Contributions to the Analysis of Economic Growth and Cross-Sectional Dependence

Jan Ditzen

Submitted for the degree of Doctor of Philosophy

Heriot-Watt University

School of Social Sciences

December 2017

Supervised by

Prof. Arnab Bhattacharjee and Prof. Mark Schaffer

Abstract

This thesis links the theoretical and the applied literature on interdependence between countries in growth models and their impact on convergence. Economic theory agrees on the existence of interactions between countries, but the empirical literature neglects these interactions. Econometric theory defines two types of dependence between units, which both needs to be taken care of when estimated.

The thesis consists of three chapters. The first chapter presents a growth model, which motivates the weaker type of dependence, spatial dependence. In this model, migration, trade and foreign direct investments act as channels for the interaction of countries. The model predicts positive effects of the interactions, especially of migration. It is common to model the second type of cross-sectional dependence in form of a multifactor error structure model in a heterogeneous slope panel. The model is estimated by the Dynamic Common Correlated Effects estimator, which approximates the dependence by time specific averages. The second chapter introduces a Stata package to compute this estimator. It discusses practical challenges in its empirical application, presents examples for the estimation and highlights the requirements for the time and cross-sectional dimensions using a Monte Carlo simulation. The final chapter combines the contributions of the first two chapters. A spatial time lag controls for spatial dependence. The growth model in the first chapter is used to motivate the choice of the weights. Strong cross-sectional dependence is taken care of by the methods explained in the preceding chapter. In addition, the chapter uses a general Lotka-Volterra model to determine the type of convergence in the presence of spatial interactions. Lastly, evidence for conditional convergence is presented for a panel of 93 countries.

To my parents and my two brothers

Acknowledgements

First of all, I would like to thank my two supervisors, Arnab Bhattacharjee and Mark Schaffer, for their support during my PhD. From the very start, they have been open to many questions and discussions, turned ideas into output and opened up many opportunities.

I am indebted to Prof. Dr. Erich Gundlach of Hamburg University. I would not have discovered my interested in growth theory and my love for Stata without his course "Growth Empirics in Stata". During my PhD I benefited from many discussion at conferences, talks or within the Economics department at Heriot-Watt; those include (in alphabetical order) Guiseppe Arbia, Parantap Basu, Philip Brock, Atanas Christev, David Drukker, Markus Eberhardt, Odded Galor, Shlomo Yitzhaki and all the users of `xtdcce2` who came back to me with many questions and suggestions.

I am thankful to my PhD fellows, friends and colleagues: Achim Ahrens (for always being helpful and having the same bad taste in movies), Antonio Carvalho (for all the electricity fun), Erkal Ersoy (for suggesting LANs), Marco Larusso (for being Italian), Kyle McNabb (for his deep knowledge of Edinburgh) and Mark Wang (for all the econometrics chats). Leo Salzmann and Flo Heine, it is great that our friendship lasted from the very first semester onwards, our weekends are unforgettable and I am excited to see where each of us will be next. Thanks to Vana Anastasiadou, Gerdis Marquardt, Jovan Vojnovic, Christina Malivitsi and Silje Sæthershagen for many coffees, dinners, drinks, movies and being there as friends. Then thanks to the HAX-crew for much fun over the last 5 years, Fran Watson for living across the courtyard, and John Sutherland for being my flatmate for four years. I could not have finished this PhD without my favourite distractions - walking up one of the 282 Munros in any sort of weather or dangling on ropes on a climbing wall, especially with Alex Collins, Chris Fensch, Charlotte Wood and Rossana Zetti.
Finally, I would like thank my family. My two brothers, my cousin Thomas, Reni and Andreas and lastly, my parents who always believed in and supported me.

ACADEMIC REGISTRY
Research Thesis Submission

Name:	Jan Ditzen		
School:	School of Social Sciences		
Version: *(i.e. First, Resubmission, Final)*	Final	Degree Sought:	PhD

Declaration

In accordance with the appropriate regulations I hereby submit my thesis and I declare that:

1) the thesis embodies the results of my own work and has been composed by myself
2) where appropriate, I have made acknowledgement of the work of others and have made reference to work carried out in collaboration with other persons
3) the thesis is the correct version of the thesis for submission and is the same version as any electronic versions submitted*.
4) my thesis for the award referred to, deposited in the Heriot-Watt University Library, should be made available for loan or photocopying and be available via the Institutional Repository, subject to such conditions as the Librarian may require
5) I understand that as a student of the University I am required to abide by the Regulations of the University and to conform to its discipline.
6) I confirm that the thesis has been verified against plagiarism via an approved plagiarism detection application e.g. Turnitin.

* *Please note that it is the responsibility of the candidate to ensure that the correct version of the thesis is submitted.*

Signature of Candidate:		Date:	

Submission

Submitted By *(name in capitals):*	
Signature of Individual Submitting:	
Date Submitted:	

For Completion in the Student Service Centre (SSC)

Received in the SSC by *(name in capitals):*			
Method of Submission *(Handed in to SSC; posted through internal/external mail):*			
*E-thesis Submitted (**mandatory for final theses)***			
Signature:		Date:	

Please note this form should be bound into the submitted thesis.
Academic Registry/Version (1) August 2016

Contents

1 Introduction **1**

1.1 Spatial Econometrics 6

1.2 Types of Cross-Sectional Dependence 10

1.2.1 Cross-Sectional Dependent Processes 10

1.2.2 Cross-Sectional Dependent Factors 11

1.2.3 Spatial Dependence and Spatial Lags 12

1.3 Dynamic Panel Data Model with Cross-Sectional Dependence 14

1.4 Test for weak cross-sectional dependence 15

1.5 Common Correlated Effects Estimators 16

1.6 Empirical Growth Models and Convergence 20

1.7 Growth Empirics and Spatial Econometrics 24

1.8 Summary 25

2 A Growth Model with Mobile Labour, Trade and Diffusion of Ideas **27**

2.1 Introduction 27

2.2 Literature 30

2.2.1 Growth Models 31

2.2.2 Diffusion of Ideas 32

2.2.3 New Economic Geography 33

2.2.4 Trade 34

2.2.5 Migration . . . 35
2.3 A Growth Model with Mobile Labour . . . 36
2.3.1 R&D Sector . . . 37
2.3.2 Manufacturing Sector . . . 38
2.3.3 Traditional Goods Sector . . . 40
2.3.4 Consumers . . . 41
2.3.5 The Growth Engine . . . 43
2.3.6 Closing the Model . . . 46
2.3.7 High Skilled Worker Migration . . . 51
2.3.8 Incorporating Migration Costs . . . 53
2.3.9 Discussion . . . 54
2.4 Simulation . . . 59
2.4.1 Baseline . . . 61
2.4.2 Migration . . . 62
2.4.3 Trade . . . 65
2.4.4 Trade and Migration . . . 67
2.4.5 Diffusion of Designs . . . 68
2.4.6 Migration, Diffusion and Trade . . . 69
2.5 Conclusion . . . 71
Appendix of Chapter 2 . . . 74
A.1 Derivation of price index and quantities demanded by consumers . . . 74
A.2 Calculation of Migration Streams with Migration Costs . . . 79

3 Estimating Common Correlated Effects Models 82

3.1 Introduction . . . 82
3.2 Estimators . . . 84
3.3 Syntax . . . 87
3.4 The Constant in xtdcce2 . . . 88

3.5 Empirical Examples . . . 89
3.5.1 Dynamic Common Correlated Effects and testing for cross-sectional dependence . . . 89
3.5.2 Pooled Mean Group . . . 97
3.5.3 Mean Group and Common Correlated Effects . . . 101
3.6 Monte Carlo Simulation . . . 103
3.7 Monte Carlo Results . . . 105
3.8 Conclusion . . . 113
Appendix of Chapter 3 . . . 117
B.1 The xtdcce2 command . . . 117
B.1.1 Syntax . . . 117
B.1.2 Version . . . 117
B.1.3 Options . . . 118
B.1.4 Saved results . . . 121
B.1.5 Postestimation . . . 121
B.1.6 xtcd2 . . . 122
B.1.7 Error Messages . . . 124
B.2 Mathematical Appendix - Delta Method . . . 124
B.3 Monte Carlo Simulation . . . 126
B.3.1 Monte Carlo Setup . . . 126
B.3.2 Further Monte Carlo Results . . . 129
4 Cross-Country Convergence in a General Lotka-Volterra Model 134
4.1 Introduction . . . 134
4.2 A General Lotka-Volterra Model . . . 135
4.3 Growth Model and Convergence . . . 139
4.3.1 Growth Model . . . 139
4.3.2 Convergence in spatial growth models . . . 141

4.4 Heterogeneous Panel Estimators 143
4.5 Empirical Equation 143
4.6 Testing for absolute convergence and inference on eigenvalues 146
4.7 Data 148
4.8 Estimations 150
4.8.1 Simultaneous Dynamic Least Squares 150
4.8.2 Dynamic Common Correlated Effects - Mean Group 151
4.9 Robustness Checks 153
4.9.1 Dynamic Common Correlated Effects - Pooled 153
4.9.2 Dynamic Common Correlated Effects - IV estimation 154
4.9.3 Alternating Number of Cross-Sectional Lags 155
4.9.4 Lagged Spatial Weights 157
4.9.5 US as common factor 158
4.9.6 Alternating length of bootstrap blocks 158
4.10 Conclusion 158
Appendix of Chapter 4 161
C.1 Countries 161
C.2 Simultaneous Dynamic Least Square estimator 161
C.3 Tables 165

5 Conclusion **177**

References **180**

List of Figures

2.1 Simulation: Baseline Model 61

2.2 Simulation: Only Migration 63

2.3 Simulation: Migration with costs 64

2.4 Simulation: Trade Only 65

2.5 Simulation: Trade and Migration 67

2.6 Simulation: Diffusion of Ideas only 69

2.7 Simulation: Migration, Trade and Diffusion of Ideas 70

3.1 Range Plot for Mean Group Estimates 95

List of Tables

1.1 Spatial Lag models . . . 7

3.1 Specifications for Monte Carlo Simulations . . . 105

3.2 Monte Carlo Results Specification 1: ϕ . . . 106

3.3 Monte Carlo Results Specification 1: β_0 . . . 107

3.4 Monte Carlo Results Specification 1: β_1 . . . 108

3.5 Monte Carlo Results Specification 2: ϕ . . . 109

3.6 Monte Carlo Results Specification 2: β_0 . . . 110

3.7 Monte Carlo Results Specification 2: β_1 . . . 111

3.8 Monte Carlo Results Specification 3: ϕ . . . 112

3.9 Monte Carlo Results Specification 3: β_0 . . . 113

3.10 Monte Carlo Results Specification 3: β_1 . . . 114

3.11 Monte Carlo Results Specification 5: β_1 . . . 115

B.1 Monte Carlo Results Specification 4: ϕ . . . 129

B.2 Monte Carlo Results Specification 4: β_0 . . . 130

B.3 Monte Carlo Results Specification 4: β_1 . . . 131

B.4 Monte Carlo Results Specification 5: ϕ . . . 132

B.5 Monte Carlo Results Specification 5: β_0 . . . 133

C.1 SDLS Estimation . . . 165

C.2 DCCE Mean Group Estimation $p_T = 3$ . . . 166

C.3 DCCE Pooled Estimation $p_T = 3$ . . . 167

C.4 DCCE IV Estimation $p_T = 3$ 168

C.5 DCCE Mean Group Estimation without cross-sectional averages . . . 169

C.6 DCCE Mean Group Estimation with $p_T = 0$ 170

C.7 DCCE Mean Group Estimation with $p_T = 1$ 171

C.8 DCCE Mean Group Estimation with lagged weight matrices 172

C.9 DCCE IV Estimation with lagged weight matrices 173

C.10 DCCE Mean Group Estimation with dominant factor 174

C.11 DCCE Mean Group Estimation and alternative bootstrap length I. . 175

C.12 DCCE Mean Group Estimation and alternative bootstrap length II. . 176

List of Abbreviations

3SLS 3 Stage Least Squares.

CCE Common Correlated Effects.

CES Constant Elasticity of Substitution.

CS-DL Cross-Sectionally Augmented Distributed Lag.

DCCE Dynamic Common Correlated Effects.

FDI Foreign Direct Investments.

FE Fixed Effects.

GDP Gross Domestic Product.

GLS Generalized Least Squares.

GMM General Method of Moments.

iid independent and identically distributed.

IIDN independent and identically distributed normal.

IIDU independent and identically distributed uniform.

IV Instrumental Variable.

MG Mean Group.

ML Maximum Likelihood.

NEG New Economic Geography.

NEGG New Economic Geography and Growth.

OECD Organisation for Economic Co-operation and Development.

OLS Ordinary Least Squares.

PMG Pooled Mean Group.

PWT Penn World Tables.

QML Quasi Maximum Likelihood.

R&D Research and Development.

RMA Recursive Mean Adjustment.

RMSE Root Mean Squared Error.

SDLS Simultaneous Dynamic Least Squares.

SSC Statistical Software Components.

SURE Seemingly Unrelated Regression.

TFP Total Factor Productivity.

US United States of America.

Chapter 1

Introduction

Why are people in the United States, Germany, and Japan so much richer today than 100 or 1000 years ago? Why are people in France and the Netherlands today so much richer than people in Haiti and Kenya? Questions like these are at the heart of the study of economic growth.

– Jones (2016, page 4)

Does the Flap of a Butterfly's Wings in Brazil Set Off a Tornado in Texas?

– Edward N. Lorenz, 1972

Charles Jones posed these questions in the introduction for the latest volume of the *Handbook of Macroeconomics*. The handbook chapter shows that the questions are still of relevance, despite the fact that the first ideas about the origin of economic growth date as least as far back as Thomas R. Malthus's seminal book *An Essay on the Principle of Population* from 1798.

The above title for a talk by Edward N. Lorenz in 1972 summarises the idea of the *butterfly effect*. The butterfly effect states that a small change somewhere in the world can have a huge cause somewhere else. Edward N. Lorenz derived this idea from simulating weather models and became one of the founders of the *Chaos*

Theory in physics.

If the *butterfly effect* applies to the weather, it may be relevant on a smaller scale for countries and their economies. In other words, a change in one country has an effect on its neighbours or other countries, meaning countries depend on each other. The combination of this statement and the questions Charles Jones posed are the focus of this thesis. Does a country's growth process depend on other countries and how can this be modelled from a theoretical and applied perspective.

Grossman and Helpman (2015) and Jones (2016) note that the theoretical literature agrees on the existence of spillovers and interactions between countries. They are modelled in many different ways such as trade (Grossman and Helpman, 1991), diffusion of knowledge (Rivera-Batiz and Romer, 1991; Lucas, 2009a) or geographical distance (Krugman, 1991). Even though theory recognises the spillovers or dependencies between countries, the empirical literature neglect them (Islam, 2003; Corrado and Fingleton, 2012). In cross-country regressions, it is common to control for the location of a country (Hall and Jones, 1999; Sachs, 2001), the openness to trade of an economy (Frankel and Romer, 1999; Wacziarg and Welch, 2007) or disease environments (Acemoglu, Johnson and Robinson, 2001). However location or its proxies are pure controls and do not represent interactions between countries. Interactions between countries need to be modelled explicitly and come out of the microfoundations of a growth model.

In a theoretical and applied econometric context the question arises how to model and control for dependence between units. The literature on panel data models differentiates between strong and weak dependence between cross-sectional units. Strong cross-sectional dependence is in the domain of multifactor error structure models and is contained as an unobserved factor in the error term. If observable, weak cross-sectional dependence can be modelled by spatial interactions in the form of weighted observations of other units. Both types of cross-sectional dependence separately received attention in the empirical literature on growth (Ertur and Koch, 2007; Eberhardt, Helmers and Strauss, 2012; Eberhardt and Teal, 2017). Only recently methods have been developed to account for both types of cross-sectional dependence in a general empirical setting (Bailey, Holly and Pesaran, 2016; Ertur and Musolesi, 2016).

Convergence is among the key questions in the empirics of economic growth and the second central topic in this thesis. Convergence is beyond pure academic relevance and has implications for policy. If countries would converge to a single equilibrium, then disparities between countries are a matter of time and policy could work towards closing those. When including spatial interactions in a growth regression, the conditions for convergence change. Arbia and Paelinck (2003a,b) suggest a growth model based on a difference equation system, which is able to handle interactions between units, for determining the type of convergence.

This thesis consists of three essays, which are linked by the above questions. The motivation of spatial interactions for an empirical model is in the centre of interest of Chapter 2. A growth model is derived, which in its outcome explicitly links countries to each other. Appropriate methods to estimate those linkages can be found in the field of spatial econometrics. An estimation procedure and the description of a Stata package to estimate dynamic panel data models with a multifactor error structure is outlined in Chapter 3. A focus is put on the requirements on the time and cross-sectional dimensions of the dataset. In addition, applied issues such as the estimation of a unit-specific constant are discussed. Chapter 4 combines the two strands of literature on cross-sectional dependence in a growth model. It uses the model derived in Chapter 2 as motivation for the spatial interactions. Strong cross-sectional dependence is addressed as described in Chapter 3. To determine the convergence in presence of spatial dependencies, the model extends the approach by Arbia and Paelinck (2003a,b). The three chapters are outlined in more detail next.

Chapter 2: "A Growth Model with Mobile Labour, Trade and Diffusion of Ideas" addresses the gap between the theoretical growth and the applied literature by developing a model that draws on features from an endogenous Romer-style growth model and a New Economic Geography (NEG) model. The model has three distinct sources of interactions between countries: mobility of high skilled workers, diffusion of knowledge and inter-country trade of goods. One novelty of the model is the focus on migration of high skilled workers and their explicit effect on economic growth. The decision to migrate comes from the micro level and therefore out of the model. The engine of growth is adapted from the endogenous growth literature. Motivated by higher wages, high skilled workers migrate to the more developed country, where they work in the Research and Development (R&D) sector. This in turn contributes towards economic growth in the more developed country, and leads to divergence between the two countries. Diffusion of knowledge links the R&D sectors of countries. Ideas developed in one country become available in the other country, resulting in an increase in productivity of the sector. Trade in the manufactured good has a positive effect on the more developed country, but only in the short run. The chapter presents an approach to model the interdependence between countries explicitly. A country's growth rate depends on characteristics of other countries and linkages are micro level founded. In its conclusion of divergence, the model brings new insights to some of the facts of economic development over the last 200 years.

Chapter 3: "Estimating Common Correlated Effects Models" introduces a new Stata command, `xtdcce2`, to estimate a dynamic model with common correlated effects and heterogeneous coefficients in a panel with a large number of observations over cross-sectional units and time periods. The package was developed as a part of this thesis. The focus of this chapter lies on the second type of cross-sectional dependence, unobservable common factors. The package computes the Dynamic Common Correlated Effects (DCCE) estimator (Pesaran, 2006; Chudik and Pesaran, 2015a), but in addition allows estimations of the Mean Group (MG) estimator (Pesaran and Smith, 1995) and the Pooled Mean Group (PMG) estimator (Shin, Pesaran and Smith, 1999). Coefficients are allowed to be heterogeneous or homogeneous. In addition, Instrumental Variable (IV) regressions and unbalanced panels are supported. A test for weak cross-sectional dependence (CD Test) is automatically calculated and presented in the estimation output. Small sample time series bias can be corrected by 'half-panel' jackknife or Recursive Mean Adjustment (RMA) correction methods. A Monte-Carlo simulation is carried out to examine the small sample properties of the estimators. It evaluates the impact of the time series bias and biases due to the heterogeneous slope coefficients and the common factors. Finally, it is intended to give the user guidance under which cross-sectional and time dimensions it is appropriate to employ the DCCE estimator.

Chapter 4: "Cross-country convergence in a general Lotka-Volterra model" combines the two preceding chapters, by estimating a convergence equation including a spatial time lag and common factors. Therefore, it controls for both types of cross-sectional dependence. The equation is estimated for 93 countries over the years 1960-2007. The spatial lag controls for spatial dependence while the common factors control for strong cross-sectional dependence. The share of high skilled migrants, trade shares and Foreign Direct Investments (FDI) are used as spatial weights matrices. A Simultaneous Dynamic Least Squares (SDLS) estimator and a DCCE estimator are employed. In the equation to be estimated, convergence depends on the country and due to the nature of the interactions, on other countries as well. A traditional approach would neglect the inter-country dependences. The growth equation is transformed into a representation of a general Lotka-Volterra model. Convergence depends on cross-sectional interactions in the differential equation system. The conditions for stability of the general Lotka-Volterra model are used to determine the existence and type of convergence. Evidence for conditional convergence is presented. Several robustness checks are carried out and confirm earlier findings. Finally, the chapter highlights the difficulty of differentiating between strong and weak dependence.

In the remaining part of this chapter, the spatial econometric theory necessary for this thesis is reviewed, followed by an outline of cross-sectional dependence.

The chapter closes with an overview of the growth empirical literature and spatial extensions of it.

Notation Throughout this thesis, the following notation applies: For the theoretical model in Chapter 2, capital letters, such as A_i refer to levels on the country level. Small letters, such as x_i refer to a firm or sector specific value. For the remaining chapters, vectors are expressed in small bold letters, for example $\mathbf{b}$, and matrices in capital bold letters or symbols, $\mathbf{B}$. $\mathbf{I}_N$ is a $N \times N$ identity matrix, $\boldsymbol{\iota}_N$ is a $1 \times N$ column vector with ones. Squared brackets denote the floor of a number, $[3.14] = 3$. N defines the cross-sectional dimension, T the time dimension. $\#(T_i \cap T_j)$ refers to the number of common time periods of cross-sectional unit i and j. $N \to \infty$ stands for convergence to infinity, $(N,T) \overset{j}{\to} \infty$ for joint convergence to infinity, $\sqrt{N}(\boldsymbol{\pi}_p - \boldsymbol{\pi}) \overset{d}{\to} N(0, \boldsymbol{\Sigma}_p)$ for convergence in distribution and $\hat{\beta} - \beta \overset{p}{\to} 0$ for convergence in probability. Δ is the first lag operator, i.e. $\Delta y_{i,t} = y_{i,t} - y_{i,t-1}$. $\|\mathbf{w}\|$ is the vector norm. $\Rightarrow$ means it follows, for example $\frac{x}{2} = y \Rightarrow y = 2x$. Lastly, Stata refers to the statistical software released by StataCorp.

1.1 Spatial Econometrics

Spatial econometrics incorporates the interaction of cross-sectional units in regression analysis. Several models are standard in the literature (Anselin, 1988; Elhorst, 2010). The most general model is the so-called Manski model, which includes spatial interactions in the independent variable, dependent variable and the error term. Table 1.1 lists a selection of spatial models, from the Manski model to the standard Ordinary Least Squares (OLS) model without any spatial components. The spatial lag and the spatial error model are the most frequent used models in the literature, of importance for this thesis and will be outlined next. In distinction of the rest of this thesis, this section looks at a cross-section and not at a panel dataset. In the

Name	Equation	Assumptions
Manski	$\mathbf{y} = \rho\mathbf{W}\mathbf{y} + \alpha\boldsymbol{\iota}_N + \mathbf{X}\boldsymbol{\beta} + \mathbf{W}\mathbf{X}\boldsymbol{\Theta} + \mathbf{u}$	
	$\mathbf{u} = \lambda\mathbf{R}\mathbf{u} + \boldsymbol{\epsilon}$	
Spatial Durbin Model	$\mathbf{y} = \rho\mathbf{W}\mathbf{y} + \alpha\boldsymbol{\iota}_N + \mathbf{X}\boldsymbol{\beta} + \mathbf{W}\mathbf{X}\boldsymbol{\Theta} + \boldsymbol{\epsilon}$	$\lambda = 0$
Spatial Durbin Error Model	$\mathbf{y} = \alpha\boldsymbol{\iota}_N + \mathbf{X}\boldsymbol{\beta} + \mathbf{W}\mathbf{X}\boldsymbol{\Theta} + \mathbf{u}$	$\rho = 0$
	$\mathbf{u} = \lambda\mathbf{R}\mathbf{u} + \boldsymbol{\epsilon}$	
Spatial Lag	$\mathbf{y} = \rho\mathbf{W}\mathbf{y} + \alpha\boldsymbol{\iota}_N + \boldsymbol{\epsilon}$	$\lambda = 0, \boldsymbol{\Theta} = \mathbf{0}$
Spatial Error Model	$\mathbf{y} = \alpha\boldsymbol{\iota}_N + \mathbf{X}\boldsymbol{\beta}\mathbf{u}$	$\rho = 0, \boldsymbol{\Theta} = \mathbf{0}$
	$\mathbf{u} = \lambda\mathbf{R}\mathbf{u} + \boldsymbol{\epsilon}$	or $\boldsymbol{\Theta} = -\rho\boldsymbol{\beta}, \lambda = \rho$
OLS Model	$\mathbf{y} = \alpha\boldsymbol{\iota}_N + \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}$	$\lambda = \rho = 0, \boldsymbol{\Theta} = \mathbf{0}$

Table 1.1: Different Spatial lag models and the relation to the Manski Model and OLS, based on Elhorst (2010, Figure 1). $\mathbf{y}, \mathbf{u}$ and $\boldsymbol{\epsilon}$ are $N \times 1$ vectors, $\mathbf{X}$ is a $N \times K$ matrix, α, ρ and λ are scalars, $\boldsymbol{\beta}$ and $\boldsymbol{\Theta}$ are $K \times 1$ and $\mathbf{W}$ and $\mathbf{R}$ are $N \times N$. N is defined as the number of observations and K the number of regressors.

spatial lag model, a weighted sum of the dependent variable is added:[1]

$$\mathbf{y} = \rho\mathbf{W}\mathbf{y} + \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}, \tag{1.1}$$

where ρ is the spatial autocorrelation coefficient and $\mathbf{W}$ is a $N \times N$ spatial weight matrix, with:

$$\mathbf{W} = \begin{pmatrix} 0 & w_{1,2} & \cdots & \cdots & w_{1,N} \\ w_{2,1} & \ddots & w_{2,3} & \cdots & w_{2,N} \\ \vdots & & \ddots & & \vdots \\ \vdots & & & \ddots & w_{N-1,N} \\ w_{N,1} & w_{N,2} & \cdots & w_{N,N-1} & 0 \end{pmatrix}. \tag{1.2}$$

The diagonal elements of the spatial weight matrix consists of zeros. $w_{i,j}$ relates unit j to unit i. It is assumed that the spatial weights are known.[2] In the case of an endogenous spatial weight matrix, the matrix needs to be estimated. Bhattacharjee and Holly (2013) and Bhattacharjee and Jensen-Butler (2013) suggest several approaches to estimate unknown spatial weights and apply it to housing markets in the United Kingdom (Bhattacharjee and Jensen-Butler, 2013), Portugal (Bhattacharjee, Maiti, Castro and Marques, 2016) and interactions in the Monetary Policy Com-

[1] $\mathbf{y}$ and $\boldsymbol{\epsilon}$ are $N \times 1$, $\mathbf{X}$ is $N \times K$, $\boldsymbol{\beta}$ is $K \times 1$ and ρ is 1×1.

[2] Alternatively the weights can be defined as exogenous or pre-specified.

mittee within the Bank of England (Bhattacharjee and Holly, 2013). Estimation for endogenous weight matrices are further discussed for example in Kelejian and Piras (2014) and Qu and Lee (2015).

Even in the case of known spatial weights, the spatial lag is endogenous if the error terms are correlated across units and OLS becomes inconsistent. Equation (1.1) can be rewritten as:

$$\mathbf{y} = (\mathbf{I}_N - \rho\mathbf{W})^{-1}\mathbf{X}\boldsymbol{\beta} + (\mathbf{I}_N - \rho\mathbf{W})^{-1}\boldsymbol{\epsilon}. \tag{1.3}$$

where $\boldsymbol{\epsilon}$ is an independent and identically distributed (iid) error term. Anselin (1988); Kelejian and Prucha (1998) show that the spatial lag model can be estimated by an Instrumental Variable (IV) and General Method of Moments (GMM) approach. Alternatively, as derived in Anselin (1988); Lee (2004); Yu, de Jong and Lee (2008) the model can be estimated by Maximum Likelihood (ML).

The weight matrix $\mathbf{R}$ has the same structure and properties as $\mathbf{W}$ and the spatial error model is defined as:

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{u}, \tag{1.4}$$

$$\mathbf{u} = \lambda\mathbf{R}\mathbf{u} + \boldsymbol{\epsilon}, \tag{1.5}$$

In a similar fashion as the spatial lag model, the spatial error model can be rewritten as:

$$\mathbf{u} = (\mathbf{I}_N - \lambda\mathbf{R})^{-1}\boldsymbol{\epsilon} \tag{1.6}$$

$$\rightarrow \mathbf{y} = \mathbf{X}\boldsymbol{\beta} + (\mathbf{I}_N - \lambda\mathbf{R})^{-1}\boldsymbol{\epsilon}, \tag{1.7}$$

The spatial error model is not efficient using OLS, but can be estimated by either Generalized Least Squares (GLS) or ML (Anselin, 1988; Kelejian and Prucha, 1999).

In order to estimate any of these models, several assumptions are placed on the spatial weight matrices, $\mathbf{W}$ and $\mathbf{R}$. It is common to row-standardise the spatial weight matrix prior to any estimation (Elhorst, 2010):

$$w_{i,j}* = \frac{w_{i,j}}{\sum_{j=1}^{N} w_{i,j}}. \tag{1.8}$$

In addition $(I_N - \rho\mathbf{W})$, respectively $(I_N - \lambda\mathbf{R})$ have to be non-singular and $|\lambda| < 1$ and $|\rho| < 1$ (Kelejian and Prucha, 1998). Kelejian and Prucha (1998, Assumption 3) state that the column sums of the matrices $\mathbf{W}$, $\mathbf{R}$, $(I_N - \rho\mathbf{W})^{-1}$ and $(I_N - \lambda\mathbf{R})^{-1}$ are bounded uniformly in absolute value. Kelejian and Prucha (2010, p. 54-55) extend this property such that, if there exists a constant $c_W < \infty$ then:

$$\max_{1\leq i\leq N} \sum_{j=1}^{N} |w_{i,j}| \leq c_W \text{ and } \max_{1\leq j\leq N} \sum_{i=1}^{N} |w_{i,j}| \leq c_W. \tag{1.9}$$

The spatial lag model can be employed in a panel data framework.[3] Adding the subscript t to each variable leads to:

$$\mathbf{y}_t = \mathbf{X}_t\beta + \rho\mathbf{W}\mathbf{y}_t + \boldsymbol{\epsilon}_t, \ t = 1, ..., T \tag{1.10}$$

For a matter of ease, it is assumed that the spatial weight matrix is constant over time.[4] Anselin et al. (2008) defines several spatial panel models with respect to the number of lags and spatial (time) lags. The most general model, the *time-space dynamic* model, includes a spatial time lag and a lag of the dependent variable:

$$\mathbf{y}_t = \phi\mathbf{y}_{t-1} + \rho_1\mathbf{W}\mathbf{y}_t + \rho_2\mathbf{W}\mathbf{y}_{t-1} + \mathbf{X}_t\boldsymbol{\beta} + \boldsymbol{\epsilon}_t. \tag{1.11}$$

This model, however, suffers from identification problems and it is more common to

[3]Spatial panel models are summarised in Anselin, Gallo and Jayet (2008) and Lee and Yu (2015).

[4]The literature on time varying spatial weights estimated by IV models or Quasi Maximum Likelihood (QML) recently received attention by Qu, Wang and Lee (2016) and Qu, Lee and Yu (2017).

restrict it to a *time-space recursive model*:

$$\mathbf{y}_t = \phi \mathbf{y}_{t-1} + \rho_2 \mathbf{W} \mathbf{y}_{t-1} + \mathbf{X}_t \boldsymbol{\beta} + \boldsymbol{\epsilon}_t. \quad (1.12)$$

This model was applied in Giacomini and Granger (2004) to spatial-time forecasting. Tao and Yu (2012) emphasise the use of spatial time lags. Leaving out a relevant lag is equivalent to an omitted variable bias problem. They show that omitting a spatial time lag causes bias and the possibility misleading inference. On the other hand, adding an irrelevant spatial time lag only causes an efficiency loss. A *time-space recursive model* without the explanatory variable will be used in Chapter 4.

1.2 Types of Cross-Sectional Dependence

Chudik, Pesaran and Tosetti (2011) and Vega and Elhorst (2016) define two forms of dependence between cross-sectional units, weak and strong dependence. Both forms of dependence can be attached to a process or a factor.

1.2.1 Cross-Sectional Dependent Processes

Assumption 2.1 and 2.2 from Chudik et al. (2011) assume the process $\mathbf{z}_{N,t} = (z_{1,t}, ..., z_{i,t}, ..., z_{N,t})'$, where $i = 1, ..., N$ and $t = 1, ..., T$ and I_t is an information set at time t, with $E(\mathbf{z}_{N,t}|I_{t-1}) = 0$ and the conditional variance $\boldsymbol{\Sigma}_{N,t} = Var(\mathbf{z}_{N,t}|I_{t-1})$ where $\boldsymbol{\Sigma}_{N,t}$ is a $N \times N$, symmetric, non-negative definite matrix. In addition, define a vector of non-stochastic weights $\mathbf{w}_{N,t} = (w_{N,1,t}, ..., w_{N,N,t})'$. The following granularity conditions have to hold:

$$\|\mathbf{w}_{N,t}\| = O(N^{-1/2}) \quad (1.13)$$

$$\frac{w_{N,j,t}}{\|\mathbf{w}_{N,t}\|} = O(N^{-1/2}), \ j = 1, ..., N. \quad (1.14)$$

Then following Definition 2.1 from Chudik et al. (2011) and the assumptions in (1.13) and (1.14), the process $z_{i,t}$ is defined to be cross-sectionally weakly dependent if

$$\lim_{N\to\infty} Var(\mathbf{w}_t'\mathbf{z}_t|I_{t-1}) = 0,\ t = 1, ..., T, \tag{1.15}$$

and cross-sectionally strongly dependent with $N \to \infty$ if

$$Var(\mathbf{w}_t'\mathbf{z}_t|I_{t-1}) \geq \kappa > 0,\ t = 1, ..., T, \tag{1.16}$$

where κ is a constant independent of N. A process can be decomposed into a strong dependent and a weakly dependent process (Chudik et al., 2011, Theorem 3.1). The process is strong if and only if there is at least one strong factor at a given point in time. Conversely, it is weakly cross-sectional dependent if it includes a weak, semi-weak or semi-strong factor at a given point in time. Thus, the driving forces for cross-sectional dependence within a process are strong and weak factors, which will be defined next.

1.2.2 Cross-Sectional Dependent Factors

The factor $f_{t,l}$, attached to the loading $\gamma_{i,l}$, is weakly cross-sectional dependent if:

$$\lim_{N\to\infty} \sum_{i=1}^{N} |\gamma_{i,l}| = \kappa < \infty \tag{1.17}$$

and strongly cross-sectional dependent if:

$$\lim_{N\to\infty} \frac{1}{N} \sum_{i=1}^{N} |\gamma_{i,l}| = \kappa > 0. \tag{1.18}$$

Cross-sectional independence is defined by $\gamma_{i,l} = 0$ $i = 1, ..., N$. Alternatively the following formulation allows to derive semi-weak and semi strong factors, using a constant α_{CSD} in the range of $0 \leq \alpha_{CSD} \leq 1$ (Chudik et al., 2011):

$$\lim_{N\to\infty} N^{-\alpha_{CSD}} \sum_{i=1}^{N} |\gamma_{i,l}| = \kappa < \infty \tag{1.19}$$

For values of $\alpha_{CSD} = 0$ the factor is weak, for $\alpha_{CSD} = 1$ the factor is strong. For intermediate values, the factor $f_{t,l}$ can be referred to as semi weak ($0 < \alpha_{CSD} < 1/2$) or semi-strong ($1/2 \leq \alpha_{CSD} < 1$).

Weak cross-sectional dependence can be thought of as the following: Even if the number of cross-sectional units increases to infinity, the sum of the effect of the common factors on the dependent variable remain constant. In the case of strong cross-sectional dependence, the sum of the effect of the common factors becomes stronger with an increase in the number of cross-sectional units.

1.2.3 Spatial Dependence and Spatial Lags

Weak cross-sectional dependence can arise due to spatial lags (Pesaran and Tosetti, 2011; Kuersteiner and Prucha, 2015). Spatial lags are the sum of weighted observations excluding the cross-sectional unit itself and therefore similar to a cross-sectional dependent process. The weights reflect typically some measure of distance, like geographical, social or economic. For example, if the nearest neighbour method is used to construct the spatial weight matrix, each cross-sectional unit is only linked to a very limited number of other units. If the number of cross-sectional units converges to infinity, then the existing linkages between the units become less important and an effect from one unit to the next one would wear out. More formally, following Pesaran and Tosetti (2011) the following spatial error model is assumed:

$$\mathbf{y} = \mathbf{X}_t\beta + \lambda\mathbf{R}\mathbf{s}_t + \epsilon, \ t = 1, ..., T, \tag{1.20}$$

where $\mathbf{R}$ is a spatial weight matrix with elements $r_{i,j}$ and has the properties as described in 1.1. The model can be rewritten on a cross-sectional level as:

$$y_{i,t} = x_{i,t}\beta + \lambda \sum_{i \neq j, j=1}^{N} r_{i,j} s_{j,t} + \epsilon_{i,t}. \tag{1.21}$$

Since **R** is bounded and using the definition in equation (1.9)

$$\max_{1\leq i\leq N}\sum_{j=1}^{N}|r_{i,j}| \leq c_R < \infty. \tag{1.22}$$

If $r_{i,j} = \gamma_{i,l}$ and $s_{j,t} = f_{l,t}$ with $l = j$, only one time period and thus $m_N = N$, then equation (1.22) is equal to the definition of weak cross-sectional dependence with $c_R = \kappa$.

The same reasoning can be applied to the spatial lag model. The spatial lag model can be written as a spatial error model with $\mathbf{R} = (I_N - \rho\mathbf{W})^{-1}$. An alternative form with the spatial lag written out is:

$$\mathbf{y}_t = \mathbf{X}_t\beta + \rho\mathbf{W}\mathbf{y}_t + \epsilon_t, \; t = 1, ..., T, \tag{1.23}$$

and as a cross-sectional level:

$$y_{i,t} = x_{i,t}\beta + \rho\sum_{i\neq j,j=1}^{N} w_{i,j}y_{j,t} + \epsilon_{i,t}. \tag{1.24}$$

As long as **W** is uniformly bounded in absolute value, $\sum_{i\neq j,j=1}^{N} w_{i,j}y_j$ can be interpreted as a weakly cross-sectional dependent factor and ρ the factor loading. The estimation of the interaction effects, namely the spatial autocorrelation coefficients ρ and λ are possible using standard spatial econometric methods. If the spatial lag model is used and the spatial weight matrix is observed, then the coefficient has a meaningful interpretation. This is an important distinction to the common factor approach and strong cross-sectional dependence, which treats the dependence as unobserved and the common factors as nuisance parameters.

This established the differences between strong and weak cross-sectional dependence and the relation between spatial and weak cross-sectional dependence. The remainder of this thesis follows Vega and Elhorst (2016) in the notation of cross-sectional dependence. Strong cross-sectional dependence is labelled as (unobserved) common factors, while weak cross-sectional dependence is referred to as spatial dependence. The only exception to this notation is the test of weak cross-sectional dependence, which will be explained next.

1.3 Dynamic Panel Data Model with Cross-Sectional Dependence

Chudik et al. (2011) and Chudik and Pesaran (2015b) assume a dynamic panel model with heterogeneous coefficients and cross-sectional dependence in the form of:

$$y_{i,t} = \alpha_i + \lambda_i y_{i,t-1} + \beta_i x_{i,t} + u_{i,t} \tag{1.25}$$

$$u_{i,t} = \sum_{l=1}^{m_f} \gamma_{i,l} f_{t,l} + \sum_{l=1}^{m_n} \nu_{i,l} n_{t,l} + e_{i,t} \tag{1.26}$$

$$\text{with } i = 1, ..., N \text{ and } t = 1, ..., T_i,$$

where $e_{i,t}$ is a cross-section unit-specific iid error term. For simplicity without loss of generality, it is assumed that only one exogenous explanatory variable exists. $\mathbf{f}_t = (f_{t,1}, ..., f_{t,m_f})'$ and $\mathbf{n}_t = (n_{t,1}, ..., n_{t,m_n})'$ are unobserved common factors. m_f and m_n are the number of factors. Both common factors are covariance stationary, have absolute summable autocovariances, are distributed independently over $e_{i,t}$ and the forth order moments are bounded. Assume that $f_{t,l}$ is a strong common factor which is possibly correlated with the regressor $x_{i,t}$ and $n_{t,l}$ is a weak, semi-weak or semi-strong common factor, which is uncorrelated with $x_{i,t}$. $\boldsymbol{\gamma}_i = (\gamma_{i,1}, ..., \gamma_{i,m_f})$ and $\boldsymbol{\nu}_i = (\nu_{i,1}, ..., \nu_{i,m_n})$ are heterogeneous factor loadings and α_i is a unit-specific fixed effect. The heterogeneous coefficients are randomly distributed around a common mean, such that $\beta_i = \beta + v_i,\ v_i \sim IID(0, \boldsymbol{\Omega}_v)$, and $\lambda_i = \lambda + a_i,\ a_i \sim IID(0, \boldsymbol{\Omega}_a)$, where $\boldsymbol{\Omega}_v$ and $\boldsymbol{\Omega}_a$ are the variance covariance matrices. λ_i lies strictly inside the unit circle. In addition, the random deviation of λ_i and β_i are independently distributed of the error term and the common factors. In the following the index t relates to the time dimension $t = 1, ..., T_i$ and index i to the cross-sections $i = 1, ..., N$. For a balanced panel it holds that $T_i = T,\ i = 1, ..., N$.

1.4 Test for weak cross-sectional dependence

If equation (1.25) is estimated without taking the error structure into account, the unobserved common factor and the heterogeneous factor loading remain a part of the error term $u_{i,t}$. In this case $u_{i,t}$ will be correlated across units, or in other words dependent across units. This renders the error not iid anymore. More important is an omitted variable bias problem, if the observed explanatory variables and the unobserved common factors are correlated, then OLS becomes inconsistent (Everaert and De Groote, 2016). Cross-sectional independence is a restrictive assumption for large panels and only strong cross-sectional dependence poses a problem (Pesaran, 2015). Thus, it is sufficient to test for weak cross-sectional dependence with the alternative of strong cross-sectional dependence.

Pesaran (2015) developed a procedure to test for weak cross-sectional dependence in large panels with N and $T \to \infty$, where N and T converge with different speeds to infinity, such that $T = O(N^{\epsilon})$ with $0 < \epsilon \leq 1$. Under the null hypothesis, the error terms are weakly cross-sectional dependent, which, using the notation from above, gives the following implicit null hypothesis:[5]

$$H_0 : 0 \leq \alpha_{CSD} < (2 - \epsilon)/4. \tag{1.27}$$

The test statistic is

$$CD = \sqrt{\frac{2T}{N\,(N-1)}} \left(\sum_{i=1}^{N-1} \sum_{j=i+1}^{N} \hat{\rho}_{ij} \right) \tag{1.28}$$

$$\hat{\rho}_{ij} = \hat{\rho}_{ji} = \frac{\sum_{t=1}^{T} \hat{u}_{i,t} \hat{u}_{jt}}{\left(\sum_{t=1}^{T} \hat{u}_{it}^2\right)^{1/2} \left(\sum_{t=1}^{T} \hat{u}_{jt}^2\right)^{1/2}} \tag{1.29}$$

[5]For a formal derivation of the null hypothesis, see Pesaran (2015).

where $\hat{\rho}_{ij}$ is the correlation coefficient.[6] In the case of an unbalanced panel, the correlation coefficient is calculated for the common sample as outlined in Chudik and Pesaran (2015b):

$$\hat{\rho}_{ij} = \hat{\rho}_{ji} = \frac{\sum_{t \in T_i \cap T_j} \left(\hat{u}_{it} - \bar{\hat{u}}_i \right) \left(\hat{u}_{jt} - \bar{\hat{u}}_j \right)}{\left[\sum_{t \in T_i \cap T_j} \left(\hat{u}_{it} - \bar{\hat{u}}_i \right)^2 \right]^{(1/2)} \left[\sum_{t \in T_i \cap T_j} \left(\hat{u}_{jt} - \bar{\hat{u}}_j \right)^2 \right]^{(1/2)}} \tag{1.30}$$

where

$$\bar{\hat{u}}_i = \frac{\sum_{t \in T_i \cap T_j} \hat{u}_{it}}{T_{ij}}, \quad T_{ij} = \# \left(T_i \cap T_j \right), \tag{1.31}$$

where $T_i \cap T_j$ are the common periods of unit i and j and $\# \left(T_i \cap T_j \right)$ is the number of common periods. The CD test statistic becomes then

$$CD = \sqrt{\frac{2}{N \left(N - 1 \right)}} \left(\sum_{i=1}^{N-1} \sum_{j=i+1}^{N} \sqrt{T_{ij}} \hat{\rho}_{ij} \right). \tag{1.32}$$

Under the null, the CD test statistic is asymptotically

$$CD \sim N(0,1) \tag{1.33}$$

distributed.

1.5 Common Correlated Effects Estimators

The previous section explained that ignoring strong cross-sectional dependence leads to inconsistent OLS estimates. Two approaches how to take out the cross-sectional dependence are discussed in the literature. The first is based on principle components and proposed by Coakley, Fuertes and Smith (2002). The common factors are approximated by the principle components of the residuals of a first-stage regression. The principle components are then used as observed explanatory variables in the second stage. Bai (2009) builds on Coakley et al. (2002) and proposes an

[6] The index for the time periods is omitted for the balanced panel.

interactive-effects estimator. The common factors are estimated by principal components, which are then used to estimate the coefficients of known explanatory variables. The advantage of the principal component approaches is that an estimation of the factor loadings is possible. However, they come at the disadvantages of the assumption of homogeneous slopes, implicitly assuming strong dependence and the number of common factors needs to be known.

This thesis focuses on heterogeneous slope models and therefore an alternative approach using cross-sectional averages is at the centre of interest. Pesaran (2006) shows that the static version of equation (1.25)

$$y_{i,t} = \alpha_i + \beta_i x_{i,t} + u_{i,t} \tag{1.34}$$

can be consistently estimated by approximating the unobserved common factors with cross-section averages $\bar{x}_t = 1/N \sum_{i=1}^{N} x_{i,t}$ under strict exogeneity of $x_{i,t}$.[7] This estimator is commonly known as the Common Correlated Effects (CCE). The underlying idea of CCE estimator is to eliminate asymptotically the differential effects of unobserved common factors by cross-sectional averages (Pesaran, 2006, p. 969). The cross-sectional averages inhibit strong cross-sectional dependence and are suitable to approximate the common factors.

The estimator was proved to be consistent under a variety of further assumptions, such as a large number of unobserved common factors (Chudik et al., 2011), serial correlation of the common factors (Kapetanios, Pesaran and Yamagata, 2011), calculation of long run coefficients (Chudik, Mohaddes, Pesaran and Raissi, 2016), approximating dominant units (Chudik and Pesaran, 2013) and pooled homogeneous models with bias correction (Everaert and De Vos, 2016). In empirical applications it was used for example in Holly, Pesaran and Yamagata (2010), Eberhardt et al. (2012), Bond and Eberhardt (2013), McNabb and LeMay-Boucher (2014) and Gundlach and Paldam (2016). The CCE estimator was designed for static panels.

[7] Unlike Pesaran (2006) the unit-specific fixed effect is kept and not partialled out. See discussion in Section 3.4.

In dynamic panels as:

$$y_{i,t} = \alpha_i + \lambda_i y_{i,t-1} + \beta_i x_{i,t} + u_{i,t}, \tag{1.35}$$

where the idiosyncratic errors $u_{i,t}$ are cross-sectionally weakly dependent and $E(\lambda_i) = \lambda$, the lagged dependent variable is no longer strictly exogenous and therefore the unit-specific estimates of β_i are asymptotically biased (Nickell, 1981). Chudik and Pesaran (2015a) show that the Mean Group (MG) estimator remains consistent if the floor of $\sqrt[3]{T}$ lags of the cross-section averages are added for both the dependent variable and the strictly exogenous variables. The number of lags is denoted by $p_T = \left[\sqrt[3]{T}\right]$. The equation to be estimated is

$$y_{i,t} = \alpha_i + \lambda_i y_{i,t-1} + \beta_i x_{i,t} + \sum_{l=0}^{p_T} \boldsymbol{\delta}'_{i,l} \bar{\mathbf{z}}_{t-l} + e_{i,t}, \tag{1.36}$$

where $\bar{\mathbf{z}}_t = (\bar{y}_t, \bar{x}_t) = (1/N \sum_{i=1}^N y_{i,t}, 1/N \sum_{i=1}^N x_{i,t})$ are the cross-sectional averages. λ_i and β_i are stacked into $\boldsymbol{\pi}_i = (\lambda_i, \beta_i)$. The MG estimates are

$$\hat{\boldsymbol{\pi}}_{MG} = \frac{1}{N} \sum_{i=1}^{N} \hat{\boldsymbol{\pi}}_i. \tag{1.37}$$

In the following, this estimator is called the Dynamic Common Correlated Effects (DCCE) estimator. $\hat{\boldsymbol{\pi}}_i$ is consistently estimated if $(N, T, p_T) \xrightarrow{j} \infty$ such that $p_T^3/T \rightarrow \varrho_1, 0 < \varrho_1 < \infty$ and $N/T \rightarrow \varrho_2, \varrho_2 > 0$ and under full rank of the factor loadings. Then $\hat{\boldsymbol{\pi}}_i$ and $\hat{\boldsymbol{\pi}}_{MG}$ converge in probability (Chudik and Pesaran, 2015a, Theorem 1 and 2, p. 397-398):

$$\hat{\boldsymbol{\pi}}_i - \boldsymbol{\pi}_i \xrightarrow{p} \mathbf{0} \tag{1.38}$$

$$\hat{\boldsymbol{\pi}}_{MG} - \boldsymbol{\pi} \xrightarrow{p} \mathbf{0}. \tag{1.39}$$

The conditions translate into a sufficient number of lags and cross-sectional averages to approximate the unobserved common factors. The requirements for consistency on the two estimators can be interpreted for each separately. The unit-specific estimates can be obtained from a simple regression on a single cross-section unit. Therefore, the requirement for consistency is $T \to \infty$. A relative expansion rate for N and T is not required. The number of cross-sectional lags is restricted in order to maintain a sufficient number of degrees of freedom and therefore the requirement on the number of lags is necessary. For consistency of the MG estimates N and T grow jointly to infinity $(N,T) \xrightarrow{j} \infty$. The cross-sectional dimension has to approach infinity due to the heterogeneous coefficients. The time dimension has to grow to reduce the time series bias due to the lagged dependent variable. The assumption on the rank of the factor loadings can be relaxed if the factors are serially uncorrelated. The MG estimates remain consistent even if the factors are serially correlated. For a more in-depth discussion of consistency, see Chapter 3 in Chudik and Pesaran (2015b) or Chudik and Pesaran (2015a).

Under these assumptions, the asymptotic variance for the mean group estimates is consistently estimated by

$$\widehat{\mathbf{Var}}(\hat{\boldsymbol{\pi}}_{MG}) = N^{-1}\hat{\boldsymbol{\Sigma}}_{MG} = \frac{1}{N(N-1)}\sum_{i=1}^{N}\left(\hat{\boldsymbol{\pi}}_i - \hat{\boldsymbol{\pi}}_{MG}\right)\left(\hat{\boldsymbol{\pi}}_i - \hat{\boldsymbol{\pi}}_{MG}\right)'. \tag{1.40}$$

The mean group estimates have the following asymptotic distribution (Chudik and Pesaran, 2015a):

$$\sqrt{N}\left(\hat{\boldsymbol{\pi}}_{MG} - \boldsymbol{\pi}\right) \xrightarrow{d} N(0, \boldsymbol{\Sigma}_{MG}). \tag{1.41}$$

Pesaran (2006) considers a pooled version of the common correlated effects estimator, with the constraint $\boldsymbol{\pi}_i = \boldsymbol{\pi}$ for $i = 1, ..., N$. In case of equal weights to all observations, $w_i = 1/N$, the common correlated effects pooled estimator for $\boldsymbol{\pi}$, $\hat{\boldsymbol{\pi}}_P$, collapses to a simple OLS estimator and is asymptotically unbiased. Everaert and De Groote (2016) show that a common correlated effects pooled estimator is consistent even in a dynamic panel as long as $(N,T) \to \infty$. For finite samples with $N \to \infty$ and T fixed, Everaert and De Groote (2016) and Everaert and De Vos (2016) suggest a restricted version of the pooled estimator.

The asymptotic distribution for the pooled estimator is

$$\sqrt{N}\left(\hat{\boldsymbol{\pi}}_P - \boldsymbol{\pi}\right) \xrightarrow{d} N(0, \boldsymbol{\Sigma}_P). \tag{1.42}$$

Following Pesaran (2006), a non-parametric variance estimator for $\hat{\boldsymbol{\pi}}_P$ is given by:

$$\widehat{\mathbf{AVar}}\left(\hat{\boldsymbol{\pi}}_P\right) = \frac{1}{N}\hat{\boldsymbol{\Psi}}^{*-1}\hat{\boldsymbol{\Xi}}^{*}\hat{\boldsymbol{\Psi}}^{*-1} \tag{1.43}$$

with

$$\hat{\boldsymbol{\Xi}}^{*} = \frac{1}{N-1}\sum_{i=1}^{N}\left(\frac{\tilde{\boldsymbol{X}}_i'\tilde{\boldsymbol{X}}_i}{T}\right)\left(\hat{\boldsymbol{\pi}}_i - \hat{\boldsymbol{\pi}}_{MG}\right)\left(\hat{\boldsymbol{\pi}}_i - \hat{\boldsymbol{\pi}}_{MG}\right)'\left(\frac{\tilde{\boldsymbol{X}}_i'\tilde{\boldsymbol{X}}_i}{T}\right) \tag{1.44}$$

where $\tilde{\boldsymbol{X}}$ are the explanatory variables with the cross-sectional averages partialled out and

$$\hat{\boldsymbol{\Psi}}^{*} = \sum_{i=1}^{N}\frac{1}{N}\left(\frac{\tilde{\boldsymbol{X}}_i'\tilde{\boldsymbol{X}}_i}{T}\right). \tag{1.45}$$

1.6 Empirical Growth Models and Convergence

Empirical Growth Models Islam (2003) defines two major strands of growth theories, the neoclassical growth theory and the new growth theory. The former includes convergence, but lacks the ability of generating long-term growth from within the model. This is mainly due to the assumption of a constant external growth driving force, such as technological growth. The new growth theory puts more emphasis on growth from within the model such as human capital, innovation or spillover effects. As the growth is endogenous, those models are called endogenous growth models.

One of the key contributions to the endogenous growth models is the Romer growth model (Romer, 1990). It assumes a Cobb-Douglas production function, with human capital, labour and an intermediate or durable good as factors of production.

The change in the stock of knowledge, which drives the number of available durable goods, depends on the existing stock of it and the stock of human capital. The returns to human capital are constant, implying the larger the stock, the more new knowledge is accumulated. The spillover between the knowledge-producing sector and the durable goods producing sector produces long run growth. On the balanced growth path, the growth rate depends solely on the stock of human capital. As the returns to human capital are constant, the model predicts unlimited growth.

The Solow growth model (Solow, 1956) based on the Harrod-Domar model is among the most widely used neoclassical growth models and offers a solid framework for growth regressions. Income is a function of physical capital and labour. Due to decreasing returns of scale, the model predicts convergence to a steady state, which depends on the saving rate for physical capital, the population growth rate and the depreciation rate of capital. It identifies as determinants of a country's output the stock of physical capital, the level of technology, and the number of workers. The input factors, physical capital and labour, are paid their marginal products. Moreover, the model assumes diminishing returns to scale with respect to the input factors and a constant and exogenous rate of technological progress. The assumptions allow an economy to converge on a transition path towards a steady state. In the steady state, the model depends on the saving rate, the depreciation rate of capital and population and technological growth. Barro and Sala-i Martin (1992) employ a simple form of the Solow model by regressing the growth rate between the initial and current level on the initial level of GDP. Mankiw, Romer and Weil (1992) estimate the Solow model in a cross-section of countries. They include the saving rate for capital and population growth rate as explanatory variables. However, their main contribution is to extend the production function of the Solow growth model by human capital and present evidence for it using cross-sectional data. Islam (1995) shows that their results are exposed to omitted variable bias and therefore invalid. He emphasises the use of panel data models to overcome this weakness. A further advantage of panel data models is, that it allows for the estimation of heterogeneous slopes. Under the assumption of parameter, or slope, homogeneity a change in an explanatory variable such as initial or lagged Gross Domestic Product (GDP) has the same effect on the growth rate for a country in Africa and the United States of America (US). Durlauf (2001) and Brock and Durlauf (2000) point out, that there is no reason to assume parameter heterogeneity in a setting with such different countries. Lee, Pesaran and Smith (1997) employ the MG estimator to a growth equation with heterogeneous slope coefficients. Their approach was further discussed in Lee, Pesaran and Smith (1998) and Islam (1998).

Convergence A key feature of a growth model and important for its policy recommendations is convergence. The literature distinguishes several types of convergence, such as across or within countries, convergence in growth rates or levels (Islam, 2003). Important for this thesis are the types in levels: absolute and conditional convergence and an intermediate between these two, club convergence. Convergence is an outcome of neoclassical growth models and requires diminished returns to the factors of production. This implies that the higher the stock of, say capital, the smaller are the gains of an additional increase. The growth rate will decline but remain positive, eventually approach zero and the level converges to a fixed number, the steady state.

Absolute convergence describes that all countries converge to the same steady state.[8] In a simple growth regression, the difference of an end of period income is regressed on an initial income (Barro, 1991; Barro and Sala-i Martin, 1992; Mankiw et al., 1992). In a panel data model, such as

$$log\left(\frac{y_{i,t}}{y_{i,t-1}}\right) = \alpha_i + \beta_i log(y_{i,t-1}) + \epsilon_{i,t}, \qquad (1.46)$$

convergence appears if $|1+\beta_i| < 1$ or $-2 < \beta_i < 0$. In the case that $\beta_i \geq 0$, then the individual country does not converge, as the growth would continue to increase. For $\beta_i \leq -2$, the country oscillates around a steady state. Absolute convergence

[8] It is common to call absolute convergence *unconditional convergence*, see (Islam, 2003, p. 312).

appears if $\alpha_i = \alpha$, for $i = 1, ..., N$, $\beta_i = \beta$ and $-2 < \beta_i < 0$, for $i = 1, ..., N$. The implication is, rich countries have a lower growth rate than poor countries, allowing the poor ones to catch up. If α_i is allowed to be heterogeneous, then each country converges to its own steady state. This concept is called conditional convergence.[9] It came out of the fact that absolute convergence was found among country groups such as the Organisation for Economic Co-operation and Development (OECD), but not for countries world-wide. Depending on the difference of α, divergence between countries becomes possible. Conditional convergence is criticised as being inappropriate in making any statements about whether differences between countries diminish. It was labelled as "hollow" (Islam, 1998, p. 1162) and "...an empty construct" (Islam, 1995, p. 326). Club convergence is a slightly restricted version of conditional convergence. It differs from conditional convergence in the sense that a group of countries is allowed to converge to the same steady state. While conditional convergence allows a different unique equilibrium for all countries, club convergence allows for multiple equilibria. Country characteristics determine to which equilibrium a country converges. In a sense, club convergence lies between unconditional and conditional convergence.

A limitation of the growth theories mentioned above is that they treat countries as isolated units and interactions between countries are ignored. This applies to the theory, the estimation of the parameters as well as to the conditions for convergence. Growth models like the Solow growth model (Solow, 1956) are a single country model and do not include any interactions or dependencies between countries. As pointed out by De Long and Summers (1991), there is no reason to believe that the growth process of a country is completely isolated from the rest of the world. Countries interact with each other by trade, knowledge transfers and humans migrate and spread their ideas. It is no coincidence that the first of the New Kaldor Facts laid out by Jones and Romer (2010) emphasizes the importance of dependence between countries. For example, Comin and Hobijn (2010) show diffusion of new technology

[9] Conditional convergence can be obtained also with a homogeneous α_i, if further explanatory variables are added. Then convergence is conditional on the explanatory variables.

intensifies over time. From an applied econometric perspective, spatial econometrics offers methods to incorporate interactions between countries in empirical models. Empirical growth models in a spatial econometric context will be discussed next.

1.7 Growth Empirics and Spatial Econometrics

Within a growth model, channels for the effect of space on growth are mobility of the factors of production or technology. Among the most notable contribution is Ertur and Koch (2007). They derive a spatial Solow growth model by extending it with a spatial lag of income per capita, and spatial lags of the break-even investments, the sum of population growth, technological change and the depreciation rate of physical capital. The model is applied to the Penn World Tables (PWT) 6.1 and estimated by ML. The spatial weights are pure geographical distance. Their findings show a significant effect of the spatial autocorrelation coefficient and they present evidence for the simple Solow growth model without human capital. On a regional level, Lopez-Bazo, Vaya and Artis (2004) and Fingleton and López-Bazo (2006) include externalities in their estimation equations in the form of a spatial lag and a spatially weighted initial income. Lesage and Fischer (2008) estimate a growth equation with a spatial Durbin model on data for 255 regions in 25 EU countries. Their findings are that characteristics of neighbouring regions, how the region is connected to its neighbours and the strength of this connection are more important than direct effects of the region itself. Koch (2008) takes space explicitly into account within a growth empirical setting, emphasizing in particular the importance of interdependence of Total Factor Productivity (TFP) across countries. In his estimation of a spatial error model, factor productivity as captured by the error term not only depends on the country itself but on other countries' productivity as well. The paper finds that if spatial effects are ignored, results using traditional methods are biased. Elhorst, Piras and Arbia (2010) estimate an extended Solow-Swan growth model using European regions. They estimate an unconstrained spatial Durbin model using GMM, ML and a mixture of both, which allows the inclusion of fixed effects

and spatial interactions. Ertur and Koch (2011) show that spatial dependencies in TFP matters for a multi-country Schumpeterian growth model augmented with technological interdependencies between countries.
Arbia and Paelinck (2003a,b) estimate regional convergence for 119 European regions using a Lotka-Volterra approach. They estimate a difference equation system using OLS, where each region is represented by one equation and use the stability conditions of this system to determine convergence. This approach implies that convergence depends on the parameters of the regions itself and of parameters of other regions.

A very common and obvious choice for the spatial weight matrix is geographical distance or a related measure. For the context of this thesis, the distinction between spatial interactions and control variables is crucial. Variables, such as climate and temperature (Sachs, 2001), disease environment (Acemoglu and Johnson, 2007), genetic distance (Spolaore and Wacziarg, 2009) control for the location, but they do not relate cross-sectional units to each other.

1.8 Summary

This chapter identified several gaps in the literature, and working towards closing them is the aim of this thesis. First of all, empirical growth models require taking interdependences between countries into account, since otherwise estimates are biased and inconsistent and cannot as such be used to refute or accept any specific theory of growth. An empirical growth model, and with it interdependence between countries, needs to be motivated from and founded in a theoretical model. In order to do so, the next chapter lays out a growth model which models interactions between countries explicitly, in the form of migration of high skilled workers, diffusion of ideas and trade in goods. The micro foundation for trade and migration is taken from the New Economic Geography and Growth (NEGG) literature. The growth rate directly depends on the number of high skilled workers moving from one to the next country and on the diffusion of ideas. The model predicts a positive effect of migration and diffusion of ideas on the growth rate, and in the short term of trade. Explosive growth and divergence are the outcomes of the model, as in the Romer model.

The econometric literature defines two types of dependence between units in a panel with a large number of observations over time and units, spatial dependence and common factors. To obtain valid results, both forms of dependence have to be taken care of. Chapter 3 develops a Stata package to estimate an equation in the presence of common factors, which is the main contribution of this chapter. The chapter explains the options of the command and presents empirical examples. A Monte Carlo simulation is carried out to shed light on the bias due to small samples, heterogeneous slope coefficients and common factors. In addition it is shown, that the bias due to a misspecification of heterogeneous slopes in the presence of homogeneous slopes is relatively small.

The last chapter estimates a growth model with both types of cross-sectional dependence. Common factors are approximated by cross-sectional averages and spatial dependence by spatial time lags. The spatial time lags are motivated by the model in Chapter 2. Accounting for both types of dependence is a recent development and has not been done in a growth empirical setting. As pointed out in the introduction, there is no reason to believe that all countries have the same slope parameters. The estimation method allows for country specific estimates of the coefficients. The equation relies on the Lotka-Volterra approach to determine the type of convergence in the presence of interdependence between countries. A further contribution is the extension of the Lotka-Volterra model with an approach to allow for testing the conditions for convergence.

Chapter 2

A Growth Model with Mobile Labour, Trade and Diffusion of Ideas

2.1 Introduction

Corrado and Fingleton (2012) argue that the frequent criticism of the current growth empirics literature in spatial econometrics, that it lacks a link to economic theory, is misplaced. Specifically, in this literature spatial externalities arise from technological interdependence among countries, where knowledge accumulated in one country depends on knowledge accumulated in other countries. However, the specific patterns of spatial interactions and dependencies in empirical growth specifications are not theoretically founded: they are assumed to depend on geographic distances. This is because a theory of why interactions between two countries matter for economic growth is largely unexplored.

This chapter aims to close this gap by combining a Romer (1990) style endogenous growth model and elements from a New Economic Geography (NEG) in the style of Krugman (1991); Forslid and Ottaviano (2003) and Baldwin and Martin (2004). The setting of the sectors, the mobility of labour and the consumption maximizing behaviour of works are drawn from the New Economic Geography models. The engine of growth originates from the Romer (1990) model. The interaction between countries is modelled in three ways. The first connection is the migration of high skilled workers, the second connection is trade of a manufactured good and the third is the diffusion of ideas. The engine of growth draws upon the endogenous growth model and therefore the model predicts divergence. Both the migration of largely high skilled workers and the prediction of divergence mimics the economic development during the Great Divergence in the 19th century. Equally, the model can relate to club divergence in current times.

The two-country model is based on three sectors in each country. The first, the Research and Development (R&D) sector, uses high skilled workers and the existing stock of designs to produce new designs. Designs can be seen as patents or ideas to found new firms in the second sector, the manufacturing sector. Before these firms can start production, they need to buy exactly one design from the R&D sector. Low skilled labour and a fraction of high skilled workers are employed to produce a firm-specific differentiated manufactured good. The last sector, the traditional sector, uses solely low skilled labour to produce a homogeneous traditional good. This sector can be interpreted as a food producing agricultural sector.

The engine of growth is the R&D sector, which accumulates designs. Designs are converted into new output enhancing firms. Therefore, the number of designs represents the number of firms operating in the manufacturing sector. Different to Romer (1990), the growth engine in this model lies in the varieties of goods produced by the manufacturing sector rather than the variety of intermediate goods (for example, machinery) which are transformed into one final homogeneous good. In line with the Romer model, a higher permanent growth rate can be achieved by an increase in high skilled workers employed in the R&D sector. In distinction to the Romer model, diffusion of ideas is taken into account as well. If ideas spread out, a permanent increase in the growth rate is possible. Hence, the model recognizes two ways to increase the growth rate: migration and diffusion of ideas.

Low skilled workers are immobile between countries but mobile between the two sectors, manufacturing and traditional. Some high skilled workers migrate to the higher wage offering country and contribute to the R&D sector there, allowing that country to grow faster. Corner solutions as well as interior solutions are possible. If the wage differential is very small, only a fraction of high skilled workers migrate. With an increase in the difference between wages, the interior solution moves towards the corner solution, in which all high skilled workers migrate at once. Trade has no direct effect on the growth rate. However, trade creates a temporary effect on the growth rate due to specialization and the re-allocation of low skilled labour between the traditional and manufacturing sector. Diffusion of ideas allows a country to catch up, if none of the other interactions are possible, or to reduce the difference between the countries.

The novelty of this model, arising from the features of the endogenous growth model and inspired by Krugman (1991), Forslid and Ottaviano (2003) and Baldwin and Martin (2004), are the three channels for interactions between countries: high skilled worker mobility, trade and the diffusion of technology, namely ideas. Similar to the Romer (1990) model, this model predicts unbounded growth, an outcome that has important implications. The country that receives high skilled workers experiences unbounded growth, while the other country becomes stagnant, implying divergence. Therefore, the model can be applied to the Great Divergence during the 19th century. A distinct feature of the Great Divergence was migration accompanied by strong economic growth in some parts of the world. The model is also consistent with economic development in the 20th century. At a broad level, it helps to understand the formation of convergence clubs, even if the convergence or catch-up processes between countries within each club is not explicitly considered.

The difference in technological development between the beginning and the end of the 19th century is vast.[1] In the 19th century, the "West" (Western Europe, US, Canada, Australia and New Zealand) started growing rapidly and left large parts of

[1] Jewkes, Sawers and Stillerman (1961) provides a comprehensive overview of important inventions in the 19th century, such as the high pressure steam engine and combustion engine, innovations in the textile industry, telegraph and telephone, electric lamps, rubber and steel.

the world behind. Maddison (2007, p. 70-71) shows that the "West" experienced a growth rate of 1.07% between 1820 and 1870 and a rate of 1.56% between 1870 and 1913. GDP per capita increased from $1,202 to $3,988 between 1820 and 1913. The rest of the world remained nearly stagnant with growth rates of 0.1% between 1820 and 1870 and 0.86% between 1870 and 1913, respectively. Over the century (93 years) GDP per capita increased only from $667 to $1,526.[2] In these times one of the main ways to distribute knowledge was movement of human capital, and hence high skilled migration was important. [3] During the 19th century the first migration wave started. By the middle of the century, approximately 300,000 migrants dared to undertake the journey from Europe to the United States each year. The numbers rose to more than 500,000 at the end of the century, and to more than a million by the beginning of the 20th century (Williamson, 2006). Even within Europe there were large migration flows. About half of the Italian emigrants stayed in Europe, especially France and Germany, while 9 percent of the population in large British cities were Irish-born (Williamson, 2006).

The remainder of this chapter is structured as follows. Section 2.2 reviews the literature, followed by the discussion of the model (Section 2.3). Section 2.4 reports on a simulation of the model and finally section 2.5 concludes.

2.2 Literature

The literature on theories of economic growth on a regional and country level is both vast and very diverse (Temple, 1999, 2003; Islam, 2003). Selected contributions to the literature are discussed next, as a motivation for, and aid towards, developing the model in section 3.

[2] See Maddison (2007), p. 70-71. GDP per capita is measured in 1990 international dollars.

[3] High skilled workers were not necessarily well educated. In the context of the 19th century, skill can be seen more as an ability to perform sophisticated jobs and to create ideas. For a more detailed definition see Chapter 2.3.

2.2.1 Growth Models

The introduction discussed two strands of the growth literature, neoclassical growth models and endogenous growth models. In the light of this thesis, the advantages of the latter are the existing spillovers from the knowledge producing sector to the durable producing sector, allowing an extension across countries. In the Romer model the growth engine is a design producing R&D sector. The designs are transferred into machines by an intermediate goods sector, which sells the machines to the final goods sector. The growth rate depends positively on the number of workers employed in the R&D sector and negatively on the interest rate for machines (capital). This implies the larger the amount of high skilled workers allocated to the R&D sector, the higher the growth rate. As the production function of the R&D sector is linear in its factors of production, the model predicts unbounded growth.

The channels to model the interaction between countries are the mobility of the factors of production and the diffusion of technology. Diffusion of technology is discussed in the next section; the literature on the mobility of labour will be summarised in Chapter 2.2.5. Another factor that influences the factors of production is geography. However, the effect is discussed more in an empirical rather than a theoretical fashion and more in the sense of a control rather than an interaction. Geography justifies the approach in Acemoglu et al. (2001) of using a settler mortality, a geography-related instrument, to account for the endogeneity of institutions. Spolaore and Wacziarg (2013) argue that the main channels of the geography effect are either directly on the factors of production or indirectly through history. Direct effects include climate and temperature (Sachs, 2001), while examples of the indirect effects are diseases (Acemoglu and Johnson, 2007), genetic distance (Spolaore and Wacziarg, 2009), and/or ancestral origin of the current population of a country (Putterman and Weil, 2010).

2.2.2 Diffusion of Ideas

Diffusion of ideas, or technology, can occur within a country or between countries.[4] The first kind is described for example in Jones (1995, 2005). Examples for the second type of diffusion are given in Eaton and Kortum (1999), Lucas (2009a) and Alvarez, Buera and Lucas (2013). What all these models have in common is that diffusion between countries in terms of time and implementation depends on an underlying distribution of the quality of ideas. In Eaton and Kortum (1999) ideas do not diffuse immediately and they differentiate between diffusion and adoption. Only the highest quality ideas, namely those closest to the technological frontier, are adopted. Lucas (2009a) extends the model by Eaton and Kortum (1999) to an overlapping generation model. Ideas are created by chance, depending on the current state of ideas, and diffuse as long as people live. Alvarez et al. (2013) build a model in which diffusion of ideas takes place via trade. This is similar to Rivera-Batiz and Romer (1991). In an extension of the Romer model, Rivera-Batiz and Romer identify different channels for the diffusion of technology or ideas. The first channel is that designs are tradeable between countries. Firms can choose between domestic or foreign designs ("Flows of Goods", Rivera-Batiz and Romer (1991)), while the R&D sector depends solely on the domestic stock of designs. Diffusion has no impact on the growth rate as the allocation of human capital between the manufacturing and research sector remains the same in this model. The second possibility is the "Flow of Ideas" (Rivera-Batiz and Romer, 1991), where foreign designs are freely available to domestic firms. As the return of human capital in the R&D sector increases, more high skilled workers are shifted into research and the growth rate increases. Comin and Hobijn (2010) for example show that the diffusion of new technology intensifies over time. They find that the average adoption lag of technology is about 45 years, with a standard deviation of 39 years, and shrinks over time.

[4]The literature is summarised for example in Grossman and Helpman (1991) and Klenow and Rodriguez-Clare (2005).

2.2.3 New Economic Geography

New Economic Geography (NEG) models, initiated by Krugman (1991), provide a natural way to integrate space into theories of economic growth. The two-region two-sector model explains the concentration of skill-intensive manufacturing firms in a region with respect to the consumption maximization behaviour of mobile, high skilled workers and thus divergence between the two regions. While the NEG models provide useful explanation for agglomeration, they do not have an explicit source or engine of growth, and are therefore not adequate for modelling the growth process in itself.

An extension of the NEG models are NEGG models, such as Baldwin and Forslid (2000), Baldwin and Martin (2004) and Cerina and Pigliaru (2007) and summarised in Bond-Smith and McCann (2014). A capital accumulating sector is added to the economy and works as the engine of growth. However, from a growth theory perspective there are some drawbacks. Firstly, the concept of capital is not clearly defined in the NEGG models. Baldwin and Martin (2004) describe immobile capital as human capital while mobile capital would be physical capital or patents. From a growth theory perspective, this ambiguity is problematic because there is considerable distinction between physical and human capital which is crucial for interpretation and policy.[5] Secondly, there is no distinction between high skilled workers and low skilled workers. This rules out the role of migration of high skilled workers as a driver of cross-country interactions. Finally, the effect of geographical agglomeration on income growth lies at the core of these models, while economic growth is only a by-product of agglomeration.

[5] The discussion of the role human capital plays in economic growth is widely discussed in the literature. See, for example, Lucas (1988), Mankiw et al. (1992), Benhabib and Spiegel (1994), Barro and Sala-i Martin (2004) and reviews by Temple (1999) or Barro (2001).

2.2.4 Trade

Trade is another possible channel for interactions between countries. The literature of the impact of trade on economic growth is heavily influenced by the overview in Grossman and Helpman (1991).[6] In an earlier work, Grossman and Helpman (1990) propose a three sector model in which the final good can be traded. They find that a reduction in tariffs only contributes towards growth if the tariff is set by the country with a comparative advantage or the country with the smaller effective labour force. Feenstra (1996) builds a two-country growth model, in which the interaction between the countries is trade in the final good. The model predicts divergence, as the larger country will gain from trade due to productivity gains from R&D, leading to a higher growth rate. In Alvarez et al. (2013) trade has a positive effect on economic growth as inefficient producers are replaced by foreign ones that are more efficient. Grossman and Helpman (2016) build a growth model with heterogeneous workers and firms. In their work, the dependence between trade, inequality within a country, funding of the R&D sector and economic growth is in the foreground. The growth driving force is a R&D sector as in the Romer model. Driven by wage, workers sort themselves into doing certain activities and match with certain types of firms, matching their skills and abilities. The sorting manifests a wage distribution within the country, which leads to inequality between the different types of workers. Allowing for trade in the intermediate good, Grossman and Helpman show that the long run growth rate and inequality rises in every country. Additionally, differences in the financial endowment of the R&D sector lead to a further diversion of growth rates across countries and inequality within countries.

The literature on the empirics of growth and trade is not in the foreground of this chapter but relevant for this thesis and so worth noting. Sachs and Warner (1995) relate openness to trade and economic growth. Frankel and Romer (1999) find a small but positive effect of trade on income. They control for country size and geographical location. In a more recent work, Wacziarg and Welch (2007)

[6] Another good summary is Ventura (2005).

find evidence based on the work by Sachs and Warner (1995) that liberalisation to trade increases income. However, the magnitude of this effect is hard to separate from other effects, such as political instability. Basu and Bhattarai (2012) relate openness and trade to human capital and educational spending. Their finding is that cognitive skills are an important driver of the relationship between education and trade. Overviews can be found for example in Rodriguez and Rodrik (2000) or Wacziarg (2001).

2.2.5 Migration

Migration and its impact on the host country as well as on the country of origin are mainly discussed from a labour or development economic perspective.[7] By contrast, the work in this thesis uses migration to model the interaction between countries, thereby focusing on the connection between high skilled migration and economic growth.

The model in this chapter is related to the works in Braun (1993), Klein and Ventura (2009) and Kennan (2013). Barro and Sala-i Martin (2004, Ch. 9.1.3) summarize the model in Braun (1993). Braun introduces migration into a Ramsey model. Costs of migration increase with the number of migrants, which then decreases the speed of convergence as it decreases per capita output in the receiving country. Klein and Ventura (2009) build a growth model for two economies with different technology, which produce a single good using capital, labour and land. Workers migrate into the more productive country and help to increase the output. They adjust their model for the enlargement of the EU and the NAFTA deepening. To offset the gains from migration an increase in capital income tax of 40% to 45% would be needed. However, their overall focus lies on the effects of an integration of two countries, rather than comparing two separate countries. Kennan (2013)

[7] See, for example, reviews by Borjas (1994) or Ottaviano and Peri (2012) for the effects on labour markets. In development economics an important topic is the brain drain from underdeveloped countries (Docquier and Rapoport, 2012; Dequiedt and Zenou, 2013). The literature has also considered the effect of migration on the income distribution within a country; see, for example, Ben-Gad (2004).

includes migration into a Heckscher-Ohlin trade model. The decision to migrate is driven by the probability of leaving the home country, which depends on the utility costs of migration. His conclusion is that if borders were open the gain from migration from a less developed into a developed country would be more than $10,000. This large gain is associated with only a small loss in real wages in the developed country. On a regional level for 27 European regions, Huber and Tondl (2012) find evidence that migration has a positive effect on the income per capita of the destination country, while it negatively effects the country of origin. This result suggests a divergence between the countries of destination and origin.

This section highlights the recognition of interactions between countries in the empirical growth literature. The literature assumes spillovers, mainly diffusion of knowledge, without providing adequate theory for how and why knowledge spreads. In addition, the above mentioned examples underline the importance of mobility of workers. Interactions between countries are modelled in the New Economic Geography literature. However, these models either lack an engine of growth or are limited by a lack of richness in modelling physical and human capital.

2.3 A Growth Model with Mobile Labour

This section develops a model of economic growth, based on the Romer model and drawing from models of the New Economic Geography literature, especially Krugman (1991), Baldwin and Martin (2004) and Forslid and Ottaviano (2003). In addition, the model is related to the work in Feenstra (1996).

This is a two-country model, with one rich or developed country, indexed by i, and one poor or less developed country, indexed by j. Following the model in Baldwin and Martin (2004), the economies in both countries consist of three sectors: a R&D, a manufacturing and a traditional goods sector. The R&D sector produces designs that are used by the manufacturing sector to found firms.[8] Each

[8]In the remainder of this thesis designs, ideas and varieties are used interchangeably.

firm in the manufacturing sector produces a unique heterogeneous good. Both, the manufacturing sector and the traditional goods sector use immobile low skilled labour, while the R&D sector employs high skilled mobile labour.

The definition of low and high skilled workers leans on the terminology in the growth and migration literature (Romer, 1990; Dequiedt and Zenou, 2013; Vargas-Silva and Rienzo, 2014). High skilled workers are either highly educated tertiary workers or workers in top positions within a firm. The defining characteristic for high skilled workers is their ability to work in the manufacturing and the R&D sector. Low skilled have little schooling or work in lower tier positions within a firm.

2.3.1 R&D Sector

The R&D sector is orientated on the Romer model, and produces designs under perfect competition. It employs high skilled workers $H_{R,i}$ as an input factor. One of the main criticisms of the Romer model is that the growth rate depends on the absolute number of workers and does not allow for diffusion of technology or spillovers (Jones, 1995). To account for diffusion, ideas or designs from abroad can be used by the domestic R&D sector, following the "Flow of Ideas" concept in Rivera-Batiz and Romer (1991). As another input factor the existing stock of domestic designs A_i and foreign designs A_j^{ρ} is used. R&D firms access foreign designs at no cost. The foreign designs increase the returns to high skilled labour, so R&D firms have an incentive to use them. This can be interpreted as the R&D sector, for example universities, working internationally together, exchanging ideas and learning from each other. The parameter ρ measures the speed or magnitude of diffusion. If $\rho = 0$, then there is no diffusion and ideas do not spread across countries. This implies that designs produced in one country remain within this country and technology or knowledge do not diffuse. However, in the open economy case, the other country has access to a higher number of varieties of the manufactured good as well. If $0 < \rho < 1$, then part of the foreign designs can be adopted, while in the case of $\rho = 1$ foreign designs are completely available to the R&D firms in the home country. The unlikely case of $\rho > 1$ implies that foreign ideas are more productive in the home country. $\dot{A}_i$ is the number of new designs in country i and produced according to

$$\dot{A}_i = \delta H_{R,i} \bar{A}_i \tag{2.1}$$

with

$$\bar{A}_i = A_i + A_j^{\rho}, \; 0 \leq \rho \leq 1. \tag{2.2}$$

The production function has two implications. First, a higher number of employees in the R&D sector implies a higher output of designs. Secondly, the designs are accumulated and increase productivity in future periods. Therefore, there are two effects of an exogenous increase in the number of high skilled workers, a growth effect and a level effect.

The firms in the R&D sector sell the designs at a price $p_{A,i}$ in a perfectly competitive market to the manufacturing firms. R&D firms pay the high skilled workers their marginal product, $w_{R,i} = \delta \bar{A}_i p_{A,i}$. Thus the profit equation of a R&D firm is:

$$\pi_{A,i} = p_{A,i} \dot{A}_i - w_{R,i} H_{R,i}. \tag{2.3}$$

2.3.2 Manufacturing Sector

A firm in the manufacturing sector produces variety s with a constant marginal product of low skilled labour.[9] Before a firm can start producing, it has to buy exactly one design from the R&D sector at price $p_{A,i}$.[10] This implies that there are exactly A_i firms in the country. Moreover each manufacturing firm requires a

[9]For simplicity, the marginal product of labour is assumed to be constant. The main advantage of this simplifying assumption is that the price does not depend on the quantity of x_i and the level of technology. Therefore, computation of the price index P_i becomes more straightforward. The assumption is in line with Forslid and Ottaviano (2003).

[10]This assumption comes from the Romer Model. In the NEG models, a fixed input requirement is included but this is not seen as a requirement for firm entry. Buying a second design would not improve the output of a firm.

share $H_{x,i} = H_i \frac{\alpha}{A_i}$ of high skilled workers, where $H_i = A_i H_{x,i} + H_{R,i}$ is the total number of high skilled workers residing in country i. The high skilled workers in the manufacturing sector can be interpreted as the administrative staff, managers or workers in top occupations, required to run a firm. In a theoretical sense, the high skilled workers provide a margin between the manufacturing and the R&D sector. Each firm produces one unique intermediate good $x_i(s)$ with the following production function:

$$x_i(s) = \phi A_i L_{x,i}. \tag{2.4}$$

$x_i(s)$ is decomposed into two parts. One part is sold in country i and the rest is exported to country j with iceberg costs τ, similar to Krugman (1991). Thus the number of manufactured goods produced in the country is $x_i(s) = x_{ii}(s) + \tau x_{ji}(s)$.[11] τ measures the number of units of the manufactured good that have to be produced in order to satisfy the foreign demand of one unit, implying $\tau \geq 1$. In the case of zero transportation costs τ equals 1. As trade costs increase to infinity, trade becomes so expensive that any trade between the countries is precluded.

Denoting the wage for low skilled labour as $w_{x,i}$, the profit equation for a firm in country i is:

$$\pi_{x,i} = p_{ii}(s)x_{ii}(s) + p_{ji}(s)x_{ji}(s) - w_{x,i}L_{x,i} - w_{H,i}H_{x,i}. \tag{2.5}$$

The cost function has two parts: the variable costs of low skilled labour, $w_{x,i}L_{x,i}$, and the fixed costs of high skilled workers, $w_{H,i}H_{x,i}$.[12] Since each firm produces a differentiated good, the market is characterised by monopolistic competition. Low skilled workers are paid the marginal product of labour in the manufacturing sector; $w_i^x = \phi A_i$. Finally, to ensure that high skilled workers have no incentive to move

[11] The first subscript refers to the country of consumption, the second to the country where the product is produced. Thus x_{ji} is produced in country i and exported to or consumed in country j. Note that by assumption the prices of x_{ii} and x_{jj} are the same in both countries, otherwise it would be possible that high skilled migrants would migrate due to a higher real income that is only due to lower prices.

[12] The wages of the high skilled workers are fixed in the R&D sector, and manufacturing firms take this as given.

from one sector to another, wages in the manufacturing and the R&D sector have to be equal, $w_{H,i} = w_{R,i} = \delta \bar{A}_i p_{A,i}$.

Manufacturing firms can only use domestic designs A_i to found firms. However, in the open economy case, spillovers via trade are possible. Consumers have the possibility to buy the foreign manufactured good and thus enjoy a larger variety of goods. These spillovers are indirect in the sense that they do not influence the production side of the home economy; they only appear on the consumption side.

2.3.3 Traditional Goods Sector

The Traditional Goods sector uses low skilled labour $L_{T,i}$ as an input and produces a homogeneous traditional good. For simplicity, it is assumed that one input produces exactly one unit of output.[13] Firms in the sector produce the good under perfect competition according to the following production function:

$$T_i = L_{T,i}. \tag{2.6}$$

The firms sell their product at a price $p_{T,i}$ and pay wages $w_{T,i}$. This gives the following profit equation:

$$\pi_i^T = p_{T,i} T_i - w_{T,i} L_{T,i}. \tag{2.7}$$

Since the competitive firms make no profits, the paid wage equals the price for the traditional good, $w_{T,i} = p_{T,i}$. In equilibrium, low skilled labour is paid the same wage in both sectors to prevent workers moving from the traditional to the manufacturing sector or vice versa. This assumption quantifies the wage in the manufacturing sector to be: $w_{x,i} = w_{T,i} = p_{T,i}$. Following standard NEG models the traditional good can be traded at no cost (Baldwin and Martin, 2004). As the good is homogeneous and consumers do not have a preference for variety of the traditional good, the country with the lower wage for low skilled workers will produce the traditional good in the

[13]This assumption follows Krugman (1991) and Forslid and Ottaviano (2003).

open economy case. As the wage for low skilled workers depends positively on the number of designs (hence on technology), the traditional good will be produced by the less productive country.

The traditional good sector is necessary to make sure that in the corner solution of all high skilled worker migrating to the home country, the foreign country still exists and the remaining, immobile, low skilled workers are employed in the traditional goods sector.

2.3.4 Consumers

As standard in the NEG and NEGG literature, the representative consumer in country i maximizes a Cobb-Douglas utility function with both goods:

$$U_i = X_i^{\mu} T_i^{\gamma} \tag{2.8}$$

with the composite X_i

$$X_i = \left[\int_0^{A_i} x_{ii}(s)^{1-\frac{1}{\sigma}} ds + \int_0^{A_j} x_{ij}(s)^{1-\frac{1}{\sigma}} ds \right]^{\frac{1}{1-1/\sigma}} \tag{2.9}$$

$$= \left[\int_0^{A_i+A_j} x_i^C(s)^{1-\frac{1}{\sigma}} ds \right]^{\frac{1}{1-1/\sigma}} \tag{2.10}$$

$$\sigma > 1. \tag{2.11}$$

The choice of the varieties depends on a Constant Elasticity of Substitution (CES) function, with σ denoting the elasticity of substitution between the different varieties. A feature of CES utility functions is that consumers prefer the consumption of a differentiated bundle ("Love for Variety"). Each consumer faces the following budget constraint, which can be expressed by two equations:

$$Y_i = w_{T,i}L_{T,i} + w_{x,i}L_{x,i} + w_{R,i}H_{R,i} + w_{H,i}H_{x,i} \tag{2.12}$$
$$= p_{T,i}T_i + \int_0^{A_i} p_{ii}(s)x_{ii}(s)ds + \int_0^{A_j} p_{ij}(s)x_{ij}(s)ds \tag{2.13}$$
$$= p_{T,i} + P_iX_i. \tag{2.14}$$

Equation (2.12) is the income of the households from working and has to equal the spending of the households (equation (2.13) and (2.14)).

Wages in the R&D sector are denoted by R, while wages in the two other sectors are denoted by x for the manufacturing and T for the traditional goods sector. H_i is the supply of high skilled workers, including both domestic and migrant foreign workers. $L_i = A_iL_{x,i} + L_{T,i}$ is the total supply of low skilled labour which is the sum of the two remaining sectors.

Further, households consume a subsistence level of the traditional good, $T_i \geq T_i^{min}$, which can be interpreted as the necessary consumption of it.

As mentioned before, high skilled workers are mobile; they can move from country j to country i.[14] They migrate if their income in the country of destination is higher compared to their home country. There is no motive for migration if the income in both countries is equal. Hence the total amount of high skilled workers

[14]It is assumed that only the high skilled workers employed in the R&D sector migrate and those who work in the manufacturing sector ($H_i\frac{\alpha}{A_i}$) remain in the country to make sure that the manufacturing sector continues to work. If these high skilled workers were to leave as well, the country would not be able to produce the manufactured good and all low skilled labour would be shifted towards the traditional goods sector. The outcome would be as in the case of trade and complete specialization; see next section and Chapter 2.4.4.

in country i is:

$$H_i = m_i(w_{H,i}, w_{H,j}) = (1 - m_{ij})L_{H,i} + m_{ji}L_{H,j} \tag{2.15}$$

$$m_{ij} = \begin{cases} 0, & \text{if } \ w_{H,i} \geq w_{H,j} \\ (0,1], & \text{if } \ w_{H,i} < w_{H,j} \end{cases} \tag{2.16}$$

$$m_{ji} = \begin{cases} 0, & \text{if } \ w_{H,j} \geq w_{H,i} \\ (0,1], & \text{if } \ w_{H,j} < w_{H,i} \end{cases}, \tag{2.17}$$

where m_{ij} is the fraction of high skilled workers (in country i) who migrate from country i to country j, and $L_{H,i}$ is the number of high skilled workers born in country i. The same holds for country j:

$$H_j = m_j(w_{H,i}, w_{H,j}) = (1 - m_{ji})L_{H,j} + m_{ij}L_{H,i}. \tag{2.18}$$

There are two potential equilibrium outcomes. In the corner solution, all high skilled workers migrate to the country with higher wages; while in the interior solution only a fraction of the high skilled workers migrate. The interior solution implies that the wage differential between the two countries is relatively small. Migration will occur until either wages are equalised or all high skilled workers have migrated, which is again the corner solution. In the interior solution, the distribution of high skilled workers is more balanced and both countries grow.

2.3.5 The Growth Engine

In the model, the sole engine of growth is the R&D sector producing new designs. If a new design is invented (or produced), a new firm in the manufacturing sector can be founded. The importance of the designs to the manufacturing firms is similar to the Romer model. In the present context, designs can be seen as patents or technologies which are used to produce a good. A firm buys the patent and has the right to produce and exclusively sell the product based on the patent, acting as a monopolist in the specific differentiated good. The number of designs increases the number of firms, making the R&D sector crucial for the growth process of an economy.

The growth of designs, $\dot{A}_i$, depends on the number of high skilled workers and the existing inventions available. A higher number of high skilled workers increases the number of new designs. Also, a large stock of designs increases the number of new designs. The first implication is that countries which are on the technological frontier have a higher production of new designs. Second, even though the firm in the R&D sector sells the design exclusively to one firm in the manufacturing sector, the research firm is allowed to use it for the production of further designs. Then, the growth rate of designs is:

$$\frac{\dot{A}_i}{A_i} = \delta H_{R,i} \left(1 + \frac{A_j^{\rho}}{A_i} \right). \qquad (2.19)$$

In the case of $\rho = 0$ equation (2.19) collapses to $A_i/A_i = \delta H_{R,i}$, implying a permanent increase in the growth rate of designs is only possible with an increase in the number of high skilled workers. Therefore, the differentiation between the growth and level effects is important. If the number of high skilled workers increases from period $t-1$ to t, the growth effect will be an increase in the growth *rate* of designs at the end of period t, while the level effect will set in from period $t+1$ in the form of a higher *stock* of designs A_i. The implication is, that two countries with the same number of workers employed in the R&D sector may have the same growth rate but may have a different number (level) of designs (or technology) and have a different level of income. This implies that the initial income level is determined by the initial level of technology. Besides raising the productivity in the R&D sector, more designs contribute to the productivity of the manufacturing sector as well.

In the more realistic case of $0 < \rho < 1$ the growth rate depends additionally on the ratio of foreign to domestic designs. The ratio A_j^{ρ}/A_i represents the additional growth due to diffusion. If the ratio is very small, or converging towards zero, hence there are more domestic designs than foreign designs, then the growth rate of the country depends only on the domestic high skilled workers. However, if the ratio is larger than 1, namely if there are more foreign designs than domestic designs, then the home country benefits from the diffusion of technology. In the limit, if $(A_i, A_j) \stackrel{j}{\to} \infty$ such that $A_i/A_j \to 1$, the growth rate is twice the rate without diffusion. One limitation remains: even though the designs are available for the R&D sector, they need to be processed by high skilled workers. Thus, if there are no high skilled worker employed in the R&D sector then, despite spillovers, the growth rate of designs remains zero, the number of firms in the manufacturing sector remains the same and there is no growth effect.

The spillovers raise the wages of the high skilled workers $w_{H,i}$, even if the number of designs in the home country does not increase. Therefore, the wage gap between two countries increases at a lower rate in comparison to the case without spillovers. The speed of the increase is lower with a larger value of ρ.

The trade concept in this model is loosely related to the "Flows of Goods" idea in Rivera-Batiz and Romer (1991). However, the effect takes place on the consumption side of the economy. Consumers have a larger variety of goods to choose from. This does not have an effect on the growth rate in either model. However, if one country solely specializes in the traditional good (complete specialisation), the high skilled workers in the manufacturing sector are not needed anymore and are available to migrate as well. Then, if trade and migration are possible, the receiving country experiences an even larger permanent growth rate. All disadvantages are borne by the country that specialises in the traditional good, as total output decrease because of lesser productivity in the traditional good sector.
In addition, in the open economy case, there is a level effect due to the relative prices of the traditional and manufactured good and accompanied reallocation of low skilled labour. The manufactured good becomes cheaper and demand rises, which shifts production from the relatively less productive traditional good sector to the more productive manufacturing sector. The increase in the production of the manufactured good increases the income level. This effect diminishes with an

increase in the number of designs and the decrease of the weight of the traditional goods sector; see next section.

NEGG models predict catastrophic agglomeration under the assumption of mobile capital and labour, meaning all firms move in the region with greater capital. In line with this prediction and the Romer model, this model shows the growth equivalent divergence, in the case of labour mobility. The country that is better equipped with human capital or designs has a higher innovation rate which translates to a higher growth rate. Instead of manufacturing firms being grouped in one region, in this model high skilled workers are grouped in the same country, fostering growth.

2.3.6 Closing the Model

Next, the behaviour of the consumers and high skilled workers is explained, the model is closed.[15] It is assumed that the size of the high and low skilled labour force ($L_{H,i}$ and L_i) is predetermined and fixed. Within each period, the number of firms (A_i) is exogenous as well.

First, the price index and quantities demanded by consumers are derived. A two-stage optimization of equation (2.8) with respect to equation (2.14), based on Dixit and Stiglitz (1977), leads to the following values for $x_{ii}(s)$, $x_{jj}(s)$, $x_{ij}(s)$ and $x_{ji}(s)$:

$$x_{ii}(s) = \frac{\sigma - 1}{\sigma}\left(\frac{Y_i}{R_i} - \frac{p_{T,i}T_i}{R_i}\right) \tag{2.20}$$

$$x_{jj}(s) = \frac{\sigma - 1}{\sigma}\left(\frac{Y_j}{R_j} - \frac{p_{T,j}T_j}{R_j}\right) \tag{2.21}$$

$$x_{ij}(s) = \tau^{-\sigma}\frac{\sigma - 1}{\sigma}\left(\frac{Y_i}{R_i} - \frac{p_{T,i}T_i}{R_i}\right) \tag{2.22}$$

$$x_{ji}(s) = \tau^{-\sigma}\frac{\sigma - 1}{\sigma}\left(\frac{Y_j}{R_j} - \frac{p_{T,j}T_j}{R_j}\right) \tag{2.23}$$

[15] Appendix A.1 contains detailed derivations.

with

$$R_i = A_i + \tau^{1-\sigma} A_j. \tag{2.24}$$

The solutions for the manufactured goods imply that if domestic income does not change, but foreign varieties and so R_i and R_j do, then demand for the domestic manufactured good (say, x_{jj} and x_{ji}) decreases by the growth rate of foreign varieties. As consumers prefer variety, they are substituting the domestic produced good with imports.

Demand for the good produced by the traditional goods sector equals:

$$T_i = \max\left(\frac{\phi L_i}{\frac{\mu}{\gamma}\frac{p_{T,i}}{P_i} + \phi},\ T_i^{min}\right) \tag{2.25}$$

$$= \max\left(\frac{\gamma}{\mu + \gamma}\frac{Y_i}{p_{T,i}},\ T_i^{min}\right). \tag{2.26}$$

The demand for the traditional good depends on the relative price of the traditional good to the manufactured good ($p_{T,i}/P_i$). The price for the traditional good is fixed at $p_{T,i} = \phi A_i$, the price index of the manufactured good will be derived next. The manufacturing firms maximize their profits using the inverse demand functions for the manufactured goods. Thus, they set the following prices for each good:

$$p_{ii}(s) = p_{ii} = p_{jj} = \frac{\sigma}{\sigma - 1}, \tag{2.27}$$

$$p_{ji}(s) = p_{ji} = p_{ij} = \frac{\sigma}{\sigma - 1}\tau = p_{ii}\tau. \tag{2.28}$$

In either case, the second equality holds because of symmetry in the production functions of the two countries. The price is independent of the quantity of the manufactured good or the technological level A_i, depending only on the elasticity of substitution between the different varieties. Moreover, the price for each variety is the same.

The price index for country i is:

$$P_i = \left[\int_0^{A_i} p_{ii}(s)^{1-\sigma} ds + \int_0^{A_j} p_{ij}(s)^{1-\sigma} ds\right]^{\frac{1}{1-\sigma}} \tag{2.29}$$

$$= \frac{\sigma}{\sigma - 1} R_i^{\frac{1}{1-\sigma}}. \tag{2.30}$$

The price index falls with an increase in the number of domestic and foreign firms. In terms of workers, a larger R&D sector with a large output implies a falling price index.

The derivation of the prices allows for an interpretation of the demand for the traditional good (equation 2.26). Assuming that the low skilled workforce remains constant, the demand for the traditional good falls if the relative price of the traditional good to the manufactured good ($p_{T,i}/P_i$) increases. As the price index for the manufactured good decreases with the number of designs while the price for the traditional good increases with the number of designs, the relative price increases if the number of designs increases. Therefore, more low skilled labour is allocated to the manufactured goods sector and the output of the corresponding sector increases.

Similar to the Romer Model, manufacturing firms make no profits because they bid for the design produced by the R&D sector. The price for a design is the discounted present value of the firm's profits at the interest rate r_i:

$$p_{A,i} = \sum_{t=0}^{\infty} \frac{1}{1 + r_i^t} \pi_{x,i,t}. \tag{2.31}$$

Under the assumption that profits are constant over time, the price collapses to:[16]

$$p_{A,i} = \frac{1}{r_i} \pi_{x,i} \tag{2.32}$$

$$\pi_{x,i} = p_{ii} x_{ii} + p_{ji} x_{ji} - w_{x,i} L_{x,i} - w_{H,i} H_{x,i} = r_i p_{A,i} \tag{2.33}$$

[16] The price for the design only changes with the number of high skilled worker, but not with the number of designs. Then in the steady state where no migration occurs, prices will not change. Secondly, any changes in migration and therefore the price for the design are unexpected for the manufacturing firms.

with

$$w_{H,i} = \delta \bar{A}_i p_{A,i} \tag{2.34}$$

$$\Rightarrow w_{H,i} = \frac{\delta \bar{A}_i x_i}{\left(r_i + \delta\alpha \frac{\bar{A}_i}{A_i} H_i\right)(\sigma - 1)}. \tag{2.35}$$

As the price for each variety is the same, all firms in the manufacturing sector produce the same quantity of the firm specific variety, implying $x_i(s) = x_i = x_{ii} + \tau x_{ji}$. Thus, the price for the designs depends negatively on the number of high skilled workers residing in the country and the interest rate.[17] The wage depends negatively on the number of high skilled workers and positively on the number of designs. This rules out the possibility that the order of wages for high skilled workers in the two countries reverses. In other words, if the domestic country has a higher wage compared to the foreign country and receives more high skilled workers in t, the wage will be higher for all following periods. Diffusion of ideas only slows the increase in the wage gap, but it is not able to close it, as long as one country has a larger number of designs ($A_i > A_j$).

Under the assumption that all labour markets cleared and that there is no unemployment:

$$H_i = A_i H_{x,i} + H_{R,i} \tag{2.36}$$

$$L_i = A_i L_{x,i} + L_{T,i}, \tag{2.37}$$

[17] The same effect of an increase of the number of high skilled workers and the productivity, A_i, on the wage of high skilled workers can be found for example in Grossmann and Stadelmann (2013).

real output can be expressed as:

$$\begin{aligned} Y_i &= A_i x_i + T_i && (2.38)\\ &= A_i \phi A_i L_{x,i} + L_i - A_i L_{x,i} && (2.39)\\ &= L_i + A_i L_{x,i} \left(\phi A_i - 1\right) && (2.40)\\ &= L_i + \frac{x_i}{\phi} \left(\phi A_i - 1\right). && (2.41) \end{aligned}$$

The time derivative is needed for the growth path:

$$\dot{Y}_i = \dot{L}_i + \frac{x_i}{\phi}\phi\dot{A}_i + \frac{\dot{x}_i}{\phi}(\phi A_i - 1) \qquad (2.42)$$

As $\dot{x}_i = \dot{L}_i = 0$:

$$\begin{aligned} \dot{Y}_i &= x_i \dot{A}_i && (2.43)\\ \Rightarrow \frac{\dot{Y}_i}{Y_i} &= \frac{\dot{A}_i x_i}{A_i x_i + L_{T,i}} && (2.44)\\ &= \frac{\dot{A}_i \phi A_i L_{x,i}}{A_i \phi A_i L_{x,i} + L_{T,i}} && (2.45)\\ &= \frac{x_i}{x_i + \frac{L_{T,i}}{A_i}} \frac{\dot{A}_i}{A_i}. && (2.46) \end{aligned}$$

Equation (2.45) shows that a reduction in the labour allocated to the traditional goods sector increases the growth rate, as the low skilled labour is more productive in the manufacturing sector. This holds as long as $\phi A_i^2 > 1$. Moreover (2.46) implies, that for a sufficient small A_i a reallocation of low skilled labour can have an impact on the growth rate. However, in the long run $L_{T,i}/A_i \to 0$ as A_i increases and $L_{T,i}$ decreases, therefore this effect wears off. The long run growth rate becomes:

$$\begin{aligned} \frac{\dot{Y}_i}{Y_i} &\approx \frac{\dot{A}_i x_i}{A_i x_i} && (2.47)\\ &= \frac{\dot{A}_i}{A_i} = \delta H_{R,i} \left(1 + \frac{A_j^{\rho}}{A_i}\right), && (2.48) \end{aligned}$$

where the last expression comes from (2.19). The implication of (2.48) is that for a sufficiently large A_i or small $L_{T,i}$ the growth rate of the economy depends solely on the growth rate of designs, which is determined by the number of high skilled workers.

2.3.7 High Skilled Worker Migration

High skilled workers migrate if the wage differential between the countries is positive.[18] Since ordering of wages is preserved and high skilled workers are homogeneous, migration flows will only be in one direction. If migration is costless, migration will happen until wages are equalised, therefore equalising the factor price of high skilled labour:

$$w_{H,i} = w_{H,j} \tag{2.49}$$

$$\delta \bar{A}_i \frac{x_i}{(\sigma-1)(r_i + \delta\alpha \frac{\bar{A}_i}{A_i} H_i)} = \delta \bar{A}_j \frac{x_j}{(\sigma-1)(r_j + \delta\alpha \frac{\bar{A}_j}{A_j} H_j)} \tag{2.50}$$

$$\bar{A}_i \frac{x_i}{r_i + \delta\alpha \frac{\bar{A}_i}{A_i}((1-m_{ij}L_{H,i} + m_{ji}L_{H,j})} = \bar{A}_j \frac{x_j}{r_j + \delta\alpha \frac{\bar{A}_j}{A_j}((1-m_{ji})L_{H,j} + m_{ij}L_{H,i})}. \tag{2.51}$$

Since labour only moves from country j to i (implying $m_{ji} > 0$ and $m_{ij} = 0$), labour movement becomes:

$$m_{ji} = \frac{1}{L_{H,j}(L_{x,i} + L_{x,j})} \left[\frac{1}{\delta\alpha} \left(L_{x,i} r_j \frac{A_i}{\bar{A}_j} - L_{x,j} r_i \frac{A_j}{\bar{A}_i} \right) + L_{H,j} L_{x,i} \frac{A_i}{A_j} - L_{H,i} L_{x,j} \frac{A_j}{A_i} \right] \tag{2.52}$$

If there is no diffusion ($\rho = 0$), equation (2.52) becomes:

$$m_{ji} = \frac{1}{L_{H,j}(L_{x,i} + L_{x,j})} \left[\frac{A_i}{A_j} L_{x,i} \left(\frac{r_j}{\delta\alpha} + L_{H,j} \right) - \frac{A_j}{A_i} L_{x,j} \left(\frac{r_i}{\delta\alpha} + L_{H,i} \right) \right]. \tag{2.53}$$

There is no labour movement if the difference between the terms in squared

[18] Since the interior solution moves towards the corner solution if the wage differential is large, only the interior solution with a maximum of 1 is discussed, which is the corner solution. Moreover, high skilled workers are informed about the wages in both countries.

brackets equals zero. Assuming that the traditional sector requires a constant share of low skilled labour, the amount of low skilled labour in the manufacturing sectors, $L_{x,i}$ and $L_{x,j}$, remains constant. As the model does not include any population growth, the total number of high skilled workers across the two countries is constant as well. Therefore, labour mobility depends solely on the stock of worldwide varieties or the number of firms. Migration increases with an increase in the difference in the number of firms between the two countries. This can easily be seen in equation (2.53), for the case without diffusion. If country i has more designs than country j, the positive part in the squared brackets increases, while the negative part decreases. If the ratio is reversed, $A_j >> A_i$, then the negative part dominates and m_{ji} can even turn negative. Thus, in terms of firms, the larger country attracts more high skilled workers. Migration flows are positive as $A_i > A_j$ implies $w_i^A > w_j^A$. An alternative interpretation is the following: wages for high skilled workers and productivity in the R&D sector increase with the number of designs. At the same time, more designs increase the output per worker in the manufacturing sector. The increase in high skilled wages drives migration, thus the country that has higher productivity in the R&D as well as in the manufacturing sector, receives more high skilled workers. This rationale of migration is similar to Klein and Ventura (2009). Hence productivity differences are a driving force of migration.

Migration can be then summarised as

$$m_{ji} = \begin{cases} \min \left(\frac{1}{L_{H,j}(L_{x,i}+L_{x,j})} \left[\frac{1}{\delta\alpha} \left(L_{x,i} r_j \frac{A_i}{A_j} - L_{x,j} r_i \frac{A_j}{A_i} \right) \right.\right. & \\ \qquad \left.\left. + L_{H,j} L_{x,i} \frac{A_i}{A_j} - L_{H,i} L_{x,j} \frac{A_j}{A_i} \right], 1 \right) & , \quad \text{if } \; w_{H,i} > w_{H,j} \\ 0 & , \quad \text{if } \; w_{H,i} \leq w_{H,j} \end{cases} \tag{2.54}$$

$$m_{ij} = \begin{cases} \min \left(\frac{1}{L_{H,i}(L_{x,j}+L_{x,i})} \left[\frac{1}{\delta\alpha} \left(L_{x,j} r_i \frac{A_j}{A_i} - L_{x,i} r_j \frac{A_i}{A_j} \right) \right.\right. & \\ \qquad \left.\left. + L_{H,i} L_{x,j} \frac{A_j}{A_i} - L_{H,j} L_{x,i} \frac{A_i}{A_j} \right], 1 \right) & , \quad \text{if } \; w_{H,j} > w_{H,i} \\ 0 & , \quad \text{if } \; w_{H,j} \leq w_{H,i} \end{cases} \tag{2.55}$$

The number of high skilled worker H_i living in country i can then be described as:

$$H_i = (1 - m_{ij})L_{H,i} + m_{ji}L_{H,j}, \tag{2.56}$$

where $L_{H,i}$ and $L_{H,j}$ are as defined before as the high skilled workers born in countries i and j. Using the fact that $H_i = A_i H_{x,i} + H_{R,i} = \alpha H_i + H_{R,i} \Rightarrow H_{R,i} = (1-\alpha)H_i$ and using equation (2.48):

$$\frac{\dot{Y}_i}{Y_i} = \frac{\dot{A}_i}{A_i} = \delta H_{R,i}\left(1 + \frac{A_j^\rho}{A_i}\right) \tag{2.57}$$

$$= \delta\,(1-\alpha)\,[(1-m_{ij})\,L_{H,i} + m_{ji}L_{H,j}]\left(1 + \frac{A_j^\rho}{A_i}\right). \tag{2.58}$$

Equation (2.58) directly relates migration of high skilled workers and the diffusion of ideas to the growth rate of a country. If migration into the country is high, the growth rate increases. Diffusion of ideas increases the growth rate as well, however the increase is larger for a country that has a relatively small level of ideas, such that A_j^ρ/A_i is large. Finally, equation (2.58) shows that trade can only have an indirect effect on the growth rate via the share of migrants (m_{ji} and m_{ij}).

2.3.8 Incorporating Migration Costs

The assumption of zero migration costs is unrealistic. Migrants incur costs to establish social networks, learn and work using a foreign language and culture. These costs can be summarised as utility costs. On the other hand, the act of migration itself bears costs. In the following, the main emphasis is put on the latter. Costs are introduced without any further assumptions about their origin. There will be migration if the wage in the host country minus costs is larger than in the domestic country:[19]

$$w_{H,i} - k > w_{H,j}. \tag{2.59}$$

[19]Detailed derivations are presented in Appendix A.2.

Ruling out migration in the opposite direction ($m_{ij} = 0$), migration flows from j to i can be described as:[20]

$$m_{ji} = \frac{l_i}{2\eta s_i s_j L_{H,j}^2} - \sqrt{\left(\frac{l_i}{2\eta s_i s_j L_{H,j}^2}\right)^2 - \frac{o_i}{2\eta s_i s_j L_{H,j}^2}}, \tag{2.60}$$

where

$$\eta = \frac{\sigma - 1}{\delta} k \tag{2.61}$$

$$s_i = \delta\alpha \frac{\bar{A}_i}{A_i} \tag{2.62}$$

$$u_i = \eta r_j s_i + \bar{A}_j x_j s_i \tag{2.63}$$

$$v_i = \eta r_i s_i - \bar{A}_i x_i s_j \tag{2.64}$$

$$o_i = u_i L_{H,i} + v_i L_{H,j} - \bar{A}_i x_i r_j + \bar{A}_j x_j r_i + \eta r_j r_i + \eta s_j s_i L_{H,i} L_{H,j} \tag{2.65}$$

$$l_i = L_{H,i} \left(u_i - v_i + \eta s_j s_i \left(L_{H,j} - L_{H,i}\right)\right). \tag{2.66}$$

Compared to the case without migration costs, interpretation of equation (2.60) is somewhat complicated.[21] The main driver of migration is still the difference in the stock of firms, since the other variables remain constant. Migration costs will have the effect of postponing migration until the wage differential is large enough to cover the costs. The influence of costs of migration is further discussed in the context of simulations in Section 2.4.

2.3.9 Discussion

This completes the description of the model. The model has the following key characteristics. First, a higher number of high skilled workers leads to an increase in the production of designs that in turn enhances growth. In addition, the richer country receives more high skilled workers and therefore the difference between the two countries increases. In the steady state, when a fraction or the entire

[20] For m_{ij} the indices i and j are interchanged.

[21] Apart from equation (2.60), there is another potential solution for m_{ji}. In this solution, the square root is positive. This solution is inadmissible as it would predict migration flows from country i to j even in the case of $w_{H,j} > w_{H,i}$.

stock of high skilled workers moves to one country, the model explains divergence. Diffusion of technology has an effect as long as the R&D sector remains active in both countries. In the case of an open economy, specialization towards the more productive good leads to a level effect and under certain circumstances to a growth effect as well.

Countries are linked to each other by three factors: migration and, in a weaker form, trade and diffusion of technology. The country with more manufacturing firms attracts high skilled workers from the other country and benefits from a higher growth rate. The country with the smaller number of manufacturing firms remains stagnant at a lower level. The second linkage is trade. The country with a lower variety of manufactured goods benefits from growth in the richer country by importing a higher number of varieties, while at the same time exporting a fixed quantity of its manufactured goods. However, if the number of domestic varieties does not increase, consumers will substitute the domestic produced manufactured good by imports. The country which produces more of the manufactured good and less of the traditional good can increase its output by enlarging the production of the manufactured good. In the case of complete specialization and together with migration a growth effect for the more productive country is possible, as all high skilled workers move and boost the output in the R&D sector. Diffusion of technology only takes place in the R&D sector. Therefore, diffusion has an effect on the growth rate as long as high skilled workers are employed in the R&D sector. In this case, the wage gap between the two countries decreases, depending on the speed of diffusion.

Like any other theoretical growth model, the above model represents an abstraction of reality. To obtain realistic interpretations it is useful to consider conceptual distinctions between cross-country or inter-region migration and growth dependencies. World history offers a huge range of examples of labour movement between countries or regions. By varying the structural parameters, the model can be applied to both the above cases.

Divergence emerges as an essential outcome of the model, a major difference in comparison to the work of Braun (1993). Additionally, the receiving country benefits from migration. Moreover the model treats migration indirectly as a determinant of economic growth rather than Kennan (2013) who focuses on individual gains from migration. The outcome of divergence is due to the growth engine and the production function of the R&D sector. As an alternative, it would be possible to model the R&D sector with decreasing returns to scale, as for example in Fernald and Jones (2014). Decreasing returns to scale would lead to convergence. However the derivation of the migration streams would become more difficult and therefore, to keep the model simple as possible, constant returns are used. Economic divergence can be seen as the complement to the concentrated agglomeration of the NEGG models. One region might end up as the manufacturing core, while the other region remains in an unproductive farm sector.

As outlined in the introduction of this chapter, in the 20th century, there is evidence of club-convergence (Baumol, 1986; Islam, 2003), where groups of countries converge to a common equilibrium, but unconditional convergence across all countries is absent.[22] Thus, the divergence outcome of the model is also consistent with contemporary evidence on convergence clubs. It may be noted, however, that there is no mechanism for catch-up or convergence within each such club. Therefore, this model should be seen more as an attempt to consider interactions between countries or regions within a growth model, rather than offering a full explanation for economic growth and convergence or divergence across countries and regions.

Diffusion of ideas only takes place in the R&D sector. This is different to Feenstra (1996) as only the R&D sector uses foreign designs. Similar to Feenstra (1996), some of the foreign ideas are available to the home country. The diffusion of ideas in this

[22] Empirical analysis in Baumol (1986) suggests convergence for 16 OECD countries, but not for a broader sample of countries. (Islam, 2003, p. 317) writes: *"[P]rodded by Romer, Baumol also considers the relationship in an extended sample of 72 countries. In this larger sample, however, he does not find evidence of convergence."* and continues in footnote 11 with *"... The numerical results of this regression were not presented, but Baumol reported that it yielded 'slightly positive slope,' indicating a process of rather divergence."*

model is between countries and not within a country.[23]

Besides relying on the work of Romer (1990), the model emphasises to the importance of human capital for economic growth, a point stressed by Lucas (2009a,b). In addition to diffusion, the spread of knowledge in this model is through migration, and thus by movement of high skilled worker. A second vehicle for the spread of technology is trade. The chapter can be understood as emphasizing the importance of labour movements in the Lucas fashion and when switching off diffusion of knowledge.

One limitation of the model is that it concentrates on high skilled workers. Many migrants were low skilled during the great divergence and after. In the context of this model, there is one limitation and one arguments for restricting the flow of unskilled workers. The limitation is that allowing low skilled workers to migrate as well, would lead the poor country to lose all its workers and to a collapse of output. Low skilled workers are paid the marginal product in the manufacturing sector ϕA_i. Therefore wages in the manufacturing sector of the richer country will be higher, giving the unskilled workers an incentive to migrate. To keep the model simple, migration of low skilled workers is not taken into account. In this particular model an increase of low skilled workers has no effect on the growth rate. In the light of the migration literature the model has a few limitations. First of all, the effect of migration on economic growth is in the foreground. Effects on the country of origin via remittances are ignored. Return migration is in principle possible, if the wage differential between the two countries would revert. However as the wage for high skilled workers is driven by the number of domestic designs, return migration becomes unlikely. The model ignores mortality of workers and thus any effects on following generations are not modelled. As the model has no population growth, depleted skills or workplaces in the country of origin are not replaced by a new generation of workers. Modelling this would call for an overlapping generation model, which would complicate the model further. Another potential critique is that the

[23] Feenstra (1996) presents some evidence such as Coe and Helpman (1995) that diffusion of technology across countries is weak.

model focuses on wage differences rather than on utility as the motive for migration. However, as the Cobb-Douglas utility function is strictly increasing in both goods and both goods are normal goods, both motives are coincident. Additionally wage differences across countries for a worker with the same education and experience are large. Clemens, Montenegro and Pritchett (2009) compare the income of a US worker with an identical worker in less developed countries. For example, wages in the US are 2.5 times higher than in Mexico. Finally, in the model the decision to migrate is purely economic. Political or geographical reasons such as conflict or natural disasters often motivate migration. In the context of the model, such motives may be viewed as reducing the migration costs, k.

The model borrows from the NEGG model in Baldwin and Martin (2004). Both models build on three sectors. In addition, both share the assumption of no depreciation of physical or human capital. Both models predict divergence between the two regions or countries. Besides these common characteristics, there are some important differences. Whereas in Baldwin and Martin (2004) the goods from all three sectors can be traded, the model in this chapter only allows trade in the traditional and the manufactured good. However, ideas can diffuse. The major distinction of the two models is related to this difference: the concept of capital. Baldwin and Martin (2004) see immobile capital as human capital and mobile capital as physical capital. This is clearly different to the model in this chapter, which includes mobility of human capital.

The model is similar to the empirical application in Ortega and Peri (2014). Ortega and Peri show that migration increases total factor productivity, which then increases income. Moreover, the authors find a positive diversity effect. A more diverse country of origin of the immigrant population has an additional positive effect on income. In contrast to migration, they do not find a robust effect of trade openness on income as soon as they control for effects such as geography. The model in this chapter can be seen as a simplification and extension of Grossman and Helpman (2016), leaving the effects of inequality within a country aside, but extending the model by migration of high skilled workers. In both models, the size of the R&D sector plays a crucial role for income.

2.4 Simulation

In this section, a simulation is carried out to understand the impact of migration, diffusion and trade on income, in the context of the economic model developed in 2.3. For this purpose, dynamics into the accumulation of designs needs to be introduced. The number of designs over subsequent periods evolves as:

$$A_{i,t+1} = \dot{A}_{i,t} + A_{i,t} \tag{2.67}$$

$$= (1 + \delta H_{R,i})A_{i,t} + \delta H_{R,i} A^{\rho}_{j,t}. \tag{2.68}$$

It is assumed that the two countries, henceforth called *developed* (i) and *undeveloped* (j), have the same number of firms (or designs) and the same number of low skilled workers, but the developed country has more high skilled workers. The following initial values for stocks of designs and workers are considered:

$$A_i = 10 \qquad A_j = 10$$

$$L_i = 8 \qquad L_j = 8$$

$$L_{H,i} = 4 \qquad L_{H,j} = 2.$$

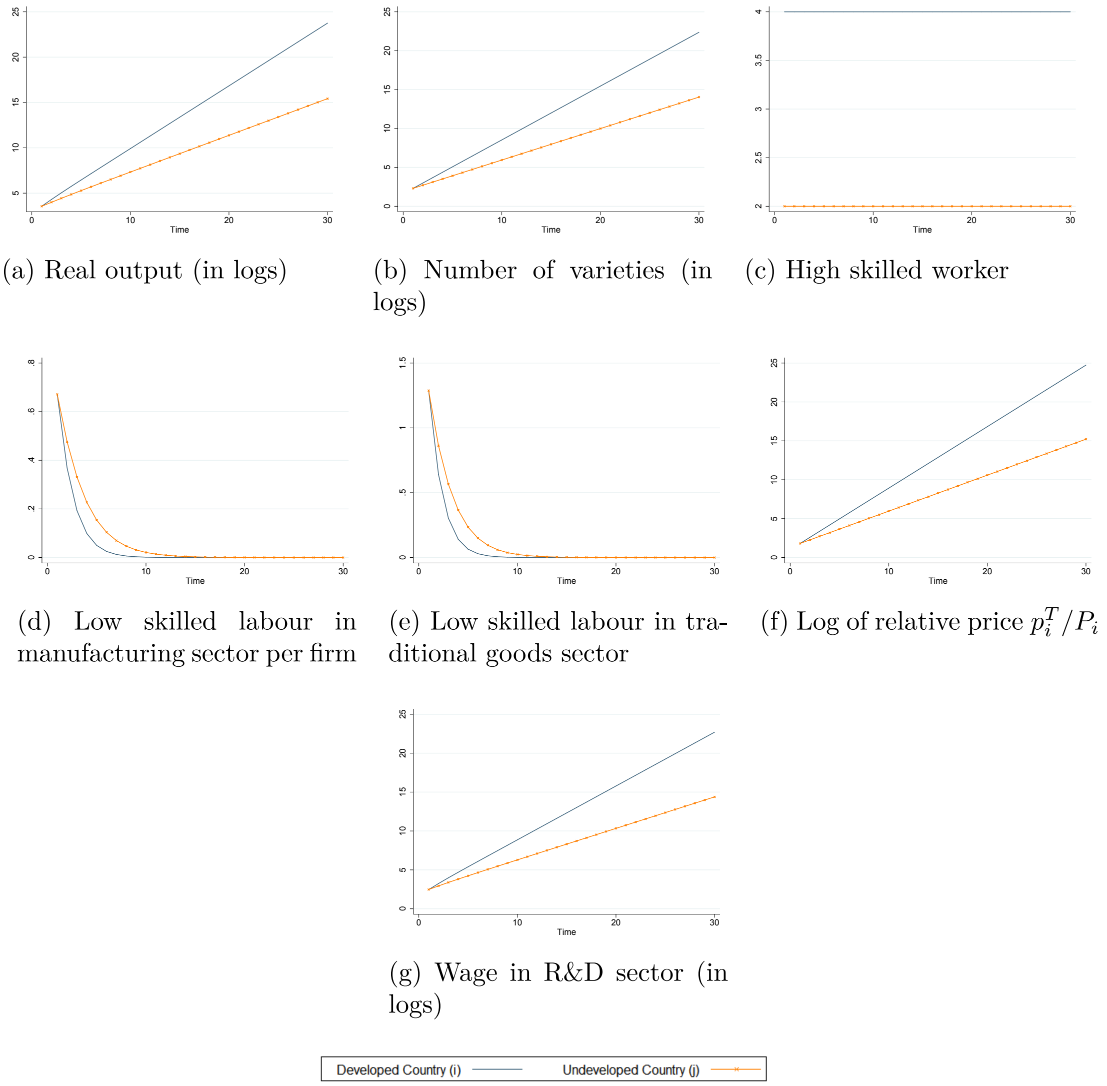
Keeping with existing notation, the above parameters imply that the developed country will have higher income due to a larger number of high skilled workers. Therefore, high skilled workers will migrate from the undeveloped to the developed country whenever such migration is allowed.

The manufacturing sector requires two units of low skilled labour for each unit produced. The R&D sector requires 4 units of high skilled workers for each newly invented design. The minimal share of high skilled workers in R&D is set to 0.0025. Jones (2016) finds that employment in the research sector for OECD countries increased from 0.15% in the 1980s to about 0.35% in 2010. The interest rates are equal in both countries at 10%. If not specified otherwise, the demand for the traditional good is unrestricted ($T_i^{min} = T_j^{min} = 0$). The marginal rate of substitution between the manufactured good and the traditional good is set according to Krugman (1991) or estimates from Broda and Weinstein (2006).[24] The share of the manufactured goods sector is set to 0.3.[25] Together with the remaining parameters, the parametrization of the simulation is as follows:

$$\begin{aligned} \delta &= 0.25 & \sigma &= 8 \\ \phi &= 0.5 & \alpha &= 0.1 \\ \mu &= 0.3 & \gamma &= 0.7 \\ \rho &= 0.0025 & r_i = r_j &= 0.1. \end{aligned}$$

For trade two cases are considered: (a) Very high transportation costs so that there is no trade, implies trade costs are set such that $\tau^{1-\sigma} = 0$; (b) otherwise, for each traded unit 1.5 units have to be produced ($\tau = 1.5$). When migration is allowed, it is introduced only from period 2 onwards. In addition, there are no constraints for migration, implying that from one period to the next, some or all high skilled workers are allowed to move between countries. Diffusion takes place immediately and set to the above value.

First, the model is simulated without migration, diffusion of ideas or trade (Figure 2.1). This case is used as a baseline and compared to simulations with migration or trade. These simulations (Figures 2.2 - 2.7) are displayed in percentage deviations from the baseline case.



(a) Real output (in logs)

(b) Number of varieties (in logs)

(c) High skilled worker

(d) Low skilled labour in manufacturing sector per firm

(e) Low skilled labour in traditional goods sector

(f) Log of relative price p_i^T/P_i

(g) Wage in R&D sector (in logs)

Figure 2.1: Baseline Model (no migration, no trade)
Parameters: $A_i = 10, L_i = 8, L_{H,i} = 4, A_j = 10, L_j = 8, L_{H,j} = 2, \delta = 0.25, \sigma = 8, \phi = 0.5, \alpha = 0.1, \mu = 0.3, \gamma = 0.7, T_i^{min} = 0, T_j^{min} = 0$. The developed country is denoted by i and the less developed by j. No migration, $m_{ji} = m_{ij} = 0$, and trade costs such that $\tau^{1-\sigma} = 0$.

2.4.1 Baseline

Figure 2.1 shows simulation results of the baseline model without migration, diffusion and trade ($\tau \to \infty, \tau^{1-\sigma} = 0$). Real output (Panel 2.1a), the number of firms (Panel 2.1b), relative prices (Panel 2.1f) and the wage in the manufacturing sector (Panel 2.1g) are transformed in logarithms.

[24] More detailed estimates for the rate of substitution up to a 12 digit HTS and 5 digit ISIC level can be found for example in Feenstra (1994).

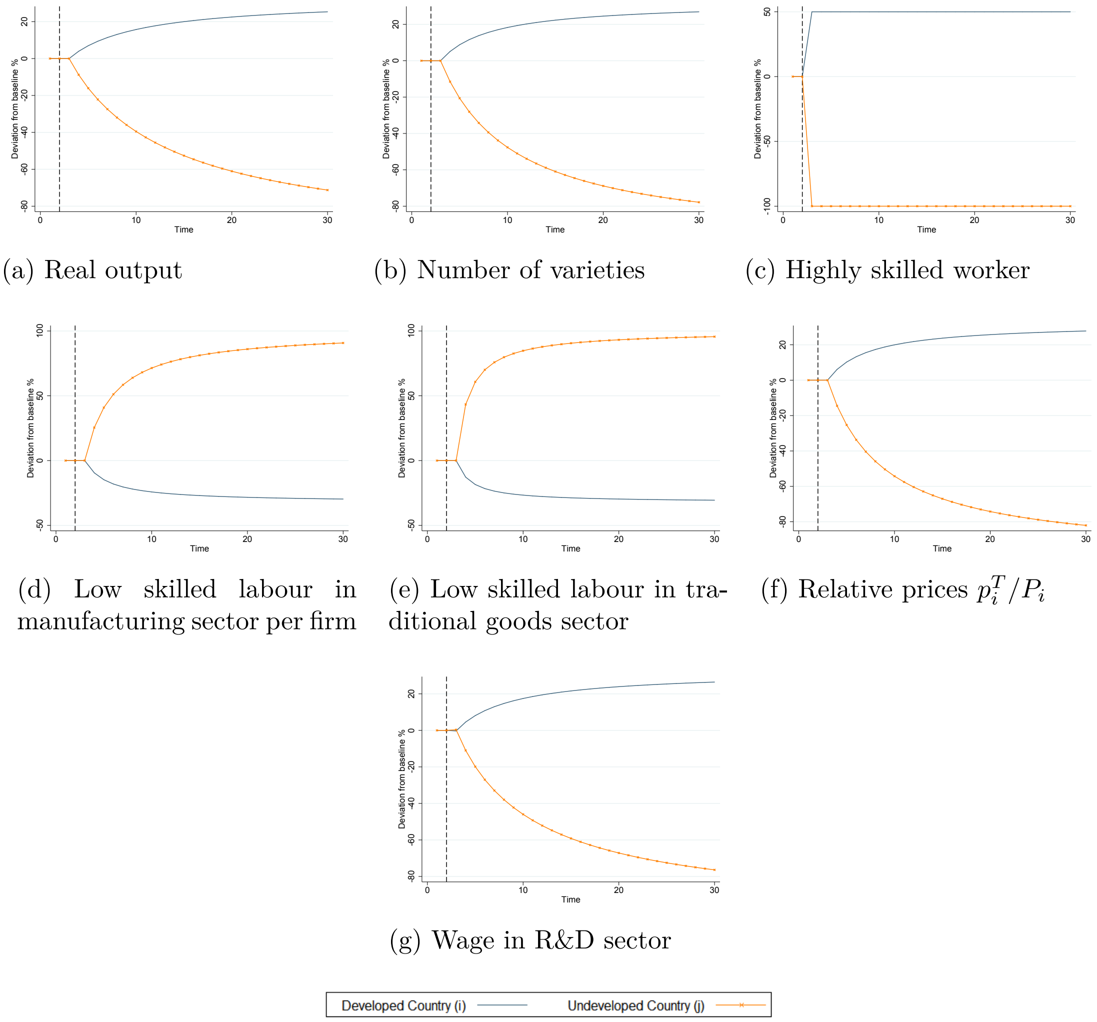
[25] Similar shares for the consumption of manufactured goods can be found for example in Herrendorf, Rogerson and Valentinyu (2014).

The model behaves as expected from a Romer model. Output increases at a constant rate, the number of firms and hence varieties increase at the same rate. As the growth rate of the developed country is larger than the rate of the undeveloped country, the difference in levels increases, leading to divergence between the two countries. The wages of high skilled workers increase at a constant rate as well. The increase is smaller than the increase of the number of varieties as the wage is scaled down by δ. Since there is no migration, the number of high skilled workers remains the same. The number of low skilled workers employed in a single firm in the manufacturing sector decreases over time as the number of firms increases. 0.7 low skilled workers are employed in each of the 10 firms of the manufacturing sector in both countries in the first period. After the first period, manufacturing firms in the developed country with the more productive manufacturing sector employ less low skilled workers per firm. The amount of low skilled labour allocated to the traditional goods sector decreases, as the countries develop (Panel 2.1e).

2.4.2 Migration

Only Migration Figure 2.2 shows a simulation with migration but no trade. The graphs show percentage deviations from the baseline model (Graph 2.1).

As soon as migration is introduced in period 2 (indicated by a vertical dashed line), high skilled workers start moving to the developed country as shown in Panel 2.2c. After two periods all high skilled workers employed in the R&D sector have moved, leaving the R&D sector in the less developed country empty. This movement boosts the amount of designs produced in the developed country while the amount of designs in the other country remain at the same level. The number of designs, and so the number of firms, do not change anymore and therefore demand remains the same. The influx of high skilled migrants leads to an increase in the growth rate of the developed country. Due to the exodus of the high skilled workers, the growth rate of the undeveloped country breaks down, intensifying divergence between the two countries in comparison to the baseline scenario and leaving the country permanently



(a) Real output

(b) Number of varieties

(c) Highly skilled worker

(d) Low skilled labour in manufacturing sector per firm

(e) Low skilled labour in traditional goods sector

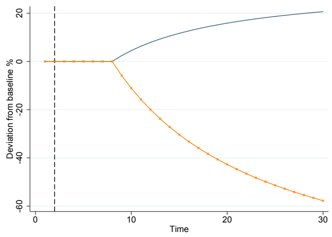
(f) Relative prices p_i^T/P_i

(g) Wage in R&D sector

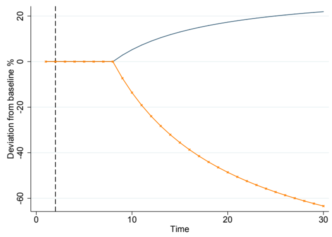
Figure 2.2: Only Migration
Migration is allowed from period 2 onwards. For parametrisation see Section 2.4 or Figure 2.1. Graphs show deviations in % from baseline.

worse off.

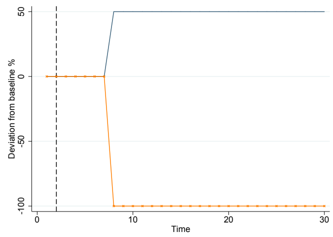
As before, the amount of low skilled labour allocated to the entire traditional sector and to each individual firm in the manufacturing sector in the developed countries decreases, even more than in the baseline scenario. For the undeveloped country, the movement is the same; however it is not as pronounced as in the baseline scenario. The wage for high skilled workers (Panel 2.2g) in the more productive country drops in comparison to the baseline case. The drop is due to the influx of new high skilled workers. However, it is not large enough to reverse the migration streams. The increase in designs then pushes the wage above the wage in the baseline scenario. The wage for high skilled workers and the relative price in the less developed country do not change anymore, because the number of designs remains constant, therefore the deviation is negative.



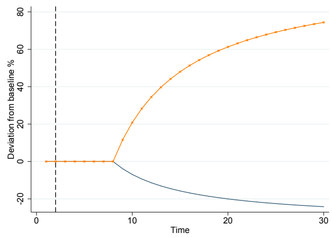
(a) Real output)



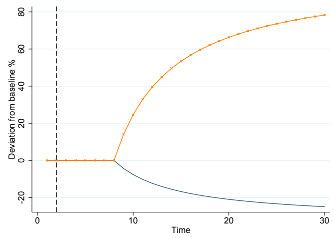
(b) Number of varieties



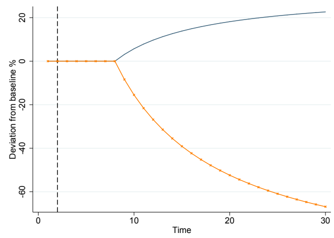
(c) High Skilled Worker



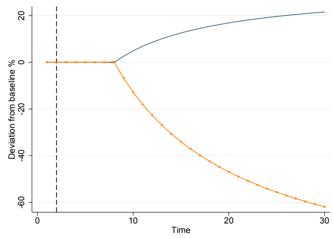
(d) Low skilled labour in manufacturing sector per firm



(e) Low skilled labour in traditional goods sector



(f) Relative prices p_i^T/P_i



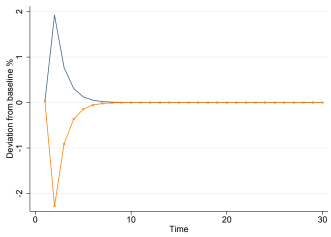
(g) Wage in R&D sector

Figure 2.3: Migration with costs
Migration is allowed from period 2 onwards, costs of migration are 300 and no trade. For parametrisation see Section 2.4 or Figure 2.1. Graphs show deviations in % from baseline.

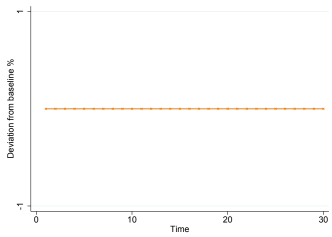
Migration Costs Figure 2.3 shows a simulation with migration costs of 300, which is three times the wage in the R&D sector in period 5. Panel 2.3c shows that migration starts with a delay of 8 periods. Again all high skilled workers from the R&D sector migrate. This is because the difference in the number of firms does not increase linearly over time. The appearance of the graphs is similar to the scenario without migration costs (Graph 2.2) and the interpretation remains the same. The most notable difference is that deviations from the baseline scenario are less pronounced as the effects of migration kick in later and countries follow their original path (without migration) for a longer time period.

The simulation clearly shows that the country which receives more high skilled worker benefits from migration by increasing the growth rate of designs and therefore experiencing a higher output growth rate. Migration costs only postpone the decision to migrate. The outcome remains the same, migration leads to divergence.

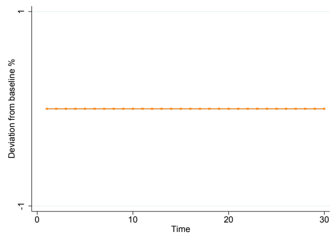
2.4.3 Trade



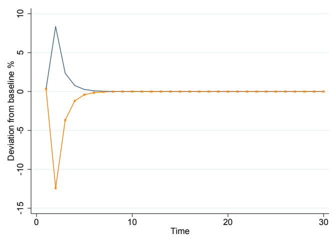
(a) Real output



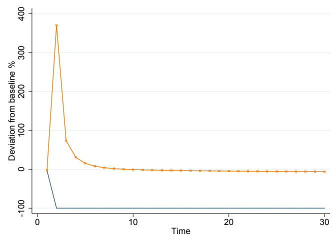
(b) Number of varieties



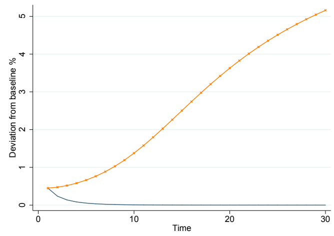
(c) High Skilled Worker



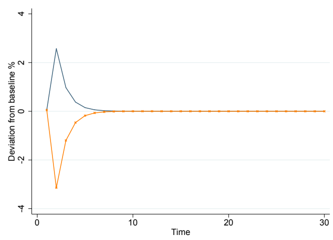
(d) Low skilled labour in manufacturing sector per firm



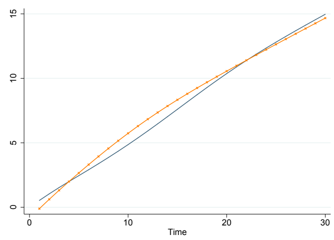
(e) Low skilled labour in traditional goods sector



(f) Relative prices p_i^T/P_i



(g) Wage in R&D sector



(h) Imports of manufactured good (absolute value)

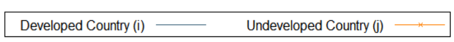


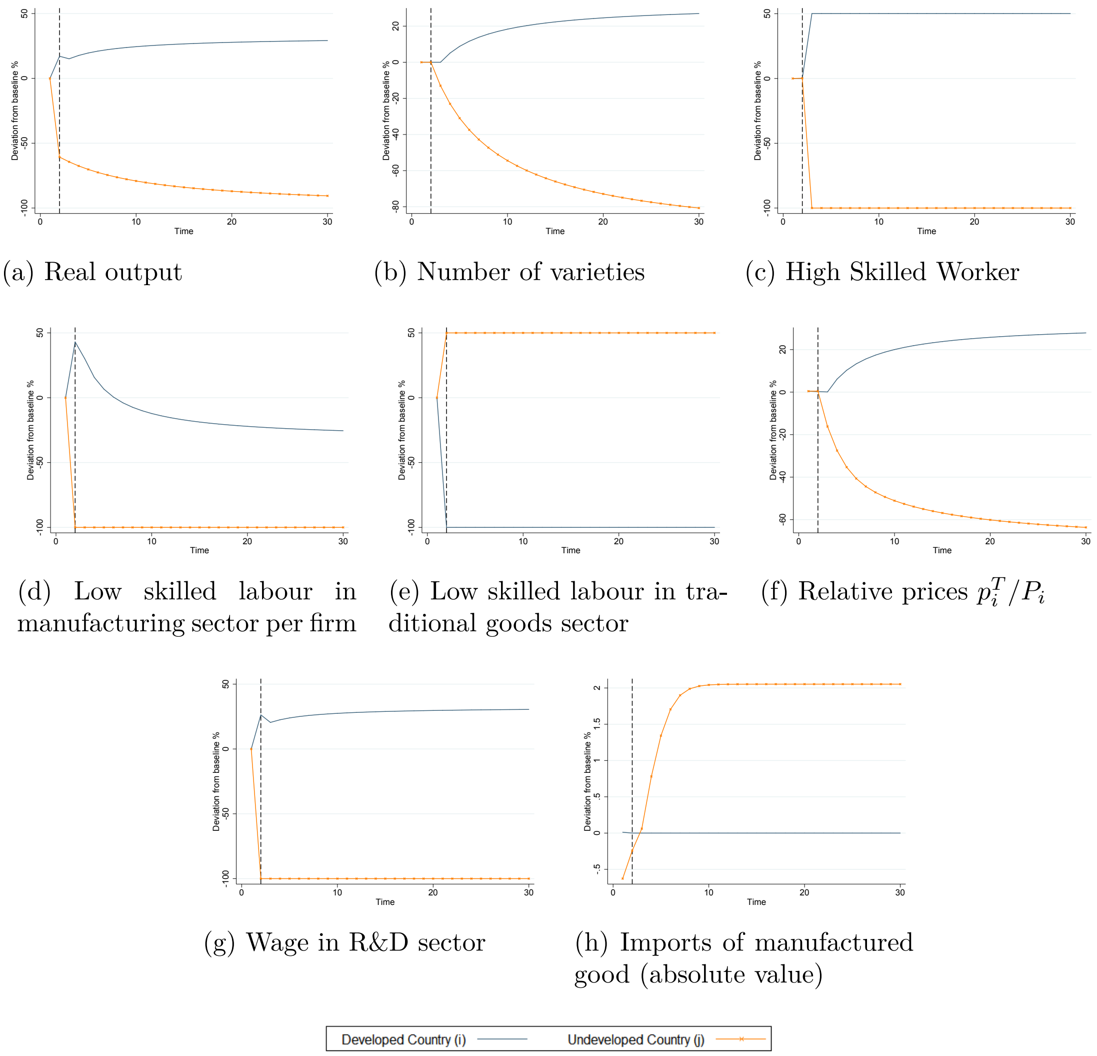
Figure 2.4: Only Trade
Trade costs are set to $\tau = 1.5$. For parametrisation see Section 2.4 or Figure 2.1. Graphs show deviations in % from baseline.

Only Trade To find out how trade affects the model, Figure 2.4 shows how the simulation with trade costs of $\tau = 1.5$ differ to the baseline scenario. High skilled workers are immobile and are not allowed to move between countries, making trade the only interaction between countries.

The number of varieties, or firms, is the same in both cases, since trade does not lead to any changes in the allocation of the high skilled workers between sectors and countries. However there is a temporary effect on the growth rate of real output. Due to the fall in the price level for the manufactured good, P_i, in the developed country (Panel 2.4f) demand rises. At the same time, the entire production of the traditional good is taken over by the relatively less productive undeveloped country as shown Panel 2.4e and 2.4d.[26] The shift results in a decrease of real output in the less productive country as visible by the downward kink in Panel 2.4a. However, with the increase of designs the allocation of low skilled workers moves away from the traditional good sector to the manufacturing sector (Panel 2.4e). Therefore in the limit real output and the growth rate of the country are not affected. As more low skilled workers are available to work in the manufacturing sector in the developed country, real output increases by the magnitude of the additional low skilled worker. This affects the level of output permanently, thus temporarily the growth rate in the short run (from period 2 to 3) but not the long run growth rate (see equation 2.46 and 2.48). Panel 2.4h shows the absolute value of imports. With an increase in designs, more goods become available to consume and trade. Moreover, both countries are growing and both countries employ the entire workforce. However, differences in the growth rate and levels persist. Hence, trade only has a positive effect on output if the gains from the reallocation of low skilled labour towards the relatively more productive sectors outweigh the trade costs. As outlined in Section 2.3.5, it is possible for the growth rate to have an effect on trade under certain circumstances - as explained next.

[26]The effect whereby an inefficient sector in one country is replaced is similar to Alvarez et al. (2013).

2.4.4 Trade and Migration



(a) Real output

(b) Number of varieties

(c) High Skilled Worker

(d) Low skilled labour in manufacturing sector per firm

(e) Low skilled labour in traditional goods sector

(f) Relative prices p_i^T/P_i

(g) Wage in R&D sector

(h) Imports of manufactured good (absolute value)

Figure 2.5: Trade with $T_i^{min} = T_i^{min} = 4$ and Migration
Trade costs are set to $\tau = 1.5$. Migration is allowed from period 2 onwards with no costs. For parametrisation see Section 2.4 or Figure 2.1. Graphs show deviations in % from baseline.

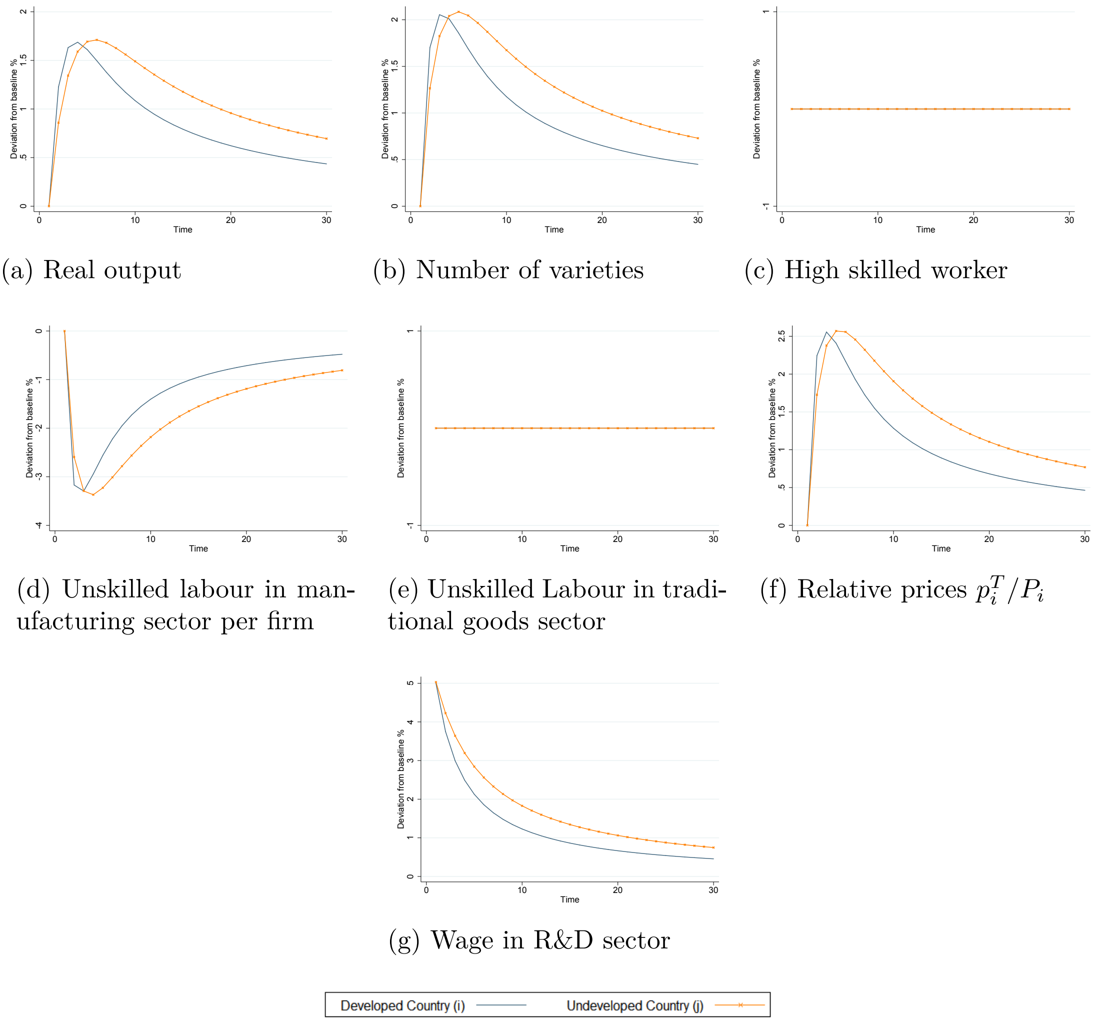
Figure 2.5 shows simulations with trade costs of $\tau = 1.5$, but in comparison to Figure 2.4 migration is allowed and the minimum consumption of the traditional good is set to $T_i^{min} = T_j^{min} = 4$. This implies that at least 8 units of low skilled labour have to be devoted to the production of the traditional good. Panel 2.5d and 2.5g show that all low skilled labour in the less productive undeveloped country is moved to the traditional good sector. The country becomes the sole producer of the traditional good. This implies that the output in absolute terms decreases and remains at 8 units. As there is no manufacturing sector in the country left, high skilled workers employed in this sector can move to the other country. This boosts the deviation of the output in comparison with the baseline model from 25% in the case of migration only to 29%. Another notable difference is that the undeveloped country does not export any manufactured goods anymore. Imports into the less productive, undeveloped country increase as consumers benefit from a larger variability of the manufactured good.

An interesting outcome can be generated if the subsistence level of $T_i^{min} = T_j^{min} = 4$ applies, trade is allowed but no migration occurs. In this case, the undeveloped country will become the sole producer of the traditional good and not produce the manufactured good. Therefore, there is no demand for further designs, making all high skilled workers unemployed.

The simulation reveals that, in comparison with migration, the effect of trade is of a smaller magnitude. In the long run the less productive country does not lose any output as long as it produces the manufactured good. The more productive developed country experiences a small level effect due to the increase in low skilled workers allocated to the manufacturing sector. If one country completely specializes and solely produces the less productive good, then this country experiences a severe drop in output, while due to migration and specialisation the other country can gain, even in the long run, from trade.

2.4.5 Diffusion of Designs

Figure 2.6 shows simulations in which designs diffuse across countries with a rate of $\rho = 0.0025$. First, migration is not allowed. In comparison with the baseline scenario, output growth and the number of designs is increasing, but the increase is not as pronounced as in the case of migration. As the number of designs is larger, the wage in the R&D sector increases as well.



(a) Real output (b) Number of varieties (c) High skilled worker

(d) Unskilled labour in manufacturing sector per firm (e) Unskilled Labour in traditional goods sector (f) Relative prices p_i^T/P_i

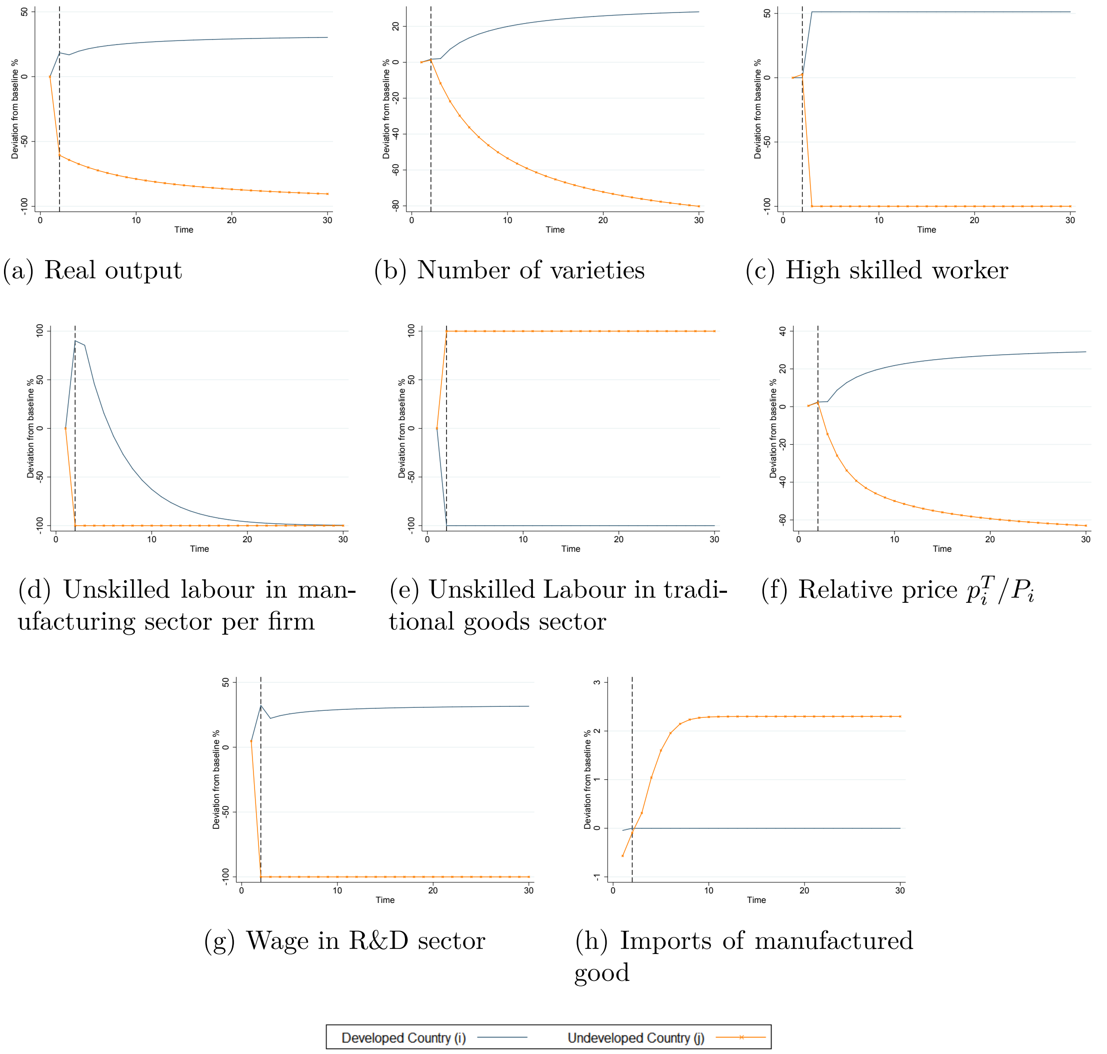
(g) Wage in R&D sector

Figure 2.6: Only Diffusion ($\rho = 0.0025$)
For parametrisation see Section 2.4 or Figure 2.1. Graphs show deviations in % from baseline.

The price level for the manufactured good increases, implying that the manufactured good becomes cheaper relative to the traditional good. Together with the fact that the manufactured sector is more productive in comparison to the baseline case, this results in a larger share of unskilled workers employed in the manufacturing sector, as seen in Panel 2.6d and 2.6e.

2.4.6 Migration, Diffusion and Trade

Finally, Figure 2.7 shows all three channels activated at the same time. Full migration is allowed from period 2, trade costs are set to $\tau = 1.5$ and diffusion of designs to



(a) Real output (b) Number of varieties (c) High skilled worker

(d) Unskilled labour in manufacturing sector per firm (e) Unskilled Labour in traditional goods sector (f) Relative price p_i^T/P_i

(g) Wage in R&D sector (h) Imports of manufactured good

Figure 2.7: Migration, Trade ($\tau = 1.5$) and Diffusion ($\rho = 0.25$). Migration is possible from period 2. For parametrisation see Section 2.4 or Figure 2.1. Graphs show deviations in % from baseline.

$\rho = 0.0025$. The minimal consumption of the traditional good is $T_i^{min} = T_j^{min} = 4$. The overall picture remains similar to before. Trade leads to a specialization, and all high skilled workers migrate to the more productive country, while the less productive country solely produces the traditional good. This enables the productive country to grow further and all unskilled workers are allocated to the manufacturing sector (Panel 2.7d and 2.7e). Therefore, the output of the country which produces the traditional good drops back to 8 units, while the output of the country producing the manufacturing good increases.

The simulations show that the country with the higher variety in the manufacturing sector attracts high skilled workers. The time it takes for all high skilled workers to migrate varies with costs; with higher costs, it takes longer for the migration to be completed. The host country experiences an economic boom, while the other country remains stagnant, implying divergence. Trade plays a minor role and, again, the more productive country is the main beneficiary.

2.5 Conclusion

Satisfactory theoretically founded spatial interactions are neglected in standard growth models in the literature. While some of the empirical spatial econometric growth literature emphasizes spatial dependence, this is typically modelled in terms of knowledge diffusion. However, theory provides no motivation as to why such diffusion exists, or even why geography should constitute the main driver for knowledge spillovers. This chapter provides an economic model to address this gap in the literature and motivate interactions between countries. Migration, diffusion of ideas and trade provide mechanisms for cross-country interactions in the presented model. The share of migrants and the trade are microfounded and thus determined within the model. The second focus of this chapter is on the effect of the interactions on economic growth.

To model interactions between countries, the restriction of immobile high skilled labour is removed. Therefore features of a New Economic Geography (NEG) model in the style of Krugman (1991) and an endogenous growth model in the style of Romer (1990) are combined to develop a new two-country economic model. The model retains the microfoundation, trade of a manufactured good and migration of labour from the New Economic Geography models. The microfoundation allows to model the interactions between the countries explicitly. Migration and trade come out of the model, while the degree of diffusion of knowledge is exogenous set under the assumption that all R&D firms have an incentive adopt foreign knowledge. On the other hand, the engine of growth and the feature of unbounded growth, which explain the divergence between two countries in this context, comes from the endogenous Romer-style growth model. Economic growth and migration are directly related and therefore the model can be used to motivate migration as a spatial weight in a cross-country growth regression.

The model has several limitations. First of all only high skilled migrants are allowed to migrate. This modelling restriction is necessary to ensure that not all workers leave a country. Secondly the model is restricted to a single generation. Following generations which would replace depleted skills, allowing the country to grow again, are not considered. In addition, return migration and the effects on remittances are not included. It is among the first models which explicitly seeks to find the role of migration together with trade and technology in generating cross-country dependence. All three are complex phenomena and naturally, there are limitations in the model. These are retained in the interest of keeping linkages between countries as simple as possible.

Without migration and trade, the model behaves like an endogenous Romer-style growth model. If high skilled workers migrate to the higher wage offering country and contribute to the R&D sector, output growth is enhanced in this country. This leads to divergence. While benefits from trade negate part of this divergence, convergence or catch-up does not emerge as an outcome. Due to reallocation of low skilled labour, trade affects growth, but only in the short run. All this leads to divergence. Simulations confirm the predictions derived from theory. High skilled workers migrate from the country that offers a lower wage and, under autarky, this country stagnates. The country that attracts high skilled workers experiences a growth boom. When trade is allowed, countries start specializing. However, this negatively affects the country which specializes in the less productive traditional good, while the other country benefits. Both effects apply only in the short run. A long run persistence effect on the growth rate is only possible if one country completely specializes in the traditional good, allowing all high skilled workers to migrate. Then the more productive country gains from trade. Diffusion has only a small effect on the level. The number of varieties, and thus output, increases in both countries.

As only a fraction of ideas diffuse, a catch-up is not possible. Further simulations show that migration costs lead to a delay in migration, but because of rising wages, this does not ultimately deter migration. Based on the migration of high skilled labour and the conclusion of divergence, the proposed model displays some of the facts of the great divergence of the 19th century, namely the finding of migration accompanied by divergence. While there is no mechanism for catch-up or convergence within the model, it is also consistent with evidence on the club-convergence in recent times.

The proposed model may be viewed as a first step towards introducing endogenous spatial interactions in appropriately theoretically founded spatial growth models. Therefore, it presents a rich agenda for future research. First, the model can be validated against cross-country economic data for the 19th and 20th centuries. Second, the R&D sector can be changed such that the outcome of the model is convergence. Finally, potential migration of low skilled labour can also be incorporated.

Appendix Chapter 2

A.1 Derivation of price index and quantities demanded by consumers

The derivation of the price index and the quantities demanded by consumers follows a two-stage optimization, after Dixit and Stiglitz (1977). Maximizing (2.8) with respect to the budget constraint (2.14) and a minimum amount of the traditional good T_i^{min}, which has to be consumed, gives the following Lagrangian :

$$\max_{X_i,T_i} X_i^{\mu} T_i^{\gamma}$$

$$\text{s.t.: } Y_i > p_{T,i}T_i + X_i P_i \text{ and } T_i \geq T_i^{min}$$

$$\mathcal{L} = X_i^{\mu} T_i^{\gamma} + \lambda_1 (Y_i - p_{T,i}T_i - X_i P_i) + \lambda_2 (T_i - T_i^{min})$$

$$\frac{\partial \mathcal{L}}{\partial X_i} = \mu X_i^{\mu-1} T_i^{\gamma} - \lambda_1 P_i = 0$$

$$\rightarrow \lambda_1 = \frac{\mu X_i^{\mu} T_i^{\gamma}}{X_i P_i} \tag{A.1}$$

$$\frac{\partial \mathcal{L}}{\partial T_i} = \gamma X_i^{\mu} T_i^{\gamma-1} - \lambda_1 p_{T,i} + \lambda_2 \geq 0 \tag{A.2}$$

$$\frac{\partial \mathcal{L}}{\partial \lambda_1} = Y_i - p_{T,i}T_i - X_i P_i = 0 \tag{A.3}$$

$$\lambda_2 \left(T_i - T_i^{min} \right) \geq 0. \tag{A.4}$$

The budget constraint is always binding, thus $\lambda_1 \neq 0$. If the constraint on the minimum value is not binding $T_i > T_i^{min}$, thus $\lambda_2 = 0$, then putting (A.1) into (A.2) and using (A.3) leads to:

$$\begin{aligned} T_i &= \frac{\phi L_i}{\frac{\mu}{\gamma}\frac{p_{T,i}}{P_i} + \phi} \\ &= \frac{\gamma}{\mu+\gamma}\frac{Y_i}{p_{T,i}} \qquad \text{(A.5)} \\ X_i &= \frac{\mu}{\gamma+\mu}\frac{Y_i}{P_i}. \qquad \text{(A.6)} \end{aligned}$$

If the constraint on the minimal consumption of the traditional good is binding, $T_i = T_i^{min}$, thus $\lambda_2 > 0$, then:

$$\begin{aligned} \Rightarrow X_i &= \frac{Y_i - p_{T,i}T_i}{P_i} \\ \frac{\gamma X_i^{\mu} T_i^{\gamma}}{T_i} &- \lambda_1 p_{T,i} + \lambda_2 > 0 \\ \lambda_2 &= \frac{\gamma X_i^{\mu} T_i^{\gamma}}{T_i} - \frac{\mu X_i^{\mu} T_i^{\gamma}}{X_i} p_{T,i} > 0 \\ \frac{Y_i}{p_{T,i}}\frac{\gamma}{\gamma+\mu} &> T_i = T_i^{min}. \qquad \text{(A.7)} \end{aligned}$$

The last condition holds as long as $\mu < 1$ and $\gamma \neq 0$. Then using (A.1), (A.3) and $T_i = T_i^{min}$:

$$Y_i = p_{T,i}T_i^{min} + X_i P_i \rightarrow X_i = \frac{Y_i - p_{T,i}T_i^{min}}{P_i}. \qquad \text{(A.8)}$$

If $\gamma = 0$, then

$$\begin{aligned} -\lambda_1 p_{T,i} + \lambda_2 > 0 \\ \lambda_2 > \lambda_1 p_{T,i} = \frac{\mu X_i^{\mu-1}}{P_i} p_{T,i} > 0. \end{aligned}$$

Hence (A.4) only holds if:

$$T_i = T_i^{min}.$$

the 2nd stage of utility maximization is:

$$\begin{aligned}
\mathcal{L} &= T_i^{\gamma}\left[\int_0^{A_i} x_{ii}(s)^{1-\frac{1}{\sigma}}ds + \int_0^{A_j} x_{ij}(s)^{1-\frac{1}{\sigma}}ds\right]^{\frac{\sigma\mu}{\sigma-1}} \\
&+ \lambda_1\left(Y_i - p_{T,i}T_i - \int_0^{A_i} p_{ii}(s)x_{ii}(s)ds - \int_0^{A_j} p_{ij}(s)x_{ij}(s)ds\right) \\
&+ \lambda_2(T_i - T_i^{min}) \\
\frac{\partial\mathcal{L}}{\partial\lambda_1} &= Y_i - p_{T,i}T_i - \int_0^{A_i} p_{ii}(s)x_{ii}(s)ds - \int_0^{A_j} p_{ij}(s)x_{ij}(s)ds = 0 \qquad \text{(A.9)} \\
\frac{\partial\mathcal{L}}{\partial x_{ii}(s)} &= \mu X_i^{\mu}T_i^{\gamma}\frac{\sigma}{\sigma-1}\frac{\sigma-1}{\sigma}x_{ii}(s)^{-\frac{1}{\sigma}}X_i^{-\frac{1}{\sigma}} - \lambda_1 p_{ii}(s) = 0 \\
\rightarrow \lambda_1 &= \mu X_i^{\mu}T_i^{\gamma}X_i^{-\frac{1}{\sigma}}x_{ii}(s)^{-\frac{1}{\sigma}}p_{ii}(s)^{-1}, \qquad \text{(A.10)}
\end{aligned}$$

with varieties s and k, multiplying with $p_{ii}(s)$ and solving for $x_{ii}(s)p_{ii}(s)$:

$$\begin{aligned}
\mu X_i^{\mu}T_i^{\gamma}X_i^{-\frac{1}{\sigma}}x_{ii}(s)^{-\frac{1}{\sigma}}p_{ii}(s)^{-1} &= \mu X_i^{\mu}T_i^{\gamma}X_i^{-\frac{1}{\sigma}}x_{ii}(k)^{-\frac{1}{\sigma}}p_{ii}(k)^{-1} \\
x_{ii}(s)p_{ii}(s) &= p_{ii}(k)^{\sigma}p_{ii}(s)^{1-\sigma}x_{ii}(k). \qquad \text{(A.11)}
\end{aligned}$$

$x_{ii}(s)$ and $x_{ij}(s)$ are symmetric and thus $\frac{\partial\mathcal{L}}{\partial x_{ij}(s)}$ leads to

$$x_{ij}(s)p_{ij}(s) = p_{ij}(k)^{\sigma}p_{ij}(s)^{1-\sigma}x_{ij}(k). \qquad \text{(A.12)}$$

In addition (A.10) implies that:

$$\begin{aligned}
\mu X_i^{\mu}T_i^{\gamma}X_i^{-\frac{1}{\sigma}}x_{ii}(k)^{-\frac{1}{\sigma}}p_{ii}(k)^{-1} &= \mu X_i^{\mu}T_i^{\gamma}X_i^{-\frac{1}{\sigma}}x_{ij}(s)^{-\frac{1}{\sigma}}p_{ij}(s)^{-1} \\
x_{ii}(k)p_{ii}(k)^{\sigma} &= x_{ij}(k)p_{ij}(k)^{\sigma}. \qquad \text{(A.13)}
\end{aligned}$$

Inserting (A.11) and (A.12) into (A.9), rearranging the integral, using the result for T_i from the first stage and defining the price index as

$$P_i = \left[\int_0^{A_i} p_{ii}(s)^{1-\sigma}ds + \int_0^{A_j} p_{ij}(s)^{1-\sigma}ds\right]^{\frac{1}{1-\sigma}}$$

leads to:

$$\begin{aligned} Y_i &= p_{T,i}T_i + \int_0^{A_i} p_{ii}(s)^{1-\sigma} p_{ii}(k)^{\sigma} x_{ii}(k) ds + \int_0^{A_j} p_{ij}(s)^{1-\sigma} p_{ij}(k)^{\sigma} x_{ij}(k) ds \\ &= p_{T,i}T_i + p_{ii}(k)^{\sigma} x_{ii}(k) \int_0^{A_i} p_{ii}(s)^{1-\sigma} ds \\ &+ p_{ij}(k)^{\sigma} x_{ij}(k) \int_0^{A_j} p_{ij}(s)^{1-\sigma} ds. \end{aligned}$$

With the results from (A.13):

$$\begin{aligned} &= p_{T,i}T_i + p_{ij}(r)^{\sigma} x_{ij}(r) \left(\int_0^{A_i} p_{ii}(s)^{1-\sigma} ds + \int_0^{A_j} p_{ij}(s)^{1-\sigma} ds \right) \\ &= p_{T,i}T_i + p_{ij}(r)^{\sigma} x_{ij}(r) P_i^{1-\sigma} \\ \to x_{ij}(s) &= \frac{1}{P_i^{1-\sigma} p_{ij}(s)^{\sigma}} \left(Y_i - p_{T,i}T_i \right), \end{aligned} \tag{A.14}$$

by symmetry it follows that:

$$\to x_{ji}(s) = \frac{1}{P_j^{1-\sigma} p_{ji}(s)^{\sigma}} \left(Y_j - p_{T,j}T_j \right) \tag{A.15}$$

and using (A.13) gives the demand for the domestic manufactured good:

$$\to x_{ii}(s) = \frac{1}{P_i^{1-\sigma} p_{ii}(s)^{\sigma}} \left(Y_i - p_{T,i}T_i \right). \tag{A.16}$$

Firms

Transporting good x_{ji} from country i to j inhibits Iceberg-Costs as in Krugman (1991). For one unit to arrive in the destination country, $\tau \geq 1$ units have to be shipped. The firm maximizes profits:

$$\pi_i(s) = p_{ii}(s)x_{ii}(s) + p_{ji}(s)x_{ji}(s) - w_{x,i}L_{x,i} - w_{H,i}H_{x,i}$$

with the inverse demand function from (A.14) respectively (A.16):

$$
\begin{aligned}
p_{ii}(s) &= \left[\frac{1}{P_i^{1-\sigma}x_{ii}(s)}\left(Y_i - p_{T,i}T_i\right)\right]^{\frac{1}{\sigma}}, \quad p_{ji}(s) = \left[\frac{1}{P_j^{1-\sigma}x_{ji}(s)}\left(Y_j - p_{T,j}T_j\right)\right]^{\frac{1}{\sigma}} \\
\rightarrow \pi_i(s) &= x_{ii}(s)\left[\frac{1}{P_i^{1-\sigma}x_{ii}(s)}\left(Y_i - p_{T,i}T_i\right)\right]^{\frac{1}{\sigma}} + x_{ji}(s)\left[\frac{1}{P_j^{1-\sigma}x_{ji}(s)}\left(Y_j - p_{T,j}T_j\right)\right]^{\frac{1}{\sigma}} \\
&\quad - w_{x,i}L_{x,i} - w_{H,i}H_{x,i} - p_i^A \\
\frac{\partial \pi_i}{\partial x_{ii}(s)} &= \left[\frac{1}{P_i^{1-\sigma}x_{ii}(s)}\left(Y_i - p_{T,i}T_i\right)\right]^{\frac{1}{\sigma}} \\
&\quad - \frac{1}{\sigma}x_{ii}(s)\left[\frac{1}{P_i^{1-\sigma}x_{ii}(s)}\left(Y_i - p_{T,i}T_i\right)\right]^{\frac{1}{\sigma}} x_{ii}(s)^{-1} - w_{x,i}\frac{\partial L_{x,i}}{\partial x_{ii}} \\
&= p_{ii}(s) - \frac{1}{\sigma}p_{ii}(s) - w_{x,i}\frac{\partial L_{x,i}}{\partial x_{ii}} = 0
\end{aligned}
$$

with $\frac{\partial L_{x,i}}{\partial x_{ii}} = \frac{1}{\phi A_i}$ and $w_{x,i} = \phi A_i$:

$$
\begin{aligned}
\rightarrow p_{ii}(s) &= \frac{\sigma}{\sigma - 1} \\
\rightarrow p_{ji}(s) &= \tau\frac{\sigma}{\sigma - 1}.
\end{aligned}
$$

Thus:

$$
p_{ii}(s) = p_{ii}, \quad p_{ji}(s) = p_{ji} = \tau p_{ii}.
$$

The second equal sign holds because of the symmetry of the two countries and because the price is the same for all varieties. It is important to make sure that $\frac{\partial L_{x,i}}{x_{ii}}$ and $\frac{\partial L_{x,i}}{x_{ji}}$ are independent of p_{ii}, p_{ji}, $x_{ii}(s)$ and $x_{ji}(s)$ in order to determine the price index and to show that each firm produces the same quantity. Then, with $p_{ii}(s) = p_{ii}$ and $p_{ji}(s) = p_{ji}$ it follows that $x_{ii}(s) = x_{ii}$ and $x_{ji}(s) = x_{ji}$.

Calculation of price aggregate P_i

Assume that $w_{x,i}\frac{\partial L_{x,i}}{\partial x_{ii}} = 1$ and $w_{x,i}\frac{\partial L_{x,i}}{\partial x_{ji}} = 1$, making prices and quantities produced constant:

$$
\begin{aligned}
P_i &= \left[\int_0^{A_i} p_{ii}^{1-\sigma}(s) ds + \int_0^{A_j} p_{ij}^{1-\sigma}(s) ds \right]^{\frac{1}{1-\sigma}} \\
&= \left[\int_0^{A_i} \left(\frac{\sigma}{\sigma - 1} \right)^{1-\sigma} ds + \int_0^{A_j} \left(\tau \frac{\sigma}{\sigma - 1} \right)^{1-\sigma} ds \right]^{\frac{1}{1-\sigma}} \\
&= \left[A_i \left(\frac{\sigma}{\sigma - 1} \right)^{1-\sigma} + A_j \left(\tau \frac{\sigma}{\sigma - 1} \right)^{1-\sigma} \right]^{\frac{1}{1-\sigma}} \\
&= \frac{\sigma}{\sigma - 1} \left(A_i + \tau^{1-\sigma} A_j \right)^{\frac{1}{1-\sigma}} = \frac{\sigma}{\sigma - 1} R_i^{\frac{1}{1-\sigma}}
\end{aligned}
$$

by symmetry P_j is:

$$
P_j = \frac{\sigma}{\sigma - 1} \left(A_j + \tau^{1-\sigma} A_i \right)^{\frac{1}{1-\sigma}} = \frac{\sigma}{\sigma - 1} R_j^{\frac{1}{1-\sigma}}.
$$

The price level decreases with A_i and A_j.[27] The higher the stock of designs, the lower the price aggregate.

A.2 Calculation of Migration Streams with Migration Costs

High skilled workers migrate to the country with the higher wage. Assume costs of k, which reduce the wage for all periods and following equation (2.59):

$$
\begin{aligned}
w_{H,i} - k &> w_{H,j} \\
\Rightarrow \delta \bar{A}_i \frac{x_i}{(r_i + \delta\alpha \frac{\bar{A}_i}{A_i} H_i)(\sigma - 1)} - k &= \delta \bar{A}_j \frac{x_j}{(r_j + \delta\alpha \frac{\bar{A}_i}{A_i} H_j(\sigma - 1)} \\
\bar{A}_i \frac{x_i}{r_i + \delta\alpha \frac{\bar{A}_i}{A_i} H_i} - \frac{\sigma - 1}{\delta} k &= \bar{A}_j \frac{x_j}{r_j + \delta\alpha \frac{\bar{A}_i}{A_i} H_j}
\end{aligned}
$$

[27] Remember that: $\sigma > 1 \rightarrow \frac{1}{1-\sigma} < -1$.

Assign the following:[28]

$$\eta = \frac{\sigma - 1}{\delta} k$$
$$s_i = \delta\alpha \frac{\bar{A}_i}{A_i}$$

Then:

$$\frac{\bar{A}_i x_i}{r_i + s_i H_i} - \eta = \frac{\bar{A}_j x_j}{r_j + s_j H_j}$$
$$\bar{A}_i x_i r_j + \bar{A}_i x_i s_j H_j - \eta(r_j + s_j H_j)(r_i + s_i H_i) = \bar{A}_j x_j r_i + \bar{A}_j x_j s_i H_i$$

with

$$\begin{aligned}
(r_j + s_j H_j)(r_i + s_i H_i) &= r_j r_i + r_j s_i H_i + s_j H_j r_j + s_j H_j s_i H_i \\
H_i H_j &= (L_{H,i} + m_{ji} L_{H,j})(1 - m_{ji}) L_{H,j} \\
&= L_{H,i} L_{H,j} - m_{ji} L_{H,i} L_{H,j} + m_{ji} L_{H,j}^2 - m_{ji}^2 L_{H,j}^2 \\
u_i &= \bar{A}_j x_j s_j + \eta r_j s_i \\
v_i &= \eta r_i s_j - \bar{A}_i x_i s_j
\end{aligned}$$

Then:

$$\begin{aligned}
\bar{A}_i x_i r_j - \bar{A}_j x_j x_j r_j - \eta r_j r_i =& \bar{A}_j x_j s_i H_i - \bar{A}_i x_i s_j H_j + \eta r_j s_i H_i + \eta s_j H_j r_i + \eta s_j H_j s_i H_i \\
=& H_i(\bar{A}_j x_j s_j + \eta r_j s_i) + H_j(\eta r_i s_j - \bar{A}_i x_i s_j) \\
=& H_i u_i + H_j v_i + \eta s_j H_j s_i H_i \\
=& (L_{H,i} + m_{ji} L_{H,j}) u_i + (1 - m_{ji}) L_{H,j} v_i + \eta s_j s_i L_{H,i} L_{H,j} \\
& - \eta s_j s_i m_{ji} L_{H,i} L_{H,j} + \eta s_j s_i m_{ji} L_{H,j}^2 - \eta s_i s_j m_{ji}^2 L_{H,j}^2 \\
=& u_i L_{H,i} + v_o L_{H,j} + \eta s_i s_j L_{H,i} L_{H,j} + m_{ji} \big(u_i L_{H,j} - v_i L_{H,j} \\
& - \eta s_i s_j L_{H,j} L_{H,i} + \eta s_i s_j L_{H,j}^2 \big) - \eta s_i s_j m_{ji}^2 L_{H,j}^2
\end{aligned}$$

[28] For s_j the indices of s_i are interchanged.

with

$$
\begin{aligned}
l_i &= u_i L_{H,j} - v_i L_{H,j} - \eta s_i s_j L_{H,j} L_{H,i} + \eta s_i s_j L_{H,j}^2 \\
\Rightarrow 0 &= u_i L_{H,i} + v_i L_{H,j} - \bar{A}_i x_i r_j \bar{A}_j x_j r_i \eta r_j r_i + \eta s_j s_i L_{H,i} L_{H,j} + m_{ji} l_i - \eta s_i s_j m_{ji}^2 L_{H,j}^2
\end{aligned}
$$

with

$$
\begin{aligned}
o_i &= u_i L_{H,i} + v_i L_{H,j} - \bar{A}_i x_i r_j + \bar{A}_j x_j r_i + \eta r_j r_i + \eta s_j s_i L_{H,i} L_{H,j} \\
\Rightarrow m_{ji}^2 &- \frac{l_i}{\eta s_i s_j L_{H,j}^2} - \frac{o_i}{\eta s_i s_j L_{H,j}^2} = 0 \\
m_{ji} &= \frac{l_i}{\eta s_i s_j L_{H,j}^2} \pm \sqrt{\frac{l_i}{2\eta s_i s_j L_{H,j}^2}^2 + \frac{o_i}{\eta s_i s_j L_{H,j}^2}}. \qquad \text{(A.17)}
\end{aligned}
$$

Only the minus sign makes sense in Eq. (A.17), as otherwise streams from country i to j are possible even if $w_{H,i} > w_{H,j}$.

The following terms are combined:

$$
\begin{aligned}
\eta &= \frac{\sigma - 1}{\delta} k \\
s_i &= \delta \alpha \frac{\bar{A}_i}{A_i} \\
u_i &= \eta r_j s_i + \bar{A}_j x_j s_i \\
v_i &= \eta r_i s_i - \bar{A}_i x_i s_j \\
o_i &= u_i L_{H,i} + v_i L_{H,j} - \bar{A}_i x_i r_j + \bar{A}_j x_j r_i + \eta r_j r_i + \eta s_j s_i L_{H,i} L_{H,j} \\
l_i &= L_{H,i} \left(u_i - v_i + \eta s_j s_i \left(L_{H,j} - L_{H,i} \right) \right).
\end{aligned}
$$

Finally, migration with costs can be summarised as:

$$
\begin{aligned}
m_{ji} &= \begin{cases} \min \left(\frac{l_i}{\eta s_i s_j L_{H,j}^2} - \sqrt{\frac{l_i}{2\eta s_i s_j L_{H,j}^2}^2 + \frac{o_i}{\eta s_i s_j L_{H,j}^2}}, 1 \right) & , \quad \text{if} \;\; w_{H,i} > w_{H,j} \\ 0 & , \quad \text{if} \;\; w_{H,i} \leq w_{H,j} \end{cases} \\
m_{ij} &= \begin{cases} \min \left(\frac{l_j}{\eta s_i s_j L_{H,i}^2} - \sqrt{\frac{l_j}{2\eta s_i s_j L_{H,i}^2}^2 + \frac{o_j}{\eta s_i s_j L_{H,i}^2}}, 1 \right) & , \quad \text{if} \;\; w_{H,j} > w_{H,i} \\ 0 & , \quad \text{if} \;\; w_{H,j} \leq w_{H,i} \end{cases} .
\end{aligned}
$$

Chapter 3

Estimating Common Correlated Effects Models

An early version of this chapter is under revise and resubmit for *The Stata Journal*. The `xtdcce2` package is available on Statistical Software Components (SSC) since August 2016.

3.1 Introduction

Estimating panels with heterogeneous coefficients in a panel with a large dimension of observations over cross-sectional units (N) and time periods (T) became standard in the last years, thanks to seminal work in theoretical econometrics (Pesaran and Smith, 1995; Shin et al., 1999). Heterogeneous slopes allow the researcher to identify effects for each cross-section separately. At the same time, the theoretical literature on how to account for unobserved dependence between cross-sectional units evolved (Pesaran, 2006; Chudik and Pesaran, 2015a). Not accounting for unobserved dependence between cross-sectional units causes the error term to be autocorrelated and leads to biased OLS regression results.

This chapter introduces and discusses a Stata package that combines these two strands of the literature. `xtdcce2` allows for Mean Group (MG) estimations in a dynamic panel with dependence between cross-sectional units. MG estimates are obtained by two steps. First, the coefficients of interest are estimated for each cross-sectional unit separately. The unit-specific estimates are averaged across all groups in a second step. `xtdcce2` approximates for cross-sectional dependence by adding cross-sectional averages and lags, as proposed by Pesaran (2006) and Chudik and Pesaran (2015a).[1] Furthermore, it tests for weak cross-sectional dependence in the error terms and allows for Instrumental Variable (IV) estimation as well. Additionally `xtdcce2` is able to correct for small sample time series bias by using the 'half-panel' jackknife correction method or the Recursive Mean Adjustment (RMA) method as proposed in Chudik and Pesaran (2015a).

`xtdcce2` differs in several ways from the existing estimation packages for common correlated effects in a heterogeneous panel. In comparison to `xtmg` (Eberhardt, 2012) it allows the consistent estimation of a dynamic panel by adding lags of the cross-sectional averages. Moreover, coefficients may be constrained to be homogeneous across all units. Additionally, unbalanced panels are supported. Compared to `xtpmg` (Blackburne and Frank, 2007), `xtdcce2` avoids ML estimations, offering the possibility to estimate models including endogenous independent variables. Hence, the main novelties within the setting of `xtpmg` and `xtmg` are the inclusion of a test for cross-sectional dependence, small time series bias correction methods and the support for IV regressions. IV regressions benefit from the `ivreg2` package. Possible applications for an IV estimation are endogenous spatial lags, which are instrumented by exogenous measures such as distance, other variables or higher order spatial lags. An application of this example will be carried out in the Chapter 4. Adding cross-sectional averages accounts for unobserved heterogeneity across units.

The `xtdcce2` package includes `xtcd2`, which tests for weak cross-sectional dependence (henceforth CD test) as proposed by Pesaran (2015) and Chudik and Pesaran (2015b). Two other programs, `xtcd` (by Markus Eberhardt) and `xtcsd` by De Hoyos and Sarafidis (2006), made the CD test already available in Stata. The novelties of `xtcd2` are the support of unbalanced panels, the possibility to test any variable for

[1] Chudik and Pesaran (2015b) give a comprehensive overview of the literature on (dynamic) common correlated effects, while Chudik and Pesaran (2015a) focuses on dynamic common correlated effects. In the following, common correlated effects cites Pesaran (2006), while dynamic common correlated effects cites Chudik and Pesaran (2015a), even though both are found in Chudik and Pesaran (2015b).

cross-sectional dependence and the option to plot the cross correlations as a kernel density plot.

The remainder of this chapter is structured as the following: the next section gives a brief introduction of the econometric methods. Then examples for an empirical application are given and the results compared to estimation procedures already available in Stata. The chapter closes with a Monte Carlo simulation and a conclusion. In order to improve the readability, the Appendix for this chapter is extensive and consists of three parts. The first part (B.1) explains the syntax, options and saved values of `xtdcce2` and `xtcd2`. The delta method to calculate the standard errors for pooled mean group estimations is explained in B.2. Finally, the setup of the Monte Carlo Simulation and tables with additional results can be found in B.3.

3.2 Estimators

For this chapter, the model from equation (1.25) is used. As common factors are the focus of this chapter, spatial dependence is set aside ($m_n = 0$) . In a Monte Carlo Simulation Chudik and Pesaran (2015a) show that the bias of the Dynamic Common Correlated Effects (DCCE) estimator for one or many common factors is similar. For a better readability and the sake of simplification, the number of common factors is set to one ($m_f = 1$). The underlying model for the DCCE estimator is then:

$$y_{i,t} = \alpha_i + \lambda_i y_{i,t-1} + \beta_i x_{i,t} + u_{i,t} \tag{3.1}$$

$$u_{i,t} = \gamma_i f_t + e_{i,t}, \tag{3.2}$$

The coefficients are defined as in the Introduction (1.3) as $\beta_i = \beta + v_i,\ v_i \sim IID(0, \mathbf{\Omega}_v)$, and $\lambda_i = \lambda + a_i,\ a_i \sim IID(0, \mathbf{\Omega}_a)$, where $\mathbf{\Omega}_v$ and $\mathbf{\Omega}_a$ are the variance covariance matrices. In the following β_i captures the short run effect of x on y. The factor loading (γ_i) and the common factors (f_t) are as defined as in Chapter 1.3.

The estimated equation is:

$$y_{i,t} = \alpha_i + \lambda_i y_{i,t-1} + \beta_i x_{i,t} + \sum_{l=0}^{p_T} \boldsymbol{\delta}'_{i,l} \bar{\mathbf{z}}_{t-l} + e_{i,t}, \tag{3.3}$$

where $\bar{\mathbf{z}}_t = 1/N \sum_{i=1}^{N} \mathbf{z}_{i,t} = (\bar{y}_t, \bar{x}_t)'$, $\bar{y}_t = 1/N \sum_{i=1}^{N} y_{i,t}$ and $\bar{x}_t = 1/N \sum_{i=1}^{N} x_{i,t}$ for $i = 1, ..., N$. For a pooled estimation the coefficients are constrained to be $\alpha_i = \alpha$, $\lambda_i = \lambda$ and $\beta_i = \beta$. The static version of the CCE estimator ($\lambda_i = 0$ for $i = 1, ..., N$) was made available in Stata by Markus Eberhard's `xtmg` command (Eberhardt, 2012).

The Pooled Mean Group (PMG) estimator (Shin et al., 1999) can be seen as an intermediate between a pure pooled estimation (homogeneous coefficients) and a MG estimation (heterogeneous coefficients). The assumptions of the Pooled Mean Group (PMG) estimator are that regressors have a homogeneous long run and a heterogeneous short run effect on the dependent variable. Equation (3.1) is transformed into an error correction model, such that

$$\Delta y_{i,t} = \phi_i (y_{i,t-1} - \theta_i x_{i,t}) + \alpha_i + \beta_i \Delta x_{i,t} + u_{i,t}. \tag{3.4}$$

$\phi_i = (\lambda_i - 1)$ is the error-correction speed of adjustment parameter and $(y_{i,t-1} - \theta_i x_{i,t})$ is the error correction term. In general, a long run relationship exists if $\phi \neq 0$ (Shin et al., 1999). β_i captures the immediate or short run effect of $x_{i,t}$ on $y_{i,t}$. The long run or equilibrium effect is captured by θ_i. The long run effect measures how the equilibrium changes and ϕ represents how fast the adjustment occurs. It can be estimated as the short run coefficient over the long run coefficient, $\theta_i = \beta_i / \phi_i = \beta_i / (\lambda_i - 1)$. In addition it is assumed to be homogeneous, while the short term dynamics are heterogeneous across units.[2] Shin et al. (1999) propose to estimate the long run coefficients by ML and the short run coefficients by OLS. The estimator is consistent as long as the disturbances are independently distributed across all individuals and time periods with a zero mean and a variance strictly larger than

[2]This notation follows Shin et al. (1999) Eq. (1) with $p = q = 1$.

zero.

The MG estimate and the variance of the short run coefficients are:

$$\hat{\boldsymbol{\delta}}_{MG} = \frac{1}{N}\sum_{i=1}^{N}\hat{\boldsymbol{\delta}}_i \tag{3.5}$$

$$\hat{\boldsymbol{\Sigma}}_{MG} = \widehat{\mathbf{Var}}(\hat{\boldsymbol{\delta}}_{MG}) = \frac{1}{N(N-1)}\sum_{i=1}^{N}\left(\hat{\boldsymbol{\delta}}_i - \hat{\boldsymbol{\delta}}_{MG}\right)^2, \tag{3.6}$$

where $\boldsymbol{\delta}_i = (\alpha_i, \beta_i)$.

The MG and the PMG estimator in the static and the dynamic version rely on large N and T. The literature on small sample time series bias corrections in dynamic heterogeneous panels is somewhat scarce, and it is for this reason that Chudik and Pesaran (2015a) focus on 'half-panel' jackknife and Recursive Mean Adjustment (RMA) bias correction methods. Neither requires knowledge of the error factor structure and can be applied to the mean group estimates.[3] The MG estimate of the 'half-panel' jackknife bias-corrected CCE estimator is

$$\tilde{\boldsymbol{\pi}}_{MG} = 2\hat{\boldsymbol{\pi}}_{MG} - \frac{1}{2}\left(\hat{\boldsymbol{\pi}}^a_{MG} + \hat{\boldsymbol{\pi}}^b_{MG}\right), \tag{3.7}$$

where $\hat{\boldsymbol{\pi}}^a_{MG}$ is the MG estimate of the first half ($t = 1, ..., \frac{T_i}{2}$) of the panel and $\hat{\boldsymbol{\pi}}^b_{MG}$ of the second half ($t = \frac{T_i}{2} + 1, ..., T_i$) of the panel.

The RMA method removes the partial mean from the all variables, meaning:

$$\tilde{\boldsymbol{\omega}}_{i,s} = \boldsymbol{\omega}_{i,s} - \frac{1}{t-1}\sum_{s=1}^{t-1}\boldsymbol{\omega}_{i,s}, \tag{3.8}$$

where $\boldsymbol{\omega}_{i,s} = (y_{i,s}, x_{i,s})$ or any other variable except the constant. In line with Chudik and Pesaran (2015a) the partial mean is lagged by one period to prevent it from being influenced by endogenous observations.

From the asymptotics of the unit-specific estimates and the MG estimates, restrictions on the dataset arise. The number of cross-sectional units and time pe-

[3] For a further discussion see Chudik and Pesaran (2015a) or Everaert and De Vos (2016).

riods is assumed to grow with the same rate. In an empirical setting, this can be interpreted as N/T being constant. A dataset with one dimension being large in comparison to the other would lead to inconsistent estimates, even if both dimension are large in numbers. For example a financial dataset on stock markets returns on a monthly basis over 30 years (T=360) of 10,000 firms would not be sufficient. While individually both dimensions can be interpreted as large, they do not grow with the same rate and the ratio would not be constant. Therefore, an estimator relying on fixed T asymptotics and large N would be appropriate. On the other hand a dataset with say $N = 30$ and $T = 34$ would qualify as appropriate, if N and T grow at the same rate.[4]

3.3 Syntax

`xtdcce2` has the following syntax:

```
xtdcce2 depvar [indepvars] [ (varlist2 = varlist_iv)] [if] ,
    crosssectional(varlist_cr) [ pooled(varlist_p) nocrosssectional cr_lags(#)
    ivreg2options(string) e_ivreg2 ivslow noisily lr(varlist_lr) lr_options(string)
    noconstant pooledconstant reportconstant trend pooledtrend jackknife
    recursive nocd showindividual fullsample ]
```

varlist_cr defines the variables for the cross-sectional averages, *varlist_p* the pooled variables. *varlist2* includes endogenous variables, instrumented by *varlist_iv*. The syntax and options are further explained in Appendix B.1 and comments about the versions can be found in B.1.2.

The syntax for `xtcd2` is:

```
xtcd2 [varname], [ noestimation rho kdensity name(string) ]
```

[4] An anonymous Stata Journal referee suggested the two examples, which is much appreciated.

varname is the name of the residual or variable to be tested. *varname* is optional in case the command is performed after an estimation command which supports `predict, residuals`. Then `xtcd2` predicts and tests the residuals for weak cross-sectional dependence.

3.4 The Constant in xtdcce2

`xtdcce2` can treat the individual-specific constants α_i in several ways. In Pesaran (2006) and Chudik and Pesaran (2015b) the individual-specific constants is partialled out as they are a part of the matrix which includes the cross-sectional averages. In common correlated effects regressions, the individual-specific constants include the factor loadings and parts of the cross-sectional averages (see Chudik and Pesaran (2015a), p. 397). Pesaran and Tosetti (2011) show that it is possible to estimate the individual-specific constant α_i if N and T converge to infinity.[5]

`xtdcce2` estimates and reports a MG estimate of the individual-specific constants if the option `reportconstant` is used. Otherwise. they are partialled out or removed from the model and not reported. Additionally `xtdcce2` allows the constants to be the same across all units by specifying the option `pooledconstant`.[6] As a final option, the constants can be completely removed from the model by using the `noconstant` option.

The individual-specific constants are removed from the model if all parameters including the constants are constrained to be homogeneous, the cross-sectional means include all variables and the dataset is strongly balanced. Loosely speaking, by partialling out the time averages of the dependent variable and all independent variables, the data is demeaned and a homogeneous constant is rendered to be zero. Thus `xtdcce2` automatically removes the constant from the model to improve the estimation. If the option `reportconstant` is used, then the constant is still estimated

[5] They show that $\hat{\alpha}_i - \alpha_i = O_p\left(\frac{1}{N}\right) + O_p\left(\frac{1}{\sqrt{NT}}\right) + O_p\left(\frac{1}{\sqrt{T}}\right)$, see Pesaran and Tosetti (2011), Eq. (31), page 189.

[6] If `pooledconstant` is used but not `reportconstant`, the constant is internally calculated but not displayed.

and reported in the output.

There is a notable difference between in the treatment of the constant in `xtdcce2` and the official StataCorp command `xtreg, fe`. If all independent variables excluding the individual-specific constants are pooled and no cross-sectional averages added, then `xtdcce2` leads to the same point estimates as `xtreg, fe` for $\boldsymbol{\pi}$. It is important to note that Stata assumes for `xtreg, fe` a model including a constant and a fixed effect (in the fashion of $y_{i.t} = \nu + \alpha_i + \beta x_{i,t} + \epsilon_{i,t}$), while the model in this chapter is described by equation 3.1. The MG estimates for the constant ($\hat{\bar{\alpha}}_{MG}$) and the constant estimated by `xtreg, fe` ($\hat{\nu}_{FE}$) are the same, the individual fixed effects obtained by `predict, u` after `xtreg, fe` differ by the MG estimate of the constant ($\hat{\alpha}_{i,FE} = \hat{\alpha}_{i,MG} - \hat{\bar{\alpha}}_{MG}$). `xtreg, fe` demeans all variables to remove the individual fixed effects, but adds the overall mean in order to obtain estimates for the constant, i.e. $\ddot{y}_{i,t} = y_{i,t} - \bar{y}_i + \bar{y}$, where $\bar{y}_i = 1/T \sum_{t=1}^{T} y_{i,t}$ and $\bar{y} = 1/(NT) \sum_{i=1}^{N} \sum_{t=1}^{T} y_{i,t}$. Then it calculates the individual effects by $\alpha_i = \bar{y}_i - \hat{\alpha} - \bar{x}_i \hat{\beta}$. The difference between the two estimators is, that `xtdcce2` identifies α_i as the (heterogeneous cross-sectional specific) constant, while `xtreg, fe` treats ν as a homogeneous constant.

3.5 Empirical Examples

In this section, three empirical examples are carried out to demonstrate the use of `xtdcce2`. A Solow model with dynamic common correlated effects and instrumental variable regression is estimated. In two other examples, estimations using `xtdcce2` are compared with the existing Stata packages `xtmg` and `xtpmg`.

3.5.1 Dynamic Common Correlated Effects and testing for cross-sectional dependence

As a first exercise, the Solow model in style of Mankiw et al. (1992), Islam (1995) and Lee et al. (1997) is estimated. The dependent variable is log GDP per capita `d.log_rgdpo` and the independent variables are lagged GDP per capita `L.log_rgdpo`, physical capital `log_ck` and the population growth rate `log_ngd`.[7] The Penn World Tables (Feenstra, Inklaar and Timmer, 2015) version 8.0 are used and restricted to the years from 1960 to 2007, which means that there is a maximum of $T = 48$ years. Both independent variables and the level on the dependent variable are added as cross-sectional averages, set by the option `crosssectional()`. The number of cross-section averages is set to $[\sqrt[3]{48}] = 3$, specified by `cr_lags()`. Together with the first lag of `log_rgdpo` and the three lags of the cross-sectional averages, 4 time periods are lost and the number of time periods used is reduced to 44. The cross-sectional dimension is in comparison to the time dimension larger ($\frac{N}{T} = \frac{93}{44} = 2.11$). To account for the small sample time series bias, the 'half-panel' jackknife bias correction method is applied using the option `jackknife`.

```
. use xtdcce2_sample_dataset.dta

. xtset id year
       panel variable:  id (strongly balanced)
        time variable:  year, 1960 to 2007
                delta:  1 unit

. xtdcce2 log_rgdpo L.log_rgdpo log_ck log_ngd ,  /*
> */ crosssectional(log_rgdpo log_ck log_ngd) cr_lags(3) jackknife
(Dynamic) Common Correlated Effects Estimator - Mean Group

Panel Variable (i): id                                 Number of obs      =      4092
Time Variable (t): year                                Number of groups   =        93
                                                       Obs per group (T)  =        44
Degrees of freedom per cross-sectional unit:
 without cross-sectional averages    =  41             F(1396, 2696)      =      5.10
 with cross-sectional averages       =  28             Prob > F           =      0.00
Number of                                              R-squared          =      0.73
 cross-sectional lags                =  3              Adj. R-squared     =      0.58
 variables in mean group regression  =  279            Root MSE           =      0.06
 variables partialled out            =  1117
                                                       CD Statistic       =      0.64
                                                          p-value         =    0.5226
-------------------------------------------------------------------------------------
           log_rgdpo |      Coef.   Std. Err.       z    P>|z|     [95% Conf. Interval]
---------------------+---------------------------------------------------------------
Mean Group Estimates:|
         L.log_rgdpo |     .61254     .028504   21.49    0.000     .5566735   .6684074
              log_ck |    .115106     .037524    3.07    0.002     .0415604   .1886508
             log_ngd |    .045182     .106571    0.42    0.672    -.1636919   .2540568
-------------------------------------------------------------------------------------
Mean Group Variables: L.log_rgdpo log_ck log_ngd
Cross-Sectional Averaged Variables: log_rgdpo log_ck log_ngd
Heterogenous constant partialled out. Jackknife bias correction used.
```

On the lower right of the upper panel, the output shows a CD test statistic of

[7] In Mankiw et al. (1992) the dependent variable is the first difference of log GDP per capital. For the purpose of a true lag, the level is used as a dependent variable. The only difference is the interpretation of the coefficient on the lagged dependent variable.

0.64 with a p-value of 0.52, so the null hypothesis of weak cross-sectional dependence fails to be rejected. Below the coefficient estimates, `xtdcce2` displays the names of the 3 mean group variables and 3 cross-sectional averages.

A regression without any pooled variables is essentially a regression run on each country separately. The degree of freedom of a regression on each country separately is shown on the left hand side under the time variable identifier. The first line shows the degree of freedom without the inclusion of cross-sectional averages, which results in the number of time periods used ($T = 44$) minus the number of variables ($K = 3$). In the line below, the degree of freedom for each country with cross-sectional averages is displayed. It equals the number of time periods ($T = 44$) minus the number of variables ($K = 3$), minus the number of cross-section averages times the number of lags ($p_T = 3$) plus one for the contemporaneous averages and minus one for the constant ($44 - 3 - 3 * (3 + 1) - 1 = 28$). In the section below, the number of lags of the cross-sectional means is displayed, together with the number of variables in the mean group regression and the number of variables partialled out, which equals the number of cross-sectional averages. `xtdcce2` displays the degree of freedom with and without the cross-sectional averages to make the researcher aware of the degree of freedom which is lost when including the cross-sectional averages. This is especially important if the number of explanatory variables or lags of the cross-sectional averages or both is large.

As the cross-sectional averages are purely treated as controls and have no interpretation, no information is lost by partialling out. Therefore, the averages are regressed on each of the explanatory variables of interest and then the residuals collected. The residuals are then used as the new explanatory and dependent variables.

The partialling out is performed in Mata. The variables to be partialled out (the cross-sectional means and, if requested, the heterogeneous intercept) are stacked in a block diagonal matrix, with zeros on the off diagonals. For a large number of units the matrix becomes sparse and calculating and inverting the cross product becomes computationally intensive, hence time consuming. To improve speed, the partialling out is done sequentially unit by unit, which is possible as long as the coefficients on the cross-sectional means, $\delta_{i,l}$, are heterogeneous.[8] Within this process, `xtdcce2` checks if the factor loadings are full rank.[9] If the check fails, the mean group estimates are still consistent, but not the cross-sectional individual estimates. Therefore `xtdcce2` shows a warning at the end of the output. For the calculation of the cross-sectional averages and the partialling out, the dataset is restricted to the observations used in the regression. In total 4 periods are lost; one for the lag of the dependent variable and a further 3 for the cross-sectional averages. So the time span for the regression are the years 1964 - 2007, making the time dimension $T = 44$.

The regression results are in favour of the Solow model. The coefficients on physical capital and the lagged dependent variable are positive and significant, while the coefficient on population growth is positive but not significant. The estimated capital share is around 23%.[10] For a more detailed discussion of the Solow model in growth empirics see Mankiw et al. (1992) or Durlauf, Johnson and Temple (2005); Jones (2016); and in a panel Islam (1995)with a focus on slope heterogeneity see Islam (1998) and Lee et al. (1997, 1998).

To predict the error term `predict, residuals` is used. Then the test on cross-sectional dependence can be done by hand to confirm the result from above.

```
. predict xtdcce2_residuals, residuals

. xtcd2 xtdcce2_residuals
Pesaran (2015) test for weak cross-sectional dependence
H0: errors are weakly cross-sectional dependent.
        CD = 0.639
```

[8]The precision lies in a negligible order of magnitude and is offsetted by the improvement in speed.
The standard solver for the calculation of the inverse of the cross product of the factor loadings is `cholsolve`. `cholsolve` cannot solve positive definite or singular matrices. In this case `qrsolve` is used.

[9]The condition of a full rank is checked on the unit-specific matrices containing the cross-sectional averages ($\bar{\mathbf{z}}_i = (\bar{\mathbf{y}}_i, \bar{\mathbf{x}}_i)$, where $\bar{\mathbf{y}}_i$ and $\bar{\mathbf{x}}_i$ are $T \times 1$ and $T \times K$ matrices containing the cross-sectional averages). This is possible as the matrix over all units is block diagonal and the rank is equal to the sum of the rank of the blocks.

[10]The calculation is $\alpha = \_b[log\_ck]/(1 - \_b[L.log\_rgdpo] + \_b[log\_ck])$. See Ditzen and Gundlach (2016) for a more detailed discussion about the estimation of the capital shares in the Solow model.

```
    p-value = 0.523
```

It is possible to calculate the residuals including the common factors, using `predict, cfresiduals`. It is important to note for this option, that the inclusion of the constant in the common factors depends on the command line. If the option `reportconstant` is used, then the constant is excluded. Therefore, the model is re-estimated and the CD test carried out:

```
. xtdcce2 log_rgdpo L.log_rgdpo log_ck log_ngd ,  /*
> */ crosssectional(log_rgdpo log_ck log_ngd) cr_lags(3) jackknife reportconstant
. predict xtdcce2_cf_residuals, cfresiduals
. xtcd2 xtdcce2_cf_residuals
Pesaran (2015) test for weak cross-sectional dependence
H0: errors are weakly cross-sectional dependent.
        CD = 18.374
   p-value = 0.000
```

As expected, the null hypothesis of weak cross-sectional dependence is rejected and the residuals including the common factors exhibit strong cross-sectional dependence. Using the option `noestimation` leads to the same result, as long as the observations which are omitted in the estimation are missing in the variable `residuals`. The advantage of `noestimation` is that it does not require a sample being set by `e(sample)`, and therefore any observable variable can be tested for weak cross-sectional dependence. For example testing the independent variable for weak cross-sectional dependence reads:

```
. xtcd2 log_rgdpo , noestimation
Pesaran (2015) test for weak cross-sectional dependence
H0: errors are weakly cross-sectional dependent.
        CD = 452.528
   p-value = 0.000
```

`estat` can be used for a graphical analysis of the MG regression results.

```
. estat rcap log_ngd log_ck if id <= 20
Combined graph saved as xtdcce2_combine.
```

In Figure 3.1, the MG estimates for the first 20 coefficients of *log_ngd* and *log_ck* are plotted using a `range plot`. The range plot exhibits the 95% confidence interval and the point estimate is depicted by the cross. Additionally, the MG point estimate along with it's 95% confidence interval is added. The function is intended to give the user of `xtdcce2` some guidance about the distribution of the unit specific estimates and the interplay between those and the MG estimates. Another advantage is that it allows an easy identification of outliers. As the mean group estimator is the unweighted average across the unit specific estimates, outliers have a huge effect in panels with a small number of cross-sectional units.

The plot underlines the difference between the unit specific and the MG estimates. It shows that the unit specific estimates are not necessarily in the confidence set of the MG estimates. Important to note is, that the MG estimates are consistent even if the unit specific estimates are not. A possible reason for the bias is the length of the time series, which is with 47 periods rather short. This biases both the unit and the MG estimates. The unit specific estimates are biased because they solely rely on the unit specific time series and have a degree of freedom as shown on the output on page 88. As discussed in Chapter 3.6, the bias of the MG estimates decreases substantially with an increase in the number of time periods. In addition, outliers of the estimated unit specific coefficients can have an effect on the MG estimates. For example, it becomes apparent that unit 8 is an outlier with respect to both coefficients. For *log_ck* the confidence interval of the unit specific estimate even lies outside of the confidence interval of the MG estimate.

In a next step, all 3 coefficients are constrained to be the same across countries, $\beta_{i,k} = \beta_k, \forall i = 1, .., N, k = 1, ..., 3$, by specifying the `pooled()` option. The constant is pooled using the `pooledconstant` option and `xtdcce2` is forced to report the constant using `reportconstant`.[11]

```
. xtdcce2 log_rgdpo L.log_rgdpo log_ck log_ngd , /*
> */ pooled(L.log_rgdpo log_ck log_ngd)   /*
> */ crosssectional(log_rgdpo log_ck log_ngd) cr_lags(3) /*
> */ pooledconstant reportconstant
(Dynamic) Common Correlated Effects Estimator - Pooled

Panel Variable (i): id                                  Number of obs      =      4092
Time Variable (t): year                                 Number of groups   =        93
                                                        Obs per group (T)  =        44
Degrees of freedom per cross-sectional unit:
 without cross-sectional averages    =  40              F(1120, 2972)      =      6.78
 with cross-sectional averages       =  28              Prob > F           =      0.00
Number of                                               R-squared          =      0.72
```

[11]Pesaran (2006) discusses the MG and the pooled version of the CCE estimator. A pooled version in the dynamic setting is not mentioned in Chudik and Pesaran (2015a).

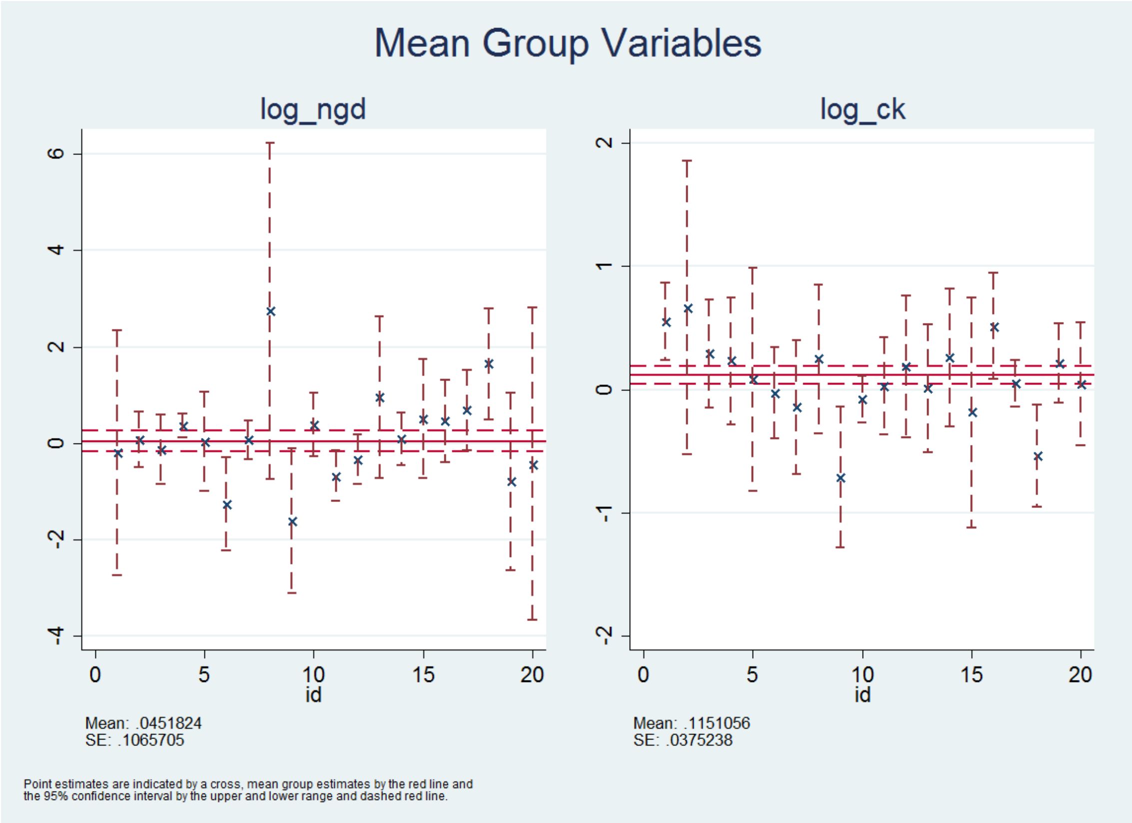


Figure 3.1: Range plot for mean group estimates of the variables log_ngd and log_ck.
Point estimates are indicated by a cross, the mean group estimates by the red line and the 95% confidence interval by the range plots.

```
cross-sectional lags                  =  3                Adj. R-squared      =      0.61
variables in mean group regression    =  4                Root MSE            =      0.06
variables partialled out              =  1116
                                                          CD Statistic        =     -0.89
                                                             p-value          =    0.3738
-------------------------------------------------------------------------------------------
              log_rgdpo |     Coef.     Std. Err.      z     P>|z|     [95% Conf. Interval]
------------------------+------------------------------------------------------------------
Pooled Variables:       |
            L.log_rgdpo |   .796726     .066135     12.05   0.000      .6671033   .9263487
                 log_ck |   .084764     .042991      1.97   0.049      .0005027   .1690251
                log_ngd |   .012159     .045954      0.26   0.791     -.0779094    .102228
                  _cons |   2.4e-14       .9422      0.00   1.000     -1.846679   1.846679
-------------------------------------------------------------------------------------------
Pooled Variables:  L.log_rgdpo log_ck log_ngd _cons
Cross-Sectional Averaged Variables: log_rgdpo log_ck log_ngd
```

The estimate for the constant is zero with probability 1. This result is expected, as outlined in the section above, because all coefficients are pooled, the panel is balanced and all variables are added as cross-sectional means. For the calculation of the standard errors, `xtdcce2` follows the approach from Pesaran (2006) as outlined in section (1.5), equation 1.43. To obtain estimates of π_i and π_{MG}, xtdcce2 performs in the background a MG estimation.

As a final exercise, assume that investments into physical capital are endogenous. Countries with a large GDP per capita can save more and therefore accumulate more capital. This leads to a reversed causality of investments into physical capital and the level of GDP; for a discussion see for example Durlauf et al. (2005) or Temple (1999). As suggested in Temple (1999) lags of the endogenous variable are used as an instrument. In order to avoid a further drop in the degree of freedom by adding more variables to the model, the first two lags are used as instruments.

In line with the syntax of ivreg2 (Baum, Schaffer and Stillman, 2003, 2007), instrumented (endogenous) variables and the instruments are enclosed in parenthesis, where the instrumented variable is followed by an equal sign and the instruments:

```
. xtdcce2 log_rgdpo L.log_rgdpo  log_ngd /*
> */ (log_ck = L.log_ck L2.log_ck) , /*
> */ crosssectional(log_rgdpo log_ck log_ngd) cr_lags(3) /*
> */ ivreg2options(noid)
(Dynamic) Common Correlated Effects Estimator - Mean Group IV

Panel Variable (i): id                                  Number of obs      =      3999
Time Variable (t): year                                 Number of groups   =        93
                                                        Obs per group (T)  =        43

Degrees of freedom per cross-sectional unit:
 without cross-sectional averages    =  40              F(1396, 2603)      =     23.27
 with cross-sectional averages       =  27              Prob > F           =      0.00
Number of                                               R-squared          =      0.72
 cross-sectional lags                =  3               Adj. R-squared     =      0.56
 variables in mean group regression  =  279             Root MSE           =      0.04
 variables partialled out            =  1117
                                                        CD Statistic       =      1.12
                                                           p-value         =    0.2618
------------------------------------------------------------------------------------------
              log_rgdpo |     Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
------------------------------------------------------------------------------------------
Mean Group Estimates:   |
                 log_ck |   .021632    .035794     0.60   0.546    -.0485226   .0917865
            L.log_rgdpo |   .599657    .024456    24.52   0.000     .5517248   .6475899
                log_ngd |   .063232    .090544     0.70   0.485    -.1142319    .240696
------------------------------------------------------------------------------------------
Mean Group Variables: L.log_rgdpo log_ngd
Cross-Sectional Averaged Variables: log_rgdpo log_ck log_ngd
Endogenous Variables:  log_ck
Exogenous Variables:  L.log_ck L2.log_ck
Heterogenous constant partialled out.
```

3.5.2 Pooled Mean Group

In the following, `xtdcce2` is compared to results from `xtpmg`, by Blackburne and Frank (2007). `xtpmg` implements the PMG estimator by Shin et al. (1999) into Stata. The two programs differ in two ways. First of all `xtpmg` estimates the following equation

$$\Delta c_{i,t} = \phi_i(c_{i,t-1} - \theta_{1,i} y_{i,t} - \theta_{2,i} \pi_{i,t}) + \delta_{0,i} + \delta_{1,i} \Delta y_{i,t} + \delta_{2,i} \Delta \pi_{i,t} + \epsilon_{i,t}, \qquad (3.9)$$

while `xtdcce2` internally estimates (leaving out any cross-sectional averages):

$$\Delta c_{i,t} = \phi_i c_{i,t-1} + \gamma_{1,i} y_{i,t} + \gamma_{2,i} \pi_{i,t} + \delta_{0,i} + \delta_{1,i} \Delta y_{i,t} + \delta_{2,i} \Delta \pi_{i,t} + \epsilon_{i,t}, \qquad (3.10)$$

Secondly, `xtpmg` calculates the long run coefficients using ML. `xtdcce2` treats the long run coefficients, defined in `lr()`, as further covariates and estimates equation (3.10) entirely by OLS. To calculate the long run coefficients, the coefficients are divided by the negative of the long run cointegration vector to match equation (3.9), $\theta_{1,i} = -\gamma_{1,i}/\phi_i$. The variances are calculated using the Delta method as described in the Appendix. B.2. Equation (3.10) and the coefficients $\gamma_{1,i}, ..., \gamma_{K,i}$ can be estimated by using `lr_options(nodivide)`.

The *jasa2* dataset is used to explain consumption with inflation and income after 1962 as in Blackburne and Frank (2007).[12] The present dataset is unbalanced as one cross-sectional unit misses an observation for the year 1993. `xtdcce2` checks if any panel misses observations before calculating the cross-sectional averages and estimating the coefficients.[13] In the case of missing observations, the panel is unbalanced and `xtdcce2` removes all observations for the cross-sectional unit in the specific time period for further calculations. In the *jasa2* dataset, the minimum

[12]The dataset is available at http://www.econ.cam.ac.uk/faculty/pesaran.

[13]Stata defines 3 types of balances for a panel: strongly balanced, weakly balanced or unbalanced. In a strongly balanced panel all cross-sectional units have the same time values. A weakly balanced panel is defined as all cross-sectional units have the same number of time values. All other panels are unbalanced. See *help tsset*.

number of time periods is 31 and the maximum number is 32 time periods. N/T can be assumed constant and therefore a common correlated effects estimator can be applied. The output from `xtdcce2` is the following:[14]

```
. use jasa2, clear

. tsset id year
       panel variable:  id (unbalanced)
        time variable:  year, 1960 to 1993
                delta:  1 unit

. eststo xtdcce1: xtdcce2 d.c d.pi d.y if year >= 1962 , /*
> */ lr(l.c pi y) p(l.c pi y) nocrosssectional lr_options(xtpmgnames)
(Dynamic) Common Correlated Effects Estimator - Pooled Mean Group

Panel Variable (i): id                                     Number of obs      =       767
Time Variable (t): year                                    Number of groups   =        24

                                                           Obs per group:
                                                                          min =        31
                                                                          avg =        32
                                                                          max =        32
Degrees of freedom per cross-sectional unit:
 without cross-sectional averages    =  26.958333          F(52, 715)         =     36.62
 with cross-sectional averages       =  25.958333          Prob > F           =      0.00
Number of                                                  R-squared          =      0.73
 cross-sectional lags                =  0                  Adj. R-squared     =      0.71
 variables in mean group regression  =  51                 Root MSE           =      0.02
 variables partialled out            =  1
                                                           CD Statistic       =      4.10
                                                              p-value         =    0.0000
------------------------------------------------------------------------------------------
                     D.c |      Coef.   Std. Err.      z     P>|z|     [95% Conf. Interval]
-------------------------+----------------------------------------------------------------
Short Run Estimates:     |
-------------------------+----------------------------------------------------------------
  Mean Group Estimates:  |
                    D.pi |  -.054823    .029859    -1.84   0.066     -.1133453   .0036984
                     D.y |   .380249    .035007    10.86   0.000      .3116365   .4488617
-------------------------+----------------------------------------------------------------
Long Run Estimates:      |
-------------------------+----------------------------------------------------------------
  Pooled Variables:      |
                      ec |  -.168358    .119581    -1.41   0.159     -.4027324    .066017
                      pi |  -.194124    .111477    -1.74   0.082     -.4126141   .0243664
                       y |   .902577    .131913     6.84   0.000      .6440312   1.161122
------------------------------------------------------------------------------------------
Pooled Variables:  ec pi y
Mean Group Variables: D.pi D.y
Long Run Variables: ec pi y
Heterogenous constant partialled out.
```

The long- and the short-run estimates are split up into two parts; one showing the results for the average long and average short run coefficients.[15] As the dataset

[14] For a later use the regression results are stored using the `estout` package (Jann, 2004).

[15] First the long run coefficients for each cross-section are computed and in a second step the individual long run coefficients are averaged. As an example, the average long run coefficient for $\hat{\bar{\theta}}_1$ is calculated as: $\hat{\bar{\theta}}_1 = 1/N \sum_{i=1}^{N} \hat{\theta}_{1,i} = 1/N \sum_{i=1}^{N} (-\hat{\gamma}_{1,i}/\hat{\phi}_i)$. If ϕ is heterogeneous, but let's say θ_1 homogeneous, then the long run coefficient γ_1 is calculated as $\hat{\gamma}_1 = -\hat{\theta}_1/(1/N \sum_{i=1}^{N} \hat{\phi}_i)$. The

is unbalanced, the minimum, average and maximum number of time periods are displayed. For the remaining regressions, `esttab` produces the following output:

```
. eststo xtpmg:  qui xtpmg d.c d.pi d.y if year>=1962, lr(l.c pi y) ec(ec) replace pmg

. eststo xtdcce2: qui xtdcce2 d.c d.pi d.y if year >= 1962 , /*
> */ lr(l.c pi y) pooled(l.c pi y) nocrosssectional lr_options(nodivide xtpmgnames)

. eststo xtdcce3: qui xtdcce2 d.c d.pi d.y if year >= 1962 , /*
> */ lr(l.c pi y) pooled(l.c pi y) crosssectional(d.c d.pi d.y)  /*
> */ cr_lags(0) lr_options(xtpmgnames)

. esttab xtpmg xtdcce1 xtdcce2 xtdcce3 /*
> */ , mtitles("xtpmg - mg" "xtdcce2 - mg" "xtdcce2 - mg" "xtdcce2 - cce" )  /*
> */ modelwidth(13)  se  s(N cd cdp)
```

	(1) xtpmg - mg	(2) xtdcce2 - mg	(3) xtdcce2 - mg	(4) xtdcce2 - cce
ec				
pi	-0.466***	-0.194	-0.0327	-0.276
	(0.0567)	(0.111)	(0.0473)	(0.195)
y	0.904***	0.903***	0.152**	0.940***
	(0.00868)	(0.132)	(0.0541)	(0.0895)
SR				
ec	-0.200***	-0.168	-0.168***	-0.184*
	(0.0322)	(0.120)	(0.0490)	(0.0901)
D.pi	-0.0183	-0.0548	-0.0548	0.0237
	(0.0278)	(0.0299)	(0.0299)	(0.0317)
D.y	0.327***	0.380***	0.380***	0.384***
	(0.0574)	(0.0350)	(0.0350)	(0.0431)
_cons	0.154***			
	(0.0217)			
N	767	767	767	767
cd		4.101	4.101	0.671
cdp		0.0000410	0.0000410	0.502

Standard errors in parentheses
* p<0.05, ** p<0.01, *** p<0.001

Column (1) shows the results using `xtpmg`, columns (2) - (4) using `xtdcce2`. Column (1) matches the results from Blackburne and Frank (2007, p. 203). As expected, the MG estimates obtained by `xtpmg` and `xtdcce2` differ due to the different estimation methods. However, the signs of the MG estimates are the same, especially for the short run coefficients. This implies that `xtdcce2` can be employed to estimate the pooled mean group model. In column (3) the option `nodivide` is

calculation of the average long run coefficients is the same as in Chudik et al. (2016). In addition, if the the differences are replaced by lags, the Cross-Sectionally Augmented Distributed Lag (CS-DL) estimator (Chudik et al., 2016) with one lag of the independent and dependent variable is obtained.

used, producing estimates for equation (3.10).[16]

As the last row indicates, using no cross-sectional averages leads to a rejection of the null hypothesis of weak cross-section dependence. Cross-sectional dependence remains in the residuals and OLS becomes inconsistent. To account for the cross-sectional dependence, cross-sectional averages are added in column (4). The p-value (row `cdp`) increases to 0.5 and the hypothesis of weak cross-sectional dependence cannot be rejected any longer.

The average short run coefficients can be restricted to be equal across all units by including them in the `pooled()` option. At the same time, the average long run coefficients can be allowed to vary as well. To test under which constraints the model is consistent, the Hausman test can be performed:

```
. eststo mg: qui xtdcce2 d.c d.pi d.y if year >= 1962 , /*
> */ lr(l.c pi y)  nocrosssectional

. eststo pmg: qui xtdcce2 d.c d.pi d.y if year >= 1962 , /*
> */ lr(l.c pi y) pooled(l.c pi y) nocrosssectional

. eststo pooled: qui xtdcce2 d.c d.pi d.y if year >= 1962 , /*
> */ lr(l.c pi y) pooled(l.c pi y d.pi d.y) nocrosssectional

. hausman mg pooled, sigmamore

                 ---- Coefficients ----
             |      (b)          (B)            (b-B)     sqrt(diag(V_b-V_B))
             |      mg         pooled        Difference          S.E.
-------------+----------------------------------------------------------------
          pi |
         D1. |  -.0253642    -.0280826        .0027184               .
           y |
         D1. |   .2337588     .3811944       -.1474357               .
           c |
         L1. |  -.3063473    -.1794146       -.1269326               .
          pi |  -.3529095     -.266343       -.0865666         .0844914
           y |   .9181344     .9120574        .0060771               .
------------------------------------------------------------------------------
                         b = consistent under Ho and Ha; obtained from xtdcce2
          B = inconsistent under Ha, efficient under Ho; obtained from xtdcce2

    Test:  Ho:  difference in coefficients not systematic

                  chi2(5) = (b-B)'[(V_b-V_B)^(-1)](b-B)
                          =        2.37
                Prob>chi2 =      0.7964
                (V_b-V_B is not positive definite)

. hausman pmg pooled, sigmamore

                 ---- Coefficients ----
             |      (b)          (B)            (b-B)     sqrt(diag(V_b-V_B))
             |      pmg        pooled        Difference          S.E.
```

[16] An alternative to obtain long run coefficients in a dynamic panel using a restricted version of the between estimator is outlined in Ditzen and Gundlach (2016).

```
             |
          pi |
         D1. |  -.0548234    -.0280826       -.0267408              .
           y |
         D1. |   .3802491     .3811944       -.0009453              .
           c |
         L1. |  -.1683577    -.1794146        .0110569              .
          pi |  -.1941238     -.266343        .0722191       .0396237
           y |   .9025766     .9120574       -.0094807              .
------------------------------------------------------------------------
                          b = consistent under Ho and Ha; obtained from xtdcce2
           B = inconsistent under Ha, efficient under Ho; obtained from xtdcce2

    Test:  Ho:  difference in coefficients not systematic

                  chi2(5) = (b-B)'[(V_b-V_B)^(-1)](b-B)
                          =        0.97
                Prob>chi2 =      0.9650
                (V_b-V_B is not positive definite)
```

The result of the Hausman test is similar to the one obtained in Blackburne and Frank (2007, Section 4.3 and 4.4). The first Hausman test implies that the pooled model is preferred over the mean group model. The second Hausman test compares the pooled mean group and the pooled model. The conclusion is different from that in Blackburne and Frank (2007), in that the pooled group model is preferred. However one difference to Blackburne and Frank (2007) and one limitation of the Hausman test are worth noting. First of all, `xtdcce2` includes in the Hausman test all coefficients, while Blackburne and Frank (2007) only include the coefficients of the long run vector (*pi* and *y*). Secondly, as Pesaran and Yamagata (2008) point out, a Hausman test lacks power in the case of pure exogenous regressors if under the null hypothesis the slope parameters are drawn from the same distribution. Also a test for slope homogeneity in multifactor error structure models in a large N and large T panel with an unknown number of factors and a lagged dependent variable has not been established.[17]

3.5.3 Mean Group and Common Correlated Effects

`xtdcce2` is able to compute the MG and CCE estimators by Pesaran and Smith (1995) and Pesaran (2006), introduced to Stata by the `xtmg` command (Eberhardt, 2012). Following Eberhardt (2012) using the dataset *manu_stata9.dta*, `xtmg` leads

[17] Ando and Bai (2015) derive a test for a multifactor error structure model, but they assume that the common factors are estimated. An estimation of the common factors is neither in Pesaran (2006) nor Chudik and Pesaran (2015b) considered and not supported by `xtdcce2`.

to the following mean group results:[18]

```
. use manu_stata9.dta

. xtset nwbcode year
       panel variable:  nwbcode (strongly balanced)
        time variable:  year, 1970 to 2002
                delta:  1 unit

. eststo xtmg95: qui xtmg ly lk, trend

. eststo xtmg06: qui xtmg ly lk, cce trend

. estout xtmg95 xtmg06   , c(b(star fmt(4)) se(fmt(4) par)) /*
> */ mlabels("xtmg - mg" "xtmg - cce" ) s(N cd cdp , fmt(0 3 3 )) /*
> */ drop(*_ly *_lk) rename(__000007_t trend) collabels(,none)

--------------------------------------------
                xtmg - mg      xtmg - cce
--------------------------------------------
lk                 0.1789*         0.3125***
                 (0.0805)        (0.0849)
trend              0.0174***       0.0108**
                 (0.0030)        (0.0035)
_cons              7.6528***       4.7860***
                 (0.8546)        (1.3227)
--------------------------------------------
N                    1194            1194
cd
cdp
--------------------------------------------

.
. eststo xtdcce95: qui xtdcce2 ly lk ,/*
> */ crosssectional(ly lk) trend nocrosssectional reportconstant

. eststo xtdcce06: qui xtdcce2 ly lk  , /*
> */ crosssectional(ly lk) cr_lags(0) trend reportconstant

. estout  xtdcce95  xtdcce06  , c(b(star fmt(4)) se(fmt(4) par)) /*
> */ mlabels("xtdcce2-mg" "xtdcce2-cce" ) s(N cd cdp , fmt(0 3 3 )) /*
> */ rename(__000007_t trend) collabels(,none)

--------------------------------------------
               xtdcce2-mg     xtdcce2-cce
--------------------------------------------
lk                 0.1789*         0.3125***
                 (0.0805)        (0.0849)
trend              0.0174***       0.0108**
                 (0.0030)        (0.0035)
_cons              7.6354***       4.7752***
                 (0.8531)        (1.3202)
--------------------------------------------
N                    1194            1194
cd                  6.686          -0.201
cdp                 0.000           0.841
--------------------------------------------
```

The first table shows the estimation results from Table 1, p. 67 in Eberhardt (2012). The lower table displays results on the same equation using `xtdcce2`. The first column shows in both tables a MG regression, the second column shows a

[18]The dataset *manu_stata9.dta* is taken from Eberhardt and Teal (2017) and is available at https://sites.google.com/site/medevecon/.

common correlated effects regression with contemporaneous cross-sectional means. The CD test statistic rejects the hypothesis of weak cross-sectional dependence in the case of the mean group regression. Including cross-sectional averages improves the statistic such that the hypothesis cannot be rejected any longer. Estimation results produced by `xtmg` and `xtdcce2` differ slightly, as seen here by the constant. The reason is, that `xtdcce2` ensures that all variables are stored as `doubles` to allow for best precision.[19]

3.6 Monte Carlo Simulation

In this section a Monte Carlo simulation is performed. The aims of this exercise are several fold. First of all the estimator is compared to a simple Fixed Effects (FE) regression. Secondly, the small time series sample adjustments are examined. Finally, the Monte Carlo will shed light on the size of the bias depending on the time and cross-sectional dimension. This can give guidance under which dimensions the DCCE estimator can be used. The Monte Carlo simulation is carried out on the lines of Chudik and Pesaran (2015a). The underlying model is:

$$y_{i,t} = c_{yi} + \phi_i y_{i,t-1} + \beta_{0i} x_{i,t} + \beta_{1i} x_{i,t-1} + u_{i,t} \tag{3.11}$$

$$u_{i,t} = \gamma_i' f_t + \epsilon_{i,t} \tag{3.12}$$

$$x_{i,t} = c_{xi} + \alpha_{xi} y_{i,t-1} + \gamma_{xi} f_t + v_{xi,t} \tag{3.13}$$

$$g_{i,t} = c_{gi} + \alpha_{gi} y_{i,t-1} + \gamma_{gi} f_t + v_{gi,t} \tag{3.14}$$

$y_{i,t}$ is the dependent variable and $x_{i,t}$ a vector of K independent variables. Without loss of generality, it is assumed that only one independent variable exists. $g_{i,t}$ is a set of covariates which are affected by the unobserved factors, but are not used to estimate $y_{i,t}$. The coefficient for the contemporaneous value of $x_{i,t}$ is drawn from a uniform distribution as $\beta_{0i} \sim IIDU(0.5, 1)$. The coefficient on the lagged value of the independent variable is set to $\beta_{1i} = -0.5$. For the lagged dependent variable two

[19] `xtdcce2` creates all variables as doubles. `xtmg` stores newly created variables as floats. Another difference versus `xtmg` is that `xtdcce2` supports time-series operators.

different scenarios are considered for the calculation of $y_{i,t}$ and $x_{i,t}$. One with low values for ϕ, $\phi_i \sim IIDU(0, 0.8)$ and $\alpha_{xi} \sim IIDU(0, 0.35)$ and one with high values $\phi_i \sim IIDU(0.5, 0.9)$ and $\alpha_{xi} \sim IIDU(0, 0.15)$. α_{gi} is in both scenarios the same: $\alpha_{gi} \sim IIDU(0, 1)$.[20] In comparison to Chudik and Pesaran (2015a), the number of common factors is restricted to one. As shown in their Monte Carlo simulation, the results are robust for a small number of common factors. The common factors f_t are potentially correlated over time ($\rho_f \neq 0$). The error $\epsilon_{i,t}$ is heteroskedastic and weakly cross-sectional dependent ($\alpha_{CSD} = 0.4$). Appendix B.3.1 describes detailed the data generating process for the Monte Carlo simulation.

xtdcce2 estimates the following equation:

$$y_{i,t} = c_{yi} + \phi_i y_{i,t-1} + \beta_{0i} x_{i,t} + \beta_{1i} x_{i,t-1} + \sum_{l=0}^{p_t} \delta'_{il} \bar{z}_{t-l} + e_{yi,t}. \tag{3.15}$$

The number of lags is set to the integer part of $T^{\frac{1}{3}}$. The cross-sectional averages contain y, x and g. Besides the unadjusted estimator, jackknife and RMA estimators are used.

Within each run of the Monte Carlo simulation, the following command line for xtdcce2 was used:[21]

```
. xtdcce2 y L.y x L.x , cr_lags(lags) cr(y x)
```

Results for four different specifications will be presented next, see Table 3.1. Tables 3.2 - 3.4 show Monte Carlo results for ϕ, β_0 and β_1 with $E(\phi_i) = 0.4$ and $\rho = 0.6$. For the following three tables (Tables 3.5 - 3.7), ρ is set to zero, implying that the common factors are $IIDN(0, 1)$ distributed and not correlated over time. For the next three tables (3.8 - 3.10) common factors are correlated again with $\rho_f = 0.6$ and $E(\phi_i) = 0.7$. For the fourth specification, ρ_f is set to 0 with high values of ϕ. For specification 5 the option pooled(L.x) is added to the command line, treating the coefficient β_1 as pooled. This result will show, if the MG estimate will

[20] ϕ_i and α_{xi} depend on each other to make sure that the series $y_{i,t}$ and $x_{i,t}$ are stationary. See Chudik and Pesaran (2015a), p. 399-400.

[21] The simulations were carried out with xtdcce2 version 1.33 and Stata 14. For the RMA and jackknife half-panel correction methods the options recursive and jackknife are used.

estimate β_1 with different bias than the pooled estimates. The last two specifications are presented in the Appendix B.3.2, with the exception of Table 3.11. The results are similar with respect to the biases from Specification 1 and Specification 2.

Specification	$E(\phi_i)$	ρ_f	Mean Group Variables	Pooled	Tables
1	0.3	0.6	`L.y x L.x`	-	Tables 3.2 - 3.4
2	0.3	0	`L.y x L.x`	-	Tables 3.5 - 3.7
3	0.7	0.6	`L.y x L.x`	-	Tables 3.8 - 3.10
4	0.7	0	`L.y x L.x`	-	Tables B.1 - B.3
5	0.3	0.6	`L.y x`	`L.x`	Tables B.4, B.5 and 3.11

Table 3.1: Specifications for Monte Carlo Simulations
The DGP is $y_{i,t} = c_{yi} + \phi_i y_{i,t-1} + \beta_{0i} x_{i,t} + \beta_{1i} x_{i,t-1} + \gamma_i' f_t + \epsilon_{i,t}$, where $\beta_{0i} \sim IIDU(0.5, 0.1)$, $\beta_{1i} = 0.5$, $c_{yi} \sim IIDN(0,1)$, $\gamma_i = \sqrt{1-\sigma_\gamma^2} + \eta_{i\gamma}$ with $\eta_{i\gamma} \sim IIDN(0, \sigma_\gamma^2)$ and $\sigma_\gamma^2 = 0.2^2$ and $f_t = \rho_f f_{t-1} + \varsigma_{ft}$ with $\varsigma_{ft} \sim IIDN(0, 1-\rho_f^2)$.

3.7 Monte Carlo Results

In a small, finite sample the bias of the DCCE estimator potentially arises from three sources: the length of the time series, cross-sectional dependence and heterogeneous slope coefficients. The first source relates to small T and the time series bias of order T^{-1} (Hurwicz bias) and is expected to decrease with $T \to \infty$. The heterogeneous coefficients are assumed to be randomly distributed around a common mean. As N converges to infinity, the mean group estimate converges to its true parameter, ignoring any influence from the other two biases. Therefore the bias due to heterogeneous coefficients is expected to decrease with an increase in N. The bias due to cross-sectional dependence needs to be separated further. For a given T and an increase in N, weak cross sectional dependence declines. Strong cross-sectional dependence does not decline when N increases, but should not pose a problem as the cross-sectional averages take it out.

Tables 3.2-3.4 present the Monte Carlo Simulation results for low values of $\phi = E(\phi_i) = 0.4$ and $\rho = 0.6$. As cross-sectional averages $\bar{y}_t, \bar{x}_t$ and $\bar{g}_t$ are added.

(N,T)	Bias (x100)					RMSE (x100)				
	40	50	100	150	200	40	50	100	150	200
Fixed Effects estimates										
40	5.21	10.20	16.51	19.98	19.76	7.21	7.60	8.36	9.16	8.94
50	4.07	8.82	15.38	18.56	19.84	6.38	6.67	7.68	8.53	8.91
100	5.23	8.91	15.70	18.34	19.54	5.13	5.81	7.22	8.01	8.37
150	5.60	9.71	16.09	18.43	20.06	4.91	5.62	7.08	7.82	8.44
200	5.16	8.66	16.55	18.08	19.16	4.50	4.98	7.23	7.68	8.01
DCCE without bias correction										
40	-42.85	-31.69	-13.52	-8.06	-5.54	18.14	13.53	6.26	4.11	3.02
50	-42.91	-30.28	-13.45	-8.25	-6.33	18.03	12.95	6.11	3.99	3.18
100	-43.33	-31.51	-13.66	-8.83	-6.17	17.86	12.96	5.85	3.90	2.78
150	-42.16	-31.11	-13.73	-8.74	-6.31	17.23	12.70	5.70	3.72	2.75
200	-43.65	-31.43	-13.69	-8.96	-6.19	17.75	12.80	5.67	3.73	2.65
DCCE with jackknife bias correction										
40	-39.57	-28.49	-10.88	-6.13	-3.98	17.01	12.46	5.42	3.56	2.61
50	-39.79	-26.83	-10.77	-6.25	-4.76	16.97	11.75	5.23	3.38	2.74
100	-40.13	-28.03	-11.03	-6.86	-4.65	16.73	11.68	4.90	3.21	2.28
150	-38.81	-27.48	-11.07	-6.76	-4.77	15.97	11.32	4.72	2.98	2.19
200	-40.52	-27.95	-11.05	-6.97	-4.64	16.56	11.48	4.66	2.98	2.09
DCCE with RMA bias correction										
40	-46.80	-34.56	-14.72	-8.80	-6.20	19.76	14.67	6.69	4.35	3.24
50	-46.80	-33.25	-14.91	-8.99	-6.94	19.60	14.13	6.65	4.24	3.38
100	-47.11	-34.36	-14.95	-9.48	-6.71	19.39	14.11	6.34	4.15	2.98
150	-46.10	-34.01	-14.87	-9.44	-6.90	18.80	13.84	6.15	3.99	2.97
200	-47.18	-34.41	-14.88	-9.69	-6.73	19.16	13.98	6.13	4.01	2.85

Table 3.2: Monte Carlo Results for Specification 1 and ϕ, with $\phi = E(\phi_i) = 0.4$. The DGP is $y_{i,t} = c_{yi} + \phi_i y_{i,t-1} + \beta_{0i} x_{i,t} + \beta_{1i} x_{i,t-1} + \gamma_i' f_t + \epsilon_{i,t}$, where $\beta_{0i} \sim IIDU(0.5, 0.1)$, $\beta_{1i} = 0.5$, $c_{yi} \sim IIDN(0,1)$, $\gamma_i = \sqrt{1-\sigma_\gamma^2} + \eta_{i\gamma}$ with $\eta_{i\gamma} \sim IIDN(0, \sigma_\gamma^2)$ and $\sigma_\gamma^2 = 0.2^2$, $f_t = \rho_f f_{t-1} + \varsigma_{ft}$ with $\varsigma_{ft} \sim IIDN(0, 1-\rho_f^2)$, $\rho_f = 0.6$. $\alpha_{CSD} = 0.4$. For a further description see Section 3.6.

The first panel of Table 3.2 shows results of the fixed effects estimator. The Fixed Effects (FE) estimator suffers from three sources of biases, the time series bias, a bias ignoring the cross-sectional dependence and the heterogeneous slope coefficients. The bias increases with T, holding the number of cross-sections fixed. With an increase in the number of cross-sections, the bias remains qualitatively similar. Surprising is the low bias for a small number of time periods. However, the bias climbs with an increase in T to a substantial level. A possible explanation is that the time series bias is offset by either or both of the others. Comparing the results to those with $\rho = 0$ will shed some light on this.

In the following panels results of the DCCE estimator, without and with small sample bias correction methods are presented. In comparison to the FE estimator,

(N,T)	Bias (x100)					RMSE (x100)				
	40	50	100	150	200	40	50	100	150	200
Fixed Effects estimates										
40	65.08	65.50	63.27	64.18	64.43	49.89	49.91	48.00	48.45	48.51
50	65.49	64.05	63.98	64.16	63.92	50.02	48.83	48.42	48.34	48.22
100	65.23	65.02	64.26	63.43	64.01	49.54	49.23	48.39	47.76	48.15
150	64.64	64.03	63.28	63.61	63.26	48.95	48.46	47.65	47.87	47.60
200	65.28	64.84	63.84	63.47	63.15	49.37	48.94	48.10	47.71	47.47
DCCE without bias correction										
40	4.34	3.37	2.05	1.30	0.67	12.47	9.21	5.62	4.40	3.66
50	4.78	3.10	2.13	1.26	1.24	11.10	8.65	5.32	4.25	3.77
100	4.27	3.71	2.17	1.15	1.18	8.54	6.55	4.10	3.11	2.62
150	3.66	3.96	2.07	1.30	1.02	7.04	5.94	3.47	2.56	2.15
200	5.21	4.07	2.34	1.39	0.96	6.71	5.17	3.18	2.37	1.89
DCCE with jackknife bias correction										
40	3.77	3.08	1.60	0.95	0.40	12.59	9.58	5.62	4.41	3.67
50	4.10	2.51	1.67	0.89	0.99	11.41	8.91	5.33	4.23	3.73
100	3.74	3.23	1.68	0.79	0.91	8.80	6.65	4.03	3.09	2.60
150	3.19	3.21	1.62	0.92	0.74	7.16	5.84	3.40	2.50	2.11
200	4.71	3.44	1.86	1.01	0.65	6.61	5.11	3.05	2.28	1.83
DCCE with RMA bias correction										
40	5.22	4.79	2.90	1.84	1.21	12.89	9.66	5.76	4.51	3.76
50	6.24	4.29	3.04	1.81	1.73	11.78	9.04	5.56	4.40	3.88
100	5.14	4.94	3.05	1.69	1.67	9.04	7.21	4.42	3.27	2.77
150	4.87	5.22	2.85	1.89	1.51	7.73	6.56	3.77	2.79	2.32
200	6.04	5.38	3.16	1.95	1.46	7.25	5.83	3.56	2.59	2.08

Table 3.3: Monte Carlo Results for Specification 1 and β_0, with $\phi = E(\phi_i) = 0.4$ and $\rho_f = 0.6$. See notes Table 3.2.

the bias decreases considerably with T. The increase due to a larger number of cross-sectional units is not that pronounced. The bias is more than halved if the number of time periods is increased from 50 to 100. One implication is that for low levels of T, the main source of the bias is the time series bias. As the bias for the largest value of N and T is around -6% and with a larger number of cross-sectional units the coefficient on the lagged dependent variable converges to its true value, the remaining bias is due to the cross-sectional dependence. Similar as the bias, the Root Mean Squared Error (RMSE) decreases with an increase in the number of time periods. For a similar bias, the RMSE of the DCCE estimator is much smaller than for the FE estimator, implying a smaller variation of the bias of the DCCE estimator.

In favour of this interpretation is the fact that the bias of the two small sample corrected estimators are in a similar region for large values of T. The jackknife bias

(N,T)	Bias (x100)					RMSE (x100)				
	40	50	100	150	200	40	50	100	150	200
Fixed Effects estimates										
40	5.90	10.02	12.43	15.37	15.26	10.64	10.77	9.54	9.98	9.65
50	5.29	7.69	12.53	13.99	14.72	10.06	9.49	9.01	9.17	9.08
100	5.90	8.94	12.53	13.23	13.76	7.31	7.72	7.84	7.91	8.06
150	5.86	8.45	11.44	12.84	14.04	6.59	6.82	7.00	7.46	7.79
200	6.90	8.60	11.59	12.85	13.60	6.48	6.33	6.77	7.21	7.45
DCCE without bias correction										
40	-8.61	-6.20	-1.68	-0.91	-0.91	12.07	10.20	6.02	4.31	3.97
50	-6.82	-5.27	-1.83	-1.02	-1.24	11.54	9.25	5.55	4.06	3.67
100	-5.12	-3.04	-1.07	-1.10	-0.16	8.14	6.41	3.75	3.15	2.66
150	-6.78	-2.76	-0.45	-0.39	-0.29	7.50	5.88	3.32	2.65	2.15
200	-5.13	-2.88	-0.44	-0.68	-0.21	6.63	5.07	2.79	2.31	1.85
DCCE with jackknife bias correction										
40	-8.43	-6.41	-1.67	-0.89	-0.93	12.59	10.62	6.15	4.39	4.01
50	-6.96	-5.55	-1.87	-1.02	-1.32	11.81	9.55	5.62	4.07	3.72
100	-5.89	-3.23	-1.04	-1.17	-0.20	8.56	6.79	3.81	3.20	2.66
150	-7.21	-2.92	-0.50	-0.50	-0.36	7.95	6.03	3.38	2.70	2.16
200	-5.75	-3.01	-0.56	-0.79	-0.32	6.96	5.21	2.82	2.32	1.87
DCCE with RMA bias correction										
40	-8.24	-5.76	-0.73	-0.30	-0.47	12.50	10.21	5.94	4.34	3.96
50	-6.56	-4.48	-1.03	-0.44	-0.69	11.64	9.35	5.51	4.02	3.61
100	-3.79	-1.72	-0.16	-0.49	0.39	8.11	6.42	3.71	3.12	2.68
150	-5.97	-1.46	0.64	0.33	0.23	7.44	5.95	3.36	2.65	2.16
200	-4.42	-1.37	0.44	-0.07	0.35	6.71	5.03	2.82	2.30	1.87

Table 3.4: Monte Carlo Results for Specification 1 and β_1, with $\phi = E(\phi_i) = 0.4$ and $\rho_f = 0.6$. See notes Table 3.2.

correction method has only little impact and decreases the bias from -42.85% to -39.57% for $N = 40, T = 40$. The bias for the estimation results with the RMA bias correction method increases, implying that the method does not lead to the desired results of a lower bias.

There are three notable differences to the study from Chudik and Pesaran (2015a), Table 2. First of all the FE estimator seems to perform better in samples with a small number of time periods. Secondly the bias of the DCCE estimator remains on a higher level than in Chudik and Pesaran (2015b). The final difference is the performance of the correction methods. While the jackknife bias correction method leads to an improvement in the precision of the estimates, the RMA method does not. Especially the latter result is a stark difference to the results from Chudik and Pesaran (2015a).

The next two tables show the simulation results for β_0 (Table 3.3) and β_1 (Table 3.4). The FE estimator has a larger bias for β_0 and is in the region of 65%. However, the bias does not decrease with neither $N \to \infty$ or $T \to \infty$. The bias for the DCCE estimates of β_0 decrease with T, as it does for ϕ. The level of the bias is smaller by a magnitude. In comparison to the RMSE of ϕ, the fall of the RMSE for β_0 is more pronounced with $N \to \infty$. The jackknife bias correction method again performs better than estimation methods without correction methods.

The differences between Table 3 in Chudik and Pesaran (2015a) and Table 3.3 have the same pattern as the one mentioned before. The biases for all estimators and specifications are larger and the RMA correction method does not lead to a decrease of the bias.

(N,T)	Bias (x100)					RMSE (x100)				
	40	50	100	150	200	40	50	100	150	200
Fixed Effects estimates										
40	-12.57	-7.98	-2.06	0.76	0.87	7.99	6.82	4.80	4.18	3.80
50	-13.33	-8.57	-2.99	-0.34	0.94	7.72	6.25	4.28	3.81	3.76
100	-11.68	-8.70	-2.41	-0.03	0.69	6.15	5.25	3.21	2.76	2.67
150	-11.10	-7.49	-2.20	0.18	1.24	5.73	4.39	2.64	2.20	2.22
200	-11.51	-8.30	-1.69	-0.22	0.84	5.60	4.44	2.34	2.04	1.94
DCCE without bias correction										
40	-36.47	-27.06	-11.58	-6.87	-4.73	15.80	11.78	5.57	3.74	2.80
50	-36.40	-25.52	-11.54	-7.02	-5.47	15.55	11.15	5.44	3.60	2.91
100	-36.79	-26.89	-11.71	-7.60	-5.30	15.31	11.18	5.12	3.47	2.48
150	-35.59	-26.51	-11.70	-7.50	-5.43	14.68	10.90	4.94	3.25	2.43
200	-36.86	-26.68	-11.72	-7.69	-5.33	15.08	10.95	4.90	3.24	2.33
DCCE with jackknife bias correction										
40	-32.74	-23.65	-8.89	-4.91	-3.15	14.60	10.66	4.77	3.23	2.43
50	-32.96	-21.86	-8.78	-4.98	-3.88	14.42	9.91	4.60	3.03	2.50
100	-33.25	-23.24	-8.99	-5.60	-3.77	14.09	9.87	4.18	2.80	2.00
150	-32.03	-22.64	-8.99	-5.49	-3.86	13.35	9.45	3.96	2.52	1.89
200	-33.40	-22.98	-8.99	-5.67	-3.77	13.79	9.56	3.88	2.51	1.79
DCCE with RMA bias correction										
40	-40.07	-29.88	-12.70	-7.59	-5.37	17.28	12.88	5.97	3.96	3.00
50	-40.02	-28.45	-13.00	-7.74	-6.07	16.99	12.29	5.97	3.83	3.10
100	-40.51	-29.50	-13.00	-8.25	-5.83	16.81	12.24	5.61	3.71	2.67
150	-39.34	-29.26	-12.80	-8.18	-5.98	16.14	11.99	5.35	3.51	2.63
200	-40.22	-29.47	-12.85	-8.39	-5.85	16.42	12.05	5.34	3.51	2.52

Table 3.5: Monte Carlo Results for Specification 2 and ϕ, with $\phi = E(\phi_i) = 0.4$ and $\rho_f = 0$. See notes Table 3.2.

The bias of the FE estimates of β_1 in Table 3.4 are smaller than the ones from

(N,T)	Bias (x100)					RMSE (x100)				
	40	50	100	150	200	40	50	100	150	200
Fixed Effects estimates										
40	77.41	77.96	76.08	77.40	77.70	58.80	58.99	57.45	58.24	58.37
50	77.41	76.65	76.84	77.63	77.12	58.71	58.05	57.96	58.35	58.07
100	77.40	77.58	77.61	76.82	77.39	58.50	58.54	58.30	57.75	58.14
150	76.85	76.52	76.71	76.92	76.97	57.96	57.74	57.67	57.80	57.85
200	77.33	77.52	77.08	76.91	76.75	58.30	58.36	57.98	57.74	57.63
DCCE without bias correction										
40	4.02	3.02	1.87	1.28	0.64	12.39	9.13	5.54	4.40	3.65
50	4.71	2.84	2.07	1.18	1.20	10.75	8.53	5.32	4.25	3.79
100	4.13	3.59	2.10	1.09	1.13	8.39	6.56	4.09	3.09	2.62
150	3.47	3.90	1.90	1.22	0.96	6.91	5.97	3.41	2.53	2.14
200	4.88	3.93	2.20	1.30	0.92	6.53	5.09	3.12	2.32	1.89
DCCE with jackknife bias correction										
40	3.40	2.71	1.40	0.91	0.36	12.53	9.42	5.56	4.40	3.66
50	3.99	2.14	1.55	0.82	0.94	11.13	8.83	5.33	4.25	3.76
100	3.63	3.10	1.57	0.71	0.86	8.61	6.68	4.01	3.08	2.60
150	2.94	3.15	1.43	0.82	0.68	7.02	5.84	3.35	2.47	2.10
200	4.33	3.31	1.70	0.91	0.60	6.46	5.02	3.00	2.24	1.84
DCCE with RMA bias correction										
40	5.05	4.42	2.65	1.77	1.14	12.68	9.60	5.67	4.49	3.73
50	6.06	4.07	2.95	1.69	1.64	11.44	8.93	5.53	4.39	3.88
100	5.14	4.79	2.93	1.58	1.58	8.91	7.17	4.40	3.25	2.75
150	4.84	5.13	2.63	1.75	1.41	7.56	6.55	3.67	2.72	2.28
200	5.83	5.13	2.96	1.82	1.37	7.08	5.70	3.47	2.53	2.05

Table 3.6: Monte Carlo Results for Specification 2 and β_0, with $\phi = E(\phi_i) = 0.4$ and $\rho_f = 0$. See notes Table 3.2.

the DCCE estimator again. Similar to the obtained results so far, the bias is more pronounced with an increase in T.[22] The DCCE estimates are again improving with an increase in the number of time periods. The bias correction methods seem not to be important for the regression estimates.

In the following (Tables 3.5 to 3.7), results for ϕ with $E(\phi_i) = 0.4$ and no autocorrelated common factors $\rho_f = 0$ are discussed. The bias for the FE estimator in Table 3.5 turns negative, so the estimates are pushed upwards due to the autocorrelation in the common factors. The bias for the DCCE specifications is only slightly smaller in comparison to the results with serially correlated factors. As the bias only changes little when the number of cross-sections is increased, but it deceases

[22]Chudik and Pesaran (2015a) do not show any simulation results for β_1. Neither in the published paper, nor in an online appendix or working paper versions.

(N,T)	Bias (x100)					RMSE (x100)				
	40	50	100	150	200	40	50	100	150	200
Fixed Effects estimates										
40	25.60	30.50	33.96	36.93	37.07	16.17	17.83	18.29	19.43	19.35
50	25.42	28.79	33.93	35.93	36.42	15.86	16.63	17.94	18.84	18.90
100	25.53	29.61	34.34	35.42	35.96	14.28	15.95	17.78	18.18	18.41
150	26.03	29.28	33.83	35.12	36.48	14.06	15.51	17.36	17.93	18.52
200	26.61	29.68	33.66	35.17	36.36	14.26	15.52	17.14	17.86	18.41
DCCE without bias correction										
40	-6.15	-4.76	-0.99	-0.53	-0.64	11.64	9.81	5.99	4.28	3.96
50	-4.57	-3.72	-1.28	-0.63	-0.94	11.13	9.05	5.50	4.06	3.66
100	-3.07	-1.73	-0.43	-0.69	0.13	7.88	6.27	3.73	3.13	2.66
150	-4.99	-1.35	0.09	-0.02	-0.01	6.99	5.74	3.33	2.62	2.13
200	-3.22	-1.56	0.13	-0.30	0.07	6.26	4.91	2.78	2.28	1.84
DCCE with jackknife bias correction										
40	-5.80	-4.96	-1.01	-0.48	-0.64	12.14	10.28	6.13	4.36	4.00
50	-4.55	-3.94	-1.31	-0.63	-1.02	11.50	9.39	5.58	4.07	3.71
100	-3.51	-1.86	-0.39	-0.76	0.09	8.30	6.59	3.80	3.19	2.66
150	-5.25	-1.42	0.04	-0.12	-0.07	7.31	5.88	3.39	2.66	2.14
200	-3.71	-1.57	0.02	-0.40	-0.03	6.56	5.04	2.82	2.29	1.85
DCCE with RMA bias correction										
40	-5.78	-4.38	-0.07	0.07	-0.23	12.26	9.91	5.96	4.32	3.96
50	-4.03	-2.88	-0.51	-0.07	-0.44	11.14	9.17	5.48	4.05	3.61
100	-1.87	-0.27	0.42	-0.12	0.64	7.91	6.35	3.71	3.13	2.70
150	-4.07	-0.04	1.13	0.65	0.47	6.97	5.81	3.40	2.64	2.14
200	-2.35	0.00	0.97	0.29	0.59	6.36	4.96	2.85	2.28	1.88

Table 3.7: Monte Carlo Results for Specification 2 and β_1, with $\phi = E(\phi_i) = 0.4$ and $\rho_f = 0$. See notes Table 3.2.

with the number of time periods, the small time series bias plays a crucial role. Theoretically, β_0 and β_1 are expected to suffer less from the small sample time series bias than ϕ. This result is confirmed again in Tables 3.6 and 3.7. It is interesting to note, that the DCCE estimator is able to identify the mean group estimates for the heterogeneous coefficient β_0 and for the homogeneous coefficient β_1 well with a low bias.

As a penultimate exercise in Table 3.8 to 3.10, high values for the autocorrelation coefficient ($E(\phi_i) = 0.7$) are considered. The bias for the FE estimates is remarkably low, but again increases with T. The DCCE estimates are smaller than with $E(\phi_i) =$ 0.4, but still extensive and decline with T. This shows again the robust behaviour of the DCCE estimator with respect to changes of the correlation over time and

(N,T)	Bias (x100)					RMSE (x100)				
	40	50	100	150	200	40	50	100	150	200
Fixed Effects estimates										
40	0.40	3.01	6.51	8.14	8.60	4.60	4.62	5.54	6.31	6.49
50	0.10	2.47	6.26	7.85	8.45	4.28	4.17	5.28	6.03	6.33
100	0.51	2.61	6.36	7.81	8.52	3.43	3.68	5.06	5.83	6.24
150	0.65	2.59	6.68	7.90	8.71	3.21	3.35	5.13	5.79	6.31
200	0.48	2.47	6.84	7.70	8.36	3.04	3.18	5.20	5.68	6.06
DCCE without bias correction										
40	-30.21	-22.13	-9.37	-5.89	-3.97	21.88	16.15	7.04	4.62	3.22
50	-31.07	-22.03	-9.52	-5.94	-4.37	22.42	16.00	7.08	4.54	3.43
100	-31.53	-22.82	-9.85	-6.23	-4.39	22.50	16.27	7.12	4.57	3.25
150	-30.87	-22.54	-9.84	-6.28	-4.52	21.94	15.98	7.02	4.52	3.28
200	-31.53	-22.61	-9.92	-6.37	-4.42	22.36	16.02	7.06	4.55	3.18
DCCE with jackknife bias correction										
40	-28.26	-19.98	-7.61	-4.57	-2.88	20.59	14.77	5.92	3.83	2.61
50	-29.04	-19.85	-7.70	-4.59	-3.27	21.12	14.59	5.91	3.70	2.78
100	-29.50	-20.58	-8.09	-4.86	-3.31	21.17	14.77	5.95	3.66	2.55
150	-28.69	-20.24	-8.06	-4.90	-3.42	20.47	14.41	5.81	3.59	2.55
200	-29.53	-20.40	-8.13	-4.98	-3.33	21.01	14.51	5.84	3.60	2.46
DCCE with RMA bias correction										
40	-35.33	-26.15	-11.15	-7.01	-4.85	25.49	18.95	8.24	5.35	3.80
50	-35.57	-25.86	-11.43	-6.98	-5.22	25.59	18.68	8.39	5.23	3.98
100	-36.29	-26.56	-11.55	-7.15	-5.17	25.85	18.88	8.29	5.19	3.79
150	-35.51	-26.27	-11.50	-7.26	-5.31	25.18	18.57	8.17	5.20	3.83
200	-36.13	-26.51	-11.57	-7.34	-5.18	25.57	18.75	8.21	5.22	3.71

Table 3.8: Monte Carlo Results for Specification 3 and ϕ, with $\phi = E(\phi_i) = 0.7$ and $\rho_f = 0.6$. For a further description see Section 3.6.

underlines its exposure to the small sample time series bias.

Lastly, Table 3.11 contains the results with the pooled coefficient ($\beta_{1,i} = \beta_1 \forall\ i$) on the lagged explanatory variable. In order to save space, the results for ϕ and β_0 can be found in the Appendix in Tables B.4 and B.5. The biases for both coefficients are in a similar order, however slightly smaller. In comparison to the results from Table 3.4, the MG estimator does a good job. The bias of the pooled estimation is smaller, but not by a large magnitude. It is interesting to note that the bias for the estimation without bias correction turns positive. Therefore, the RMSE for the large panels is larger than the one for the MG estimation.

Overall, the Monte Carlo simulations show the following: first for the common correlated effects estimators and the number of time periods decreases the bias more substantially than the number of cross-sectional units. This implies that the time

(N,T)	Bias (x100)					RMSE (x100)				
	40	50	100	150	200	40	50	100	150	200
Fixed Effects estimates										
40	72.05	73.21	71.75	72.50	72.87	55.07	55.63	54.29	54.60	54.76
50	72.48	71.18	72.19	72.72	72.35	55.25	54.15	54.54	54.69	54.50
100	72.44	72.66	72.67	72.28	72.73	54.97	54.98	54.67	54.38	54.67
150	72.02	71.77	71.83	72.41	72.42	54.50	54.29	54.06	54.45	54.46
200	72.61	72.68	72.41	72.27	72.08	54.87	54.83	54.53	54.29	54.15
DCCE without bias correction										
40	1.03	1.07	1.13	0.72	0.37	11.82	8.84	5.55	4.36	3.69
50	1.56	0.86	1.23	0.62	0.80	10.56	8.60	5.21	4.17	3.71
100	1.41	1.43	1.07	0.48	0.73	8.00	6.11	3.93	3.01	2.55
150	0.73	1.50	1.05	0.63	0.50	6.53	5.30	3.23	2.45	2.05
200	2.16	1.80	1.26	0.69	0.49	5.78	4.44	2.83	2.23	1.81
DCCE with jackknife bias correction										
40	0.98	1.24	0.95	0.57	0.23	12.21	9.40	5.61	4.41	3.73
50	1.28	0.80	1.10	0.47	0.71	11.09	8.96	5.31	4.21	3.71
100	1.47	1.42	0.90	0.34	0.61	8.38	6.43	3.98	3.05	2.58
150	0.75	1.32	0.93	0.49	0.40	6.94	5.47	3.26	2.46	2.06
200	2.23	1.66	1.10	0.54	0.34	6.00	4.59	2.86	2.23	1.81
DCCE with RMA bias correction										
40	2.76	3.84	3.24	2.19	1.79	12.17	9.54	6.00	4.59	3.98
50	3.70	3.10	3.32	2.09	2.11	11.35	9.09	5.80	4.53	4.01
100	2.93	3.80	3.03	1.82	2.01	8.59	6.89	4.50	3.34	2.94
150	2.61	3.78	2.89	2.02	1.79	7.01	6.03	3.85	2.92	2.46
200	3.78	4.19	3.12	2.01	1.72	6.35	5.34	3.61	2.69	2.22

Table 3.9: Monte Carlo Results for Specification 3 and β_0, with $\phi = E(\phi_i) = 0.7$ and $\rho_f = 0.6$. See notes Table 3.2.

series bias is the main driver of the overall bias. Therefore the simulation provided here implies it is reasonable to employ the estimator with $T = 100$. Secondly, the bias on the lagged dependent variable appears to be more pronounced than the coefficients on exogenous explanatory variables. The last result is in line with Chudik and Pesaran (2015a). Similar to Chudik and Pesaran (2015a), bias correction methods have only a small impact and seem not to matter a lot. As Specification 5 showed, the MG estimator identifies a pooled variable equally well.

3.8 Conclusion

The package `xtdcce2` for Stata introduces new routines to estimate a heterogeneous panel model using dynamic common correlated effects in a large N and T panel. It combines estimation procedures proposed in Pesaran and Smith (1995) and Shin et al. (1999) with those in Pesaran (2006) and Chudik and Pesaran (2015a). It allows

(N,T)	Bias (x100)					RMSE (x100)				
	40	50	100	150	200	40	50	100	150	200
Fixed Effects estimates										
40	12.05	18.48	24.84	28.80	30.60	12.58	13.60	14.44	15.77	16.39
50	11.34	15.99	25.48	28.77	29.62	11.88	12.24	14.30	15.55	15.65
100	12.82	18.22	25.96	28.50	29.73	9.95	11.52	13.91	14.87	15.35
150	12.83	17.03	25.71	28.40	30.50	8.87	10.32	13.57	14.69	15.62
200	13.83	18.08	25.95	28.29	29.88	9.08	10.38	13.49	14.52	15.24
DCCE without bias correction										
40	-27.50	-19.98	-7.54	-4.69	-3.15	18.37	14.25	7.19	5.22	4.45
50	-26.79	-19.25	-7.86	-4.81	-3.90	17.69	13.23	6.94	5.01	4.16
100	-25.57	-17.75	-7.79	-5.28	-3.15	15.14	11.08	5.59	4.18	3.18
150	-26.03	-17.74	-7.12	-4.60	-3.48	14.83	10.56	4.95	3.59	2.88
200	-25.23	-17.56	-7.25	-5.08	-3.21	14.13	10.12	4.56	3.53	2.51
DCCE with jackknife bias correction										
40	-25.60	-18.25	-6.26	-3.70	-2.39	18.08	14.01	6.99	5.15	4.37
50	-25.22	-17.76	-6.39	-3.78	-3.16	17.44	12.97	6.56	4.80	4.03
100	-24.24	-15.96	-6.43	-4.27	-2.38	14.80	10.72	5.22	3.92	3.01
150	-24.25	-15.94	-5.77	-3.59	-2.70	14.22	9.94	4.50	3.31	2.67
200	-23.88	-15.83	-5.98	-4.09	-2.48	13.67	9.43	4.09	3.20	2.30
DCCE with RMA bias correction										
40	-27.31	-18.79	-5.33	-3.17	-1.73	18.60	14.06	6.76	5.00	4.28
50	-26.47	-17.58	-6.07	-3.23	-2.42	17.71	12.94	6.56	4.70	3.87
100	-24.62	-15.62	-5.72	-3.74	-1.58	14.87	10.38	4.99	3.79	2.92
150	-25.63	-16.07	-4.86	-2.92	-2.02	14.74	10.05	4.32	3.17	2.58
200	-24.99	-15.52	-5.36	-3.57	-1.80	14.17	9.43	3.96	3.08	2.17

Table 3.10: Monte Carlo Results for Specification 3 and β_1, with $\phi = E(\phi_i) = 0.7$ and $\rho_f = 0.6$. See notes Table 3.2.

coefficients to be pooled or estimated as mean groups. Furthermore, it supports unbalanced panels, estimation of instrumental variables, small sample time series bias corrections and a test for cross-sectional dependence, using the included `xtcd2` routine. An empirical example estimating a growth regression is given. Monte Carlo simulation results show that the DCCE estimator is robust to changes in the autocorrelation coefficients of the common factors and the dependent variable. The main driver of the bias for the coefficient of the dependent variable, ϕ, is the small sample time series bias. The coefficients of further explanatory variables are well estimated with a low bias. The results imply, that the DCCE estimator can be comfortably used in a setting with 100 time periods. If the number of time periods is smaller, especially the coefficient on the lagged dependent variable is exposed to a bias. In addition, the simulation showed, that the bias due to the falsely assumption of heterogeneity is negligible.

(N,T)	Bias (x100)					RMSE (x100)				
	40	50	100	150	200	40	50	100	150	200
Fixed Effects estimates										
40	5.90	10.02	12.43	15.37	15.26	10.64	10.77	9.54	9.98	9.65
50	5.29	7.69	12.53	13.99	14.72	10.06	9.49	9.01	9.17	9.08
100	5.90	8.94	12.53	13.23	13.76	7.31	7.72	7.84	7.91	8.06
150	5.86	8.45	11.44	12.84	14.04	6.59	6.82	7.00	7.46	7.79
200	6.90	8.60	11.59	12.85	13.60	6.48	6.33	6.77	7.21	7.45
DCCE without bias correction										
40	-4.31	-3.37	-0.93	-0.30	-0.70	11.89	10.43	6.16	4.54	4.17
50	-2.98	-2.89	-1.02	-0.13	-0.84	11.15	9.24	5.63	4.31	3.85
100	-1.06	0.21	0.03	-0.55	0.44	8.17	6.48	3.84	3.23	2.78
150	-2.44	0.64	0.77	0.43	0.21	7.08	5.76	3.54	2.72	2.28
200	-1.05	-0.12	0.67	0.02	0.46	6.34	5.11	2.83	2.34	1.94
DCCE with jackknife bias correction										
40	-5.53	-4.45	-1.51	-0.70	-1.02	12.21	10.64	6.26	4.57	4.20
50	-4.22	-4.05	-1.70	-0.54	-1.21	11.32	9.49	5.72	4.33	3.89
100	-2.81	-1.10	-0.66	-1.10	0.04	8.34	6.70	3.85	3.30	2.77
150	-3.98	-0.68	0.00	-0.17	-0.26	7.39	5.84	3.53	2.71	2.29
200	-2.94	-1.36	-0.13	-0.59	-0.03	6.58	5.21	2.84	2.36	1.94
DCCE with RMA bias correction										
40	-3.38	-2.45	0.09	0.41	-0.19	12.10	10.38	6.09	4.56	4.17
50	-1.94	-1.60	-0.13	0.49	-0.24	11.47	9.44	5.63	4.31	3.80
100	0.61	1.89	1.07	0.15	1.06	8.20	6.67	3.89	3.23	2.84
150	-1.12	2.28	1.97	1.22	0.78	7.09	6.04	3.67	2.78	2.32
200	0.28	1.80	1.69	0.70	1.07	6.47	5.33	2.96	2.37	2.01

Table 3.11: Monte Carlo Results for Specification 5 and β_1, with $\phi = E(\phi_i) = 0.3$ and $\rho_f = 0.6$. The coefficient on the lagged independent variable `L.x` is pooled. See notes Table 3.2.

`xtdcce2` can be further developed in many ways. First of all, there is room for speed improvements. In particular, pooled and IV estimations can be time consuming. Features such as alternative variance and covariance estimators or bias corrections methods as proposed in in Everaert and De Groote (2016) and Everaert and De Vos (2016) can be easily implemented in the current framework. Another feature can be the estimation of the average long run effects described in Chudik et al. (2016).

Within the literature of cross-sectional dependence, several estimation procedures are unavailable in Stata. The first is the two-stage estimation procedure in Bailey, Holly and Pesaran (2016) (this estimation strategy is summarised in the next chapter). Econometric theory discusses the estimation of factor loadings. Factor loadings are important for the estimation of stochastic frontier models (Filippini and Tosetti, 2014) or gravity models (Serlenga, Shin, Gunnella and Mastromarco, 2013). It might be no coincidence that none of these papers is not published yet and point towards the gap in the literature. Estimating the common factors is a challenging for several reasons. First of all, it is necessary to establish the number of factors. Sarafidis and Wansbeek (2012) developed a promising approach, but it has never been applied to the best of my knowledge. If the number of common factors is known, the next question that arises is, how to approximate the common factors. Are cross-sectional averages sufficient and how to deal if the number of factors is not equal to the number of cross-sectional averages. The literature on overidentification or weak instruments can be a good starting point. Finally, econometric theory and so is the application in Stata, is missing a test for slope homogeneity in large dynamic panels with common factors. Developing such a test and implementing it in Stata would give researchers more confidence selecting between pooled and mean group estimations.

Appendix Chapter 3

B.1 The xtdcce2 command

B.1.1 Syntax

`xtdcce2` *depvar* [*indepvars*] [(*varlist2* = *varlist_iv*)] [*if*] ,

`crosssectional(`*varlist_cr*`)` [`pooled(`*varlist_p*`)` `nocrosssectional` `cr_lags(`#`)`

`ivreg2options(`*string*`)` `e_ivreg2` `ivslow` `noisily` `lr(`*varlist_lr*`)` `lr_options(`*string*`)`

`noconstant` `pooledconstant` `reportconstant` `trend` `pooledtrend` `jackknife`

`recursive` `nocd` `showindividual` `fullsample`]

Data has to be [TS] **xtset** before using `xtdcce2`. *depvar*, *indepvars*, *varlist2*, *varlist_iv*, *varlist_cr*, *varlist_p* and *varlist_lr* may contain time-series operators, see [TS] **tsvarlist**, and factor variables, see [U] **11.4.3 Factor variables**. `xtdcce2` requires the `moremata` package by Jann (2005). *varlist2* are the endogenous variables and *varlist_iv* are the instruments.

B.1.2 Version

The very first version on SSC was called `xtdcce`. As of November 2017 `xtdcce2` is available on SSC archive as version 1.32. The results in this chapter, including the Monte Carlo simulation, were carried out using a beta of version 1.33. The version is currently under work (November 2017) and continuously updated. An overview of the current working version is available from within Stata using `net install`:

```
. net from http://www.ditzen.net/Stata/xtdcce2_beta

. net install xtdcce2133 , from(http://www.ditzen.net/Stata/xtdcce2_beta)
```

In Stata this version can be called by the command line `xtdcce2133`. Alternatively a wrapper for all beta versions is available from the above shown command lines. Using the beta versions requires the installation of `xtcd2` separately, either from SSC, or from:

```
. net install xtcd2 , from(http://www.ditzen.net/Stata/xtdcce2_beta)
```

B.1.3 Options

`crosssectional(`*varlist_cr*`)` defines the variables which are included in $\mathbf{z}_t$ and added as lagged cross-sectional averages ($\bar{\mathbf{z}}_{t-l}$) to the equation. The coefficients of the lagged cross-sectional averages are treated as nuisance parameters, which have no interpretation and are therefore partialled out.
`crosssectional(_all)` adds the levels of the variables from *depvar*, *indepvars*, *varlist2*, *varlist_iv* and *varlist_lr* as cross-sectional averages. No cross-sectional averages are added if `crosssectional(_none)` is used, which is equivalent to `nocrosssectional`.
`crosssectional()` is a required option but can be substituted by `nocrosssectional`. Variables in `crosssectional()` may be included in `pooled()`, `exogenous_vars()`, `endogenous_vars()` and `lr()`.

`pooled(`*varlist_p*`)` specifies homogenous coefficients. For these variables the estimated coefficients are constrained to be equal across all units ($\beta_i = \beta$ for $i = 1, ..., N$). Variable may occur in *indepvars*, *varlist2*, *varlist_iv*, *varlist_cr*, and *varlist_lr*.

`cr_lags(`#`)` specifies the number of lagged cross-sectional averages. For example, `cr_lags(2)` includes the contemporaneous cross-sectional averages and the first and second lag of the cross-sectional averages. If not defined, but `crosssectional()`

contains *varlist_cr* or `cr_lags(0)` , then only contemporaneous cross-sectional averages are added, but no lags.

`nocrosssectional` prevents adding cross-sectional averages. Results will be equivalent to the Pesaran and Smith (1995) Mean Group estimator, or if `lr(`*varlist*`)` specified to the Shin et al. (1999) Pooled Mean Group estimator.

`xtdcce2` allows IV regression. *varlist_2* specifies the endogenous variables and *varlist_iv* exogenous variables from IV regression using `ivreg2` by Baum et al. (2003, 2007). The use of *varlist_iv* and *varlist_2* require the prior installation of `ivreg2`.

`ivreg2options(`*string*`)` passes further options on to `ivreg2`. See `ivreg2` for more information.

`e_ivreg2` posts all available results from `ivreg2` in `e()` with prefix *ivreg2_*.

`noisily` shows the output of wrapped `ivreg2` regression command.

`ivslow` requests to use `ivreg2` for the calculation of auxiliary regressions rather than a faster `mata` routine. For the calculation of standard errors for pooled coefficients an auxiliary regression is performed. In this regression all coefficients are heterogeneous. If option `ivslow` is used , then `xtdcce2` calls `ivreg2` for the auxiliary regression. This is advisable as soon as `ivreg2` specific options are used, which influence point estimates.

`lr(`*varlist*`)`: Variables to be included in the long-run cointegration vector in addition to the error-correcting speed of adjustment term. Using the notation from Eq. (3.4) with the error correction term as $(y_{i,t-1} - \theta_i x_{i,t})$ the option would read: `lr(L.y x)`.

`lr_options(`*string*`)` Options for the long run coefficients. Options may be:

`nodivide`, coefficients are not divided by the error correction speed of adjustment vector (i.e. estimate equation 3.10).

`xtpmgnames`, coefficients names in `e(b)` and `e(V)` match the name convention from `xtpmg`.

`noconstant` suppress the constant term.

`pooledconstant` restricts the constant to be the same across all groups ($\beta_{0,i} = \beta_0,\ i = 1, ..., N$).

`reportconstant` reports the constant. If not specified, the constant is treated as a part of the cross-sectional averages and partialled out.

`trend` adds a linear unit-specific trend t_i. It cannot be combined with `pooledtrend`.

`pooledtrend` adds linear common trend. It cannot be combined with `trend`.

`jackknife` applies the 'half-panel' jackknife bias correction for small sample time series bias. It cannot be combined with `recursive`.

`recursive` applies the recursive mean adjustment method to correct for small sample time series bias. It cannot be combined with `jackknife`.

`nocd` suppresses calculation of CD test statistic.

`showindividual` reports cross-sectional unit-specific estimates in output.

`fullsample` uses the entire sample available for calculation of cross-sectional averages. Any observations which are lost due to lags will be included calculating the cross-sectional averages, but are not included in the estimation itself. This option is only helpful in case of small panels.

B.1.4 Saved results

`xtdcce2` saves the following in `e()`:

Scalars			
`e(N)`	number of observations	`e(N_g)`	number of groups
`e(T)`	number of time periods	`e(df_m)`	model degrees of freedom
`e(K_partial)`	number of variables partialled out	`e(K_mg)`	number of regressors (excluding partialled out)
`e(K_pooled)`	number of pooled variables	`e(K_omitted)`	number of omitted variables
`e(cr_lags)`	number of lags	`e(mss)`	model sum of square
`e(rss)`	residual sum of squares	`e(rmse)`	root mean squared error
`e(F)`	F statistic	`e(df_r)`	residual degree of freedom
`e(r2)`	R-squared	`e(r2_a)`	R-squared adjusted
`e(cd)`	CD test statistic	`e(cdp)`	p-value of CD test statistic
Scalars (unbalanced panel)			
`e(Tmin)`	minimum time	`e(Tmax)`	maximum time
`e(Tbar)`	average time		
Macros			
`e(tvar)`	name of time variable	`e(idvar)`	name of unit variable
`e(depvar)`	name of dependent variable	`e(indepvar)`	name of independent variables
`e(omitted)`	name of omitted variables	`e(lr)`	long run variables
`e(pooled)`	name of pooled variables	`e(cmdline)`	command line
`e(cmd)`	returns command (`xtdcce2`)		
Macros (iv-specific)			
`e(insts)`	instruments (exogenous) variables	`e(instd)`	instrumented (endogenous) variables
Matrices			
`e(b)`	coefficient vector (mean group)	`e(V)`	variance–covariance matrix (mean group)
`e(bi)`	coefficient vector (individual and pooled)	`e(Vi)`	variance–covariance matrix (individual and pooled)
Functions			
`e(sample)`	marks estimation sample		

B.1.5 Postestimation

`predict` and `estat` can be used after `xtdcce2`. The syntax of `predict` following `xtdcce2` is:

`predict` [`type`] *newvarname*[*if*][*in*][`, xb residuals cfresiduals stdp coefficients se partial`]

The default option is `xb` and calculates the fitted values. `residuals` calculates the residuals and `cfresiduals` calculates the residuals including the comman factors. Important to note is, that if the option `reportconstant` is not used, then the common factors include the constant. `stdp` calculates the standard error of the prediction. `coefficients` creates a separate variable for each coefficient with the unit-specific estimate. `se` creates in the same fashion a variable with the standard errors. `partial` partials out the cross-sectional averages and saves the variables. The new variables have the name *newvarname_varname*.

`estat` following `xtdcce2` draws a box, bar or range plot of the MG coefficients. The syntax is:

`estat` ***graphtype*** [*varlist*] [*if*] [*in*] [, `combine(`*string*`)` `individual(`*string*`)` `nomg` `cleargraph`].

graphtype is either *bar* for a bar plot, *box* for a box plot or *rcap* for a range plot. *varlist* is optional and if not specified, all mean group coefficients are included. If the bar or range plot is drawn, then a bar plot for each mean group coefficient defined by *varlist* is created and all are combined in the end. The option `individual` passes further graph options to the individual graphs and `combine` to the combined graph. If a box plot is drawn, option `individual` controls the appearance of the graph. A confidence interval around the mean of the mean group estimate is added to the range plot. Option `nomg` prevents including the confidence interval. The option `cleargraph` clears the option of the graph command and is best used in combination with the `combine()` and `individual()` options. Options `combine()` and `individual()` are used without a leading and ending quotation marks. The name of the graph is saved as *r(graph_name)*.

B.1.6 xtcd2

Included in the `xtdcce2` package is the `xtcd2` command, which tests for weak cross-sectional dependence. The command supports balanced as well as unbalanced panels.[23] For a discussion of the test statistic, see section 1.4.

[23] `xtcd2` differs from existing routines as `xtcsd` or `xtcd` in the sense that it follows the computation of the correlation coefficients in Pesaran (2015), while other routines rely on Stata's correlation function. Therefore, a difference can occur if the average of the variable within a cross-section is non zero.

Syntax

`xtcd2` [*varname*], [`noestimation rho kdensity name(`*string*`)`]

varname is the name of the residuals or variables to be tested. *varname* is optional in case the command is performed after an estimation command which supports `predict, residuals`. Then `xtcd2` predicts and tests the residuals for weak cross-sectional dependence.

Options

If `noestimation` is specified, then `xtcd2` is not run as a postestimation command and does not require `e(sample)` to be set. This option allows any variable to be tested. If not set, then `xtcd2` either uses the variable specified in *varname* or predicts the residuals using `predict, residuals`. In both cases the sample is restricted to `e(sample)`.

`kdensity` plots a kernel density graph of the cross correlations. The number of observations, the mean, percentiles, minimum and maximum of the cross correlations are reported. If `name(`*string*`)` is set, then the graph is saved and not drawn.

`rho` saves the matrix with the cross correlations in `r(rho)`.

Saved results

Scalars	
`r(CD)`	value of the CD statistic
`r(p)`	p-value
Matrices	
`r(rho)`	cross correlations matrix, if requested

B.1.7 Error Messages

`xtdcce2` produces the following error codes:

`r(109)`	`ivreg2` not installed	`r(184)`	options `noconstant` and `pooledconstant`, `trend` and `trendconstant` or `jackknife` and `recursive` are combined.
`r(2001)`	More variables than observations.		

B.2 Mathematical Appendix - Delta Method

For the calculation of the long run coefficients the estimates of $\hat{\gamma}_{k,i}$ obtained by OLS from equation (3.10) are divided by estimates of the long run cointegration vector $\hat{\phi}_i$. To calculate the variance covariance matrix, the delta method is used. The delta method allows the calculation of an approximate probability distribution for a matrix function $\mathbf{a}(\boldsymbol{\beta})$ based on a random vector with a known variance (see for example Hayashi, 2000, p. 93). Suppose that for the random vector $\boldsymbol{\beta}_i \rightarrow_p \boldsymbol{\beta}$ and $\sqrt{n}(\boldsymbol{\beta}_i - \boldsymbol{\beta}) \rightarrow_d N(0, \sigma)$. Denote the first derivatives of $\mathbf{a}(\boldsymbol{\beta})$ as

$$\mathbf{A}(\boldsymbol{\beta}) \equiv \frac{\partial a(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}'}.$$

Then the distribution of the function $\mathbf{a}()$ is

$$\sqrt{n}\left[\mathbf{a}(\boldsymbol{\beta}_i) - \mathbf{a}(\boldsymbol{\beta})\right] \rightarrow_d N\left(0, \mathbf{A}(\boldsymbol{\beta})\boldsymbol{\Sigma}\mathbf{A}(\boldsymbol{\beta})'\right).$$

For the calculation of the long run coefficients and using the notation from equation (3.10), assume that

$$\boldsymbol{\beta}_i = (\phi_i, \gamma_{1,i}, \gamma_{2,i}, \delta_{1,i}, \delta_{2,i})'$$

The variance covariance matrix is:

$$\mathbf{\Sigma} = \begin{pmatrix} V(\phi_i) & Cov(\phi_i\gamma_{1,i}) & Cov(\phi_i\gamma_{2,i}) & Cov(\phi_i\delta_{1,i}) & Cov(\phi_i\delta_{2,i}) \\ Cov(\phi_i\gamma_{1,i}) & V(\gamma_{1,i}) & Cov(\gamma_{1,i}\gamma_{2,i}) & ... & . \\ . & & . & & . \\ . & & & . & . \\ Cov(\phi_i\delta_{2,i}) & ... & ... & ... & V(\delta_{2,1}) \end{pmatrix}$$

The function $\mathbf{a}()$ maps the long run coefficients and leaves the short run coefficients:

$$\begin{aligned} \mathbf{a}(\boldsymbol{\beta}_i) &= (\phi_i, -\gamma_{1,i}/\phi_i, -\gamma_{2,i}/\phi_i, \delta_{1,i}, \delta_{2,i})' \\ &= (\phi_i, \theta_{1,i}, \theta_{2,i}, \delta_{1,i}, \delta_{2,i})' \end{aligned}$$

The first derivative of $\mathbf{a}()$ is then:

$$\mathbf{A}_i(\boldsymbol{\beta}) = \begin{pmatrix} \frac{\partial\phi_i}{\partial\phi_i} & \frac{\partial\theta_{1,i}}{\partial\phi_i} & \frac{\partial\theta_{2,i}}{\partial\phi_i} & \frac{\partial\delta_{1,i}}{\partial\phi_i} & \frac{\partial\delta_{2,i}}{\partial\phi_i} \\ \frac{\partial\phi_i}{\partial\gamma_{1,i}} & \frac{\partial\theta_{1,i}}{\partial\gamma_{1,i}} & \frac{\partial\theta_{2,i}}{\partial\gamma_{1,i}} & \frac{\partial\delta_{1,i}}{\partial\gamma_{1,i}} & \frac{\partial\delta_{2,i}}{\partial\gamma_{1,i}} \\ \frac{\partial\phi_i}{\partial\gamma_{2,i}} & \frac{\partial\theta_{1,i}}{\partial\gamma_{2,i}} & \frac{\partial\theta_{2,i}}{\partial\gamma_{2,i}} & \frac{\partial\delta_{1,i}}{\partial\gamma_{2,i}} & \frac{\partial\delta_{2,i}}{\partial\gamma_{2,i}} \\ \frac{\partial\phi_i}{\partial\delta_{1,i}} & \frac{\partial\theta_{1,i}}{\partial\delta_{1,i}} & \frac{\partial\theta_{2,i}}{\partial\delta_{1,i}} & \frac{\partial\delta_{1,i}}{\partial\delta_{1,i}} & \frac{\partial\delta_{2,i}}{\partial\delta_{1,i}} \\ \frac{\partial\phi_i}{\partial\delta_{2,i}} & \frac{\partial\theta_{1,i}}{\partial\delta_{2,i}} & \frac{\partial\theta_{2,i}}{\partial\delta_{2,i}} & \frac{\partial\delta_{1,i}}{\partial\delta_{2,i}} & \frac{\partial\delta_{2,i}}{\partial\delta_{2,i}} \end{pmatrix}$$

$$= \begin{pmatrix} 1 & -\frac{\gamma_{1,i}}{\phi_i^2} & -\frac{\gamma_{2,i}}{\phi_i^2} & 0 & 0 \\ 0 & -\frac{1}{\phi_i} & 0 & 0 & 0 \\ 0 & 0 & -\frac{1}{\phi_i} & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix}$$

All components of the variance covariance matrix are then known and it can be calculated as:

$$\mathbf{\Sigma}_a = \mathbf{A}(\boldsymbol{\beta})\mathbf{\Sigma}\mathbf{A}(\boldsymbol{\beta})'$$

B.3 Monte Carlo Simulation

B.3.1 Monte Carlo Setup

As in Chudik and Pesaran (2015a) the data generating processes are the following:

$$y_{i,t} = c_{yi} + \phi_i y_{i,t-1} + \beta_{0i} x_{i,t} + \beta_{1i} x_{i,t-1} + u_{i,t}$$

$$u_{i,t} = \gamma_i' f_t + \epsilon_{i,t}$$

$$x_{i,t} = c_{xi} + \alpha_{xi} y_{i,t-1} + \gamma_{xi} f_t + v_{xi,t}$$

$$g_{i,t} = c_{gi} + \alpha_{gi} y_{i,t-1} + \gamma_{gi} f_t + v_{gi,t}$$

$y_{i,t}$ is the dependent variable and $x_{i,t}$ the only independent variable. For a matter of ease, it is assumed that only one explanatory variable exists. $g_{i,t}$ is another independent variable, which is affected by the unobserved factors and $y_{i,t}$, but not used to estimate it.

Common Factors The common factors are calculated as below:

$$f_t = \rho_f f_{t-1} + \varsigma_{ft}, \varsigma_{ft} \sim IIDN(0, 1 - \rho_f^2)$$

$$v_{xi,t} = \rho_{xi} v_{xi,t-1} + \varsigma_{xi,t}, \varsigma_{xi,t} \sim IIDN(0, \sigma_{vxi}^2)$$

$$v_{gi,t} = \rho_{gi} v_{gi,t-1} + \varsigma_{gi,t}, \varsigma_{gi,t} \sim IIDN(0, \sigma_{vgi}^2)$$

$$\rho_{xi} \sim IIDU(0, 0.95)$$

$$\rho_{gi} \sim IIDU(0, 0.95)$$

$$\rho_f = 0 \text{ if serially uncorrelated factors, or if correlated } \rho_f = 0.6$$

$$\sigma_{vxi}^2 = \sigma_{vgi}^2 = \sigma_{vi}^2 = \left(\beta_{0i} \sqrt{1 - [E(\rho_{xi})]^2}\right)^2$$

Fixed Effects The cross-section specific fixed effects are generated as:

$$
\begin{aligned}
c_{yi} &\sim IIDN(1,1)\\
c_{xi} &= c_{yi} + \varsigma_{c_x i}, \varsigma_{c_x i} \sim IIDN(0,1)\\
c_{gi} &= c_{yi} + \varsigma_{c_g i}, \varsigma_{c_g i} \sim IIDN(0,1).
\end{aligned}
$$

Dependence between $x_{i,t}, g_{i,t}$ and c_{yi} is introduced by adding c_{yi} to the equations for c_{xi} and c_{gi}.

Coefficients The coefficient for the contemporaneous value of $x_{i,t}$ is drawn from a uniform distribution as $\beta_{0i} \sim IIDU(0.5, 1)$. The coefficient on the lagged value of the independent variable is set to $\beta_{1i} = -0.5$. For the lagged dependent variable two different scenarios are considered for the calculation of $y_{i,t}$ and $x_{i,t}$. One with low values for ϕ, $\phi_i \sim IIDU(0, 0.8)$ and $\alpha_{xi} \sim IIDU(0, 0.35)$ and one with high values $\phi_i \sim IIDU(0.5, 0.9)$ and $\alpha_{xi} \sim IIDU(0, 0.15)$. α_{gi} is in both scenarios the same: $\alpha_{gi} \sim IIDU(0, 1)$.[24]

Factor Loadings

$$
\begin{aligned}
\gamma_i &= \gamma + \eta_{i\gamma}, \eta_{i\gamma} \sim IIDN(0, \sigma_\gamma^2)\\
\gamma_{xi} &= \gamma_x + \eta_{i\gamma x}, \eta_{i\gamma x} \sim IIDN(0, \sigma_{\gamma x}^2)\\
\gamma_{gi} &= \gamma_g + \eta_{i\gamma g}, \eta_{i\gamma g} \sim IIDN(0, \sigma_{\gamma g}^2)\\
\sigma_\gamma^2 &= \sigma_{\gamma x}^2 = \sigma_{\gamma g}^2 = 0.2^2\\
\gamma &= \sqrt{b_\gamma}, b_\gamma = \frac{1}{m} - \sigma_\gamma^2\\
\gamma_x &= \sqrt{b_x}, b_x = \frac{2}{m(m+1)} - \frac{2}{m+1}\sigma_{\gamma x}^2\\
\gamma_g &= \sqrt{b_g}, b_g = \frac{1}{m^2} - \frac{\sigma_g^2}{m}
\end{aligned}
$$

where m is the number of unobserved factors. In comparison to Chudik and Pesaran (2015a) it is restricted to 1.

[24] ϕ_i and α_{xi} depend on each other to make sure that the series $y_{i,t}$ and $x_{i,t}$ are stationary. See Chudik and Pesaran (2015a), p. 399-400.

Error Term The errors are generated such that heteroskedasticity and weakly cross-sectional dependence is allowed.

$$
\begin{aligned}
\boldsymbol{\epsilon}_t &= \alpha_{CSD}\mathbf{S}_\epsilon\boldsymbol{\epsilon}_t + \mathbf{e}_{\epsilon t} \\
\Rightarrow \boldsymbol{\epsilon}_t &= \left(1 - \alpha_{CSD}\mathbf{S}_\epsilon\right)^{-1}\mathbf{e}_{\epsilon t} \\
\mathbf{e}_{\epsilon t} &\sim IIDN(0, \frac{1}{2}\sigma_i^2), \text{ with } \sigma_i^2 \sim \chi^2(2) \\
\mathbf{S}_\epsilon &= \begin{pmatrix}
0 & \frac{1}{2} & 0 & 0 & \dots & 0 \\
\frac{1}{2} & 0 & 1 & 0 & & 0 \\
0 & 1 & 0 & \ddots & & \vdots \\
0 & 0 & \ddots & \ddots & 1 & 0 \\
\vdots & & & 1 & 0 & \frac{1}{2} \\
0 & 0 & \dots & 0 & \frac{1}{2} & 0
\end{pmatrix}
\end{aligned}
$$

B.3.2 Further Monte Carlo Results

(N,T)	Bias (x100)					RMSE (x100)				
	40	50	100	150	200	40	50	100	150	200
Fixed Effects estimates										
40	-4.79	-2.04	1.64	3.10	3.63	5.60	4.30	3.23	3.38	3.43
50	-4.98	-2.48	1.29	2.93	3.46	5.42	4.02	2.83	3.10	3.28
100	-4.38	-2.43	1.53	2.93	3.57	4.24	3.26	2.34	2.70	2.96
150	-4.22	-2.12	1.74	3.05	3.72	3.99	2.71	2.10	2.57	2.92
200	-4.39	-2.28	1.91	2.86	3.53	3.89	2.64	2.04	2.42	2.75
DCCE without bias correction										
40	-26.26	-19.45	-8.43	-5.31	-3.61	19.19	14.30	6.42	4.26	3.01
50	-26.97	-19.24	-8.54	-5.36	-3.97	19.61	14.07	6.43	4.16	3.18
100	-27.34	-20.03	-8.81	-5.63	-3.97	19.58	14.35	6.41	4.16	2.98
150	-26.64	-19.73	-8.76	-5.65	-4.09	19.01	14.03	6.28	4.09	2.99
200	-27.36	-19.72	-8.84	-5.72	-3.99	19.45	14.00	6.31	4.10	2.89
DCCE with jackknife bias correction										
40	-24.13	-17.22	-6.63	-3.98	-2.51	17.82	12.88	5.30	3.48	2.42
50	-24.85	-16.95	-6.69	-3.98	-2.86	18.26	12.59	5.26	3.34	2.55
100	-25.13	-17.72	-7.02	-4.24	-2.89	18.14	12.81	5.22	3.26	2.29
150	-24.39	-17.32	-6.95	-4.26	-2.98	17.49	12.39	5.06	3.16	2.26
200	-25.25	-17.40	-7.02	-4.33	-2.89	18.03	12.43	5.08	3.16	2.17
DCCE with RMA bias correction										
40	-31.08	-23.25	-10.17	-6.41	-4.48	22.59	16.95	7.59	4.96	3.56
50	-31.43	-23.00	-10.43	-6.39	-4.82	22.74	16.70	7.72	4.84	3.72
100	-31.83	-23.57	-10.46	-6.55	-4.74	22.72	16.82	7.54	4.78	3.49
150	-31.02	-23.23	-10.32	-6.60	-4.86	22.04	16.46	7.36	4.74	3.51
200	-31.52	-23.32	-10.38	-6.65	-4.72	22.35	16.52	7.38	4.74	3.39

Table B.1: Monte Carlo Results for Specification 4 and ϕ, with $\phi = E(\phi_i) = 0.7$ and $\rho_f = 0$. For a further description, see Section 3.6.

(N,T)	Bias (x100)					RMSE (x100)				
	40	50	100	150	200	40	50	100	150	200
Fixed Effects estimates										
40	79.89	80.73	79.10	79.97	80.30	60.59	61.01	59.67	60.12	60.26
50	80.10	79.16	79.70	80.32	79.73	60.69	59.90	60.06	60.31	59.99
100	80.04	80.32	80.42	79.85	80.21	60.45	60.56	60.39	60.00	60.23
150	79.72	79.29	79.67	79.85	80.01	60.10	59.79	59.86	59.97	60.12
200	80.16	80.39	79.90	79.84	79.69	60.39	60.50	60.08	59.92	59.82
DCCE without bias correction										
40	0.90	1.03	1.11	0.80	0.40	11.87	8.84	5.52	4.35	3.67
50	1.82	0.83	1.27	0.65	0.84	10.33	8.53	5.22	4.21	3.72
100	1.42	1.57	1.12	0.51	0.75	7.94	6.19	3.94	3.00	2.56
150	0.82	1.67	1.03	0.65	0.53	6.37	5.40	3.21	2.43	2.06
200	2.18	1.91	1.25	0.70	0.52	5.70	4.43	2.83	2.23	1.83
DCCE with jackknife bias correction										
40	0.74	1.20	0.93	0.65	0.27	12.14	9.36	5.57	4.40	3.70
50	1.55	0.71	1.11	0.51	0.75	10.81	8.96	5.30	4.25	3.72
100	1.46	1.58	0.94	0.36	0.63	8.25	6.51	3.98	3.04	2.58
150	0.91	1.48	0.89	0.51	0.43	6.74	5.54	3.25	2.44	2.06
200	2.18	1.84	1.08	0.54	0.38	5.88	4.59	2.87	2.23	1.83
DCCE with RMA bias correction										
40	2.84	3.82	3.11	2.16	1.73	12.26	9.59	5.94	4.57	3.93
50	4.01	3.24	3.33	2.03	2.05	11.04	9.01	5.75	4.53	4.00
100	3.13	4.07	3.02	1.80	1.95	8.55	6.97	4.51	3.33	2.92
150	2.98	4.07	2.80	1.99	1.74	6.92	6.15	3.79	2.86	2.42
200	3.92	4.29	3.07	1.98	1.69	6.32	5.34	3.57	2.67	2.21

Table B.2: Monte Carlo Results for Specification 4 and β_0, with $\phi = E(\phi_i) = 0.7$ and $\rho_f = 0$. See notes Table 3.2.

(N,T)	Bias (x100)					RMSE (x100)				
	40	50	100	150	200	40	50	100	150	200
Fixed Effects estimates										
40	28.49	35.05	41.72	44.96	46.49	17.49	19.79	21.92	23.20	23.82
50	28.57	33.30	41.83	45.34	45.46	17.32	18.64	21.64	23.28	23.18
100	29.12	34.49	42.52	44.76	45.98	16.19	18.36	21.77	22.71	23.26
150	29.56	34.19	42.58	44.81	46.65	15.73	17.92	21.65	22.65	23.51
200	29.83	34.94	42.45	44.71	46.43	15.86	18.12	21.49	22.55	23.36
DCCE without bias correction										
40	-23.19	-17.04	-6.33	-4.00	-2.66	16.71	13.11	6.89	5.08	4.39
50	-22.46	-16.57	-6.67	-4.06	-3.38	16.08	12.28	6.63	4.83	4.04
100	-21.16	-14.79	-6.55	-4.50	-2.60	13.26	9.87	5.19	3.96	3.05
150	-21.95	-14.91	-5.86	-3.81	-2.92	13.06	9.35	4.54	3.34	2.72
200	-21.19	-14.62	-5.99	-4.29	-2.66	12.36	8.87	4.09	3.25	2.33
DCCE with jackknife bias correction										
40	-20.92	-15.19	-5.02	-2.95	-1.88	16.44	12.89	6.72	5.02	4.32
50	-20.68	-14.90	-5.15	-3.02	-2.61	15.85	12.07	6.30	4.66	3.92
100	-19.47	-12.78	-5.16	-3.47	-1.82	12.91	9.44	4.86	3.73	2.90
150	-19.91	-12.92	-4.50	-2.78	-2.13	12.31	8.74	4.15	3.09	2.54
200	-19.63	-12.76	-4.68	-3.29	-1.92	11.84	8.17	3.67	2.95	2.15
DCCE with RMA bias correction										
40	-23.01	-16.02	-4.25	-2.58	-1.36	17.19	12.99	6.56	4.90	4.25
50	-21.81	-14.81	-4.96	-2.54	-2.01	15.95	12.07	6.30	4.59	3.79
100	-20.13	-12.54	-4.61	-3.04	-1.16	13.01	9.19	4.66	3.63	2.86
150	-21.41	-13.06	-3.62	-2.21	-1.57	12.95	8.78	3.98	3.00	2.47
200	-20.56	-12.38	-4.09	-2.81	-1.31	12.25	8.12	3.54	2.85	2.05

Table B.3: Monte Carlo Results for Specification 4 and β_1, with $\phi = E(\phi_i) = 0.7$ and $\rho_f = 0$. See notes Table 3.2.

(N,T)	Bias (x100)					RMSE (x100)				
	40	50	100	150	200	40	50	100	150	200
Fixed Effects estimates										
40	5.21	10.20	16.51	19.98	19.76	7.21	7.60	8.36	9.16	8.94
50	4.07	8.82	15.38	18.56	19.84	6.38	6.67	7.68	8.53	8.91
100	5.23	8.91	15.70	18.34	19.54	5.13	5.81	7.22	8.01	8.37
150	5.60	9.71	16.09	18.43	20.06	4.91	5.62	7.08	7.82	8.44
200	5.16	8.66	16.55	18.08	19.16	4.50	4.98	7.23	7.68	8.01
DCCE without bias correction										
40	-38.75	-28.93	-12.53	-7.49	-5.18	16.55	12.52	5.92	3.91	2.90
50	-38.99	-27.87	-12.63	-7.68	-5.93	16.54	12.05	5.82	3.78	3.04
100	-39.47	-29.02	-12.73	-8.32	-5.74	16.35	11.98	5.49	3.72	2.62
150	-38.59	-28.54	-12.78	-8.17	-5.95	15.83	11.69	5.34	3.50	2.61
200	-39.82	-28.92	-12.76	-8.42	-5.80	16.23	11.81	5.30	3.52	2.50
DCCE with jackknife bias correction										
40	-36.40	-26.41	-10.34	-5.87	-3.83	15.77	11.69	5.24	3.46	2.56
50	-36.46	-25.14	-10.39	-5.97	-4.57	15.67	11.12	5.10	3.27	2.66
100	-36.97	-26.20	-10.52	-6.62	-4.42	15.49	10.95	4.70	3.13	2.19
150	-35.93	-25.63	-10.53	-6.46	-4.60	14.83	10.59	4.51	2.87	2.13
200	-37.39	-26.08	-10.51	-6.70	-4.44	15.32	10.74	4.44	2.88	2.02
DCCE with RMA bias correction										
40	-42.21	-31.61	-13.69	-8.22	-5.82	17.97	13.58	6.33	4.15	3.11
50	-42.45	-30.61	-14.04	-8.39	-6.53	17.94	13.13	6.35	4.02	3.23
100	-42.80	-31.61	-13.95	-8.95	-6.28	17.71	13.03	5.95	3.95	2.82
150	-42.05	-31.20	-13.87	-8.85	-6.53	17.20	12.73	5.76	3.77	2.82
200	-42.95	-31.65	-13.90	-9.12	-6.33	17.48	12.89	5.75	3.79	2.70

Table B.4: Monte Carlo Results for Specification 5 and ϕ, with $\phi = E(\phi_i) = 0.3$ and $\rho_f = 0.6$. The coefficient on the lagged independent variable `L.x` is pooled. See notes Table 3.2.

(N,T)	Bias (x100)					RMSE (x100)				
	40	50	100	150	200	40	50	100	150	200
Fixed Effects estimates										
40	65.08	65.50	63.27	64.18	64.43	49.89	49.91	48.00	48.45	48.51
50	65.49	64.05	63.98	64.16	63.92	50.02	48.83	48.42	48.34	48.22
100	65.23	65.02	64.26	63.43	64.01	49.54	49.23	48.39	47.76	48.15
150	64.64	64.03	63.28	63.61	63.26	48.95	48.46	47.65	47.87	47.60
200	65.28	64.84	63.84	63.47	63.15	49.37	48.94	48.10	47.71	47.47
DCCE without bias correction										
40	4.24	3.15	1.83	1.32	0.66	11.92	8.92	5.49	4.32	3.66
50	4.81	3.08	2.16	1.31	1.20	10.96	8.36	5.26	4.21	3.76
100	4.23	3.92	2.17	1.16	1.21	8.40	6.42	3.96	3.05	2.54
150	3.62	4.04	2.14	1.36	1.12	6.81	5.73	3.42	2.57	2.13
200	4.90	3.92	2.38	1.46	1.03	6.48	5.07	3.18	2.35	1.86
DCCE with jackknife bias correction										
40	3.59	2.84	1.37	0.93	0.34	11.96	9.17	5.48	4.30	3.67
50	3.97	2.40	1.66	0.89	0.90	11.06	8.44	5.27	4.21	3.72
100	3.55	3.29	1.59	0.75	0.88	8.61	6.40	3.84	3.00	2.50
150	2.99	3.22	1.59	0.91	0.78	6.80	5.54	3.30	2.47	2.07
200	4.20	3.15	1.81	1.02	0.64	6.25	4.91	3.02	2.24	1.78
DCCE with RMA bias correction										
40	5.17	4.60	2.72	1.89	1.20	12.51	9.35	5.66	4.45	3.76
50	6.16	4.37	3.09	1.86	1.69	11.64	8.85	5.49	4.39	3.86
100	5.09	5.17	3.05	1.71	1.70	8.94	7.13	4.29	3.21	2.70
150	4.62	5.27	2.93	1.96	1.62	7.43	6.44	3.75	2.80	2.32
200	5.75	5.26	3.23	2.01	1.53	7.00	5.74	3.60	2.59	2.06

Table B.5: Monte Carlo Results for Specification 5 and β_0, with $\phi = E(\phi_i) = 0.3$ and $\rho_f = 0.6$. The coefficient on the lagged independent variable `L.x` is pooled. See notes Table 3.2.

Chapter 4

Cross-Country Convergence in a General Lotka-Volterra Model

An early version of this chapter is published in *Spatial Economic Analysis*, 2018, Volume 13, Issue 2 (Ditzen, 2018).[1]

4.1 Introduction

This chapter reverts back to the questions posed in the Introduction. In the core lies the question if interdependence between countries matter for their economic development. The convergence model in Arbia and Paelinck (2003a,b) is extended. In addition the chapter borrows from the literature on multifactor error structure models. Following Arbia and Paelinck (2003a,b) a general Lotka-Volterra Model is used to determine the type of convergence in a set of 93 countries rather than regions. Absolute convergence is tested by estimating the steady states for each country and then testing those for equality. The stability conditions of the Lotka-Volterra model are used in conjunction with a bootstrap to test for conditional convergence. This chapter advances the empirical spatial growth literature in the form that both forms of cross-sectional dependence are controlled for. Weak cross-sectional dependence, or spatial dependence, is controlled for by adding spatial time lags of the

[1] I am grateful to the editor Paul Elhorst and two anonymous referees for their comments and feedback.

dependent variable. Chapter 2 pointed at three different channels for interactions between countries. This motivates the use of high skilled migration, exports and Foreign Direct Investments (FDI) as spatial weight matrices. Strong cross-sectional dependence is taken care of in the form of a multifactor error structure model. The common effects are approximated by cross-sectional averages as explained in Section 1.5.

There are several notable findings. A Lotka-Volterra approach on a classical convergence equation without cross-sectional interactions confirms earlier empirical findings in the literature. However strong cross-sectional dependence remains in these specifications, invalidating these results. Adding cross-sectional averages and spatial lags controls sufficiently for cross-sectional dependence. Lastly, this chapter presents evidence for conditional convergence.

The remainder of the chapter is structured as follows. The next section introduces a general Lotka-Volterra model, followed by a discussion about convergence. Then a growth model is outlined and the estimation strategy and the empirical equation are discussed. In the following section, the empirical results are presented and checked against alternative specifications. The chapter closes with a conclusion.

4.2 A General Lotka-Volterra Model

Lotka-Volterra models are used in mathematical biology to model the evolution of two dependent species and have been developed by Lotka (1920) and Volterra (1926).[2] The general two equation Lotka-Volterra model describes the evolution of two species, x and y, with the help of a differential equation system. The population of x depends negatively on the level of the other, y, but grows positively in the absence of y. The population of y depends positively on the population size of

[2]The textbook example of the Lotka-Volterra Model is on the interaction of a predator, y, and a prey, x. In absence of the predator, the population of the prey increases by xa. As the predator needs the prey to survive its population decreases by yc. The larger the population of the predator, more prey are hunted and their population decreases by by. On the other hand, large numbers of prey lead to an increase of the population of the predators by dx.

species x, but decreases if the size is zero. The path over time of the two populations can be described by the following two equations:[3]

$$\frac{dx}{dt} = x(a - by)$$
$$\frac{dy}{dt} = y(-c + dx)$$

where $a > 0, b > 0, c > 0, d > 0$. a and c capture the effect the specie has on its own, while b and d represent the interaction between both.

Among the first contributions to an economic interpretation of the Lotka-Volterra model are Goodwin (1967) and Samuelson (1971). Samuelson (1971) presented the standard Lotka-Volterra model in an economic context, generalised it to more than two species and derived its equilibrium behaviour. Goodwin (1967) applied the Lotka-Volterra model to an endogenous growth model with Harrod-Domar technology. In the model income is divided into a worker's share, or the wage in efficiency terms, and investments. The worker's share of income preys on the employment rate. If the worker's share of income is high, investments are low, which leads to a contracting employment and reduces the employment rate. On the other hand, if the employment rate is high, income needs to be spread across a larger number of workers and therefore the worker's share of income is low.

In a similar fashion, instead of the two species or worker's share and the employment rate, assume two countries R and P, with output y_R and y_P. Similar to the population Lotka-Volterra model, one is "preying" upon the other. Say, the output per capita of country P increases if country R's increases. However the opposite does not hold and country P's output is reduced with the others'. The two interactions can be captured by the terms $-\rho_R w_R$ and $\rho_P w_P$. For later use, ρ_R and ρ_P are coefficients to be estimated and w_R and w_P observed weights. The system can be

[3] The equation follows Murray (2002, Chapter 3.1).

then expressed by:

$$\frac{dy_R}{dt} = y_R(a - \rho_R w_R y_P) \Rightarrow \frac{\dot{y}_R}{y_R} = a - \rho_R w_R y_P$$
$$\frac{dy_P}{dt} = y_P(-c + \rho_P w_P y_R) \Rightarrow \frac{\dot{y}_P}{y_P} = -c + \rho_P w_P y_R$$

The two equations can be stacked in matrices:

$$\begin{aligned}
\begin{pmatrix} \frac{\dot{y}_R}{y_R} \\ \frac{\dot{y}_P}{y_P} \end{pmatrix} &= \begin{pmatrix} a \\ -c \end{pmatrix} + \begin{pmatrix} -\rho_R w_R y_P \\ \rho_P w_P y_R \end{pmatrix} \\
&= \begin{pmatrix} a \\ -c \end{pmatrix} + \begin{pmatrix} -\rho_R & 0 \\ 0 & \rho_P \end{pmatrix} \begin{pmatrix} 0 & w_R \\ w_P & 0 \end{pmatrix} \begin{pmatrix} y_R \\ y_P \end{pmatrix} \\
&= \mathbf{c} + \boldsymbol{\rho}\mathbf{W}\mathbf{y} \\
&= \mathbf{c} + \mathbf{D}\mathbf{y}.
\end{aligned}$$

Adding time indices gives

$$\Delta \mathbf{y}_t = \mathbf{c} + \mathbf{D}\mathbf{y}_{t-1}. \tag{4.1}$$

If the growth rate converges to zero, two solutions can be derived from $\mathbf{y}_t = \mathbf{y}_{t-1} = \mathbf{y}^* = -\mathbf{D}^{-1}\mathbf{c}$. The first solution is $(y_R, y_P) = (0, 0)$ and in the prevailing context meaningless as it would imply that the income of a country is zero. The second solution is $(y_R, y_P) = (c/(\rho_P w_P), a/(\rho_R w_R))$ and has a more meaningful interpretation. If $y_R = y_P$ then the countries converge to the same output per capita. Both solutions are stable if the real parts of the eigenvalues of the community matrix $\mathbf{D}$ are negative (Arbia and Paelinck, 2003b; Griffith and Paelinck, 2011).[4] Thus in the case of $y_R \neq y_P$ and stability, both countries would converge to their own equilibria and conditional convergence occurs.

[4]The condition is derived from a Liapunov function; for further details see Arbia and Paelinck (2003b); Griffith and Paelinck (2011). In mathematical biology, the matrix $\mathbf{D}$ in equation (4.1) is often called a *community matrix* (Murray, 2002). This chapter follows this notation.

Using matrix notation, the steady state under convergence would be:

$$\Delta \mathbf{y}_t = 0 \tag{4.2}$$

$$\Rightarrow \mathbf{y}_t = \mathbf{y}_{t-1} = \mathbf{y}^* = -\mathbf{D}^{-1}\mathbf{c}. \tag{4.3}$$

The assumptions on the signs of the interactions $-\rho_R w_R$ and $\rho_P w_P$ can be relaxed, such that $\rho_R w_R$ and $\rho_P w_P$ are positive or negative. Then Lotka-Voterra models can be applied to a more general setting, as for example interactions between countries. In addition, the model is not limited to two countries or equations. In the case of N countries, there would be N difference equations. Still, it would be possible that some countries benefit, while others lose from the interactions. Samuelson (1971) describes such an extension with $n > 2$ predators and prey. Convergence would depend on the properties of the eigenvalues of the community matrix $\mathbf{D}$.

The predator and prey relationship in the general Lotka-Volterra model or the economic interpretation in Goodwin (1967) is a null-sum game. If the population size of the predator increases, the number of prey decrease. Even if there is an increase in the population of the prey, the increase in the number of the predators will decrease the population of the prey. In the country version of the Lokta-Volterra model as described by equation (4.1) with relaxed assumptions on the sign of the interactions, countries can benefit from each other and a null-sum game is not a necessity. While this is different to the classical Lotka-Volterra models, the same assumption applies to the model in Arbia and Paelinck (2003a,b). A Lotka-Volterra in which both species benefit from each other is described in a biological context as a *Mutualism* or *Symbiosis* Lotka-Volterra model in Murray (2002, Chapter 3.6 - 3.7).

4.3 Growth Model and Convergence

4.3.1 Growth Model

Based on a Cobb-Douglas production function, Barro and Sala-i Martin (1992) derive a closed form solution for a simple growth model in the fashion of:

$$\Delta y_{i,t} = \beta_i + \alpha y_{i,t-1} + u_{i,t}, \tag{4.4}$$

where $y_{i,t}$ is the log of GDP per capita at time period t in country i, β_i is a country specific technology term and $u_{i,t}$ is an iid error term. In general, there is convergence if the coefficient α on the lagged income per capita level is between -1 and 0 (Mankiw et al., 1992; Islam, 1995). The interpretation is, the higher the income per capita in period t, the lower the growth rate. Therefore, each country converges as the growth rate of income per capita declines. Under the assumption that technology is the same for all countries $\beta_i = \beta$ for $i = 1, ..., N$, unconditional or absolute convergence occurs if the coefficient α on the lagged income per capita level is between $-1 < \alpha < 0$ (Mankiw et al., 1992; Islam, 1995). Unconditional convergence implies that the gap between countries will be depleted. In the case that β_i is different for countries and $-1 < \alpha < 0$, countries converge to their own steady state, implying conditional convergence. Poor countries do not necessarily catch up and differences in income levels can persist with conditional convergence. In both cases, poor countries with a smaller level of GDP grow with a larger growth rate, as the coefficient on lagged GDP per capita is negative.

Chapter 2 describes how observable dependencies between countries can be modelled in a growth model. In a more applied fashion, Gallo and Fingleton (2014) extend the growth model from equation (4.4) by a spatial component to account for spatial dependence.[5] To model strong dependence, an unobserved common factor,

[5]For an overview of spatial growth models see Fingleton and López-Bazo (2006) and Abreu, De Groot and Florax (2004).

f_t, and an heterogeneous unit-specific factor loading, γ_i, are added. In addition, the coefficient on the lagged dependent variable is allowed to vary across countries:

$$\Delta y_{i,t} = \beta_i + \alpha_i y_{i,t-1} + \rho_i \sum_{i \neq j, j=1}^{N} w_{i,j} y_{j,t-1} + \epsilon_{i,t}, \tag{4.5}$$

$$\epsilon_{i,t} = \gamma_i f_t + u_{i,t} \tag{4.6}$$

where ρ_i is the spatial autocorrelation coefficient, $w_{i,j} y_{i,t-1}$ the spatial time lag and $w_{i,j}$ observed spatial weights. The spatial time lag accounts for observed spatial dependence as explained in the Introduction. In the same chapter, the motivation for the inclusion of spillovers and dependencies is discussed. As pointed out, there is no reason to believe that countries are isolated from each other. Furthermore, there is evidence for the increase in integration across countries presented in the *New Kaldor Facts* in Jones and Romer (2010). Chapter 2 presented a growth model with interactions between countries. The model underlines the crucial role interactions play for the development of a country.

In equation (4.5) the spillovers are captured by the spatial time lag and are therefore global. This means that the dependent variable depends on time lags of the dependent variable of the other cross-sectional units. In addition, the spatial weight matrices capture the effect of the neighbours and of neighbours of a higher order as well. Chapter 2 suggests trade in goods, diffusion of ideas and high skilled migration as possible channels for interactions between countries. Trade allows for specialization, FDI raises or lowers the accumulation of capital and migration in the form of human capital produces more ideas. The common factors, or strong dependence, can be for example common aggregate shocks or time specific fixed effects (Pesaran, 2006; Kuersteiner and Prucha, 2015).

In addition, the coefficients on the lagged dependent variable and the spatial autocorrelation coefficient are allowed to vary across countries. As noted in the Introduction, there are no reasons why countries have the same slope coefficients. The same argument applies to the spatial autocorrelation coefficient.

4.3.2 Convergence in spatial growth models

Determining convergence solely based on the coefficient α_i, would miss out the interactions between countries represented by the spatial time lag. Therefore, it is necessary to include the spatial time lag into the condition for convergence. Equation (4.5) can be rewritten in matrix form as:

$$\Delta \mathbf{y}_t = \boldsymbol{\beta}_0 + \mathbf{diag}(\boldsymbol{\alpha})\mathbf{y}_{t-1} + \mathbf{diag}(\boldsymbol{\rho})\mathbf{W}\mathbf{y}_{t-1} + \boldsymbol{\epsilon}_t, \tag{4.7}$$

with

$$\mathbf{y}_t = (y_{1,t}, ..., y_{N,t})' \qquad \boldsymbol{\epsilon}_t = (\epsilon_{1,t}, ..., \epsilon_{N,t})'$$

$$\boldsymbol{\alpha} = (\alpha_1, ..., \alpha_N)' \qquad \boldsymbol{\rho} = (\rho_i, ..., \rho_N)'$$

$$\mathbf{W} = \begin{pmatrix} 0 & w_{1,2} & \dots & w_{1,N} \\ w_{2,1} & \ddots & & \vdots \\ \vdots & & \ddots & \vdots \\ w_{N,1} & \dots & \dots & 0 \end{pmatrix} \qquad \boldsymbol{\beta}_0 = (\beta_{0,1}, ..., \beta_{0,N})'$$

In the next step $\mathbf{diag}(\boldsymbol{\alpha})$ and $\mathbf{diag}(\boldsymbol{\rho})\mathbf{W}$ are put together in a matrix and the system becomes:

$$\Delta \mathbf{y}_t = \mathbf{b} + \mathbf{A}\mathbf{y}_{t-1} + \boldsymbol{\epsilon}_t, \tag{4.8}$$

and

$$\mathbf{b} = \boldsymbol{\beta}_0 \tag{4.9}$$

$$\mathbf{A} = \mathbf{diag}(\boldsymbol{\alpha}) + \mathbf{diag}(\boldsymbol{\rho})\mathbf{W} = \begin{pmatrix} \alpha_1 & \rho_1 w_{1,2} & \dots & \rho_1 w_{1,N} \\ \rho_2 w_{2,1} & \alpha_2 & & \vdots \\ \vdots & & \ddots & \vdots \\ \rho_N w_{N,1} & \dots & \dots & \alpha_N \end{pmatrix} \tag{4.10}$$

The last equation is an economic growth representation of the Lotka-Volterra model as presented in equation (4.1) and used by Arbia and Paelinck (2003a). Similar conditions for convergence as for the Lotka-Volterra model outlined in Section 4.2 apply. For stability, the real parts of the eigenvalues of the community matrix **A** have to be $-2 < \lambda_1, ..., \lambda_N < 0$.[6] In addition equation, (4.5) can be re-written if $y_{i,t} = y_{i,t-1} \equiv y_i^*$ and $\Delta y_{i,t} = 0$ as:

$$y_i^* = -\frac{\beta_{0,i} + \rho_i \sum_{j=1,i\neq j} w_{i,j} y_j^*}{\alpha_i}. \tag{4.11}$$

Equation (4.11) illustrates several requirements for convergence. As α_i is expected to be negative, the sum of the numerator has to be positive, such that y_i^* is positive as well. If one country does not converge, all other countries do not converge as well, if $\rho_i \neq 0$ and $w_{i,j} \neq 0$. If either $\rho_i = 0$ or $w_{i,j} = 0$, then convergence for the other countries is possible, despite the entire system not being stable. Moreover, a negative ρ_i implies a larger steady state. For unconditional convergence, the entries of matrix $\mathbf{A}^{-1}\mathbf{b}$ are the same for each country, i.e. $y_i^* = y_j^* = y^*, \ i, j = 1, ..., N$. Thus the steady state under absolute convergence is:

$$y_i^* = y^* = -\frac{\beta_{0,i}}{\alpha_i + \rho_i \sum_{j=1,i\neq j} w_{i,j}}. \tag{4.12}$$

The outlined approach is different to those prevailing in the time series literature. In a pure time series framework with no cross-sectional units, a unit root test on the lag of the dependent variable and tests for stationary would be the equivalent to the cross-sectional test for convergence. For a time series panel, Pesaran (2007) derives a pair-wise test for output convergence basing on stationary of the output differences of a set of countries. However, this test does not take spatial interactions between countries into account.[7]

[6] The requirement for stability for a discreet dynamic equation is, $|\lambda_i| < 1, \ i = 1, ..., N$, see Galor (2007). This is the case if the level occurs on both sides; here on the left hand side the growth rate occurs.

[7] A unit root test in a panel with cross-sectional dependence is derived in Pesaran, Smith and Yamagata (2013).

4.4 Heterogeneous Panel Estimators

In this chapter, two estimators which allow estimations of unit-specific coefficients are considered: the DCCE and Simultaneous Dynamic Least Squares (SDLS) estimator. The DCCE estimator was introduced in Chapter 1 and its application discussed in Chapter 3. Arbia and Paelinck (2003a,b) employ the SDLS estimator. Therefore, the estimator SDLS is used in this chapter to compare the results with earlier findings in the literature. SDLS minimizes the squared deviations between an observed and an endogenously computed value and is similar to Seemingly Unrelated Regression (SURE) or a 3 Stage Least Squares (3SLS) estimator without endogenous variables. It is a generalised reduced form estimator, a ML estimator if the error term is homoscedastic, normally distributed with mean zero and a variance $\sigma^2\mathbf{I}$. For a further description see the Appendix C.2 or Griffith and Paelinck (2011, Chapter 11).

Both estimators produce consistent estimates of the unit-specific coefficients as necessary for the construction of the matrices $\mathbf{A}$ and $\mathbf{b}$ in equation (4.8). While DCCE allows controlling for both types of cross-sectional dependence, SDLS can only handle spatial dependence using spatial lags.

4.5 Empirical Equation

Rewriting equation (4.5) with the cross-sectional averages instead of the common factors, the following equation is estimated:[8]

$$\Delta y_{i,t} = \beta_{0,i} + \alpha_i y_{i,t-1} + \sum_{k=1}^{K} \rho_{(k),i} \sum_{j=1,i\neq j}^{N} w_{(k),i,j} y_{j,t-1} + \sum_{l=0}^{p_T} \vartheta_i \bar{y}_{i,t-l} + u_{i,t} \qquad (4.13)$$

where $\sum_{k=1}^{K} \rho_{(k),i} \sum_{j=1,i\neq j}^{N} w_{(k),i,j} y_{j,t-1}$ captures spatial dependence. $\bar{y}_{i,t-l}$ is the l-th of p_T lags of the cross-sectional averages of $y_{i,t}$ and ϑ_i the coefficient. The sum

[8]Note that the model can be rewritten in levels as $y_{i,t} = \beta_{0,i} + (1 + \alpha_i) y_{i,t-1} + \sum_{k=1}^{K} \rho_{(k),i} \sum_{j=1,i\neq j}^{N} w_{(k),i,j} y_{j,t-1} + \epsilon_{i,t}$. The point estimates between the two models do not differ and the dependent variable occurs as a spatial time lag.

$\sum_{l=0}^{p_T} \vartheta_i \bar{y}_{i,t-l}$ represents all the lags of the cross-sectional averages and accounts for the common factors. The model can be written in matrix notation:

$$\Delta \mathbf{y}_t = \mathbf{b} + \mathbf{A}\mathbf{y}_{t-1} + \mathbf{u}_t, \tag{4.14}$$

with $\mathbf{u}_t$ containing the cross-sectional averages. The difference between equation (4.13) and equation (4.5) is that the latter allows for multiple spatial interactions, namely K spatial weights matrices. Equation (4.13) follows Arbia and Paelinck (2003a,b) extended by cross-sectional averages and has several insights. First, it is relatively easy to compute. Secondly, no country-specific effects have to be considered. The equation can be estimated for each country separately and the constant contains the country specific effects. This procedure avoids removing the country-specific effect by transforming the convergence equation into first differences, which would be then exposed to endogeneity.[9] The spatial time lag ensures that the term is exogenous, as neither $y_{i,t}$ nor $y_{i,t-1}$ are included in $y_{j,t-1}$. Endogeneity is not present as long as the errors are not autocorrelated and further explanatory variables are exogenous. Finally, it accounts for spatial dependence using the spatial time lag and common factors by the multifactor error structure, respectively the cross-sectional averages.

The combination of spatial dependence and common factors recently received attention in the literature and was summarised in Elhorst, Abreu, Amaral, Bhattacharjee, Corrado, Fingleton, Fuerst, Garretsen, Igliori, Le Gallo, McCann, Monastiriotis and Yu (2016). Ertur and Musolesi (2016) estimate total factor productivity with a multifactor error structure and spatial dependence in the form of geographical distance. Bailey, Holly and Pesaran (2016) develop a two-step estimator to control for spatial dependence and common factors. In the first step, the common factors are factored out. The weakly cross-sectional dependent observations are then used in the second step in a spatial model. Bailey, Holly and Pesaran (2016) apply it to a

[9] For a more detailed discussion on endogeneity in a empirical growth context, see Hauk and Wacziarg (2009); Gallo and Fingleton (2014)

model of house prices and estimation of the spatial weight matrix. Vega and Elhorst (2016) apply an extension of the method to a regional unemployment model.

The approach in this chapter differs in several ways from the empirical spatial growth models. First of all it takes both sources of cross-sectional dependence into account. The common factors capture unobservable shocks, which hit all countries at the same time.

Another difference is the inclusion of a spatial time rather than a contemporaneous spatial lag. Lopez-Bazo et al. (2004); Fingleton and López-Bazo (2006); Ertur and Koch (2007); Lesage and Fischer (2008); Elhorst et al. (2010) and Ertur and Koch (2011) use a spatial lag model and emphasise the importance to include a contemporaneous spatial lag. A notable exception in the context of spatial growth models is Ho, Wang and Yu (2013). They follow the approach from Ertur and Koch (2007) and estimate a spatial Solow model extended by a spatial time lag. A contemporaneous spatial lag is not possible with a Lotka-Volterra model. Estimating a spatial lag model requires methods that do not rely on OLS. The model is rewritten into a closed form solution with the inverse of an identity matrix minus the spatial weight matrix multiplied to the right hand side. Usually this model requires a maximum likelihood estimator (Lee, 2004; Yu et al., 2008) or a GMM estimator (Kelejian and Prucha, 1998, 1999) and cannot be used in a multifactor error structure. Thus, adding a contemporaneous spatial lag, would rule out the possibility of using the DCCE estimator to control for common factors. Because of the spatial time lag, the type of endogeneity in this chapter differs from the one mainly occurring in the literature. In a spatial lag model with an exogenous spatial weight matrix, reversed causality occurs as the observations of the dependent variable from other cross-sectional units are added as explanatory variables. This type of endogeneity in a spatial lag model is usually estimated with an IV or GMM approach (Kelejian and Prucha, 1998, 1999). In this chapter, endogeneity stems from the spatial weight matrix and thus the spatial time lag becomes endogenous.

A third difference to the works above and to Barro and Sala-i Martin (1992); Mankiw et al. (1992); Islam (1995) is the assumption of slope heterogeneity. In an empirical setting, parameter heterogeneity was discussed in Lee et al. (1997, 1998); Islam (1998) and Eberhardt and Teal (2011). Especially in a cross-country setting, the assumption of slope homogeneity is rather strong. Slope heterogeneity is a prerequisite for the Lotka-Volterra model as the model requires estimates for each cross-sectional unit.

Another difference is that the growth model in equation (4.5) does not include any further covariates, such as physical or human capital or a measure of the size of the population. In the case of a Lotka-Volterra model, further covariates would be required to be stable as well and would complicate the conditions for convergence.

The exclusion of a spatial lag and further covariates are important limitations of the Lotka-Volterra approach. Including those would complicate the estimation method as described above. The focus of this chapter is on the Lotka-Volterra approach in a cross-country setting with a multifactor error structure and controlling for both types of cross-sectional dependence. Therefore, the model and the estimation method are kept as simple as possible. The Lotka-Volterra model is not widely adopted in empirical economics. A possible explanation is, that testing the conditions for convergence is difficult as explained next.

4.6 Testing for absolute convergence and inference on eigenvalues

In order to state the type of convergence, the conditions for it have to be assessed. The country specific estimates are used to calculate the steady states for each country $\hat{y}_i^*$. Then a Wald Test with the hypothesis $H_0 : y_1^* = y_2^* = ... = y_N^*$ is performed to test for absolute convergence. If the null is rejected, the steady states are different from each other and no absolute convergence is found.

Testing for conditional convergence requires more care, as instead of the steady states, the negativity of the eigenvalues is tested. A vast literature deals with inference of eigenvalues of covariance matrices in principal component analysis. However, in contrast to inference on covariance matrices of principal components, the eigenvalues in this chapter are allowed to be negative. Hence, the large sample distribution theory as derived for example in Anderson (2003) is not valid.

An approximate distribution of the eigenvalues can be derived using the Delta Method. Under the assumption that the eigenvalues are a function of $\boldsymbol{\omega}$, i.e. $\boldsymbol{\lambda}(\boldsymbol{\omega}) = (\lambda_1, ..., \lambda_N)'$, with $\boldsymbol{\omega} = (\boldsymbol{\alpha}, \boldsymbol{\rho})$, $\omega_i = (\alpha_i, \rho_i)$ and $\boldsymbol{\Omega}$ the true value, then $\boldsymbol{\omega} \rightarrow_p \boldsymbol{\Omega}$ and $\sqrt{n}(\boldsymbol{\omega} - \boldsymbol{\Omega}) \rightarrow_d N(0, \boldsymbol{\Sigma})$.[10] The distribution of the eigenvalues using the delta method is then $\sqrt{(n)}(\boldsymbol{\lambda}(\boldsymbol{\omega}) - \boldsymbol{\lambda}(\boldsymbol{\Omega})) \rightarrow_d N(0, \boldsymbol{\Lambda}(\boldsymbol{\Omega})\boldsymbol{\Sigma}\boldsymbol{\Lambda}(\boldsymbol{\Omega})')$.[11] Then a possible hypothesis would be $H_0 : -2 < \lambda_i < 0 \; i = 1, ..., N$. However, this joint test would involve inequalities under the null hypothesis. As an alternative, the maximum of an eigenvalue can be tested. The maximum of eigenvalues follows an extreme value distribution. This approach is infeasible, because the moments of the extreme value distribution would rely on a single observation.

As an alternative a bootstrap is carried out, where the cross-sectional dimension (N) is fixed and $T \rightarrow \infty$. Moreover as this is a heterogeneous slope model, the unit-specific coefficients ω_i depend on the observed data $\mathbf{X} = (\mathbf{X}_1, ..., \mathbf{X}_T)$, implying that $\boldsymbol{\omega}$ is a function of the data, $\boldsymbol{\omega} = \boldsymbol{\omega}(\mathbf{X})$. Hence the eigenvalues are indirectly a function of the observed data as well, $\boldsymbol{\lambda}(\boldsymbol{\omega}(\mathbf{X}))$. For the bootstrap the following steps are performed:[12,13]

[10] Additionally the weight matrix $\mathbf{W}$ and their weights have to be uncorrelated with $\boldsymbol{\omega}$. $\boldsymbol{\Lambda}(\boldsymbol{\Omega})$ is the first derivative of the eigenvalues.

[11] See Hayashi (2000) Lemma 2.5.

[12] For an overview of bootstrap methods see Horowitz (2001) and for the block bootstrap, see Härdle, Horowitz and Kreiss (2003). A non-overlapping block bootstrap is preferred over the more complicated overlapping block bootstrap. Härdle et al. (2003) report that the numerical results of the two approaches are similar.

[13] The argumentation can be put differently: The community matrix depends on the estimated coefficients of the lagged dependent variable and the spatial autocorrelation coefficient. The Monte Carlo results in Chapter 3 proof, that the biases of the coefficient on the lagged dependent variable and on exogenous explanatory variables decrease with T. In addition, the sum of the spatial autocorrelation coefficient and the spatial time lag is weakly cross-sectional dependent. Both favour the cross-sectional dimension to be fixed and to draw from the time dimension.

1. $\mathbf{X} = (\mathbf{X}_1, ..., \mathbf{X}_T)$ is split into country specific blocks of length B_l.

2. A sample $\mathbf{X}^*$ is drawn with replacement.

3. Equation (4.13) is estimated, matrix A constructed and the eigenvalues computed.

4. The number of eigenvalues $-2 < \lambda_i < 0$ is stored in T_r.

5. Steps 1 - 4 are repeated R times

6. $\bar{T} = \frac{1}{R}\sum_{r=1}^{R} T_r$ is calculated.

Under the null hypothesis all eigenvalues are between -2 and 0, implying convergence, thus $H_0 : \bar{T} = N$, against the alternative $H_0 : \bar{T} < N$. The length of the blocks is 10 periods ($B_l = 10$) periods. It is alternated to 5 and 20 periods as a robustness check.

4.7 Data

GDP per capita originates from the Penn World Tables 8 (Feenstra et al., 2015) and is transformed into logarithms. The yearly data is restricted to the years 1960 to 2007. Even though the Penn World Tables are available until 2011, data from 2008 onwards is excluded due to the financial crisis. Using yearly observations in cross-country regressions is rather unusual. It is common to use 5-year averages (Islam, 1995; Caselli, Esquivel and Lefort, 1996; Hauk and Wacziarg, 2009). However, this will shorten the panel such that the DCCE estimator is not applicable. The reason for the period averages is to smooth short run effects such as the business cycle. However, those should be captured by the common factors of the DCCE estimator.[14]

Spatial Weights Three different row normalised spatial weight matrices are used in this study. The first one is based on migration streams of high skilled workers. The weight matrix is constructed out of data for 2010 from the IAB brain drain

[14]This topic is somewhat related to the questions of continuous-time models and issues of aggregation over time, which is well beyond the topic of this chapter and thesis.

dataset (Brücker, Capuano and Marfouk, 2013). The rationale is that high skilled workers contribute to the economic performance of a country and strengthen the ties between countries.[15] A positive spatial autocorrelation coefficient $\rho_{(1)}$ would imply that high skilled migrants are carriers of positive spillover effects. Equation (2.58) in Chapter 2 directly relates migration streams and economic growth.

Mayda (2010) finds using bilateral migration flows from the OECD International Migration Statistics that pull factors (income in destination country) increase the size of emigration rates. Barro and Sala-i Martin (2004) assume a smaller rate of convergence if the regression contains migration. They find a small negative effect of migration on income per capita growth rates.

On a regional level for 27 European regions, Huber and Tondl (2012) find evidence that migration has a positive effect on the income per capita of the destination country, while it negatively effects the country of origin. Ozgen, Nijkamp and Poot (2010) summarize the effect of migration on convergence on a regional level in form of a meta study.

The second spatial weights matrix are trade shares between countries, originating from the Direction of Trade Statistics (DOTS) of the IMF. The data ranges from 2009 to 2013 in yearly intervals. The latest available year is used as the baseline spatial weights matrix, as done for migration. Trade is commonly used as a spatial weight for economic distance (LeSage and Pace, 2008). If two countries trade a lot, spillovers between these two will be larger. A more theoretical rationale is that countries use their comparative advantage in the production of specific goods, which are then traded. Moreover, it is suggested in the literature (Grossman and Helpman, 1991; Coe and Helpman, 1995) that international trade is a major channel for diffusion of technology. Both theories strongly suggest an estimated spatial

[15] Only high skilled migration into OECD countries is taken into account, from OECD countries as well as from non-OECD countries. The first reason is that the amount of high skilled migrants into non OECD countries is likely to be very small. Secondly, data for streams between non OECD countries is largely unavailable. Alternatively, it would be possible to restrict the study to OECD countries, with the cost of losing 3/4 of the observations. A list of the countries can be found in the Appendix C.1.

autoregressive coefficient for the trade weight matrix larger than zero. In a growth empirical setting, trade as a spatial weights matrix to measure economic distance can be found for example in Ertur and Koch (2011) or Ho et al. (2013).

As a third weight matrix, FDI from the Coordinated Direct Investment Survey (CDIS) of the IMF is considered. The data spans the years 1980 until 2010 in 5 year intervals. In line with the other two spatial weights, the latest available year is used as a spatial weight. FDI represents technological diffusion. In a spatial context, it is surveyed in Abreu et al. (2004).

4.8 Estimations

In the next section, regression results are presented. First estimates obtained by the SDLS estimator are discussed and compared to earlier findings in the literature. Then regression results using the DCCE estimator are analysed.[16]

4.8.1 Simultaneous Dynamic Least Squares

SDLS results are presented in Table C.1. Panel A shows the results without any spatial weights. The p-values from the bootstrap with the hypothesis that all eigenvalues are negative are shown in squared brackets in the column labelled "Eigenval." The p-value for the test of absolute convergence is shown in parenthesis. In the last row of each panel, the spatial weight is removed from the computation of the community matrix **A** and the number of negative eigenvalues and the corresponding p-value of the bootstrap is shown.

The regression results for the classical model in Panel A are not in favour of conditional or absolute convergence, as the hypothesis of all eigenvalues being between -2 and 0, and all steady states being the same are rejected. However, the estimate of the coefficient on the lagged dependent variable is negative and significant. Therefore, in a classical sense, conditional convergence is found.

[16] The estimations are performed in Stata using the `xtdcce2` package (Ditzen, 2017), version 1.32 from July 2017.

Adding spatial weights changes the overall picture. All values for $\boldsymbol{\alpha}$ and the real parts of the eigenvalues of $\mathbf{A}$ are negative. In the case of migration and exports as spatial weights conditional convergence is found as all eigenvalues are negative. This evidence is supported by the bootstrap. Moreover the spatial autocorrelation coefficient using exports as a weight is significant at a level of 10%. This result is in line with the results from Arbia and Paelinck (2003a,b), who find convergence once spatial interactions are accounted for. The last panel includes all three spatial weight matrices. The bootstrap just rejects the hypothesis of conditional convergence. However not all eigenvalues are negative and none of the spatial weights are significant.

As the last column indicates, the hypothesis of spatial dependence in the residuals is rejected for the case without spatial lags and if the FDI migration weight matrix and all three weight matrices are used. This implies that OLS is inconsistent. Therefore, as a next step, the DCCE estimator is used to control for unobserved cross-sectional dependence together with spatial interactions to account for observed dependence.

4.8.2 Dynamic Common Correlated Effects - Mean Group

Table C.2 displays results using the DCCE estimator. Cross-sectional averages of the independent variable and $\left[\sqrt[2]{47}\right] = 3$ lags of it are added in Panels B - F.

Panel A shows results without any cross-sectional averages or spatial weights. The mean of the coefficient on the lagged dependent variable, α_i is negative and highly significant. However only 83 eigenvalues are between -2 and 0, indicating no conditional convergence. The bootstrapped p-value of the test that all eigenvalues are between -2 and 0 is 0.988, thus supporting the hypothesis of negative eigenvalues. The hypothesis of absolute convergence is rejected. The p-value of the CD test in the final column rejects the hypothesis of weak cross-sectional dependence, suggesting the estimates are inconsistent.

If cross-sectional averages are added (Panel B), the overall results appear to be similar. However, the hypothesis of weak cross-sectional dependence in the error term cannot be rejected. The value of the test statistic decreases from 31.55 to 0.25. The value implies that not much of cross-sectional dependence is left. Only one country has an eigenvalue not between -1 and 0, implying that most countries show a tendency towards convergence.

In the regressions in Panel C to Panel E a single spatial interaction is included. The spatial autocorrelation coefficient for migration and FDI is significant and positive while the coefficient on exports is insignificant. The significance of the coefficients implies that spatial dependence exists and underlines the importance of accounting for both types of dependence. The interpretation of the coefficients is, that the growth rate increases for countries that have strong ties via migration or FDI. For migration, the result is in contradiction to earlier findings as in Barro and Sala-i Martin (2004) who find a negative relationship. The number of eigenvalues between -2 and 0 does not change if the coefficients relating to the spatial interactions are removed from community matrix A. The column on the very right shows the value of the CD test statistic and the p-values. The hypothesis of weak cross-sectional dependence cannot be rejected in all three cases.

Panel F includes all three spatial weights. Out of the spatial interactions, only FDI has a significant positive effect. This result can be explained that only little cross-sectional dependence remains in Panel B and that the remaining dependence is split among the three channels. The number of eigenvalues between -2 and 0 is stable around 88, the bootstrapped p-values imply that the hypothesis of conditional convergence cannot be rejected. None of the regressions show any evidence for absolute convergence as the p-values are 0.

The estimated convergence equation is similar to the models estimated by Baumol (1986) and Barro and Sala-i Martin (1992). Baumol finds for a larger set of countries club convergence, but for a smaller absolute convergence. While the former fits loosely into conditional convergence, the latter is clearly contradicted by the results. Barro and Sala-i Martin find conditional convergence for 98 countries. In a more recent study, Barro (2015) finds conditional convergence in large panel as well. Therefore, the results show that a Lotka-Volterra approach can be placed into the existing literature.

4.9 Robustness Checks

Several robustness checks were carried out and lead to similar conclusions, confirming conditional convergence but no evidence for absolute convergence.

4.9.1 Dynamic Common Correlated Effects - Pooled

In Table C.3 the spatial autocorrelation coefficients are constrained to be equal across countries ($\rho_{k,i} = \rho_k,\ i = 1, ..., N$). The overall picture remains similar to the results from the DCCE-MG estimation. For comparison, Panel A and B are the same as in Table C.2. None of the three spatial autocorrelation coefficients is significant. The number of eigenvalues between -2 and 0 is between 88 and 90, the bootstrapped p-values imply however that all eigenvalues are between -2 and 0. The number of eigenvalues between -2 and 0 reacts only very little to changes in the interactions. This result is somewhat surprising. However taking the size of the coefficients into account and comparing it to the size of the coefficient on the lagged dependent variable α_i, the coefficients on ρ appear to be much smaller. The hypothesis of weak cross-sectional dependence cannot be rejected in all cases with spatial weights.

In the absence of a test for heterogeneous slopes, it only possible to speculate which model is the preferred one. Given the similarity of the Monte Carlo results in Chapter 3 and the significance of the MG results, the MG estimation results are preferred.

4.9.2 Dynamic Common Correlated Effects - IV estimation

Consistency of the DCCE estimator depends on the assumption that the variables are (weakly-) exogenous. However, if the dependent variable is an economic measure such as GDP and the spatial weight is a measure of 'economic distance' such as trade volumes, then the spatial weights are likely to be correlated with the dependent variable (Qu and Lee, 2015). The reversed causality leads to endogeneity of the spatial (time) lag. Kelejian and Piras (2014) propose regressing the spatial weights on exogenous variables. This implies that the weight matrix itself is estimated. This chapter follows an extension of this approach suggested by Qu and Lee (2015). A simple IV regression is performed on the sum of the weighted spatial autocorrelations. The first stage is

$$\begin{aligned} v_{i,t} = \sum_{j=1,j\neq i}^{N} w_{(1),i,j} y_{j,t-1} = \gamma_{0,i} + \gamma_{1,i} \sum_{j=1,j\neq i}^{N} w_{(2),i,j} y_{j,t-1} \\ + \gamma_{2,i} \sum_{j=1,j\neq i}^{N} w_{(3),i,j} y_{j,t-1} + \gamma_3 s_{i,t} + \xi_{i,t}, \end{aligned} \tag{4.15}$$

where $w_{(1),i,j}$ is the endogenous weight and $w_{(2),i,j}$ and $w_{(3),i,j}$ are the elements of the exogenous weight matrices. $v_{i,t}$ is the sum over the elements i, j of the spatial weight matrix times the spatial lag. $s_{i,t}$ are further exogenous covariates from the second stage, such as cross-sectional averages. $\hat{v}_i$ is then used in the second stage:

$$\Delta y_{i,t} = \beta_{0,i} + \alpha_i y_{i,t-1} + \sum_{k=1}^{K} \rho_{(k),i} \sum_{j=1,i\neq j}^{N} \hat{v}_{i,t} + \sum_{l=0}^{p_T} \vartheta_i \bar{y}_{i,t-l} + u_{i,t}. \tag{4.16}$$

The weight matrices of migration of high skilled workers, FDI and exports are instrumented by distance and distance squared. In equation (4.15), $w_{(1),i,j}$ contains either of the three matrices, $w_{(2),i,j}$ and $w_{(3),i,j}$ are the exogenous weights distance and distance squared.

The location of each country and its borders is taken from the World Borders

Dataset (Sandvik, 2008). The location of a country can be assumed to be exogenous as for example described in Eaton and Kortum (1999); Klenow and Rodriguez-Clare (2005).

In Table C.4 the spatial weight matrices are assumed to be endogenous. All three weight matrices are constructed using data from 2010 and therefore likely to be an outcome of the dependent variable. The coefficients on the migration and FDI spatial weights matrices are significant. The autocorrelation coefficient on the migration weights matrix is positive and significant. This implies that an influx of high skilled migrants increases the growth rate. From the results, it is evident that migration has a positive effect on the growth rate. The coefficient on direct investment is positive and significant at a level of 1%. More foreign direct investments imply an increase in the growth rate. As only two instruments are available, it is not possible to treat all three spatial weight matrices as endogenous at the same time. The bootstrapped p-values are in comparison to earlier results much smaller, but with the exception of the last two Panels, still in a non-rejection area. Again, there is no evidence for absolute convergence. The CD test for all regressions including spatial weights cannot reject the hypothesis of weak cross-sectional dependence.

The overall picture remains the same, FDI and to a lesser extent migration are the main drivers of spatial interactions. Both coefficients are positive and significant, but the coefficient on FDI is much larger. The coefficient on Exports remains insignificant and negative. While FDI clearly improves growth rates, exports have a dampening effect. The bootstrapped p-values testing the hypothesis that all eigenvalues are between -2 and 0 are in a non rejection region.

4.9.3 Alternating Number of Cross-Sectional Lags

In Tables C.5 to C.7 the number of cross-sectional averages and its lags are alternated.

No Cross-sectional Averages An OLS regression with heterogeneous slope coefficients is estimated in Table C.5. This is essentially a MG regression. All spatial autocorrelation coefficients appear to be positive and statistically significant different from zero at a level of 1%. The bootstrapped p-values indicate conditional convergence. However as the last column indicates, the test for weak cross-sectional dependence is rejected and therefore cross-sectional dependence occurs and renders the point estimates inconsistent. The results clearly show that accounting for spatial dependence by spatial weights is not sufficient to remove all spatial dependence.

Only contemporaneous averages In the next table, contemporaneous cross-sectional averages are added, but no lags. This setting would suffice a static panel model. Interestingly the MG estimates for the weight matrices turn negative and are with the exception of the autocorrelation coefficient on the migration weight matrix significant. The evidence for conditional convergence is even stronger as the p-values for the test on the negativity of the eigenvalues are larger. Similar to the case without cross-sectional averages, the CD test rejects the hypothesis of weak cross-sectional dependence in all cases. This implies that if only spatial dependence is accounted for, the common factors are still in the error terms. In addition, it is remarkable that the CD test statistic hardly reacts to the inclusion of the spatial terms.

One Lag Table C.7 shows the results of the same specification as in Table C.2, but with 1 instead of 3 lags of the cross-sectional averages ($p_T = 1$). As expected, the size of the CD test statistic increases and therefore the p-values decrease and are below 0.1 for all specifications but the one without any spatial weights. The reason for an increase in the CD test statistics is that, when leaving out the cross-sectional lags, the error terms are more exposed to unaccounted cross strong sectional dependence. The coefficients on the migration and FDI weight matrices are significant and positive, a result which is in line to earlier findings. The number of negative eigenvalues remains high and the results from the bootstrap imply conditional convergence.

Overall, the regressions in Tables C.5 and C.6 highlight several findings. First of all, they emphasise the sensitivity of the CD test to a change in the number of cross-sectional averages. Furthermore, it shows the importance to account for common factors by using a sufficient number of lags. Even though the MG estimates are significant for most cases, the estimates are inconsistent as the error terms inhibit strong cross-sectional dependence. The change of the sign of the spatial autocorrelation coefficients in the case of only contemporaneous cross-sectional averages can be interpreted in two ways. The first is due to the small sample bias in a finite sample. The second interpretation is that the cross-sectional averages take out some of the spatial dependence and render the coefficient to change its sign.

4.9.4 Lagged Spatial Weights

An important assumption is that the spatial weight matrix is exogenous. In the previous Tables the migration and export weight matrix dates from 2010, the FDI matrix from 2013. Similar to the argumentation in Section 4.9.2 the endogeneity of the spatial weights is questionable. As an alternative method to the one in Section 4.9.2 it is possible to use spatial weight matrices from an earlier point in time. Therefore, in Tables C.8 and C.9 the migration and export weights dates from 1980 and FDI from 2010.

Table C.8 is the analog to Table C.2 and with the same setting. Overall the MG estimations for $\boldsymbol{\alpha}$ are in similar regions. Most notable is that the coefficient on FDI almost halved and the one on exports increased in absolute value and is significant. Moreover it is negative and implying a dampening effect of exports on growth. The overall picture remains similar. Conditional convergence is found and the CD test statistics are in a non-rejection region.

In a similar fashion, Tables C.4 and C.9 can be compared. The level of the MG estimate for migration is similar. FDI becomes smaller and insignificant and exports have a strong negative effect. Notable is that the p-values for the test on conditional convergence are much smaller and almost lie in a non-rejection area. This implies that with the lagged spatial weight matrices, no conditional convergence is found.

4.9.5 US as common factor

To rule out any effects of dominating units on the spatial interactions, the US are removed from the weight matrix and modelled as a common factor in Table C.10. The approach is similar to Holly, Pesaran and Yamagata (2011) and Chudik and Pesaran (2013). None of the spatial autocorrelation coefficients is significant. This underlines the importance of the US for spatial dependence in terms of FDI, trade and migration. The bootstrapped p-values are all in a non-rejection region, implying conditional convergence. Again, there is no evidence for absolute convergence. Including spatial weights and the cross-section averages, leads to weakly cross-sectional dependent error terms. This might raise the question once more that the cross-sectional averages take out too much of the spatial dependence and lead to insignificant results for coefficients which are meant to capture spatial dependence.

4.9.6 Alternating length of bootstrap blocks

As a final robustness check, the length of the blocks for the bootstrap is alternated. If the blocks are small, the drawn data resembles the observed data closer, while with large blocks specific time periods can appear multiple times.

The results in Tables C.11 and C.12 are similar to those obtained by the baseline model in C.2. The p-values differ only in a very small order. The result implies that the length of the blocks has only a minor effect on the number of eigenvalues between -1 and 0.

4.10 Conclusion

This chapter studies convergence in a cross-country framework using a general Lotka-Volterra Model to detect the type of convergence. Two types of cross-sectional dependence are accounted for. Common factors are incorporated by adding cross-sectional averages of the dependent and independent variables. As a second type, spatial dependence is accounted for by using a spatial time lag with three different weight matrices. The model is estimated with two different types of estimators, a Simultaneous Dynamic Least Square estimator and the Dynamic Common Correlated Effects Estimator, in the mean group as well as in the pooled version.

The chapter extends the current literature manifold. First of all the approach by Arbia and Paelinck (2003a,b) is advanced by taking both types of cross-sectional dependence into account. The conditions for absolute and conditional convergence are tested directly and by using a bootstrap method. The type of convergence is identified using a difference equation system, rather than a single parameter. This is further novelty in cross-country regressions. Finally, the use of the DCCE estimator in combination with spatial weights to account for observed as well as unobserved (global) factors in a empirical growth model is new to the best of my knowledge. Using this rather simple estimation technique comes at the cost of one important limitation. A spatial time lag is used rather than a more general spatial lag model.

There are several notable findings. A Lotka-Volterra approach on a classical convergence equation without spatial interactions confirms earlier empirical findings in literature. Conditional convergence appears to be more frequent if both types of cross-sectional dependence are controlled for. This allows the conclusion that interactions between countries contribute towards convergence. This finding however has to taken with caution. Conditional convergence in the prevailing context implies that countries converge towards their own equilibrium. Secondly, this argument underlines the importance of considering spatial dependencies between countries in growth empirics. In particular FDI and Migration appear to play a strong role in connecting countries and somewhat surprisingly to a lesser extent trade. However this is in line with the theoretical model in Chapter 2. Finally, the non rejection of weak cross-sectional dependence, if controlled for by spatial interactions and cross-sectional averages, is a strong argument to account for those. To put it very succinctly, a growth regression without taking dependence between between countries into account is likely to be inconsistent.

There are some possible extensions to the estimation approach of this chapter. In a growth empirical setting, applying the model to estimate the rate of convergence and the question of club convergence is of interest. In particular the matrix environment of the Lotka-Volterra model has the potential to determine convergence clubs. This could shed further light on the blurry definition of conditional convergence. An interesting and easy extension would be to apply the model to a subset of countries, however this would come at a cost, the loss of many cross-sectional observations. Three further applications of the Lotka-Volterra model in combination with the cross-sectional dependence literature are possible. First, the spatial Solow model developed by Ertur and Koch (2007) with physical and human capital and population growth as explanatory variables can be estimated adding cross-sectional averages to control for strong dependence. Secondly, a contemporaneous spatial lag can be added. This would make the model more general, but the estimation procedure would be more complicated. Finally, the two-step estimation procedure proposed by Bailey, Holly and Pesaran (2016) and the GMM procedure developed by Kuersteiner and Prucha (2015) can be applied to the settings above. Further work is required to encompass the two strands into a Lotka-Volterra model, which is beyond the scope of this thesis and left for further research.

Appendix Chapter 4

C.1 Countries

Argentina, Australia, Austria, Bangladesh, Barbados, Belgium, Benin, Bolivia, Botswana, Brazil, Burundi, Cameroon, Canada, Central African Republic, Chile, China, Colombia, Republic of Congo, Costa Rica, Cote dÌvoire, Cyprus, Denmark, Dominican Republic, Ecuador, Egypt, El Salvador, Fiji, Finland, France, Gabon, Gambia, Germany, Ghana, Greece, Guatemala, Honduras, Hong Kong, Iceland, Indonesia, Iran, Ireland, Israel, Italy, Jamaica, Japan, Jordan, Kenya, Republic of Korea, Lesotho, Luxembourg, Malawi, Malaysia, Mali, Malta, Mauritania, Mauritius, Mexico, Morocco, Mozambique, Namibia, Nepal, Netherlands, New Zealand, Niger, Norway, Pakistan, Panama, Paraguay, Peru, Philippines, Portugal, Rwanda, Senegal, South Africa, Spain, Sri Lanka, Sweden, Switzerland, Syria, Taiwan, Tanzania, Thailand, Togo, Trinidad & Tobago, Tunisia, Turkey, Uganda, United Kingdom, United States, Uruguay, Venezuela, Zambia, Zimbabwe

C.2 Simultaneous Dynamic Least Square estimator

Arbia and Paelinck (2003a,b) estimate regional convergence for 119 European regions. They use an iterative SDLS, which minimizes the squared deviations between an observed and an endogenously computed value, to estimate the parameters of a Lotka-Volterra equation. The SDLS estimator is similar to SURE or a 3SLS estimator without endogenous variables. It is a generalised reduced form estimator, a ML estimator if $\boldsymbol{\epsilon} \sim N(0, \sigma^2 \mathbf{I})$ and consistent estimator under homoscedastic errors (see Griffith and Paelinck (2011, Chapter 11)). The equilibrium conditions for Lotka-Volterra models are then used to determine the existence of convergence. Arbia and Paelinck estimate the following equation:

$$\Delta y_{r,t} = a_r y_{r,t-1} + b_r y^*_{r,t-1} + c_r y^{**}_{r,t-1} + c_r + e_0 \tag{C.1}$$

$$\Delta \mathbf{y} = \mathbf{A}\mathbf{y} + \mathbf{c}$$

where $y^*_{r,t-1}$ and $y^{**}_{r,t-1}$ are spatial time lags of first and second order (in the spatial dimension) and $y_{r,t-1}$ lagged income. c_r and e_0 are regional specific effects. The second equation is the Lotka-Volterra representation of Eq. (C.1).

In a similar fashion, the same approach will be used in this chapter to estimate convergence among 93 countries during the years 1960 to 2007. In line with Arbia and Paelinck (2003b) the following steps will be performed:

First, the income variable is set relative to the overall average. If all countries move towards the same level of income, or converge in absolute terms, then the relative income in each country should converge to the same value.
Secondly, the following model is estimated for each country separately by simple OLS:

$$y_{i,t} - y_{i,t-1} = \alpha_i y_{i,t-1} + \sum_{k=1}^{K} \rho_{(k),i} \sum_{j=1, i \neq j}^{N} w_{(k),i,j} y_{j,t-1} \tag{C.2}$$

$$+ \beta_{0,i} + e_{i,0} + \epsilon_{i,t} \tag{C.3}$$

where

$$\boldsymbol{\beta}_i = (\beta_{1,i}, \beta_{2,i}, \beta_{3,i})' \tag{C.4}$$

and $e_{0,i}$ is a dummy for the initial difference. $w_{(1),i,j}, w_{(2),i,j}$ and $w_{(3),i,j}$ are spatial weights constructed out of migration streams, trade data and the inverse of the distance between countries. $\rho_{(1)}, \rho_{(2)}$ and $\rho_{(3)}$ are the spatial autoregressive parameters. The entire system can be expressed in matrices:

$$\Delta \mathbf{y}_t = \mathbf{diag}(\boldsymbol{\alpha})\mathbf{y}_{t-1} + \sum_{k=1}^{K} \mathbf{diag}(\boldsymbol{\rho}_{(k)})\mathbf{W}_{(k)}\mathbf{y}_{t-1} + \boldsymbol{\beta}_0 + \mathbf{e}_0 + \boldsymbol{\epsilon}_t \tag{C.5}$$

with

$$\mathbf{y}_t = (y_{1,t}, ..., y_{N,t})' \qquad \boldsymbol{\alpha} = (\alpha_1, ..., \alpha_N)' \tag{C.6}$$

$$\boldsymbol{\rho}_{(k)} = (\rho_{(k),i}, ..., \rho_{(k),N})' \qquad \mathbf{e}_0 = (e_{1,0}, ..., e_{N,0})' \tag{C.7}$$

$$\boldsymbol{\epsilon}_t = (\epsilon_{1,t}, ..., \epsilon_{N,t})' \tag{C.8}$$

$$\mathbf{W}_{(k)} = \begin{pmatrix} 0 & w_{(k),1,2} & \cdots & w_{(k),1,N} \\ w_{(k),2,1} & \ddots & & \vdots \\ \vdots & & \ddots & \vdots \\ w_{(k),N,1} & \cdots & \cdots & 0 \end{pmatrix} \tag{C.9}$$

The coefficients are saved as $\hat{\boldsymbol{\alpha}}^{(1)}, \hat{\boldsymbol{\beta}}^{(1)}, \hat{\boldsymbol{\rho}_{(k)}}^{(1)}, ...$ etc. for use in the next step. In the following a hat indicates an estimated value, while the superscript refers to the number of repetitions.

Thirdly, income in t is iteratively computed for each period using the estimated coefficients from step 1. For the first period, $y_{i,1}$, the dummy on the initial difference $e_{0,i}$ is used instead of the lagged left hand side variable:

$$\hat{\mathbf{y}}_1^{(1)} = \hat{\boldsymbol{\beta}}_0^{(1)} + \mathbf{e}_0^{(1)} \tag{C.10}$$

The income of the subsequent periods, $t > 1$, is then derived from:

$$\hat{\mathbf{y}}_t^{(1)} = \left(\mathbf{diag}(\hat{\boldsymbol{\alpha}})^{(1)} + 1\right)\hat{\mathbf{y}}_{t-1}^{(1)} + \sum_{k=1}^{K} \mathbf{diag}(\hat{\boldsymbol{\rho}}_{(k)})\mathbf{W}_{(k)}\hat{\mathbf{y}}_{t-1} \tag{C.11}$$

Finally, the first three steps are repeated until the coefficients converge. The final estimation results are put into a system of differential equations, which has in addition to Eq. (4.1) a time varying factor. Thus

$$\Delta\mathbf{y}_t = \mathbf{A}\mathbf{y}_{t-1} + \mathbf{b} + \epsilon_t \tag{C.12}$$

where

$$\mathbf{A} = \mathbf{diag}(\boldsymbol{\alpha}) + \sum_{k=1}^{K} \mathbf{diag}(\boldsymbol{\rho}_{(k)})\mathbf{W}_{(k)}$$

$$\mathbf{b} = \boldsymbol{\beta}_0 + \mathbf{e}_0$$

The matrix $\mathbf{A}$ and vector $\mathbf{b}$ are then used to determine the type of convergence. The diagonal elements of $\mathbf{A}$ are the autocorrelation coefficients α_i, while on the off diagonal elements the sum of the spatial interactions ($\sum_{k=1}^{K} \rho_{(k),i} w_{(k),i,j}$) appear. This implies that that the diagonal represents the effects over time from within a country, while the off diagonal elements stand for cross-country or spatial interactions.

Equation (4.13) allows for different convergence equations. If all spatial interactions and possible explanatory variables are omitted ($\rho_{(k),i} = \beta_i = 0,\ i = 1, ..., N$ and $k = 1, ..., K$), then the classical convergence equation (Barro and Sala-i Martin, 1992) is obtained. If only the spatial interactions are set to zero ($\rho_{(k),i} = 0,\ k = 1, ..., K$), then the model is closer to the work of Mankiw et al. (1992). A spatial convergence model is obtained if $\rho_{(k),i} \neq 0,\ k = 1, ..., K$.

C.3 Tables

Statistic	α_i	$\rho_{1,i}$	$\rho_{2,i}$	$\rho_{3,i}$	$\beta_{0,i}$	Eigenval.	CD Test, p-val
Panel A: No weights							
Mean	-0.0344***				0.0344***	-0.0344	5.6036
SE	0.0026				0.0098	0.0249	0.0000
$\# < 0^{\dagger}$	84				63	84 [0.0000] (0.00)	
Panel C: Migration Weight Matrix							
Mean	-0.6365***	0.1218			0.5114	-0.6365	-1.2999
SE	0.0116	0.1042			0.5900	0.1065	0.1936
$\# < 0^{\dagger}$	93	7			9	93 [1.0000] (0.00)	
$\# -2 < \lambda_i < 0^{\dagger\dagger}$		93 [1.0000]					
Panel D: FDI Weight Matrix							
Mean	-0.0936***		-0.1097		0.2313	-0.0936	6.5453
SE	0.0132		0.1870		0.2483	0.1282	0.0000
$\# < 0^{\dagger}$	75		35		74	76 [0.0000] (0.00)	
$\# -2 < \lambda_i < 0^{\dagger\dagger}$			75 [0.0000]				
Panel E: Exports Weight Matrix							
Mean	-0.9752***			0.0213*	0.9525	-0.9752	0.2940
SE	0.0194			0.0149	0.9261	0.1874	0.7688
$\# < 0^{\dagger}$	93			17	33	93 [1.0000] (0.00)	
$\# -2 < \lambda_i < 0^{\dagger\dagger}$				93 [1.0000]			
Panel F: Migration, Exports and FDI Matrices							
Mean	-0.6964***	-0.4819	0.6428	-0.3328	0.8427	-0.6964	12.0440
SE	0.0386	0.7105	0.6906	0.3071	0.7986	0.3704	0.0000
$\# < 0^{\dagger}$	90	12	46	27	47	90 [0.0540] (0.00)	
$\# -2 < \lambda_i < 0^{\dagger\dagger}$		90 [0.0540]	90 [0.0000]	91 [0.0000]			

Table C.1: Simultaneous Dynamic Least Squares Estimation
Row *Mean* displays the mean group estimates, i.e. $\alpha = \frac{1}{N}\sum_{i=1}^{N}\alpha_i$. The last row in each panel counts how often the individual coefficients (eg.: α_i, $i = 1, ..., N$) are smaller than zero. The three steps are 50 times repeated to ensure convergence of the coefficients. For a more detailed discussion, see Appendix C.2. Significance levels: $^{*}p < 0.1$, $^{**}p < 0.05$, $^{***}p < 0.01$.

Statistic	α_i	$\rho_{1,i}$	$\rho_{2,i}$	$\rho_{3,i}$	$\beta_{0,i}$	Eigenval.	CD Test, p-val
Panel A: No weights							
Mean	-0.0466***				0.3882***	-0.0466	31.5479
SE	0.0054				0.0384	0.0520	0.0000
$\# < 0^{\dagger}$	20				0	83 [0.9860] (0.00)	
Panel B: No weights but with cross section averages (CCE)							
Mean	-0.1411***				0.0249	-0.1411	0.2534
SE	0.0116				0.1714	0.1116	0.7999
$\# < 0^{\dagger}$	34				13	90 [1.0000] (0.00)	
Panel C: Migration Weight Matrix and CCE							
Mean	-0.1513***	0.0311**			0.0862	-0.1513	0.5497
SE	0.0124	0.0154			0.1760	0.1197	0.5825
$\# < 0^{\dagger}$	32	0			12	89 [1.0000] (0.00)	
$\# -2 < \lambda_i < 0^{\dagger\dagger}$		89 [1.0000]					
Panel D: FDI Weight Matrix and CCE							
Mean	-0.1863***		0.1002***		0.2277	-0.1863	0.0243
SE	0.0152		0.0401		0.2252	0.1521	0.9806
$\# < 0^{\dagger}$	36		5		8	88 [1.0000] (0.00)	
$\# -2 < \lambda_i < 0^{\dagger\dagger}$			88 [1.0000]				
Panel E: Exports Weight Matrix and CCE							
Mean	-0.1851***			-0.0054	0.2258	-0.1851	0.9099
SE	0.0142			0.0533	0.2127	0.1377	0.3629
$\# < 0^{\dagger}$	35			13	9	91 [0.9960] (0.00)	
$\# -2 < \lambda_i < 0^{\dagger\dagger}$				91 [0.9960]			
Panel F: Migration, Exports and FDI Matrices and CCE							
Mean	-0.2300***	-0.0127	0.0887*	0.0273	0.1816	-0.2300	0.1731
SE	0.0170	0.0385	0.0590	0.0796	0.2912	0.1725	0.8626
$\# < 0^{\dagger}$	41	0	8	13	6	88 [0.9980] (0.00)	
$\# -2 < \lambda_i < 0^{\dagger\dagger}$		90 [0.9960]	89 [0.9980]	88 [0.9920]			

Table C.2: Dynamic Common Correlated Effects Estimation, $p_T = 3$ Row *Mean* displays the mean group estimates, i.e. $\bar{\alpha} = \frac{1}{N}\sum_{i=1}^{N}\alpha_i$ for $\bar{\alpha}$ and $\bar{\beta}_0$, and row *SE* its standard error. Row $\# < 0$ counts how often the individual coefficients are significantly smaller than zero at a level of 5% $H_0 : \alpha_i < 0$. † Bootstrapped values for the test of conditional convergence are in squared brackets, p-value for test of absolute convergence ($H_0 : y_1^* = ... = y_N^*$) in parenthesis. †† Spatial Weight removed. For a more detailed discussion, see Section 4.4. Significance levels: $^{*}p < 0.1$, $^{**}p < 0.05$, $^{***}p < 0.01$.

Statistic	α_i	$\rho_{1,i}$	$\rho_{2,i}$	$\rho_{3,i}$	$\beta_{0,i}$	Eigenval.	CD Test, p-val
Panel A: No weights							
Mean	-0.0466***				0.3882***	-0.0466	31.5479
SE	0.0054				0.0384	0.0520	0.0000
$\# < 0^{\dagger}$	20				0	83 [0.9860] (0.00)	
Panel B: No weights but with cross section averages (CCE)							
Mean	-0.1411***				0.0249	-0.1411	0.2534
SE	0.0116				0.1714	0.1116	0.7999
$\# < 0^{\dagger}$	34				13	90 [1.0000] (0.00)	
Panel C: Migration Weight Matrix and CCE							
Mean	-0.1444***	0.0649			0.0613	-0.1444	0.5417
SE	0.0114	0.0616			0.1705	0.1102	0.5881
$\# < 0^{\dagger}$	34	0			13	90 [1.0000] (0.00)	
$\# -2 < \lambda_i < 0^{\dagger\dagger}$		90 [1.0000]					
Panel D: FDI Weight Matrix and CCE							
Mean	-0.1519***		0.0535		0.1760	-0.1519	0.8194
SE	0.0115		0.0635		0.1696	0.1118	0.4125
$\# < 0^{\dagger}$	36		0		9	88 [1.0000] (0.00)	
$\# -2 < \lambda_i < 0^{\dagger\dagger}$			88 [1.0000]				
Panel E: Exports Weight Matrix and CCE							
Mean	-0.1413***			0.0018	0.0313	-0.1413	0.2654
SE	0.0116			0.0620	0.1713	0.1115	0.7907
$\# < 0^{\dagger}$	33			0	11	90 [1.0000] (0.00)	
$\# -2 < \lambda_i < 0^{\dagger\dagger}$				90 [1.0000]			
Panel F: Migration, Exports and FDI Matrices and CCE							
Mean	-0.1527***	0.0482	0.0606	-0.0343	0.1047	-0.1527	0.7837
SE	0.0116	0.1259	0.0876	0.0938	0.1706	0.1124	0.4332
$\# < 0^{\dagger}$	38	0	0	0	11	88 [1.0000] (0.00)	
$\# -2 < \lambda_i < 0^{\dagger\dagger}$		88 [1.0000]	88 [1.0000]	88 [1.0000]			

Table C.3: Dynamic Common Correlated Effects Pooled Estimation $p_T = 3$
All spatial autocorrelation coefficients are constrained to be the same across all unites ($\rho_i = \rho,\ i = 1, ..., N$).
See notes Table C.2.

Statistic	α_i	$\rho_{1,i}$	$\rho_{2,i}$	$\rho_{3,i}$	$\beta_{0,i}$	Eigenval.	CD Test, p-val
Panel A: No weights							
Mean	-0.0466***				0.3882***	-0.0466	31.5479
SE	0.0054				0.0384	0.0520	0.0000
$\# < 0^{\dagger}$	20				0	83 [0.9860] (0.00)	
Panel B: No weights but with cross section averages (CCE)							
Mean	-0.1411***				0.0249	-0.1411	0.2534
SE	0.0116				0.1714	0.1116	0.7999
$\# < 0^{\dagger}$	34				13	90 [1.0000] (0.00)	
Panel C: Migration Weight Matrix and CCE							
Mean	-0.1442***	0.0351*			0.1674	-0.1442	0.6538
SE	0.0154	0.0230			0.1909	0.1482	0.5133
$\# < 0^{\dagger}$	35	0			13	87 [0.2800] (0.00)	
$\# -2 < \lambda_i < 0^{\dagger\dagger}$		87 [0.9180]					
Panel D: FDI Weight Matrix and CCE							
Mean	-0.1363***		0.2469***		-0.2091	-0.1363	1.4086
SE	0.0529		0.0915		0.4309	0.5573	0.1590
$\# < 0^{\dagger}$	35		7		8	83 [0.0040] (0.00)	
$\# -2 < \lambda_i < 0^{\dagger\dagger}$			83 [0.2320]				
Panel E: Exports Weight Matrix and CCE							
Mean	-0.1931***			-0.0861	-0.0419	-0.1931	1.4399
SE	0.0217			0.1619	0.7519	0.2013	0.1499
$\# < 0^{\dagger}$	34			11	8	82 [0.0440] (0.00)	
$\# -2 < \lambda_i < 0^{\dagger\dagger}$				84 [0.3720]			

Table C.4: Dynamic Common Correlated Effects IV Estimation $p_T = 3$
Spatial weights are instrumented by distance and squared distance.
See notes Table C.2.

Statistic	α_i	$\rho_{1,i}$	$\rho_{2,i}$	$\rho_{3,i}$	$\beta_{0,i}$	Eigenval.	CD Test, p-val
Panel A: No weights							
Mean	-0.0466***				0.3882***	-0.0466	31.5479
SE	0.0054				0.0384	0.0520	0.0000
$\# < 0^{\dagger}$	20				0	83 [0.9860] (0.00)	
Panel C: Migration Weight Matrix and CCE							
Mean	-0.0639***	0.0185***			0.3890***	-0.0639	31.8716
SE	0.0063	0.0052			0.0406	0.0610	0.0000
$\# < 0^{\dagger}$	21	0			0	85 [0.9840] (0.00)	
$\# -2 < \lambda_i < 0^{\dagger\dagger}$		85 [0.9840]					
Panel D: FDI Weight Matrix and CCE							
Mean	-0.1238***		0.0657***		0.4393***	-0.1238	32.3911
SE	0.0100		0.0149		0.1131	0.0996	0.0000
$\# < 0^{\dagger}$	26		6		2	91 [0.9960] (0.00)	
$\# -2 < \lambda_i < 0^{\dagger\dagger}$			91 [0.9960]				
Panel E: Exports Weight Matrix and CCE							
Mean	-0.1246***			0.0576***	0.5065***	-0.1246	32.4542
SE	0.0105			0.0169	0.1270	0.1027	0.0000
$\# < 0^{\dagger}$	27			7	3	92 [0.9960] (0.00)	
$\# -2 < \lambda_i < 0^{\dagger\dagger}$				92 [0.9960]			
Panel F: Migration, Exports and FDI Matrices and CCE							
Mean	-0.1567***	0.0008	0.0013	0.1194*	0.1957	-0.1567	32.1247
SE	0.0132	0.0357	0.0543	0.0733	0.2712	0.1307	0.0000
$\# < 0^{\dagger}$	25	1	6	6	4	89 [1.0000] (0.00)	
$\# -2 < \lambda_i < 0^{\dagger\dagger}$		88 [0.9980]	89 [0.9480]	89 [0.9440]			

Table C.5: Dynamic Common Correlated Effects Estimation with no cross-sectional averages
All spatial autocorrelation coefficients are constrained to be the same across all unites $(\rho_i = \rho,\ i = 1, ..., N)$.
See notes Table C.2.

Statistic	α_i	$\rho_{1,i}$	$\rho_{2,i}$	$\rho_{3,i}$	$\beta_{0,i}$	Eigenval.	CD Test, p-val
Panel A: No weights							
Mean	-0.0466***				0.3882***	-0.0466	31.5479
SE	0.0054				0.0384	0.0520	0.0000
$\# < 0^{\dagger}$	20				0	83 [0.9860] (0.00)	
Panel B: No weights but with cross section averages (CCE)							
Mean	-0.1435***				0.0020	-0.1435	22.0185
SE	0.0115				0.2081	0.1111	0.0000
$\# < 0^{\dagger}$	33				10	92 [1.0000] (0.00)	
Panel C: Migration Weight Matrix and CCE							
Mean	-0.1437***	-0.0097			-0.0228	-0.1437	21.1757
SE	0.0118	0.0107			0.2111	0.1138	0.0000
$\# < 0^{\dagger}$	31	1			10	92 [1.0000] (0.00)	
$\# -2 < \lambda_i < 0^{\dagger\dagger}$		91 [1.0000]					
Panel D: FDI Weight Matrix and CCE							
Mean	-0.1630***		-0.0481**		-0.2487	-0.1630	17.5179
SE	0.0130		0.0285		0.2790	0.1274	0.0000
$\# < 0^{\dagger}$	32		10		12	91 [1.0000] (0.00)	
$\# -2 < \lambda_i < 0^{\dagger\dagger}$			91 [1.0000]				
Panel E: Exports Weight Matrix and CCE							
Mean	-0.1653***			-0.1116***	-0.2772	-0.1653	16.9437
SE	0.0137			0.0382	0.2439	0.1331	0.0000
$\# < 0^{\dagger}$	31			10	11	90 [1.0000] (0.00)	
$\# -2 < \lambda_i < 0^{\dagger\dagger}$				90 [1.0000]			
Panel F: Migration, Exports and FDI Matrices and CCE							
Mean	-0.1969***	-0.0343	-0.0600	-0.0015	-0.5297*	-0.1969	12.8127
SE	0.0166	0.0347	0.0540	0.0761	0.3397	0.1625	0.0000
$\# < 0^{\dagger}$	32	1	9	11	9	89 [1.0000] (0.00)	
$\# -2 < \lambda_i < 0^{\dagger\dagger}$		87 [1.0000]	87 [0.9980]	86 [1.0000]			

Table C.6: Dynamic Common Correlated Effects Estimation with contemporaneous cross-sectional averages $p_T = 0$ All spatial autocorrelation coefficients are constrained to be the same across all unites ($\rho_i = \rho,\ i = 1, ..., N$). See notes Table C.2.

Statistic	α_i	$\rho_{1,i}$	$\rho_{2,i}$	$\rho_{3,i}$	$\beta_{0,i}$	Eigenval.	CD Test, p-val
Panel A: No weights							
Mean	-0.0466***				0.3882***	-0.0466	31.5479
SE	0.0054				0.0384	0.0520	0.0000
$\# < 0^{\dagger}$	20				0	83 [0.9860] (0.00)	
Panel B: No weights but with cross section averages (CCE)							
Mean	-0.1318***				-0.0057	-0.1318	1.5571
SE	0.0109				0.1972	0.1047	0.1195
$\# < 0^{\dagger}$	29				9	93 [1.0000] (0.00)	
Panel C: Migration Weight Matrix and CCE							
Mean	-0.1439***	0.0318**			0.0496	-0.1439	1.9596
SE	0.0114	0.0158			0.2025	0.1110	0.0500
$\# < 0^{\dagger}$	29	0			8	93 [1.0000] (0.00)	
$\# -2 < \lambda_i < 0^{\dagger\dagger}$		93 [1.0000]					
Panel D: FDI Weight Matrix and CCE							
Mean	-0.1713***		0.0618**		-0.0102	-0.1713	1.6612
SE	0.0135		0.0343		0.2871	0.1345	0.0967
$\# < 0^{\dagger}$	31		5		6	90 [0.9980] (0.00)	
$\# -2 < \lambda_i < 0^{\dagger\dagger}$			90 [0.9980]				
Panel E: Exports Weight Matrix and CCE							
Mean	-0.1697***			0.0040	0.1094	-0.1697	2.3107
SE	0.0137			0.0438	0.2402	0.1333	0.0209
$\# < 0^{\dagger}$	26			9	6	89 [1.0000] (0.00)	
$\# -2 < \lambda_i < 0^{\dagger\dagger}$				89 [0.9980]			
Panel F: Migration, Exports and FDI Matrices and CCE							
Mean	-0.2108***	0.0008	0.0122	0.1030	-0.1484	-0.2108	2.1680
SE	0.0160	0.0427	0.0657	0.1032	0.3728	0.1611	0.0302
$\# < 0^{\dagger}$	32	0	7	10	6	87 [1.0000] (0.00)	
$\# -2 < \lambda_i < 0^{\dagger\dagger}$		89 [1.0000]	88 [0.9980]	88 [1.0000]			

Table C.7: Dynamic Common Correlated Effects Estimation with contemporaneous and one lag of the cross-sectional averages $p_T = 1$
All spatial autocorrelation coefficients are constrained to be the same across all unites ($\rho_i = \rho,\ i = 1, ..., N$).
See notes Table C.2.

Statistic	α_i	$\rho_{1,i}$	$\rho_{2,i}$	$\rho_{3,i}$	$\beta_{0,i}$	Eigenval.	CD Test, p-val
Panel A: No weights							
Mean	-0.0466***				0.3882***	-0.0466	31.5479
SE	0.0054				0.0384	0.0520	0.0000
$\# < 0^{\dagger}$	20				0	83 [0.9860] (0.00)	
Panel B: No weights but with cross section averages (CCE)							
Mean	-0.1411***				0.0249	-0.1411	0.2534
SE	0.0116				0.1714	0.1116	0.7999
$\# < 0^{\dagger}$	34				13	90 [1.0000] (0.00)	
Panel C: Migration Weight Matrix and CCE							
Mean	-0.1558***	0.0294**			0.0702	-0.1558	0.4474
SE	0.0128	0.0153			0.1766	0.1237	0.6546
$\# < 0^{\dagger}$	32	0			13	89 [1.0000] (0.00)	
$\# -2 < \lambda_i < 0^{\dagger\dagger}$		89 [1.0000]					
Panel D: FDI Weight Matrix and CCE							
Mean	-0.1853***		0.0575*		0.0250	-0.1853	0.8562
SE	0.0157		0.0386		0.1961	0.1567	0.3919
$\# < 0^{\dagger}$	36		13		9	85 [1.0000] (0.00)	
$\# -2 < \lambda_i < 0^{\dagger\dagger}$			86 [0.9980]				
Panel E: Exports Weight Matrix and CCE							
Mean	-0.1818***			-0.0834*	0.0452	-0.1818	0.5089
SE	0.0138			0.0514	0.1789	0.1340	0.6109
$\# < 0^{\dagger}$	35			11	13	89 [1.0000] (0.00)	
$\# -2 < \lambda_i < 0^{\dagger\dagger}$				89 [1.0000]			
Panel F: Migration, Exports and FDI Matrices and CCE							
Mean	-0.2311***	-0.0111	-0.0258	-0.0168	0.0298	-0.2311	0.5738
SE	0.0176	0.0324	0.0650	0.0980	0.2274	0.1764	0.5661
$\# < 0^{\dagger}$	40	1	11	16	9	87 [0.9960] (0.00)	
$\# -2 < \lambda_i < 0^{\dagger\dagger}$		87 [0.9800]	86 [0.9620]	85 [0.9420]			

Table C.8: Dynamic Common Correlated Effects with lagged spatial weight matrices and $p_T = 3$
All spatial autocorrelation coefficients are constrained to be the same across all unites ($\rho_i = \rho,\ i = 1, ..., N$).
See notes Table C.2.

Statistic	α_i	$\rho_{1,i}$	$\rho_{2,i}$	$\rho_{3,i}$	$\beta_{0,i}$	Eigenval.	CD Test, p-val
Panel A: No weights							
Mean	-0.0466***				0.3882***	-0.0466	31.5479
SE	0.0054				0.0384	0.0520	0.0000
$\# < 0^{\dagger}$	20				0	83 [0.9860] (0.00)	
Panel B: No weights but with cross section averages (CCE)							
Mean	-0.1411***				0.0249	-0.1411	0.2534
SE	0.0116				0.1714	0.1116	0.7999
$\# < 0^{\dagger}$	34				13	90 [1.0000] (0.00)	
Panel C: Migration Weight Matrix and CCE							
Mean	-0.1452***	0.0259*			0.1119	-0.1452	0.3127
SE	0.0149	0.0197			0.1826	0.1455	0.7545
$\# < 0^{\dagger}$	35	0			13	86 [0.1780] (0.00)	
$\# -2 < \lambda_i < 0^{\dagger\dagger}$		85 [0.9420]					
Panel D: FDI Weight Matrix and CCE							
Mean	-0.1825***		0.0577		-0.0834	-0.1825	1.2840
SE	0.0207		0.0574		0.2705	0.1888	0.1992
$\# < 0^{\dagger}$	32		12		10	83 [0.0120] (0.00)	
$\# -2 < \lambda_i < 0^{\dagger\dagger}$			83 [0.3180]				
Panel E: Exports Weight Matrix and CCE							
Mean	-0.1858***			-0.2184**	-0.3173	-0.1858	0.5189
SE	0.0174			0.1106	0.3688	0.1722	0.6038
$\# < 0^{\dagger}$	33			13	13	87 [0.0540] (0.30)	
$\# -2 < \lambda_i < 0^{\dagger\dagger}$				89 [0.3880]			

Table C.9: Dynamic Common Correlated Effects IV Estimation with lagged spatial weight matrices and$p_T = 3$ All spatial autocorrelation coefficients are constrained to be the same across all unites $(\rho_i = \rho,\ i = 1, ..., N)$. See notes Table C.2.

Statistic	α_i	$\rho_{1,i}$	$\rho_{2,i}$	$\rho_{3,i}$	$\beta_{0,i}$	Eigenval.	CD Test, p-val
Panel A: No weights							
Mean	-0.0466***				0.3882***	-0.0466	31.5479
SE	0.0054				0.0384	0.0520	0.0000
$\# < 0^{\dagger}$	20				0	83 [0.9860] (0.00)	
Panel B: No weights but with cross section averages (CCE)							
Mean	-0.1411***				0.0249	-0.1411	0.2534
SE	0.0116				0.1714	0.1116	0.7999
$\# < 0^{\dagger}$	34				13	90 [1.0000] (0.00)	
Panel C: Migration Weight Matrix and CCE							
Mean	-0.1661***	0.0087			0.0442	-0.1661	0.3929
SE	0.0129	0.0157			0.1785	0.1254	0.6944
$\# < 0^{\dagger}$	32	0			12	90 [1.0000] (0.00)	
$\# -2 < \lambda_i < 0^{\dagger\dagger}$		90 [1.0000]					
Panel D: FDI Weight Matrix and CCE							
Mean	-0.2030***		0.0290		0.1661	-0.2030	-0.4635
SE	0.0157		0.0564		0.2330	0.1552	0.6430
$\# < 0^{\dagger}$	37		7		8	88 [0.9980] (0.00)	
$\# -2 < \lambda_i < 0^{\dagger\dagger}$			88 [0.9980]				
Panel E: Exports Weight Matrix and CCE							
Mean	-0.1950***			-0.0563	0.1819	-0.1950	-0.2218
SE	0.0161			0.0707	0.2875	0.1581	0.8245
$\# < 0^{\dagger}$	33			13	10	89 [0.9980] (0.00)	
$\# -2 < \lambda_i < 0^{\dagger\dagger}$				89 [0.9980]			
Panel F: Migration, Exports and FDI Matrices and CCE							
Mean	-0.2395***	0.0129	0.0506	-0.0297	0.1871	-0.2395	-0.3409
SE	0.0184	0.0308	0.0579	0.0910	0.3345	0.1877	0.7332
$\# < 0^{\dagger}$	36	0	7	9	3	88 [0.9940] (0.00)	
$\# -2 < \lambda_i < 0^{\dagger\dagger}$		88 [0.9960]	88 [0.9960]	87 [0.9960]			

Table C.10: Dynamic Common Correlated Effects Estimation with dominant factor and $p_T = 3$
USA is taken out of spatial weight matrix and added as an additional common factor. See notes Table C.2.

Statistic	α_i	$\rho_{1,i}$	$\rho_{2,i}$	$\rho_{3,i}$	$\beta_{0,i}$	Eigenval.	CD Test, p-val
Panel A: No weights							
Mean	-0.0466***				0.3882***	-0.0466	31.5479
SE	0.0054				0.0384	0.0520	0.0000
$\# < 0^{\dagger}$	20				0	83 [0.9860] (0.00)	
Panel B: No weights but with cross section averages (CCE)							
Mean	-0.1411***				0.0249	-0.1411	0.2534
SE	0.0116				0.1714	0.1116	0.7999
$\# < 0^{\dagger}$	34				13	90 [1.0000] (0.00)	
Panel C: Migration Weight Matrix and CCE							
Mean	-0.1513***	0.0311**			0.0862	-0.1513	0.5497
SE	0.0124	0.0154			0.1760	0.1197	0.5825
$\# < 0^{\dagger}$	32	0			12	89 [1.0000] (0.00)	
$\# -2 < \lambda_i < 0^{\dagger\dagger}$		89 [1.0000]					
Panel D: FDI Weight Matrix and CCE							
Mean	-0.1863***		0.1002***		0.2277	-0.1863	0.0243
SE	0.0152		0.0401		0.2252	0.1521	0.9806
$\# < 0^{\dagger}$	36		5		8	88 [1.0000] (0.00)	
$\# -2 < \lambda_i < 0^{\dagger\dagger}$			88 [1.0000]				
Panel E: Exports Weight Matrix and CCE							
Mean	-0.1851***			-0.0054	0.2258	-0.1851	0.9099
SE	0.0142			0.0533	0.2127	0.1377	0.3629
$\# < 0^{\dagger}$	35			13	9	91 [0.9960] (0.00)	
$\# -2 < \lambda_i < 0^{\dagger\dagger}$				91 [0.9960]			
Panel F: Migration, Exports and FDI Matrices and CCE							
Mean	-0.2300***	-0.0127	0.0887*	0.0273	0.1816	-0.2300	0.1731
SE	0.0170	0.0385	0.0590	0.0796	0.2912	0.1725	0.8626
$\# < 0^{\dagger}$	41	0	8	13	6	88 [0.9980] (0.00)	
$\# -2 < \lambda_i < 0^{\dagger\dagger}$		90 [0.9960]	89 [0.9980]	88 [0.9920]			

Table C.11: Dynamic Common Correlated Effects Estimation $p_T = 3$
Bootstrap with 5 year blocks rather than 10 year blocks. See notes Table C.2.

Statistic	α_i	$\rho_{1,i}$	$\rho_{2,i}$	$\rho_{3,i}$	$\beta_{0,i}$	Eigenval.	CD Test, p-val
Panel A: No weights							
Mean	-0.0466***				0.3882***	-0.0466	31.5479
SE	0.0054				0.0384	0.0520	0.0000
$\# < 0^{\dagger}$	20				0	83 [0.9860] (0.00)	
Panel B: No weights but with cross section averages (CCE)							
Mean	-0.1411***				0.0249	-0.1411	0.2534
SE	0.0116				0.1714	0.1116	0.7999
$\# < 0^{\dagger}$	34				13	90 [1.0000] (0.00)	
Panel C: Migration Weight Matrix and CCE							
Mean	-0.1513***	0.0311**			0.0862	-0.1513	0.5497
SE	0.0124	0.0154			0.1760	0.1197	0.5825
$\# < 0^{\dagger}$	32	0			12	89 [1.0000] (0.00)	
$\# -2 < \lambda_i < 0^{\dagger\dagger}$		89 [1.0000]					
Panel D: FDI Weight Matrix and CCE							
Mean	-0.1863***		0.1002***		0.2277	-0.1863	0.0243
SE	0.0152		0.0401		0.2252	0.1521	0.9806
$\# < 0^{\dagger}$	36		5		8	88 [1.0000] (0.00)	
$\# -2 < \lambda_i < 0^{\dagger\dagger}$			88 [1.0000]				
Panel E: Exports Weight Matrix and CCE							
Mean	-0.1851***			-0.0054	0.2258	-0.1851	0.9099
SE	0.0142			0.0533	0.2127	0.1377	0.3629
$\# < 0^{\dagger}$	35			13	9	91 [0.9960] (0.00)	
$\# -2 < \lambda_i < 0^{\dagger\dagger}$				91 [0.9960]			
Panel F: Migration, Exports and FDI Matrices and CCE							
Mean	-0.2300***	-0.0127	0.0887*	0.0273	0.1816	-0.2300	0.1731
SE	0.0170	0.0385	0.0590	0.0796	0.2912	0.1725	0.8626
$\# < 0^{\dagger}$	41	0	8	13	6	88 [0.9980] (0.00)	
$\# -2 < \lambda_i < 0^{\dagger\dagger}$		90 [0.9960]	89 [0.9980]	88 [0.9920]			

Table C.12: Dynamic Common Correlated Effects Estimation $p_T = 3$
Bootstrap with 20 year blocks rather than 10 year blocks. See notes Table C.2.

Chapter 5

Conclusion

This thesis has studied the fields of applied research in economic growth and the connection to interdependence between countries. Economic theory recognises interdependence across countries, but in applied empirical works those interdependence are neglected. Econometric and spatial econometric theory identify two types of interdependence between countries, strong and weak cross-sectional dependence. Strong dependence is modelled in form of unobservable common factors. Spatial econometrics offers methods to account for weak cross-sectional dependence, or spatial dependence. Therefore, to obtain consistent estimates for a growth regression, an appropriate method needs to pay attention to both sources of interdependence. The main message of a growth regression in the long run interpretation is convergence. Conditions for it need to be adjusted if interactions between countries are included in a growth model.

One of the aims of this thesis is to close the gap between the theoretical and the applied growth literature on the implementation of interdependences between countries. Chapter 2 derives a growth model by combining an endogenous Romer model with features from the New Economic Geography and Growth literature. Countries interact via three different channels: migration of high skilled workers, diffusion of ideas and trade of goods. The countries' growth rate depends positively on all three factors. Especially migration of high skilled workers has a large effect on the growth rate and is founded on the micro level. Wage differentials are the motivation for high skilled workers to migrate. The engine of growth is taken from the endogenous Romer growth model. Therefore, the model predicts explosive growth and divergence across countries. Left for further research is to change the production function of the R&D sector to decreasing returns of scale. A possible outcome of this would be convergence.

Before it is possible to use the model from above as motivation for interdependence between countries for an empirical model, it is important to discuss the second type of cross-sectional dependence, common factors. In the literature it is standard to use the Common Correlated Effects estimators for the estimation of the multi-factor error structure models. Chapter 3 describes a Stata package, which allows the estimation of the Dynamic Common Correlated Effects estimator. The package allows for the estimation of Mean Group and pooled models. It accounts for common factors by adding contemporaneous values and lags of cross-sectional averages, following Pesaran (2006) and Chudik and Pesaran (2015a). The chapter discusses the integrated routine to test for weak cross-sectional dependence and the estimation strategy for a unit-specific fixed effect. Monte Carlo simulations point towards the requirements for the time and cross-sectional dimension the estimator can be applied to. The main driver for the bias of the coefficient of the lagged dependent variable is the time series bias. The coefficient is biased in small samples and the bias decreases only with a reasonable time series length of around 100 periods. The bias for coefficients of further exogenous explanatory variables remains low, even for small samples.

The final chapter draws on the preceding chapters. A growth equation with both types of cross-sectional dependence is estimated. The spatial weight matrices, migration of high skilled workers, trade and foreign direct investments, are motivated by the model from Chapter 2. The estimation procedure follows on the lines of the method described in Chapter 3. To determine the type of convergence a general Lotka-Volterra model is employed. Evidence for conditional convergence is found, especially if cross-sectional dependence is controlled for. FDI and migration play a strong role for the interdependence between countries. The robustness checks underline the importance to control for both types of dependence, otherwise the regression results are endangered to be invalidated. To put it succinctly, results from growth regression without interdependence between the units are likely to be inconsistent. The results highlight the potential use of a general Lotka-Volterra model, respectively of a differential equation model, for establishing convergence in the presence of interactions between countries.

`xtdcce2` is an ongoing project. On the top of the agenda are improvements of the speed of the estimations, new features such as the CS-DL and CS-ARDL estimator (Chudik et al., 2016) and alternative variance/covariance estimators for pooled coefficients and individual coefficients in a mean group regression. Further extensions can incorporate the bias correction methods for the pooled estimator suggested in Everaert and De Groote (2016) and Everaert and De Vos (2016). The literature on cross-sectional dependence increased in the last years. However, the implementation of the newly developed methods into programs such as Stata lags behind. This excludes many researchers from using the methods in applied research. Left in this field is for example the two-stage approach described in Bailey, Holly and Pesaran (2016). `xtdcce2` gives the researcher the possibility to estimate the sum of the common factors, with or without the constant. Estimating the factor loadings would be beneficial in several ways. First of all, the common factors are of use for example in stochastic frontier models (Filippini and Tosetti, 2014) or estimation of gravity models (Serlenga et al., 2013). A key for the estimation of the factors is to determine the number of factors. Sarafidis and Wansbeek (2012) propose methods to estimate the number of the common factors. From there the question arises, if the cross-sectional averages are suitable to approximate the common factors for an estimation of the factor loadings. Important is a test for slope homogeneity. To the best of my knowledge, no such test for a panel with large N and T exists. Developing a test would be highly beneficial for applied econometrics. If the common factors are estimated, then a test on the lines of Ando and Bai (2015) can be derived.

The model in Chapter 4 has several limitations. The model relies on a spatial time rather than a spatial lag. A spatial lag requires more sophisticated methods. The Lotka-Volterra approach is in the focus and therefore a simple method preferred. In addition to a spatial lag, further explanatory variables are missing. Including them would require further conditions for convergence. This can go in hand with a Lotka-Volterra representation of a neoclassical growth model as the Solow model. However the question of the meaning of conditional convergence remains. It does not make a statement about the persistence of income differences. Of interest would be the identification of convergence clubs and the Lotka-Volterra approach might be of use. There are alternatives to the difference equation Lotka-Volterra model. One alternative for the pair-wise approach to test for convergence is described in Pesaran (2007). However, it is not obvious how the test would perform in an equation with spatial dependence or common factors. Another limitation is a clear econometric guidance when to include spatial lags to control for spatial dependence. The CD test captures strong dependence, but is not be able to detect independence. A method to specify this could be along the lines of estimating α_{CSD} as proposed in Bailey, Kapetanios and Pesaran (2016).

Bibliography

Abreu, Maria, Henri L. F. De Groot and Raymond J. G. M. Florax. 2004. "Space and growth: a survey of empirical evidence and methods." Tinbergen Institute Working Paper TI 04-129.

Acemoglu, Daron and Simon Johnson. 2007. "Disease and Development: The Effect of Life Expectancy on Economic Growth." Journal of Political Economy 115(6):925–985.

Acemoglu, Daron, Simon Johnson and James A. Robinson. 2001. "The Colonial Origins of Comparative Development: An Empirical Investigation." The American Economic Review 91(5):1969–1401.

Alvarez, Fernando E., Francisco J. Buera and Robert E. Lucas. 2013. "Idea Flows, Economic Growth, and Trade." NBER Working Paper 19667.

Anderson, Theodore W. 2003. An Introduction to Multivariate Statistical Analysis. 3rd ed. Hoboken, New Jersey: John Wiley & Sons.

Ando, Tomohiro and Jushan Bai. 2015. "A simple new test for slope homogeneity in panel data models with interactive effects." Economics Letters 136:112–117.

Anselin, Luc. 1988. Spatial Econometrics: Methods and Models. Dordrecht, Boston, London: Kluwer Academic Publishers.

Anselin, Luc, Julie Le Gallo and Hubert Jayet. 2008. "Spatial Panel Econometrics." Advanced Studies in Theoretical and Applied Econometrics 46:625–660.

Arbia, Giuseppe and Jean H. P. Paelinck. 2003a. "Economic convergence or divergence? Modeling the interregional dynamics of EU regions, 1985-1999." Journal of Geographical Systems 5(3):291–314.

Arbia, Giuseppe and Jean H. P. Paelinck. 2003b. "Spatial Econometric Modeling of Regional Convergence in Continuous Time." International Regional Science Review 26(3):342–362.

Bai, Jushan. 2009. "Panel Data Models With Interactive Fixed Effects." Econometrica 77(4):1229–1279.

Bailey, Natalia, G. Kapetanios and M. Hashem Pesaran. 2016. "Exponent of Cross-Sectional Dependence: Estimation and Inference." Journal of Applied Econometrics 31:929–960.

Bailey, Natalia, Sean Holly and M. Hashem Pesaran. 2016. "A Two-Stage Approach to Spatio-Temporal Analysis with Strong and Weak Cross-Sectional Dependence." Journal of Applied Econometrics 31(1):249–280.

Baldwin, Richard E and Philippe Martin. 2004. Agglomeration and Regional Growth. In Handbook of Regional and Urban Economics: Cities and Geography, ed. Jacques Olivier and Jacques-Francois Thisse. 4 ed. Elsevier chapter 60, pp. 2671–2711.

Baldwin, Richard E. and Rikard Forslid. 2000. "The Core - Periphery Model and Endogenous Growth : Stabilizing and Destabilizing Integration." Economica 67(267):307–324.

Barro, Robert J. 1991. "Economic growth in a cross section of countries." Quarterly Journal of Economics 106(2):407.

Barro, Robert J. 2001. "Human capital and growth." The American Economic Review 91(2):12–17.

Barro, Robert J. 2015. "Convergence and modernisation." Economic Journal 125(585):911–942.

Barro, Robert J. and Xavier Sala-i Martin. 1992. "Convergence." Journal of Political Economy 100(2):223–251.

Barro, Robert J and Xavier Sala-i Martin. 2004. Economic growth. 2nd ed. Cambridge, Massachusetts: The MIT Press.

Basu, Parantap and Keshab Bhattarai. 2012. "Cognitive Skills, Openness and Growth." Economic Record 88(280):18–38.

Baum, Christopher F., Mark E. Schaffer and Steven Stillman. 2003. "Instrumental variables and GMM: Estimation and testing." Stata Journal 1(3):1–31.

Baum, Christopher F, Mark E. Schaffer and Steven Stillman. 2007. "Enhanced routines for instrumental variables/generalized method of moments estimation and testing." Stata Journal 7(4):465–506.

Baumol, William J. 1986. "Productivity growth, convergence, and welfare: what the long-run data show." The American Economic Review 76(5):1072–1085.

Ben-Gad, Michael. 2004. "The economic effects of immigration - a dynamic analysis." Journal of Economic Dynamics and Control 28(9):1825–1845.

Benhabib, Jess and Mark M. Spiegel. 1994. "The role of human capital in economic development evidence from aggregate cross-country data." Journal of Monetary Economics 34(2):153–173.

Bhattacharjee, Arnab and Chris Jensen-Butler. 2013. "Estimation of the spatial weights matrix under structural constraints." Regional Science and Urban Economics 43(4):617–634.

Bhattacharjee, Arnab and Sean Holly. 2013. "Understanding Interactions in Social Networks and Committees." Spatial Economic Analysis 8(1):23–53.

Bhattacharjee, Arnab, Taps Maiti, Eduardo Castro and Joao Marques. 2016. "Endogenous Spatial Regression and Delineation of Submarkets: A New Framework with Application to Housing Markets." Journal of Applied Econometrics 31(1):32–57.

Blackburne, Edward F. and Mark W. Frank. 2007. "Estimation of nonstationary heterogeneous panels." The Stata Journal 7(2):197–208.

Bond-Smith, Steven and Philip McCann. 2014. Incorporating Space in the Theory of Endogenous Growth: Contributions from the New Economic Geography. In Handbook of Regional Science, ed. Michael Wegener and Peter Nijkamp. Berlin Heidelberg: Springer-Verlag chapter 12, pp. 741–758.

Bond, Stephen R. and Markus Eberhardt. 2013. "Accounting for unobserved heterogeneity in panel time series models." Unpublished mimeo University of Oxford.

Borjas, GJ. 1994. "The economics of immigration." Journal of Economic Literature 32(4):1667–1717.

Braun, Juan. 1993. Essays on Economic Growth and Migration , phd thesis Harvard University.

Brock, William A. and Steven N. Durlauf. 2000. "Growth economics and reality." NBER Working Paper 8041.

Broda, C. and D. E. Weinstein. 2006. "Globalization and the Gains From Variety." The Quarterly Journal of Economics 121(2):541–585.

Brücker, Herbert, Stella Capuano and Abdeslam Marfouk. 2013. "Education, gender and international migration: insights from a panel-dataset 1980-2010.".
URL: *http://www.iab.de/en/daten/iab-brain-drain-data.aspx*

Caselli, Francesco, Gerardo Esquivel and Fernando Lefort. 1996. "Reopening the convergence debate: a new look at cross-country growth empirics." Journal of Economic Growth 1(3):363–389.

Cerina, Fabio and Francesco Pigliaru. 2007. Agglomeration and Growth in the NEG: a critical assessment. In New Directions in Economic Geography, ed. Bernard Fingleton. Edward Elgar Publishing chapter 5.

Chudik, Alexander, Kamiar Mohaddes, M. Hashem Pesaran and Mehdi Raissi. 2016. Long-Run Effects in Large Heterogeneous Panel Data Models with Cross-Sectionally Correlated Errors. In Essays in Honor of Aman Ullah (Advances in Econometrics, Volumn 36), ed. R. Carter Hill, Gloria GonzÁlez-Rivera and Tae-Hwy Lee. Emerald Group Publishing Limited pp. 85–135.

Chudik, Alexander and M. Hashem Pesaran. 2013. "Econometric Analysis of High Dimensional VARs Featuring a Dominant Unit." Econometric Reviews 32(5-6):592–649.

Chudik, Alexander and M. Hashem Pesaran. 2015a. "Common correlated effects estimation of heterogeneous dynamic panel data models with weakly exogenous regressors." Journal of Econometrics 188(2):393–420.

Chudik, Alexander and M. Hashem Pesaran. 2015b. Large Panel Data Models with Cross-Sectional Dependence: A Survey. In The Oxford Handbook Of Panel Data, ed. Badi H. Baltagi. Oxford: Oxford University Press chapter 1, pp. 2–45.

Chudik, Alexander, M. Hashem Pesaran and Elisa Tosetti. 2011. "Weak and strong cross-section dependence and estimation of large panels." The Econometrics Journal 14(1):C45–C90.

Clemens, Michael A., Claudio E. Montenegro and Lant Pritchett. 2009. "The place premium: wage differences for identical workers across the US border." Serie Documentos de Trabajo SDT 321 Universidad de Chile.

Coakley, Jerry, Ana-María Fuertes and Ron Smith. 2002. "A principal components approach to cross-section dependence in panels." Unpublished mimeo 2002.

Coe, David T. and Elhanan Helpman. 1995. "International R&D spillovers." European Economic Review 39(5):859–887.

Comin, Diego and Bart Hobijn. 2010. "An Exploration of Technology Diffusion." The American Economic Review 100(5):2031–2059.

Corrado, Luisa and Bernard Fingleton. 2012. "Where Is the Economics in Spatial Econometrics?" Journal of Regional Science 52(2):210–239.

De Hoyos, Rafael E. and Vasilis Sarafidis. 2006. "Testing for cross-sectional dependence in panel-data models." The Stata Journal 6(4):482–496.

De Long, J. Bradford and Lawrence Summers. 1991. "Equipment Investment and Economic Growth." The Quarterly Journal of Economics 106(2):445–502.

Dequiedt, Vianney and Yves Zenou. 2013. "International migration, imperfect information, and brain drain." Journal of Development Economics 101:117–132.

Ditzen, Jan. 2017. "XTDCCE2: Stata module to estimate heterogeneous coefficient models using common correlated effects in a dynamic panel.".
URL: *https://ideas.repec.org/c/boc/bocode/s458204.html*

Ditzen, Jan. 2018. "Cross Country Convergence in a General Lotka-Volterra Model." Spatial Economic Analysis 13(2):191–211.

Ditzen, Jan and Erich Gundlach. 2016. "A Monte Carlo study of the BE estimator for growth regressions." Empirical Economics 51(1):31–55.

Dixit, Avinash K. and Joseph E. Stiglitz. 1977. "Monopolistic competition and optimum product diversity." The American Economic Review 67(3):297–308.

Docquier, Frédéric and Hillel Rapoport. 2012. "Globalization, brain drain, and development." Journal of Economic Literature 50(3):681–730.

Durlauf, Steven N. 2001. "Manifesto for a growth econometrics." Journal of Econometrics 100(1):65–69.

Durlauf, Steven N., Paul A. Johnson and Jonathan R. W. Temple. 2005. Growth econometrics. In Handbook of Economic Growth, ed. Philippe Aghion and Steven N. Durlauf. Volume 1 ed. chapter 8, pp. 555–677.

Eaton, Jonathan and Samuel Kortum. 1999. "International technology diffusion: Theory and measurement." International Economic Review 40(3).

Eberhardt, Markus. 2012. “Estimating panel time series models with heterogeneous slopes.” The Stata Journal 12(1):61–71.

Eberhardt, Markus, Christian Helmers and Hubert Strauss. 2012. “Do Spillovers Matter When Estimating Private Returns to R&D?” Review of Economics and Statistics 95(2):436 – 448.

Eberhardt, Markus and Francis Teal. 2011. “Econometrics for Grumblers: A New Look At the Literature on Cross-Country Growth Empirics.” Journal of Economic Surveys 25(1):109–155.

Eberhardt, Markus and Francis Teal. 2017. “The magnitude of the task ahead: Productivity analysis with heterogeneous technology.” Globalisation and Economic Policy (GEP) GEP 17/16.

Elhorst, J. Paul. 2010. “Applied Spatial Econometrics: Raising the Bar.” Spatial Economic Analysis 5(1):9–28.

Elhorst, P., M. Abreu, P. Amaral, A. Bhattacharjee, L. Corrado, B. Fingleton, F. Fuerst, H. Garretsen, D. Igliori, J. Le Gallo, P. McCann, V. Monastiriotis and J. Yu. 2016. “Raising the bar (4).” Spatial Economic Analysis 11(4):355–360.

Elhorst, Paul, Gianfranco Piras and Giuseppe Arbia. 2010. “Growth and Convergence in a Multiregional Model with Space-Time Dynamics.” Geographical Analysis 42(3):338–355.

Ertur, Cem and Antonio Musolesi. 2016. “Weak and Strong Cross-Sectional Dependence: A Panel Data Analysis of International Technology Diffusion.” Journal of Applied Econometrics 47(4):36–37.

Ertur, Cem and Wilfried Koch. 2007. “Growth, Technological Interdependence and Spatial Externalities: Theory and Evidence.” Journal of Applied Econometrics 22(6):1033–1062.

Ertur, Cem and Wilfried Koch. 2011. "A contribution to the theory and empirics of Schumpeterian growth with worldwide interactions." Journal of Economic Growth 16(3):215–255.

Everaert, Gerdie and Ignace De Vos. 2016. "Bias-corrected Common Correlated Effects Pooled estimation in homogeneous dynamic panels." Working Paper 2016/920.

Everaert, Gerdie and Tom De Groote. 2016. "Common Correlated Effects Estimation of Dynamic Panels with Cross-Sectional Dependence." Econometric Reviews 35(3):1–31.

Feenstra, Robert C. 1994. "New Product Varieties and the Measurement of International Prices." The American Economic Review 84(1):157–77.

Feenstra, Robert C. 1996. "Trade and uneven growth." Journal of Development Economics 49(1):229–256.

Feenstra, Robert C., Robert Inklaar and Marcel Timmer. 2015. "The Next Generation of the Penn World Table." The American Economic Review 105(10):3150–82. **URL:** *www.ggdc.net/pwt*

Fernald, By John G and Charles I Jones. 2014. "The Future of US Economic Growth." AEA Papers and Proceedings 104(5):44–49.

Filippini, Massimo and Elisa Tosetti. 2014. "Stocastic Frontier Models for Long Panel Data Sets: Measurement of the Underlying Energy Eciency for the OECD Countries." 13/198 Economics Working Paper Series ETH Zurich.

Fingleton, Bernard and Enrique López-Bazo. 2006. "Empirical growth models with spatial effects." Papers in Regional Science 85(2):177–198.

Forslid, Rikard and Gianmarco I. P. Ottaviano. 2003. "An analytically solvable core-periphery model." Journal of Economic Geography 3(3):229–240.

Frankel, Jeffrey A. and David Romer. 1999. "Does trade cause growth?" The American Economic Review 89(3):379–399.

Gallo, Julie Le and Bernard Fingleton. 2014. Regional Growth and Convergence Empirics. In Handbook of Regional Science, ed. Manfred M. Fischer and Peter Nijkamp. Berlin Heidelberg: Springer chapter 16, pp. 291–315.

Galor, Oded. 2007. Discrete Dynamical Systems. Berlin, Heidelberg: Springer-Verlag.

Giacomini, Raffaella and Clive W.J. Granger. 2004. "Aggregation of space-time processes." Journal of Econometrics 118(1-2):7–26.

Goodwin, Richard. 1967. A growth cycle. In Socialism, capitalism, and economic growth, ed. Carl Feinstein. Cambridge University Press.

Griffith, Daniel A. and Jean H. Paul Paelinck. 2011. A Mixed Linear-Logarithmic Specification for Lotka-Volterra Models with Endogenously Generated SDLS-Variables. In Non-standard Spatial Statistics and Spatial Econometrics. 1 ed. Advances in Geographic Information Science Berlin, Heidelberg: Springer chapter 12, pp. 179–187.

Grossman, Gene M. and Elhanan Helpman. 1990. "Comparative Advantage and Long-Run Growth." The American Economic Review 80(4):796–815.

Grossman, Gene M. and Elhanan Helpman. 1991. Innovation and Growth in the Global Economy. Cambridge, MA: The MIT Press.

Grossman, Gene M. and Elhanan Helpman. 2015. "Globalization and Growth." The American Economic Review 105(5):100–104.

Grossman, Gene M. and Elhanan Helpman. 2016. "Growth, Trade, and Inequality." mimeo.

Grossmann, Volker and David Stadelmann. 2013. “Wage effects of high-skilled migration: international evidence.” World Bank Policy Research Working Paper WPS6317.

Gundlach, Erich and Martin Paldam. 2016. “Socioeconomic transitions as common dynamic processes.” Economics Working Papers Department of Economics and Business Economics, Aarhus University.

Hall, Robert E. and Charles I. Jones. 1999. “Why do some countries produce so much more output per worker than others?” The Quarterly Journal of Economics 114(1):83–116.

Härdle, Wolfgang, Joel Horowitz and Jens-Peter Kreiss. 2003. “Bootstrap Methods for Time Series.” International Statistical Review 71(2):435–459.

Hauk, William R. and Romain Wacziarg. 2009. “A Monte Carlo study of growth regressions.” Journal of Economic Growth 14(2):103–147.

Hayashi, Fumio. 2000. Econometrics. Princeton: Princeton University Press.

Herrendorf, Berhold, Richard Rogerson and Akos Valentinyu. 2014. Growth and Structural Transformation. In Handbook of Economic Growth, ed. Philippe Aghion and Steven Durlauf. 1 ed. Vol. 2 Elsevier chapter 6, pp. 855–941.

Ho, Chun-Yu, Wei Wang and Jihai Yu. 2013. “Growth spillover through trade: A spatial dynamic panel data approach.” Economics Letters 120(3):450–453.

Holly, Sean, M. Hashem Pesaran and Takashi Yamagata. 2010. “A spatio-temporal model of house prices in the USA.” Journal of Econometrics 158(1):160–173.

Holly, Sean, M. Hashem Pesaran and Takashi Yamagata. 2011. “The spatial and temporal diffusion of house prices in the UK.” Journal of Urban Economics 69(1):2–23.

Horowitz, Joel L. 2001. The Bootstrap. In Handbook of Econometrics, ed. James J Heckman and Edward Leamer. Vol. 5 Amsterdam: chapter 52, pp. 143–144.

Huber, Peter and Gabriele Tondl. 2012. "Migration and regional convergence in the European Union." Empirica 39(4):439–460.

Islam, Nazrul. 1995. "Growth Empirics: A Panel Data Approach." The Quarterly Journal of Economics 110(4):1127 – 1170.

Islam, Nazrul. 1998. "Growth empirics: a panel data approach - A Comment." The Quarterly Journal of Economics 113(1):325–329.

Islam, Nazrul. 2003. "What have We Learnt from the Convergence Debate?" Journal of Economic Surveys 17(3):309–362.

Jann, Ben. 2004. "estout: Stata module to export estimation results from estimates table.".
URL: *http://ideas.repec.org/c/boc/bocode/s439301.html*

Jann, Ben. 2005. "moremata: Stata module (Mata) to provide various functions.".
URL: *https://ideas.repec.org/c/boc/bocode/s455001.html*

Jewkes, John, David Sawers and Richard Stillerman. 1961. The Sources of Invention. 2 ed. Palgrave Macmillan UK.

Jones, Charles I. 1995. "R & D-based models of economic growth." Journal of Political Economy 103(4):759–784.

Jones, Charles I. 2005. Growth and Ideas. In Handbook of Economic Growth, ed. Philippe Aghion and Steven N. Durlauf. Vol. 1B Elsevier B.V. chapter 16, pp. 1063–1111.

Jones, Charles I. 2016. The Facts of Economic Growth. In Handbook of Macroeconomics, ed. John B. Taylor and Harald Uhlig. Vol. 2A Elsevier B.V. chapter 1, pp. 3–69.

Jones, Charles I. and Paul M. Romer. 2010. "The New Kaldor Facts: Ideas, Institutions, Population, and Human Capital." American Economic Journal: Macroeconomics 2(1):224–245.

Kapetanios, G., M. Hashem Pesaran and T. Yamagata. 2011. "Panels with non-stationary multifactor error structures." Journal of Econometrics 160(2):326–348.

Kelejian, Harry H. and Gianfranco Piras. 2014. "Estimation of spatial models with endogenous weighting matrices, and an application to a demand model for cigarettes." Regional Science and Urban Economics 46:140–149.

Kelejian, Harry H. and Ingmar R. Prucha. 1998. "A generalized spatial two-stage least squares procedure for estimating a spatial autoregressive model with autoregressive disturbances." The Journal of Real Estate Finance and Economics 17(1):99–121.

Kelejian, Harry H. and Ingmar R. Prucha. 1999. "A generalized moments estimator for the autoregressive parameter in a spatial model." International Economic Review 40(2):509–536.

Kelejian, Harry H. and Ingmar R. Prucha. 2010. "Specification and estimation of spatial autoregressive models with autoregressive and heteroskedastic disturbances." Journal of Econometrics 157(1):53–67.

Kennan, John. 2013. "Open borders." Review of Economic Dynamics 16(2):L1–L13.

Klein, Paul and Gustavo Ventura. 2009. "Productivity differences and the dynamic effects of labor movements." Journal of Monetary Economics 56(8):1059–1073.

Klenow, Peter J. and Andres Rodriguez-Clare. 2005. Externalities and growth. In Handbook of economic growth, ed. Philippe Aghion and Steven N. Durlauf. 1a ed. Elsevier B.V. chapter 11, pp. 818–861.

Koch, Wilfried. 2008. "Development Accounting with Spatial Effects." Spatial Economic Analysis 3(3):321–342.

Krugman, Paul. 1991. "Increasing Returns and Economic Geography." Journal of Political Economy 99(3):483–499.

Kuersteiner, Guido M and Ingmar R Prucha. 2015. "Dynamic Spatial Panel Models : Networks , Common Shocks , and Sequential Exogeneity." CESifo Working Paper Series (5445):1 – 40.

Lee, Kevin, M. Hashem Pesaran and Ron Smith. 1997. "Growth and Convergence in a Multi-Country Empirical Stochastic Solow Model." Journal of Applied Economics 12(4):357–392.

Lee, Kevin, M. Hashem Pesaran and Ron Smith. 1998. "Growth Empirics: A Panel Data Approach - A Comment." The Quarterly Journal of Economics 113(1):319–323.

Lee, Lung-Fei. 2004. "Asymptotic Distributions of Quasi-Maximum Likelihood Estimators for Spatial Autoregressive Models." Econometric Theory 72(6):1899–1925.

Lee, Lung-Fei and Jihai Yu. 2015. Spatial Panel Data Models. In The Oxford Handbook Of Panel Data, ed. Badi H. Baltagi. Oxford University Press chapter 12, pp. 363–401.

Lesage, James P. and Manfred M. Fischer. 2008. "Spatial Growth Regressions: Model Specification, Estimation and Interpretation." Spatial Economic Analysis 3(3):275–304.

LeSage, James P. and R. Kelley Pace. 2008. "Spatial econometric modeling of origin-destination flows." Journal of Regional Science 48(5):941–967.

Lopez-Bazo, Enrique, Esther Vaya and Manuel Artis. 2004. "Regional Externalities and Growth : Evidence From European Regions." Journal of Regional Science 44(1):43–73.

Lotka, Alfred J. 1920. "Analytical Note on Certain Rhythmic Relations in Organic Systems." Proceedings of the National Academy of Sciences of the United States of America 6(7):410–415.

Lucas, Robert E. 1988. "On the mechanics of economic development." Journal of Monetary Economics 22(1):3–42.

Lucas, Robert E. 2009a. "Ideas and Growth." Economica 76(301):1–19.

Lucas, Robert E. 2009b. "Trade and the Diffusion of the Industrial Revolution." American Economic Journal: Macroeconomics 1(1):1–25.

Maddison, Angus. 2007. Contours Of The World Economy, 1-2030 AD. 1st ed. Oxford University Press.

Mankiw, N Gregory, David Romer and David N Weil. 1992. "A Contribution to the Empirics of Economic Growth." The Quarterly Journal of Economics 107(2):407–437.

Mayda, Anna Maria. 2010. "International migration: A panel data analysis of the determinants of bilateral flows." Journal of Population Economics 23(4):1249–1274.

McNabb, Kyle and Philippe LeMay-Boucher. 2014. "Tax Structures, Economic Growth and Development." ICTD Working Paper 22 International Centre for Tax and Development.

Munro, Sir Hugh. 1891. "Table giving all the Scottish mountains exceeding 3000 feet in height, compiled by Munro." S.M.C. Journal 1(6).

Murray, James D. 2002. Mathematical Biology : I . An Introduction. 3rd ed. New York: Springer-Verlag.

Nickell, Stephen. 1981. "Biases in Dynamic Models with Fixed Effects." Econometrica 49(6):1417–1426.

Ortega, Francesco and Giovanni Peri. 2014. "Openness and income: The roles of trade and migration." Journal of International Economics 92(2):231–251.

Ottaviano, Gianmarco I. P. and Giovanni Peri. 2012. "Rethinking the Effect of Immigration on Wages." Journal of the European Economic Association 10(1):152–197.

Ozgen, Ceren, Peter Nijkamp and Jacques Poot. 2010. “The effect of migration on income growth and convergence: Meta-analytic evidence.” Papers in Regional Science 89(3):537–561.

Pesaran, M. Hashem. 2006. “Estimation and inference in large heterogeneous panels with a multifactor error structure.” Econometrica 74(4):967–1012.

Pesaran, M. Hashem. 2007. “A pair-wise approach to testing for output and growth convergence.” Journal of Econometrics 138(1):312–355.

Pesaran, M. Hashem. 2015. “Testing Weak Cross-Sectional Dependence in Large Panels.” Econometric Reviews 34(6-10):1089–1117.

Pesaran, M. Hashem and Elisa Tosetti. 2011. “Large panels with common factors and spatial correlation.” Journal of Econometrics 161(2):182–202.

Pesaran, M. Hashem and Ron Smith. 1995. “Estimating long-run relationships from dynamic heterogeneous panels.” Journal of Econometrics 68(1):79–113.

Pesaran, M. Hashem and Takashi Yamagata. 2008. “Testing slope homogeneity in large panels.” Journal of Econometrics 142(1):50–93.

Pesaran, M. Hashem, Vanessa L. Smith and Takashi Yamagata. 2013. “Panel unit root tests in the presence of a multifactor error structure.” Journal of Econometrics 175(2):94–115.

Putterman, Louis and David N. Weil. 2010. “Post-1500 population flows and the long run determinants of economic growth and inequality.” The Quarterly Journal of Economics 125(4):1627–1682.

Qu, Xi and Lung-Fei Lee. 2015. “Estimating a spatial autoregressive model with an endogenous spatial weight matrix.” Journal of Econometrics 184(2):209–232.

Qu, Xi, Lung-Fei Lee and Jihai Yu. 2017. “QML estimation of spatial dynamic panel data models with endogenous time varying spatial weights matrices.” Journal of Econometrics 197(2):173–201.

Qu, Xi, Xiaoliang Wang and Lung-Fei Lee. 2016. "Instrumental variable estimation of a spatial dynamic panel model with endogenous spatial weights when T is small." Econometrics Journal 19(3):261–290.

Rivera-Batiz, Luis A. and Paul M. Romer. 1991. "Economic integration and endogenous growth." The Quarterly Journal of Economics 106(2):531–555.

Rodriguez, Francisco and Dani Rodrik. 2000. Trade Policy and Economic Growth: A Skeptic's Guide to the Cross-National Evidence. In NBER Macroeconomics Annual 2000, ed. Ben S. Bernanke and Kenneth Rogoff. Vol. 15 MIT Press pp. 261–338.

Romer, Paul M. 1990. "Endogenous technological change." Journal of Political Economy 98(5):71 – 102.

Sachs, Jeffrey D. 2001. "Tropical underdevelopment." NBER Working Paper 8119.

Sachs, Jeffrey D. and Andrew Warner. 1995. "Economic reform and the process of global integration." Brookings papers on economic activity 1995(1):1–118.

Samuelson, Paul A. 1971. "Generalized Predator-Prey Oscillations in Ecological and Economic Equilibrium." Proceedings of the National Academy of Sciences 68(5):980–983.

Sandvik, Bjorn. 2008. "World Borders Dataset.".
URL: *http://thematicmapping.org/downloads/world_borders.php*

Sarafidis, Vasilis and Tom Wansbeek. 2012. "Cross-Sectional Dependence in Panel Data Analysis." Econometric Reviews 31(5):483–531.

Serlenga, Laura, Yongcheol Shin, Vanessa Gunnella and Camilla Mastromarco. 2013. "The Euro Effect on Intra-EU Trade: Evidence from the Cross Sectionally Dependent Panel Gravity Models." Unpublished mimeo (March):1–30.

Shin, Yongcheol, M. Hashem Pesaran and Ron P. Smith. 1999. "Pooled Mean Group Estimation of Dynamic Heterogeneous Panels." Journal of the American Statistical Association 94(446):621 –634.

Solow, RM. 1956. "A contribution to the theory of economic growth." The Quarterly Journal of Economics 70(1):65–94.

Spolaore, Enrico and Romain Wacziarg. 2009. "The diffusion of development." The Quarterly Journal of Economics 124(2):469–529.

Spolaore, Enrico and Romain Wacziarg. 2013. "How Deep Are the Roots of Economic Development?" Journal of Economic Literature 51(2):325–369.

Tao, Ji and Jihai Yu. 2012. "The spatial time lag in panel data models." Economics Letters 117(3):544–547.

Temple, Jonathan. 1999. "The New Growth Evidence." Journal of Economic Literature 37(1):112–156.

Temple, Jonathan. 2003. "The Long-Run implications of Growth Theories." Journal of Economic Surveys 17(3):497–510.

Vargas-Silva, Carlos and Cinzia Rienzo. 2014. "Highly Skilled Migration to the UK 2007-2013: Policy Changes, Financial Crises and a Possible 'Balloon Effect'?" The Migration Observatory, University of Oxford.

Vega, Solmaria Halleck and J. Paul Elhorst. 2016. "A regional unemployment model simultaneously accounting for serial dynamics, spatial dependence and common factors." Regional Science and Urban Economics 60:85–95.

Ventura, Jaume. 2005. A Global View of Economic Growth. In Handbook of Economic Growth, ed. Philippe Aghion and Steven N. Durlauf. Vol. 1B chapter 22, pp. 1419–1497.

Volterra, Vito. 1926. "Variazioni e fluttuazioni del numero d'individui in specie animali conviventi." Memorie della R. Accademia dei Lincei 6(2):31–113.

Wacziarg, Romain. 2001. "Measuring the Dynamic Gains from Trade." The World Bank Economic Review 15(3):393–429.

Wacziarg, Romain and Karen H. Welch. 2007. "Trade Liberalization and Growth: New Evidence." The World Bank Economic Review 22(2):187–231.

Williamson, Jeffrey G. 2006. "Global Migration." Finance and Development 43(3).

Yu, Jihai, Robert de Jong and Lung-Fei Lee. 2008. "Quasi-maximum likelihood estimators for spatial dynamic panel data with fixed effects when both n and T are large." Journal of Econometrics 146(1):118–134.