mathbf{A}'_{xfi}(L) + \mathbf{H}_3^-(L) \mathbf{H}_1^- \boldsymbol{\lambda}_i \boldsymbol{\alpha}'_{xi}$ that is an infinite order lag-polynomial. First differences of cross sectional averages appear in equation (4.22) for two reasons. First, because of the augmentation of the lagged levels of the unobserved factors, which is reflected by the second term of $\mathbf{A}'_{xfi}(L)$ and secondly because of the augmentation of the lagged first differences of the unobserved factors that is reflected by the first term in the definition of $\mathbf{A}'_{xfi}(L)$. Under the null hypothesis of weak exogeneity the second term of $\mathbf{A}'_{xfi}(L)$ will be zero. Since the lag polynomials in the summands of $\mathbf{A}'_{xfi}(L)$ are obtained by an inversion of a lag-polynomial, the lag order of the polynomial is potentially infinite and hence will have to be truncated to make the procedure feasible. The truncation requires exponentially decaying coefficients, see Chapter 3 for a discussion of this issue.

Since $\mathbf{A}'_{xfi}(L)$ reduces to $\mathbf{A}'_{xfi}(L) = \mathbf{H}_4^-(L) \mathbf{A}'_{xfi}(L)$ under the null hypothesis, the fact that it has exponentially decaying coefficients can be proven by using similar arguments as in Chapter 3, which makes use of the relevant results in Chudik and Pesaran (2013a) and Chudik and Pesaran (2013c). Here we assume that there exists real positive constants $K < \infty$ and $\rho_A < 1$ such that, $|\mathbf{A}'_{xfi\ell}| < K \rho_A^\ell$, for any $\ell \in \{0, 1, 2, \dots\}$.

In choosing the lag-length we need to consider two constraints that have opposite effects. This yields an optimal solution. An optimal lag-length can be selected by considering the asymptotic effect of the first differences that are not modelled

parametrically, hence left out in the error term and the loss of degrees of freedom arises by using higher order lags. Two two constraints make the problem an optimization problem. In this chapter we show that using a proper truncated lag order and leaving the rest of the lagged first differences of the cross-sectional averages in the error term do not distort the asymptotic behaviour of the test statistics under certain conditions, which are specified in the theorems. In the next section we rewrite the model in (4.22) by using a hypothetical truncation-lag order for the first differences of the cross sectional averages. We shall call the truncation lag order p_T hereafter, to emphasise that it depends on the time series dimension.

LM test with unobserved factors

To obtain the LM_i statistic for the i^{th} cross-section unit we follow a two step approach. The first step is to obtain the estimated error correction term by using the augmented conditional model for Δy_{it} . We denote the estimated error correction term obtained by estimating the augmented conditional model for Δy_{it} as $\widehat{\mathbf{U}}_{i,-1}$. From Chapter 3 we know that it is a consistent estimate of $\mathbf{U}_{i,-1}$. The consistency of $\widehat{\mathbf{U}}_{i,-1}$ stems from the T -consistent estimates of the long-run parameters denoted by $\widehat{\boldsymbol{\beta}}_{xi}$ and $\widehat{\mathbf{H}}_1^- \boldsymbol{\lambda}_i$ ($\widehat{\boldsymbol{\lambda}}_i$) that are indirect least squares estimators of $\boldsymbol{\beta}_{xi}$ and $\boldsymbol{\lambda}_i^*$ respectively. Although $\widehat{\mathbf{H}}_1^- \boldsymbol{\lambda}_i$ is not the estimator for the factor loadings in the long-run relation and has no value in terms of interpretation, using this estimate allows us to obtain the consistent fitted error correction term.

The model with the truncated lag order for the first differences of the cross-sectional averages can be written as

$$\begin{aligned} \Delta \mathbf{X}_i &= (\mathbf{y}_{i,-1} - \mathbf{X}_{i,-1} \boldsymbol{\beta}_{xi} - \overline{\mathbf{Z}}_{-1} \boldsymbol{\lambda}_i^*) \boldsymbol{\alpha}'_{xi} + \mathbf{G}_{yi} \boldsymbol{\Gamma}_{xyi} + \mathbf{G}_{xi} \boldsymbol{\Gamma}_{xxi} \\ &\quad + \overline{\mathbf{G}}_z \boldsymbol{\Gamma}_{xfi}^* + \tilde{\boldsymbol{\varepsilon}}_{xi}, \end{aligned} \quad (4.23)$$

where

$$\begin{aligned} \tilde{\boldsymbol{\varepsilon}}_{xi} &= \boldsymbol{\varepsilon}_{xi} + \mathbf{V}_{A,-1} \mathbf{H}_1^- \boldsymbol{\lambda}_i \boldsymbol{\alpha}'_{xi} - \mathbf{V}_{B,-1} \mathbf{H}_4^-(L) \mathbf{A}'_{xfi}(L) \\ &\quad + \overline{\Delta \mathbf{Z}}_{-p_T-1} \mathbf{A}^{**'}_{xfi}(L), \end{aligned} \quad (4.24)$$

and where $\Delta \mathbf{y}_i$ is a $(T - p_T) \times 1$ vector, $\mathbf{X}_{i,-1}$, \mathbf{G}_{yi} , \mathbf{G}_{xi} are $(T - p_T) \times m$, $(T - p_T) \times (p - 1)$ and $(T - p_T) \times m(p - 1)$ matrices respectively and defined

similarly to the corresponding matrices in Section 4.4.1. Furthermore, we define

$$\begin{aligned} \bar{\mathbf{Z}}_{-1} &= \begin{pmatrix} \bar{\mathbf{z}}'_{p_T} \\ \vdots \\ \bar{\mathbf{z}}'_{T-1} \end{pmatrix}, & \overline{\Delta\mathbf{Z}}_{-p_T-1} &= \begin{pmatrix} \overline{\Delta\mathbf{z}}'_0 \\ \vdots \\ \overline{\Delta\mathbf{z}}'_{T-p_T-1} \end{pmatrix}, \\ \tilde{\boldsymbol{\varepsilon}}_{xi} &= \begin{pmatrix} \tilde{\boldsymbol{\varepsilon}}_{xi,p_T} \\ \vdots \\ \tilde{\boldsymbol{\varepsilon}}_{xiT} \end{pmatrix}. \end{aligned}$$

For the current and lagged values of the approximations we have

$$\bar{\mathbf{G}}_z = \begin{pmatrix} \bar{\mathbf{g}}'_{z,p_T+1} \\ \vdots \\ \bar{\mathbf{g}}'_{zT} \end{pmatrix}, \quad \bar{\mathbf{g}}_{zt} = \begin{pmatrix} \overline{\Delta\mathbf{z}}_t \\ \vdots \\ \overline{\Delta\mathbf{z}}_{t-p_T} \end{pmatrix},$$

and where $\mathbf{\Gamma}_{xfi}^* = (\mathbf{A}_{xfi0}^*, \mathbf{A}_{xfi1}^*, \dots, \mathbf{A}_{xfi,p_T}^*)'$ and $\mathbf{A}_{xfi}^{**}(L) = \sum_{\ell=0}^{\infty} \mathbf{A}_{xfi,\ell+p_T}^* L^\ell$. The error term of the model now is $\tilde{\boldsymbol{\varepsilon}}_{xi}$ that is defined in (4.24), which includes the regression error $\boldsymbol{\varepsilon}_{xi}$, the factor approximation errors and the truncation error. The first term is the regression error, the second and the third terms are induced by the factor approximations for the levels and for the first differences of the unobserved factors respectively.

For the case in which the factors are approximated by the cross-sectional averages of the observables, defining the error correction term as

$$\mathbf{U}_{i,-1} = \mathbf{y}_{i,-1} - \mathbf{X}_{i,-1}\boldsymbol{\beta}_{xi} - \mathbf{F}_{-1}\boldsymbol{\lambda}_i, \quad (4.25)$$

the fitted error correction term can be written as

$$\widehat{\mathbf{U}}_{i,-1} = \mathbf{y}_{i,-1} - \mathbf{X}_{i,-1}\widehat{\boldsymbol{\beta}}_{xi} - \bar{\mathbf{Z}}_{-1}\widehat{\mathbf{H}}_1^{-1}\boldsymbol{\lambda}_i, \quad (4.26)$$

where $\widehat{\boldsymbol{\beta}}_{xi}$ and $\widehat{\mathbf{H}}_1^{-1}\boldsymbol{\lambda}_i$ are the indirect least squares estimators of $\boldsymbol{\beta}_{xi}$ and $\boldsymbol{\lambda}_i^*$ respectively and where $\boldsymbol{\lambda}_i^*$ is defined subsequent to (4.22). These estimators are obtained using the augmented conditional model for Δy_{it} . Now by using (4.19), which gives the approximation error for the levels of the unobserved factors we can write

$$\begin{aligned} \widehat{\mathbf{U}}_{i,-1} - \mathbf{U}_{i,-1} &= \mathbf{X}_{i,-1}(\boldsymbol{\beta}_{xi} - \widehat{\boldsymbol{\beta}}_{xi}) + \bar{\mathbf{Z}}_{-1}(\mathbf{H}_1^{-1}\boldsymbol{\lambda}_i - \widehat{\mathbf{H}}_1^{-1}\boldsymbol{\lambda}_i) \\ &\quad - \overline{\Delta\mathbf{Z}}_{-1}\mathbf{H}'_3(L)\mathbf{H}_1^{-1}\boldsymbol{\lambda}_i - \mathbf{V}_{A,-1}\mathbf{H}_1^{-1}\boldsymbol{\lambda}_i. \end{aligned} \quad (4.27)$$

While the effect of estimating β_{xi} and λ_i still vanishes at rate T due to super-consistency, see Chapter 3, we have two additional terms whose properties depend on the rate of convergence of the factor estimates and final term due to the short-run effects of the approximation of the unobserved factors. Substituting the estimated error correction term (4.26) in the augmented marginal model (4.23), we obtain

$$\Delta \mathbf{X}_i = \widehat{\mathbf{U}}_{i,-1} \alpha'_{xi} + \mathbf{G}_{yi} \Gamma_{xyi} + \mathbf{G}_{xi} \Gamma_{xxi} + \overline{\mathbf{G}}_z \Gamma_{xfi}^* + \tilde{\mathbf{e}}_{xi}, \quad (4.28)$$

where

$$\begin{aligned} \tilde{\mathbf{e}}_{xi} = & \varepsilon_{xi} + (\tilde{\mathbf{U}}_{i,-1} - \widehat{\mathbf{U}}_{i,-1}) \alpha'_{xi} + \mathbf{V}_{A,-1} \mathbf{H}_1^- \lambda_i \alpha'_{xi} \\ & + \mathbf{V}_{B,-1} \mathbf{H}_4^-(L) \mathbf{A}'_{xfi}(L) + \overline{\Delta \mathbf{Z}}_{-PT} \mathbf{A}^{**'}_{xfi}(L). \end{aligned} \quad (4.29)$$

The set-up can be simplified if we impose the null hypothesis of weak exogeneity. The first term on the right-hand side of (4.29) and the third term vanish under the null hypothesis that $\alpha_{xi} = 0$. Thus, without short run dynamics namely if $\mathbf{A}_{xfi}(L) = 0$, the LM (also Wald) type tests given above will be nuisance parameter free and have the same distribution as if the common factors were observed. However, in the presence of additional short-run dynamics, the second term and the last term, $\mathbf{V}_{B,-1} \mathbf{H}_4^-(L) \mathbf{A}'_{xfi}(L)$ and $\overline{\Delta \mathbf{Z}}_{-PT-1} \mathbf{A}^{**'}_{xfi}(L)$ will induce nuisance parameters on the limiting distributions for finite N .

In order to show under which conditions the limiting distribution will be nuisance parameter free in the presence of short-run dynamics, we start with writing the estimator for α_{xi} that is obtained from the estimation of the model in (4.28). We have

$$\widehat{\alpha}'_{xi} = \left(\widehat{\mathbf{U}}'_{i,-1} \widetilde{\mathbf{M}}_i \widehat{\mathbf{U}}_{i,-1} \right)^{-1} \left(\widehat{\mathbf{U}}'_{i,-1} \widetilde{\mathbf{M}}_i \Delta \mathbf{X}_i \right), \quad (4.30)$$

where

$$\widetilde{\mathbf{M}}_i = \mathbf{I}_{T-p_T} - \widetilde{\mathbf{G}}_i (\widetilde{\mathbf{G}}'_i \widetilde{\mathbf{G}}_i)^{-1} \widetilde{\mathbf{G}}'_i, \quad (4.31)$$

with $\widetilde{\mathbf{G}}_i = (\mathbf{G}_{yi}, \mathbf{G}_{xi}, \overline{\mathbf{G}}_z)$ and the definitions for \mathbf{G}_{yi} , \mathbf{G}_{xi} and $\overline{\mathbf{G}}_z$ can be found subsequent to (4.23). The Lagrange Multiplier test statistic can then be written as

$$\widetilde{LM}_i = \widehat{\alpha}'_{xi} \left[\widehat{\mathbf{V}}(\widehat{\alpha}_{xi}) \right]^{-1} \widehat{\alpha}_{xi}, \quad (4.32)$$

where $\widehat{\mathbf{V}}(\widehat{\boldsymbol{\alpha}}_{xi}) = \widehat{\boldsymbol{\Sigma}}_{xxi}(\widehat{\mathbf{U}}'_{i,-1}\widetilde{\mathbf{M}}_i\widehat{\mathbf{U}}_{i,-1})^{-1}$, and $\widehat{\boldsymbol{\Sigma}}_{xxi} = T^{-1}\widehat{\mathbf{e}}_{xi}^{0'}\widehat{\mathbf{e}}_{xi}^0$ where $\widehat{\mathbf{e}}_{xi}^0$ are the residuals from estimating (4.28) under the null hypothesis, imposing $\boldsymbol{\alpha}_{xi} = 0$. Under the null of weak exogeneity the error term of the marginal model (4.28) becomes

$$\widetilde{\mathbf{e}}_{xi} = \boldsymbol{\varepsilon}_{xi} + \mathbf{V}_{B,-1}\mathbf{H}_4^-(L)\mathbf{A}'_{xfi}(L) + \overline{\Delta\mathbf{Z}}_{-p_T-1}\mathbf{A}^{**'}_{xfi}(L), \quad (4.33)$$

where $\mathbf{V}_{B,-1}$ is the approximation error defined in (4.16) and

$$\mathbf{A}^{**}_{xfi}(L) = \sum_{\ell=0}^{\infty} \mathbf{A}^*_{xfi,\ell+p_T} L^\ell,$$

where $\mathbf{A}^{*'}_{xfi}(L) = \sum_{\ell=0}^{\infty} \mathbf{A}^{*'}_{xfi\ell} L^\ell = \mathbf{H}_4^-(L)\mathbf{A}'_{xfi}(L)$ under the null hypothesis of weak exogeneity. Using (4.33) and (4.28) in (4.30), under the null hypothesis, the estimator for $\boldsymbol{\alpha}'_{xi}$ becomes

$$\begin{aligned} \widehat{\boldsymbol{\alpha}}'_{xi} &= \left(\widehat{\mathbf{U}}'_{i,-1}\widetilde{\mathbf{M}}_i\widehat{\mathbf{U}}_{i,-1} \right)^{-1} \\ &\quad \times \left[\widehat{\mathbf{U}}'_{i,-1}\widetilde{\mathbf{M}}_i(\boldsymbol{\varepsilon}_{xi} + \mathbf{V}_{B,-1}\mathbf{H}_4^-(L)\mathbf{A}'_{xfi}(L) + \overline{\Delta\mathbf{Z}}_{-p_T-1}\mathbf{A}^{**'}_{xfi}(L)) \right]. \end{aligned}$$

Noting that we have a single cointegrating relation, which makes $\widehat{\mathbf{U}}'_{i,-1}\widetilde{\mathbf{M}}_i\widehat{\mathbf{U}}_{i,-1}$ a scalar and substituting this into the expression for LM_i , we have

$$\begin{aligned} \widetilde{LM}_i &= \left(\widehat{\mathbf{U}}'_{i,-1}\widetilde{\mathbf{M}}_i\widehat{\mathbf{U}}_{i,-1} \right)^{-1} \\ &\quad \times \left[\widehat{\mathbf{U}}'_{i,-1}\widetilde{\mathbf{M}}_i(\boldsymbol{\varepsilon}_{xi} + \mathbf{V}_{B,-1}\mathbf{H}_4^-(L)\mathbf{A}'_{xfi}(L) + \overline{\Delta\mathbf{Z}}_{-p_T-1}\mathbf{A}^{**'}_{xfi}(L)) \right] \\ &\quad \times \widehat{\boldsymbol{\Sigma}}_{xxi}^{-1} \left[(\boldsymbol{\varepsilon}_{xi} + \mathbf{V}_{B,-1}\mathbf{H}_4^-(L)\mathbf{A}'_{xfi}(L) + \overline{\Delta\mathbf{Z}}_{-p_T-1}\mathbf{A}^{**'}_{xfi}(L))' \widetilde{\mathbf{M}}_i\widehat{\mathbf{U}}_{i,-1} \right]. \end{aligned} \quad (4.34)$$

A further analysis of the right-hand side of the equation above gives us the extent to which the asymptotic properties of the LM_i statistic is affected by the approximation method we use for the unobserved common factors. We first present four important lemmas which are directly used in the theorem presented later in the chapter. The proofs of the lemmas can be found in the Appendix. We need some additional notation to present the results in the lemmas in a compact way. For

this purpose, we define the orthogonal projection matrix

$$\ddot{\mathbf{M}}_i = \mathbf{I} - \ddot{\mathbf{G}}_i(\ddot{\mathbf{G}}_i' \ddot{\mathbf{G}}_i)^+ \ddot{\mathbf{G}}_i, \quad (4.35)$$

where $\ddot{\mathbf{G}}_i = (\mathbf{G}_{yi}, \mathbf{G}_{xi}, \ddot{\mathbf{G}}_F)$ with $\ddot{\mathbf{G}}_F = (\ddot{\mathbf{g}}_{f1}, \dots, \ddot{\mathbf{g}}_{fT})'$ is a $T \times p_T(1+m)$. Each row of $\ddot{\mathbf{G}}_F'$ is defined as

$$\ddot{\mathbf{g}}_{ft} = \begin{pmatrix} \sum_{\ell=0}^{\infty} \mathbf{H}'_{4\ell} \Delta \mathbf{f}_{t-\ell-1} \\ \vdots \\ \sum_{\ell=0}^{\infty} \mathbf{H}'_{4\ell} \Delta \mathbf{f}_{t-p_T-\ell-1} \end{pmatrix}. \quad (4.36)$$

Note that a $(m+1) \times 1$ row of $\ddot{\mathbf{g}}_{ft}$ represents the rotated first differences true factors. Taking into account the approximation result stated in (4.16), this definition is useful to find the effect of the approximation on the Lagrange multiplier statistic. Note that the elements of $\ddot{\mathbf{G}}_F$ are not known, however the asymptotic properties of the elements are known. It can be shown that using $\ddot{\mathbf{M}}_i$ instead of \mathbf{M}_i does not alter the asymptotic distribution of the LM_i statistic. So we can adapt the result mentioned in Section 4.4.1 and write

$$L\ddot{\mathbf{M}}_i \sim \chi^2(m), \quad (4.37)$$

where $L\ddot{\mathbf{M}}_i$ is the Lagrange Multiplier test statistic obtained using $\widehat{\boldsymbol{\alpha}}_{xi}$, where

$$\widehat{\boldsymbol{\alpha}}_{xi} = \left(\mathbf{U}'_{i,-1} \ddot{\mathbf{M}}_i \mathbf{U}_{i,-1} \right)^{-1} \left(\mathbf{U}'_{i,-1} \ddot{\mathbf{M}}_i \Delta \mathbf{X}_i \right),$$

with $\widehat{V}(\widehat{\boldsymbol{\alpha}}_{xi}) = \widehat{\boldsymbol{\Sigma}}_{xxi}(\mathbf{U}'_{i,-1} \ddot{\mathbf{M}}_i \mathbf{U}_{i,-1})^{-1}$ and $\widehat{\boldsymbol{\Sigma}}_{xxi} = T^{-1} \widehat{\mathbf{e}}_{xi}^{0'} \widehat{\mathbf{e}}_{xi}^0$ where $\widehat{\mathbf{e}}_{xi}^0$ are the residuals from estimating

$$\Delta \mathbf{X}_i = \mathbf{G}_{yi} \boldsymbol{\Gamma}_{xyi} + \mathbf{G}_{xi} \boldsymbol{\Gamma}_{xxi} + \ddot{\mathbf{G}}_F \boldsymbol{\Gamma}_{xfi}^* + \ddot{\mathbf{e}}_{xi}. \quad (4.38)$$

In the following lemmas we provide the asymptotic results regarding the elements of \widetilde{LM}_i in relation to the elements of $L\ddot{\mathbf{M}}_i$.

Lemma 4.4.1. *Consider the long-run coefficients in the error correction term in (4.26), $\widehat{\boldsymbol{\beta}}_{xi}$ and $\widehat{\boldsymbol{\lambda}}_i$. Under Assumptions 4.1-4.5, under the null hypothesis of weak exogeneity and as $(T, N, p_T) \rightarrow \infty$, the indirect least squares estimators of $\widehat{\boldsymbol{\beta}}_{xi}$ and $\widehat{\boldsymbol{\lambda}}_i$ obtained by estimating the augmented, truncated conditional error correction*

model for Δy_{it} have the following properties regarding their probability orders:

$$\begin{pmatrix} \widehat{\beta}_{xi} - \beta_{xi} \\ \widehat{\lambda}_i - \lambda_i^* \end{pmatrix} = O_p\left(\frac{1}{T}\right) + O_p\left(\frac{1}{\sqrt{N}}\right) + O_p\left(\frac{p_T^2}{T^2}\right), \quad (4.39)$$

where β_{xi} and λ_i^* are the true values of the long-run coefficients of the augmented error correction term.

Lemma 4.4.2. *As $(N, T, p_T) \rightarrow \infty$, if $\frac{T}{N} \rightarrow 0$, $\frac{p_T^2}{T} \rightarrow 0$, then under the null hypothesis of weak exogeneity we have*

$$(i) \quad \frac{\widehat{\mathbf{U}}'_{i,-1} \widetilde{\mathbf{M}}_i \widehat{\mathbf{U}}_{i,-1}}{T} - \frac{\mathbf{U}'_{i,-1} \ddot{\mathbf{M}}_i \mathbf{U}_{i,-1}}{T} = o_p(1), \quad (4.40)$$

$$(ii) \quad \frac{\widehat{\mathbf{U}}'_{i,-1} \widetilde{\mathbf{M}}_i \mathbf{e}_{xi}}{\sqrt{T}} - \frac{\mathbf{U}'_{i,-1} \ddot{\mathbf{M}}_i \mathbf{e}_{xi}}{\sqrt{T}} = o_p(1), \quad (4.41)$$

$$(iii) \quad \frac{\widehat{\mathbf{U}}'_{i,-1} \widetilde{\mathbf{M}}_i \mathbf{V}_{B,-1} \mathbf{H}_4^-(L) \mathbf{A}'_{xfi}(L)}{\sqrt{T}} = o_p(1), \quad (4.42)$$

and in addition to the conditions stated above, if $\sqrt{T} \rho_A^{p_T} \rightarrow 0$, we have

$$(iv) \quad \frac{\widehat{\mathbf{U}}'_{i,-1} \widetilde{\mathbf{M}}_i \overline{\Delta \mathbf{Z}}_{p_T-1} \mathbf{H}_4^-(L) \mathbf{A}^{**'}_{xfi}(L)}{\sqrt{T}} = o_p(1). \quad (4.43)$$

Lemma 4.4.3. *The difference between the estimated variance of the residuals of the two cases is negligible asymptotically only if $(T, N, p_T) \rightarrow \infty$, $\frac{p_T^2}{T} \rightarrow 0$ and $\frac{p_T^2}{N} \rightarrow 0$. This result can be stated as*

$$\widehat{\boldsymbol{\Sigma}}_{xxi} - \widetilde{\boldsymbol{\Sigma}}_{xxi} = o_p(1). \quad (4.44)$$

The proofs of the lemmas can be found in the Appendix.

Lemma 4.4.1 gives the resulting probability order of the long-run coefficient estimates from the augmented conditional model when they are estimated by the indirect least squares method. This result is a byproduct of Theorem 4 of Chapter 3, which suggests that the long-run coefficients can be estimate super-consistently as $\frac{T}{\sqrt{N}} \rightarrow 0$ and $\frac{p_T^2}{T} \rightarrow 0$ as $(N, T, p_T) \rightarrow \infty$. Lemma 4.4.2(i) asserts that, under the conditions of the lemma, the approximation does not affect the second moments of the error correction term when projected on the orthogonal space spanned by the first differences of the augmented marginal model. Lemma 4.4.2(ii) shows that, under the conditions of the lemma, the approximation does not affect the

covariance between the error correction term and the regression error when projected on the orthogonal space spanned by the first differences of the augmented marginal model. Lemma 4.4.2(iii) shows that the approximation error does not affect the LM_i statistic asymptotically and lastly, Lemma 4.4.2 (iv) shows that the truncation error does not affect the LM_i statistic asymptotically, under the conditions stated for that result.

Using these lemmas, we can state our theorem regarding the asymptotic behaviour of the LM_i test statistic for the i^{th} individual, for the case with approximated factors.

Theorem 4.1. *Suppose that \mathbf{z}_{it}^* for $i = 1, \dots, N$ and $t = 1, \dots, T$ are generated by (4.1) and Assumptions 4.1-4.3, 4.4 and 4.5 hold, consider the Lagrange multiplier test statistic in (4.32). Let $\frac{T}{N} \rightarrow 0$, $\frac{p_T^2}{T} \rightarrow 0$, and $\sqrt{T}\rho_A^{p_T} \rightarrow 0$ as $(N, T, p_T) \rightarrow \infty$, then the difference between $\check{L}\check{M}_i$ defined in (4.37) and $\widetilde{L}\widetilde{M}_i$ is asymptotically negligible, so we have*

$$\widetilde{L}\widetilde{M}_i \sim \chi^2(m).$$

Proof of Theorem 4.1: The result in the theorem is obtained using the results in the Lemmas 4.4.1-4.4.3.

Remark 4.4.10. The necessary conditions regarding the truncated lag order for these results to hold are $\frac{p_T^2}{T} \rightarrow 0$ and $\sqrt{T}\rho_A^{p_T} \rightarrow 0$. As also explained in Remark 2 of Chudik and Pesaran (2013a), the first condition guarantees the existence of a sufficient degree of freedom which makes consistent and efficient estimation possible. The second condition guarantees that the omitted variable problem arises from ignoring the lags after the p_T^{th} order is asymptotically negligible.

Remark 4.4.11. The required relative rate of increase for N and T is $\frac{T}{N} \rightarrow 0$ as $(N, T) \rightarrow \infty$ jointly. This condition suggests that the limit theory provided here is more useful for the cases where the N -dimension is larger than the T -dimension. In most of the panel applications this requirement can be met easily. For example in macroeconomic applications one can increase the cross-section dimension by including more countries whereas the time dimension can not be increased that much due to availability problems.

4.4.3 Pooled Test

It has been shown that pooling information across cross-section units improves the power of the tests under consideration. In this section we study the pooled weak exogeneity tests. To be specific, we are interested in testing the pooled null hypothesis

$$H_0^p : \alpha_{xi} = 0, \text{ for all } i \in \{1, \dots, N\},$$

against the alternative,

$$H_A^p : \alpha_{xi} < 0, \text{ for some } i \in \{1, \dots, N\}.$$

In the following lines, we mention two pooling methods that are commonly used in the literature and have been proven to have standard asymptotic distributions, good size and power properties in the absence of cross-sectional dependence. One of the methods that is used for panel tests is inspired by Fisher (1934) and is based on combining p-values of individual test statistics. This method is extensively used in meta-analysis. We denote the significance level (p-value) of the LM_i statistic by $p_i(\cdot)$ for cross-section unit i . Maddala and Wu (1999) proposes to use Fisher's p-value combination method for finite N , that is

$$P^{MW} = -2 \sum_{i=1}^N \ln(p_i(\cdot)), \quad (4.45)$$

which has a χ^2 distribution with $2N$ degrees of freedom. Choi (2001) extends this method to an infinite N setting and proposes to use

$$P^C = \frac{-2 \sum_{i=1}^N \ln(p_i(\cdot)) - 2N}{\sqrt{4N}}, \quad (4.46)$$

which has a standard normal asymptotic distribution. In addition to the p-value combination method one can also use a group-mean estimator to test the null hypothesis stated above. The group-mean statistic is based on (standardized) averages of the individual test statistics and can be written as

$$GM = \frac{1}{\sqrt{N}} \sum_{i=1}^N \frac{LM_i - m}{\sqrt{2m}}, \quad (4.47)$$

where m is the number of regressors that are being tested for weak exogeneity. If the individual test statistics are independent over cross-section units the resulting statistic, GM will have a standard normal asymptotic distribution.

In our set-up the common factor structure introduces cross-sectional dependence and invalidates the aforementioned testing procedures. The literature suggests several methods to take into account the cross-sectional dependency in the test statistics. Pesaran (2007) proposes a method to test unit roots in panels with cross-section dependence in a single factor residual set-up. They take simple averages of the individual, cross-sectionally augmented Dickey-Fuller test statistics, then truncate the resulting statistics to avoid the effects of extreme outcomes. They show that conditional on the factor structure, the limiting distribution of the truncated and cross-sectionally augmented statistics will be independent across cross-section units. They further show that the unconditional limiting distributions exist and are free of nuisance parameters. This method is extended to a model with a multi-factor residual structure in Pesaran et al. (2013). Gengenbach et al. (2014) develops a panel data test to test for the null hypothesis of no error correction. Noting that the resulting individual test statistics are dependent across cross-section units they follow Pesaran (2007) and propose to use a truncated version of the average of the individual tests.

Another method to combine dependent test statistics is proposed by Hartung (1999). The author assumes that the dependence can be characterized by the correlation of individual statistics that are obtained by inverse normal transformation and suggests that combining the probits by using a weighting scheme yields a pooled test statistic, which has a standard normal asymptotic distribution. This method requires a single certain parametric formulation of the correlation. Demetrescu et al. (2006) show that, with a modification, the method proposed by Hartung (1999) can be made valid with more general correlation matrices and in their paper, they provide an application of this method to panel tests.

Another possibility to get around the difficulty that arises from the presence of cross-correlation between the individual test statistics is to use a bootstrap version of the pooled tests. Bootstrapping the panel unit root tests is also suggested by Maddala and Wu (1999). They show, through simulations, that under cross-section dependence the panel unit root test they propose suffers from substantial size distortions and further they show that the bootstrap version of the test does not have such size distortions. Chang (2004) applies a bootstrap methodology to unit root tests with cross-correlations and proves the validity of the method. Gonçalves

and Perron (2014) focus on bootstrapping factor augmented regression models, they propose and theoretically justify bootstrapping methods in the context of factor-augmented regression models. They prove the asymptotic validity of their bootstrapping method under similar assumptions that are considered by Bai and Ng (2006). Inspired by the work of Chang (2004), Gonçalves and Perron (2014) and Palm et al. (2007), we use the residual bootstrap procedure to deal with the cross-sectional dependency. In the rest of this section we discuss the details of the bootstrap procedure.

Bootstrap procedure

We propose to use a bootstrap inference method to test the null hypothesis stated above. As it is also mentioned above, cross-sectional dependence between the individual test statistics yields non-standard distributions for the average of the individual test statistics. However it is possible to approximate the sample distribution of the test statistic by using a bootstrap procedure.

To construct the bootstrapped tests we employ the following algorithm. For expositional simplicity we assume that the lag order for the first differences of the variables in the conditional error correction model and in the marginal model is equal to 1.

Step 1. Obtain the residuals and the coefficient estimates from the conditional model and the marginal model with the relevant lag order specification and under the null hypothesis of weak exogeneity by estimating the augmented models by OLS. Obtain the fitted residual for each cross-section unit i at time t as

$$\begin{aligned} \widehat{e}_{y.xit} = & \Delta y_{it} - \widehat{\alpha}_{yi}(y_{i,t-1} - \widehat{\beta}_{xi}\mathbf{x}_{it} - \widehat{\lambda}\bar{\mathbf{z}}_{t-1}) - \widehat{A}_{y.yi1}\Delta y_{i,t-1} - \widehat{A}_{y.xi0}\Delta \mathbf{x}_{it} \\ & - \widehat{A}_{y.xi1}\Delta \mathbf{x}_{i,t-1} - \widehat{A}_{y.fi0}\bar{\Delta \mathbf{z}}_t - \widehat{A}_{y.fi1}\bar{\Delta \mathbf{z}}_{t-1} - \dots - \widehat{A}_{y.fip_T}\bar{\Delta \mathbf{z}}_{t-p_T}, \end{aligned}$$

where $(\widehat{\alpha}_{yi}, \widehat{A}_{y.yi1}, \widehat{A}_{y.xi0}, \widehat{A}_{y.xi1}, \widehat{A}_{y.fi0}, \widehat{A}_{y.fi1}, \dots, \widehat{A}_{y.fip_T})$ are obtained by the OLS estimation of the augmented, truncated, conditional model for Δy_{it} and $(\widehat{\beta}_{xi}, \widehat{\lambda})$ are obtained by indirect least squares estimation. For details see the relevant section of Chapter 3. Under the null hypothesis of weak exogeneity, obtain the residuals for cross-section unit i and at time t from the marginal model as

$$\widehat{e}_{xit} = \Delta \mathbf{x}_{it} - \widehat{A}_{xyi1}\Delta y_{it} - \widehat{A}_{xxi1}\Delta \mathbf{x}_{i,t-1} - \widehat{A}_{xfi1}\bar{\Delta \mathbf{z}}_{t-1} - \dots - \widehat{A}_{xfip_T}\bar{\Delta \mathbf{z}}_{t-p_T},$$

where $(\widehat{\mathbf{A}}_{xyi1}, \widehat{\mathbf{A}}_{xxi1}, \widehat{\mathbf{A}}_{xfi1}, \dots, \widehat{\mathbf{A}}_{xfip_T})$ are obtained by OLS estimation of the augmented, truncated marginal model.

Step 2. To make sure that the resulting process has a zero mean, recenter the residuals for each cross-section unit i , as $\widehat{e}_{y.xit}^0 = \widehat{e}_{y.xit} - \bar{\widehat{e}}_{y.xi}$, $\widehat{e}_{xit}^0 = \widehat{e}_{xit} - \bar{\widehat{e}}_{xi}$, where $\bar{\widehat{e}}_{y.xi}$ and $\bar{\widehat{e}}_{xi}$ are the averages over time of the residuals.

Step 3. In this step the index vector for the resampling scheme needs to be generated. There are two issues involved. First of all, to preserve the cross-sectional dependence of the resulting test statistics the fitted residuals should be resampled only across the t dimension. Secondly, the truncation of the lag order for the first differences of the cross-sectional averages may induce serially correlated residuals. For this reason one may use a block bootstrap procedure. For the block bootstrap procedure with block length, bl , following Palm et al. (2011), draw j_0, \dots, j_{s-1} i.i.d. from $U[1, T - p_T - bl]$, where $s = \lfloor (T - p_T - 1)/bl \rfloor + 1$ is the number of blocks. The case of the i.i.d. bootstrap is also covered in this notation by setting $bl = 1$.

Step 4. Take the T -vector $\widehat{\mathbf{e}}_{y.xi}^0 = (\widehat{e}_{y.xi1}^0, \dots, \widehat{e}_{y.xiT}^0)'$ and $T \times m$ matrix $\widehat{\mathbf{e}}_{xi}^0 = (\widehat{e}_{xi1}^0, \dots, \widehat{e}_{xiT}^0)'$ for each i and reorder the elements of the vectors by using the index numbers that are obtained in Step 3. This can be achieved by $\mathbf{e}_{y.xit}^* = \widehat{e}_{y.xi, j_{g+r}}^0$, $\mathbf{e}_{xit}^* = \widehat{e}_{xi, j_{g+r}}^0$, where $g = \lfloor (t - p_T - 1)/bl \rfloor$ and $r = t - g \times bl - 1$.

Step 5. Denoting the bootstrap samples of y_{it} and x_{it} as y_{it}^* and x_{it}^* respectively for each cross-section unit, construct y_{it}^* and x_{it}^* by using the conditional and the marginal models with the same lag order specification as the models used in the first step, the resampled residuals obtained in the previous step and the estimated coefficients obtained in Step 1. Given the assumption of the strong exogeneity of the factors the levels and also the first differences of the factors (or their approximations) can be treated as given. Denoting the maximum of p and p_T as p_{max} , for the initial values, y_{i0}^* and x_{i0}^* , use the $(p_{max} + 1)^{th}, \dots, (2 \times p_{max} + 1)^{th}$ rows of the original data to sustain the integration order of the reconstructed version of the data.

Step 6. Compute the LM_i^* statistic and the p-value for each LM_i^* from the inverse of the cumulative distribution function of the $\chi^2(m)$ distribution for each i .

Step 7. For the group mean estimator take the cross-sectional average of the LM_i statistics by

$$\overline{LM}^* = \frac{1}{N} \sum_{i=1}^N LM_i^*.$$

For the bootstrap version of the Fisher's p-value combination test, use the formula in (4.46) to obtain the pooled statistic. This statistic is denoted by P^{C*} .

Step 8. Repeat Steps 1–8 B times to obtain the bootstrap test statistics \overline{LM}^{*b} and P^{C*b} , $b = 1, \dots, B$ and where B is the number of bootstrap replications. Select the bootstrap critical values that are equal to the $(1 - \alpha)$ -quantile of the ordered \overline{LM}^{*b} and P^{C*b} statistics, where α is the nominal value of the test. We denote the bootstrap critical values as C_{LM}^B and C_P^B for the group mean and for the Fisher-type (given in (4.46)) tests respectively.

Finally, compute the LM_i statistic using the original data for each unit $i = 1, \dots, N$. Take the average of the LM_i 's to obtain the group mean statistic \overline{LM} . Use the inverse of the distribution function of the $\chi^2(m)$ distribution for each LM_i and use the formula in (4.46) to obtain the Fisher type pooled statistic, P^C . Reject the null hypothesis of weak exogeneity if $\overline{LM} > C_{LM}^B$ for the group-mean test and if $P^C > C_P^B$ for the Fisher test.

In the next section we compare the size and power properties of the bootstrap versions of the pooled test statistics GM and P^C with the size and power properties of the original testing procedures through Monte Carlo simulations.

4.5 Monte Carlo Simulations

In this section we assess the finite sample performances of the test statistics. We consider the individual specific statistics for the individual cross-section unit, LM_i and the pooled tests given in the previous section namely the Fisher's p-value combination test and the group mean test. As discussed in the previous section, considering the dependence of the LM_i statistics we propose to use bootstrap version of the pooled tests. To assess the properties of the bootstrap method for the group-mean and the Fisher-type statistics, we provide the size and the power estimates of the bootstrap method. Furthermore, we investigate the small sample performance of the Wald test statistic.

We consider three different cases regarding the factor structure and the approximation method. As a benchmark case we provide the results for the test statistics mentioned above for the case in which the factors are not present in the DGP. In this case the cross-section units are independent. Then the estimation boils down to a standard case. We do not provide the results for the bootstrap versions of the tests for this case. In addition to that, we also provide the results for the case in which we assume that the factors are observed. In this case the dependence across cross-section units is present but observed, hence no approximation is necessary. Finally to evaluate the finite sample performance of the method proposed in this chapter, we consider the case with the unobserved common factor structure, which are approximated by the cross-sectional averages of the observed variables in a way that is proposed in the approximation section of the current chapter.

We consider two different methods for the selection of the lag order. We run estimations, first, with true lag lengths ignoring the necessity for using the truncation lag order for the first differences of the cross sectional averages. In addition to these, considering the theoretical results of the current chapter, for the case where the factors are not observed but approximated by cross-sectional averages we use a truncation lag order for the first differences of the cross-sectional averages. Following Chudik and Pesaran (2013a) we choose p_T (the truncation lag) as equal to the integer part of $T^{1/3}$, denoted by $[T^{1/3}]$.

In the next section we discuss the DGP used in the Monte Carlo simulations.

4.5.1 Data Generating Process

We generate data according to the following model. We set the number of idiosyncratic, weakly exogenous variables equal to 1 and the number of common factors, k , equal to 2. The idiosyncratic variables for each i are then generated as

$$\begin{aligned}\Delta y_{it} &= \alpha_{yi}(y_{i,t-1} - \beta_{xi}x_{i,t-1} - \lambda_i f_{t-1}) + A_{yyi}\Delta y_{i,t-1} + A_{yxi}\Delta x_{i,t-1} \\ &\quad + A_{yf1i}\Delta f_{1,t-1} + A_{yf2i}\Delta f_{2,t-1} + \varepsilon_{yit} + b\varepsilon_{xit}, \\ \Delta x_{it} &= A_{xyi}\Delta y_{i,t-1} + A_{xxi}\Delta x_{i,t-1} + A_{xf2i}\Delta f_{2,t-1} + \varepsilon_{xit}.\end{aligned}$$

The common factors are generated as

$$\begin{aligned}\Delta f_{1,t} &= A_{ff11}\Delta f_{1,t-1} + A_{ff12}\Delta f_{2,t-1} + \varepsilon_{f1t}, \\ \Delta f_{2,t} &= A_{ff21}\Delta f_{1,t-1} + A_{ff22}\Delta f_{2,t-1} + \varepsilon_{f2t}.\end{aligned}$$

We draw the error terms as $\varepsilon_{fjt} \sim IIDN(0,1)$ for $j = 1, 2$, $\varepsilon_{yit} \sim IIDN(0,1)$ and $\varepsilon_{xit} \sim IIDN(0,1)$. ε_{fjt} , ε_{yit} and ε_{xit} are generated independently of each other, whereas the latter two are generated independently across i . We set $b_i \sim U[0.1, 0.3]$.

We draw the long-run coefficients as $\beta_{xi} \sim IIDU[-1.2, -0.8]$ and $\lambda_{ji} \sim IIDU[-1.2, -0.8]$ for $j = 1, 2$. $\alpha_{yi} \sim IIDU[-1.2, -0.8]$. For size studies, we set $\alpha_{xi} = 0$, while $\alpha_{xi} = -0.15$ to assess power. As for the coefficients of the short-run dynamics in the models for the idiosyncratic variables we use

$$\begin{pmatrix} A_{yyi} & A_{xyi} \\ A_{xyi} & A_{xxi} \end{pmatrix} \sim IID \begin{pmatrix} U[-0.50, 0.50] & U[-0.50, 0.50] \\ U[-0.50, 0.50] & U[-0.50, 0.50] \end{pmatrix},$$

$$\begin{pmatrix} A_{yfi} & A_{yfi} \\ A_{xf1i} & A_{xf2i} \end{pmatrix} \sim IID \begin{pmatrix} U[-0.50, 0.50] & U[-0.50, 0.50] \\ 0 & U[-0.50, 0.50] \end{pmatrix}.$$

The parameters of the lagged first differences of the common factors in the models for Δf_{1t} and Δf_{2t} are equal to

$$\begin{pmatrix} A_{ff11} & A_{ff12} \\ 0 & A_{ff22} \end{pmatrix} = IID \begin{pmatrix} U[-0.50, 0.50] & U[-0.50, 0.50] \\ U[-0.50, 0.50] & U[-0.50, 0.50] \end{pmatrix}.$$

The coefficients of the short-run dynamics drawn by making sure that the modulus of the characteristic roots are smaller than 1. We report results for $N = 30, 50, 100, 150$ and $T = 30, 50, 100, 150$. All tests are performed at a 5% nominal level. Reported rejection frequencies are based on 1000 replications and we draw $B = 499$ bootstrap replications to obtain the bootstrap critical values.

4.5.2 Monte Carlo Results

We report the results of 4 experiments, each with different choices for the factor structure, lag order of the first differences of the factors and the block length chosen for the bootstrap procedure. The experiments are summarised in the table below.

Table 4.1: Description of Experiments

Experiment	Table	Factors	Lag order	Bootstrap
E1	4.E1	No factors	1	i.i.d.
E2	4.E2	Factors Known	1	i.i.d.
E3	4.E3	CSA	1	bl = 2
E4	4.E4	CSA	$\lfloor T^{1/3} \rfloor$	i.i.d.

“CSA” is the abbreviation for cross sectional averages

- E1.** This experiment is considered as a benchmark, to observe the size and power properties of the Lagrange multiplier test in the absence of cross-sectional dependence. The results in Table 4.E1 shows that size of the individual LM_i and the pooled versions, both the group mean and the Fisher’s p-value combination tests are close to the nominal value, even for small N and T values. In terms of size, Lagrange multiplier test performs better than the Wald test. The pooled tests have a high power as expected by the theory, since the individual statistics are distributed independently of each other in the absence of a factor structure. We observe that individual tests have low power in this set-up.
- E2.** In this experiment, we assume that the factors are observed. Since there is no approximation we expect that the proposed tests to have the correct size and high power when we use the lag order we used in the DGP and a simple residual bootstrap procedure. As we see from Table 4.E2., the results meet our expectations, in general. Especially when we compare the results with the results of E1. The individual LM_i test have size close to the nominal value for all N and T combinations and it is getting closer to 5 as N and T increase. We do not expect the simple group mean and the p-value combination methods to work, since the factor structure induces dependence on the LM_i statistics across cross-section units. This is what we see also observe in the simulation results. Especially for the Wald test the original pooled tests are oversized. We see an improvement in terms of size in the bootstrap versions of the tests. As for the power of the tests, the results provided in the continued Table 4.E2, suggest that the individual test has low power. However, we see an increase in the power as T increases, as expected. For the pooled tests, the bootstrap version of the group mean and

the p-value combination tests are at most as powerful as the original group mean and p-value combination tests, which is also expected by the theory and promising for the success of the bootstrap procedure we propose.

- E3.** In this experiment, we augment the models with cross-sectional averages of the observables in a way that is suggested in the chapter to replace the factors that are assumed to be unobserved. Noting that ignoring the necessity to add further lags of the first differences of the cross-sectional averages may induce serially correlated residuals, we use a block bootstrap procedure and set the block length equal to 2. The results of this experiment can be found in Table 4.E3. The individual LM_i and $WALD_i$ tests, in this case, are slightly oversized. However, compared to the results of the previous experiment, the increase in the size is not dramatic. The original group mean and the p-value combination tests are now substantially oversized. This outcome suggests that when the true factors are replaced by their approximations the effect of the cross-sectional dependency of the individual statistics on the pooling procedure increases, at least in small samples. Bootstrap versions of the pooled tests usually have a size close to the nominal value, which asserts that the bootstrap procedure is successful. However, as N increases we observe a slight increase in the size. As for the power properties of the tests, the individual tests have low power as in the previous experiments. Moreover, the bootstrap version of the pooled tests have low power for small T , however as T increases the power also increases. We observe that for large values of T , the bootstrap versions of the tests are as powerful as the original pooled tests.
- E4.** In this experiment we use a truncated lag order for the first differences of the cross-sectional averages in the augmented model. The truncated lag order is a function of T and the functional form is $T^{1/3}$. The results indicate a worsening of the size of the tests that are being considered. However, the results suggests that the bootstrap procedure yield an improvement in the size compared to the original pooled tests. We observe that for large values of N and T , the bootstrap versions of the tests are as powerful as the original pooled tests.

4.6 Conclusion

We consider a LM type statistic for weak exogeneity in a cointegrated panel data model with non-stationary common factors, extending results of Boswijk and Urbain (1997). If the common factors are observed, the long-run parameters can be super-consistently estimated from the individual specific conditional ECMs. This is a standard result in the literature for the models without the factors and it can be easily extended to a case with observed factors in, for example see Chapter 3. Boswijk and Urbain (1997) suggests that a test for no error correction in the marginal model will be nuisance parameter free and have standard chi-squared distributions. In this chapter we use the result provided by Chapter 3, which suggests that if the unobserved common factor structure has to be approximated, the long-run parameters can still be super-consistently estimated. Then, we show that the approximation of the factors also in the marginal model induces nuisance parameters in the T -asymptotic distributions of the statistics, for finite N . But if N and T jointly go to infinity with the restriction that $\frac{T}{N} \rightarrow 0$ and provided that lag order of the augmented first differences of the cross-sectional averages observed variables is large enough, the asymptotic distribution of the LM test statistic for the individual unit i , becomes nuisance parameter-free and converges to a χ^2 with m degrees of freedom.

The truncation of the lag order is a delicate issue. The theoretical conditions we propose to choose the lag order for the first differences of the cross-sectional averages of the observed variables are $\frac{\rho^{pT}}{T} \rightarrow 0$, $\frac{\rho^{pT}}{\sqrt{N}} \rightarrow 0$ and $\sqrt{T}\rho^{pT} \rightarrow 0$, the intuition behind these conditions is explained in Remark 4.4.10.

We also consider the pooled tests. The standard results for this case work only under the independence of the LM_i statistics. However, the presence of the common factor structure invalidates the use of standard results in our case. To overcome this problem we propose a bootstrap method to approximate the sample distributions of the pooled test statistics. We consider the bootstrap versions of a group mean statistics, which is simply the average of the individual statistics and a p-value combination statistics, which is inspired by the seminal work of Fisher (1934).

We evaluate the small sample performance of the tests we propose through Monte Carlo simulations. The results indicate that pooled tests are substantially oversized. This is due to the dependence present in the data, which is explained by a factor structure. The factor structure is still present in the LM_i statistics

and this creates a dependence across cross-section units. Bootstrap versions of the pooled tests have size close to the nominal value, which indicates that the bootstrap procedure we propose in the chapter works well in small samples. Power of the bootstrap versions of the tests are close to the original pooled tests. As for the individual tests, we observe that the size is slightly affected by approximating \mathbf{f}_t , especially when N is small.

4.A Tables

Table 4.E1: No Factors, Lag order = 1

T/N	Size							
	LM				Wald			
	30	50	100	150	30	50	100	150
	Individual							
30	5.6	4.9	5.3	5.7	7.1	7.1	6.4	7.4
50	4.7	4.5	4.5	4.8	5.9	6.1	6.0	5.6
100	5.4	4.9	5.3	6.1	6.1	5.4	5.8	6.5
150	7.2	5.7	5.4	4.9	7.4	5.8	5.8	5.5
	Group Mean							
30	5.3	5.3	6.5	5.4	15.2	17.6	27.0	35.0
50	4.7	6.3	5.0	5.5	9.2	13.1	13.0	13.9
100	5.3	5.0	5.4	5.1	7.8	6.9	7.6	8.7
150	4.1	4.1	4.5	4.8	5.4	5.5	5.3	6.8
	p-value combination							
30	5.4	5.6	6.9	6.4	14.6	16.3	25.1	30.9
50	4.5	6.5	5.3	6.0	8.4	12.1	12.1	12.9
100	4.8	4.8	5.6	4.9	6.9	6.2	7.7	8.1
150	4.3	4.2	4.7	4.9	5.3	5.4	5.6	6.0

Table 4.E1 cont'd: No Factors, Lag order = 1

T/N	Power							
	LM				Wald			
	30	50	100	150	30	50	100	150
	Individual							
30	10.3	11.1	11.6	12.1	13.4	14.0	14.4	14.5
50	15.1	18.8	18.1	18.7	16.8	21.3	20.0	20.6
100	33.8	33.7	31.0	34.9	34.9	34.5	32.7	36.2
150	45.6	46.7	48.6	44.1	46.2	47.6	49.3	45.6
	Group Mean							
30	30.0	43.2	67.6	81.3	53.1	71.5	90.8	97.3
50	67.7	85.8	97.5	99.5	79.6	92.7	99.2	100.0
100	99.5	100.0	100.0	100.0	99.5	100.0	100.0	100.0
150	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
	P-value Combination							
30	30.0	43.5	68.2	82.5	50.9	69.3	90.3	96.5
50	68.7	86.1	98.1	99.6	78.8	92.0	99.2	100.0
100	99.2	100.0	100.0	100.0	99.5	100.0	100.0	100.0
150	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0

Notes: LM and Wald refer to the Lagrange multiplier and Wald tests respectively. The results for the 'Individual', 'group mean' and 'p-value combination' tests are obtained based on the formulas provided in (4.28), (4.46) and (4.47) respectively.

Table 4.E2: True Factors, i.i.d. Bootstrap, Lag order = 1

	Size							
	LM				Wald			
T/N	30	50	100	150	30	50	100	150
	Individual							
30	6.5	8.2	5.1	6.5	8.3	10.4	8.0	8.9
50	7.1	5.2	6.8	5.0	7.6	6.8	7.4	5.8
100	7.0	5.1	5.7	3.8	7.3	5.7	6.2	4.4
150	5.1	5.4	5.3	5.6	5.3	5.8	5.4	6.0
	Group Mean Bootstrap							
30	4.4	4.6	4.7	3.5	4.1	4.7	4.7	3.7
50	6.1	4.3	4.5	5.6	5.4	4.3	5.1	5.6
100	4.8	5.5	5.0	5.0	4.9	5.7	4.9	4.8
150	4.4	5.1	4.7	6.1	4.4	5.0	4.6	5.9
	P-value Combination Bootstrap							
30	3.3	5.0	4.6	4.2	3.7	4.8	4.6	4.2
50	5.7	4.9	4.8	4.7	5.8	4.8	4.9	5.2
100	4.7	5.7	5.6	4.9	4.5	5.6	5.7	4.8
150	4.5	4.8	4.5	5.9	4.5	4.8	4.5	5.8
	Group Mean							
30	11.1	10.3	16.7	19.3	25.7	30.8	50.7	61.8
50	7.0	5.8	6.4	8.4	12.8	13.2	17.9	23.2
100	5.1	6.6	5.2	5.7	6.8	8.8	8.0	9.6
150	4.3	4.4	4.2	6.9	5.3	6.0	5.8	8.2
	P-value Combination							
30	10.5	10.6	17.8	20.3	22.7	27.8	45.6	57.2
50	6.8	5.6	6.8	8.7	11.6	11.9	16.3	20.9
100	5.3	6.7	5.4	5.9	6.6	8.2	7.3	9.0
150	4.3	4.5	4.2	6.1	5.1	5.3	5.5	8.2

Table 4.E2 cont'd: True Factors, i.i.d. Bootstrap, Lag order= 1

T/N	Power							
	LM				Wald			
	30	50	100	150	30	50	100	150
	Individual							
30	8.4	7.9	8.2	7.1	10.5	9.8	10.4	9.1
50	10.2	11.4	12.4	9.9	11.4	12.5	13.7	11.1
100	18.9	21.8	18.9	18.4	19.9	22.7	19.7	19.5
150	29.1	27.2	28.7	27.4	30.4	28.4	29.4	28.4
	Group Mean Bootstrap							
30	15.3	22.5	38.5	53.2	11.5	18.9	32.2	46.2
50	59.5	80.4	98.6	100.0	57.4	78.5	98.3	100.0
100	99.2	100.0	100.0	100.0	98.8	100.0	100.0	100.0
150	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
	P-value Combination Bootstrap							
30	14.8	23.0	39.9	53.3	12.2	19.6	34.2	48.0
50	60.9	81.3	98.7	100.0	58.8	80.2	98.5	99.9
100	99.4	100.0	100.0	100.0	99.3	100.0	100.0	100.0
150	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
	Group Mean							
30	39.2	56.2	82.8	93.2	63.8	79.3	96.5	99.2
50	74.6	91.4	99.8	100.0	83.9	96.1	100.0	100.0
100	99.5	100.0	100.0	100.0	99.5	100.0	100.0	100.0
150	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
	P-value Combination							
30	39.8	55.2	83.7	93.7	59.6	77.8	95.7	99.2
50	74.4	91.9	99.8	100.0	83.4	96.2	100.0	100.0
100	99.5	100.0	100.0	100.0	99.5	100.0	100.0	100.0
150	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0

Notes: 'Group Mean Bootstrap' and 'P-value Combination Bootstrap' are based on the bootstrap method presented in Section 4.4. See Table 4.E1 for an explanation of the rest.

Table 4.E3: CSA, Block length = 2, Lag order = 1

T/N	Size							
	LM				Wald			
	30	50	100	150	30	50	100	150
	Individual							
30	6.1	7.4	6.1	7.1	9.3	9.2	7.7	9.0
50	5.6	6.7	6.5	6.4	6.6	7.6	7.5	7.1
100	7.4	6.1	7.1	6.4	8.1	6.6	7.8	6.9
150	6.7	7.6	7.0	7.3	7.2	7.7	7.5	7.9
	Group Mean Bootstrap							
30	5.5	5.0	5.9	5.5	5.2	5.1	7.2	5.9
50	4.1	5.9	5.6	6.9	4.0	5.7	5.5	7.7
100	4.2	4.3	6.2	7.6	4.0	4.5	6.3	7.4
150	5.1	5.0	8.6	8.9	5.1	4.8	8.9	8.7
	P-value Combination Bootstrap							
30	5.0	4.8	6.1	5.5	5.3	5.4	6.7	6.9
50	4.5	6.0	6.0	7.1	4.2	5.9	5.7	7.3
100	4.1	5.0	6.3	7.4	4.0	5.0	6.2	7.4
150	5.1	5.4	8.1	9.5	5.3	5.3	8.1	9.3
	Group Mean							
30	11.1	12.7	18.3	26.9	27.1	34.0	51.5	67.1
50	7.3	10.0	10.6	14.7	12.5	16.1	25.5	30.9
100	9.4	11.4	15.1	16.0	12.6	15.6	20.3	22.6
150	10.6	12.5	19.3	18.50	12.5	14.7	23.7	24.1
	P-value Combination							
30	11.7	12.5	19.2	27.7	24.9	31.5	47.5	63.7
50	7.0	9.9	10.7	14.3	11.6	16.1	22.1	28.5
100	8.4	11.1	15.6	15.9	11.7	14.4	19.2	21.3
150	10.5	12.7	19.5	17.7	12.5	14.7	23.1	21.7

Table 4.E3 cont'd: CSA, Block length = 2, Lag order = 1

T/N	Power							
	LM				Wald			
	30	50	100	150	30	50	100	150
	Individual							
30	9.1	8.3	8.5	8.1	11.1	10.2	10.2	10.0
50	10.6	8.8	10.2	9.2	12.3	10.5	11.6	11.0
100	15.2	17.5	15.1	15.5	16.0	18.0	16.0	16.1
150	19.7	20.6	21.5	22.3	20.7	21.4	22.2	23.2
	Group Mean Bootstrap							
30	5.7	6.9	8.9	10.3	4.7	5.5	7.6	7.7
50	22.3	32.3	64.4	77.5	21.0	28.9	61.1	75.6
100	82.0	96.8	100.0	100.0	81.2	96.4	100.0	100.0
150	99.1	100.0	100.0	100.0	99.0	100.0	100.0	100.0
	P-value Combination Bootstrap							
30	5.5	6.6	8.9	9.7	4.6	5.6	7.3	8.1
50	23.0	34.0	65.1	77.9	21.7	32.3	64.1	76.4
100	83.2	97.0	100.0	100.0	83.0	96.9	100.0	100.0
150	99.2	100.0	100.0	100.0	99.2	100.0	100.0	100.0
	Group Mean							
30	26.6	39.1	65.9	79.7	48.8	66.1	90.4	96.9
50	51.7	68.6	94.6	99.2	64.2	81.9	98.4	99.8
100	93.2	99.4	100.0	100.0	95.3	99.7	100.0	100.0
150	99.6	100.0	100.0	100.0	99.9	100.0	100.0	100.0
	P-value Combination							
30	26.4	38.8	66.1	79.8	45.6	64.0	89.2	96.0
50	51.4	70.8	94.6	99.2	61.6	80.6	98.1	99.8
100	93.8	99.5	100.0	100.0	95.1	99.7	100.0	100.0
150	99.7	100.0	100.0	100.0	99.9	100.0	100.0	100.0

Notes: See Table 4.E2 for an explanation.

Table 4.E4: CSA, i.i.d. Bootstrap, Lag order = $T^{1/3}$

T/N	Size							
	LM				Wald			
	30	50	100	150	30	50	100	150
Individual								
30	9.5	10.7	9.9	10.4	12.4	13.8	12.7	13.7
50	6.9	7.5	7.5	7.7	8.3	9.0	9.2	8.7
100	8.3	7.3	7.6	5.8	9.3	8.1	8.1	6.6
150	8.5	8.1	8.4	8.0	9.0	8.3	8.6	8.5
Group Mean Bootstrap								
30	5.8	6.6	9.6	11.4	6.6	7.8	13.1	16.2
50	5.6	5.6	6.2	6.3	5.3	5.7	5.8	6.4
100	6.4	8.5	10.8	11.7	6.4	8.5	11.0	11.6
150	8.7	11.0	16.5	16.5	8.6	10.9	16.2	16.8
P-value Combination Bootstrap								
30	5.5	6.9	9.1	12.1	6.4	8.4	12.6	16.2
50	5.0	5.6	6.0	6.5	5.4	5.5	6.0	6.2
100	7.2	8.5	10.6	11.9	7.0	8.6	10.8	11.5
150	9.1	11.4	17.4	16.3	9.1	11.3	17.1	16.5
Group Mean								
30	46.3	57.5	81.6	92.4	69.6	82.9	96.3	99.5
50	13.2	19.0	29.8	37.3	23.7	32.0	49.4	62.2
100	14.2	16.7	25.3	29.0	18.5	22.8	33.7	38.0
150	14.4	18.6	30.0	33.0	17.3	23.0	34.9	40.7
P-value Combination								
30	45.4	56.6	81.9	92.3	66.5	80.8	96.1	99.4
50	12.7	18.6	28.9	37.4	21.8	29.2	47.1	58.4
100	13.3	16.5	25.1	27.8	17.0	21.0	31.8	36.5
150	14.3	17.6	28.7	32.1	15.9	20.3	33.4	38.3

Table 4.E4 cont'd: CSA, i.i.d. Bootstrap, Lag order= $\lfloor T^{1/3} \rfloor$

	Power							
	LM				Wald			
T/N	30	50	100	150	30	50	100	150
	Individual							
30	12.4	11.3	12.0	13.3	15.8	14.6	15.2	15.7
50	10.5	10.1	10.8	9.3	11.8	11.4	11.5	11.0
100	14.4	15.0	17.5	13.1	14.9	15.7	18.4	14.0
150	20.1	20.7	22.4	21.9	20.6	21.3	23.1	22.5
	Group Mean Bootstrap							
30	5	3.8	3.9	2.5	3.7	3.6	3.6	3.6
50	22.5	28.2	57.1	68.8	20.7	25.7	52.4	63.7
100	82.1	95.7	99.9	100.0	81.9	95.0	99.9	100.0
150	98.4	99.9	100.0	100.0	98.3	99.9	100.0	100.0
	P-value Combination Bootstrap							
30	5.2	3.7	3.6	2.4	3.6	3.1	3.8	3.5
50	22.4	28.8	57.8	69	21.6	27	53.8	65.1
100	84.1	96.0	99.9	100.0	83.1	95.7	99.9	100.0
150	98.5	99.9	100.0	100.0	98.5	99.9	100.0	100.0
	Group Mean							
30	56.8	74.1	95.2	99.5	80.2	91.4	99.6	100
50	54.5	72.1	94.5	97.7	67.5	83.4	98.4	99.4
100	93.0	98.4	100.0	100.0	95.1	99.3	100.0	100.0
150	99.5	100.0	100.0	100.0	99.6	100.0	100.0	100.0
	P-value Combination							
30	56.1	72.9	95.1	99.1	77.3	90.9	99.5	100
50	55.2	70.8	94.5	98.9	66.3	82.9	98.1	99.4
100	93.0	98.5	100.0	100.0	94.5	99.2	100.0	100.0
150	99.6	100.0	100.0	100.0	99.6	100.0	100.0	100.0

Notes: See Table 4.E2 for an explanation.

4.B Appendix

4.B.1 Preliminary Results

In this section we provide some preliminary results to be used in the proofs of the lemmas provided in the chapter. The lemmas established in this section involve the limits and/or the rates of convergence of the cross product terms appearing in the expressions we come across while proving Lemmas 4.4.1-4.4.3. The preliminary results presented in this section are grouped in eight lemmas. The first five lemmas, B1-B5 are related to the cross products that involve the observed processes, the approximation error, truncation errors. For example in Lemma 4.B.1 we provide the orders of the cross products of the levels of the observed variables only, whereas Lemma 4.B.2 deals with the limit results of the cross products involving the approximation errors $\mathbf{V}_{A,-1}$, $\mathbf{V}_{B,-1}$. In Lemmas 4.B.3-4.B.4 we present the rates of convergence of the cross products that involve the truncation errors. 4.B.5 we present the order results for the cross products involving the difference between the true error correction term and the estimated error correction term corrected for the first differences of the cross-sectional averages of the observed variables. The last three lemmas, B6-B8, deals with the orthogonal projection matrix that appears in the LM_i estimator.

Lemma 4.B.1.

$$\frac{\mathbf{X}'_{i,-1}\mathbf{X}_{i,-1}}{T} = O_p(T), \quad (4.B.1)$$

$$\frac{\mathbf{X}'_{i,-1}\bar{\mathbf{Z}}_{-1}}{T} = O_p(T), \quad (4.B.2)$$

$$\frac{\mathbf{U}'_{i,-1}\mathbf{X}_{i,-1}}{T} = O_p(1), \quad (4.B.3)$$

$$\frac{\mathbf{U}'_{i,-1}\bar{\mathbf{Z}}_{-1}}{T} = O_p(1), \quad (4.B.4)$$

$$\frac{\bar{\mathbf{Z}}'_{-1}\bar{\mathbf{Z}}_{-1}}{T} = O_p(T), \quad (4.B.5)$$

$$\frac{\mathbf{X}'_{i,-1}\boldsymbol{\varepsilon}_{xi}}{T} = O_p(1), \quad (4.B.6)$$

$$\frac{\bar{\mathbf{Z}}'_{-1}\boldsymbol{\varepsilon}_{xi}}{T} = O_p(1). \quad (4.B.7)$$

for all $i = 1, \dots, N$.

Proof of Lemma 4.B.1. Given the properties of the processes involved in the above cross products, the results stated above are the results in the standard literature for stationary and non-stationary processes. Hence we do not provide the proofs here. ■

The lemma below provides the results we use in dealing with the terms appearing in the test statistics due to the approximation of the unobserved common factors. First note the definitions for the approximation errors which are obtained in Section 4.4.2.

$$\mathbf{v}_{At} = \bar{\mathbf{z}}_t - \mathbf{H}'_1 \mathbf{f}_t - \mathbf{H}_3(L) \overline{\Delta \mathbf{z}}_t,$$

where \mathbf{v}_{A1t} is defined in (4.20). Furthermore we have

$$\mathbf{v}_{Bt} = \overline{\Delta \mathbf{z}}_t - \mathbf{H}'_4(L) \Delta \mathbf{f}_t = \frac{1}{N} \sum_{i=1}^N (\mathbf{C}_{zzi} + \mathbf{C}^*_{zzi}(L)(1-L)) \boldsymbol{\varepsilon}_{zit}.$$

We stack the time dimension of these vectors as $\mathbf{V}_{A,-1} = (\mathbf{v}_{A,p_T+1}, \dots, \mathbf{v}_{A,T-1})'$, $\mathbf{V}_{B,-1} = (\mathbf{v}_{B,p_T+1}, \dots, \mathbf{v}_{B,T-1})'$.

Lemma 4.B.2.

$$\frac{\mathbf{X}'_{i,-1} \mathbf{V}_{A,-1}}{T} = O_p \left(\frac{T}{\sqrt{N}} \right), \quad (4.B.8)$$

$$\frac{\bar{\mathbf{z}}'_{-1} \mathbf{V}_{A,-1}}{T} = O_p \left(\frac{T}{\sqrt{N}} \right), \quad (4.B.9)$$

$$\frac{\mathbf{V}'_{A,-1} \mathbf{V}_{A,-1}}{T} = O_p \left(\frac{T}{N} \right), \quad (4.B.10)$$

$$\frac{\mathbf{V}'_{A,-1} \boldsymbol{\varepsilon}_{xi}}{\sqrt{T}} = O_p \left(\frac{\sqrt{T}}{\sqrt{N}} \right), \quad (4.B.11)$$

$$\frac{\mathbf{U}'_{i,-1} \mathbf{V}_{A,-1}}{T} = O_p \left(\frac{1}{\sqrt{N}} \right), \quad (4.B.12)$$

$$\frac{\mathbf{U}'_{i,-1} \mathbf{V}_{B,-1}}{\sqrt{T}} = O_p \left(\frac{\sqrt{T}}{\sqrt{N}} \right), \quad (4.B.13)$$

for all $i = 1, \dots, N$.

Proof of Lemma 4.B.2. Proofs of (4.B.8), (4.B.10), (4.B.13) are provided in Chapter 3. For the rest one can follow similar steps to prove the results. ■

Now we consider the difference between the estimated error correction term obtained by augmenting the conditional model by cross-sectional averages and the true error correction term. We write

$$\widehat{\mathbf{U}}_{i,-1} = \mathbf{U}_{i,-1} + (\widehat{\mathbf{U}}_{i,-1} - \mathbf{U}_{i,-1}),$$

and by using the definitions of $\widehat{\mathbf{U}}_{i,-1}$ and $\mathbf{U}_{i,-1}$ given in (4.26) and (4.25), the second term of the right-hand side can be written as

$$\begin{aligned} \widehat{\mathbf{U}}_{i,-1} &= \mathbf{U}_{i,-1} + \mathbf{X}_{i,-1}(\boldsymbol{\beta}_{xi} - \widehat{\boldsymbol{\beta}}_{xi}) + \overline{\mathbf{Z}}_{-1}(\mathbf{H}_1^- \boldsymbol{\lambda}_i - \widehat{\mathbf{H}}_1^- \boldsymbol{\lambda}_i) - \overline{\Delta \mathbf{Z}}_{-1} \mathbf{H}_3'(L) \mathbf{H}_1^- \boldsymbol{\lambda}_i \\ &\quad - \mathbf{V}_{A,-1} \mathbf{H}_1^- \boldsymbol{\lambda}_i, \end{aligned} \quad (4.B.14)$$

where the t^{th} row of $\mathbf{V}_{A,-1}$ is defined in (4.20). Considering the fourth term on the right-hand side of (4.B.14) and the orthogonal projection matrix $\widetilde{\mathbf{M}}_i$ that is defined in (4.31) we can write

$$\widetilde{\mathbf{M}}_i \overline{\Delta \mathbf{Z}}_{-1} \mathbf{H}_3'(L) \mathbf{H}_1^- \boldsymbol{\lambda}_i = \widetilde{\mathbf{M}}_i \mathbf{V}_{C,-p_T}, \quad (4.B.15)$$

where $\mathbf{V}_{C,-p_T}$ is the part which is not orthogonal to the space spanned by $\widetilde{\mathbf{G}}_i$ and can be defined as

$$\mathbf{V}_{C,-p_T} = \begin{pmatrix} (\sum_{\ell=p_T+1}^{\infty} \mathbf{H}_{3\ell} \overline{\Delta \mathbf{z}}_{p_T+1-\ell})' \\ (\sum_{\ell=p_T+1}^{\infty} \mathbf{H}_{3\ell} \overline{\Delta \mathbf{z}}_{p_T+2-\ell})' \\ \vdots \\ (\sum_{\ell=p_T+1}^{\infty} \mathbf{H}_{3\ell} \overline{\Delta \mathbf{z}}_{T-\ell})' \end{pmatrix}, \quad (4.B.16)$$

where $\mathbf{H}_{3\ell}$ is the coefficient of the infinite order lag polynomial $\mathbf{H}_3(L) = \sum_{\ell=0}^{\infty} \mathbf{H}_{3\ell} L^\ell$ for the ℓ^{th} lag order. We state the limit results necessary for the cross products involving $\mathbf{V}_{C,-p_T}$ in the following lemma.

In Section 4.4.2 we define an exponential decay rate for the coefficients of the lag-polynomial $\mathbf{A}_{xfi}^{*'}(L)$, that is ρ_A^ℓ . It can be shown that $\mathbf{H}_3(L)$ also has exponentially decaying coefficients and similarly we can define the exponential decay rate as $|\mathbf{H}_{3\ell}| < K \rho_H^\ell$, where $K < \infty$ and $\rho_H < 1$ are real positive constants.

Lemma 4.B.3. *Noting the definition for $\mathbf{V}_{C,-p_T}$ given in (4.B.16), we have*

$$\frac{\mathbf{U}'_{i,-1} \mathbf{V}_{C,-p_T}}{T} = O_p(\rho_H^{p_T}), \quad (4.B.17)$$

$$\frac{\mathbf{V}'_{C,-p_T} \mathbf{V}_{C,-p_T}}{T} = O_p(\rho_H^{2p_T}), \quad (4.B.18)$$

$$\frac{\mathbf{V}'_{C,-p_T} \boldsymbol{\varepsilon}_{xi}}{\sqrt{T}} = O_p(\rho_H^{p_T}), \quad (4.B.19)$$

$$\frac{\mathbf{V}'_{C,-p_T} \mathbf{V}_{B,-1}}{\sqrt{T}} = O_p\left(\frac{\sqrt{T}\rho_H^{p_T}}{\sqrt{N}}\right), \quad (4.B.20)$$

for all $i = 1, \dots, N$.

Proof of Lemma 4.B.3: It can be shown that the coefficients of the infinite order lag-polynomial $\mathbf{H}_3(L)$ are exponentially decaying. (See Chapter 3 for a detailed discussion.) We use this property in the proofs. The proofs are provided in Chapter 3 and will not be repeated here. ■

The second infinite order lag polynomial we encounter in the proof of the theorem is $\mathbf{A}_{xfi}^{**}(L)$, which is defined subsequent to (4.33). As it is also mentioned in Section 4.4.2 it can be shown that $\mathbf{A}_{xfi}^{**}(L)$ has exponentially decaying coefficients, by using similar arguments as in Chapter 3. This lag-polynomial appears in the error term, which can be written under the null hypothesis as

$$\tilde{\boldsymbol{\varepsilon}}_{xi} = \boldsymbol{\varepsilon}_{xi} - \mathbf{V}_{B,-1} \mathbf{H}_4^-(L) \mathbf{A}'_{xfi}(L) + \overline{\Delta \mathbf{Z}}_{-p_T} \mathbf{A}_{xfi}^{**'}(L).$$

Using the definition $\mathbf{A}_{xfi}^{**}(L) = \sum_{\ell=0}^{\infty} \mathbf{A}_{xfi,\ell+p_T}^* L^\ell$, where $\mathbf{A}_{xfi}^{*'}(L) = \sum_{\ell=0}^{\infty} \mathbf{A}_{xfi\ell}^{*'} L^\ell = \mathbf{H}_4^-(L) \mathbf{A}'_{xfi}(L)$ under the null hypothesis of weak exogeneity we define

$$\mathbf{V}_{(T-p_T) \times (m+1)}^{Ei,-p_T} = \begin{pmatrix} (\sum_{\ell=p_T+1}^{\infty} \mathbf{A}_{xfi\ell}^* \overline{\Delta \mathbf{z}}_{p_T+1-\ell})' \\ (\sum_{\ell=p_T+1}^{\infty} \mathbf{A}_{xfi\ell}^* \overline{\Delta \mathbf{z}}_{p_T+2-\ell})' \\ \vdots \\ (\sum_{\ell=p_T+1}^{\infty} \mathbf{A}_{xfi\ell}^* \overline{\Delta \mathbf{z}}_{T-\ell})' \end{pmatrix}, \quad (4.B.21)$$

the error term becomes,

$$\tilde{\boldsymbol{\varepsilon}}_{xi} = \boldsymbol{\varepsilon}_{xi} - \mathbf{V}_{B,-1} \mathbf{H}_4^-(L) \mathbf{A}'_{xfi}(L) + \mathbf{V}_{Ei,-p_T}.$$

Below we provide the probability orders of the cross products involving $\mathbf{V}_{Ei,-p_T}$.

Lemma 4.B.4. *Noting the definition of $\mathbf{V}_{Ei,-p_T}$ provided in (4.B.21), we have*

$$\frac{\mathbf{U}'_{i,-1} \mathbf{V}_{Ei,-p_T}}{\sqrt{T}} = O_p(\sqrt{T}\rho_A^{p_T}), \quad (4.B.22)$$

$$\frac{\mathbf{V}'_{C,-p_T} \mathbf{V}_{Ei,-p_T}}{\sqrt{T}} = O_p(\sqrt{T}\rho_A^{2p_T}), \quad (4.B.23)$$

for all $i = 1, \dots, N$.

Proof of Lemma 4.B.4: A similar result is derived in Chapter 3 for slightly different processes. The derivation can be easily adapted, hence we refer to Chapter 3 for the details. ■

Next, considering (4.B.14) we define

$$\mathbf{d}_{i,-1}^u = \widehat{\mathbf{U}}_{i,-1} - \mathbf{U}_{i,-1} + \overline{\Delta \mathbf{Z}}_{-1} \mathbf{H}'_3(L) \mathbf{H}_1^- \boldsymbol{\lambda}_i, \quad (4.B.24)$$

then we can write

$$\mathbf{d}_{i,-1}^u = \mathbf{X}_{i,-1} (\boldsymbol{\beta}_{xi} - \widehat{\boldsymbol{\beta}}_{xi}) + \overline{\mathbf{Z}}_{-1} (\mathbf{H}_1^- \boldsymbol{\lambda}_i - \widehat{\mathbf{H}}_1^- \boldsymbol{\lambda}_i) - \mathbf{V}_{A,-1} \mathbf{H}_1^- \boldsymbol{\lambda}_i.$$

The left-hand side of the equation above is encountered frequently in the proofs. In the following lemma we provide probability orders of the cross products that appear in the test statistic, which involve $\mathbf{d}_{i,-1}^u$.

Lemma 4.B.5.

$$\frac{\mathbf{U}'_{i,-1} \mathbf{d}_{i,-1}^u}{T} = O_p \left(\frac{1}{T} \right) + O_p \left(\frac{1}{\sqrt{N}} \right), \quad (4.B.25)$$

$$\frac{\boldsymbol{\varepsilon}'_{xi} \mathbf{d}_{i,-1}^u}{\sqrt{T}} = O_p \left(\frac{1}{\sqrt{T}} \right) + O_p \left(\frac{\sqrt{T}}{\sqrt{N}} \right), \quad (4.B.26)$$

$$\frac{\mathbf{V}'_{C,-pT} \mathbf{d}_{i,-1}^u}{T} = O_p \left(\frac{\rho_H^{pT}}{T} \right) + O_p \left(\frac{\rho_H^{pT}}{\sqrt{N}} \right), \quad (4.B.27)$$

$$\frac{\mathbf{V}'_{Ei,-pT} \mathbf{d}_{i,-1}^u}{\sqrt{T}} = O_p \left(\frac{\rho_A^{pT}}{\sqrt{T}} \right) + O_p \left(\frac{\sqrt{T} \rho_A^{pT}}{\sqrt{N}} \right), \quad (4.B.28)$$

$$\frac{\mathbf{d}_{i,-1}^{u'} \mathbf{d}_{i,-1}^u}{T} = O_p \left(\frac{T}{N} \right) + O_p \left(\frac{1}{T} \right) + O_p \left(\frac{1}{\sqrt{N}} \right), \quad (4.B.29)$$

$$\frac{\mathbf{V}'_{B,-1} \mathbf{d}_{i,-1}^u}{\sqrt{T}} = O_p \left(\frac{1}{\sqrt{TN}} \right) + O_p \left(\frac{1}{T} \right) + O_p \left(\frac{\sqrt{T}}{N} \right), \quad (4.B.30)$$

for all $i = 1, \dots, N$.

Proof of Lemma 4.B.5 In order to prove (4.B.25) we use the triangular inequality and the Cauchy-Schwarz inequality that yields

$$\begin{aligned} \|T^{-1} \mathbf{U}'_{i,-1} \mathbf{d}_{i,-1}^u\| &\leq \|T^{-1} \mathbf{U}'_{i,-1} \mathbf{X}_{i,-1}\| \|(\boldsymbol{\beta}_{xi} - \widehat{\boldsymbol{\beta}}_{xi})\| \\ &\quad + \|T^{-1} \mathbf{U}'_{i,-1} \overline{\mathbf{Z}}_{-1}\| \|(\mathbf{H}_1^- \boldsymbol{\lambda}_i - \widehat{\mathbf{H}}_1^- \boldsymbol{\lambda}_i)\| \end{aligned}$$

$$\begin{aligned}
& -\|T^{-1}\mathbf{U}'_{i,-1}\mathbf{V}_{A,-1}\|\|\mathbf{H}_1^-\boldsymbol{\lambda}_i\| \\
& = O_p(T^{-1}) + O_p(N^{-1/2}).
\end{aligned}$$

Using the result stated in Lemma 4.4.1, (4.B.3) and (4.B.4), the orders of the first and the second term can be shown to be equal to $O_p(T^{-1}) + O_p(N^{-1/2})$ as $(N, T, p_T) \rightarrow \infty$ and $\frac{p_T^2}{T^2} \rightarrow 0$. For the third term we use (4.B.12) and the boundedness of the product $\mathbf{H}_1^-\boldsymbol{\lambda}_i$ to show that the order is $O_p(N^{-1/2})$. Combining these results gives the required result.

The proof of (4.B.26) follows similarly to the proof of (4.B.25). Using the Cauchy-Schwarz inequality and the triangular inequality, we obtain

$$\begin{aligned}
\|T^{-1/2}\boldsymbol{\varepsilon}'_{xi}\mathbf{d}_{i,-1}^u\| & \leq \|T^{-1/2}\boldsymbol{\varepsilon}'_{xi}\mathbf{X}_{i,-1}\|\|(\boldsymbol{\beta}_{xi} - \widehat{\boldsymbol{\beta}}_{xi})\| \\
& \quad + \|T^{-1/2}\boldsymbol{\varepsilon}'_{xi}\bar{\mathbf{Z}}_{-1}\|\|(\mathbf{H}_1^-\boldsymbol{\lambda}_i - \widehat{\mathbf{H}}_1^-\boldsymbol{\lambda}_i)\| \\
& \quad - \|T^{-1/2}\boldsymbol{\varepsilon}'_{xi}\mathbf{V}_{A,-1}\|\|\mathbf{H}_1^-\boldsymbol{\lambda}_i\| \\
& = O_p(T^{-1/2}) + O_p(T^{1/2}N^{-1/2}),
\end{aligned}$$

as $\frac{p_T^2}{T^{3/2}} \rightarrow 0$. The result is obtained using the results in Lemma 4.4.1, (4.B.6), (4.B.7), (4.B.11) and the boundedness of the product $\mathbf{H}_1^-\boldsymbol{\lambda}_i$. (4.B.27), (4.B.28) can be proven in a similar way using the related preliminary results.

To prove (4.B.29), using (4.B.25) rewrite the expression as

$$\begin{aligned}
\|T^{-1}\mathbf{d}_{i,-1}^{u'}\mathbf{d}_{i,-1}^u\| & \leq \|(\boldsymbol{\beta}_{xi} - \widehat{\boldsymbol{\beta}}_{xi})\|\|T^{-1}\mathbf{X}'_{i,-1}\mathbf{X}_{i,-1}\|\|(\boldsymbol{\beta}_{xi} - \widehat{\boldsymbol{\beta}}_{xi})\| \\
& \quad + \|(\boldsymbol{\beta}_{xi} - \widehat{\boldsymbol{\beta}}_{xi})\|\|T^{-1}\mathbf{X}'_{i,-1}\bar{\mathbf{Z}}_{-1}\|\|(\mathbf{H}_1^-\boldsymbol{\lambda}_i - \widehat{\mathbf{H}}_1^-\boldsymbol{\lambda}_i)\| \\
& \quad - \|(\boldsymbol{\beta}_{xi} - \widehat{\boldsymbol{\beta}}_{xi})\|\|T^{-1}\mathbf{X}'_{i,-1}\mathbf{V}_{A,-1}\|\|\mathbf{H}_1^-\boldsymbol{\lambda}_i\| \\
& \quad + \|(\mathbf{H}_1^-\boldsymbol{\lambda}_i - \widehat{\mathbf{H}}_1^-\boldsymbol{\lambda}_i)\|\|T^{-1}\bar{\mathbf{Z}}'_{-1}\mathbf{X}_{i,-1}\|\|(\boldsymbol{\beta}_{xi} - \widehat{\boldsymbol{\beta}}_{xi})\| \\
& \quad + \|(\mathbf{H}_1^-\boldsymbol{\lambda}_i - \widehat{\mathbf{H}}_1^-\boldsymbol{\lambda}_i)\|\|T^{-1}\bar{\mathbf{Z}}'_{-1}\bar{\mathbf{Z}}_{-1}\|\|(\mathbf{H}_1^-\boldsymbol{\lambda}_i - \widehat{\mathbf{H}}_1^-\boldsymbol{\lambda}_i)\| \\
& \quad - \|(\mathbf{H}_1^-\boldsymbol{\lambda}_i - \widehat{\mathbf{H}}_1^-\boldsymbol{\lambda}_i)\|\|T^{-1}\bar{\mathbf{Z}}'_{-1}\mathbf{V}_{A,-1}\|\|\mathbf{H}_1^-\boldsymbol{\lambda}_i\| \\
& \quad - \|\mathbf{H}_1^-\boldsymbol{\lambda}_i\|\|T^{-1}\mathbf{V}'_{A,-1}\mathbf{X}_{i,-1}\|\|(\boldsymbol{\beta}_{xi} - \widehat{\boldsymbol{\beta}}_{xi})\| \\
& \quad - \|\mathbf{H}_1^-\boldsymbol{\lambda}_i\|\|T^{-1}\mathbf{V}'_{A,-1}\bar{\mathbf{Z}}_{-1}\|\|(\mathbf{H}_1^-\boldsymbol{\lambda}_i - \widehat{\mathbf{H}}_1^-\boldsymbol{\lambda}_i)\| \\
& \quad + \|\mathbf{H}_1^-\boldsymbol{\lambda}_i\|\|T^{-1}\mathbf{V}'_{A,-1}\mathbf{V}_{A,-1}\|\|\mathbf{H}_1^-\boldsymbol{\lambda}_i\| \\
& = O_p(TN^{-1}) + O_p(T^{-1}) + O_p(N^{-1/2}),
\end{aligned}$$

as $\frac{p_T^2}{TN^{1/2}} \rightarrow 0$ and $\frac{p_T^2}{T^{3/2}} \rightarrow 0$. The final order is obtained by using the results given in Lemma 4.4.1, (4.B.1), (4.B.2), (4.B.5), (4.B.8), (4.B.9), (4.B.10) and by noting that $\mathbf{H}_1^- \boldsymbol{\lambda}_i$ is bounded.

The orthogonal projection matrix $\ddot{\mathbf{M}}_i = \mathbf{I} - \ddot{\mathbf{G}}_i(\ddot{\mathbf{G}}_i' \ddot{\mathbf{G}}_i)^+ \ddot{\mathbf{G}}_i'$ defined in (4.35) is used to obtain the infeasible LM test-statistic for the i^{th} group denoted by $L\ddot{M}_i$. In the theorem, we show that the feasible LM test statistic denoted by \widetilde{LM}_i is asymptotically equivalent to $L\ddot{M}_i$ under certain conditions. In order to prove this result we need some preliminary order results for the cross products which involve $\ddot{\mathbf{G}}_i$, where $\ddot{\mathbf{G}}_i = (\mathbf{G}_{yi}, \mathbf{G}_{xi}, \ddot{\mathbf{G}}_F)$ and

$$\ddot{\mathbf{g}}_{ft} = \begin{pmatrix} \sum_{\ell=0}^{\infty} \mathbf{H}'_{4\ell} \Delta \mathbf{f}_{t-\ell-1} \\ \vdots \\ \sum_{\ell=0}^{\infty} \mathbf{H}'_{4\ell} \Delta \mathbf{f}_{t-p_T-\ell-1} \end{pmatrix},$$

where $\mathbf{H}_{4\ell}$ is the coefficient of the lag polynomial $\mathbf{H}_4(L)$ for the ℓ^{th} lag order, which is defined subsequent to (4.17).

Lemma 4.B.6. *As $(N, T, p_T) \rightarrow \infty$, where $\frac{p_T^2}{T} \rightarrow 0$, we have*

$$\frac{\Delta \mathbf{X}'_i \ddot{\mathbf{G}}_i}{T} = O_p(1), \quad (4.B.31)$$

$$\frac{\mathbf{U}'_{i,-1} \ddot{\mathbf{G}}_i}{T} = O_p(1), \quad (4.B.32)$$

$$\frac{\ddot{\mathbf{G}}_i' \ddot{\mathbf{G}}_i}{T} = O_p(1), \quad (4.B.33)$$

$$\frac{\boldsymbol{\varepsilon}'_{xi} \ddot{\mathbf{G}}_i}{\sqrt{T}} = O_p(1), \quad (4.B.34)$$

$$\frac{\mathbf{V}'_{B,-1} \ddot{\mathbf{G}}_i}{\sqrt{T}} = O_p\left(\frac{\sqrt{T}}{\sqrt{N}}\right), \quad (4.B.35)$$

$$\frac{\ddot{\mathbf{G}}_i' \mathbf{V}_{C,-p_T}}{T} = O_p(\rho_H^{p_T}), \quad (4.B.36)$$

$$\frac{\ddot{\mathbf{G}}_i' \mathbf{V}_{Ei,-p_T}}{\sqrt{T}} = O_p(\sqrt{T} \rho_A^{p_T}), \quad (4.B.37)$$

$$\frac{\ddot{\mathbf{G}}_i' \mathbf{d}_{i,-1}^u}{T} = O_p\left(\frac{1}{T}\right) + O_p\left(\frac{1}{\sqrt{N}}\right), \quad (4.B.38)$$

for all $i = 1, \dots, N$.

Proof of Lemma 4.B.6. The results above can be proven by standard arguments. Condition $\frac{p_T^2}{T} \rightarrow 0$ is required to obtain well defined limits as $(N, T, p_T) \rightarrow \infty$. This issue is addressed in Berk (1974) and also in Chudik and Pesaran (2013c) in a similar set-up. Adapting their result to our set-up the results in this lemma can be proven straightforwardly. Hence, detailed proofs will not be provided here. It is worth noting that the sum of the coefficients of the lag-polynomial $\mathbf{H}_4(L)$ are bounded and are exponentially decaying. (4.B.32) and (4.B.33) are $O_p(1)$ because the cross product terms involved are functions of covariance stationary processes for an individual unit, say i . For (4.B.34) it is sufficient to mention that since the elements in $\boldsymbol{\varepsilon}_{xi}$ are i.i.d. over t , they are independent of the processes in $\ddot{\mathbf{G}}_i$, which are covariance stationary processes. For (4.B.35) we use the fact that $\mathbf{V}_{A,-1}$ and $\mathbf{V}_{B,-1}$ are of orders $O_p\left(\frac{\sqrt{T}}{\sqrt{N}}\right)$ and $O_p\left(\frac{1}{\sqrt{N}}\right)$ respectively as proven in Lemma B.1 and Lemma B.2 of Chapter 3. ■

For the next lemma, consider the orthogonal projection matrices $\widetilde{\mathbf{M}}_i$ and $\ddot{\mathbf{M}}_i$ that are defined in (4.31) and (4.35) respectively.

Lemma 4.B.7. *Note that $\widetilde{\mathbf{G}}_i$ and $\ddot{\mathbf{G}}_i$ are $(T - p_T) \times (m + 1)p_T$ random matrices, where in our set-up T and p_T are deterministic sequences nondecreasing in N . Let $\boldsymbol{\delta}_{\widetilde{\mathbf{G}},i1}$ and $\boldsymbol{\delta}_{\ddot{\mathbf{G}},i2}$ represent any vectors belonging to the column space of $\ddot{\mathbf{G}}_i$. Similarly let $\boldsymbol{\delta}_{\widetilde{\mathbf{G}},i2}$ and $\boldsymbol{\delta}_{\ddot{\mathbf{G}},i1}$ represent any vectors belonging to the column space of $\widetilde{\mathbf{G}}_i$. Suppose there exist vectors \mathbf{c}_{F1} , \mathbf{c}_{F2} , \mathbf{c}_{Z1} and \mathbf{c}_{Z2} which satisfy $\boldsymbol{\delta}_{\widetilde{\mathbf{G}},i1} = \ddot{\mathbf{G}}_i \mathbf{c}_{F1}$, $\boldsymbol{\delta}_{\ddot{\mathbf{G}},i2} = \ddot{\mathbf{G}}_i \mathbf{c}_{F2}$, $\boldsymbol{\delta}_{\widetilde{\mathbf{G}},i2} = \widetilde{\mathbf{G}}_i \mathbf{c}_{G1}$ and $\boldsymbol{\delta}_{\ddot{\mathbf{G}},i1} = \widetilde{\mathbf{G}}_i \mathbf{c}_{G2}$ such that $\|\mathbf{c}_{F1}\| < K$, $\|\mathbf{c}_{F2}\| < K$, $\|\mathbf{c}_{G1}\| < K$ and $\|\mathbf{c}_{G2}\| < K$, where $K < \infty$ is a constant which does not depend on N . Under these conditions we have the following results:*

$$\begin{aligned} \|\widetilde{\mathbf{M}}_i \boldsymbol{\delta}_{\widetilde{\mathbf{G}},i2}\| &= O_p\left(\frac{p_T \sqrt{T - p_T}}{\sqrt{N}}\right), \\ \|\ddot{\mathbf{M}}_i \boldsymbol{\delta}_{\ddot{\mathbf{G}},i1}\| &= O_p\left(\frac{p_T \sqrt{T - p_T}}{\sqrt{N}}\right). \end{aligned}$$

Furthermore, we obtain the following results regarding the inner products of the vectors:

$$\begin{aligned} \langle \widetilde{\mathbf{M}}_i \boldsymbol{\delta}_{\widetilde{\mathbf{G}},i2}, \widetilde{\mathbf{M}}_i \boldsymbol{\delta}_{\widetilde{\mathbf{G}},i2} \rangle &= O_p\left(\frac{p_T^2 (T - p_T)}{N}\right), \\ \langle \ddot{\mathbf{M}}_i \boldsymbol{\delta}_{\ddot{\mathbf{G}},i1}, \ddot{\mathbf{M}}_i \boldsymbol{\delta}_{\ddot{\mathbf{G}},i2} \rangle &= O_p\left(\frac{p_T^2 (T - p_T)}{N}\right), \end{aligned}$$

where $\widetilde{\mathbf{M}}_i$ and $\ddot{\mathbf{M}}_i$ are the orthogonal complement matrices projected onto the orthogonal complements of the column spaces of $\widetilde{\mathbf{G}}_i$ and $\ddot{\mathbf{G}}_i$ respectively.

Proof of Lemma 4.B.7: This lemma is an adaptation of Lemma A.1 provided in Chudik and Pesaran (2013a) to our set-up. We define

$$\mathbf{V}_{B,(p_T)}^* = (0_{1 \times p-1+m(p-1)}, \mathbf{V}_{B,-1}, \mathbf{V}_{B,-2}, \dots, \mathbf{V}_{B,-(m+1)}) \quad (4.B.39)$$

$(T-p_T) \times [p-1+m(p-1)+(m+1)p_T]$

By considering the definitions of $\widetilde{\mathbf{G}}_i$ and $\ddot{\mathbf{G}}_i$, we can write

$$\widetilde{\mathbf{G}}_i = \ddot{\mathbf{G}}_i + \mathbf{V}_{B,(p_T)}^*. \quad (4.B.40)$$

Note that $\|\widetilde{\mathbf{g}}_{i\ell}\| = O_p(\sqrt{T-p_T})$, $\|\ddot{\mathbf{g}}_{i\ell}\| = O_p(\sqrt{T-p_T})$ and $\|\mathbf{V}_{B,(p_T)}^*\| = O_p(\sqrt{T-p_T}N^{-1/2})$ for all $\ell = 1, 2, \dots, (m+1)p_T$, where $\widetilde{\mathbf{g}}_{i\ell}$, $\ddot{\mathbf{g}}_{i\ell}$ and $\mathbf{V}_{B,(p_T)}^*$ are the ℓ^{th} columns of $\widetilde{\mathbf{G}}_i$, $\ddot{\mathbf{G}}_i$, $\mathbf{V}_{B,(p_T)}^*$ respectively. By an application of the Hilbert projection theorem (see Rudin (1987)), the proof of the result follows from the proof of Lemma A.1 Chudik and Pesaran (2013). \blacksquare

Lemma 4.B.8. Suppose $(N, T, p_T) \rightarrow \infty$ and $\frac{p_T^2}{T} \rightarrow 0$. Let \mathbf{A}_i and \mathbf{B}_i be any two matrices with random elements of size $(T-p_T) \times q_A$ and $(T-p_T) \times q_B$ respectively. Denoting the individual columns of \mathbf{A}_i and \mathbf{B}_i by \mathbf{a}_{ij} and \mathbf{b}_{is} , where $j = 1, \dots, q_A$ and $s = 1, \dots, q_B$, let \mathbf{A}_i and \mathbf{B}_i to satisfy

$$\begin{aligned} (\sqrt{T-p_T})^{-1} \|\ddot{\mathbf{M}}_i \mathbf{a}_{ij}\| &= O_p(R_A), \\ (\sqrt{T-p_T})^{-1} \|\ddot{\mathbf{M}}_i \mathbf{b}_{is}\| &= O_p(R_B), \\ (\sqrt{T-p_T})^{-1} \|\widetilde{\mathbf{M}}_i \mathbf{a}_{ij}\| &= O_p(R_A), \\ (\sqrt{T-p_T})^{-1} \|\widetilde{\mathbf{M}}_i \mathbf{b}_{is}\| &= O_p(R_B), \end{aligned}$$

for all individual columns, where R_A and R_B are deterministic functions of T , N , p_T . Then we have

$$\frac{\mathbf{A}'_i \widetilde{\mathbf{M}}_i \mathbf{B}_i}{T} - \frac{\mathbf{A}'_i \ddot{\mathbf{M}}_i \mathbf{B}_i}{T} = O_p\left(\frac{p_T \times \max\{R_A, R_B\}}{\sqrt{N}}\right) + O_p\left(\frac{p_T^2}{N}\right), \quad (4.B.41)$$

uniformly in i .

Proof of Lemma 4.B.8: This lemma is an adaptation of Lemma A.6 of Chudik and Pesaran (2013a). Hence the proofs follow similarly. We first use the Cauchy-

Schwarz inequality to write

$$\mathbf{a}'_{ij} \widetilde{\mathbf{M}}_i \mathbf{b}_{is} - \mathbf{a}'_{ij} \ddot{\mathbf{M}}_i \mathbf{b}_{is} \leq \|\widetilde{\mathbf{M}}_i \mathbf{a}_{ij}\| \|\widetilde{\mathbf{M}}_i \mathbf{b}_{is}\| - \|\ddot{\mathbf{M}}_i \mathbf{a}_{ij}\| \|\ddot{\mathbf{M}}_i \mathbf{b}_{is}\|. \quad (4.B.42)$$

Then by using the Hilbert projection theorem we can write

$$\|\widetilde{\mathbf{M}}_i \mathbf{a}_{ij}\| \|\widetilde{\mathbf{M}}_i \mathbf{b}_{is}\| \leq \|\mathbf{a}_{ij} - \boldsymbol{\delta}_{zai}\| \|\mathbf{b}_{is} - \boldsymbol{\delta}_{zbi}\|,$$

for any vectors $\boldsymbol{\delta}_{zai}$ and $\boldsymbol{\delta}_{zbi}$ belonging to the column space of $\widetilde{\mathbf{G}}_i$. For our purposes we choose $\boldsymbol{\delta}_{zai}$ and $\boldsymbol{\delta}_{zbi}$ as

$$\begin{aligned} \boldsymbol{\delta}_{zai} &= \ddot{\mathbf{P}}_i \mathbf{a}_{ij} - \widetilde{\mathbf{M}}_i \ddot{\mathbf{P}}_i \mathbf{a}_{ij}, \\ \boldsymbol{\delta}_{zbi} &= \ddot{\mathbf{P}}_i \mathbf{b}_{is} - \widetilde{\mathbf{M}}_i \ddot{\mathbf{P}}_i \mathbf{b}_{is}, \end{aligned}$$

where $\ddot{\mathbf{P}}_i = \mathbf{I} - \ddot{\mathbf{M}}_i$. Note that $\boldsymbol{\delta}_{zai} = (\mathbf{I}_{T-p_T} - \widetilde{\mathbf{M}}_i) \ddot{\mathbf{P}}_i \mathbf{a}_{ij}$ and $\boldsymbol{\delta}_{zbi} = (\mathbf{I}_{T-p_T} - \widetilde{\mathbf{M}}_i) \ddot{\mathbf{P}}_i \mathbf{b}_{is}$ indeed belong to the column space of $\widetilde{\mathbf{G}}_z$. Then we can write

$$\begin{aligned} \|\widetilde{\mathbf{M}}_i \mathbf{a}_{ij}\| \|\widetilde{\mathbf{M}}_i \mathbf{b}_{is}\| &\leq \|\mathbf{a}_{ij} - \ddot{\mathbf{P}}_i \mathbf{a}_{ij} + \widetilde{\mathbf{M}}_i \ddot{\mathbf{P}}_i \mathbf{a}_{ij}\| \|\mathbf{b}_{is} - \ddot{\mathbf{P}}_i \mathbf{b}_{is} + \widetilde{\mathbf{M}}_i \ddot{\mathbf{P}}_i \mathbf{b}_{is}\| \\ &\leq \|\ddot{\mathbf{M}}_i \mathbf{a}_{ij} + \widetilde{\mathbf{M}}_i \ddot{\mathbf{P}}_i \mathbf{a}_{ij}\| \|\ddot{\mathbf{M}}_i \mathbf{b}_{is} + \widetilde{\mathbf{M}}_i \ddot{\mathbf{P}}_i \mathbf{b}_{is}\| \\ &\leq \|\ddot{\mathbf{M}}_i \mathbf{a}_{ij}\| \|\ddot{\mathbf{M}}_i \mathbf{b}_{is}\| + \|\ddot{\mathbf{M}}_i \mathbf{a}_{ij}\| \|\widetilde{\mathbf{M}}_i \ddot{\mathbf{P}}_i \mathbf{b}_{is}\| \\ &\quad + \|\widetilde{\mathbf{M}}_i \ddot{\mathbf{P}}_i \mathbf{a}_{ij}\| \|\ddot{\mathbf{M}}_i \mathbf{b}_{is}\| + \|\widetilde{\mathbf{M}}_i \ddot{\mathbf{P}}_i \mathbf{a}_{ij}\| \|\widetilde{\mathbf{M}}_i \ddot{\mathbf{P}}_i \mathbf{b}_{is}\| \\ &= \|\ddot{\mathbf{M}}_i \mathbf{a}_{ij}\| \|\ddot{\mathbf{M}}_i \mathbf{b}_{is}\| + \epsilon_{iU,NT}. \end{aligned}$$

The third inequality yields an upper bound for $\|\widetilde{\mathbf{M}}_i \mathbf{a}_{ij}\| \|\widetilde{\mathbf{M}}_i \mathbf{b}_{is}\|$. An upper bound for $\|\ddot{\mathbf{M}}_i \mathbf{a}_{ij}\| \|\ddot{\mathbf{M}}_i \mathbf{b}_{is}\|$ can be obtained in a similar way as

$$\begin{aligned} \|\ddot{\mathbf{M}}_i \mathbf{a}_{ij}\| \|\ddot{\mathbf{M}}_i \mathbf{b}_{is}\| &\leq \|\widetilde{\mathbf{M}}_i \mathbf{a}_{ij}\| \|\widetilde{\mathbf{M}}_i \mathbf{b}_{is}\| + \|\widetilde{\mathbf{M}}_i \mathbf{a}_{ij}\| \|\ddot{\mathbf{P}}_i \mathbf{b}_{is}\| \\ &\quad + \|\ddot{\mathbf{P}}_i \mathbf{a}_{ij}\| \|\widetilde{\mathbf{M}}_i \mathbf{b}_{is}\| + \|\ddot{\mathbf{P}}_i \mathbf{a}_{ij}\| \|\ddot{\mathbf{P}}_i \mathbf{b}_{is}\| \\ &= \|\widetilde{\mathbf{M}}_i \mathbf{a}_{ij}\| \|\widetilde{\mathbf{M}}_i \mathbf{b}_{is}\| + \epsilon_{iL,NT}, \end{aligned}$$

where $\ddot{\mathbf{P}}_i = \mathbf{I} - \widetilde{\mathbf{M}}_i$. The last two inequalities provide a lower and an upper bound for the expression on the right-hand side of (4.B.42), such that

$$\epsilon_{iL,NT} \leq \|\widetilde{\mathbf{M}}_i \mathbf{a}_{ij}\| \|\widetilde{\mathbf{M}}_i \mathbf{b}_{is}\| - \|\ddot{\mathbf{M}}_i \mathbf{a}_{ij}\| \|\ddot{\mathbf{M}}_i \mathbf{b}_{is}\| \leq \epsilon_{iU,NT}. \quad (4.B.43)$$

Evaluating the orders of the bounds and showing that the orders are equal to each other will give us the order of the term we are interested in. By setting $\boldsymbol{\delta}_{\widetilde{\mathbf{G}},i2} = \ddot{\mathbf{P}}_i \mathbf{a}_{ij}$, $\boldsymbol{\delta}_{\widetilde{\mathbf{G}},i2} = \ddot{\mathbf{P}}_i \mathbf{b}_{is}$, $\boldsymbol{\delta}_{\widetilde{\mathbf{G}},i1} = \ddot{\mathbf{P}}_i \mathbf{a}_{ij}$ and $\boldsymbol{\delta}_{\widetilde{\mathbf{G}},i2} = \ddot{\mathbf{P}}_i \mathbf{b}_{is}$, we can make use

of Lemma 4.B.7 to establish the order results below.

$$\begin{aligned}\|\ddot{\mathbf{M}}_i \tilde{\mathbf{P}}_i \mathbf{a}_{ij}\| &= O_p\left(\frac{p_T \sqrt{T - p_T}}{\sqrt{N}}\right), \\ \|\ddot{\mathbf{M}}_i \tilde{\mathbf{P}}_i \mathbf{b}_{is}\| &= O_p\left(\frac{p_T \sqrt{T - p_T}}{\sqrt{N}}\right), \\ \|\widetilde{\mathbf{M}}_i \ddot{\mathbf{P}}_i \mathbf{a}_{ij}\| &= O_p\left(\frac{p_T \sqrt{T - p_T}}{\sqrt{N}}\right), \\ \|\widetilde{\mathbf{M}}_i \ddot{\mathbf{P}}_i \mathbf{b}_{is}\| &= O_p\left(\frac{p_T \sqrt{T - p_T}}{\sqrt{N}}\right).\end{aligned}$$

The orders of $\|\widetilde{\mathbf{M}}_i \mathbf{a}_{ij}\|$, $\|\widetilde{\mathbf{M}}_i \mathbf{b}_{is}\|$, $\|\ddot{\mathbf{M}}_i \mathbf{a}_{ij}\|$ and $\|\ddot{\mathbf{M}}_i \mathbf{b}_{is}\|$ are stated in the lemma. By combining those conditions with the order results above, we obtain

$$\epsilon_{i\ell, NT} = O_p\left(\frac{p_T(T - p_T)R_A}{\sqrt{N}}\right) + O_p\left(\frac{p_T(T - p_T)R_B}{\sqrt{N}}\right) + O_p\left(\frac{p_T^2(T - p_T)}{N}\right),$$

for $\ell = L_B, U_B$. Using this result in (4.B.43) yields

$$\begin{aligned}& \left\| \frac{\widetilde{\mathbf{M}}_i \mathbf{a}_{ij}}{\sqrt{T}} \right\| \left\| \frac{\widetilde{\mathbf{M}}_i \mathbf{b}_{is}}{\sqrt{T}} \right\| - \left\| \frac{\ddot{\mathbf{M}}_i \mathbf{a}_{ij}}{\sqrt{T}} \right\| \left\| \frac{\ddot{\mathbf{M}}_i \mathbf{b}_{is}}{\sqrt{T}} \right\| \\ &= O_p\left(\frac{p_T(T - p_T)R_A}{T\sqrt{N}}\right) + O_p\left(\frac{p_T(T - p_T)R_B}{T\sqrt{N}}\right) + O_p\left(\frac{p_T^2(T - p_T)}{TN}\right).\end{aligned}$$

This result gives us the orders of each element in $\frac{\mathbf{A}'_i \widetilde{\mathbf{M}}_i \mathbf{B}_i}{T} - \frac{\mathbf{A}'_i \ddot{\mathbf{M}}_i \mathbf{B}_i}{T}$. We further simplify the orders by noting that $T - p_T \rightarrow T$, which yields the result stated in the lemma. \blacksquare

Lemma 4.B.9. *We have the following results regarding the processes appearing in the estimators:*

$$\begin{aligned}(\sqrt{T - p_T})^{-1} \|\ddot{\mathbf{M}}_i \mathbf{u}_{i,-1}\| &= O_p(1), \\ (\sqrt{T - p_T})^{-1} \|\ddot{\mathbf{M}}_i \mathbf{g}_{ij1}\| &= O_p(1), \\ (\sqrt{T - p_T})^{-1} \|\ddot{\mathbf{M}}_i \varepsilon_{xij2}\| &= O_p(1), \\ (\sqrt{T - p_T})^{-1} \|\ddot{\mathbf{M}}_i \mathbf{x}_{ij3}\| &= O_p(\sqrt{T - p_T}), \\ (\sqrt{T - p_T})^{-1} \|\ddot{\mathbf{M}}_i \mathbf{f}_{j4}\| &= O_p(\sqrt{T - p_T}), \\ (\sqrt{T - p_T})^{-1} \|\ddot{\mathbf{M}}_i \mathbf{v}_{Aj5}\| &= O_p(\sqrt{T - p_T} N^{-1/2}), \\ (\sqrt{T - p_T})^{-1} \|\ddot{\mathbf{M}}_i \mathbf{v}_{Bj6}\| &= O_p(N^{-1/2}), \\ (\sqrt{T - p_T})^{-1} \|\ddot{\mathbf{M}}_i \mathbf{v}_{Cj7}\| &= O_p(\rho_H^{p_T}),\end{aligned}$$

$$\begin{aligned} (\sqrt{T-p_T})^{-1} \|\ddot{\mathbf{M}}_i \mathbf{v}_{Ej_8}\| &= O_p(\rho_A^{p_T}), \\ (\sqrt{T-p_T})^{-1} \|\ddot{\mathbf{M}}_i \mathbf{d}_{i,-1}^u\| &= O_p(\sqrt{T-p_T} N^{-1/2}), \end{aligned}$$

uniformly in $i = 1, \dots, N$, where $j_1, j_2, j_3, j_4, j_5, j_6, j_7, j_8$ are used to denote the individual columns of the corresponding matrices.

Proof of Lemma 4.B.9: The results are established by using the Pythagoras inequality, which implies $\|\ddot{\mathbf{M}}_i \mathbf{a}_{ij}\| \leq \|\mathbf{a}_{ij}\|$. ■

4.B.2 Proofs of Lemmas 4.4.1-4.4.3

Proof of Lemma 4.4.1.

Using the results provided by Chapter 3 the result can be obtained straightforwardly. ■

Proof of Lemma 4.4.2(i).

The estimated error correction term augmented with the cross-sectional averages can be expanded as

$$\begin{aligned} \widehat{\mathbf{U}}_{i,-1} &= \mathbf{U}_{i,-1} + (\widehat{\mathbf{U}}_i - \mathbf{U}_{i,-1} - \overline{\Delta \mathbf{Z}}_{-1} \mathbf{H}'_3(L) \mathbf{H}_1^- \boldsymbol{\lambda}_i) \\ &\quad + \overline{\Delta \mathbf{Z}}_{-1} \mathbf{H}'_3(L) \mathbf{H}_1^- \boldsymbol{\lambda}_i, \end{aligned} \quad (4.B.44)$$

and by noting that the second term on the right-hand side can be replaced by $\mathbf{d}_{i,-1}^u$ defined in (4.B.24), for the first term on the left-hand side of (4.40) we can write,

$$\begin{aligned} T^{-1} \widehat{\mathbf{U}}'_{i,-1} \widetilde{\mathbf{M}}_i \widehat{\mathbf{U}}_{i,-1} &= T^{-1} \mathbf{U}'_{i,-1} \ddot{\mathbf{M}}_i \mathbf{U}_{i,-1} + T^{-1} \mathbf{U}'_{i,-1} (\widetilde{\mathbf{M}}_i - \ddot{\mathbf{M}}_i) \mathbf{U}_{i,-1} \\ &\quad + T^{-1} \mathbf{U}'_{i,-1} \widetilde{\mathbf{M}}_i \mathbf{d}_{i,-1}^u + T^{-1} \mathbf{U}'_{i,-1} \widetilde{\mathbf{M}}_i \mathbf{V}_{C,-p_T} \\ &\quad + T^{-1} \mathbf{d}'_{i,-1} \widetilde{\mathbf{M}}_i \mathbf{U}_{i,-1} + T^{-1} \mathbf{d}'_{i,-1} \widetilde{\mathbf{M}}_i \mathbf{d}_{i,-1}^u \\ &\quad + T^{-1} \mathbf{d}'_{i,-1} \widetilde{\mathbf{M}}_i \mathbf{V}_{C,-p_T} + T^{-1} \mathbf{V}'_{C,-p_T} \widetilde{\mathbf{M}}_i \mathbf{U}_{i,-1} \\ &\quad + T^{-1} \mathbf{V}'_{C,-p_T} \widetilde{\mathbf{M}}_i \mathbf{d}_{i,-1}^u \\ &\quad + T^{-1} \mathbf{V}'_{C,-p_T} \widetilde{\mathbf{M}}_i \mathbf{V}_{C,-p_T}, \end{aligned} \quad (4.B.45)$$

where $\overline{\Delta Z}_{-1} \mathbf{H}'_3(L) \mathbf{H}_1^- \boldsymbol{\lambda}_i$ is replaced with $\mathbf{V}_{C,-p_T}$ by using the result in (4.B.15) and the definition in (4.B.16).

In order to prove the lemma, we need to find the orders of the last nine terms on the right-hand side of the equation above. Starting with the second term, by using the triangle inequality we write

$$\begin{aligned} \left\| \frac{\mathbf{U}'_{i,-1} (\widetilde{\mathbf{M}}_i - \ddot{\mathbf{M}}_i) \mathbf{U}_{i,-1}}{T} \right\| &\leq \left\| \frac{\mathbf{U}'_{i,-1} \widetilde{\mathbf{M}}_i \mathbf{U}_{i,-1}}{T} \right\| - \left\| \frac{\mathbf{U}'_{i,-1} \ddot{\mathbf{M}}_i \mathbf{U}_{i,-1}}{T} \right\| \\ &= O_p(p_T N^{-1/2}) + O_p(p_T^2 N^{-1}), \end{aligned}$$

where the order is obtained by using Lemmas 4.B.9 and 4.B.8. The third term of (4.B.45) can be written in two terms as

$$T^{-1} \mathbf{U}'_{i,-1} \widetilde{\mathbf{M}}_i \mathbf{d}_{i,-1}^u = T^{-1} \mathbf{U}'_{i,-1} \ddot{\mathbf{M}}_i \mathbf{d}_{i,-1}^u + T^{-1} \mathbf{U}'_{i,-1} (\widetilde{\mathbf{M}}_i - \ddot{\mathbf{M}}_i) \mathbf{d}_{i,-1}^u.$$

For the first term, we have $\|T^{-1} \mathbf{U}'_{i,-1} \ddot{\mathbf{M}}_i \mathbf{d}_{i,-1}^u\| \leq \|T^{-1} \mathbf{U}'_{i,-1} \mathbf{d}_{i,-1}^u\|$ by the Pythagoras theorem. Then using (4.B.25) yields $o_p(1)$ as $N, T \rightarrow \infty$ as the order of the first term. To find the order of the second term we first use the triangle inequality and then Lemmas 4.B.8 and 4.B.9 and obtain

$$\begin{aligned} \left\| \frac{\mathbf{U}'_{i,-1} (\widetilde{\mathbf{M}}_i - \ddot{\mathbf{M}}_i) \mathbf{d}_{i,-1}^u}{T} \right\| &\leq \left\| \frac{\mathbf{U}'_{i,-1} \widetilde{\mathbf{M}}_i \mathbf{d}_{i,-1}^u}{T} \right\| - \left\| \frac{\mathbf{U}'_{i,-1} \ddot{\mathbf{M}}_i \mathbf{d}_{i,-1}^u}{T} \right\| \\ &= O_p(p_T T^{1/2} N^{-1}) + O_p(p_T^2 N^{-1}). \end{aligned}$$

In order to find the order of the fourth term we write the left-hand side as

$$T^{-1} \mathbf{U}'_{i,-1} \widetilde{\mathbf{M}}_i \mathbf{V}_{C,-p_T} = T^{-1} \mathbf{U}'_{i,-1} \ddot{\mathbf{M}}_i \mathbf{V}_{C,-p_T} + T^{-1} \mathbf{U}'_{i,-1} (\widetilde{\mathbf{M}}_i - \ddot{\mathbf{M}}_i) \mathbf{V}_{C,-p_T}.$$

The order of the first term is obtained by first applying the Pythagoras theorem, which yields $\|T^{-1} \mathbf{U}'_{i,-1} \ddot{\mathbf{M}}_i \mathbf{V}_{C,-p_T}\| \leq \|T^{-1} \mathbf{U}'_{i,-1} \mathbf{V}_{C,-p_T}\|$ then by using (4.B.17) the order is shown to be equal to $O_p(\rho_H^{p_T})$, which is equivalent to $o_p(1)$ as $p_T \rightarrow \infty$. For the second term we first use the triangular inequality as before and then use Lemmas 4.B.9 and 4.B.8, which gives $O_p(p_T^2 N^{-1})$. The fifth term of (4.B.45) is the transpose of the third term, so the order is the same. The order of the sixth term can be found similarly to the third and the fourth term by using (4.B.29) and Lemmas 4.B.9 and 4.B.8, which yields

$$T^{-1} \mathbf{d}_{i,-1}^{u'} \widetilde{\mathbf{M}}_i \mathbf{d}_{i,-1}^u = O_p(TN^{-1}) + O_p(p_T T^{1/2} N^{-1}) + O_p(p_T^2 N^{-1}).$$

as $(N, T, p_T) \rightarrow \infty$. The order of the seventh term can be found in a similar way using (4.B.27) and Lemmas 4.B.9 and 4.B.8, which yield

$$T^{-1} \mathbf{d}_{i,-1}^{u'} \widetilde{\mathbf{M}}_i \mathbf{V}_{C,-p_T} = O_p(p_T^2 N^{-1}),$$

as $(N, T, p_T) \rightarrow \infty$. The eighth and the ninth terms are the transposes of the fourth and the seventh terms respectively so they must have the same order. Finally, to find the order of the tenth term we again use a similar method. Using (4.B.18) and Lemmas 4.B.9 and 4.B.8 we obtain the order, which is equal to $O_p(p_T^2 N^{-1})$. Combining the orders of the last nine terms of (4.B.45) gives

$$\begin{aligned} T^{-1} \widehat{\mathbf{U}}'_{i,-1} \widetilde{\mathbf{M}}_i \widehat{\mathbf{U}}_{i,-1} &= T^{-1} \mathbf{U}'_{i,-1} \ddot{\mathbf{M}}_i \mathbf{U}_{i,-1} + O_p\left(\frac{T}{N}\right) + O_p\left(\frac{p_T}{\sqrt{N}}\right) \\ &\quad + O_p\left(\frac{p_T^2}{N}\right) + O_p\left(\frac{p_T \sqrt{T}}{N}\right). \end{aligned}$$

But under the conditions of the lemma, which are $\frac{p_T}{T} \rightarrow 0$ and $\frac{T}{N} \rightarrow 0$ the order becomes $o_p(1)$. ■

Proof of Lemma 4.4.2(ii)

Using (4.B.44), (4.B.15) and (4.B.24), under the null hypothesis of weak exogeneity we can write

$$\begin{aligned} \frac{\widehat{\mathbf{U}}'_{i,-1} \widetilde{\mathbf{M}}_i \boldsymbol{\varepsilon}_{xi}}{\sqrt{T}} &= \frac{\mathbf{U}'_{i,-1} \ddot{\mathbf{M}}_i \boldsymbol{\varepsilon}_{xi}}{\sqrt{T}} + \frac{\mathbf{U}'_{i,-1} (\widetilde{\mathbf{M}}_i - \ddot{\mathbf{M}}_i) \boldsymbol{\varepsilon}_{xi}}{\sqrt{T}} \\ &\quad + \frac{\mathbf{d}_{i,-1}^{u'} \widetilde{\mathbf{M}}_i \boldsymbol{\varepsilon}_{xi}}{\sqrt{T}} + \frac{\boldsymbol{\lambda}'_i \mathbf{H}_1^{-1} \mathbf{V}'_{C,-p_T} \widetilde{\mathbf{M}}_i \boldsymbol{\varepsilon}_{xi}}{\sqrt{T}}. \end{aligned} \quad (4.B.46)$$

We need to find the orders of the last three terms to prove the lemma. We use the same methods as the ones we used for the proof of the previous lemma. The order of the second term is $O_p(p_T N^{-1/2}) + O_p(p_T^2 N^{-1})$, which is obtained by using the triangle inequality and Lemmas 4.B.9 and 4.B.8. For the third term applying the Pythagoras theorem, triangle inequality and using (4.B.26) and Lemmas 4.B.9 and 4.B.8, we obtain the order as $O_p(T^{1/2} N^{-1/2}) + O_p(p_T N^{-1/2}) + O_p(p_T^2 N^{-1}) + O_p(p_T T^{1/2} N^{-1})$. The order of the fourth term can be obtained similarly, this time by using (4.B.19) as $O_p(p_T N^{-1/2}) + O_p(p_T^2 N^{-1}) + O_p(p_T N^{-1/2})$. Combining the

orders obtained for the last four terms of (4.B.46) yields

$$\begin{aligned} \frac{\widehat{\mathbf{U}}'_{i,-1} \widetilde{\mathbf{M}}_i \mathbf{e}_{xi}}{\sqrt{T}} &= \frac{\mathbf{U}'_{i,-1} \ddot{\mathbf{M}}_i \boldsymbol{\varepsilon}_{xi}}{\sqrt{T}} + O_p\left(\frac{\sqrt{T}}{\sqrt{N}}\right) \\ &+ O_p\left(\frac{p_T}{\sqrt{N}}\right) + O_p\left(\frac{p_T^2}{N}\right) + O_p\left(\frac{p_T \sqrt{T}}{N}\right) \end{aligned}$$

Then, under the conditions of the lemma the right-hand side can be written as $o_p(1)$. \blacksquare

Proof of Lemma 4.4.2(iii)

For the first part of the Lemma, using equation (4.B.44), the definition in (4.B.24) we can write

$$\begin{aligned} \frac{\widehat{\mathbf{U}}'_{i,-1} \widehat{\mathbf{M}}_i \mathbf{V}_{B,-1} \mathbf{H}_4^-(L) \mathbf{A}'_{xfi}(L)}{\sqrt{T}} &= \frac{\mathbf{U}_{i,-1} \widetilde{\mathbf{M}}_i \mathbf{V}_{B,-1} \mathbf{H}_4^-(L) \mathbf{A}'_{xfi}(L)}{\sqrt{T}} \\ &+ \frac{\mathbf{d}^{ul}_{i,-1} \widetilde{\mathbf{M}}_i \mathbf{V}_{B,-1} \mathbf{H}_4^-(L) \mathbf{A}'_{xfi}(L)}{\sqrt{T}} \\ &+ \frac{\boldsymbol{\lambda}'_i \mathbf{H}_1^{-\prime} \mathbf{V}'_{C,-p_T} \widetilde{\mathbf{M}}_i \mathbf{V}_{B,-1} \mathbf{H}_4^-(L) \mathbf{A}'_{xfi}(L)}{\sqrt{T}}. \end{aligned}$$

In order to find the orders of the three terms on the right-hand side of (4.B.47) we use the same steps as we used in the proof of Lemmas 4.4.2(i) and 4.4.2(ii). For the first term of (4.B.47), by using (4.B.13), Lemmas 4.B.9 and 4.B.8 we obtain

$$\begin{aligned} \frac{\mathbf{U}_{i,-1} \widetilde{\mathbf{M}}_i \mathbf{V}_{B,-1} \mathbf{H}_4^-(L) \mathbf{A}'_{xfi}(L)}{\sqrt{T}} &= O_p(T^{1/2} N^{-1/2}) + O_p(p_T N^{-1/2}) \\ &+ O_p(p_T^2 N^{-1}). \end{aligned}$$

As for the second term using (4.B.30) and Lemmas 4.B.9 and 4.B.8 we have

$$\begin{aligned} \frac{\mathbf{d}^{ul}_{i,-1} \widetilde{\mathbf{M}}_i \mathbf{V}_{B,-1} \mathbf{H}_4^-(L) \mathbf{A}'_{xfi}(L)}{\sqrt{T}} &= O_p(T^{1/2} N^{-1}) + O_p(p_T T^{1/2} N^{-1}) \\ &+ O_p(p_T^2 N^{-1}). \end{aligned}$$

As for the last term using (4.B.20) and Lemmas 4.B.9 and 4.B.8 we have

$$\frac{\boldsymbol{\lambda}'_i \mathbf{H}_1^{-\prime} \mathbf{V}'_{C,-p_T} \widetilde{\mathbf{M}}_i \mathbf{V}_{B,-1} \mathbf{H}_4^-(L) \mathbf{A}'_{xfi}(L)}{\sqrt{T}} = O_p(T^{1/2} \rho^{p_T} N^{-1/2})$$

$$+O_p(p_T^2 N^{-1}).$$

Combining the last three results above and considering (4.B.47) yields

$$\frac{\widehat{\mathbf{U}}'_{i,-1} \widehat{\mathbf{M}}_i \mathbf{V}_{B,-1} \mathbf{H}_4^-(L) \mathbf{A}'_{xfi}(L)}{\sqrt{T}} = O_p\left(\frac{\sqrt{T}}{\sqrt{N}}\right) + O_p\left(\frac{p_T^2}{N}\right) + O_p\left(\frac{p_T \sqrt{T}}{N}\right),$$

which is equivalent to $o_p(1)$ under the conditions of the Lemma. This completes the proof of the first part.

Proof of Lemma 4.4.2(iv)

We start with rewriting the left-hand side of equation (4.43) using (4.B.44), (4.B.24) and (4.B.21) as

$$\begin{aligned} \frac{\widehat{\mathbf{U}}'_{i,-1} \widetilde{\mathbf{M}}_i \overline{\Delta \mathbf{Z}}_{P_{T-1}} \mathbf{H}_4^-(L) \mathbf{A}^{**'}_{xfi}(L)}{\sqrt{T}} &= \frac{\mathbf{U}_{i,-1} \widetilde{\mathbf{M}}_i \mathbf{V}_{Ei,-p_T}}{\sqrt{T}} + \frac{\mathbf{d}_{i,-1}^{u'} \widetilde{\mathbf{M}}_i \mathbf{V}_{Ei,-p_T}}{\sqrt{T}} \\ &\quad + \frac{\boldsymbol{\lambda}'_i \mathbf{H}_1^{-1} \mathbf{V}'_{C,-p_T} \widetilde{\mathbf{M}}_i \mathbf{V}_{Ei,-p_T}}{\sqrt{T}}. \end{aligned} \quad (4.B.47)$$

For the first term on the right-hand side of (4.B.47), by using (4.B.22) and Lemmas 4.B.9 and 4.B.8 we have

$$\frac{\mathbf{U}_{i,-1} \widetilde{\mathbf{M}}_i \mathbf{V}_{Ei,-p_T}}{\sqrt{T}} = O_p(\sqrt{T} \rho^{p_T}) + O_p(p_T N^{-1/2}) + O_p(p_T^2 N^{-1}).$$

For the second term of (4.B.47), using (4.B.28) and Lemmas 4.B.9 and 4.B.8 we obtain

$$\frac{\mathbf{d}_{i,-1}^{u'} \widetilde{\mathbf{M}}_i \mathbf{V}_{Ei,-p_T}}{\sqrt{T}} = O_p(T^{1/2} \rho^{p_T} N^{-1/2}) + O_p(p_T^2 N^{-1}) + O_p(p_T T^{1/2} N^{-1}).$$

The order of the last term of (4.B.47) can be obtained by using (4.B.23) and Lemmas 4.B.9 and 4.B.8, as

$$\frac{\boldsymbol{\lambda}'_i \mathbf{H}_1^{-1} \mathbf{V}'_{C,-p_T} \widetilde{\mathbf{M}}_i \mathbf{V}_{Ei,-p_T}}{\sqrt{T}} = O_p(\sqrt{T} \rho^{2p_T}) + O_p(p_T^2 N^{-1}).$$

Combining the last three results yields

$$\frac{\widehat{\mathbf{U}}'_{i,-1} \widetilde{\mathbf{M}}_i \overline{\Delta \mathbf{Z}}_{P_{T-1}} \mathbf{H}_4^-(L) \mathbf{A}^{**'}_{xfi}(L)}{\sqrt{T}} = O_p(\sqrt{T} \rho^{p_T}) + O_p\left(\frac{p_T}{\sqrt{N}}\right) + O_p\left(\frac{p_T^2}{N}\right)$$

$$+O_p\left(\frac{p_T\sqrt{T}}{N}\right),$$

which is equivalent to $o_p(1)$ under the conditions of the lemma. \blacksquare

Proof of Lemma 4.4.3

If we assume that $\ddot{\mathbf{G}}_f = (\ddot{\mathbf{g}}_{f1}, \dots, \ddot{\mathbf{g}}_{fT})'$ is also observed that is factors and the rotation matrix are observed, where,

$$\ddot{\mathbf{g}}_{ft} = \begin{pmatrix} \sum_{\ell=0}^{\infty} H_{4\ell} \Delta \mathbf{f}_{t-\ell-1} \\ \vdots \\ \sum_{\ell=0}^{\infty} H_{4\ell} \Delta \mathbf{f}_{t-p_T-\ell-1} \end{pmatrix}$$

as defined before, under the null of weak exogeneity we can write

$$\begin{aligned} \Delta \mathbf{X}_i &= \mathbf{G}_{yi} \Gamma_{xyi} + \mathbf{G}_{xi} \Gamma_{xxi} + \ddot{\mathbf{G}}_f \Gamma_{xfi} + \ddot{\mathbf{e}}_{xi}^0, \\ \ddot{\mathbf{M}}_i \Delta \mathbf{X} &= \widehat{\mathbf{e}}_{xi}^0, \end{aligned}$$

and for the case in which the factors are not observed and approximated by the method proposed in the chapter, we can write

$$\begin{aligned} \Delta \mathbf{X}_i &= \mathbf{G}_{yi} \Gamma_{xyi} + \mathbf{G}_{xi} \Gamma_{xxi} + \overline{\mathbf{G}}_z \Gamma_{xfi} + \widetilde{\mathbf{e}}_{xi}^0, \\ \widetilde{\mathbf{M}}_i \Delta \mathbf{X} &= \widehat{\mathbf{e}}_{xi}^0. \end{aligned}$$

Using these equations we can write

$$\begin{aligned} \widehat{\Sigma}_{xxi} - \widetilde{\Sigma}_{xxi} &= T^{-1} \widehat{\mathbf{e}}_{xi}^{0'} \widehat{\mathbf{e}}_{xi}^0 - T^{-1} \widetilde{\mathbf{e}}_{xi}^{0'} \widetilde{\mathbf{e}}_{xi}^0 \\ &= \Delta \mathbf{X}_i' (\widetilde{\mathbf{M}}_i - \ddot{\mathbf{M}}_i) \Delta \mathbf{X}_i. \end{aligned}$$

The next step is to find the order of this expression. By using the triangle inequality we can write

$$\begin{aligned} \left\| \frac{\Delta \mathbf{X}_i' (\widetilde{\mathbf{M}}_i - \ddot{\mathbf{M}}_i) \Delta \mathbf{X}_i}{T} \right\| &\leq \left\| \frac{\Delta \mathbf{X}_i' \widetilde{\mathbf{M}}_i \Delta \mathbf{X}_i}{T} \right\| - \left\| \frac{\Delta \mathbf{X}_i' \ddot{\mathbf{M}}_i \Delta \mathbf{X}_i}{T} \right\| \\ &= O_p(p_T N^{-1/2}) + O_p(p_T^2 N^{-1}), \end{aligned}$$

where the last equation is obtained by using Lemmas 4.B.9 and 4.B.8. Then, under the conditions of the lemma we obtain the required result. \blacksquare

Chapter 5

Conclusion

In this chapter we briefly summarize the findings of this thesis. We discuss the methodological contributions and the limits of each chapter. Furthermore, we review the simulation results of every single chapter to give an indication of the robustness of the methods developed in this thesis in small samples. We then provide a discussion for possible extensions of each work. Given that all the chapters in this thesis have their own conclusion sections, to avoid repetitions, we only briefly summarize the findings without going into details.

This thesis deals with cross-sectional dependence in panel data analysis. In each chapter we contribute to the literature on this branch in different set-ups. While Chapter 2 considers a stationary panel data model, Chapter 3 and Chapter 4 consider panels in a non-stationary set-up. A common feature of all chapters of this thesis is that, we use a common factor structure approach to account for cross-sectional dependence in the model we study. Moreover, in all the chapter we make use of the Common Correlated Effects (CCE) method that is originally developed by Pesaran (2006) to estimate cross-sectionally dependent panels.

In Chapter 2, we consider the estimation of factor-augmented static panel data regression models with homogeneous slope coefficients. One of the most popular approaches towards this end is the pooled common correlated effects (CCE) estimator of Pesaran (2006). For this estimator to be consistent at the usual \sqrt{NT} rate, where N and T denote the number of cross-section and time series observations, respectively, the number of factors cannot be larger than the number of observables. This is a problem in the typical application involving only a small number of regressors. Chapter 2 proposes a simple extension to the CCE proce-

ture by which the requirement can be relaxed. We name this procedure as the C^3E method. The CCE approach is based on taking the cross-section average of the observables as an approximation to the common factors. The idea put forth in Chapter 2 is to consider not only the average but also other cross-section combinations.

The theoretical contributions of the second chapter can be summarized as follows. The C^3E method is shown to yield \sqrt{NT} -consistent and asymptotically normally distributed pooled estimates of the slope coefficients of the model under the condition that $\frac{T}{N} \rightarrow 0$ as $(N, T) \rightarrow \infty$. For the case where $\frac{T}{N} \rightarrow \tau$, where τ is a non-zero constant that does not depend on N , we give an estimate of the bias in terms of the model parameters and propose a bias corrected estimator. The result we obtain regarding the bias correction asserts that under the conditions $\frac{\sqrt{T}}{N} \rightarrow 0$ and $\frac{\sqrt{N}}{T} \rightarrow 0$ and as $(N, T) \rightarrow \infty$, bias correction is successful and the it does not contribute to the limiting variance. Then, we propose a procedure to select the combinations among many candidates. The selection procedure can be summarized as follows. The first step is to order the candidates according to their correlation with the cross-sectional averages of the observed variables, which are considered as an approximation for the space spanned by the factor loadings in this case. Then, to use an information criteria as a second step, which can be seen as an adapted version of one of the *IC* criteria of Bai and Ng (2002). We show that the selection procedure proposed in the chapter consistently selects the appropriate number of combinations that are required to approximate the underlying factor structure. We conduct experiments in various set-ups to investigate the small sample performance of the C^3E method. For comparison purposes, we provide the bias and size results for the *OLS*, *CCE* and the *PC* (principal components) estimators. First we consider the case, in which the rank assumption of Pesaran (2006) is satisfied and the loadings are highly correlated with the candidates that are used to combine the observed variables. We observe that in this case the C^3E estimator performs equally well as the *CCE* estimator and the bias correction seems to be successful in reducing the bias in small samples. In a second experiment we generate the candidates for the combinations independently of the factor loadings. In this case the C^3E estimator still performs as well as the *CCE* estimator in terms of bias. Then for the rest of the experiments, we generate the loadings in a way that would ensure that the rank condition to be violated. In the i.i.d. loadings case, the bias of the pooled CCE estimator is close to the bias of the C^3E estimator but higher for most of the N and T combinations, however

when we compare bias of the CCE estimator across different N and T combinations, we observe that the consistency rate reduce from \sqrt{NT} to N as expected by the theoretical results. For the case in which the loadings are not i.i.d. the CCE estimator breaks down, the C^3E estimator continues to perform well.

Overall, the theoretical results and the Monte Carlo simulations suggest that the method we propose to account for cross-sectional dependence work under more general assumptions regarding the factor loadings and the number of factors. This is an important contribution to the literature, considering the fact that the factors and their loadings are unobserved and can not be estimated separately. This makes it difficult to test the assumptions placed on those two. In this chapter we consider a static model in a stationary set-up. A natural extension might be to consider the method we propose in a non-stationary and/or cointegrated set-up. Furthermore, we assume the presence of strong cross-section dependence only. This work can be extended to a set-up where there is weak cross-section dependence even after accounting for the dependence with a factor structure. Apart from this extension to weak cross-sectional dependence, one could also think about an extension towards heterogeneous β_i . To underpin this conjecture we refer to the paper by Pesaran and Smith (1995). This paper shows that in a static case, with heterogeneous β_i the pooled estimator gives unbiased estimates.

In Chapter 3, we consider the estimation of long run effects in dynamic heterogeneous panel models with I(1) variables and cross-sectional dependence. In particular, we focus on a dynamic cointegrated panel data model with unobserved common factors. Simple estimators for the individual long-run parameters and a mean group estimator for the mean of the long-run parameters in conditional error correction models are proposed, and their properties are evaluated. The estimator can be viewed as belonging to the class of common correlated effect estimators, as initially proposed by Pesaran (2006), that makes use of cross-section averages to control for cross-sectional dependence.

After adapting the results in the existing literature to our set-up and notation by assuming that the factors are observed, we show that a dynamic version of the approximation method that is proposed by Pesaran (2006) can be used to approximate the space spanned by the unobserved common factors and the resulting approximation error is negligible if $\frac{T}{N} \rightarrow 0$, as $(N, T) \rightarrow \infty$. A second result of the chapter is that the coefficients of the augmented conditional model can be estimated by the OLS method. The resulting estimates are consistent and asymptotically normally distributed without any nuisance parameters as $(N, T, p_T) \rightarrow \infty$

and as $\frac{T}{N} \rightarrow 0$, $\frac{p_T^2}{T} \rightarrow 0$ and $\sqrt{T}\rho_A^{p_T} \rightarrow 0$, where p_T is the truncated lag-order for the lagged first differences that are to be included in the augmented model to account for the cross-sectional dependence and $\rho_A < 1$. $\rho_A^{p_T}$ indicates the exponential decay rate of the coefficients of the first differences of the cross-sectional averages in the augmented conditional model as $p_T \rightarrow \infty$. The third important contribution of the paper is related to the estimation of the long-run coefficients for the case, in which the assumed common factor structure is not observed. We show that the indirect least squares estimation of the long-run coefficients yields T -consistent estimates as $(N, T, p_T) \rightarrow \infty$ and as $\frac{T}{\sqrt{N}} \rightarrow 0$, $\frac{p_T^2}{T} \rightarrow 0$ and $\sqrt{T}\rho_A^{p_T} \rightarrow 0$ and the estimators have nuisance parameter free mixed normal asymptotic distributions. Lastly, we provide a result on the mean group estimator. We find that as $(N, T, p_T) \rightarrow \infty$, the mean group estimator is consistent if $\frac{p_T^2}{T} \rightarrow 0$ and asymptotically normally distributed, if, in addition to the conditions for the consistency result, $\frac{\rho_A^{p_T}\sqrt{N}}{T} \rightarrow 0$. As it can be seen from the results summarized above, our results hold conditional on the relative growth rates of N , T and p_T . It is straightforward to show that these conditions do not contradict with each other. Monte Carlo simulations are in line with the theoretical findings of the chapter. We conduct experiments for different set-ups regarding the long-run heterogeneity and the weak exogeneity of the conditioning variables. We compare the performance of the estimator for the long-run coefficients we propose with the static CCE estimator that is proposed by Kapetanios et al. (2011) and with the infeasible estimator with various lag-order selection methods. Overall, we observe that the method we propose performs well in small samples and an increase in the cross-section and the time dimensions improves the performances.

The results of Chapter 3 are obtained under the assumption of weak exogeneity. This assumption is crucial for our analysis and in applications it is needed to be verified. We propose a testing method for the weak exogeneity of the conditioning variables for the long-run coefficients in Chapter 4. From the standard results in the literature we know that the mixed normality of the long-run coefficient estimates from the conditional model can only be ensured under this assumption. A more comprehensive study, in particular a study of the unconditional model (i.e. the panel VAR) can be considered as a possible extension of this work. In the paper we assume that there is only one cointegrating vector for each cross-section unit. In future work, this assumption might be relaxed and a mixed panel case can be considered, where only some of the units have a cointegrating relation. Furthermore, this study can be extended by considering the case with the presence

of deterministic components both in the short-run and in the long-run. Also considering a pooled mean group estimator of Pesaran et al. (1999), obtaining the estimator within a likelihood set-up can be viewed as an extension of Pesaran et al. (1999) to a case with unobserved factor structure.

In Chapter 4, we propose Lagrange-multiplier type tests for weak exogeneity in panel error correction models with unobserved common factors. We consider a test for the absence of error correcting behavior from the marginal model for individual units, as well as a joint panel test. We use the common-correlated effects method to approximate the unobserved common factors. We derive the asymptotic distributions of individual specific statistics and propose bootstrap group-mean type panel tests.

The results obtained in Chapter 4 can be summarized as follows. We show that the test statistics are not affected by the approximation method we use under certain conditions placed on the relative growth rates of N , T and p_T , where p_T denotes again the truncated lag-order for the first differences of the cross-sectional averages that we propose to augment the conditional and the marginal model with. In particular, we find that as $(N, T, p_T) \rightarrow \infty$, the Lagrange multiplier test statistic is not affected by the approximation of the unobserved common factors given that the following conditions hold: $\frac{T}{N} \rightarrow 0$, $\frac{p_T^2}{T} \rightarrow 0$ and $\sqrt{T}\rho_A^{p_T} \rightarrow 0$, where $\rho_A < 1$ and $\rho_A^{p_T}$ gives the exponential decay rate of the coefficients of the lagged first differences of the cross-sectional averages in the marginal model as $p_T \rightarrow \infty$. Moreover, the test statistic we obtain for individual cross-section units are shown to be cross-sectionally dependent, which invalidates the standard pooling methods to use standard asymptotic results for the pooled test statistics. To get around that difficulty we propose a bootstrap procedure. The finite sample properties of the tests are assessed in a Monte Carlo simulation. We provide the results of four different experiments. In the first experiment, we generate the data independently across cross-section units, in the second experiment we assume that the common factors are observed and obtain the test statistics accordingly. In the third and the fourth experiments, we use the approximation method suggested in the paper. The third and the fourth experiments differ based on the way we treat the potential serial correlation in the error terms of the augmented models. In the third experiment we use a block bootstrap procedure in the fourth experiment we use a truncation lag order. A comparison of the results of the first and the third experiments suggest that the individual Lagrange multiplier statistic is indeed not affected by the approximation. We observe substantial size distortion in the

standard pooled tests as expected. However the bootstrap versions of the tests performs well in small samples and yield size results close to the nominal value of the tests. As for the power properties of the tests, we observe low power for the individual tests, whereas the power increases substantially as we move from individual tests to bootstrap versions of the pooled tests, again as expected by the theory.

A potential extension of the problem considered in Chapter 4 can be built on the pooled tests. Although we propose to use a bootstrap procedure for the dependent test statistics we do not verify the validity of the bootstrap procedure. This might be considered as a limitation of our work and can be pursued as an extension.

In conclusion, although there are still open questions regarding the issues considered in this thesis, we hope that the results provided here can provide contributions to the literature on cross-sectionally dependent panels and can be a stepping stone for future theoretical research. Also we hope that empirical researchers can benefit from the results of this thesis.

Bibliography

- Acemoglu, D., M. Arellano, and E. DeKel (2013). *Advances in Economics and Econometrics: Economic theory*, Volume 1. Cambridge University Press.
- Ahn, S. C. and A. R. Horenstein (2013). Eigenvalue ratio test for the number of factors. *Econometrica* 81, 1203–1227.
- Andrews, D. W. K. (2005). Cross-section regression with common shocks. *Econometrica* 73, 1551–1586.
- Anselin, L. (1988). *Spatial econometrics: methods and models*, Volume 4. Springer.
- Bai, J. (2003). Inferential theory for factor models of large dimensions. *Econometrica* 71, 135–171.
- Bai, J. (2009). Panel Data Models With Interactive Fixed Effects. *Econometrica* 77, 1229–1279.
- Bai, J. and J. L. Carrion-i Silvestre (2009, July). Testing panel cointegration with unobservable dynamic common factors. MPRA Paper 35243, University Library of Munich, Germany.
- 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.

BIBLIOGRAPHY

- Bai, J. and S. Ng (2002). Determining the number of factors in approximate factor models. *Econometrica* 70, 191–221.
- Bai, J. and S. Ng (2004). A PANIC attack on unit roots and cointegration. *Econometrica* 72, 1127–1177.
- Bai, J. and S. Ng (2006). Confidence intervals for diffusion index forecasts and inference for factor-augmented regressions. *Econometrica* 74, 1133–1150.
- Bai, J. and S. Ng (2007). Determining the number of primitive shocks in factor models. *Journal of Business & Economic Statistics* 25.
- Bailey, N., G. Kapetanios, and M. H. Pesaran (2012). Exponent of Cross-sectional Dependence: Estimation and Inference. CESifo Working Paper.
- Baltagi, B. (2008). *Econometric analysis of panel data*, Volume 1. John Wiley & Sons.
- 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. and J. L. Carrion-i Silvestre (2012). Testing for panel cointegration using common correlated effects estimators. Discussion Papers 11-16, Department of Economics, University of Birmingham.
- Bårdsen, G. (1989). Estimation of long run coefficients in error correction models. *Oxford Bulletin of Economics and Statistics* 51, 345–50.
- Berk, K. N. (1974). Consistent autoregressive spectral estimates. *The Annals of Statistics* 2, 489–502.
- Boswijk, H. (1992). *Cointegration, Identification and Exogeneity*, Vol. 37 of *Tinbergen Institute Research Series*. Amsterdam: Thesis Publishers.
- Boswijk, H. P. (1995). Efficient inference on cointegration parameters in structural error correction models. *Journal of Econometrics* 69, 133–158.
- Boswijk, P. H. and J.-P. Urbain (1997). Lagrange multiplier tests for weak exogeneity: A synthesis. *Econometric Reviews* 16, 21–38.
- Breitung, J. and I. Choi (2013). Factor models. *Handbook Of Research Methods And Applications In Empirical Macroeconomics*.

- Breitung, J. and M. H. Pesaran (2008). *Unit roots and cointegration in panels*. Springer.
- Burger, R. P. and F. J. Teal (2014). The effect of schooling on worker productivity: evidence from a south african industry panel. Technical report.
- Cesa-Bianchi, A. (2013). Housing cycles and macroeconomic fluctuations: A global perspective. *Journal of International Money and Finance* 37, 215–238.
- 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.
- Chudik, A. and M. H. Pesaran (2011). Infinite-dimensional vars and factor models. *Journal of Econometrics* 163, 4–22.
- Chudik, A. and M. H. Pesaran (2013a). Common correlated effects estimation of heterogeneous dynamic panel data models with weakly exogenous regressors. CESifo Working Paper.
- Chudik, A. and M. H. Pesaran (2013b). Common correlated effects estimation of heterogeneous dynamic panel data models with weakly exogenous regressors.
- Chudik, A. and M. H. Pesaran (2013c). Econometric analysis of high dimensional vars featuring a dominant unit. *Econometric Reviews* 32, 592–649.
- Chudik, A. and M. H. Pesaran (2013d). Large panel data models with cross-sectional dependence: a survey. CESifo Working Paper.
- Chudik, A., M. H. Pesaran, and E. Tosetti (2011). Weak and strong cross-section dependence and estimation of large panels. *Econometrics Journal* 14, C45–C90.
- Coakley, J., A.-M. Fuertes, and R. Smith (2002). A Principal Components Approach to Cross-Section Dependence in Panels. 10th International Conference on Panel Data, Berlin, July 5-6, 2002 B5-3, International Conferences on Panel Data.

BIBLIOGRAPHY

- Davidson, J. (1994). *Stochastic Limit Theory: An Introduction for Econometricians*. Oxford university press.
- Davidson, J. (2000). *Econometric Theory*. Blackwell Publishers Ltd.
- Demetrescu, M., U. Hassler, and A.-I. Tarcolea (2006). Combining significance of correlated statistics with application to panel data. *Oxford Bulletin of Economics and Statistics* 68, 647–663.
- Eberhardt, M., C. Helmers, and H. Strauss (2013). Do spillovers matter when estimating private returns to r&d? *Review of Economics and Statistics* 95, 436–448.
- Eberhardt, M. and F. Teal (2008). Modeling technology and technological change in manufacturing: how do countries differ? Technical report.
- Engle, R. F., D. F. Hendry, and J.-F. Richard (1983). Exogeneity. *Econometrica: Journal of the Econometric Society*, 277–304.
- Fisher, R. A. (1934). Statistical methods for research workers.
- Forni, M. and M. Lippi (2001). The generalized dynamic factor model: representation theory. *Econometric theory* 17, 1113–1141.
- Gengenbach, C., F. C. Palm, and J.-P. Urbain (2006). Cointegration testing in panels with common factors. *Oxford Bulletin of Economics and Statistics* 68, 683–719.
- Gengenbach, C., J.-P. Urbain, and J. Westerlund (2013). Alternative representations for cointegrated panels with global stochastic trends. *Economics Letters* 118, 485–488.
- Gengenbach, C., J. Westerlund, and J.-P. Urbain (2014). Error correction testing in panels with global stochastic trends. Unpublished Manuscript.
- Gonçalves, S. and B. Perron (2014). Bootstrapping factor-augmented regression models. *Journal of Econometrics* 182, 156–173.
- Hartung, J. (1999). A note on combining dependent tests of significance. *Biometrical journal* 41, 849–855.

- Herzer, D. and P. Nunnenkamp (2012). The effect of foreign aid on income inequality: Evidence from panel cointegration. *Structural Change and Economic Dynamics* 23, 245–255.
- Hsiao, C. (2007). Panel data analysis advantages and challenges. *Test* 16, 1–22.
- Im, K. S., M. H. Pesaran, and Y. Shin (2003). Testing for unit roots in heterogenous panels. *Journal of Econometrics* 115, 53–74.
- Johansen, S. (1991). Estimation and hypothesis testing of cointegration vectors in gaussian vector autoregressive models. *Econometrica* 59, 1551–1580.
- Johansen, S. (1992). Cointegration in partial systems and the efficiency of single-equation analysis. *Journal of Econometrics* 52, 389–402.
- Kao, C. (1999). Spurious regression and residual-based tests for cointegration in panel data. *Journal of Econometrics* 90, 1–44.
- Kao, C., L. Trapani, and G. Urga (2012). Asymptotics for panel models with common shocks. *Econometric Reviews* 31, 390–439.
- Kapetanios, G. (2004). A new method for determining the number of factors in factor models with large datasets. Technical report, Working Paper, Department of Economics, Queen Mary, University of London.
- Kapetanios, G. (2010). A testing procedure for determining the number of factors in approximate factor models with large datasets. *Journal of Business & Economic Statistics* 28, 397–409.
- Kapetanios, G., M. H. Pesaran, and T. Yamagata (2011). Panels with nonstationary multifactor error structures. *Journal of Econometrics* 160, 326–348.
- Karabiyik, H., J.-P. Urbain, and J. Westerlund (2014). CCE estimation of factor-augmented regression models with more factors than observables. Research Memorandum 007, Maastricht University, Graduate School of Business and Economics (GSBE).
- 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. and J. Lyhagen (2007). Inference in panel cointegration models with long panels. *Journal of Business and Economic Statistics* 25, 473–483.

BIBLIOGRAPHY

- Larsson, R., J. Lyhagen, and M. Lötgren (2001). Likelihood-based cointegration tests in heterogenous panels. *Econometrics Journal* 4, 109–142.
- Lee, L.-f. and J. Yu (2010). Some recent developments in spatial panel data models. *Regional Science and Urban Economics* 40, 255–271.
- Levin, A., C. F. Lin, and C. S. J. Chu (2002). Unit root tests in panel data: Asymptotic and finite-sample properties. *Journal of Econometrics* 108, 1–24.
- Lütkepohl, H. (1996). *Handbook of Matrices*. Wiley.
- Lütkepohl, H. (2007). *New Introduction to Multiple Time Series Analysis*. Springer.
- 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.
- Moral-Benito, E. and L. Servén (2013). Testing weak exogeneity in cointegrated panels.
- Onatski, A. (2009). Testing hypotheses about the number of factors in large factor models. *Econometrica* 77, 1447–1479.
- Palm, F. C., S. Smeeke, and J.-P. Urbain (2007). A sieve bootstrap test for cointegration in a conditional error correction model. METEOR Research Memorandum RM/07/054, Maastricht University.
- Palm, F. C., S. Smeeke, and J.-P. Urbain (2011). Cross-sectional dependence robust block bootstrap panel unit root tests. *Journal of Econometrics* 163, 85–104.
- 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. (1999). Critical values for cointegration tests in heterogenous panels with multiple regressors. *Oxford Bulletin of Economics and Statistics* 61, 653–670.
- Pedroni, P. (2004). Panel cointegration; asymptotic and finite sample properties of pooled time series tests, with an application to the PPP hypothesis. *Econometric Theory* 20, 597–625.

-
- Pesaran, H. M., Y. Shin, and R. Smith (1999). Pooled mean group estimation of dynamic heterogeneous panels. *Journal of the American Statistical Association* 94, 621–634.
- Pesaran, M. and Y. Shin (1995). An autoregressive distributed lag modelling approach to cointegration analysis. Technical report.
- Pesaran, M. H. (2006). Estimation and inference in large heterogeneous 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 R. Smith (1995). Estimating long-run relationships from dynamic heterogeneous panels. *Journal of Econometrics* 68, 79–113.
- Pesaran, M. H. and E. Tosetti (2011). Large panels with common factors and spatial correlation. *Journal of Econometrics* 161, 182–202.
- Pesaran, M. H., L. Vanessa Smith, and T. Yamagata (2013). Panel unit root tests in the presence of a multifactor error structure. *Journal of Econometrics* 175, 94–115.
- Phillips, P. C. B. and H. R. Moon (1999). Linear regression limit theory for nonstationary panel data. *Econometrica* 67, 1057–1111.
- Rudin, W. (1987). *Real and Complex Analysis*. Tata McGraw-Hill Education.
- Said, S. E. and D. A. Dickey (1984). Testing for unit roots in autoregressive-moving average models of unknown order. *Biometrika* 71, 599–607.
- Sarafidis, V. (2009). Gmm estimation of short dynamic panel data models with error cross-sectional dependence.
- Sarafidis, V. and T. Wansbeek (2012). Cross-sectional dependence in panel data analysis. *Econometric Reviews* 31, 483–531.
- Song, M. (2013). Asymptotic theory for dynamic heterogeneous panels with cross-sectional dependence and its applications. Technical report, Mimeo, January.
- Stock, J. H. and M. W. Watson (2002). Forecasting using principal components from a large number of predictors. *Journal of the American Statistical Association* 97, 1167–1179.

BIBLIOGRAPHY

- Trapani, L. (2014). Testing for exogeneity in cointegrated panels. *Oxford Bulletin of Economics and Statistics*.
- Urbain, J.-P. (1992). On weak exogeneity in error correction models. *Oxford Bulletin of Economics and Statistics* 54, 187–206.
- Wagner, M. and J. Hlouskova (2010). The performance of panel cointegration methods: Results from a large scale simulation study. *Econometric Reviews* 29, 182–223.
- Westerlund, J. and R. Larsson (2009). A note on the pooling of individual PANIC unit root tests. *Econometric Theory* 25, 1851–1868.
- Westerlund, J. and J.-P. Urbain (2013a). Cross-sectional averages versus principal components. Unpublished Manuscript.
- Westerlund, J. and J.-P. Urbain (2013b). On the estimation and inference in factor-augmented panel regressions with correlated loadings. *Economics Letters* 119, 247–250.
- Zellner, A. (1962). An efficient method of estimating seemingly unrelated regressions and tests for aggregation bias. *Journal of the American statistical Association* 57, 348–368.

Nederlandse samenvatting

Paneldata is een verzameling van gegevens waarbij verschillende eenheden worden gevolgd gedurende een bepaalde tijdsperiode. De data in een panel bevat dus zowel een tijdsdimensie als ook een zogenaamde individuele dimensie. Het woord ‘individuele in de voorafgaande zin moet ietwat ruimer worden opgevat dan de naam aanvankelijk doet vermoeden. De individuele dimensie kan inderdaad bestaan uit gegevens betreffende specifieke individuen. Ze kan echter ook refereren naar verschillende bedrijven, industrietakken of zelf landen. Het bruto binnenlands product (BBP) zoals waargenomen voor enkele landen in de Europese Unie over de tijdsperiode 2000-2014 is een voorbeeld van paneldata. Recente analyse heeft aangetoond dat de individuele eenheden vaak afhankelijkheid vertonen. Deze afhankelijkheid kan verregaande gevolgen hebben, als ze wordt genegeerd in de modellering. Er is dan ook theoretisch onderzoek gedaan ter beantwoording van de vraag: hoe kan individuele afhankelijkheid in paneldata (het best) gemodelleerd worden? Gedurende de laatste jaren is de gerelateerde literatuur gestaag gegroeid. Toch is er nog steeds interesse in nieuwe methoden. Methoden die werken onder minder strenge aannamen, of statistische toetsen of te zien of deze aannamen waarschijnlijk zijn. Dit proefschrift levert een bijdrage aan de ontwikkeling van zulke vernieuwende methoden.

De statistische afhankelijkheid in de individuele dimensie kan veroorzaakt worden door het ontbreken van relevante variabelen in het model, de aanwezigheid van globale verstoringen, geografische effecten, etc. Stel dat er een variabele is die niet onafhankelijk is over de individuele eenheden. Het ontbreken van deze variabele in de modellering impliceert ogenblikkelijk dat de individuele verstoringen in het model deze variabele zullen bevatten. Dit resulteert in de genoemde

afhankelijkheid in de individuele dimensie. De aanwezigheid van globale verstoringen is vergelijkbaar. In dit geval is er a priori een verstoring die alle individuele eenheden beïnvloedt. Laten we terugkeren naar ons voorbeeld: het bbp van de landen in de Europese Unie. Volgens empirische studies is het aannemelijk dat het bruto binnenlands product wordt beïnvloed door zulke globale verstoringen. In andere woorden, het bbp van de verschillende landen in de Europese Unie is niet alleen een voorbeeld van paneldata, maar zelfs een voorbeeld van paneldata met afhankelijk in de individuele dimensie.

De globale verstoringen kunnen gemodelleerd worden door middel van factor modellen. De achterliggende gedachte is dat een klein aantal niet waargenomen variabelen (de factoren) alle eenheden in de individuele dimensie gezamenlijk beïnvloedt. Het feit dat deze factoren niet worden waargenomen impliceert dat deze alleen benaderd kunnen worden. Twee populaire benaderingen zijn de *Principale Componenten (PC)* benadering en de *Common Correlated Effects (CCE)* benadering. De *CCE* approximatie is voorgesteld in Pesaran (2006) voor stationaire factoren en stationaire variabelen. De niet geobserveerde factoren worden benaderd door gemiddelden over de individuele eenheden. Deze methode is toepasbaar onder verschillende aannamen. Een voordeel van de *CCE* methode is de eenvoud: gemiddelden volstaan.

De *CCE* approximatie staat centraal in dit proefschrift. Hoofdstuk 1, de introductie, geeft meer informatie over deze benadering en behandelt de concepten die herhaaldelijk gebruikt zullen worden. Daarnaast informeert deze introductie de lezer over de doelstelling en de motivatie onderliggende dit proefschrift. De wetenschappelijke bijdragen die dit proefschrift levert worden ook vermeld.

Hoofdstuk 2 van dit proefschrift bevat de eerste discussie van paneldata met afhankelijkheid in de individuele dimensie. Het vertrekpunt van deze discussie is n van de aannamen die cruciaal is in Pesaran (2006). Deze aanname vereist dat het aantal niet waargenomen factoren niet groter mag zijn dan het aantal geobserveerde variabelen. Deze veronderstelling is noodzakelijk voor de consistente en efficiënte schatting van het model. We stellen voor niet alleen gemiddelden te nemen in de individuele dimensie, maar ook andere combinaties van de waargenomen variabelen toe te staan. Het zal blijken dat de genoemde aanname door deze modificatie overbodig is.

In Hoofdstuk 3 beperken we ons tot de gemiddelden over de individuele eenheden. We beschouwen niet langer andere combinaties van de variabelen. De uitbreiding en uitdaging is nu de toepassing op een specifiek model, namelijk het

Error Correction Model (ECM). Een extra aanname, *weak exogeneity*, maakt het mogelijk ons zonder verlies van efficiëntie te concentreren op het *Conditional Error Correction Model (CECM)*. In dit *CECM* laten we vervolgens afhankelijkheid toe in de individuele dimensie. Om met deze afhankelijkheid te kunnen omgaan verrijken we het model met de huidige en vertraagde eerste verschillen van de individuele dimensie gemiddelden. Zowel de lange als ook de korte termijn parameters kunnen nu consistent geschat worden met de (indirecte) kleinste kwadraten methode. We zullen ook aantonen dat een andere consistente schatter van de lange termijn parameters verkregen kan worden door de lange termijn parameters voor iedere individuele eenheid afzonderlijk te schatten, en de resultaten te middelen.

Het gebruik van het *CECM* was gebaseerd op de *weak exogeneity* aanname. In Hoofdstuk 4 beschouwen we Lagrange-Multiplier toetsen om te testen voor *weak exogeneity* van sommige variabelen voor de lange termijn parameters. Allereerst leiden we een test procedé af dat geldig is voor de individuele eenheden afzonderlijk. Hierna construeren we uit de toetsingsgrootheden voor de individuele eenheden nieuwe toetsingsgrootheden die de testen voor de *weak exogeneity* van alle individuele eenheden gezamenlijk. De toegestane afhankelijkheid van de individuele eenheden leidt echter ook tot afhankelijkheid van de individuele toetsingsgrootheden. De standaard methoden die gebruikt worden om individuele toetsen te combineren tot een gezamenlijke toets zijn niet langer direct toepasbaar. We beredeneren dat als gevolg hiervan bootstrap methoden noodzakelijk zijn.

Dit proefschrift vormt hopelijk een waardevolle bijdrage tot de literatuur over paneldata met afhankelijkheid in de individuele dimensie. Een literatuur waarin vele vragen nog onbeantwoord zijn. We hopen dat zowel de theoretische als ook de toegepaste onderzoeker zal profiteren van de inhoud van dit proefschrift. De theoreticus ontdekt wellicht vernieuwende inzichten, of zelfs een opstapje voor een geheel nieuw eigen onderzoek. De toegepaste onderzoeker zou zich kunnen verheugen op een nieuw stukje gereedschap om paneldata te onderzoeken. Echter geen gewoon stukje gereedschap, maar een stukje gereedschap dat expliciet rekening houdt met mogelijke afhankelijkheid in de individuele dimensie.

Valorization

Knowledge valorization of a scientific work is particularly concerned with the societal relevance of that work, thus with its contribution to real life phenomena, in particular on social and economic issues. In this addendum, I outline the knowledge valorization of the results obtained in this thesis. I do so, by first focusing on the importance of research in econometric theory, then, by providing an overview of the societal relevance of the results reported in this dissertation, the target groups that can benefit from these results and the activities that these results can be translated into. Further, I discuss the innovative characteristics of the current research and its implementability.

This thesis is concerned with developing methodologies for the estimation and inference of cross-sectionally dependent panel data models. Although, the research reported in thesis is theoretical and methodological and the results can be considered as contributions to the field of econometric theory, its knowledge valorization is far from being inconsiderable. The results provided by econometric theory are used by researchers in fields of applied sciences, who aim at improving societal and economical conditions by giving answers to socially and economically relevant questions. Whenever a set of data is at a researcher's disposal, the usage of econometric and statistical techniques is in order. From this point of view, progress in econometric theory, which provides advanced tools to empirical researchers, directly effects the world of applied research.

Technological advancements associated with obtaining and recording data in societal and economic fields have made huge data sets available to researchers in relevant fields. Apart from that, new communication technologies, developing financial markets, agreements between financial entities, increasing world popula-

tion, cultural accumulation within societies have lead to an increase in the number and complexity of economically and socially relevant questions that can be investigated by social scientists. Accumulation of data and increase in the number and complexity of questions, have together increased the demand for technical tools that can be used to answer those questions by using available data more effectively. For instance, questions regarding causal relationships between certain variables, effects of certain variables on others or dynamic relations between economic indicators may be answered by using available data related to the questions and proper econometric tools. The knowledge valorization of research on econometric theory should be viewed by reckoning with these arguments.

Panel data analysis is used to investigate the behaviour of a set of entities that is periodically observed over time. It has many applications in social sciences, particularly in the fields of behavioural sciences like psychology or sociology, where one studies the characteristics of groups of people over time, or in economics, where one investigates the behaviour of economic and financial indicators of multiple entities; for instance, people, firms, regions, countries, products and markets over time. Although, panel data analysis is used in many other fields such as health sciences, educational research, etc., in this knowledge valorization addendum I focus on its applications in the fields of economics and sociology.

The discussion until this point illustrates the social and economic relevance and importance of the research in econometric theory and panel data econometrics in general whereas the research reported in this dissertation investigates a special topic in panel data analysis. In particular, in this thesis, we study stationary and non-stationary panel data models with cross-sectional dependence. In order to narrow the discussion down to the specific topic of the research reported in this dissertation, consider a data set that enlists the gross domestic products (GDP) of countries in the Middle East, for twenty years, say 1994-2014. This is an example of a panel data set. Empirical research has shown that the GDP data series are usually non-stationary. We say that a data series is non-stationary if the underlying distribution changes over time. Presence of non-stationarity necessitates the utilization of methodologies that are specifically developed for that phenomena. Estimation and inference of cross-sectionally dependent non-stationary panels are considered in Chapter 3 of this thesis. To illustrate the concept of cross-sectional dependence suppose that the determination of the GDPs of each country at each time period depends on several factors. Those factors can be grouped in two main categories, country specific factors such as inflation, interest rates, private con-

sumption, imports and exports etc., and factors that are common to all countries such as oil prices, interest rates in the U.S., etc. In such a setup, if data sets meet certain criteria, econometric modelling techniques and econometric analysis can be used to analyse the determination of the GDP, the effects of changes in the factors on changes in GDP and the causal relations between the variables considered. For the results of econometric analysis to be reliable, it is important that the data sets and the models that are being considered to meet certain assumptions of the econometric techniques that are being used. Classical methods of panel data analysis assume that the innovation processes that generate the data are cross-sectionally independent. In the current example, let me name the part of the GDP that remains after accounting for the effects of the individual specific factors and observed common factors as the error term. Cross-sectional independence requires that the error terms of each country to be independent of other countries. However, most of the panel data sets do not satisfy this assumption. For instance, some of the common factors, such as global shocks, might not be observed, hence cannot be included in the models when investigating the determination of the gross domestic product. Unobserved common factors that are left in the error terms of each country, induce a dependence between the error terms. In the literature this phenomena is called the cross-sectional dependence and it necessitates using special techniques to obtain reliable results from econometric analysis of panel data.

The social and economic relevance of this research can be explained better by considering some examples of applied research that attempt to answer socially and economically relevant questions by using cross-sectionally dependent panel data. Herzer and Nunnenkamp (2012) investigate the short-run and long-run affects of foreign aid on income inequality by considering 21 recipient countries over the period 1970-1995. They consider the potential cross-sectional dependence by recognizing that aid flows to each country might be affected by global shocks such as business cycles, global natural disasters, regional wars and income inequality in each country might be affected by technological progress. Long-run and short-run affects of foreign aid on income inequality can be investigated by using panel cointegration techniques. However, cross-sectional dependence has to be accounted for properly, in order to obtain reliable estimates. In Chapter 3 of this dissertation we provide techniques to be used in situations similar to this example. Other examples of applications of the techniques investigated in this dissertation can be listed as follows. Burger and Teal (2014) measure the effect of schooling on individual

earning in African countries where cross-sectional dependence in data is present because of global productivity shocks. Eberhardt and Teal (2008) investigate the cointegrating relationship between output and a set of inputs. Cesa-Bianchi (2013) investigates the effects of housing demand shocks on real economic activity. Questions such as above demand empirical researchers to use techniques that are specifically developed for cross-sectionally dependent data. In particular, developing models and estimating those models can only be done in a reliable way by using techniques that take into account the potential cross-sectional dependence. This fact reveals the social and economical relevance of the research reported in this thesis.

The results of this thesis can be useful to researchers that are doing empirical research in almost any field, for instance finance, marketing, demography and climatology. In all these fields panel data analysis is a common practice and the likelihood of coming across with non-stationary and cross-sectionally dependent data is considerably high. For example, in climatology, the temperature measured in several locations over time may be non-stationary due to the increase of CO₂ emissions, where the level of CO₂ emission has the effect of changing the underlying distribution of temperature over time. Outcomes of this thesis can be interesting for government officials, policy makers in investigating, forecasting the effects of their policies. They can be used by the researchers in central banks to forecast and estimate the effects of certain decisions regarding interest rates, exchange rates, money supply and other financial products central banks control. They may be and are used by applied macroeconomists for instance at the European Central Bank or other central banks, in testing validity of macroeconomic theories. For example, to see if the Fisher Effect is a real phenomena, one can use the results of Chapter 3 to estimate the cointegrating relation between real interest rates, nominal interest rates and inflation. Firms might also benefit from the results of this dissertation, if, for example they are interested in the factors that are affecting their sales in different countries over time. Social organisations constitute another indirect target group of this research, in the sense that these results can be used by them in understanding the effects of certain variables on others and how do the effects change among social groups. In this case, using a more accurate technique is crucial in choosing the right action.

Apart from their applicability in applied social sciences, the results of this thesis provide a range of options for future research in econometric theory, that

is to say the researchers in econometric theory are the natural audience of this research.

The above discussion provides an overview of the target groups of the research presented in this thesis. Now I shall discuss the products that the research in this thesis can be translated into. Each chapter can be seen as an introduction of a toolbox that an empirical researcher can use. These constitute the set of immediate products of the research in this thesis. Nevertheless, the consequences of employment of the outcomes of this thesis by potential audiences mentioned above should be counted as well. For instance, social organisations that are aiming at improving societal conditions can take actions based on their research that is conducted by using the toolboxes developed in this thesis. Other potential final products of this research can be listed as, inter alia, firms' pricing, marketing and segmentation decisions, central bankers' decisions regarding the interest rates, money supply, improvements in macroeconomic theory.

The literature on cross-sectionally dependent panel data analysis has become voluminous in the last decades as applied research has repeatedly proven that cross-sectional dependence is present in most of the panel data sets. The contribution of this thesis to the literature of cross-sectionally dependent panel data models is discussed in Chapter 1 in detail. The research reported in this thesis extends the existing results and allows the empirical researchers to answer more sophisticated questions that can be answered by analysis of panel data sets. For instance, investigating the short-run and long-run effects of certain variables on some others simultaneously, in a cross-sectionally dependent panel can be done by using the results of this thesis. The method developed in Chapter 2 allows the empirical researchers to work under less restrictive assumptions. The result of Chapter 4 provides a method to test for one of the crucial assumptions of the method developed in Chapter 3. The results of this thesis can be called innovative in the sense that they provide a wider range of opportunities for empirical researchers in modelling and estimation of cross-sectionally dependent panel data.

The implementation of the results of this thesis can be considered as a continuous process. The toolboxes developed in this thesis can be used by empirical researchers whenever they have a panel data in which they suspect presence of cross-sectional dependence. Besides that, the results can be used as a stepping stone for further theoretical research. For further extensions of the results I refer the reader to the conclusion part of this thesis.

Curriculum Vitae

Hande Karabıyık was born in July 3, 1986 in Nevsehir, Turkey. She attended high school between 1997 and 2004 at 60. Yıl Anadolu Lisesi in İzmir. Subsequently, she started studying in the Economics and Business Administration program that is jointly run by University of London and Istanbul Bilgi University. In August 2008 she obtained her B.Sc. degree from London School of Economics and Political Science (with first class honours) as a consequence. In August 2010, she received a M.Sc. degree in Economics. Her master thesis is entitled “The Effect of Exchange Rate Volatility on Trade Volume: A Panel Data Application”.

After graduation, Hande joined the Department of Quantitative Economics as a Ph.D. candidate in September 2010, under the supervision of Prof. dr. Jean-Pierre Urbain and Prof. dr. Franz Palm. The results of her research are presented in this thesis. Hande presented her work at various international conferences and parts of this thesis are under revision in international refereed academic journals.