CEGE-SCHRIFTEN

Jens Südekum

Agglomeration and Regional Unemployment Disparities

PETER LANG

CEGE-SCHRIFTEN

Jens Südekum

Agglomeration and Regional Unemployment Disparities

In the European Union, unemployment rates differ markedly across regions, both within and across nations. This study presents a coherent theoretical approach to explain the emergence and persistence of such regional unemployment disparities. The analysis builds on the wage curve literature, and on regional agglomeration theories like the new economic geography. These theoretical strings are combined and extended, in order to provide a unified framework.

Jens Südekum was born in Goslar, Germany, in 1975. Since 1996 he studied economics at the University of Göttingen and the University of California at Los Angeles (UCLA). He received his diploma in 2000 and his PhD in Economics from the University of Göttingen in 2003.

Agglomeration and Regional Unemployment Disparities

CeGE-Schriften

Center for Globalization and Europeanization of the Economy
Zentrum für Globalisierung und Europäisierung der Wirtschaft
Georg-August-Universität Göttingen

Band 6

PETER LANG
Frankfurt am Main · Berlin · Bern · Bruxelles · New York · Oxford · Wien

Jens Südekum

Agglomeration and Regional Unemployment Disparities

A Theoretical Analysis with Reference to the European Union

PETER LANG
Europäischer Verlag der Wissenschaften

Bibliographic Information published by Die Deutsche Bibliothek
Die Deutsche Bibliothek lists this publication in the Deutsche Nationalbibliografie; detailed bibliographic data is available in the internet at <http://dnb.ddb.de>.

Zugl.: Göttingen, Univ., Diss., 2003

Gedruckt auf alterungsbeständigem, säurefreiem Papier.

D 7
ISSN 1617-741X
ISBN 3-631-51745-9
US-ISBN 0-8204-6521-6
ISBN 978-3-631-75686-7 (eBook)

Printed in Germany 1 2 4 5 6 7

www.peterlang.de

Foreword of the editor

Unemployment rates differ dramatically within the European Union. On a national level, these differences are typically explained by different national labour market institutions, economic policies, and country specific shocks. However, when looking at the European Union on the level of regions, one can have doubts if it is really useful to think about unemployment only in national dimensions. Within most EU countries, there are marked, if not dramatic intra-national unemployment disparities. Moreover, unemployment in Europe tends to be organized in a trans-national spatial pattern. On average, unemployment rates are low in the so called "European banana", i.e. in the economic core belt in the middle of the continent, where production and income are highly agglomerated. Vice versa, unemployment rates are extraordinarily high in the peripheral regions at the outside borders of EU-15.

In view of this, it seems necessary to reach out for some alternative theoretical approach, where differences in unemployment rates across jurisdictions are not primarily explained by institutional differences, which are mostly negligible for regions within the same country. It rather seems more adequate to attribute inter- and intra-national unemployment disparities to purely geographical factors. Recently, there was a rapidly growing interest in theoretical explanations for regional income disparities. Most notably, the "new economic geography" has shown how a core-periphery-structure of economic activity can endogenously emerge and persist within an integrated economic area. However, this literature does not incorporate an explicit analysis of labour market institutions and abstracts from unemployment. On the other hand, theories that explicitly aim at regional unemployment disparities, e.g. the wage curve theory, do not take into account the arguments for agglomeration and regional core-periphery patterns.

This gap in the literature is the starting point of the present study by Jens Suedekum. After a careful review of the wage curve literature and the new regional divergence theories, he combines essential elements of these different strings and offers a coherent theoretical analysis of inter- and intra-national unemployment disparities. He shows how trans-national unemployment clusters can emerge due to agglomeration economies, and he thereby shifts the focus away from institutional differences towards geographical factors as the main determinant for unemployment differences across administrative units. Even though the technical level of the book is advanced in some parts, the reader can always follow the economic intuition of the arguments. The study of Jens Suedekum in my view is a substantial contribution to the literature, and it can also inspire fruitful further research in this important area.

Peter Ruehmann
Goettingen, July 2003

Acknowledgements of the author

The present book has been accepted as my PhD-Thesis at the University of Goettingen on July 2nd, 2003. It was accomplished during the time period starting in November 2000, when I began to work as the research assistant of Professor Peter Ruehmann at the Department of Economics, who also was my PhD advisor.

I am indebted to his advice, his experience and his patience in long and very fruitful discussions about the contents of this book. He encouraged me to work on the issue of regional unemployment disparities, and I greatly benefited from his own research on the topic of unemployment.

My thanks also go to Professor Wolfgang Koenig, who was my first economics teacher ever and the second referee of my thesis, as well as to the third referee, Professor Wolfgang Benner.

The work I present in this book benefited a lot from discussions with numerous colleagues, both from the Department of Economics in Goettingen and from the outside. From this group of people, I would like to thank Andreas Haufler and Salvatore Barbaro with special emphasis, since each of them inspired and gave direction to my work in a substantial way.

I am grateful also to Professor Renate Ohr, the director of the *Center for Globalization and Europeanization of the Econcomy (CeGE)* at the University of Goettingen, who agreed to include my thesis in the institute's dissertation series.

My thanks go to my good friend Helge Carl, who not only provided excellent technical support, but also – as a non-economist – forced me numerous times to express my ideas in a way that is accessible also to people from other intellectual backgrounds.

Finally, it must be mentioned that working on this book would not have been possible without the care and personal support from my girlfriend Barbara, my father Dieter, and my deceased mother Ingrid, to whom I devote this book.

Jens Suedekum
Goettingen, July 2003

Forewords 5

Table of contents 9

List of tables, figures and maps 13

Introduction 15

A) Spatial economic disparities within the European Union: The evidence 19

A1) Preface: Level of spatial disaggregation and the choice of territorial units 19

A2) Gross domestic product (GDP) 20
A2.1.) GDP of European NUTS II-regions, 1999 20
A2.2.) Regional convergence versus divergence in Europe 21

A3) Regional unemployment in Europe 26
A3.1.) Unemployment rates in NUTS2-regions, 2000 28
A3.2.) Convergence versus divergence of regional unemployment rates 31

A4) Other regional indicators 32
A4.1.) Employment Growth 32
A4.2.) Population Density, Population Changes and Migration 34
A4.3.) Education 39
A4.4.) Innovation and research activities 39

A5) A closer look at the West German Länder 42

A6) Summing up the evidence 45

B) Macroeconomic theories of unemployment and the "European labour market model" 47

B1) Introduction 47

B2) A brief historical overview about macroeconomics 47
B2.1.) The ´classics´ 47
B2.2.) Keynes and the neoclassical synthesis 48
B2.3.) Friedman and the ´natural rate of unemployment´ 49

B2.4.) ´New classical macroeconomics´ and rational expectations 50
B2.5.) The Keynesian response 52

B3) The "European labour market model (ELMM)" 54
B3.1.) Why is there insider power in the labour market 55
B3.2.) The aggregate wage setting curve 58
B3.3.) The aggregate price setting curve 60
B3.4.) Equilibrium in the ELMM 63
B3.5.) Some further issues of the ELMM 67

C) The wage curve 71

C1) Introduction 71

C2) The wage curve as an empirical regularity 72

C3) Wage curve theory: The Blanchflower/Oswald-model 75
C3.1.) The partial equilibrium foundation of the wage curve 75
C3.2.) General equilibrium in the Blanchflower/Oswald-model 79
C3.3.) Critique of the Blanchflower/Oswald-model 82

C4) The model of Blien (2001) 83
C4.1.) Partial labour market equilibrium in the Blien-model 83
C4.2.) The product market and general equilibrium in the Blien-model 84
C4.3.) Critique of the Blien-model 86

C 5) Conclusion on wage curve theory and motivation for an own approach 88

D) Regional agglomeration theory and ´new economic geography´ 91

D1) Introduction 91

D2) Scale economies, externalities, and market competition 93

D3) The Marshallian agglomeration economies 95

D4) Centrifugal forces and other location factors 98

D5) The core-periphery model of ´new economic geography´ 100
D5.1.) Consumer behaviour 101
D5.2.) Production 103
D5.3.) Equilibrium conditions 104
D5.4.) Sustainability 107
D5.5.) Stability 109
D5.6.) ´New economic geography´ and the new trade theory 112

D6) Other ´new economic geography´-models 113
D6.1.) Venables (1996) and Krugman/Venables (1995) 113
D6.2.) Housing scarcity: Helpman (1998) 117
D6.3.) Analytically tractable models 120
D6.4.) More sectors, more regions 120
D6.5.) Dynamic models 121
D6.6.) Empirics, politics, and other unsettled issues 121

D7) Regional costs-of-living: An extension of the Krugman-model 122
D7.1.) Regional costs-of-living: the evidence 124
D7.2.) The basic structure of the extended model 126
D7.3.) Sustainability and stability 128
D7.4.) Conclusion of our approach 136

Appendix 138

E) Regional agglomeration and regional unemployment 139

E1) Introduction 139

E2) The existing literature 139
E2.1.) The model of Matusz (1996) 139
E2.2.) The model of Peeters/Garretsen (2000) 140

E3) Regional agglomeration and the wage curve: The model 142
E3.1.) The closed-economy setting 142
E3.2.) The two-region case with imperfect trade 148
E3.3.) The impact of labour mobility 155

E4) Critical discussion of our model approach 159
E4.1.) Discussion of the model from a theoretical point of view 159
E4.2.) Discussion of the model from an empirical point of view 164

E5) Further issues 169

F) Internal migration and regional disparities 171

F1) Introduction 171

F2) The causes of internal migration 172

F3) The consequences of internal migration: The neoclassical view 173

F4) Internal migration and regional divergence: alternative views 175
F4.1.) Increasing returns to scale 176
F4.2.) Selective labour migration 179

F5) Selective migration in a two-region model 183
F5.1.) The full-employment case 183
F5.2.) Union wage setting and unskilled unemployment 186
F5.3.) Conclusion of the model with constant returns to scale 190

F6) Agglomeration, worker heterogeneity and national union wage setting 191
F6.1.) The full employment case 193
F6.2.) The case with unemployment 205

F7) Discussion of the theoretical models and their empirical relevance 207
F7.1.) Discussion from a theoretical point of view 207
F7.2.) Discussion from an empirical point of view 208

Concluding remarks 211

List of references 215

List of tables, figures and maps

Chapter A

Map A1:	GDP per head by region (PPS), 1999	22
Map A2:	GDP per person employed (EURO), 1999	23
Table A1:	Disparities in GDP per head in PPS by region within Member States, 1989-1999	25
Table A2:	Transition probability matrices of GDP per capita (PPS) and unemployment rates of European NUTS2-regions relative to EU-15 average	27
Map A3:	Unemployment rate by region (2000)	30
Figure A4:	Disparities in unemployment, EU-15 1970-1999	31
Map A5:	Cumulative employment growth 1976-1998	33
Map A6:	Population density by NUTS3-region, 1999	35
Map A7a:	Crude rate of total population change, average 1995-1997	36
Map A7b:	Crude rate of net migration, average 1995-1997	37
Figure A8:	Features of regions with population decline in EU-15, 1993-1998	38
Map A9:	Education level by region, 2000	40
Map A10:	European patent applications, average 1997 to 1999	41
Figure A11:	West Germany - Gross wage bill per employee in % of national average, 1970-1999	43
Figure A12:	West Germany – Unemployment rates in North and South, 1967-1999	43
Figure A13:	West Germany – Net internal migration 1988-1999	44

Chapter B

Figure B1:	The wage setting curve in the ELMM	58
Figure B2:	Equilibrium in the ELMM	63
Figure B3:	Adjustment in the ELMM	65

Chapter C

Figure C1:	The wage curve	73
Figure C2:	Full equilibrium in the B/O-model	80

Chapter D

Figure D1:	Sustainability of a core-periphery structure	108
Figure D2:	The stability of the symmetric equilibrium	110
Figure D3:	The bifurcation diagram of the Krugman-model	111
Figure D4:	Location of the industrial sector in the Krugman/Venables-model	115
Figure D5:	Real wages in the Krugman/Venables-model	116
Figure D6:	The Helpman-model	119
Table D1:	Income, size and cost-of-living in US metropolitan areas	125
Table D2 :	The correlation matrix	126
Figure D7:	Relative CPI ψ_2/ψ_1 as a function of γ	130
Figure D8:	Real wage quotient ω_2/ω_1 as a function of γ	131
Figure D9:	Stability of the symmetric equilibrium	134
Figure D10:	Different scenarios of symmetry breaking	134-135
Figure D11:	The bifurcation diagram of our NEG-model	136

Chapter E

Figure E1:	Equilibrium in the closed economy	146
Figure E2:	Product market equilibrium in both regions	153
Figure E3:	Equilibrium in the two-region economy with immobile agents	154
Figure E4:	Total agglomeration with perfectly mobile labour	156
Figure E5:	Distribution of the labour force with intrinsic regional preferences	158

Chapter F

Figure F1:	Migration in a neoclassical two-region model	174
Figure F2:	Migration in a two-region model with an aggregate externality	178
Figure F3:	The determination of μ^*	201
Figure F4:	The determination of μ^* with endogenous $w_{2,t}$	205

One of the best ways to understand how the international economy works is to start by looking at what happens inside nations. (Paul Krugman)

Introduction

If we had to describe the economic geography of the European Union (EU-15) in only one phrase, we would say that there is a high degree of spatial agglomeration of economic activity in the EU-area and very pronounced core-periphery divides along various economic dimensions.

There are marked differences in per capita output and income levels across European regions. People from the richest European regions have e.g. an average real purchasing power about five times higher than people from the poorest areas. Spatial divides and regional disparities are even larger with respect to other indicators of economic activity. In this book, we will specifically be concerned with regional unemployment rates. In the European Union today, regions with practically full employment and regions with excessive mass unemployment coexist. In many cases, they even coexist within the same country. Germany, Italy and Spain are the most prominent examples of national economies, where some regions have unemployment rates below 5 per cent, whereas others are stuck with figures well above 20 per cent. Such spatial unemployment disparities within and across countries exist for decades. In recent years, they even tended to increase.

In chapter A of this book we will give an overview about the spatial structure of economic activity in the European Union. We will argue that regional unemployment rates in the EU follow a quite distinct, trans-national pattern and closely resemble the core-periphery-structure of regional GDP per capita. Regional unemployment rates are low in the rich core regions of the European Union, where population, production and income are agglomerated. On the contrary, high unemployment rates are found in the sparsely populated and economically peripheral regions with low levels of output and income per capita.

The main aim of this book is to explain this stylised fact. More specifically, we aim to explain the spatial structure of regional unemployment rates within an integrated economic area in relation to the corresponding spatial structure of output and income. This is a largely unexplored issue in economic theory.

Theorizing about unemployment has always been one of the most prominent tasks for macroeconomists, who predominantly think about the phenomenon of unemployment in its national dimension. Actually, different schools of thought within macroeconomics are still distinguished by the way how they think about the emergence of unemployment, especially in relation to the rate of inflation, and by the

the implications they derive for economic policy. Regional issues traditionally play a minor role in this debate.

However, regional labour market analysis has gained some prominence during the last ten years. One useful *regional* approach comes from David Blanchflower and Andrew Oswald (1990, 1996). These two authors have compiled a great deal of empirical evidence about regional labour markets and claim to have distilled an "empirical law" of economics from the data, according to which "doubling the unemployment rate of some region will drive down the regional wage level by roughly ten per cent". This law, known as the *wage curve*, and the theoretical work that is associated with it, will play an important role for our analysis. In order to justify the existence of a wage curve theoretically, one has to work with concepts of imperfect competition in the labour market. So do Blanchflower/Oswald in the theoretical parts of their work, which is build on one specific macroeconomic approach that is often labelled the "European labour market model (ELMM)". This important macroeconomic precursor will be introduced and discussed in chapter B of this book. Chapter C will then be devoted to the theory of the wage curve, that in many respects is the regional pendant of the ELMM.

The wage curve theory is useful for our purposes, since it draws an inherent link between key labour market variables on a regional level, the unemployment rate and the real wage level. But the existing wage curve models alone are insufficient to account for the observed spatial agglomeration of economic activity in the EU. The existing literature is useful to understand how regional unemployment rates develop if the corresponding regional levels of wages, output and income are exogenously given. But wage curve theory is ill-equipped to address *why* there are so pronounced disparities in the real world.

Economic agglomeration theories are the second string in the literature that is related to our theoretical analysis. It is known since Alfred Marshall (1890) that there are forces in the world that push for spatial concentration of economic activity in only a few locations or regions within an economy. Again, such spatial and regional issues traditionally have been neglected by mainstream economists. Of course, important contributions have been made in location economics, in urban and regional economics in the course of more than a century.[1] But all in all, this field of economics has eked a shadow existence within the discipline. This has dramatically changed in the last years. The revival of interest in spatial and regional issues as a piece of mainstream economics is somehow symbolised by the seminal contributions of Krugman (1980, 1991a,b) that led to the theories known as "new trade theory (NTT)" and "new economic geography (NEG)". Especially

[1] Location economics is actually seen as a "German tradition" and dates back to Heinrich von Thünen (1826), and followed by several other German writers such as Weber (1909), Christaller (1966), Loesch (1954). For a brief overview of some history of economic thought see Fujita/Krugman/Venables (1999), ch. 2+3.

the latter can be seen as a modern theory of regional agglomeration that explicitly addresses why core-periphery divides of economic activity can endogenously emerge and persist within an integrated area.

Chapter D will be devoted to the discussion of regional agglomeration theories, with special emphasis on NEG. The core model of Krugman (1991a) will be introduced, and the main developments of this theory will be traced until the current research frontier. Still we conclude that there are several unexplored and open issues. One of these issues, which is not directly related to the analysis of regional unemployment disparities, concerns unrealistic predictions that current NEG-models make about the development of regional costs-of-living during the process of regional agglomeration. This issue is taken up in chapter D, where we derive an own model that improves on the current state of art in the NEG-literature.

The main critique we formulate against the new agglomeration theories, however, is that they have nothing to say about unemployment. All standard models of NTT and NEG assume that labour markets always automatically clear. The phenomenon of regional unemployment disparities can thus not be analysed with these models. This neglect is peculiar, given that regional differences in unemployment rates are at least as pronounced, if not even more dramatic, than income disparities in the EU-15.

The chapters E and F we will therefore propose theoretical frameworks that attempt to close this gap in economic theory. In chapter E we will marry a wage curve, which is thought of as a labour market equilibrium curve, with a product market that exhibits the essential features of the new regional agglomeration theories. The innovation that comes from this model is twofold: Firstly, it can be seen as a general equilibrium model with a wage curve, where the regional disparities can develop endogenously. And secondly, the model can be seen as an attempt to introduce the element of unemployment to the new regional agglomeration theories.

The main result of our model framework is that the spatial structure of unemployment follows the spatial agglomeration pattern of overall economic activity. Large core regions with high per capita income levels have low unemployment rates. Vice versa, small regions with low income levels have high unemployment rates. Hence, our theoretical framework implies results that are consistent with the stylised facts about regional unemployment disparities and regional agglomeration in the EU.

Chapter F will specifically address the issue of labour mobility. If there are intranational economic disparities, what happens if people migrate from the blurring to the blooming areas? Is labour mobility an adjustment force that gradually leads to an erosion of existing disparities? Or do differences even get larger and more pronounced when workers migrate? The conventional viewpoint on this matter, known from the neoclassical theory of factor migration, is that labour mobility will lead to a convergence of wages and income levels. But even slight departures

from the neoclassical world lead to fundamentally different conclusions about the impacts and effects of interregional labour mobility.

We will specifically be concerned with the issue of *selective* labour migration. There are good theoretical and empirical reasons to assume that the group of workers who are mobile across space mainly consists of young and well educated individuals. The economic consequences of this *selective* labour migration are then derived using two alternative theoretical models. In the first version, we abstracts from endogenous agglomeration forces and formulate a straightforward neoclassical approach with a Cobb-Douglas production function. Still the model shows that high skilled labour migration will lead to a regional divergence process with respect to the real incomes of immobile low-skilled workers. In a second version, we again introduce endogenous agglomeration economies and show that selective migration and localised increasing returns are two distinct, but complementary arguments why labour mobility leads to regional divergence instead of convergence.

Both frameworks of chapter F are then taken further to analyse the implications of national union wage setting for regional unemployment outcomes. The models specifically address intra-national unemployment disparities that arise because of undifferentiated union wage setting irrespective of regional productivity differentials. The regional unemployment disparities that arise are magnified through selective labour migration. All in all, the models of chapter F constitute an alternative, but complementary view about the relation of regional agglomeration and regional unemployment. The book will be finished by some concluding remarks and a very brief discussion of some policy issues.

A) Spatial economic disparities within the European Union: The evidence

A1) Preface: Level of spatial disaggregation and the choice of territorial units

At what level of spatial disaggregation should we measure the magnitude and the development of regional disparities over time? Mainly for reasons of data availability one is constrained to choose among different stages of administrative units. The statistical office of the European commission (Eurostat) has developed the NUTS-division scheme (NUTS="Nomenclature of Statistical Territorial Units"). Herein, four levels of gradation are distinguished: the level NUTS0 is identical with the 15 current member countries of the European Union. Below this, there are 77 subordinate NUTS1, 211 NUTS2 and 1031 NUTS3 in the EU-15. These units respectively correspond to the German jurisdiction levels Bundesländer, Regierungsbezirke and Landkreise.

Aggregate economic analysis traditionally is concerned with the national (=NUTS0) level. There are good reasons for this, since nations are mostly well defined entities with a coherent institutional and political structure for which data availability is usually warranted. Furthermore, the development of aggregate theories requires that the area under consideration must be large enough in order to "convexify undeniable human indivisibilities and micro fixed costs" (Boldrin/Canova, 2001:212), i.e. to allow the abstraction from several microeconomic influences that macroeconomic theory necessarily has to make. The large NUTS0-regions surely have this desired property.

For an illustration of spatial economic disparities within the EU-15, however, NUTS0- regions are too unspecific. It will become apparent in this chapter that economic activity is often utterly unequal within member countries. Moreover, for today's European Union, one would have to rely on an unsatisfactorily low number of 15 observations, which even comprise cases that are hardly comparable at all (e.g. Luxemburg versus Germany). Therefore it makes sense to choose some smaller unit size to document the economic geography of the European Union. By using a more disaggregated NUTS-level, one can "gain variance" and base intra-European illustrations on a higher number of observations. Secondly, by using smaller spatial units, the problem of comparability is moderated, since the differences in size of the single NUTS2-regions are of course much smaller than for the single NUTS0-regions.

However, one should also be aware of different problems when disaggregation is pushed too far. NUTS3 areas e.g. might simply be too divergent in their underlying characteristics, natural endowments and historical preconditions as to conduct a meaningful comparison between them. Moreover, data problems and measurement errors can arise because of the entanglement of the single regions. For example, the

GDP per employee can be overstated in urban NUTS3-regions through daily commuters who live in surrounding areas.
In other words, the territorial units must be numerous enough, the spatial areas must at least be roughly comparable, and also large enough to cope with the problems mentioned above. A commonly accepted level of disaggregation for documenting economic and social cohesion (and also for the conduct of European regional policy) is the NUTS2-level.[1] Germany consists of 40 NUTS2 regions, whereas whole Denmark and Luxemburg are considered to be a single NUTS2.[2]

A2) Gross domestic product (GDP)
We will first consider the geographical distribution of GDP per capita across European NUTS2 regions measured in purchasing power standards (PPS) for the average of the years 1997-1999 . Later we trace out the development over the last decades.
GDP per capita is the most prominent indicator for measuring income inequalities within the EU, also heavily used for the conduct of European regional policies (see Suedekum, 2002a). From a theoretical point of view this measure might be problematic, since it measures output, not the disposable income level of the regional population. But for our documentation of stylized facts, we stick to this commonly used variable.

A2.1.) GDP of European NUTS2-regions, 1997-1999
Europe's richest NUTS2 region is Inner London with an income level of 246.3 relative to EU-15 (=100), followed by Brussels (223.1), Hamburg (183.4) and Luxemburg (179.7). At the other extreme, one finds regions like Ipeiros in Greece (47.3), Extremadura in Spain (50.3) or the Acores in Portugal (52.2). Put differently, individuals from the richest European regions have an average real purchasing power that is about five times higher than for inhabitants of the poorest areas.
But income differences by no means only occur across the EU-15 member states. Also within countries there are substantial differences: the earnings in Hamburg are about three times higher than in Dessau (63.2), albeit the two regions are less than 300 kilometers apart. Other extreme examples include Lombardia (136.5) versus Calabria (61.9) in Italy, or Ile de France (154.1) versus Languedoc-

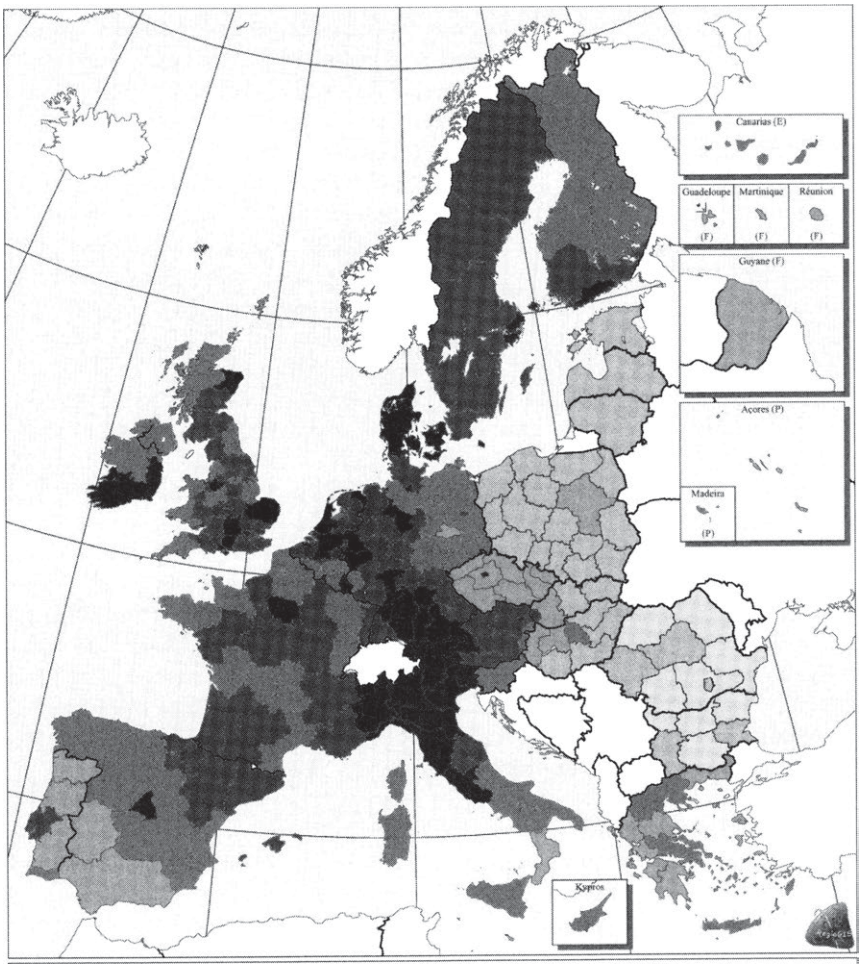
[1] Note however the critique of Boldrin/Canova (2001:212) who imply that NUTS2 are too small and too inherently different.
2 A full list of all European NUTS2 regions, including various economic summary statistics can be found in the Cohesion Report of the EU Commission (2001), available under http://www.europa.eu.int/comm/regional_policy/sources/docoffic/official/reports/pdf/taba_en.pdf

Rousillon (77.2).[3] Cornwall in the UK (66.6) or Hainaut in Belgium (71.8) are also substantially poorer than their corresponding capital regions.
A more comprehensive picture on regional income disparities in the EU is presented in map A1. The prospective accession countries from Eastern Europe are included in this map, so that the regional GDP levels are relative to the EU-27 average. It can be seen that most entrant regions have incomes even below the poorest current member regions. But we want to focus here solely on the EU-15.
The map illustrates that the regional distribution of GDP per capita follows a quite distinct spatial pattern: the rich regions are located roughly in the middle of the continent, in the so-called "European Banana", ranging from Southern UK over Benelux, East France and West Germany up to Northern Italy. Surrounding the economic core belt is a group of regions with medium per capita incomes, e.g. North-West Germany, Northern UK, Scandinavia as well as large parts of France. The economically lagging parts of Europe are all at the outside borders of EU-15. Most notably these are Southern Italy, East Germany, the Burgenland (AT), Greece and nearly all of Spain and Portugal. Together this group is eligible for structural funding from the EU Commission under "objective 1" until at least 2006, i.e. they match the criterion that their GDP per capita falls short of 75% of the EU-15 average.[4]
The image of a "core – periphery" structure in the spatial distribution of GDP remains unchanged when we consider GDP per employed person (map A2).[5] This measure more directly reflects the productivity level in European regions through adjusting for different employment rates. The image of the economic backwardness of Portugal and Spain is magnified to some degree, whereas the situation looks somehow more friendly for the case of Southern Italy. All in all, however, the magnitude and the geographic structure of spatial divides presented in this figure complements the impression from map A1.

A2.2.) Regional convergence versus divergence in Europe

The illustration of the descriptive cross-section evidence does not cope with the development of relative regional income levels over time. Has the difference between the richest and the poorest regions been narrowing, or did the disparities tend to grow over the last years?

[3] This excludes the French overseas departments that are even poorer (55.1).
[4] For an overview of European regional policy see e.g. Boldrin/Canova (2001), Suedekum (2002a), Puga (2002)
[5] Due to data limitations, the map only traces NUTS1 (or even NUTS0) regions.



Map A1: GDP per head by region (PPS), 1999

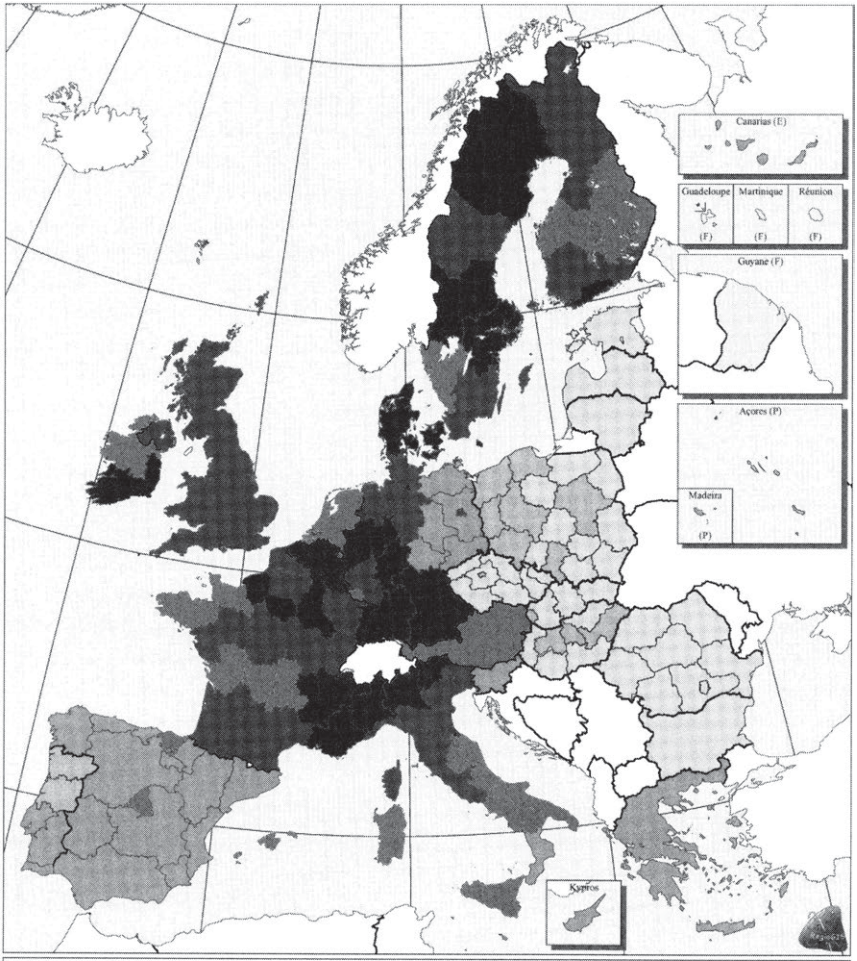
Index, EUR-27 = 100

- < 30
- 30 - 50
- 50 - 75
- 75 - 100
- 100 - 125
- >= 125

Source: Eurostat

0 100 500km

© EuroGeographics Association for the administrative boundaries



Map A2: GDP per person employed (EUR), 1999

Index, EUR-27 = 100

- < 25
- 25 - 50
- 50 - 100
- 100 - 120
- 120 - 130
- >= 130
- no data

EUR-27 = 100
Standard deviation = 45.6

D : NUTS1
B, EL, NL, A, UK, BG : NUTS0

Sources: Eurostat and NSI

0 100 500km

One instructive way to assess this question is to look at the (weighted) standard deviation of the relative income levels. This is done in table A1. Data is presented for the EU-15 as a whole, as well as for all EU member countries that consist of more than one NUTS2 region. According to this table, regional income dispersion across EU-15 in 1999 is approximately on the same level as in 1992 (28.4 vs. 28.6). The substantially lower value of 1989 (26.4) is due to the fact that the East German Länder were not yet included. Similarly, the value for 1991 (29.4) is probably so high, because it reflects the immediate economic shock from reunification. All in all, one can say that there have been no apparent signs of regional convergence within the EU over the last decade, particularly not since 1995. Matters are different if one looks at the income dispersion across nations. The standard deviation has been more or less constantly declining since 1992. Especially if one considers the last five years, it looks as if there has been a process of national income convergence parallel with a persistence of regional income disparities.

This at first sight puzzling development can be understood by looking at what has happened within the single EU member states. Table A1 indicates that the economic dispersion across German, French, Italian, Greek, Austrian and Belgian regions remained roughly stable or was even slightly declining since 1995. In the same time period, however, there has been a strong regional income divergence in other countries, namely in Spain, Portugal, Ireland, the Netherlands, Finland, Sweden and the UK. This divergence process specifically occurred, because the large and economically most advanced regions in these countries grew significantly faster than the poorest areas. For example, the area around Lissabon caught up substantially with the rest of the European community, whereas the relative income of Alentejo was even lower in 1999 than in 1995.

Since the large central regions within any nation have a higher weight for aggregate figures, their economic progress leads to a closing national income gap relative to the EU-average. This convergence at the country level, however, was accompanied by a process of intra-national differentiation. This view is supported by various authors, such as Gianetti (2002), Esteban (2000), Magrini (1999), P. Martin (1998), Fagerberg/Verspagen (1996) or Neven/Gouyette (1995). For the EU-15 as a whole, however, it is more appropriate to speak of a persistence of regional disparities rather than of regional divergence (Boldrin/Canova, 2001; R. Martin, 2001; Lopez-Baso et. al., 1999; Canova/Marcet, 1995), simply because some countries experienced at least a moderate reduction of spatial divides. But in line with the arguments above, those countries that experiences the sharpest increase in regional inequalities, had the highest catching-up speed at the national level (Quah, 1997).

Table A1: Disparities in GDP per head by region within EU-member states
(weighted standard deviation of index EU-15 = 100)

Member State	1989	1990	1991	1992	1993	1994	1995	1996	1997	1998	1999
B	24,7	25,1	25,1	26	27,1	25,9	40,1	40,8	40,7	40,8	39,4
D			38,6	35,8	32,4	31,3	25,8	25,4	25,2	25,4	25,4
excl. New Länder	21	21,8	22,7	23	22,8	23,4					
EL	6,5	6,3	6,1	6,6	7,6	7,8	10,4	10,2	9,5	9,5	9,7
E	14,9	14,9	16	15,9	15,2	15,9	16,8	17,2	17,8	18	18,9
F	28,6	28,9	29,9	28,9	29,9	30,8	28,3	28,2	27,6	27	27,2
IRL							13,8	13,5	16,7	16,6	17,4
I	25,8	24,8	24,7	24,9	24,7	25,5	28,5	28,6	27,7	28,1	27,7
NL	10,6	10,6	11,8	11,3	11,5	10,8	13,4	14,3	15,4	16,1	16,2
A	27	27,5	28,6	28,7	30,3	28,1	25,5	24,9	23,8	22,3	22,4
P	17,7	13,5	15	13,6	14,3	13,8	16	16,5	18,2	19,1	19,1
FIN	17,7	17,9	17,7	15,4	17	17,1	19,5	20,9	20,8	23,9	24,2
S	10,9	10,8	12	10,9	12,8	11	13,1	14	16,2	17	16,4
UK	20,7	20,2	19,2	19,6	20,6	18,3	31,5	31,6	34	35,6	34,2
EU (15) - by			29,4	28,6	27,7	27,5	28,7	28,5	28,6	29	28,4
excl. New Länder	26,4	26,5	26,4	26,5	26,3	26,5					
EU (15) - by member state			13,1	13,2	12,5	12,7	12,4	11,9	11,5	11,5	10,8
excl. New Länder	15,3	15,4	15,5	15,6	14,6	14,6					

Source: Eurostat; European Commission, DGRegio.

Looking only at weighted standard deviations can be quite misleading. A constant value can be interpreted in two completely different ways: either the relative position of the single regions actually remained constant, or it is only the overall dispersion of regional incomes that is unchanged, whereas the position of single regions within this distribution has changed significantly. One way to discriminate between these two possibilities, suggested by Quah (1997), is to construct a *transition probability matrix*. For this matrix, an arbitrary ranking scheme must be defined according to which all individual regions are grouped. We have chosen the same ranking scheme as Quah (1997) and Puga (2002), i.e. we have divided all 211 NUTS2-regions into five sub-groups. The poorest group entails regions with GDP per capita below 60 per cent of the EU-15 average, in the rich group there are all regions with an income level above 130 per cent etc. Then we have determined to which group either of regions has belonged according to the GDP per capita levels in 1995, and in 1999. The numeric values in table A2a indicate the relative

frequencies of group membership according to 1999 GDP figures, given the information about 1995 GDP. Reading along the main diagonal yields the fraction of regions that remained in their respective income range.
The values indicate that there has been a high degree of stability in the spatial distribution of GDP across NUTS2 regions in the time period from 1995 to 1999. The fraction of regions that remain in the same income range is usually over 80 per cent. For example, from the 87 regions with initial incomes between 75 and 100 per cent in the year 1995, more than 86 per cent ended up in the same range also in 1999. 11 per cent have moved upwards to the group with incomes from 100 to 130 per cent, and 2 per cent of the regions moved to the lower group. Not a single of the 211 NUTS2 regions has switched its position to a non-neighboring group.
These figures support the interpretation that the roughly constant value of the weighted standard deviation in table A1 is actually due to a strong income persistence: the position of single regions within the overall regional income distribution has changed little over the last years. The most notable exceptions are Southern and Eastern Ireland in the positive, and Berlin in the negative sense. However, one should also be aware of some methodological problems associated with the transition probability matrix. This firstly concerns the arbitrary definition of the group borders, as well as the neglect of any changes that occur within the single subgroups. These problematic aspects notwithstanding, table A2 is still an instructive way to look at the degree of regional income persistence in the last years. And it is interesting to compare our own calculations for this recent time period with the analogous calculations that Puga (2002) has made for the years 1987 and 1995. Although his results already revealed regional income persistence across the European regions, this inertia seemingly has been strengthened recently, particularly for regions from the medium income range.[6] This is again consistent with the findings obtained from table A1, since the standard deviation of regional per capita incomes exhibited inertia especially since 1995.

A3) Regional unemployment in Europe

The second basic indicator of economic activity that we are concerned with is the unemployment rate. Again, we firstly adopt a cross-section view of unemployment rates in the single European NUTS2-regions for the year 2000 before we look at the development over time.

[6] One should note, however, that Puga (2002) looks at a longer time period of 8 years, during which regional convergence processes can be more pronounced than for a four-year time period. Our own figures are thus not directly comparable with those of Puga.

Table A2: Transition probability matrices of GDP per capita (PPS) and unemployment rates of European NUTS2-regions relative to EU-15 average.

a) GDP per capita (PPS) 1995 and 1999

1995 GDP per cap.	[n]		1999 GDP per cap. [13] < 0.6	[34] 0.6-0.75	[91] 0.75-1.0	[54] 1.0-1.3	[19] >1.3
	[16]	**< 0.6**	0.813	0.188	0.000	0.000	0.000
	[34]	**0.6–0.75**	0.000	0.853	0.147	0.000	0.000
	[87]	**0.75-1.0**	0.000	0.023	0.862	0.115	0.000
	[58]	**1.0-1.3**	0.000	0.000	0.190	0.724	0.086
	[16]	**> 1.3**	0.000	0.000	0.000	0.125	0.875

b) GDP per capita (PPS) 1987 and 1995 (from Puga, 2001)

1987 GDP per cap.		1995 GDP per cap. < 0.6	0.6-0.75	0.75-1.0	1.0-1.3	>1.3
	< 0.6	0.83	0.17	0.00	0.00	0.00
	0.6–0.75	0.21	0.47	0.32	0.00	0.00
	0.75-1.0	0.00	0.18	0.68	0.14	0.00
	1.0-1.3	0.00	0.00	0.13	0.72	9.15
	> 1.3	0.00	0.00	0.00	0.17	0.83

c) Unemployment rates 1986 and 1996 (from Puga, 2001)

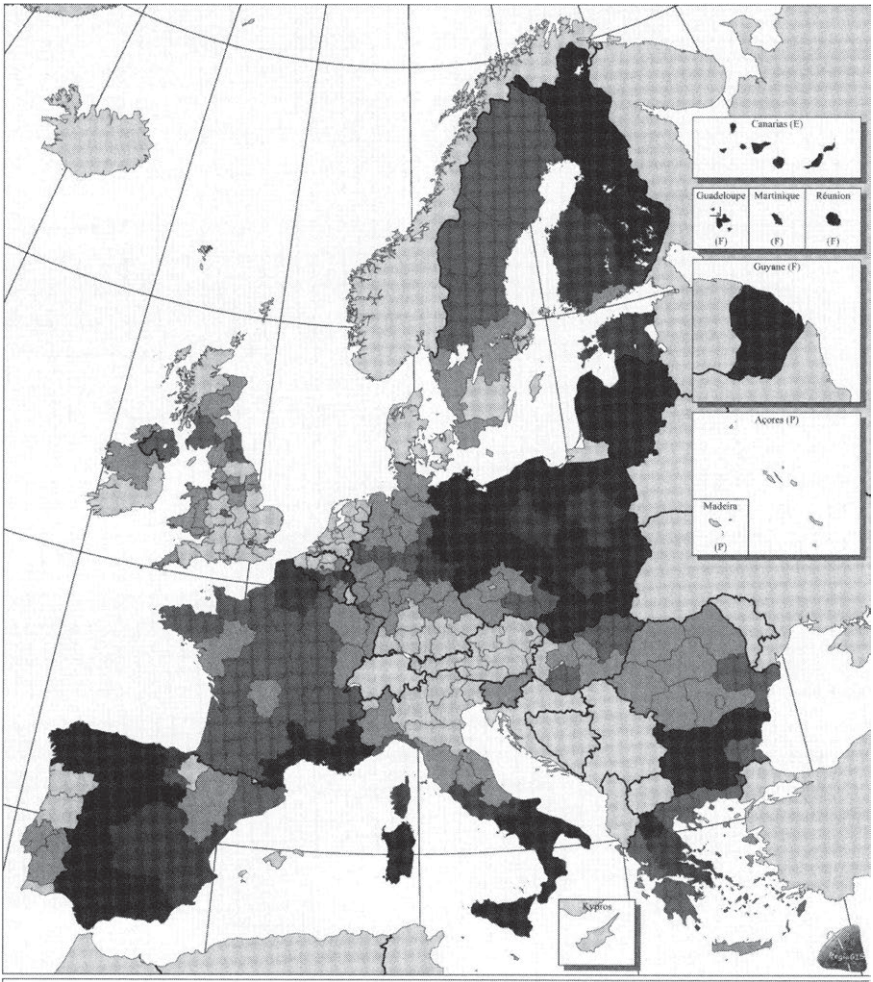
1986 Unemployment		1996 Unemployment rate < 0.6	0.6-0.75	0.75-1.0	1.0-1.3	>1.3
	< 0.6	0.81	0.19	0.00	0.00	0.00
	0.6–0.75	0.52	0.26	0.09	0.09	0.04
	0.75-1.0	0.24	0.29	0.26	0.21	0.00
	1.0-1.3	0.06	0.22	0.34	0.19	0.19
	> 1.3	0.00	0.00	0.16	0.22	0.62

A3.1) Unemployment rates in NUTS2-regions, 2000

The first impression that emerges from map A3 is that unemployment rates are also utterly unequal across Europe, both across and within member states. There are several regions in the EU-15 where there is in principle full employment, like Aland in Finland (unemployment rate 1.7 in 2000), Berkshire (1.9) or Utrecht (2.1). At the same time there are regions with excessive mass unemployment, like Calabria (27.7), Ceuta y Mellila (25.5) or Halle (19.2). Probably the most puzzling feature of the map A3, however, is the extremity of intra-national differences in unemployment rates. In Italy, Spain and Germany there are areas which – in the sharpest possible contrast to Calabria, Ceuty y Mellila and Halle – are very close to full employment levels, e.g. Trentino-Alto Adige (3.1), Navarra (4.9), or Oberbayern (3.5). These three countries, Italy, Spain and Germany, constitute the most extreme cases of regional unemployment disparities. Nevertheless, in almost all countries intra-national differences exist to a no-negligible degree. In France, the range goes from 5.3 (Alsace) to 16.1 (Languedoc-Rousillon), where we have neglected the overseas departments Guadeloupe, Guyane, Martinique and Reunion with excessive unemployment rates around 30 per cent. In the UK, Merseyside (11.2) contrasts the low-unemployment experience of Berkshire, just as Ita-Suomi (15.5) contrasts Aland for the case of Finland. Also small countries exhibit marked differences. In Greece they amount to more than 9 percentage points between Ionia Nisia (5.1) and Dytiki Makedonia (14.7), in Belgium the range goes from 2.9 (Vlaams Brabant) to 13.1 (Hainaut). Smaller differences around 3-5 percentage points exist in Austria, Portugal, and Sweden. And practically the only country without regional disparities are the Netherlands, where unemployment is ubiquitously low at levels around 2-4 per cent.[7]

There is a long tradition in economic research to attribute the observed cross-sectional variation of unemployment rates to the impact of institutional arrangements, most notably to various forms of labour market rigidities. This research strategy has mostly been applied to comparisons of national unemployment rates (e.g. Nickell, 1997). This is straightforward, since most institutions that are relevant for the labour market, the tax system, welfare state arrangements, labour laws etc., differ only across but now within countries. Yet, when looking at map A3, it becomes questionable if it is really useful to predominantly think of unemployment along country borders. Identical institutions on the national level produce utterly different regional unemployment rates. In view of the evidence, there seems to be a more meaningful grouping scheme to describe the geographical configuration of unemployment in the EU that does not obey to national borders.

[7] The same general picture, wide regional unemployment disparities both across and within the member states in the EU, also emerges when using OECD-data on NUTS2 level, see e.g. OECD (2000).



Map A3: Unemployment rate by region, 2000

% of labour force

< 4.95
4.95 - 7.85
7.85 - 10.75
10.75 - 13.65
>= 13.65

EUR-27 = 9.3
Standard deviation = 5.74

Sources: Eurostat and NSI

0 100 500km

Regional unemployment rates in fact closely resemble the "core-periphery" structure of regional GDP per capita that has been described above. Low unemployment is centered around the "European Banana", i.e. in Southern UK, Netherlands, Flanders, Southern Germany and Northern Italy. Similarly, all areas with mass unemployment belong to the poor peripheral parts of EU-15, the so called "objective 1"-regions: Southern Italy, East Germany, South and East Spain, Northern Finland.[8] Most medium income regions also belong to the group with intermediate unemployment rates.

Thus, the membership of a single region to one of the three "clubs" (Banana, objective 1, intermediate) seems to be a more reliable indicator for the regional unemployment rate than the pure assignment to some EU member country. This notion is supported e.g. by Overman/Puga (2002), who find that "*the unemployment outcomes of individual regions are much closer to the outcomes of their neighbors, than to the average outcomes of other regions within the same EU-member state*". This "neighboring effect" leads them to conclude that unemployment clusters have been shaped within EU-15, and that there is truly a spatial dimension in European unemployment.

The relation between income per capita and unemployment rate is surely not one-by-one, as several counterexamples are at hand: Portugal entirely consists of poor regions, especially when considering GDP per employed person. Portuguese unemployment on the other hand is low by European standards. Greek unemployment is also not as high as one might expect given its GDP figures. One might therefore put it this way: belonging to the "objective 1"-group is a necessary, but not a sufficient condition for exhibiting extreme regional unemployment rates of above 15 per cent or so. Nevertheless, the average unemployment rate for all "objective 1"-regions is markedly higher than the EU-15 average (15.8 vs. 9.7 for 1999). There are also some rather rich central regions with quite high unemployment rates (e.g. Nord-Pas de Calais in France). But the general conclusion, that the spatial pattern of unemployment rates resembles the spatial pattern of GDP per capita, seems hardly disputable.

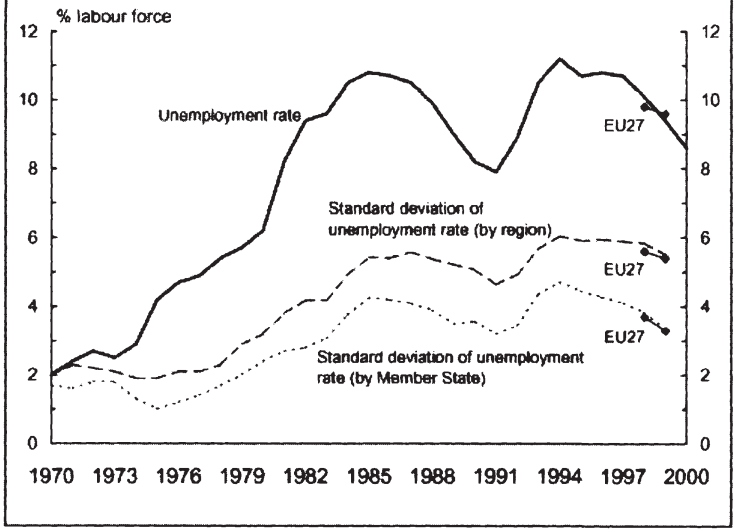
[8] This view is supported e.g. by CER (1998). They report that "the high unemployment regions in Europe have a low per capita income (30% below the EU average) and a similar production structure, in which manufacturing represents a lower than average share of output and is characterized by technologically stagnant industries such as food, mining, leather and apparel. On the contrary, the low unemployment regions are characterized by a 10% higher than average per capita income and a production structure in which manufacturing is prominent and diversified, with a prevalence of industries such as machinery, precision instruments and electronics".

A3.2.) Convergence versus divergence of regional unemployment rates

Looking at the history and the development of regional unemployment rate disparities within Western Europe, one finds much clearer evidence for regional divergence than it was the case for regional GDP per capita levels.

At first glance this can be seen in fig.A4, which indicates a strong upward trend of the regional unemployment disparities since the 1970s, measured by the weighted standard deviation of all regions now belonging to EU-15. Both regional as well as national differences rose pro-cyclically with the overall unemployment rate of EU-15, but the rise of regional disparities has been more pronounced. Since the mid 1980s, regional unemployment disparities rose only moderately on average, but surely revealed no sign of deterioration. The OECD (2000:32) subscribes to this view as it points out that "*variation in regional unemployment rates increased in many countries during the 1970s and early 1980s. [...] This variation generally remained stable or increased between 1985 and 1997.*"

Fig. A4: Disparities in unemployment, EU15 1970-99; EU27 1998-99



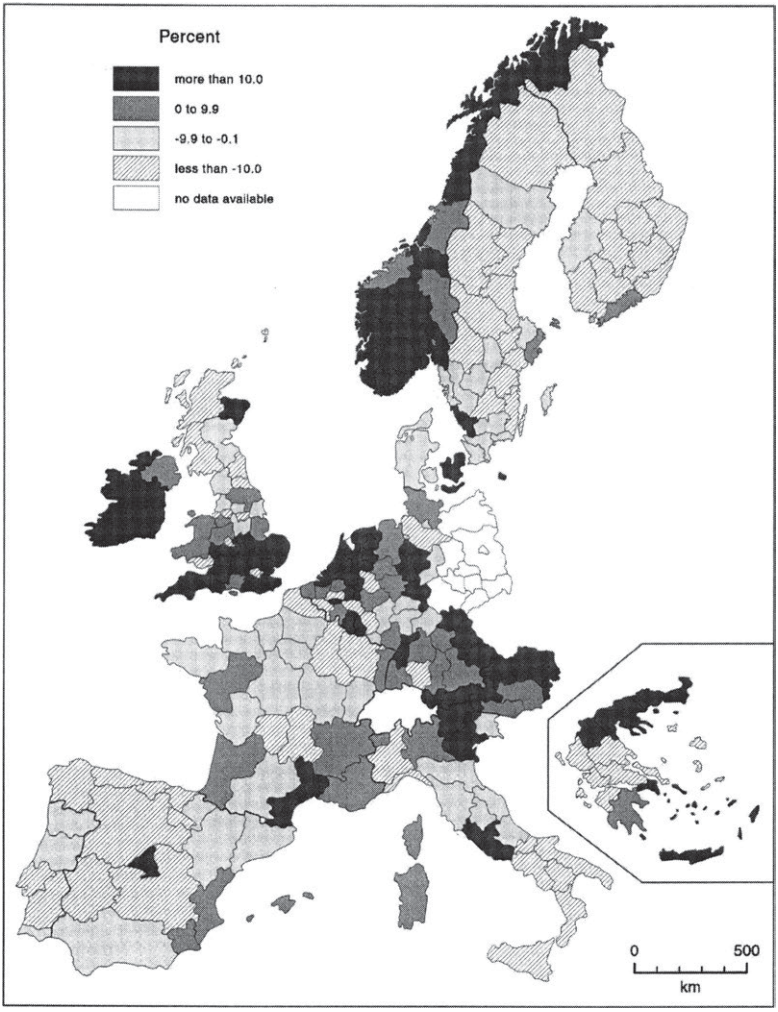
However, there is also evidence for another process that can not be revealed by using simplistic summary statistics like standard deviations. Overman/Puga (2002)

show that regional unemployment in the EU was subject to a process of polarization since the mid 1980s. The transition probability matrix for unemployment rates of European NUTS2 regions relative to the EU-15 average at times 1986 and 1996, that is reported in table A2c), illustrates this fact. There has been a quite high degree of persistence for the group of regions with very high and very low unemployment rates. That is, regions with low (high) unemployment rates in 1986 mostly belonged to the same group in 1996. This inertia, however, was absent for many regions who belonged to the group with intermediate unemployment rates in 1986. The degree of persistence for the three sub-groups with relative unemployment rates ranging from 0.6 to 1.3 is substantially lower than for the other two groups. Many regions with intermediate unemployment levels in 1986 moved to either of the two extremes. For example, several Italian, French and Spanish regions that used to have relative unemployment rates somewhere around the EU average experienced a significant deterioration since 1986, whereas some other regions from the same countries saw their unemployment rates drop substantially.
As Overman/Puga (2002) show, this polarization process led to the geographical configuration of unemployment that we have described above, where areas with high and low unemployment were not divided by national borders, but distinguished by cross-country unemployment clusters. To obtain a better understanding of this evolution, it is helpful to distinguish what has driven this polarization process. Unemployment rates by definition change either because of changes in the size of the labour force (i.e. labour supply changes), or changes in employment (i.e. labour demand), or a combination of the two. It seems thus natural to look for regional employment growth patterns, as well as for population changes and migration flows within EU-15 in the next section.

A4) Other regional indicators

A4.1.) Employment growth

An influential study on regional employment growth in Europe comes from Martin/Tyler (2000). The authors look at the evolution of cumulative relative employment growth, defined as the cumulative annual change in the log of regional employment minus the cumulative annual change in the log of Europe-16 employment, which consists of the usual EU-15 countries plus Norway. The study comprises annual data for the period 1975-98, the disaggregation corresponds whenever possible to the NUTS2 level. This measure thus captures the cumulative differential employment growth experience of the regions.
The authors show that there has been marked regional divergence in employment growth within virtually all countries, particularly during the 1980s. Although one



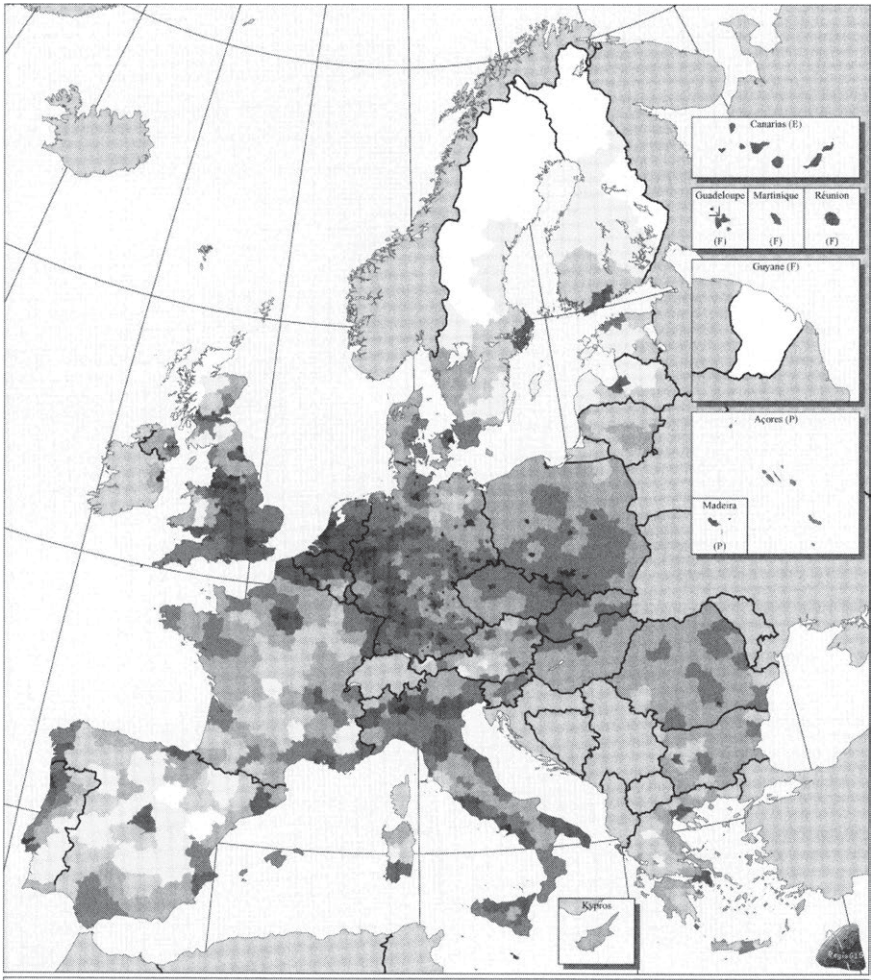
Map A5: Cumulative employment growth, 1976-98 (from Martin/Tyler, 2000)

can also single out three groups of countries that faced different employment growth at the national level,[9] the most distinguishing feature for the time period under consideration is the substantial intra-national divergence in cumulative employment growth. The degree of regional divergence varies across countries. It is particularly large in Italy, Spain, the UK and Greece. In these states, there has been a difference of 50% or more between the cumulative employment growth of leading and lagging regions (Martin/Tyler, 2000:605 ff.).
From a bird's perspective, these regional evolutions again follow a specific geographical pattern (see map A5). The economic core belt (the "European banana") is again visible in this picture and reveals by and large cumulative employment growth well above the European average. With a few outliners and exceptions, the poor peripheral regions of the EU (the "objective 1" - regions) constituted the area with slow employment growth. In between, there is again the group of regions with intermediate income levels and unemployment rates, which on average also faced employment growth of medium strength.
The regional characteristics ´low unemployment rate´ and ´strong employment growth´ thus tended to coincide. In relation to the results from the last section, it follows that the polarization process of regional unemployment rates was mainly driven by differential intra-national employment growth paths. This result is supported by Overman/Puga (2002), who find that regions that recently ended up with relatively low unemployment had relatively high employment growth over the last decades. The authors therefore conclude that " *[...] contrary to labour force changes, employment changes have worked for polarization [of unemployment rates]. It is employment changes that have driven high unemployment regions to their high rates and low unemployment regions to their low rates*".

A4.2.) Population Density, Population Changes and Migration

In the same vein, one can show that population changes and migration on average have worked *against* unemployment polarization, since most regions that ended up with relatively low unemployment revealed relatively high labour force growth. An increase in labour supply translates into more unemployment if it outperforms growth in labour demand. However, the evidence indicates that the opposite has happened in the last decade: the increase in labour demand has been stronger than the increase in labour supply for the regions with good performance in reducing unemployment.

[9] Martin/Tyler (2000) show that five states (Ireland, Luxemburg, Netherlands, Austria and Norway) had national employment growth rates significantly above the Europe-16 average, a second group consisting of the large EU-core countries West Germany, France, Italy and UK as well as of Belgium and Denmark exhibited employment growth rate in line with the European average. On the bottom of the scale, employment growth has been markedly below average in the geographically remote countries Spain, Portugal, Greece, Sweden and Finland.



Map A6: Population density by NUTS3 region, 1999

Inhabitants/km²

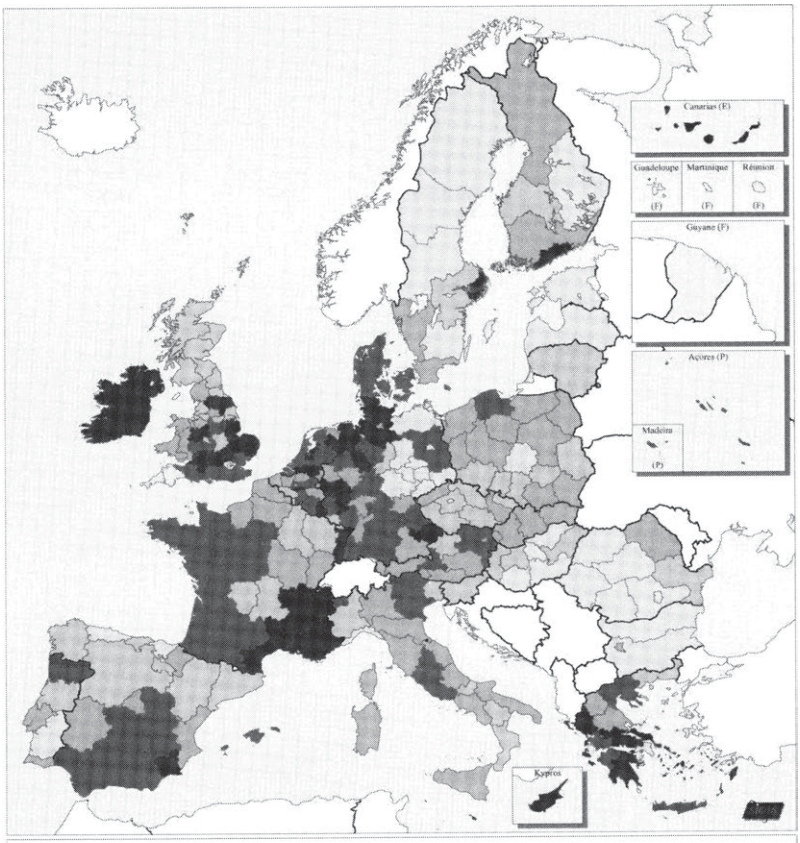
- < 10
- 10 - 25
- 25 - 50
- 50 - 100
- 100 - 200
- 200 - 500
- 500 - 1000
- >= 1000

EUR-27 = 112

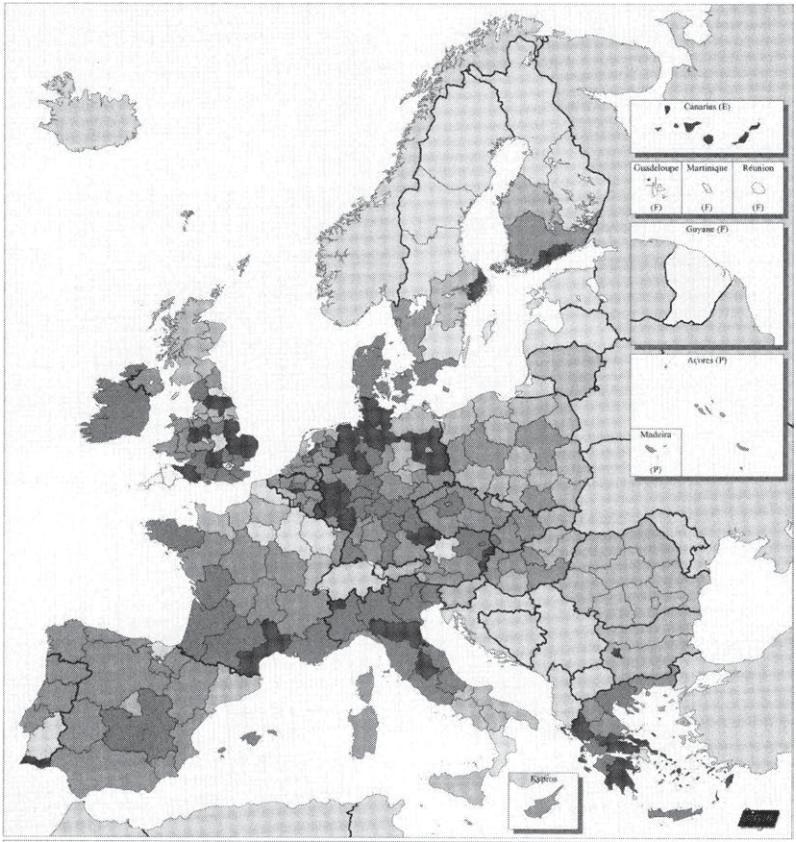
EL, F, CY: 1998

Source: Eurostat and NSI

0 100 500km



Map A7a: Crude rate of total population change, average 1995-1997



Map A7b: Crude rate of net migration, average 1995-1997

< -2.5
-2.5 - 0.0
0.0 - 2.5
2.5 - 5.0
> 5.0
no data

Sachsen, Wales, Scotland: NUTS1
IRL: NUTS0

Source: Eurostat

0 100 200 km

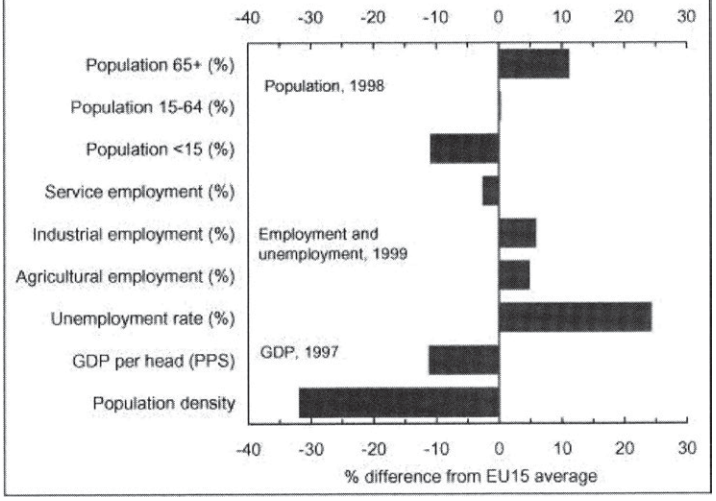
© MEGRIN for the administrative boundaries

To assess the regional dimension of labour supply in Europe, it is instructive to look first at the population density across regions in the EU-15. Map A6 does so on a rather disaggregated basis, for NUTS3 units in the year 1999. The strong concentration in the "European banana" becomes particularly obvious in this map. It is fair to say that there is a strong agglomeration of labour supply in the economic core belt of EU-15. The peripheral parts of EU-15 on the contrary are characterized by a much lower population density.

Moreover, there is indication that the poor and sparsely populated areas within EU-15 have even lost inhabitants over the last years. This is most evidently so for East Germany, Northern Spain and Southern Italy, as can be seen in map 7a.

The population decline is due both to natural processes as well as to emigration, depicted in map 7b. According to that, East Germany is subject to a rather large wave of emigration,[10] whereas the Spanish population decline apparently has had mainly other causes. The greatest population gains were recorded in Ireland, Southern France, and Southern UK. The latter area together with some West German regions has received the largest waves of immigration.

Fig.A8: Features of regions with population decline in EU15, 1993-98



[10] with one exception: the substantial wave of immigration to the region "Brandenburg" is a singularity due to a sub-urbanization process of the Berlin area.

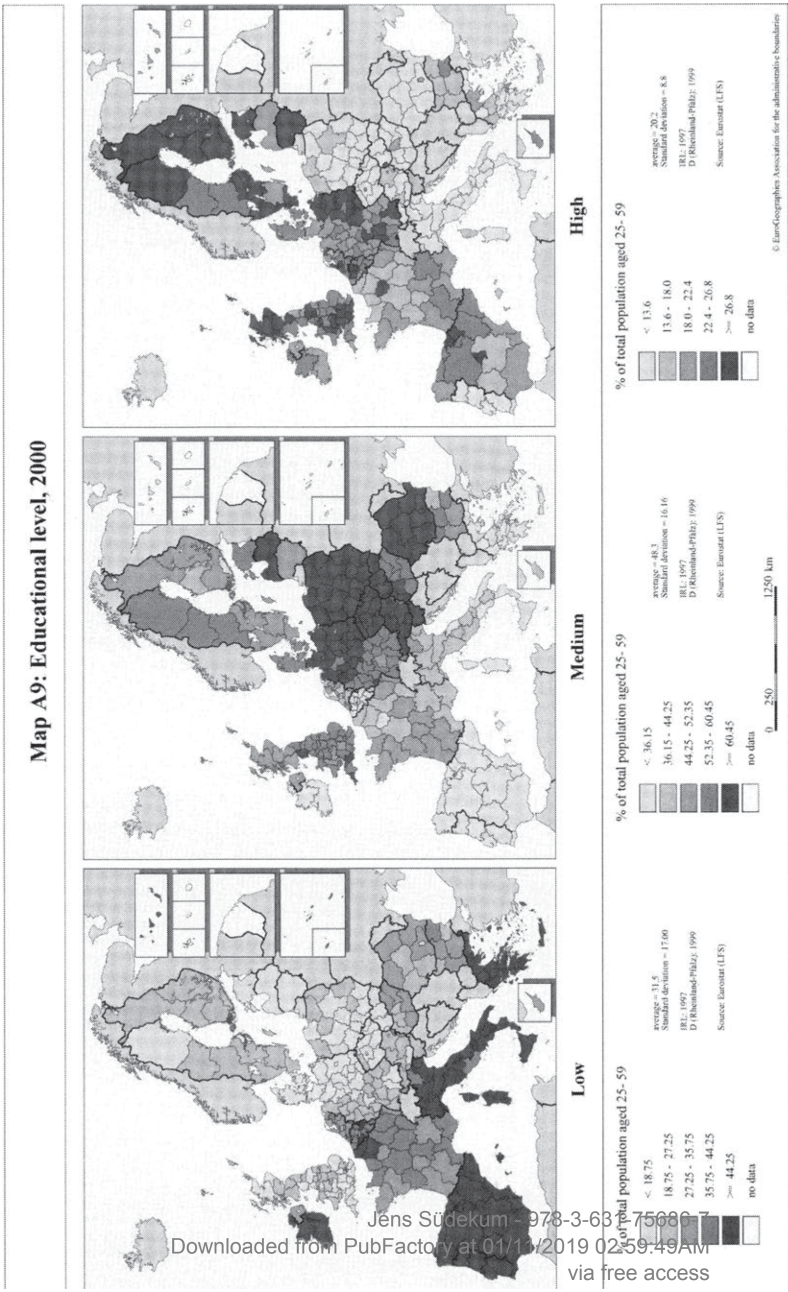
All in all, some regularities of regions that exhibited a population decline between 1993 and 1998 can be distinguished and highlighted. Fig. A8 illustrates common characteristics of NUTS2 regions that were subject to population decline between 1993 and 1998. It shows that on average this subgroup of regions was characterized by comparatively high unemployment rates (about 25% above the EU-15 average) and below-average income levels. Interestingly, regions which had a low population density to begin with, were subject to further population decline.
The general picture that emerges from the maps and figures A6-A8 is that migration and population flows maybe do not reveal such a clear-cut spatial pattern as income levels and unemployment rates. But on average, rich regions with low and declining unemployment rates received immigrants, whereas regions with poor prospects have lost population. The labour markets in the "objective 1"-regions thus faced relaxed competition from the supply side. This confirms the view that the rising unemployment rates in these areas were driven by labour demand through a slack in employment growth. We come back to the issue of labour migration in more detail in chapter F.

A4.3.) Education

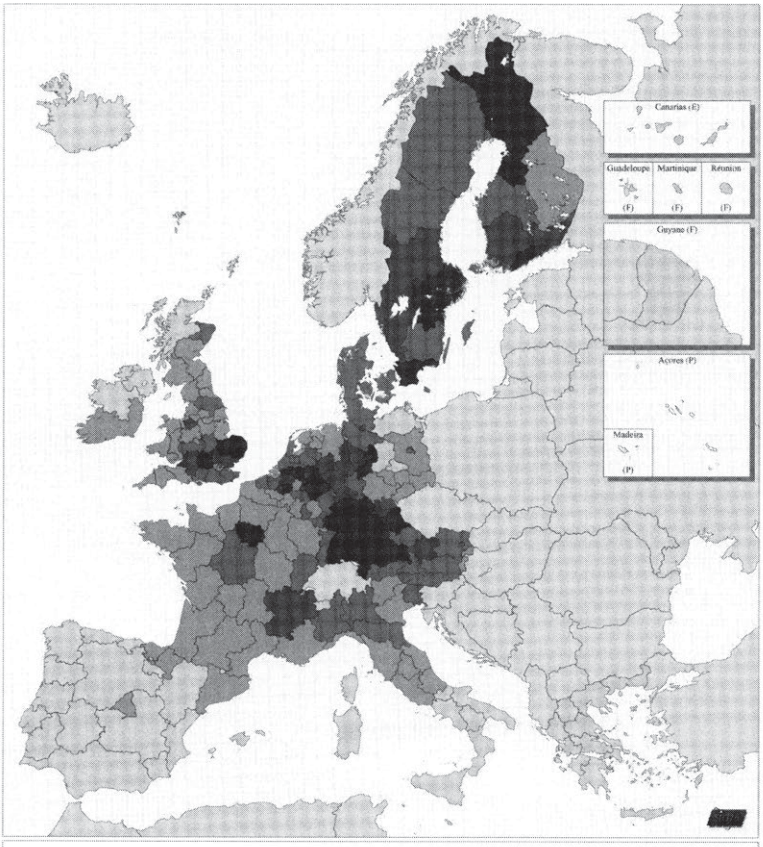
Not only the pure population mass is distributed unequally across the European regions. Also important characteristics of the regional labour forces differ markedly. Map A9 distinguishes three categories of (formal) education levels and depicts the single shares of the qualification groups in the working-age population of EU-15 regions. We just want to highlight two salient features of this map.
Firstly, the labour force in Portugal, Spain, Greece and Southern Italy is significantly characterized by low education levels. But not all "objective 1"-regions have a predominantly low-skilled population. East Germany is a notable exception, which is probably due to the high levels of formal education dating back from years prior to reunification. Secondly, high skilled workers are predominantly found in the economic core belt of the EU, but also in Scandinavia. All in all, the map is consistent with the view that the economic centers are inhabited by a workforce with above-average education levels. Similarly, low education is a predominant feature of the labour force in most "objective 1"-regions.

A4.4.) Innovation and research activities

A last, but nonetheless interesting indicator for the economic geography of Europe is the number of patent applications per million inhabitants per year, shown in map A10. Given the implications from recent endogenous growth theory, this is supposedly an important measure for the prospects of regional growth.



Map A9: Educational level, 2000



Map A10: European patent applications, average 1997 to 1999

Number per million inhabitants

< 30
30 - 90
90 - 150
150 - 210
≥ 210
no data

EUR15 = 119.4
Standard deviation = 116.3
D (Sachsen): NUTS1

Source: Eurostat

© MEGRIN for the administrative boundaries

The theory stresses the role of technological change and innovation for the growth process in modern industrialized economies. In a spatial context, the location of patent applications is important, because it indicates which regions are the innovative centers of an integrated economic area. Since technological knowledge usually spills only imperfectly to other locations, the innovation centers will tend to be the "fast growers".[11]

According to map A10, innovations are extremely locally concentrated in regions from the "European Banana". In the peripheral regions, innovation activity is particularly low. The rich core regions are thus the "growth engines" according to the theoretical implications of the new endogenous growth theory. This is plausible, because there are more resources available to invest in R&D-activities. Map A10 can thus raise some serious concerns about the prospects of poor regions within the EU.

A5) A closer look at the West German Länder

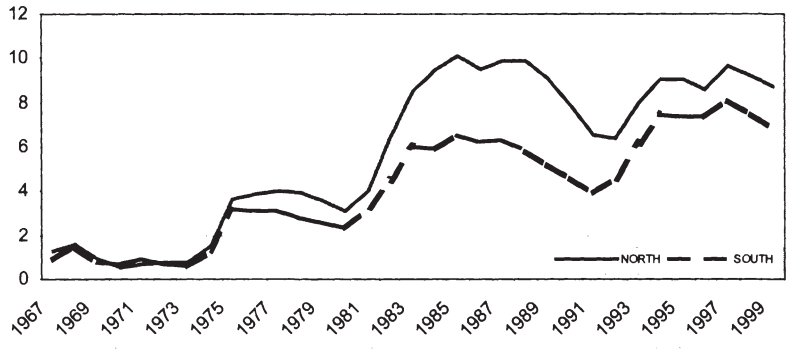
For obvious reasons we are specifically concerned with Germany. It has become apparent in the preceding sections that the main divides within Germany go along the former border between East and West. The existing economic disparities might still, even after more than 10 years since reunification, reflect the extreme structural differences that were build up in the 40 years of division, and some undigested parts from the ´reunification shock´. We therefore want to shift the focus a little bit, away from East versus West Germany and towards the regional disparities within West Germany, which are also substantial and remarkably persistent over the last decades.

The purpose of this section is to give a short illustration, using data from different sources than before. We have annual unemployment data for the West German Länder (i.e. NUTS1) covering the period 1967-1999 from the German microcensus, as well as regional real earnings data directly obtained from the national accounts. The earnings data do not reflect GDP, but rather the gross wage bill per employee. This measure is thus more directly related with the factor labour.

Since the theoretical approaches in the following chapters are often constructed as two-region models, we have chosen a special aggregation scheme for West Germany for expositional purposes. We distinguish two regions, NORTH and SOUTH. The region NORTH contains the five West German Länder Schleswig-Holstein, Niedersachsen, Hamburg, Bremen and Nordrhein-Westfahlen. The regions SOUTH consists of Bayern, Baden-Württemberg, Rheinland-Pfalz, Saarland and Hessen. The area of West-Berlin is excluded. This division in NORTH and SOUTH is instructive, since at the beginning of the observation period, in 1967, both areas roughly had the same population size of about 28 million people.

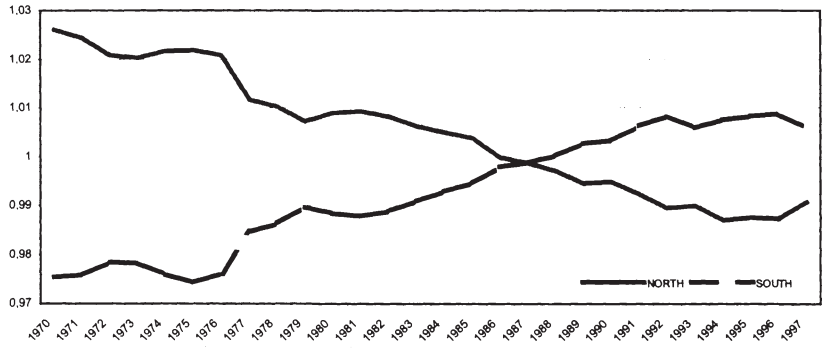
[11] see Grossmann/Helpman (1991, 1994) or Romer (1994) for an overview of this strand of theory

Fig. A11: Unemployment Rates in North and South Germany, 1967-1999



Source: Own calculations based on Stat. Bundesamt, Microcensus (2000)

Fig. A12: Gross wage bill per employee in % of national average



Source: Own calculations based on Stat. Landesamt Baden-Württemberg, VGR of the German Laender

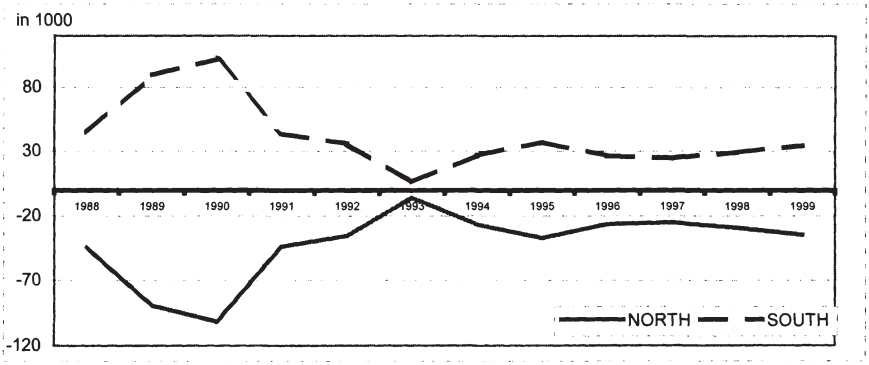
Figures A12 and A13 depict the gross wage bill per employee and the unemployment rate of NORTH and SOUTH relative to the West German average. Starting from a situation of identical unemployment rates until 1975, the macroeconomic shocks of the 70s and 80s caused wide divergence of the regional unemployment rates amounting to 4 percentage points in 1986. After the macroeconomic turbulences calmed down, NORTH and SOUTH now face a stable difference of about 2 percentage points which is persisting since 10 years and shows no tendency to vanish.

Looking at the corresponding spatial structure of effective earnings, it tends to be the case that low-unemployment regions reveal high earnings and vice versa. This observation will play an important role in the following chapters, where we also consider the German system of wage formation in greater detail. For the time being it is important to note that since 1977 the relative gross wage bill is steadily increasing in the SOUTH, and steadily decreasing in the NORTH. Since 1987, the SOUTH has both higher earnings and lower unemployment, whereas the NORTH is lagging behind in both respects.

Not surprisingly, migration flows within West Germany went from NORTH to SOUTH. Figure A14 shows net internal migration 1988-1999. The data refer only to migration within West Germany. Exchange of workers with East Germany and whole Berlin has been subtracted. Obviously the North has been constantly loosing population to the South in all years of the observed time period.[12]

A last, but important characteristic are the spatial differences in skill levels within West Germany. Based on data from the Federal Statistical Office for 1997, we have calculated that approximately 61% of all "high skilled" persons in West Germany were located in the southern part, only 39% in the NORTH. This indicates a slightly higher average skill level of the southern workforce, since in the same year only 59% of total employment has been in the SOUTH.

Figure A13: Net Internal Migration 1988-1999 – West Germany without Berlin



Source: Own calculations based on Migration Data from Stat.Bundesamt (2000)

[12] We come back to the issue of internal labour migration in West Germany in section F2, where we estimate a net internal migration equation based on the data presented in this section.

A6) Summing up the evidence

In this last section, we want to summarize the combined descriptive evidence from the sections A2)-A5) in form of some stylised facts.

- The area of EU-15 can roughly be divided into three groups: The economic core belt "European Banana", the poor and peripheral "objective 1"-regions, and the "intermediates". The important economic indicators that we have considered reveal a spatial pattern that more or less clearly follows this division scheme.

- Regions in the European Banana reveal on average high levels of GDP per capita and GDP per employee, low unemployment rates, high employment growth rates, a high population density, net inward migration, a relatively skilled labour force and high innovative activity.

- The "objective 1"-regions on the contrary have low income levels, high unemployment rates, low employment growth, a low population density with an even declining tendency, as well as a labour force with a high fraction of low skilled workers. Innovation activities are almost negligible.

- The "intermediates" lie in between the two groups with respect to all economic indicators under consideration.

- There has been a convergence process of national income levels, but no regional income convergence over the last decade. Quite contrarily, regional incomes have even (strongly) diverged in some European countries. For the EU as a whole, the time period since 1989 was characterised by a persistence of regional income disparities.

- High and low unemployment areas in the EU are not divided by national borders, but distinguished by cross-country unemployment clusters. These clusters emerged through a polarization process of unemployment rates that was driven by differential employment growth paths of core and peripheral regions within the EU.

- The situation in West Germany, which has been subject to a special examination, is also characterised by large and persistent regional disparities. The area in the south has both lower unemployment rates and higher earnings than the area in the north, which is lagging behind in both respects since approximately 15 years. Internal migration in West Germany went from the North to the South.

B) Macroeconomic theories of unemployment and the “European labour market model”

B1) Introduction

We try to understand the structure of unemployment rates across space in relation to the phenomenon of regional agglomeration. We thus need theoretical background from various fields in economics. We need a theoretical framework to analyse the issue of unemployment, with special reference to its regional dimension, a theory of regional agglomeration, and finally we need to combine the two.
We shall at first start with the theory of unemployment in this chapter. In the introduction it was already said that our analysis will build on a specific regional labour market approach that was pioneered by Blanchflower/Oswald (1990, 1996), and that is comprehensively labelled the *wage curve*. This regional theory is closely linked to one specific macroeconomic approach which is known as the “European labour market model (ELMM)” or the “imperfect competition approach to macroeconomics”.
The purpose of this chapter is to introduce this macroeconomic model, since many essential features of the wage curve theory are built directly on it. In order to demonstrate how the ELMM relates to the long history of economic thought in macroeconomics, we intend to give at first a very brief overview of major developments in macroeconomics over the last century or so, with a special emphasis on theoretical explanations for the phenomenon of unemployment. This historical sketch that is provided in section B2 is by no means an attempt to survey all that has been done in this field. It only tries to present some essential ideas of a few major contributions in a very simplified manner. With this historical information it is easier to see what is actually new about the “imperfect competition approach” to macroeconomics which is then presented afterwards in section B3.

B2) A brief historical overview about macroeconomics

B2.1) The ´classics´

The very traditional perspective in macroeconomics comes from the so called ´classics´. In their view, the real and the monetary sphere of an economy can be completely separated from each other. Inflation is interpreted as a purely monetary phenomenon, with the rate of inflation given only by the growth rate of money supply. The real sector on the other hands is seen as a system of perfectly competitive markets which all tend to clear because of the invisible hand of the market.

B2.2) Keynes and the neoclassical synthesis

This consensus view of the economy as a whole, however, was put heavily under strain with the occurrence of the Great Depression in the late 1920s. It was impossible to explain this event with the traditional economic models at hand, since they were not ready to address why there could be persistent and involuntary mass unemployment in an economy. In this historical period, the macroeconomic theory of Keynes (1936) appeared. The contribution of Keynes was to show that economies can be in an equilibrium where full employment is not reached, because the prevailing level of effective demand is insufficient to render full utilization of existing production capacities. In other words, there exists the possibility of involuntary unemployment in the economy.

It is still subject to ongoing and possibly never ending debates *what exactly* is at the root of involuntary unemployment in the view of Keynes himself.[1] But the most influential interpretation following the *General Theory*, that came to dominate macroeconomic theory and policy for a very long period of time, was the *neoclassical synthesis*. This Keynesian theory is crucially based on nominal wage rigidity, i.e. market failures on the labour market .

The demand side in this model was modelled in the fashion of the famous IS/LM-model (Hicks, 1937). Economists then added a neoclassical production function and a standard neoclassical labour market to the analysis, but (in the name of Keynes) imposed nominal wages to be downwardly rigid. This was then seen as the major source of unemployment, since real wages were prevented to adjust to market clearing levels by these nominal rigidities.[2] The important implication for economic policy that comes from this well-known model is that the government can effectively achieve increases in production and employment through demand side policies (i.e. expansionary fiscal or monetary policy). As long as workers (or respectively, unions) are only concerned with nominal wages in a situation in which involuntary unemployment exists, the government can stimulate aggregate demand and hereby increase the price level. With nominal wages given, this translates into falling real wages and induces an increase in economic activity, possibly to a level where full employment is restored.

If taken further, an interpretation of the (modified) Phillips-curve in the pure vein of the neoclassical synthesis can be based on this logic: with nominal wage inertia, the government can determine the price level through inflationary policies, thereby the real wage and hence the level of employment in this economy.[3] In other words, the government is subject to a downward sloping schedule in the

---

[1] For an overview of the debate about "what Keynes really meant", see Kromphardt (1991), p. 168-179.

[2] For a comprehensive representation of the model see Jarchow (1998), chapter IV.

[3] For this interpretation of the Phillips-curve within the neoclassical synthesis see Carlin/Soscice (1990), ch. I.2; Felderer/Homburg (1993), S. 265 ff.

inflation/unemployment rate-space from which it can choose a desired combination. Although the concept of the Philipps-curve traditionally has been introduced in a different manner into the Keynesian system, [4] this simplifying notion illustrates the main channels of the neoclassical synthesis. Effectively the model predicts that quantity adjustments were needed in this economy to restore equilibrium, because of sluggish nominal variables. Thus, if demand from the private sector is insufficient to render full capacity utilization and prices can not adjust so as to restore equilibrium, the government can step in and take maintenance for low unemployment at the cost of higher inflation.

B2.3) Friedman and the ´natural rate of unemployment´

This neoclassical synthesis was widely accepted in economic theory and policy during the first 20 to 25 years in the post-war period. However, the Keynesian orthodoxy was gradually challenged by political and economic developments and by advances in economic theory. In the beginning of the 1970s, major industrial countries saw inflation and unemployment rise simultaneously in the aftermath of the oil crisis. This development, called *stagflation*, was inconsistent with the conventional Phillips-curve. But prior to this, several scholars have criticised the Keynesian orthodoxy also from a theoretical point of view. In particular Milton Friedman (1968) emphasised that the validity of the Keynesian system effectively rested on money illusion by the workers. In the mechanism outlined above, expansionary fiscal or monetary policy drives up the price level, i.e. it causes inflation. This has only positive employment effects as long as nominal wages are rigid, so that real wages fall and restore full employment. But why should nominal wages remain constant in view of positive inflation? This would require that workers do not recognize the inflation pressure brought about by the government policy, and that they base labour supply decisions on nominal rather than on real wages. Friedman's argument was that this confusion of nominal with real wages can only be true in the short run, because workers might have incomplete information about inflation and might thus do not instantaneously perceive a fall in real wages.

---

[4] Usually, the Phillips-curve was introduced in a pure *ad-hoc* way based on Phillips's (1958) empirical observation that the level of unemployment and growth rate of money wages was negatively correlated. The theoretical concept of the modified Phillips-curve included mark-up pricing (normal costs pricing) on the firms part. If firms set prices simply as mark-ups over wage costs, and wage increases depend negatively on the rate of unemployment (see Lipsey, 1960), a negative correlation between the unemployment rate and the rate of inflation is implied (Samuelson/ Solow, 1960). Thus, the rate of price and the rate of wage increases were identical, but the nominal wages is not necessarily rigid. For the introduction of an ad hoc Phillips-curve in this spirit into the neoclassical synthesis see Carlin/Soscice (1990:69 ff).

Once they do, however, workers will adapt their nominal wage claims so as to restore the level of real wages that was prevailing before the policy intervention. Unemployment will thereby also move back to its old level. Effectively, all that has changed is that prices and nominal wages are now on a higher level after the political intervention.
In the long run, there is no trade-off between inflation and unemployment, but rather a vertical Philipps-curve above some unemployment rate that Friedman called the ´natural rate of unemployment´. This natural rate according to Friedman (1968) is *"[...]the level which would be ground out by the Walrasian system of general equilibrium equations, provided that there is imbedded in them the actual structural characteristics of the labour and commodity markets, including market imperfections, stochastic variability in demands and supplies, the cost of gathering information about job vacancies and labour availabilities, the costs of mobility, and so on"* . In the long-run, unemployment is thus determined in the "classical" way, i.e. on a competitive labour market where unemployment is either voluntary or due to rigidities and the malfunctioning of labour market institutions. With the use of macroeconomic demand management policies, policymakers can keep unemployment below this natural level only at the cost of steadily accelerating inflation, which led to the name of a "non-accelerating inflation rate of unemployment (NAIRU)" as a synonym. Put differently, demand side policies can not be used systematically to reduce unemployment, they are neutral in the long-run and have only short term real effects when agents confuse nominal and real wages.

B2.4) ´New classical macroeconomics´ and rational expectations
The short term effects of monetary and fiscal policy in the Friedman-model were due to the workers' temporary misconception of nominal and real wages. This was brought about by systematically wrong expectations about inflation, which in the Friedman-model were formed in an adaptive way. In the 1970s and 1980s, however, the ´new classical macroeconomics´ appeared that questioned this construction and urged for a more radical reassessment of macroeconomics.
Economic agents build expectations about future developments. Since the future is by definition uncertain, no agent can make predictions that are always absolutely correct, as there is always the possibility for stochastic and unexpected events. But the proponents of the ´new classical macroeconomics´, most notably Robert Lucas, Thomas Sargent, and Edward Prescott, believed that agents will not make *systematic* errors. Friedman's use of adaptive expectations that left room for short run policy effects was replaced with the construction of *rational expectations*. This means that agents incorporate all available information that is at their disposal at any moment into their expectation formation.
The consequence is that workers, when faced with inflationary government policies, would not misperceive nominal and real wages, but would correctly expect

falling real wages. Consequently, there is no room for the sort run mechanism á la Friedman, but inflationary policies are offset right away by changes in nominal wages in order to maintain the prevailing level of real wages. Put differently, because of rational expectations, output and employment in the models of new classical macroeconomics are permanently at the natural rate, except maybe for random disturbances (Lucas, 1972). The inherently stable economic system leaves no room for stabilization policy. Any attempt by policymakers to influence the output level via macroeconomic policies would only lead to inflation, or to a complete crowding out of private activities.

The ´new classical macroeconomics´ can be viewed at as a modern and mathematically more articulate version of the pre-Keynesian classical economics. Both rest essentially on Walrasian general equilibrium theory with perfect competition and instantaneous price adjustments mandated by an imaginary auctioneer, and the ´new classical´ school has added rational expectations to this model. As it is well known, the Walrasian price *tâtonnement* process leads to the clearing of all markets in the economy, including the labour market.

The principal problem of the ´new classics´ is thus more or less the same as it has been for the ´old classics´. General equilibrium theory needs to answer why there can be business cycle fluctuations in a model where output is always at its natural level and agents have rational expectations.[5] How can the observed short-run fluctuations of output and employment be explained in such an environment? Similarly, also the ´new classical macroeconomics´ would have to come up with a convincing story why something like the Great Depression can unravel when markets work that perfectly.[6]

One way to deal with business cycle fluctuations in the context of ´new classical macroeconomics´ was developed in the early 1980s and became known as "real business cycle theory (RBC)". RBC-models, pioneered by Kydland/Prescott (1982), are market clearing models with agents who form rational expectations. Fluctuations in output and employment are seen as resulting from exogenous technological shocks hitting the production function, and thereby the marginal productivity of labour. Agents in these models decide on intertemporal labour supply paths by equating the obtainable wage rate (the marginal product of labour) with the value of leisure. In view of adverse technological shocks, agents may find it optimal to engage in leisure or non-market activities, as falling wages can fall short of reservation wages. A typical RBC-model thus implies that output is always at its natural level, and this level fluctuates due to technological shocks. These models are therefore often called "equilibrium business cycle models".

Any unemployment is entirely voluntary and simply determined by the optimal intertemporal allocation of time. But the use of the term "voluntary unemploy-

[5] This puzzle was already noticed by Hayek (1933), p. 33 for the old general equilibrium theory.

[6] For this debate see e.g. Lucas (1980), Tobin (1977) and Modigliani (1977).

ment" needs some further classification. Recorded unemployment rates by definition only include such persons who actively seek for jobs, i.e. no persons who have chosen not to belong to the labour force. However, this critique does not necessarily flaw the predictions of RBC-models and all other classical models that view unemployment as "voluntary". Firstly, these models might refer to a broader measure of joblessness (e.g. the "non-employment rate") that also includes labour force drop-outs that *in principal* are interested in accepting a job, but not at going wage rates. And secondly, unemployed workers in reality have some degrees of freedom to signal job search activities, even though they are currently not interested to work. The main implication of classical and RBC models is simply that unemployment can not be viewed at as a phenomenon where individuals would be willing to accept jobs at going wage rates, but receive no job offers and thus remain unwillingly unemployed.

B2.5) The Keynesian response

The ´new classical´ model of macroeconomics as well as RBC-models make extreme assumptions with respect to the rationality of agents and their state of knowledge, but also with respect to the functioning of markets. In particular, the models rest very heavily on the instantaneous price adjustments brought about by the Walrasian auctioneer.

The first wave of Keynesian criticism against the new classical model was mainly concerned with the artificiality of the price tâtonnement. What are the implications for the economy if this process is not functioning? Recall that in theory the Walrasian auctioneering process works such that some initial price vector for all commodities in the economy is announced, all market participants signal the quantities they are willing to buy and sell at these prices, the auctioneer calculates excess demands and excess supplies on all markets, and calls out a new pricing vector that brings quantitative supply and demand closer together. This process is repeated until the market clearing price vector is found. Only then do transactions take place, at the equilibrium prices that equilibrate supply and demand on all markets.

Criticism against the artificiality of this theoretical construct was articulated by Clower (1965) and Leijonhufvud (1967, 1968) even before the emergence of the ´new classical´ macroeconomics, with the original intention of "re-interpreting Keynes" in opposition against the neoclassical synthesis. Their works, however, where the precursors to much of the Neo-Keynesian critique in the 1970s, e.g. by Malinveaud (1977). Central to this literature was the analysis of economic systems when an auctioneer is absent and transactions take place at "wrong" prices.[7]

[7] For a comprehensive overview of this school of Neo-Keynesian economics see Felderer/Homburg (1993:287ff.) or Carlin/Soscice (1990:106ff.).

Clower (1965) has emphasised that it is highly artificial to assume that transactions only take place once the market clearing price vector is announced. In reality, there is continuous trading also at non-market clearing prices. Because markets do not instantaneously clear, agents face quantity constraints on some markets, which via a budget constraint also affects economic plans with respect to other markets. For example, if an initial price vector exists where the wage rate is higher than the theoretical market clearing level, agents can not sell the desired amount of labour, but rather face a quantity constraint on the labour market. Nevertheless, trading already starts at this pricing vectors and agents do not wait for an imaginary auctioneer to calculate market-clearing prices. Because agents supply less labour than intended, they can also not realize their desired consumption plans. They are forced to demand a lower quantity of goods, because actual income is below desired income. This is the essence of the "dual decision hypothesis". Agents first form desired economic plans. After having perceived quantity constraints, they have to adjust to market conditions and choose their actual economic plans. These actual plans are the optimal household decisions subject to the quantity constraints at the given non market-clearing price vector.

The critical question is then whether there exist market forces that realign prices such that agents can gradually come to realize their desired economic plans. As shown by Dréze (1975), this is difficult or even impossible if there is no auctioneer and all agents are atomistic. In other words, an economy might be trapped in a sub-optimal equilibrium where markets do not clear, because the Pareto-optimal Walrasian equilibrium is unknown to single agents and an adjustment process can not be triggered.

Leijonhufvud (1967) emphasised that at best it takes a considerable amount of time for an economy to converge from a Keynesian equilibrium with quantity constraints towards the Walrasian general equilibrium. During the transition, there is room for government action, including the exploitation of multiplier effects through stabilization policy.

To sum up, this first wave of Neo-Keynesian economics with the so-called "fixed-price models" was concerned with the analysis of non-Walrasian equilibria.[8] Involuntary unemployment can follow in these models, just as there is room for macroeconomic policy under some circumstances.[9] It must, however, be stressed that the theoretical framework in this version of Keynesian economics was one of

---

[8] Felderer/Homburg (1993:300) point out that the term "fixed-price" might be misleading. It means nothing more that there is no tâtonnement and no price auctioneer.

[9] In particular Malinvaud (1977) was concerned with the distinction of ´classical´ versus ´Keynesian´ unemployment, i.e. with the question if rising real wages lead to more unemployment because of the neoclassical labour cost argument, or to less unemployment because of a boost in demand. Governments should therefore thoroughly analyse if unemployment is ´classical´ or ´Keynesian´ before taking any action.

atomistic agents and perfect competition on goods and labour markets. The bottom-line message was that within this very conventional market system, economies can still end up in sub-optimal equilibria with unemployment, because price adjustments do not work well.

B3) The "European labour market model (ELMM)"

During the late 1980s, a different macroeconomic approach was developed mainly in Europe that was popularised under the name ELMM, or "imperfect competition approach to macroeconomics". It is generally considered to be an approach in the Keynesian tradition, which is supported by the following citation that comes from one of its most famous adherents

> *"My interpretation of the empirical evidence is that the magnitude and persistence of changes in statistically recorded unemployment are too large to be explained as variations in search or frictional unemployment, intertemporal substitution of leisure or a misinterpretation among economic agents regarding inflation or relative price and wage changes in the context of market-clearing models. The apparent unhappiness of many unemployed workers do not suggest that they have simply, in an optimal fashion, reallocated leisure in response to perceived temporal or intertemporal wage changes. [...] My inference from all this is that market-clearing approaches to the labour market cannot possibly be appropriate for an analysis of short- and medium-term macroeconomic developments."*
>
> Lindbeck (1992:209f.)

Probably all Neo-Keynesians would subscribe to this viewpoint. But the new school departed from the "fixed-price model" described above, since there was a growing dissatisfaction with its micro-foundations. It appeared as if this literature has examined the properties of Walrasian systems with price rigidities, described non-Walrasian equilibria and classified ´new classical macroeconomics´ as a very special case, where the prevailing price vector happens to be the market-clearing one.

But this theory did not point to the *origins* and *sources* of price stickiness.[10] Moreover and more importantly, the use of perfect competition as the reference system of markets became increasingly unsatisfactory to economists. In particular for European countries it seemed much more appropriate to acknowledge that

[10] This issue was later developed in much more detail in a different string of "Neo Keynesianism" that was concerned to formulate rigid micro-foundations for price stickiness in the context of Walrasian market-clearing models (menu costs etc.). For an overview of this literature see Gordon (1990).

goods markets and especially labour markets were not working perfectly, but were rather characterised by substantial and *systematic* imperfections. Thus, theoretical effort of the ELMM was devoted to the formulation of an alternative economic model that was based on the concept of imperfect competition. The explicit departure from the concept of perfect competition is the important feature that distinguishes the ELMM from other theoretical macro approaches.
Working with imperfect competition has the important advantage that it explicitly involves wage and price setting of economic agents. The Walrasian auctioneer is no longer needed, nor are discussions about possible deficiencies of the tâtonnement process: it does no longer exist. Furthermore, it is much more straightforward in this environment to discuss the sources of price stickiness, which might result from the market power of agents operating in imperfectly competitive markets.

B3.1) Why is there insider market power in the labour market?
The labour market does not clear at a given price level, because nominal wages (and hence real wages) do not adjust to the market-clearing level. The outsiders in the labour market are involuntarily unemployed, because it is not possible for them to replace the employed insiders by underbidding wages. Why is that? The general answer is that insiders have some degree of market power that they can use to defend their position against outsiders. Lindbeck (1992) distinguishes four broad channels why employed insiders have the power to maintain wages at higher than market clearing levels and prevent underbidding by outsiders.

a) Social norms: A first channel might be that wage underbidding is regarded by society as an unacceptable form of behaviour. The unemployed might refrain from offering labour at lower than current wages, because they fear social stigmatisation for having “stolen jobs”. Similarly, firms might also not want to accept low wage offers by outsiders, because they fear negative drawbacks on their reputation. If consumers regard the firm’s recruitment behaviour as unacceptable, they might become negatively conditioned against the products of the particular firm. Lindbeck (1992) calls these considerations the implicit eleventh and twelfth commandment, whereby individuals are told “thou shalt not steal jobs from thy comrades by underbidding their wages” and respectively firms are urged “not to encourage nor accept job theft by way of underbidding”. This view of human behaviour is of course inconsistent with the assumption of a purely selfish homo oeconomicus, since otherwise outsiders would have no problem whatsoever to underbid wages. Recent evidence from experimental economics (see e.g. Fehr/Fischbacher, 2002), however, points to the fact that individuals in the real world are also driven by non-selfish motives like social norms.

Nevertheless, it is surely insufficient to explain nominal wage rigidity, insider market-power and hence unemployment only through the honourable social behaviour of the unemployed who would rather suffer than cause wage decreases for somebody else. There must be other sources of insider market-power.

b) Union wage setting: Wages are not determined by market forces, but rather are set by trade unions, or respectively unions and employer associations bargain over wages. This type of wage determination is specifically relevant for continental Europe, where union wages typically apply to more than 80% of all employees (Ochel, 2000). Unions represent the interests and bargain on behalf of their members. It is the primary concern of unions to maximize their members' utility function, which includes their wage income subject to the interrelated constraints on employment. Typically this optimisation process will lead unions to set wages above the market clearing level (Layard/Nickell/Jackman, 1991:83ff.).[11] But this alone does not explain why there is no wage underbidding in the economy. Even though unions might bargain high wages for their members, non-union members in principle have the freedom to offer labour to employers at lower salaries.
This has led Lindbeck (1992) to infer that union models alone do not explain why labour markets do not clear. However, this underestimates the role that unions play as political pressure groups. Unions do not only bargain about wages on behalf of their members, they also seek to prevent underbidding through political lobbying activities. In Germany e.g., bargaining results may under certain circumstances be declared as generally binding by the government for all firms and workers in a specific sector, which legally enforces union wages and prevents underbidding (Kirsch/Bispink, 2002).
In general, the high rates of bargaining coverage in continental Europe that are far greater than union membership rates indicate that union wages have a high factual compulsoriness. Outsiders are both legally and factually restrained from underbidding in many important sectors of the economy.

c) Turnover costs: A natural complement to union models is the insider-outsider-theory by Lindbeck/Snower (1986). This theory specifically stresses the *source* of insider market power. Employed insiders know that they can not easily be replaced with outsiders, because of various costs associated with the turnover. The most direct form of turnover costs are hiring and firing costs as

[11] It is commonly believed that the deviation of union wages from market-clearing wages particularly depends on the level of aggregation at which bargaining takes place, because of the different abilities of a union to exter-nalise the employment consequences of its wage claim (Calmfors/Driffill, 1988).

well as training costs. But also more subtle costs might accrue to firms, if e.g. the productivity of prospective entrants is lower because old insiders engage in harassment behaviour and refuse to cooperate with new entrants.
Unions, as the interest group of insiders, again might lobby for various political steps to increase turnover costs and thereby the market power of employed workers, e.g. through advocating employment protection legislation.
The existence of these costs grants insiders with market power, as they can claim higher wages than outsiders. Of course this market power is not unlimited. But within a feasible range, insiders can carry through wages which on an aggregate level lead to non-market clearing on the labour market.

d) Efficiency wages: Firms might also consciously decide to pay wages above the market clearing level in order to boost the productivity or profitability of their incumbent workforce. This might be so for various reasons, highlighted by various different efficiency wage models.
All models have in common that they take labour to be a very special input in the production process, hardly comparable to capital and machinery. Since workers are human beings, they are very specialized with respect to their characteristics and skills. A firm might thus look at its incumbent workforce as an asset portfolio that is worth to retain and to develop. Moreover, workers are able to vary their personal work performance. The possible range starts at zero performance, and goes up to the maximum performance that a worker can supply with given skill level and with the given capital stock at hand. Unlike machinery, workers will consciously adjust their effort in response to firm internal and external incentives. Firms can only incompletely monitor the personal performance of single workers. Within this "human" environment, the wage paid by a firm becomes an instrument to govern the performance and the motivation of the incumbent workforce (Stiglitz, 2002).
As Layard/Nickell/Jackman (1991:150ff.) put it, firms use efficiency wages in order to "recruit, retain and motivate" its workers, which refers to different particular efficiency wage models as e.g. Schlicht (1978); Salop (1979); Akerlof (1982); Shapiro/Stiglitz (1984) or Akerlof/Yellen (1990).
We do not want to introduce nor survey the details of the different efficiency wage models.[12] For the issue of wage underbidding, however, all existing models share qualitatively similar predictions: Since employers pay higher than market wages out of free will, they are not interested in accepting low wage offers by outsiders. They can use outside unemployment as a pressure device, but motivation, fairness or screening considerations (depending on the specific model) induce the firm to keep on paying higher wages. Consequently, outsiders can not enter jobs at these firms and must stay unemployed.

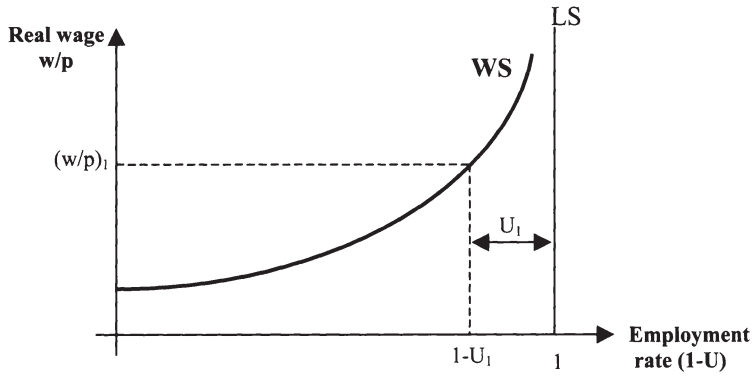
[12] For an overview see e.g. Yellen (1984).

It is quite conceivable that the single arguments a)-d) are not mutually exclusive, but rather complementary. The sources of insider market-power highlighted above are at the core of the aggregate description of the labour market under imperfect competition, because they clarify how wage setting in imperfect labour markets can cause unemployment. It is now important to go also the other way around and show how the level of unemployment in turn influences the wage setting behaviour. To this issue we turn next.

B3.2) The aggregate wage setting curve

The aggregate description of the labour market in the ELMM does no longer consist only of an aggregate labour supply and an labour demand function as under perfect competition. A central relation in the ELMM is the *wage setting curve* (WS)[13], illustrated in fig. B1. The curve WS is an upward sloping curve in the real wage/employment rate space. It shows all combinations of real wages and employment rate that are consistent with equilibrium in the imperfectly competitive labour market in an economy. The combination $(w/p)_1$ and $(1-U_1)$ is e.g. one such combination.

Figure B1: The wage setting curve in the ELMM



For simplicity we have assumed that labour supply (LS) is perfectly inelastic, so that the LS curve is vertical above the full employment situation with $(1-U_1)=1$. The unemployment rate for any given wage rate is thus given by the vertical distance between the WS and the LS curve.

[13] There is no unique terminology for labelling the WS curve. Carlin/Soscice (1990) use the term "bargained real wage curve". Lindbeck (1992) calls WS also simply a "wage setting curve". Blanchard/Katz (1997) call WS a "supply wage relation" (WS). Layard/Nickell/Jackman (1991) also use the most simple terminology, wage setting curve (WS).

There are various ways to provide economic intuition for this positive slope, which correspond to the intuition about imperfect competition in the labour market that was given in section B3.1.). Two "stories" have become most common in recent years, union models and efficiency wages, although several authors (e.g. Blanchard/Katz, 1997) also base WS curves on different micro-foundations.

The first explanation for the positive slope of the WS curve rests on union wage setting and insider-outsider theory (points b and c from section B3.1.). Suppose the labour market at stake is heavily unionised. The union sets nominal wages in relation to the given or expected price level, which in the short run implies a real wage. It is quite intuitive that the union's bargaining power and thereby the bargained real wage is a positive function of the employment rate. With high unemployment, insider power is low. Outsiders are willing to underbid wages more aggressively, and insiders can rely on turnover costs to a smaller extent. In tight labour markets, unions can negotiate higher real wages. Strike announcements are more credible, the pool of remaining outsiders is characterised by unfavourable personal characteristics. And thus insiders can rely more heavily on their market power.[14]

The second common foundation for the positive slope of the WS curve comes from efficiency wage theory (approach d in B3.1.). It is also straightforward to see why employers have to increasingly rely on efficiency wages in tight labour markets. If unemployment is high, it alone acts as a disciplining or motivating device for workers who fear to loose their jobs.[15] Employers do not have to use additional instruments. If the unemployment rate is low and the outflow rate from unemployment is high, joblessness can be perceived to be a minor thread by workers. They might consequently feel induced to engage in shirking behaviour, low commitment to employers, low investments in firm-specific human capital etc. If a particular job pays above the market level, however, the commitment and the performance of workers can be maintained through efficiency wages, because workers fear to loose their privileged position at their current employer. The WS curve then represents the level of real wages that firms are willing to pay in order to achieve their motivation or screening objectives for any given employment rate.

All in all, there exist plausible economic explanations for a WS curve as depicted in fig. B1. In this section we have presented only intuitive arguments rather than presenting the details of one particular model for a WS curve. In the next chapter, when we discuss regional labour markets and introduce the wage curve, we will

---

[14] For a more complete discussion of this approach, and for a formal derivation of the WS curve under union wage setting see Layard/Nickell/Jackman (1991:83ff.) or Carlin/Soskice (1990:387 ff.).

[15] Blanchard/Katz (1997:53 f.) rightly notice, it is really the outflow rate from unemployment that determines the strength the perceived penalty, not so much the overall unemployment rate. However, the overall unemployment rate is commonly used as a proxy for the labour market prospects of the unemployed.

use one specific theoretical approach to explicitly derive a wage curve. This will be one particular efficiency wage model, namely the *shirking approach* by Shapiro/Stiglitz (1984). It is worth mentioning that the same shirking approach could also be used to justify a macroeconomic WS curve. And one should also keep in mind that a WS curve (and a wage curve) can be based on more than one plausible economic argument.

B3.3) The aggregate price setting curve

The WS curve represents equilibrium combinations of (w/p) and U stemming from the labour market. The second central relation in the ELMM is the aggregate representation of product market equilibrium, which in our terminology will be called the *price setting curve* (PS). It is a distinguishing feature of the ELMM to assume imperfect competition in the labour market and derive an upward sloping WS curve. For the product market, however, there exist approaches both with perfect and with imperfect competition.

a) Perfect competition in the product market

With perfect competition in the product market, there is again price taking behaviour of atomistic firms, as well as the requirement that firms make zero profits. It is well known that under these circumstances competitive firms pay labour according to its marginal product.

Any firms maximizes profits $\pi = p\,Y - w\,N - r\,K$, where Y is output, N is labour and K is capital. It has to take factor prices r and w as given. Suppose that capital is internationally traded and thus capital costs r are determined on world markets. The nominal wage rate w on the other hand is dictated by the imperfectly competitive labour market. Output prices p can not be changed by any single firm, but they ground out from the Walrasian tâtonnement process.

In equilibrium, there is only one price level that realigns real wages such that there is equilibrium in the product market. This real wage must be consistent with efficient production along the minimum cost schedules, profit maximization and zero profits for all firms. Under perfect competition this real wage is given by

$$\frac{w}{p} = \frac{\partial Y}{\partial N} = MPL \qquad \text{(B.1)}$$

Equation (B.1) is the equilibrium condition for the product market and describes the *PS curve,* which for the case of perfect competition can also simply be labelled

the labour demand curve.[16] Its shape is determined entirely by the marginal product of labour (MPL), i.e. by the properties of the underlying production function.
If the capital stock is fixed and the production function exhibits neoclassical features, one would typically expect that labour faces diminishing marginal returns. If this is so, the PS curve under perfect competition becomes a downward sloping curve in the real wage/employment rate-space. With a low employment rate, the marginal product of labour is high. Product market equilibrium is obtained at a high real wage level. As the economy comes closer to full employment, labour faces diminishing returns, and product market equilibrium is only compatible with a lower real wage.
However, it is subject to considerable dispute if marginal costs are indeed rising on an aggregate level (e.g. Blanchard/Fisher, 1989:463 ff.). For several reasons it is also conceivable that the MPL is actually constant. This will particularly be the case if the production function is such that capital and labour are used in fixed proportions. If this is the case, as in Blanchard/Katz (1997:55), the real wage consistent with product market equilibrium is independent of the employment rate. The PS curve is then simply a horizontal line in figure B1. The curve would also be horizontal if labour is seen as the only variable output, and constant marginal returns are assumed.
But no matter if the PS curve is flat or downward sloping, the general equilibrium in this economy would lie at the intersection point of the PS curve with the upward sloping WS curve (at point B in figure B2). At this point, a combination of the real wage and the employment rate prevails that is consistent with equilibrium in both the product and the labour market.[17] We will come back to the case with perfect competition when discussing the wage curve approach of Blachflower/Oswald (1996). But now we turn to the (probably) empirically more relevant and the theoretically more interesting case with imperfect product market competition.

[16] Under perfect competition the terminology *PS curve* it not fully appropriate, since the firm sector has no market power to set prices. However, the labelling is chosen in order to highlight the analogy with the imperfect competition case.
[17] The issue of market clearing is this more is a bit more complex and requires some discussion. Since there is excess supply in the labour market, by Walras' law there must be excess demand on some other market. As Lindbeck (1992:192) has put it, there is indeed ´notional´ excess demand in the product market, since workers are income constrained because of unemployment. They can not demand as much goods as they would desire based on full employment considerations, i.e. at the Walrasian equilibrium represented by point C in fig. B2. The ´dual decision´ hypothesis of Clower (1965) thus pops up again. However, perfect competition in the product market here simply means that there is no additional quantity rationing stemming from market power in the goods market.

b) Imperfect competition in the product market

With imperfect competition, firms have market power and can actively set prices. Suppose that there is a large number of firms in the economy, each producing a distinct but symmetrical good under monopolistic competition. Each firm now maximizes profits subject to a downward sloping demand curve for its specific product. For simplicity we assume that labour is the only variable input of the firm. Profits are given by

$$\pi = p(Y(N))Y(N) - wN \tag{B.2}$$

where Y(N) describes the short run production function for any firm, depending only on employment N. The downward sloping demand function is given by p(Y(N)). Maximizing (B.2) with respect to N yields the familiar rule that prices are a mark-up over marginal costs, which by definition is equal to the nominal wage divided by the MPL

$$p = \frac{1}{1 - 1/\sigma} \frac{w}{MPL} \tag{B.3}$$

σ is simply the absolute value of the price elasticity of demand. To arrive at an expression for the PS curve under imperfect competition, we simply have to rearrange (B.3) in order to obtain

$$\frac{w}{p} = (1 - 1/\sigma)MPL \tag{B.4}$$

The slope of the PS curve in the real wage/employment rate-space depends on two factors: a) whether the MPL is constant or declining in the aggregate employment rate, and b) whether the demand elasticity is a function of aggregate employment or not. As far as the MPL is concerned, the same consideration apply as for the case with perfect competition. Labour might face diminishing returns if the capital stock becomes a binding factor, but the evidence for this proposition in aggregate data seems scant. With respect to the demand elasticity σ, matters are also controversial. It has become common to work with iso-elastic demand functions e.g. of a CES-type. With these functions, the slope of the PS curve is entirely determined by the slope of the MPL. But some authors have argued that demand elasticity tends to be a pro-cyclical variable (Bils 1987, 1989). If this is so, the mark-up $(1-1/\sigma)^{-1}$ is decreasing in the employment rate and not constant.

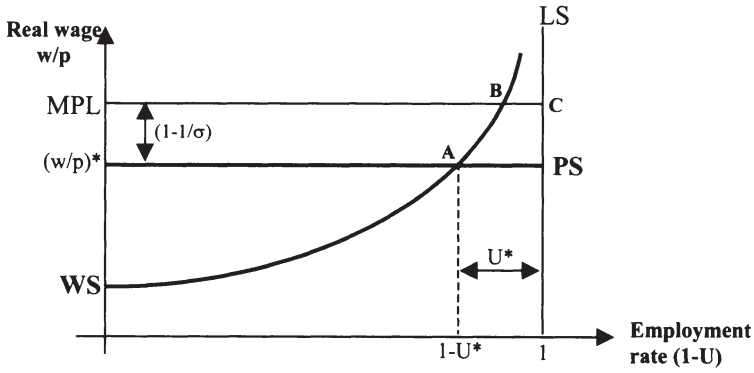
Neither theory nor empirical evidence are fully decisive on the slope of the PS curve under imperfect competition, since neither the slope of the marginal cost curve nor the behaviour of the mark-up is known on an aggregate level. For most insights of the ELMM it is, however, not essential whether the PS curve is horizontal or downward sloping.

B3.4) Equilibrium in the ELMM

For simplicity and expositional purposes we will work with a flat PS curve. This slope can follow either because both the MPL and the demand elasticity σ are constant, or if the MPL is declining whereas σ is rising in (1-U). We will consider here the former possibility. Figure B2 graphically summarizes the ELMM, and highlights the differences between a competitive Walrasian equilibrium (point C), and the equilibrium points with imperfect competition in both product and labour markets (A), or only in the labour market (B).

The Walrasian equilibrium with perfect competition in goods and labour markets is simply at the intersection of labour supply (LS) and labour demand (MPL). Of course, full employment is rendered under this competitive general equilibrium. But we have argued above that this Pareto-optimal situation might not develop, because of systematic market imperfections.

Figure B2: Equilibrium in the ELMM



Product market equilibrium with perfect competition in the goods market is graphically described by the labour demand curve (MPL). For the imperfect competition case, the pricing behaviour of firms must be taken into account that is described by equation (B.4). The PS curve, the "labour demand determined real wage", will lie below the MPL. Equilibrium in the imperfectly competitive labour market is described by the upward sloping WS curve. For a general equilibrium, both the labour and the product market need to be in equilibrium.

With imperfections in both markets this is the case at point A. The real wage (w/p)* must be equal to the level that is determined by the PS curve. The associated unemployment rate U* can be called the "equilibrium rate of unemployment" (Carlin/Soskice, 1990), or, in reminiscence to Friedman's expression, even the "natural rate of unemployment" (Blanchard/Katz, 1997).[18]

It is important to understand in what respect U* is an *equilibrium* rate of unemployment. Probably the best way to think about this issue has been proposed by Layard/Nickell/Jackman (1991:8 ff.). They put strongest emphasis on the fact that *equilibrium* in the imperfect competition approach does not imply *market clearing*. It only means that private economic plans are compatible with each other, and that a system will return to the equilibrium configuration in case of a random disturbance.

The WS and the PS curves represent such private plans. The WS curve represents how nominal wages are set in relation to goods prices in the labour market. The PS curve shows how firms on goods markets set output prices in relation to costs (i.e. nominal wages). Both price- and wage setters make their economic plans in *real* terms, and accordingly use their nominal action parameters. Suppose the WS curve is based on a monopoly union model. The union's bargaining power, and hence the real wage claim, positively depends on the employment rate. The union claims a nominal wage so as to achieve the desired level of the real wage, taking into account (or forming expectations about) the prevailing price level. Equally, price setters are only concerned with the real value of their profits. They set prices in relation to the nominal wage in pursue of their economic plans in real terms. Only at the equilibrium level (w/p)* are the private economic plans compatible with each other. The adjustment mechanism that brings about the consistency of the real claims is the level of unemployment.

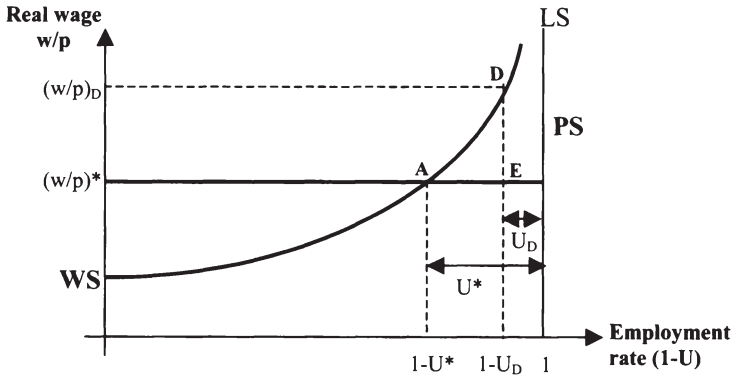
The working of the ELMM can be seen best when considering an example. In figure B3 the general equilibrium is given by point A. This point represents the constellation of real wage and (un)employment, where output claims of firms and union are compatible. Suppose the union expects output prices to rise by 2% in the next year. Since the economy is in its long-run steady state in point A, the unions will claim nominal wage increases also equal to 2 %. The real wage is thus constant, as well as the real level of profits and the unemployment rate.

Now suppose that a shock pulls the economy out off the equilibrium constellation A and to some point D that is associated with an unemployment rate U_D below the equilibrium level. The low unemployment positively affects the union's bargaining power. With adaptive expectations, the union expects inflation to remain at 2% in the next period. But at U_D the union will want to increase nominal wage by

[18] One can see that the equilibrium rate of unemployment is lower with perfect than with imperfect product market competition. This is of course due to the fact that imperfect competition in the product market imply quantity rationing from firms, and thus lower labour demand.

more than 2%, say by 5%, in order to reach the desired real wage level $(w/p)_D$. However, this real wage level would put the real profits under strain. Recall that firms set prices as a mark-up over wages, and the equilibrium conditions for the product market are only satisfied at the real wage level (w/p)*.
At point D there is an inconsistency of economic plans. Put differently, the desired real wage income and the desired real profit level do exceed real output. Since this is not possible, an adjustment needs to occur that brings the competing claims of firms and unions for real output shares in line again. This will occur at first through a change in output prices. Typically the timing of the pricing decisions is assumed to be the following: first the union make the nominal wage claim based on expectations about future inflation and unemployment. Afterwards, pricing decisions of firms are made on the basis of the nominal wage claim, which then determine the actual rate inflation.[19] Since the firms seeks to maintain their real level of profits, they react to the nominal wage claim and increase prices also by 5% instead of 2%. With inflation equal to 5%, the real wage level is consistent again with product market equilibrium. Graphically this situation is given at point E in fig. B3, where the equilibrium real wage (w/p)* prevails, but on a higher nominal level than before.

Figure B3: Adjustment in the ELMM



At point E, unemployment is still at the level U_D below the natural rate U*. If in the next period the union still seeks to increase real wages, it would now have to claim nominal wage increases of more than 5%. Again, firms would respond by

[19] In case the WS curve is based on efficiency wages rather than on a union model, both nominal wages and output prices are set simultaneously by firms that just need to bring together goods market considerations and motivation objectives. The timing of pricing decisions is discussed more intensively in Carlin/Soskice (1990:163).

price increases proportional to the nominal wage increase. The ELMM thus predicts that unemployment can be kept below the natural rate only at the costs of accelerating inflation. Similarly, unemployment rates can remain above U* only in association with permanently falling inflation. Only at the equilibrium A are real wages claimed by unions and firms compatible with each other, and there is no need for a change in the rate of inflation. The unemployment rate U* can therefore be understood as the non-accelerating inflation rate of unemployment (NAIRU). However, note that inflation at U* does not necessarily have to be zero. It is just implied that the rate of change of nominal wages and output prices is identical.

But will the system be endogenously driven back to the equilibrium value U* after an exogenous shock? In other words, are there endogenous forces that bring unemployment back from U_D to U*? So far we have only shown that maintaining unemployment at U_D imposes costs in form of an accelerating wage-price spiral. But unemployment dynamics have not been explicitly spelled out.

Many authors such as Carlin/Soskice (1990) or Lindbeck (1992) are in fact not very explicit about this issue. However, a long-run convergence of unemployment to its natural level U* is the only economically plausible possibility. Otherwise, as pointed out by Layard/Nickell/Jackman (1991:12f.), there would be an everlasting wage-price spiral. The intuition for the transition from U_D to U* is easier to provide in the context of efficiency wage models at the core of the WS curve. Since here both wage and price setting is actually done by firms, real wages and (un)employment are determined simultaneously. The combination of an unemployment rate U_D and a real wage (w/p)* as in point E is not sustainable, because unemployment is too low for given real wages to assure the optimal level of worker morale and effort. The firm sector will thus fire workers until the economy is at the equilibrium rate of joblessness, U*.

In the context of a union model, firms do not set nominal wages. They set nominal prices in relation to the unions' wage claims. But they also choose the level of employment. Because we have assumed a flat MPL curve, firms can not speculate on a higher marginal productivity of workers at a lower employment rate. If the PS curve were to slope downwards, it would again be easy to see how the economy would converge to its long-run steady state at point A. But the basic trade-off implied by a downward-sloping labour demand curve, that plays a very prominent role in other collective bargaining models such as McDonald/Solow (1981), is absent with this linear technology. However, there is no reason to believe that firms will keep on reacting to excessive nominal wage increases only with rising output prices. There will also be a reduction in employment to the equilibrium level (1-U*) in order to avoid the wage-price-spiral.

The equilibrium combination of unemployment U* and real wages (w/p)* is thus going to prevail in the long-run. In the short-run, unemployment can differ from its natural level. But changes in inflation and subsequent changes in employment

will gradually restore the general equilibrium, which is characterised by a consistency of claims for real output shares of wage and price setters.

B3.5.) Some further issues of the ELMM

Is the ELMM really a Keynesian model? Since we have placed the ELMM in the tradition of Neo-Keynesianism in the brief history of economic thought that was presented in section B2, one might somehow be biased to say that the ELMM is Keynesian. This is correct insofar as it is by nature a non-market clearing approach with involuntary unemployment.[20] However, the model also has a NAIRU and a long-run vertical Phillips-curve as an integral part, which sounds pretty familiar from Friedman's (1968) classical model. It would thus probably be most appropriate to view the ELMM as a model with both Keynesian and classical features. For the remainder of this chapter, we want to illustrate this claim a little bit further by comparing the ELMM directly with the Friedman-model and other essential approaches introduced in this chapter.

Even though both the ELMM and the Friedman-model imply the existence of a "natural" rate of unemployment at which inflation is remaining constant, there are still some notable differences. Most notably, these are the underlying microfoundations and the general concept about how the labour market works.

This can be seen best by an example that highlights the differences: Both the Friedman-model and the ELMM share the policy implication that a deprivation of union bargaining power would lead to a lower equilibrium rate of unemployment. In the Friedman-model this would directly reduce the NAIRU, which is reflecting structural characteristics of the labour market. In the ELMM, lower union bargaining power ceteris paribus shifts down the WS curve, so that the intersection with the PS curve occurs at a lower level of U. However, the ELMM would still not imply that the deprivation of union power would transform the labour market into a perfectly competitive ideal. As argued in section B3.1., there are additional arguments that are inherent to the labour market itself why wage underbidding and perfect competition among workers will not occur: efficiency wage considerations as well as inevitable turn-over costs on which insiders can rely also without unions. The ELMM derives from a fundamentally different view with respect to the working of labour markets. Friedman's reference model is a Walrasian system with perfect product and labour markets. The equilibrium rate of unemployment is thus seen as a measure of market malfunctioning, or as the degree of deviation

[20] This does by the way not imply that all observed unemployment in industrialized economies is entirely involuntary. Lindbeck (1992) points out, that the ELMM might just describe the rationing process of in a primary labour market. Individuals who are kept away from 'good jobs' still have to decide whether they accept jobs in an unregulated and usually flexible secondary labour market that can more realistically be described in the usual fashion as a market with perfect competition. It is easy to see how institutions and welfare state arrangements play a role for this individual decision problem.

from the desirable ideal of perfect competition. The ELMM on the other hand acknowledges that the labour market is systematically, by its very nature, characterised by imperfect competition and has no endogenous tendency to clear.
Differences between the ELMM and the Friedman-model also concern the endogenous inflation-unemployment dynamics. Recall that the initial inflationary impulse in Friedman (1968) comes from a monetary injection that stimulates aggregate demand. Short run real effects arise because workers on instance misperceive higher nominal wages with higher real wages. The lower unemployment can only be sustained if authorities keep on injecting money into the economy and systematically keep on confusing the individuals' perceptions.
The mechanisms of the short-run trade-off between unemployment and inflation in the ELMM are fundamentally different. If unemployment is below its natural level (e.g. because of a demand stimulus), inflationary pressure arises because the bargaining positions of wage and price setters have shifted. Through inflation and employment changes, the real claims on output shares are brought back in line again. Monetary accommodation is not at all needed for this inflation mechanism, nor is the explicit introduction of money. Inflation in the ELMM is not a monetary phenomenon, but is rather stemming from the "battle of the mark-ups" (Layard/Nickell, 1986) between wage and price setters.
The similarities between Friedman and the ELMM are more or less exhausted with the common result that unemployment can be kept below the equilibrium rate only at the cost of accelerating inflation, and with the use of adaptive expectations. Both approaches do not work with the extreme construction of rational expectations as ´new classical macroeconomics´. If this were the case in the ELMM, the inflation-unemployment dynamics we have just spelled out would not develop. The economy would rather always be at the equilibrium constellation U* and (w/p)*. Note, however, that the use of rational expectations does not imply that the emergence of equilibrium unemployment vanishes. The underlying causes, the market imperfections in labour and goods markets, are more general and do not hinge on the type of expectation formation.
It should be mentioned that the ELMM can be used as a framework to analyse the impact of various exogenous shocks or policy interventions on real wages and unemployment. Among the issues that have been analysed in the vein of the ELMM are structural change, a change in labour productivity, employment subsidies, taxation, profit-sharing schemes, monetary and fiscal policy etc.[21]
We have argued in this chapter that the main motivation for developing the ELMM has been the insight that labour markets in the real world can not, or at least should not be modelled as a perfectly competitive market. There is "some-

[21] For an extensive coverage of policy implications of the ELMM, see: Layard/Nickell/Jackman (1991); Blanchard/Katz (1997); Bean (1994); and various others.

thing special" about the labour market (Solow, 1990). Approaches with imperfect competition seemingly are more appropriate for describing labour markets from a theoretical point of view. Stated in more technical terms, the major insight of the ELMM has been the replacement of a standard competitive labour supply function with an upward sloping WS curve, which has required micro-foundations radically different from the usual neoclassical ones. The economic arguments that are at the root of the WS curve also apply to the wage curve on the basis of regions, to which we turn now.

Chapter C) The wage curve

C1) Introduction

In the last chapter we have given a brief overview about major developments in macroeconomics and introduced in some detail the "European Labour Market Model (ELMM)". This macro approach is distinguished from others by its explicit use of imperfect competition on the labour market (and usually also on the product market). In many respects, the wage curve literature (Blanchflower/Oswald[1], 1990, 1996) is directly build upon this macroeconomic approach. The wage curve can actually be seen as a regional pendant of the WS curve that was derived in the ELMM.

Why are regions considered instead of nations? This is by and large due to the fact that the research focus of B/O is mainly an empirical one. It has been argued in chapter A that the use of regions is a good strategy, because one can "gain variance" for empirical research. B/O have gathered and compiled a great deal of empirical evidence about regional wages and unemployment rates from various, European and non-European, countries. The authors themselves describe as their main achievement that they have distilled from this data an "empirical law of economics" (B/O, 1996:1). This law, which seems to hold for virtually all countries and time periods under consideration, can be formulated as follows:

> *"Doubling the regional unemployment rate will drive down the regional wage level by roughly 10 per cent."*

Within any country for which a wage curve exists, there are regions that have both low unemployment and high wages as well as regions that have high unemployment and low wages. This empirical evidence is hard to bring in accordance with standard neoclassical theory of competitive labour markets. This theory would rather predict that unemployment is high where wages are high and vice versa. However, the observed empirical evidence is consistent with the theoretical implications of the ELMM, which B/O have applied and partly reformulated to serve for a regional context.

In this chapter, we want to briefly introduce the empirical research strategy of B/O and contrast their empirical findings with neoclassical regional models in section C2. Afterwards, we discuss in greater detail the theoretical part of the wage curve literature. We introduce one original approach of B/O in section C3, and one more recent approach by Blien (2001) in section C4. This theoretical work is useful for our purposes, since the wage curve literature is concerned with the regional dimension of unemployment in relation to the spatial structure of other important

[1] Hereafter labelled simply as B/O

economic variables. However, we will argue that the current state of the wage curve theory alone is not sufficient to understand the geography of unemployment in the EU, most importantly because it fails to account for the highly relevant feature of spatial agglomeration of economic activity. The problems and shortcomings of the current wage curve literature will be comprehensively summarized in section C5. From this critique we derive the motivation to extent and modify this string of theory.

C2) The wage curve as an empirical regularity

B/O have worked with large scale microeconomic datasets (e.g. the "International Social Survey Programme"). They used the available individual earnings data and ran in principle standard wage equations á la Mincer (1974), only with the regional unemployment rate as an additional explanatory variable.[2] The basic wage curve equation has the following form

$$\ln (wage_{irt}) = \beta_0 + \beta_1 \ln (unemployment_{rt}) + \beta_2 x_{irt} + \text{other terms} \qquad (C.1)$$

The indices i,r,t represent single individuals, regional units and the time period respectively. It is well understood that the earnings level of an individual i ($wage_{irt}$) will depend on personal characteristics (x_{irt}), like his or her level of education, the work experience, the gender etc. Other factors that influence individual earnings can stem from time factors, e.g. the state in the business cycle, or other forms of fixed effects.

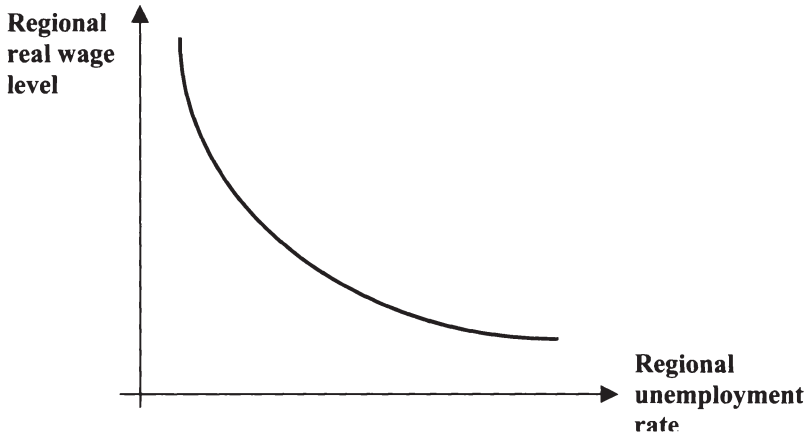
The main finding of B/O is that, when controlling for all these characteristics, the estimator β_1 is significantly negative for virtually all OECD countries and time periods under consideration. Even more surprising, the magnitude of the effect seems to be roughly identical for all countries, -0,10. The estimator β_1 captures the partial effect of the unemployment rate of region of residence on the individual earnings level. The best way to think about the observed significantly negative effect is the following: take two observationally equivalent individuals with the same formal skill and experience level who live in different regions of a given country. The individual living in a low-unemployment region will earn a higher wage than the individual in the high-unemployment area. The magnitude of the partial wage effect of unemployment is given by the coefficient β_1.

In an aggregate sense, the wage curve observation implies that regional wage levels and regional unemployment rates within any given country are robustly negatively correlated. At any point in time, there exist regions with both high wages and low unemployment rates, and regions with low wages and high unemployment rates. Frequently this relationship is graphically represented. Qualitatively

[2] Econometric issues, how to estimate wage curves etc. are intensively discussed in Blien (2001:129 ff.)

the wage curve is a non-linear downward sloping curve in the real wage/unemployment rate-space as presented in fig. C1.

Figure C1: The wage curve



This implication stands in sharp contrast to those models that were dominating research about the relation of wages and unemployment across space all over the 1970s and 1980s. The literature that descended from the work of Harris/Todaro (1970) and Hall (1970, 1972) implies that regional wage levels and regional unemployment rates are positively correlated. The basic intuition for this result is quite simple and clearly stems from neoclassical considerations: individuals living in regions with high unemployment rates must be compensated with a relatively high regional wage level.

This point has been made first by Harris/Todaro (1970) in the field of development economics. The authors tried to understand the process of ongoing rural-to-urban migration in Africa despite the high rates of urban unemployment. They construct a two-sector model with a rural agricultural and an urban industrial sector. The wage in the rural labour market is flexible, whereas the urban wage is exogenously given by some parameter w. The urban wage by assumption exceeds the rural level. Industrial firms are price-takers under a neoclassical technology and demand workers up to the point where the value marginal product of labour is equal to the exogenous wage. Jobs are contracted for a fixed time period, after which workers again enter the pool of unemployed and look out for the next industrial job. Rural migrants also have to enter the pool of urban unemployed and can only apply for industrial jobs after they have migrated to the urban area. Any unemployed person has a certain chance of entering a job at the exogenous wage w. This probability can be approximated by the employment rate (1-U), and is

decreasing in w. With probability U, the worker remains unemployed and has zero income.

Under this quite restrictive model construction, the endogenous variable U, the unemployment rate, is positively correlated with the parameter w. Rural workers keep on migrating until the expected urban income (1-U)w is equal to the agricultural income. Put differently, agricultural workers are constantly employed at a low wage, whereas urban industrial workers are only employed for some time spells, but then at a higher urban wage. The high urban wage level compensates for the urban unemployment rate.

In the context of modern industrialized economies, the regional "compensating differentials" equilibrium has been introduced by Hall (1970, 1972). He considers a multi-location economy subject to unemployment, populated with mobile workers who choose in which location to settle. In equilibrium, all migration incentives must have vanished and therefore all locations must offer the same expected income to workers, which similarly is given by the regional wage times the regional employment rate. An area with high unemployment must therefore offer higher regional wages. Other papers that subscribe to the idea of a compensating equilibrium between unemployment rates and wages are e.g. Reza (1978), Adams (1985), Marston (1985) and Roback (1987).[3]

This string of theory, which implies a positive correlation between regional wage levels and unemployment rates, is inconsistent with the empirical findings of B/O. The authors go as far as to point out that "*this hypothesis [of a positive correlation] is as decisively rejected by the international microeconomic data as it is possible to imagine*" (B/O, 1994:9). Matters, however, are not quite as clear as B/O have put it. There is still an ongoing debate, ambiguous empirical findings, or even some evidence in favour of the Harris-Todaro-Hall hypothesis (Patridge/Rickman, 1997). These debates notwithstanding, it seems safe to conclude that today the majority of studies concludes that a wage curve in fact exists in most OECD countries.[4] Unfortunately, there is no wage curve estimate for the European Union as a whole. This is due to the fact that the estimation would require consistent and huge scale microeconomic datasets on the basis of European regions (like the Microcensus or the Socio-economic panel for the case of Germany). To our knowledge such data does not yet exist, which rules out a rigorous estimation. The descriptive evidence presented in chapter A, however, seems to suggest that a wage curve will also exist in the EU. High unemployment and high per capita income (which should be highly positively correlated with wage levels) coincide geographically to an extend that supposedly does not simply reflect the degree of heterogeneity in individuals' underlying characteristics.

---

[3] An overview is provided by Elhorst (2000). The main ideas of the most famous models of this class are also comprehensively introduced in B/O (1996:15 ff.).

[4] See e.g. the overviews of Blien (2001: ch.8), Buettner (1999: ch. 5 +6) or Card (1995).

C3) Wage curve theory: The Blanchflower/Oswald-model

If one accepts the wage curve empirically, one has to think about a consistent theoretical model. Competitive models of the labour market seem inapplicable to rationalize a wage curve, which is interpreted by B/O as a long-run equilibrium curve in regional labour markets. However, theoretical rationale for a downward-sloping labour market equilibrium curve in the real wage/unemployment rate-space is readily available from the macroeconomic "imperfect competition" approach, the ELMM.

We now turn to the theory of the wage curve as introduced by B/O (1996). It will be shown at first that the wage curve is actually equivalent to the WS curve from the ELMM and describes a partial labour market equilibrium relation. We will derive one particular wage curve based on efficiency wages in section C3.1.). The next step is to integrate this partial equilibrium relation with a description of the regional product markets. Essentially, this is the same as deriving a PS curve in the ELMM, and will be done in section C3.2.). However, since the wage curve is concerned with the regional dimension of an economy, some further issues do now arise. Namely, on a regional level one must take factor mobility into account. In section C3.2.) we will therefore introduce the full interregional general equilibrium model of the two-region economy. The section C3 ends with a critical discussion of the wage curve model of B/O in C3.3.).

C3.1) The partial equilibrium foundations of the wage curve

The wage curve in figure C1 looks similar to the WS curve that was introduced in the last chapter (figure B1). With a fixed and perfectly inelastic labour force (which we will continue to assume), the two concepts of a WS curve and a wage curve are analytically equivalent. Saying that the equilibrium wage in the labour market positively depends on the employment rate is the same as saying that it negatively depends on the regional unemployment rate.

In the theoretical part of their work, B/O present three plausible stories why in partial equilibrium a higher regional unemployment rate might depress the regional wage level. The first approach is based on the idea of implicit contracts. This approach will not be introduced here, because it is the most complicated but the least convincing of the three (see Card, 1995:796; Blien, 2001: 84). The other two are already familiar from the last chapter. The foundation of the wage curve can either be a collective bargaining model, or an efficiency wage model. The underlying logic for either of the two is identical with the underlying intuitions of the WS curve that was provided in section B3.2., only that one has to think now on a regional level. If regional unions bargain about regional wages, their power relative to employers will likewise depend positively on the regional employment rate, i.e. negatively on the regional unemployment rate.

a) What is the appropriate micro-foundation of a wage curve?

To be more explicit, we derive a wage curve based on one particular efficiency model, namely the shirking-approach of Shapiro/Stiglitz (1984) in the slightly modified version of B/O (1996:64 ff.). The reason why we use this particular model is mostly its analytical simplicity. However, there is also another reason for considering an efficiency wage-model instead of a union model, given the institutional background e.g. of the German labour market.

It is often spelled out that union models have been used in the ELMM to reflect the institutional situation in continental Europe, whereas the efficiency wage models apply more to the "flexible" labour markets in the UK and the USA. Since we are mostly interested in the European Union, one might thus expect that a collective bargaining approach should be chosen when becoming explicit about the microeconomic foundations of the wage curve. However, recall that from now we are concerned with the regional dimension of an economy. It is true that e.g. the German labour market is highly unionised (e.g. Ochel, 2000). But at the same time the German labour market is also characterised by a very low degree of regional differentiation of union wages. Collective bargaining in Germany takes place at the sectoral level. Formally, the regional sub organizations of the nationwide unions and employers associations bargain on the level of the German Bundesländer in most sectors. De facto, however, this hardly means anything. Typically there is one pilot agreement in one region, which subsequently is applied without any notable modification all over the nation (see e.g. Buettner, 1999:99 ff.; Bispink, 1995; iwd 40/2002). Regional wage differentiation in Germany does not occur through regional differences in union wages, but mostly through differences in *effective earnings* (e.g. Schnabel, 1995).

In other words, an approach that bases a wage curve on regional differences in bargaining strength of inherently regional unions is not appropriate for the institutional structure of the German labour market. This does of course not imply that union wage setting is irrelevant for explaining unemployment in Europe. One can argue that it is precisely the low degree of regional differentiation of union wages that causes intra-national unemployment disparities. We will come back to this argument in chapter F. For the wage curve, however, efficiency wages seem to be the more appropriate micro-foundation. Regional wage differentiation occurs, because firms from different regions pay different effective wages above the (uniform) contracted level. The reasons for this behaviour presumably may be found in efficiency wage considerations (Blien, 2001:86).

b) A wage curve based on efficiency wages

In their monograph,. B/O use a modified version of the shirking approach of Shapiro/Stiglitz (1984). We will use an even more simplified version of the Shapiro/Stiglitz-model in this paper. We consider an economy in continuous time consisting of two regions r={1,2}, and we assume risk-neutral workers, who gain

utility from wage income w_r, but disutility from work-effort e_r. Utility V_r is assumed to be linear.

$$V_r = w_r - e_r. \tag{C.2}$$

Effort at work is assumed to be a technologically fixed number $e_r > 0$. Individuals can choose to "shirk" at work and spend zero effort $e_r=0$. Shirking individuals run the risk of being detected and then fired. The exogenous detection and firing probability $(1-\gamma_r) < 1$ is less than perfect. Once fired, an individual enters the pool of the unemployed. Yet, following Shapiro/Stiglitz (1984), there is also some exogenous destruction rate of firms $R_r > 0$ that likewise leads to an inflow from employment to unemployment. For simplicity, we assume that unemployed persons have no source of income.[5]

The unemployed have a chance α_r of re-entering into a job. This endogenous variable depicts the flow from unemployment back into the pool of the employed. In the steady state equilibrium, the two labour market flows must be equal. Given that nobody will shirk in equilibrium, we can write this condition as $R_r N_r = \alpha_r (L_r-E_r)$, where L_r is the labour force and E_r is employment. The definition of the unemployment rate is $U_r = 1- E_r/L_r$. This determines the function α_r to be $\alpha_r = (R_r / U_r) - R_r$. Thus, the outflow probability from unemployment is decreasing in the regional unemployment rate U_r. With these assumptions, the (expected) utility of an unemployed individual (V_{ur}) is given by

$$V_{ur}= \alpha_r (w_r - e_r). \tag{C.3}$$

Non-shirking employed workers and shirkers have utility levels V_{enr} and V_{esr} respectively

$$V_{enr} = w_r - e_r \tag{C.4}$$
$$V_{esr} = \gamma_r w_r + (1-\gamma_r)(\alpha_r(w_r - e_r)). \tag{C.5}$$

The firm has an interest to prevent shirking and will thus pay efficiency wages that are just sufficient to ensure equal utility for shirkers and non-shirkers, i.e. $V_{esr}=V_{enr}$. Equating (C.4) and (C.5) yields after some manipulations the following expression

[5] In most parts of the efficiency wage framework of B/O, they assume that regions might differ with respect to the level of unemployment benefits. We do not consider these cases, because it is irrelevant for most continental European countries. Unemployment benefits are generally not differentiated across regions. We therefore have assumed that unemployment benefits b_r are equalized on the level $b_r=0$. This normalization, however, is only for analytical simplifications.

$$w_r = e_r + \frac{\gamma_r e_r}{(1-\gamma_r)(1-\alpha_r(U_r))} \qquad \text{(C.6)}$$

Equation (C.6) is the regional wage curve and can be interpreted as the aggregate non-shirking condition in region r. It shows the efficiency wage that is sufficient to prevent shirking for any given unemployment rate, and given the structural parameters e_r and γ_r. The graphical representation of (C.6) qualitatively looks like in figure C1: the required efficiency wage is lower, the higher is the regional unemployment rate U_r. The intuition for this result is clear. At any given shirking detection probability, individuals become more reluctant to shirk when the unemployment rate is high. Becoming unemployed is perceived to be a strong penalty. Consequently, firms do not have to rely on a wage premium to prevent their incumbent workforces from shirking. As the regional unemployment rate decreases, becoming unemployed is less of a threat for single workers, and shirking becomes a more viable option. To prevent shirking, firms pay efficiency wages $w_r > e_r$, since workers then put stronger value on their specific, well-paid jobs and abstain from shirking. The required efficiency wage is higher at any level of U_r the higher is the disutility of effort e_r and the lower is the shirking detection rate $(1-\gamma_r)$.[6]

What B/O essentially do is to bring this aggregate non-shirking condition in a two-region context. They assume that both regions are structurally identical, meaning that R_r, e_r and γ_r are the same in both regions.[7] This implies that both regions face the same wage curve, i.e. the same labour market equilibrium curve given by equation (C.6).

B/O then analyse what happens if one region is intrinsically more attractive than the other, e.g. because of climatic and cultural circumstances. This region offers an utility supplement ξ to each individual who lives and works there. They show that regardless of ξ, both regions will still face exactly the same equilibrium locus with respect to unemployment and wages, the identical wage curve fig. C1. This

[6] With perfect information on workers effort ($\gamma_r=0$), the firm would only need to pay $w_r=e_r$ for any unemployment rate to prevent shirking. Note further that with imperfect monitoring ($\gamma_r>0$), full employment is not possible in equilibrium, since $\alpha_r<1$ requires that $U_r>R_r/(1+R_r)$. Put differently, the required efficiency wage would have to become infinite.

[7] One might potentially introduce differences in the structural parameters e_r and γ_r across regions. These differences might e.g. be thought of as differences in labour market institutions, although the disutility level of effort or shirking detection rates are typically not the kind of institutions that seem directly relevant labour market comparisons. Model extensions are conceivable where γ_r is influenced by regional employment protection laws, or e_r is some sort of reservation wage dependent on regional welfare state arrangements. But the same argument as for unemployment benefits applies: usually the degree of institutional variation across regions within the same country is very little in continental Europe.

can be seen by considering the equilibrium condition (value of shirking equal to value of non-shirking) for the intrinsically attractive region.

$$w_r - e + \xi = \gamma (w_r + \xi) + (1-\gamma_r) \{ \alpha_r (w_r - e) + \xi \}$$

In the process of substitution, the term ξ will cancel out, and the attractive region will face the same labour market equilibrium locus (C.6) as the unattractive region.

But, as will be become more clear in the next section, B/O also assume mobile workers. If the "economic variables" w_r and U_r were the same in both regions, workers would want to move to the intrinsically more attractive area, since here they are rewarded with an utility bonus ξ. B/O (1996:69) show that in an inter-regional equilibrium, which is characterized by a situation where there is no incentive for further migration, the attractive region will exhibit a higher expected unemployment rate and a lower expected wage. In other words, the intrinsically unattractive region has to compensate for its missing amenities by offering better "economic" values. The unattractive region will thus be located on the upper left part of the wage curve in fig. C1. The attractive region on the other hand will find itself on the lower right tail of the wage curve. In observable regional data, a negative correlation between regional unemployment rates and wages is visible, since both areas are located on the same wage curve.

C3.2) General equilibrium in the Blanchflower/Oswald-model

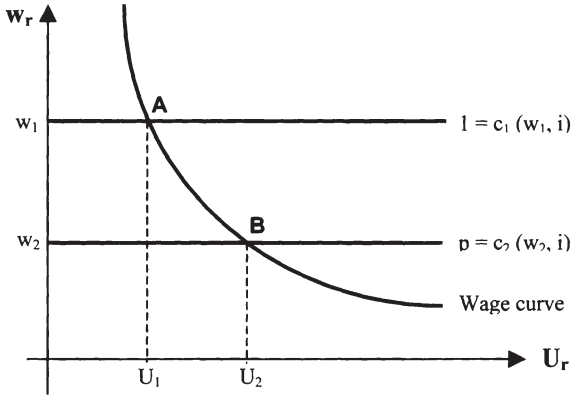
This notion already gives an idea of how B/O will establish the stability of the wage curve as a long-run equilibrium curve. But these implications will only become fully visible when moving to a general equilibrium characterization of this economy. The wage curve in figure C1 only represents "one half" of equilibrium, analogous to the WS curve in the ELMM. It needs to be determined exactly where on the wage curve the single regions r=1 and r=2 will end up. This will be a matter of goods market equilibrium. So, the full equilibrium in the B/O-model is also determined by joint equilibrium on labour and goods market, in full analogy to the ELMM.

B/O assume that each of the two regions produces a distinct tradable commodity under constant returns to scale and perfect competition. The product market equilibrium they derive is therefore qualitatively identical to the case from section B3.3a). They assume that the production function for the regional tradable good Y_r is given by $Y_r = f(N_r, K_r)$. Capital (K_r) is assumed to be an essential input of production, for which the price i is determined on world markets. Labour and capital in both regions have to be used in fixed proportions. Firms in both regions will thus face the following minimum cost function C_r

$$C_r(Y_r, w_r, i) = \min_{N_r, K_r} \{ w_r N_r / Y_r + i K_r / Y_r \} = Y_r c_r(w_r, i) \qquad \text{(C.7)}$$

Under this limitational production function with constant returns to scale, total minimum costs are simply the product of minimum unit cost (c_r) times the quantity of output Y_r. Perfect competition implies that minimum unit costs $c_r(w_r,i)$ need to equal the product price p_r in order for profits to be zero. The goods market equilibrium is given by the condition $p_r = c_r(w_r , i)$.
The product prices p_1 and p_2 for the two regional commodities Y_1 and Y_2 again ground out from a Walrasian tâtonnement process, and are thus given to any single firm. Without loss of generality, B/O normalize the given product price for the good that is exclusively produced in region 1 to unity. The price of the product from region 2 is denoted p. Equilibrium in goods markets then requires that $1 = c_1(w_1,i)$ and $p=c_2(w_2,i)$.
General equilibrium in either region is reached when product and labour market are jointly in equilibrium. Since both regions face the same wage curve locus, the graphical representation of the general equilibrium in both regions can be illustrated in only one diagram, fig. C2.

Figure C2: Full equilibrium in the Blanchflower/Oswald-model



The horizontal curves represent the product market equilibrium conditions for region 1 and 2 at the given output prices 1 and p. Full equilibrium in either region is obtained at the intersection points with the wage curve, i.e. at points A and B. At point A, firms make zero super-normal profits and shirking is deterred in region 1. The same is true for region 2 at point B.
If the parameter constellation is such that $p_2<p_1$ ($p<1$), nominal wages are higher and unemployment is lower in region 1. Note that with freely tradable goods, workers from both regions face the same consumer price index, and nominal wage differences are thus equal to real wage differences. Hence, for this constellation of exogenous product prices in fig. C2, region 1 is advantaged over region 2 along

two dimensions: the real wage is higher in region 1, and the unemployment rate is lower.

But for a full interregional equilibrium, also all migration incentives must have vanished. In the situation depicted in the diagram, this is not yet the case. Individuals from region 2 have an incentive to move to region 1.

B/O (1996:81) are not very explicit about the technological effects on labour productivity and wages if migration occurs. This will of course crucially depend on the properties of the underlying production function. If firms can adjust the essential capital input proportionally with the additional stock of workers, the MPL, and ultimately the zero profit curves in fig. C1 would remain unchanged. The incentive for migration would remain constant, as the regional wage gap is independent of the number of migrants and only depends on the exogenous product prices. Matters are different if the capital stock can not be adjusted. Under this circumstance, every additional worker has a marginal productivity of zero, since the technology is limitational. However, the total amount of labour (measured e.g. in working hours) that is technologically efficient for the given capital stock can simply be shared among a higher number of workers. The total wage income in both regions would thus remain constant, but the wage *per worker* in region 1 and 2 would converge through labour migration.

Migration would thus lead to convergence of per capita remunerations and ultimately to an erosion of the wage curve relation.[8] However, B/O partly avoid these discussions about technological effects of migration by going back to their construction with intrinsic regional characteristics that was mentioned earlier. Recall that they have developed the implications of differences in the intrinsic attractiveness of regions. In partial equilibrium, these intrinsic differences were exogenous. Now they endogenise the utility supplement ξ and make it negatively proportional to the population density of a region. In other words, as workers move into region 1 because of the better economic situation, the place becomes gradually crowded and thereby unattractive.

With this construction of congestion, it is possible to construct general equilibria with the zero-migration condition satisfied. These will look like figure C2. Regions one the "bad side" of the wage curve will compensate individuals with inherent regional amenities, in this case by the fact that there is little congestion. In terms of the observable economic variables w and U, however, a wage curve is visible in the data. The regional values of w and U do not collapse into one point, because the compensating amenities make up for the regional differences in wages and unemployment. The wage curve is thus seen as a long-run equilibrium phenomenon, because single regions are placed at different points along the labour market equilibrium schedule. Regional wages and regional unemployment are

[8] A more detailed discussion about the effects of migration in a model with a neoclassical production function is provided in chapter F.

negatively correlated in equilibrium, and these regional disparities persist and show no tendency to vanish.

C 3.3) Critique of the Blanchflower/Oswald-model

Even though B/O intended to depart from the work of Harris/Todaro (1970), both models share some important common characteristics. Both subscribe to the idea of an equilibrium with compensating differentials, which is reached when incentives for migration have vanished. In the Harris/Todaro-world, unemployment rates and wages together form an equilibrium of compensating differentials and therefore are positively correlated. In the world with a wage curve, intrinsic regional amenities make up for the combination of unemployment and wages, which now are negatively correlated.

The first critique against the Blanchflower/Oswald-model is that the substantial origin of regional differences remains an open issue. Regions are assumed to produce different final goods and sell them at different exogenous product prices under perfect competition. Thus, two principally identical regions still manufacture different commodities, which consequently leads to disparate regional development. The whole analysis is incapable of pointing to the very reason for regional differences in depth. Why can regions with low selling goods not switch over and manufacture better commodities? What is the reason that one local entity produces a "better" good than the other, which ultimately is rewarded with a lower unemployment rate and a higher wage? These questions remain unanswered.

The second problematic aspect with exogenously given product prices is the apparent identification of regions with sectors, or at least with specific products. Because this identification is much more explicitly developed in the Blien-model that will be discussed in the next section, this critique will be postponed to a later point.

The third critique concerns the analysis of labour mobility. As noted before, all individuals have a principle interest to move to the region with the "better" final product, which has both lower unemployment and higher real wages. If nothing else is added, the wage curve relation in the B/O-model would erode, and it would not be a stable long-run equilibrium relation as the authors imply. But B/O assume, in an "ad-hoc" way, that regional preferences are operating as an opposing factor. The critical point with this ad-hoc-construction is that the long-run stability of the wage curve (the main contention of B/O) crucially hinges on it. From a theoretical point of view, this does not seem fully convincing.

Blien (2001:96ff.) points to another critical aspect: within the general equilibrium model, the original causality of the wage curve running from unemployment to wages suddenly has changed without explicit notice. The wage curve theory meant to provide rationale for negative wage effects of unemployment. But in the model of B/O, now exogenous product prices determine a wage rate via the zero profit condition for firms. Only in a secondary step is unemployment determined,

but in an unusual way such that high wages are now associated with low unemployment. We subscribe to the critique of Blien, even though one can think of regional wages and unemployment rates as being simultaneously determined. It is still one more representation of the fact that essentially everything is driven by the exogenously given product prices in the model of B/O.

C 4) The model of Blien (2001)

To cope with several critical aspects of the B/O-model, Blien (2001) presents an own approach to wage curve theory. This concerns both the partial equilibrium foundations of the wage curve, and the integration of the wage curve into a full model with a product market.

C4.1.) Partial labour market equilibrium in the Blien-model

Blien's motivation to base the wage curve on different micro-foundations are unrealistic features of the Shapiro/Stiglitz-world. Individuals in reality do not decide just whether to shirk completely or to provide full work effort. Moreover, strict legislation on dismissal policies often prevent firms from firing shirking individuals in Europe, specifically because the definition of shirking is difficult to formulate in reality. Therefore Blien's main idea is that firms do not try to solve an "information problem" stemming from imperfect monitoring possibilities. Instead, they solve an "enforcement problem" and try to use efficiency wages to motivate employees to spend more work effort. His partial equilibrium model of the wage curve is inclined by the labour turnover approach of efficiency wage theory, and specifically builds on work by Schlicht (1978). The core idea of this approach is that individuals do not simply decide whether to supply full work effort or to shirk completely, but that the chosen effort level is a continuous function. Consider some individual worker, who has a job at some firm j. The worker's effort is given by the following function A

$$A = A\left(\frac{w^j}{\bar{w}}, U_r\right) \tag{C.8}$$

with $\partial A / \partial w^j > 0$, $\partial A/\partial U_r > 0$. The worker will spend more effort the higher is the wage of firm j relative to some exogenous market wage $\bar{w}$, and the higher is the unemployment rate is the respective region of residence. The intuition for the derivatives is straightforward: The penalty of loosing the particular job is greater for the worker the better she is paid at the particular firm j, and the worse are the outside prospects approximated by the regional unemployment rate. Work effort is an insurance for the worker against an individual lay-off. Therefore, the worker will spend more effort the higher is her interest in remaining employed at this particular firm.

The firm's problem is to maximize profits, taking into account the effects of efficiency wages on workers' performance. When aggregating over the wage setting decisions of identical firms in region r, Blien (2001) ultimately also arrives at a wage curve relation. We just present the analytical expression of the wage curve that is given by equation (C.9)

$$ln\, w_r = ln\, \overline{w} - \beta ln\, U_r + \beta ln \overline{U}$$

or

$$w_r = \frac{\overline{w}\ \overline{U}^{\beta}}{U_r^{\beta}} \qquad \text{(C.9)}$$

U_r and w_r are regional values of unemployment rate and wages, whereas $\overline{U}$ and $\overline{w}$ are exogenously given values of national averages. Equation (C.9) is also a negatively sloped (non-linear) wage curve that can be graphed like in fig.C1. It analogously provides the labour market equilibrium relation for every region.

C4.2.) The product market and general equilibrium in the Blien-model

To move towards a full interregional general equilibrium model, Blien needs to specify the product market conditions. He assumes that each region is specialised in a single, distinct product or industry, just like B/O. But he adds a dynamic component to the product markets, by applying the idea of product cycles. Let demand and supply for the good of region r be given by the following two simple equations

$$P_r = a - b\, Y_r \qquad \text{(demand curve)} \qquad \text{(C.10)}$$

$$P_r = M\, w_r / D_r \qquad \text{(supply curve)} \qquad \text{(C.11)}$$

a, b are exogenous parameters for the demand side. M is also exogenous and captures a mark-up of prices over wages to pay for the rental rate of capital. Y_r is regional income, D_r is labour productivity, defined as $D_r = Y_r/N_r$, where N_r is employment. Upon substitution, one can obtain the following expression.

$$N_r = \frac{a}{bD_r} - \frac{M}{bD_r^2} w_r \qquad \text{(C.12)}$$

Under the use of the definition of the unemployment rate $U_r = (L_r - N_r) / L_r$ this expression can be rewritten as

$$U_r = 1 + \frac{Mw_r}{bL_r D_r^2} - \frac{a}{bL_r D_r} \qquad \text{(C.13)}$$

If we think of the wage w_r as being exogenously given, equation (C.13) could be used to analyse the impacts of productivity improvements on unemployment.[9] This type of analysis has been introduced by Appelbaum/Schettkat (1995). Their main result can be summarized as follows: A productivity improvement leads to an increase in employment, if the elasticity of demand on the product market is greater than one. Similarly, unemployment will increase in response to productivity improvements if product demand reacts inelastic.[10]

The idea of product cycles enters in the following way. New products tend to face largely unsatisfied demand, and their price elasticity therefore is high. But specific products age over time, demand becomes largely satisfied. Price elasticities decline, and productivity improvements and lower prices do not translate any longer into an increase in total production.

Think of the regional consequences of product cycles, if the single regions are completely specialized. The regional development is then driven by the dynamics of the market for the region-specific product. Blien (2001) assumes that improvements in labour productivity are exogenously given and identical for all regions. This has different effects on the single regions, depending on the state of the specific products within the cycle. Those who specialize in old products will be harmed, as higher productivity effectively leads to more unemployment. The opposite is true for regions with young products at the beginning of the cycle. There is a high elasticity of demand for the specific commodity, and productivity improvements (=falling prices) translate into higher employment.

Equation (C.13) is the "second half" of full equilibrium, since it characterises the product market equilibrium and is a representation of labour demand. Graphically this is an upward sloping line in the (U,w)-space. For full equilibrium, equations (C.9) and (C.13) need to be integrated. Upon substitution, we obtain the following implicit function Z

$$Z \equiv U_r - \frac{M \;\; \bar{w}\, \bar{U}^{\beta}\, U_r^{-\beta}}{b\, L_r\, D_r^2} + \frac{a}{bL_r D_r} - 1 = 0 \qquad \text{(C.14)}$$

The central insight of the product market dynamics remains unchanged: Whether productivity improvements *at given wages* decrease or increase the regional unemployment rate depends on the elasticity of product demand. But now the wage curve comes into play. Changes in unemployment will have wage effects, which in turn will again influence labour demand. For example, if there are productivity improvements and inelastic product demand, unemployment will increase. Due to

[9] Be aware that since M is a fixed number, the relative factor intensity is assumed to remain constant even with improved labour productivity.

[10] See Blien (2001:120 ff.) for a formal elaboration.

the wage curve, the necessity to pay efficiency wages is relaxed to some extend and equilibrium wages fall. This drop in wages then stimulates labour demand and consequently lowers unemployment, but to a smaller extend than the initial loss.[11]
The final scope now is to integrate these ideas into a full interregional equilibrium. So far, Blien (2001) has shown how product market dynamics drive the equilibrium values of wages and unemployment rates for the single regions. In the long run, however, individuals from regions specialized in the production of commodities at the end of the product cycle do have an incentive to emigrate to booming areas.
Recall that B/O have argued that individuals in lagging locations are compensated by "non-economic" amenities, and that the wage curve is thereby stable over time. Blien pursues a different path here. He acknowledges that workers will move to those areas where wages are high and unemployment is low. But he rightly points out that migration is not an instantaneous reaction to small differences in economic variables. It is a costly and slow process. Blien argues that migration will gradually take place in response to regional inequalities. Because of this, the regional disparities will slowly fade away, other things being equal. The wage curve is thus not a long-run equilibrium curve in his model, but rather one of temporary short-run equilibrium. But he does not present a formal integration of this argument into his model. He just states the principal tendency of the wage curve to erode due to labour mobility. In regional data, however, wage curve relationship is visible, because the equilibrating forces are weak, and frequent impulses from product markets keep the labour market in motion permanently. A wage curve that is detected in the data is a representations of permanent disequilibrium and sluggish adjustment in the labour market.

C 4.3) Critique of the Blien-model

The main innovation of the Blien-model is the integration of a dynamic element, the product cycle, into the model. By this construction, a boom for a region has the same origin as a possible subsequent downturn: the state within the product cycle. This dynamic element has great merits compared to the rather static approach of B/O with exogenously given product prices. By inspection, one can think of various examples where the economic situation of a region was inevitably linked to one very characteristic product: shipbuilding in the German harbour cities in the North, coalmining in the Ruhr area, and so on. But the identification of

[11] One can see this argument more clearly by analysing the impacts of a demand shock on unemployment with and without wage reactions. The derivative $\partial U_r/\partial b$ in equation (C.13) gives the unemployment reaction if wages are fixed, the same derivative for (5) shows the reactions if wages are endogenously determined by the wage curve reaction. One can show that with fixed wages reactions are more drastic. The wage curve hence smoothes out some effects, since wages move in the opposite direction as unemployment.

regions with industries that is inherent to the approaches of both Blien and B/O is problematic for two reasons, one empirical, one theoretical.

On the empirical front one has to take into account that regions in Europe are far from being specialized in one or only a few products. By the same token, specific industries are not very much concentrated in only one region. Take the German example: car production is located not only around Stuttgart and Munich, but also in Wolfsburg, in Eisenach, in Leipzig and so on, let alone other sectors that are even more dispersed. Even if industries are very much concentrated in some regions, these regions in general host more than just one sector. For example, there is more than just car production in Stuttgart. There is more than just financial businesses in Frankfurt. And think of agriculture, the service sector, tourism. The regional concentration is not so high that one can really set regions equal with industries. Never mind, one could argue, regional unemployment rates could then be derived by using sectoral unemployment rates (that in principle could be obtained with Blien's methodology), weighted with the sectoral shares of each region. This approach has been proposed for example by Armstrong/Taylor (1993). However, it seems to be a well established empirical fact that differences in regional unemployment rates can only weakly be attributed to the sectoral specialization patterns of regions. Evidence on this point is provided e.g. by OECD (2000), R.Martin (1997), Taylor/Bradley (1983), Dixon/Thirlwall (1975), or Elhorst (2000) who concludes that

> *"Most empirical applications have indicated that spatial differences in industry mix account for little, if any, of the variation in unemployment rates between regions. The same industry seems to experience different unemployment rates in different regions."*

Hence, one should aim to develop a theory where regional disparities are not simply reflections of the sectoral structure. At least this empirical point should be seen as a caveat not to rely too exclusively on sectoral explanations.

But apart from this empirical critique, also a theoretical point needs to be stressed. Just like in the B/O-model, the sectoral specialization is completely exogenous in the Blien-model. It is unspecified *why* regions specialize in certain products and *why* they do not change their specialization patterns if products grow old and times go bad. Furthermore, why are there exogenous improvements in labour productivity that are independent of the stage in the product cycle? The appealing the idea of product cycles may be, the working of the model is very much driven by factors that are coming from outside the model and are assumed rather than explained.

Finally, the long-run implications remain an open issue. Blien argues that workers are in principal mobile across regions. Thereby the initial regional disparities along the wage curve locus would gradually erode over time. The visibility of a

wage curve in the data is attributed to the frequency of exogenous shocks and the slowness of the migration process. In a world without further shocks, regions would thus converge with respect to wages and unemployment rates. But unlike B/O, Blien does not specify the long-run equilibrium, maybe because he is not working with a production function. And he contradicts one of B/O's main conclusion, namely that the wage curve is more than just a representation of permanent disequilibrium. It has been central to the wage curve literature that regional disparities along the wage curve locus are stable over time.

There is thus a dissent between B/O and Blien on this issue, and ultimately it will be a matter of further empirical research to determine who is right. But given that Blien does not tell *why* he does not believe in the long-run character of the wage curve, apart maybe from the fact that the theoretical explanation of B/O is not fully convincing, we feel more inclined with the original result from B/O. The persistence of regional labour market disparities has been so high in recent decades that sluggish migration in our view fails to be a convincing explanation.

C 5) Conclusion on wage curve theory and motivation for an own approach

In this chapter we have discussed in quite some detail (albeit mostly non-technical) the two main contributions to wage curve theory. The analogies with the macroeconomic ELMM were obvious. The micro-foundations of the wage curve as a partial labour market equilibrium locus were merely identical with those of the WS curve. We have argued above that in principle many suitable stories can be tailored to rationalize the characteristic slope of the wage curve in imperfectly competitive labour markets.

We have pointed to several problematic aspects of the existing two wage curve approaches. These aspects all had to do with the specification of the *product market*, i.e. with the analogue of the PS curve. Our critique shall now be summarized in order to yield a schedule how to proceed in the next chapters of this book.

The principal problem of the two approaches seems to be that they can not explain the endogenous emergence of regional disparities. Regions in both the B/O- and the Blien-model were identified with industries or sectors, and regional unemployment is merely compounded of the sectoral decomposition. More than that, the sectoral specialization in both models and thus the source of regional disparities only exist *by assumption*, as regions are simply assigned to the production of specific goods.

This *complete* exogeneity is not even that critical. One can think of model extensions where regions are characterised by different factor endowments or different availability of natural resources. In this case, the sectoral specialization would shape because of comparative cost advantages in spirit of the neo-classical trade theory. Exogenous endowments would then be at the core of regional disparities, as they would determine the specialization patterns that in a secondary step drive

the regional development paths. However, such an approach would probably still be insufficient.

The are both empirical and theoretical reasons for this proposition. On the empirical front we have indicated that spatial unemployment differences can not be explained well by the industry mix (OECD, 2000; Elhorst, 2000). There is a truly *regional* dimension. Moreover, it is most definitely so that the economic landscape in Europe is not only driven by comparative advantages. We have shown in chapter A that the reality in the EU-15 is characterised by a clear core-periphery structure of economic activity. Production is distributed very unevenly across space. The compelling evidence on this matter is inconsistent with the view that rich core regions have their status only because of underlying regional characteristics. Instead, it has become common to believe that today's spatial configuration is also the result from endogenous cumulative processes and circular causation mechanisms.

Economic theory in recent years has seen a great revival of theories that analyse why economic activity in market economies seemingly tends to be organized unevenly across space, why there is a tendency for agglomeration, and why regional disparities can *endogenously* unravel and persist. As will be described in the next chapter, these theories have departed from traditional assumptions that are typical for neoclassical models, and that were also used by B/O and Blien. This namely concerns the assumptions that production is characterised by a technology exhibiting constant returns to scale and that goods are perfectly tradable across space. In the new trade and location approaches, one rather works with localized increasing returns to scale and transportation costs. Under this set of assumptions it is possible to explain the *endogenous* emergence and persistence of core-periphery-structures.

Our plan is thus to reformulate the general equilibrium approach of the wage curve. We will leave the partial labour market equilibrium curve unchanged. But we will change the respective product market structure and work with a technology that takes the salient feature of regional agglomeration into account. At first, however, we will introduce the main ideas and features of the modern theories of regional agglomeration in the next chapter.

Chapter D)
Regional agglomeration theory and the ´new economic geography´

D1) Introduction

Typical neoclassical models are characterised by a technology exhibiting constant returns to scale, substitutability of input factors and diminishing marginal returns in case of partial factor variations. Markets are characterised by perfect competition and abstract from transaction costs such as spatial transportation costs. The neoclassical paradigm is ill-equipped to study spatial questions like agglomeration of economic activity, which is one of the most salient and obvious features of real world economies (Ottaviano/Thisse, 2001).[1]

This has led researchers to gradually depart from the ´perfect´ neoclassical world by relaxing its restrictive assumptions. One of the most prominent tasks has been to challenge the central neoclassical assumption of constant returns to scale. Recall the meaning of a scale elasticity equal to one: it means that theoretically any production process can be divided into infinitely many parts without any loss of final output. If this were true, if there are no economies of scale at stake, the world should be characterised by some sort of "backyard capitalism". A casual look at the map, however, indicates that something else must be going on in the real world. There must be advantages of agglomeration, because otherwise we simply would not observe this enormous degree of spatial unevenness.[2] By pooling many firms, many workers and many consumers, per capita output and productivity must be higher vice versa a spatial configuration where everything is widely spread out.[3]

Ironically, the ´father´ of neoclassical economics himself, Alfred Marshall (1890), was the first to point to distinct mechanisms for what in modern terminology might be called ´agglomeration advantages´ or ´increasing returns to scale from

---

[1] Formally this argument has been made by Starret (1978), who has developed the "spatial impossibility theorem", according to which there does not exist any competitive equilibrium with trade between distant locations if space is homogenous and markets are perfect. The problems of neoclassical economics to deal with location problems has long been recognized. Koopmans (1954) e.g. points out that "*without recognizing indivisibilities -in human person,in residences, plants,equipment,and in transportation -urban location problems, down to those of the smallest village,cannot be understood.*"

[2] Fujita/Krugman/Venables [F/K/V] (1999) and Ottaviano/Puga (1998) note that a neoclassical model with constant returns to scale would not necessarily imply complete spatial homogeneity. There still could be some unevenness due to differences in natural environments, climate and the like, i.e. because of comparative advantages. But these disparities in underlying spatial characteristics are surely insufficient to explain the actual core-periphery divides in the real world.

[3] Empirical support for the existence of various pooling advantages has been provided e.g. by Ciccone (2002) or Ciccone/Hall (1996).

spatial concentration´. These ideas, and in particular one of them, is at the root of the modern approaches in trade and location theories that are inevitably tied with the name of Paul Krugman.

The particularly hard task of introducing increasing returns to scale to a theoretical model is that it usually requires to depart also from another neoclassical paradigm, namely that of perfect competition. A convincing general equilibrium framework with increasing returns and imperfect competition was not available to the economics profession until the contribution of Dixit/Stiglitz (1977), who have build on the monopolistic competition approach of Chamberlin (1950) and formalised it rigorously. The first impact of this seminal model on trade and location theory was the development of the "new trade theory" by Krugman (1980).[4] This approach is in many respects the direct forerunner of the core model of "new economic geography" (NEG) developed by Krugman (1991a, b) which then explicitly addressed the issue of endogenous regional agglomeration.

In order to highlight some of the theoretical issues at stake, and in order to provide some better preparation for the arguments in the remainder of this chapter, we start with a clarification of some terminology concerning the phrase ´increasing returns to scale´ and derive some implications for competition and market structures in section D2. Afterwards we will briefly discuss the intuition of the Marshallian agglomeration economies, as well as introduce other arguments for spatial agglomeration coming from the field of endogenous growth theory in section D3. All these arguments essentially constitute channels for a centripetal tendency of economic activity that has been neglected by neoclassical models. The intuitive arguments have been used by older theorists like Myrdal (1957), Hirschman (1958), Pred (1966), or Kaldor (1970), who have described regional divergence processes before, but without a compelling micro-foundation with respect to the underlying market structures.

Of course there are not only centripetal, but also centrifugal forces and other influence factors that shape economic landscapes in the real world. We discuss the intuition of some of these in section D4. During our discussion of various agglomeration and dispersion forces, we put some emphasis on the particular mechanisms that show up in the seminal core model of NEG. On the centripetal side these are "endogenous market size effects" or "market linkages" due to increasing returns at the firm level. The opposing dispersion force in NEG stems from local immobility of demand in conjunction with transportation costs.

After having introduced the intuitive arguments for spatial agglomeration and the possible opposing forces, we turn to the analytical part of this chapter. We will introduce the standard core-periphery model of Krugman (1991a) in section D5. We also highlight in this section in what respect the NEG is derived from the

[4] For an appraisal of the impact of the Dixit-Stiglitz-model see e.g. Krugman (1998) or Brakman/Garretsen/Marrewijk (2991:69ff.).

"new trade theory" (NTT) of Krugman (1980). Ever since Krugman (1991a,b), NEG became a rapidly growing field of interest in academic economics. Various other models have appeared that elaborated on different agglomeration and dispersion forces, used different model assumptions, worked with different solution techniques etc. We will highlight and survey some distinguishing ideas and modelling strategies in section D6 and discuss the most recent developments in the field of NEG.

Still, we come to the conclusion that the state of the art in NEG still has left various open and unsettled issues, out of which we concentrate on two. The first one directly concerns the main topic of this book. NEG models almost without exception assume full employment and automatic labour market clearing. They are therefore incapable of analysing regional unemployment disparities. This neglect will be taken up in chapter E, where we try to pull the pieces of regional agglomeration theory and regional unemployment together.

However, the neglect of unemployment is not the only omission of NEG as it stands today. In section D7 we point to another strange implication of the Krugman-model that is at odds with real world evidence. The standard NEG-model predicts that the industrial agglomeration centres have lower overall costs-of-living than rural peripheral areas. This empirically unsatisfactory result has not been addressed in the literature so far. We therefore take up this issue and develop an own NEG-approach that is capable of reproducing the main results of the Krugman-model, but that has more reasonable implications with respect to regional costs-of-living. This issue is a bit off-topic with respect to the analysis of regional unemployment disparities. Nevertheless, it improves the state of art in agglomeration theories, which are "one half" of our theoretical backbones. Therefore the model developed in section D7 should be seen as an additional and complementary contribution that comes from this book.

D2) Scale economies, externalities and market competition

Increasing returns to scale describe a situation in which an increase in the output level implies a decrease in the average costs per output unit. One can distinguish between "internal" and "external" economies of scale (Scitovsky, 1954). The former concept refers to the case where a single firm faces a downward sloping average cost curve when increasing its own output level. This type of scale economies arises e.g. if production incurs fixed costs and marginal costs are constant. Internal scale economies at the plant level play a dominant role in the Dixit-Stiglitz-model and therefore also in NEG. It is clear that internal scale economies are inconsistent with perfect competition, since merging production of two atomistic firms would always be dominant over separating production activities. The Dixit-Stiglitz-model of increasing returns therefore works with *monopolistic competition*: each firm produces a distinct commodity under increasing returns. The single varieties are linked in the sense that consumers can substitute between these single

goods. The elegant feature of the Dixit-Stiglitz-model', that allows for many of its derivations, is the critical assumption that all single commodities enter symmetrically into the utility function of the representative consumer and that the elasticity of substitution between commodities is constant.

The concept of "external" increasing returns on the other hand refers to the case that scale economies arise on an aggregate (spatial or industry) level, making average costs per output unit a decreasing function of aggregate output. One can distinguish between *pure* and *pecuniary* externalities. With pure (technological) externalities, an increase in aggregate output changes the technological relation between input and output for each firm, i.e. it affects the production function on the micro level. Typically, the aggregate externality enters the production function through some term capturing the total factor productivity A. Consider e.g. the production function of a single firm to be

$$y^j = A(Y)k^{\alpha}n^{1-\alpha}$$

where y^j is firm output, k and n are the firms capital stock and employment respectively, and Y is some aggregate variable (e.g. aggregate output) beyond disposal of any single firm. An increase in Y has a positive effect on the total factor productivity A and thus on the firm's output y for given input levels of k and n. Since the aggregate externality enters parametrically into the function y, this type of increasing returns is consistent with perfect competition (Chipman, 1970). This modelling strategy has a long tradition in particular in the first wave of endogenous growth theory, where the aggregate externality represents e.g. some form of "learning-by-doing" (Arrow, 1962), when A is a function of the aggregate capital stock K, or a human capital externality (Lucas, 1988), where A is increasing in the aggregate stock of human capital.[5]

Pecuniary externalities, or pecuniary external economies of scale, are instead transmitted through the market price mechanism and do not influence the relation between inputs and outputs (i.e. the production function) on the level of single firms. An increase in aggregate output rather has impacts on goods or factor prices e.g. through a "love-for-variety effect" on the consumer side, or through a greater variety of industrial inputs. This reasoning is often labelled a "market linkage" or "market interdependence" and is the second essential feature of the Dixit-Stiglitz-model of monopolistic competition. This seminal model thus works with internal scale economies on the plant level, as well as with a pecuniary externality on an aggregate level. The details of this proposition will become clear below when we

[5] For an overview of this first wave of endogenous growth theory, also in opposition to the second wave of "innovation based" growth models, see Romer (1994). Essentially the effort of this string of theory was to look for ways to overcome the "tragedy of diminishing returns" inherent to neoclassical production functions, but inconsistent with real-world evidence on income convergence.

discuss the core model of NEG, which is essentially a spatial version of the Dixit-Stiglitz framework.
To sum up, increasing returns to scale usually require that markets are imperfect. Marshall (1890) was probably not fully aware of this requirement when he developed three distinct arguments why economic activity tends to push for spatial agglomeration. To these intuitive arguments we turn next.

D3) The Marshallian agglomeration economies
The three Marshallian arguments for industrial agglomeration can be labelled in modern terminology as follows. Agglomeration of economic activity ensures a) the availability of large markets for specialized inputs, b) the presence of knowledge spillovers, and c) forward and backward linkages stemming from a large market size for final products. In this section, we will briefly introduce the main ideas for either of these forces, as well as introduce another argument from endogenous growth theory.

a) Large markets for specialized inputs
In simple neoclassical models, input factors are mostly homogenous. This means that all workers are assumed to be identical, with no differences in skill, education etc. In modern real–world societies, however, production processes are highly sophisticated and require very specialized skills and intermediate inputs. These are not easily available everywhere. For example, computer specialists are concentrated in Silicon Valley and hard to find in Nebraska. Financial market insiders live in London or Frankfurt, not in Finland. Modern firms thus have an incentive to locate where such specialized human capital is available. On the other hands, specialists in any industry have an incentive to move to areas where there is not only one potential employer, but a high number of them. Thus, there is a reciprocal advantage for local concentration: skilled workers go where firms are, and firms go where workers are. Note that this ´cumulative causation´ mechanism particularly applies to high skilled workers. For many low skilled professions, there is often not THE place to go. But this does not mean that low skilled workers would not to gain from local concentration of specialists in their region. For example, if many sophisticated specialists pool in some region A, also the low skilled workers from that area will benefit, because different types of labour are complementary (Matsuyama, 1995). Theoretical issues on the labour pooling argument come e.g. from Krugman (1991b). Evidence for the empirical relevance of the argument has been provided e.g. by Audretsch/Feldmann (1996). Formally, the presence of a large pool of specialized inputs can be seen as a pecuniary external scale economy for any single firm, since being close to large factor markets enables firms to hire specific inputs at lower costs.

b) Knowledge spillovers and externalities

A different argument why firms from the same industry like to cluster together are external knowledge flows and spillovers that require personal proximity. For example, it is easier to gather information about competitors, about sector specific news, to make deals, to negotiate contracts etc. if many firms from the same industry cluster in the same location. Essentially, the argument can be expressed in formal terms like in section D2, where we have introduced the purely external form of increasing returns. Knowledge spill-overs imply that firms have an advantage when operating in some location A with a high density of firms from the same industry rather than in isolation. The pooling advantages shift the efficiency frontier of any single firm through a total factor productivity term as described above.

The use of pure externalities is therefore convenient if one is preoccupied to maintain a market structure of perfect competition. F/K/V and Krugman (1991b), however, criticize its use as a modelling strategy that "looks like assuming one's conclusion". Since spill-overs are not market-mediated but exogenously imposed, the concept lacks some theoretical substance and might suffer from "ad-hocery". This does not imply that it is irrelevant.[6] But the availability of imperfect competition models like the Dixit-Stiglitz-approach at least made it possible from a theoretical point of view to advance to other approaches of increasing returns that were more sufficing.

c) Market size effects and linkages

The last of the classical Marshallian arguments for agglomeration are linkages stemming from large markets for final products. This third mechanism is actually at the core of the Dixit-Stiglitz-model and thus of NTT and NEG. It will be discussed in much more detail below. Nevertheless we want to briefly describe its intuition here.

The linkage argument heavily rests on two essential assumptions: firstly, there are internal economies of scale at the plant level through the existence of fixed costs that restrict firms to only one location. And secondly, it is assumed that there exist transportation costs for final goods. Firms thus have an incentive to locate close to large markets with many customers nearby in order to economize on transportation costs.[7] Similarly, customers also like to be close to the firms, because they enjoy a greater variety of local consumption goods for which prices are not blown up by transportation costs. Additional linkages can arise within the production sector, as some firms might use the products of other firms as intermediate inputs.

---

[6] In particular urban economists have stressed the importance of external knowledge spill-overs for the shaping of cities and urban locations. See also Ciccone/Hall (1996)

[7] Transportation costs are relevant for the firm even if they are fully rolled over on prices, simply because demand and thereby profits will drop if prices are blown up by transportation costs.

If this is the case, the firm sector has a motive to concentrate spatially, since increasing returns ensure a greater variety of inputs that are available locally and do not need to be costly imported.
It is straightforward to see how ´cumulative causation´ can result: since firms like to go where customers are, and customers (=workers) like to go where firms are located, there is an endogenous centripetal tendency. We will develop further on this agglomeration channel when we discuss the core model of NEG.

d) Growth and innovation
Not belonging to the canonical Marshallian arguments for agglomeration, but still an interesting perspective comes from the so called "new growth theory". The second wave of endogenous growth theory also builds on the Dixit-Stiglitz-model, i.e. on imperfect competition in goods markets and increasing returns to scale.[8] It is viewing growth as a phenomenon of innovation and technological and structural change, not mainly as a phenomenon of accumulation. Contrary to old growth theories, technological progress is no longer seen as something that is „falling from the sky", but rather as the result of specific (and mostly private) R&D-investments. The innovators must thus firstly have rents in order to finance these investments, and secondly they must be able to extract temporary monopoly rents in case of a successful innovation. Both these requirements illustrate why a model set-up with monopolistic competition is much more appropriate than one with perfect competition and zero profits if one thinks about these Schumpetrian processes.
In spatial terms, the temporary monopoly rents accrue in the location where the innovation has been made. The newly created technological knowledge spills only imperfectly into other regions. If this logic is then combined with a tendency of increasing returns in the R&D-sector, it follows that innovation activities will reveal a high tendency towards spatial concentration. This concentration will perpetuate growth in those regions where the innovative research centres are located, and the diffusion of this growth into other regions will be imperfect. Because of that, the centre regions again have more resources available to invest in further, and even more sophisticated R&D. Growth and agglomeration might then end up in a cumulative causation mechanism, since they are mutually reinforcing processes (Martin/Ottaviano, 2001). This view is supported e.g. Audretsch/Feldman (1996), who show that in the US the geographical structures of innovation and production are quite similar, but that the innovation sector is stronger concentrated. In section A4.4.) we have shown that the same is true for the EU, where there is also a heavy spatial concentration of innovation in only a few European NUTSII-regions (EU-Commission, 2001). The considerations from this section

[8] Famous adherents of this class of "innovation based growth models" are Romer (1990) and Grossman/Helpman (1991).

highlight why the high spatial concentration of innovation might be a primary concern of policymakers interested in territorial equity.
The arguments presented in this section all claim that there are advantages from spatial concentration of economic activity, and may be labelled *centripetal forces*. None of the stories can be reconciled with the standard neoclassical paradigms (except to some extent the pure externality argument b). But of course there are not only centripetal forces in the world, but also centrifugal forces.

D4) Centrifugal forces and other location factors
If the world were only characterised by agglomeration forces, economic activity would only take place in one single location in the world. This is of course nonsensical. The degree of core-periphery divides observed in the real world suggests that centripetal tendencies might be quite strong. But of course there are also opposing dispersion forces. Together they will determine the spatial equilibrium structure of an economy, and a high degree of equilibrium agglomeration simply reflects that agglomeration forces are strong relative to dispersion forces. Furthermore, the economic landscape in the real world is not only determined by the balance of "pure" centrifugal and centripetal tendencies, but is also influenced by other factors, and subject to considerable inertial forces.

a) "Pure" centrifugal forces
The dispersion forces used in the seminal NEG-model of Krugman (1991a), but also in most other standard NEG-models like e.g. Venables (1996) or Puga (1999), is local immobility of demand in conjunction with transportation costs for final goods. The argument rests on the fact that some individuals (consumers) are assumed to be regionally immobile and inevitably tied to specific locations. They typically engage in some sort of basic production activity under constant returns and perfect competition, generally labelled "agriculture". They nevertheless demand also commodities that are produced by the manufacturing sector. The increasing returns in this sector push for concentration in only one location (the "centre"). But the demand from peripheral farmers is opposing this tendency. Since manufacturing firms are interested in satisfying also this demand segment, it is ambiguous whether the increasing returns outweigh the transportation costs that are necessary when the "peripheral" demand is served through exports. This centrifugal force is typical for NEG-models. Its use is driving many of the results of NEG-models that will become apparent below.
But demand immobility is by no means the only conceivable dispersion force. Actually, old location theorists like Isard (1956) or Weber (1909) have claimed that probably the most important boundary for spatial economic concentration is the limited availability of usable land and housing opportunities. Suppose that the housing stock in any region is fixed. In the process of agglomeration in one specific location, housing prices will be bid up. At some point will the agglomeration

advantages stemming from increasing returns not make up any more for the high housing and land prices. Surprisingly, however, most standard NEG-models neglect this highly relevant and plausible dispersion force. An exception is Helpman (1998), who explicitly sets up a model where increasing returns explicitly face housing scarcity.

Further arguments against spatial agglomeration are congestion costs, broadly defined. Metropolis areas might be characterised by high crime rates, pollution, the lack of traditional social networks and personal interactions, and all sorts of other sociological factors. But individuals are surely heterogeneous with respect to the weight they put on these congestion costs in their respective utility functions. Some individuals might even perceive specific "urban" characteristics like anonymity etc. to be intrinsically attractive. These individuals would thus not perceive "costs" of congestion, but rather benefits from it.

b) Status-quo inertia

Apart from centrifugal forces that directly oppose spatial agglomeration, there are other factors that in general produce a status-quo bias with respect to the spatial economic structure. Mobility costs are a good example. Neither relocation of firms, nor of workers is costless. Not only physical relocation costs (moving production facilities or furniture) have to be considered. Also more subtle, psychological mobility costs arise: the costs for adapting to new environments, the costs for gathering information about suitable housing etc. Mobility costs do not specifically favour or prevent agglomeration. They rather invoke a bias for any spatial economic structure that exists initially. Mobility costs can be seen as a inertial force that prevents the instantaneous adjustment of the actual spatial structure to the equilibrium structure that is determined by the balance of centripetal and centrifugal forces.

Related to mobility costs is the concept of a "regional preferences". Workers might intrinsically prefer to live in a specific area (e.g. the area of birth, in which case we would speak of a "home bias"). The reasons might be that they are accustomed to the local culture and language, they can draw on intact social networks etc. The presence of these biases leads workers not to respond instantaneously to interregional differences in observable economic variables with migration. Since living elsewhere is subject to a discount factor in the individuals' utility function, the difference in economic prospects must be sufficiently large in order to invoke a relocation of workers.

c) Other location factors

One should note that technological centripetal and centrifugal forces are always overlapped with unchangeable regional characteristics like the natural environment, the fertility of land, the climate etc. that might be summarized by the term "regional comparative advantages". These underlying features are at the root of

neoclassical trade and location models (as well as of the wage curve models discussed in the last chapter). We have argued that comparative advantages *alone* are insufficient to explain the spatial economic structure (and the pattern of sectoral specialization) in the real world, as they leave no room for cumulative processes. This, however, does not mean that comparative advantages are irrelevant. The interplay between exogenous comparative advantages and endogenous location forces has been discussed e.g. by Krugman (1999) and should always be kept in mind.

The enumerative list of various centripetal, centrifugal and other influence factors provided in sections D3 and D4 is supposedly still incomplete and subject to further expansion. The core-model of NEG, however, has essentially singled out one centrifugal and one centripetal force, and abstracted from all other location or inertial forces. We not turn to the discussion of this seminal model of Krugman (1991a).

D5) The core-periphery model of 'new economic geography'

Krugman (1991a) considers what he calls a "2 x 2 x 2"-model, i.e. he analyses an economy consisting of two regions r=1,2, two sectors (labelled agriculture A_r and manufacturing M_r) and two types of workers, farmers and manufacturing workers, who are specific inputs to the respective sector. Both regions are identical with respect to technology, preferences and endowments. There are no inherent comparative advantages or intrinsic regional differences. Agricultural production is assumed to operate under perfect competition and constant returns to scale. The products from the agricultural sector can be freely shipped across space. The manufacturing sector on the other hand is characterised by monopolistic competition and internal increasing returns to scale á la Dixit-Stiglitz (1977). Interregional transportation of manufacturing products imposes so called 'iceberg'-transportation costs that were introduced first by Samuelson (1952).

Furthermore, farmers are assumed to be regionally immobile and equally split across the two regions. Manufacturing workers on the other hands are mobile, but initially also split across the two regions. We have already pointed out above that the centrifugal force in this model will be the manufacturing demand of the regionally tied farmers. Let us thus shortly consider this critical assumption on the differential mobility behaviour of farmers and manufacturing workers, as this will drive the results in the Krugman-model.

One can more generally think of the agricultural sector as a basic services sector in which increasing returns does not play any role. The farmers who are employed in this sector are low skilled workers. The manufacturing, or "modern" sector however exhibits scale economies and employs mobile, high-skilled workers. There seems to be the implicit assumption that skilled workers from the modern sectors are in principle mobile, whereas low skilled workers are not. In chapter F

of this book, where we take up the issue of selective labour mobility more explicitly, we show that this assumption is in principle not unreasonable.

D5.1.) Consumer behaviour

The representative consumer in each region r={1,2} respectively maximizes a Cobb-Douglas utility function

$$U_r = M_r^{\mu} A_r^{1-\mu} \tag{D.1}$$

where A_r are units of the agricultural good that can be freely shipped across regions, and the aggregate M_r is a symmetric CES function over a continuum of $(n_r + n_s)$ single consumption varieties m(i) from the manufacturing sector. The parameter μ reflects the manufacturing share.

$$M_r = \left[n_r \left(m_{rr}(i) \right)^{\rho} + n_s \left(m_{sr}(i) \right)^{\rho} \right]^{\frac{1}{\rho}} \tag{D.2}$$

n_r indicates the number of varieties that are produced and consumed in the respective region itself (m_{rr}), and n_s is the number of varieties imported from the other region (m_{sr}). The parameter $0< \rho <1$ measures how close substitutes the single varieties are. The lower is ρ, the more differentiated are the single consumption goods, and the lower is the elasticity of substitution $\sigma = 1 / (1-\rho)$. Consumers maximize (D.1) subject to the budget constraint

$$p_r^A A_r + \left(n_r p_r m_{rr}(i) + n_s T\, p_s m_{sr}(i) \right) = Y_r \quad , \tag{D.3}$$

where Y_r is total regional income, and p_r^A is the price of the agricultural commodity. The ´iceberg´-costs that are involved in the transportation of manufactured goods across regions are captured by the parameter T. For one unit of $m_{rs}(i)$ to arrive, $T > 1$ units need to shipped and the rest "melts away" during the transportation.

Since the manufacturing aggregate M_r consists of symmetrical commodities, the mill prices $p_r(i)$ within each region are the same for all varieties, i.e. $p_r(i) = p_r$ and $p_s(j) = p_s$ for all i, j. Prices for the good A_r are equalized across regions due to the free tradability. Therefore p_r^A can serve as the numeraire and is normalized to one. Since the utility function is Cobb-Douglas, expenditure on each of the three aggregates is a constant share of income. The demand for the agricultural good is simply $A_r = (1-\mu)Y_r$, and the demand for the consumption aggregate M_r can be written as $M_r = \mu Y_r / G_r$. The variable G_r is the composite price index for manufactured goods in region r. It can be understood as an expenditure function dual to the

quasi-utility function M_r, i.e. it describes the minimum costs of purchasing one unit of the manufacturing aggregate M_r. The function G_r can be computed as[9]

$$G_r = \left[n_r (p_r)^{1-\sigma} + n_s (T\, p_s)^{1-\sigma} \right]^{\frac{1}{1-\sigma}}. \qquad \text{(D.4)}$$

The manufacturing price index G_r is decreasing in n_r and n_s, with the partial effect of an increase in n_r being stronger. These negative partial derivatives represent a "love for variety" effect and can be seen as a pecuniary external scale effect. If the number of available consumption varieties increases, the costs for purchasing one unit of an aggregate M_r, and thus the costs for attaining a given utility level, decline. This effect is stronger if the increase in varieties is local, because no transportation costs apply to the new varieties.
The demand functions for single local consumption commodities and for imported varieties are respectively given by

$$m_{rr}(i) = \mu Y_r \frac{p_r^{-\sigma}}{G_r^{-(\sigma-1)}},$$

$$m_{rs}(i) = \mu Y_r \frac{(T\, p_s)^{-\sigma}}{G_r^{-(\sigma-1)}} \qquad \text{(D.5)}$$

The parameter σ not only gives the elasticity of substitution between varieties, but is also equal to the price elasticity of any single manufacturing good. This construction is an artefact of the Dixit-Stiglitz-model frequently criticised in the literature.[10] It is also important to note that it is assumed that each single producer neglects the influence of price change for his specific commodity on the price index G_r.
Let $q_{i,r}$ be the total sales of a manufacturing firm i in region r. Using (D.5), and taking into account the fraction of the product that melts away during interregional transportation, sales per firm amount to $q_{i,r} = m_{rr} + Tm_{rs}$, or

[9] For details about the derivation of a composite manufacturing price index dual to a CES-function see F/K/V (1999:46f.)
[10] See e.g. Ottaviano/Tabuchi/Thisse (2002): "Though convenient from an analytical point of view, such a result conflicts with research in spatial competition which shows that demand elasticity varies with distance while prices change with the level of demand and the intensity of competition. Moreover, the iceberg assumption also implies that any increase in the price of the transported good is accompanied by a proportional increase in its trade cost, which is unrealistic."

$$q_{i,r} = \mu\left(Y_r \frac{p_r^{-\sigma}}{G_r^{-(\sigma-1)}} + T Y_s \frac{(p_r T)^{-\sigma}}{G_s^{-(\sigma-1)}} \right) \tag{D.6}$$

D5.2.) Production

The agricultural sector A_r operates under perfect competition and a very simple constant returns to scale technology: we say that in both sectors one unit of labour is transformed into one unit of output. It is convenient to normalize the size of the total agricultural workforce to $(1-\mu)$, such that $\frac{1}{2}(1-\mu)$ farmers live in either region. It is warranted with this technology that the nominal wages is equal to the respective product price, i.e. $p^A = w^A = 1$ in both regions.

Monopolistically competitive firms in the manufacturing sector on the other hand face fixed costs on the plant level. To produce q_i units of a consumption variety m(i), a firm at any location needs

$$\ell_i = F + \beta q_i \tag{D.7}$$

manufacturing workers, where F is a fixed and β a marginal input requirement. The increasing returns and people's preference for variety ensure that every single good m(i) will be produced by only one firm. Hence, the number of active manufacturing firms and the number of varieties produced in region r are identical. Manufacturers in region r earn a nominal wage w_r. Each manufacturing firm makes profits

$$\pi_{i,r} = p_r q_i - w_r(F + \beta q_i). \tag{D.18}$$

As noted above, the Dixit-Stiglitz-model abstracts from strategic interactions of manufacturing firms. Each firm knows that it is small relative to the whole market, and thus regards the influence of its price setting decision of the manufacturing price index G_r as negligible. This Chamberlinian "large group" assumption implies the familiar result that monopolistically competitive firms set prices as a constant mark-up over marginal costs, i.e. $p_r(1-1/\sigma) = \beta w_r$.

The next essential feature of the Dixit-Stiglitz-model is the assumption that profits will always be equal to zero due to the free entry of potential competitors. Thus, substituting the pricing rule into (D.18), we find that all manufacturing firms operate at a unique scale of output in both location. The output per firm is given by $q^* = F(\sigma-1)/\beta$, and the associated labour requirement is $\ell^* = F\sigma$. The number of active manufacturing firms n_r and n_s is then entirely determined by the regional labour supplies in the manufacturing sector.

Let the total manufacturing workforce in the economy have the size $L^M=\mu$. A fraction $\lambda_r\mu$ lives and works in region r, with $\lambda_s=(1-\lambda_r)$. For the moment, we can treat λ_r as a given parameter. The number of varieties produced in region r is then simply given by $n_r = \lambda_r\mu / F\sigma$. Without loss of generality, we can normalize units such that $\beta = (\sigma-1)/\sigma$ and $F = \mu / \sigma$. It follows that the output and the labour demand per firm are identically given by

$$q^* = \ell^* = \mu \tag{D.8}$$

Moreover, nominal manufacturing wages in region r are equal to the respective mill price of the symmetrical consumption varieties, and the number of active firms is directly given by the share of manufacturing labour in the respective region

$$w_r = p_r \tag{D.9}$$

$$n_r = \lambda_r \tag{D.10}$$

The equations (D.8) - (D.10) summarize neatly the central insights of the Dixit-Stiglitz-model. Even though there are internal economies of scale at the firm level, the presence of potential competition and the requirement of zero long-run profits imply that all firms operate at a unique scale of output. All scale economies in the model are driven by the number of manufacturing firms rather than through the individuals firm scales. This maximum number of varieties is restricted by labour supply, or respectively by employment. The second crucial point is that the mark-up of firms is also not affected by the market size. This is due to the assumption of complete symmetry and constant elasticity of substitution among all manufacturing varieties that is identical with the single price elasticities.

D5.3) Equilibrium conditions

We have laid out supply and demand for both sectors in the preceding two sections and can now move forward to determine the equilibrium conditions. This is most simple for the agricultural sector. Since A is freely tradable across space and the expenditure on A in a constant share of income, the market automatically clears at the numeraire price $p^A=1$.

Determining the equilibrium prices in the manufacturing sector is slightly more difficult. Equilibrium requires that markets for all consumption varieties simultaneously clear. From (D.8) we know that all firms supply a unique quantity of output $q^* = \mu$. This needs to equal total sales per firm, given by (D.6). Equating these two, and manipulating terms yields the following expressions for the regional mill prices p_r that by (D.9) must be equal to the nominal manufacturing wages w_r

$$p_r = w_r = [Y_r G_r^{\sigma-1} + Y_s G_s^{\sigma-1} T^{1-\sigma}]^{1/\sigma} \quad \text{(D.11)}$$

Equilibrium prices in the M-sector not only increase with the income level in the respective region itself, but also depend on income in the other region as well as on the composite regional price indices G_r and G_s. But the variables Y_r, Y_s, G_r, and G_s in turn also depend endogenously on w_r. Income is high in regions where wages are high:

$$Y_r = \lambda_r \mu w_r + \frac{1-\mu}{2}. \quad \text{(D.12)}$$

Similarly, the regional manufacturing price indices are increasing functions of the single commodity prices. This can be seen in (D.13), which is obtained by using (D.9) and (D.10) in (D.4)

$$G_r = [\lambda_r (w_r)^{1-\sigma} + (1-\lambda_r)(w_s T)^{1-\sigma}]^{\frac{1}{1-\sigma}} \quad \text{(D.13)}$$

By substituting (D.12) and (D.13) into (D.11), we can find the nominal wages w_r at which the manufacturing sector clears in both regions.
These equilibrium values only depend on the parameters T, μ, σ and on the endogenous variable λ_r. Recall that we have assumed that the manufacturing workforce is initially equally distributed across the two regions, i.e. $\lambda_r = ½$. It is, however, crucial for the derivation of the geographical equilibrium structure to know how wages w_r react upon changes in λ_r. Do equilibrium nominal wages rise in region r if more workers concentrate in that location, thus constituting a motive for agglomeration? Due to the involved non-linearities of equations (D.11)-(D.13), it is not possible to compute a closed form solution for w_r as functions of λ_r. But one can grasp some intuitive insights and single out different effects, summarized in proposition D1.

Proposition D1. *Since* $\partial G_r / \partial \lambda_r < 0$ *and* $\partial Y_r / \partial \lambda_r > 0$, *the sign of* $\partial w_r / \partial \lambda_r$ *is uncertain, because* $\partial w_r / \partial G_r > 0$ *and* $\partial w_r / \partial Y_r > 0$.

Intuitively, an increase of λ_r has two effects on nominal manufacturing wages w_r (see also F/K/V, 1999:53; Neary, 2001:542ff.). If more workers concentrate in region r, competition on the regional goods market increases. This "competition effect" is represented by the fact that the composite manufacturing price index G_r is decreasing in λ_r. Demand for each single manufacturing commodity is shifted

downwards, as can be seen in (D.5). Ceteris paribus, this implies that more factor market competition among manufacturing workers reduces manufacturing wages. At the same time, there is a so-called "backward linkage effect" operating in the other direction. With more manufacturing labour in region r, the number of firms n_r and thus income Y_r increase. The increase in n_r implies a rise in labour demand and puts incipient pressure on manufacturing wages. This in turn increases the goods demand in region r, giving rise to a secondary upward pressure on w_r.
Which of the competing effects is stronger depends on exogenous parameters. But we do not further develop on the effects on nominal wages, since ultimately we are interested in the reaction of *real wages* that will drive the location decision of the mobile manufacturing workers. Real wages ω_r can be obtained by deflating the nominal wages w_r with the aggregate regional price index. In this standard NEG-model, the regional cost-of-living index is simply identical with the composite manufacturing price index G_r, since the price for agricultural products has been normalized to unity. Real wages are thus given by

$$\omega_r = \frac{w_r}{(G_r)^{\mu}} \tag{D.14}$$

Similar to the nominal wage, also the real wage ω_r endogenously only depends on λ_r. Considering the real wage, there is an additional "forward linkage" effect. Since G_r is decreasing in λ_r, the establishment of new firms in the region implies an increase in real wages by lowering the living expenses for manufacturing commodities.
There are thus three competing effects on ω_r: The "competition effect" implies that real wages drop if more workers enter the regional factor market. This would imply that an initially symmetric situation with $\lambda_r = ½$ is inherently stable and agglomeration can not develop. However, there are the "forward" and "backward" linkages that operate in the other direction. Concentration of workers (=firms) in one region implies a greater variety of locally produced manufacturing commodities. Through the pecuniary external scale effect described above as the "love for variety" effect, this translates into higher nominal wages, and through the decrease for regional costs-of-living it complementarily implies higher real wages.
Whether linkage effects are strong enough to overcompensate the competition effect heavily depends on the exogenous parameters. If it is warranted, a symmetric initial situation will break up and a core-periphery structure with all manufacturing labour concentrated in one region develops endogenously. There are only two possible structures that can prevail in the long run. The economy will either be characterised by a complete agglomeration of all manufacturing labour in one

region (λ_r=1), [11] or the two regions are completely symmetrical. F/K/V have proposed an elegant way to determine the actual equilibrium structure as a function of exogenous parameters. First they take a situation as initially given where all manufacturing labour is pooled in region 1 (λ=1) and check under which parameter constellations this is sustainable.[12] If there are sustainable core-periphery-structures, the next step is to ask whether the economy can be driven to this equilibrium by the dynamics of the system itself. This is done by checking the stability around the *symmetric* steady state with $\lambda = ½$.

D5.4) Sustainability

When all manufacturing labour is pooled in region 1 (λ=1), the regional income levels in the core (Y_1) and in the periphery (Y_2) are

$$Y_1 = \mu w_1 + \frac{1-\mu}{2}$$

$$Y_2 = \frac{1-\mu}{2} \qquad \text{(D.15)}$$

By using (D.15) in (D.11) one can easily see that the equilibrium nominal wage in this completely agglomerated situation is simply w_1=1. The manufacturing price indices from (D.13)are given by G_1=1 and $G_2 = T > 1$ respectively. This clarifies that in the Krugman-model the peripheral region 2 is actually the area with the higher aggregate price index, since overall costs-of-living are identified in this approach with the composite manufacturing price index. We come back to this issue in section D7).

To check sustainability, we need expressions for the regional real wages. We already know that ω_1=1. Since there is no manufacturing in region 2 in the initial situation, no manufacturing wages are actually paid. However, we can derive the *theoretical value* for the real wage ω_2 that one single manufacturing worker could earn if he were to move from region 1 to region 2. This is done by inserting (D.15) into (D.11), and then by deflating this expression for the nominal wage w_2 with the regional CPI $G_2 = (T)^{-\mu}$. We obtain the following expression for the real wage ω_2.

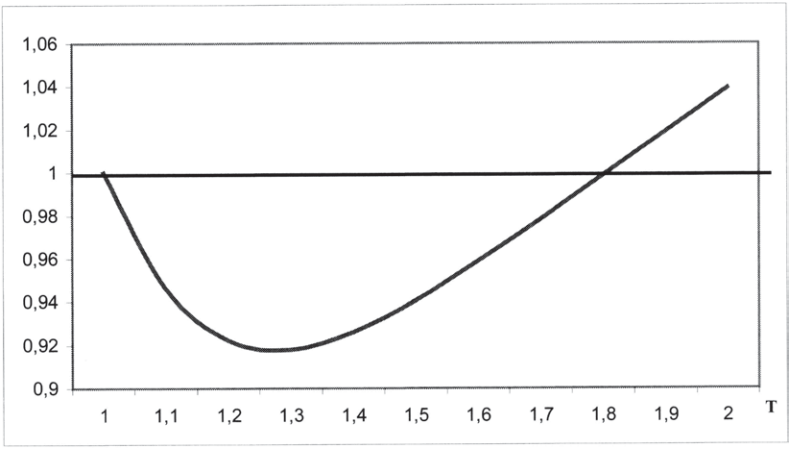
[11] Of course it is also possible that λ_r=0, and all mobile workers pool in region 2. But this completely analogous case of agglomeration will not be considered.

[12] We denote this situation simply by λ=1 and drop the subscript r from now on. Thus, $\lambda\mu$ manufacturing workers are located in region 1, and $(1-\lambda)\mu$ in region 2.

$$\omega_2 = T^{-\mu}\left[\frac{1+\mu}{2}T^{1-\sigma} + \frac{1-\mu}{2}T^{\sigma-1}\right]^{1/\sigma} < 1 \qquad \text{(D.16)}$$

The existing core-periphery-structure is sustainable if no manufacturing worker is better off by migrating from the core to the periphery, i.e. as long as $\omega_2 < 1$. Figure D1 plots the real wage ω_2 from (D.16) as a function of T and for given exogenous parameters μ and σ. The pre-existing geographical structure can be maintained for those values of T where the curve runs below the thick horizontal line $\omega_2=1$.

Figure D1: Sustainability of a core-periphery structure (σ=5, μ=0.4)



As can be seen, the function ω_2 has a "u-shape" with respect to T. Sustainability is warranted for low and intermediate values of T. With high transportation costs, however, a core-periphery structure is not sustainable, since workers would be better off migrating to the peripheral region 2. The reason for this result is straightforward: recall that ½ (1-μ) farmers are inevitably tied to region 2. These consumers spend a fraction μ of their income on manufacturing commodities. With high transportation costs, it is too costly to satisfy this demand through exports from the core region 1. The centripetal force stemming from the localized increasing returns in the M-sector do not compensate for the high amount of resources wastes on transportation.

Vice versa, if T is on a low or intermediate level, the increasing returns are more pervasive than the transportation costs. This is particularly so if T is on an intermediate level. If T is close to 1, i.e. if there is a situation of costless (or "free")

trade, the location of production does not matter any longer and there is perfect wage equalization across regions. The two other parameters affecting the equilibrium are μ and ρ (i.e. σ). Ceteris paribus, sustainability is more likely the lower is ρ, i.e. the more important are increasing returns in manufacturing, and the higher is μ, i.e. the more important is the manufacturing sector (F/K/V, 1999: ch. 5).

D5.5.) Stability

Sustainability only means that an initially existing core-periphery-structure can be maintained for given parameter values. This, however, does not address the process of agglomeration, i.e. whether a c-p-structure can emerge endogenously in the model. To deal with this issue, one has to assume that a symmetric equilibrium with $\lambda = ½$ exists initially. Around this steady state, all endogenous variables, including ω_1 and ω_2, are identical in both regions. F/K/V then check whether this steady state is stable or unstable for given parameters. For the emergence of agglomeration, instability is necessary: one single worker must be better off when migrating from region 2 to region 1. If he is, a circular and cumulative process develops, at the end of which the whole manufacturing labour force will be located in region 1.

Formally this condition can be written as $d\omega_1/d\lambda > 0$ and $d\omega_2/d\lambda < 0$ around $\lambda = ½$. Since under initial symmetry small changes of endogenous variables in one region are exactly mirrored by changes in the opposite direction in the other region ($d\omega_1 = -d\omega_2$), we can write the condition for instability simply as $d\omega/d\lambda > 0$. The analytical expression for $d\omega/d\lambda$ as a function of exogenous variables can be obtained by computing the values of all endogenous variables (D.11) - (D.13) for $\lambda = ½$, substitute those into the real wage equation (D.14) and then differentiate this expression with respect to λ. It yields the following expression[13]

$$\frac{d\omega}{d\lambda} = 2ZG^{-\mu}\left(\frac{1-\rho}{\rho}\right)\left[\frac{\mu(1+\rho) - Z(\mu^2+\rho)}{1-\mu Z(1-\rho) - \rho Z^2}\right] \qquad \text{(D.17)}$$

where $Z \equiv \frac{\left(1-T^{1-\sigma}\right)}{(1+T^{1-\sigma})} = \frac{\left(1-T^{1-\sigma}\right)}{2G^{1-\sigma}}$.

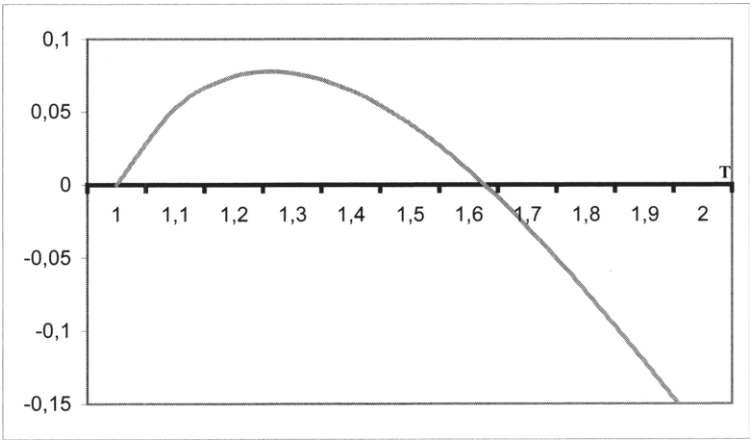
This condition (D.17) is graphically represented in figure D2 as a function of T, with exogenous parameters μ and σ as in fig. D1.

The symmetrical equilibrium is unstable if (D.17) is positive, i.e. if the curve runs above the thick horizontal line. This is the case for low and intermediate transportation costs. At some critical value of T, the property of symmetry breaking will

[13] Details of the derivation are given in F/K/V (1999: 76f.).

vanish. It can be shown that this critical level, the break point (T_B), is strictly lower than the critical level of T beyond which sustainability does not hold anymore, the sustain point (T_S). This is the so-called 'locking-in-effect' that gives rise to a tomahawk bifurcation and multiple equilibria. It can be seen best in the bifurcation diagram in figure D3).

Figure D2) The stability of the symmetric equilibrium



It describes the geographical structure of this economy (given by the spatial distribution of the manufacturing workforce) as a function of transportation costs T. Solid lines represent stable equilibria, whereas dotted lines represent unstable ones.

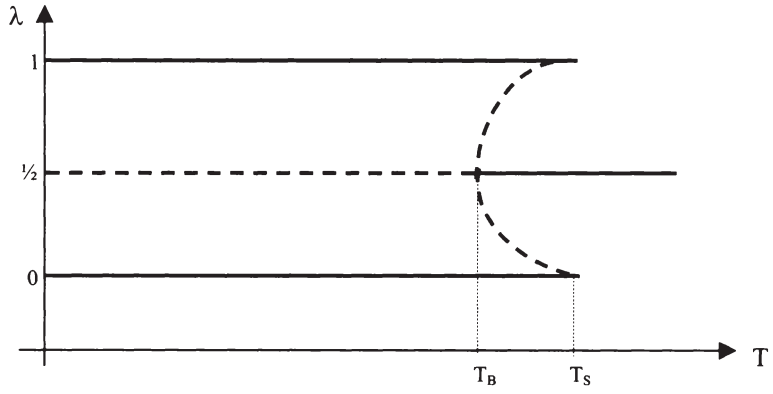
Figure D3 summarizes the working of the Krugman-model. Various points should be noted. Firstly, for high trade impediments ($T > T_S$) there is a unique and stable geographical equilibrium that is characterised by complete regional symmetry. The manufacturing workforce is equally split across the two regions ($\lambda = ½$), and all endogenous variables are identical in both areas. We have described the reason for this result above. For values of T below T_B, the whole manufacturing workforce is concentrated in only one region (which one is indeterminate), because the centripetal increasing returns in the M-sector are more pervasive.

At intermediate levels of T ($T_B < T < T_S$), the model exhibits multiple equilibria and local hysteresis. In this range, a symmetrical equilibrium would not break, but a pre-existing core-periphery equilibrium would be sustainable. The actual equilibrium configuration thus depends on the initial conditions and is "path-dependent".

It has been central to the NEG-literature to analyse the effects of secularly declining transportation costs. The term is broadly defined and captures all sorts of spatial transaction costs, including information costs, trade impediments etc. Falling transport costs can either represent regional integration (for example in the EU), or it can be used as a proxy for "globalisation", or it might reflect the decline of spatial transaction costs brought about by road infrastructure investments or innovations in the telecommunication sector.

The important implication of the Krugman-approach is that declining trade impediments can result in regional divergence. When trade costs decline from a level that is initially very high, they first pass the level T_S. This has at first no consequences, as the symmetrical equilibrium is still stable. Once they cross the critical level T_B, however, the symmetrical configuration breaks and a core-periphery structure develops endogenously. A circular spiral of cumulative causation sets in: one manufacturing worker is better off when changing locations from region 2 to region 1. Through this initial move, the incentives for manufacturers to leave region 2 as well increase. Eventually all manufacturers will have emigrated from region 2. In the theoretical model, this cumulative causation mechanism does not unravel gradually, but the emergence of the core-periphery structure happens spontaneously. The model implies a "catastrophic" agglomeration.

Figure D3) The bifurcation diagram of the Krugman-model



Another important feature, that can not be seen directly in figure D3, is that the agglomeration rents for the centre are hump-shaped with respect to T. We have commented on this fact when discussing figure D1: the real wage premium of the centre fades away as trade costs T approach the free trade case with T=1. The disparity between core and periphery is highest for intermediate transportation costs.

In the original Krugman-model, this u-shape is immaterial, as region 2 will never regain manufacturing. The substance of the hump-shaped agglomeration rents, however, has shown up much clearer in other NEG-models in the aftermath of Krugman (1991a) which will be discussed below.

D5.6.) ´New economic geography´ and the new trade theory

We have argued above that the basic model of NEG is a direct follower of the core model of the NTT developed by Krugman (1980). Since the connection between the two is actually very close, and since we will also rely on insights from NTT later, we will briefly discuss this approach here.

In short, what distinguishes NEG from NTT is the assumption that high-skilled manufacturing workers are mobile across regions. In the two-country model of Krugman (1980), all agents are assumed to be immobile. Moreover, there is only one consumption good sector in the economy that is identical to the manufacturing sector in the above NEG model: there is a large number of monopolistically competitive firms, each producing a distinct consumption commodity under increasing returns to scale by employing labour only. The maximum number of varieties that a country can potentially produce is, as above, restricted by labour supply. Consumers "value variety", i.e. they have symmetric CES-preferences where utility increases in the number of consumption varieties N.

$$U_i = \left[N(c_j)^{\rho} \right]^{\frac{1}{\rho}} \qquad 0 < \rho < 1$$

In such a model, the increasing returns become an independent source of (intra-industrial) trade, apart from the traditional arguments known from neoclassical trade theory (comparative advantage, preference heterogeneity etc.). Allowing for international trade enlarges the number of available consumption varieties. Since consumers value variety, the introduction of trade increases welfare in both trading countries. If trade is costless, there will be no welfare differences between countries, since all consumption varieties are equally available everywhere.

In a second step, Krugman (1980) then assumes the presence of ´iceberg´-transportation costs. If this is the case, large regions have an advantage over small regions, since they will produce a larger number of varieties locally and hence save on transportation costs.. Via a zero profit condition similar to (D.18), this scale advantage of large nations is absorbed by a higher equilibrium wage rate of the larger country.

Since labour is assumed to be immobile in Krugman (1980), this wage differential does not lead to migration. The step from the NTT towards NEG is taken if one allows for labour mobility, which is an endogenous agglomeration channel. Viewed at it in this light, the introduction of the constant-returns sector ("agricul-

ture") in the core model of NEG, which has not been introduced in the core model of NTT, was simply needed in order to have a centrifugal force in the model. Without the agricultural sector, theoretically all labour would flow from the small to the large country and there would be no break on the spatial concentration process. All in all, one can see from this section that the core models of NEG and NTT are structurally very similar, if not identical.[14]

D6) Other ´new economic geography´-models

In this section, we will briefly describe the main ideas and results of some other NEG-models. The models that will be discussed in this section were all motivated in one way or the other by the seminal Krugman (1991a)-model and extended and changed it in various directions.[15]

D6.1.) Venables (1996) and Krugman/Venables (1995)

One important approach that is actually by now considered the second "standard" model of NEG has been developed by Venables (1996). His main motivation to depart from the Krugman-model has been the insight that labour mobility is low in the European Union, across regions but particularly across countries (see Decressin/Fatas, 1995; Puhani, 1999). The agglomeration channel in the Krugman-model, however, critically hinges on the mobility of manufacturing workers. Venables (1996) formulated a model, where labour is immobile and still core-periphery structures can develop endogenously (contrarily to Krugman, 1980). In his model, the agglomeration channel does not occur through labour mobility, but rather through the sectoral specialization pattern of regions.

Venables (1996) argues that firms like to be close to each other because of direct input-output linkages amongst themselves. In the model set-up, Venables introduces an explicit input-output-structure of the manufacturing sector, where upstream industries produce intermediate inputs for downstream industries. Both vertically linked industrial sectors are operating under imperfect competition and exhibit internal increasing returns to scale. Additionally, there is again a competitive sector (agriculture) in the two-region economy. Labour is now assumed to be immobile across regions, but mobile across sectors. Interregional transportation of commodities imposes the usual ´iceberg´ transaction costs. A similar model set-up as in Venables (1996) has already been proposed in Krugman/Venables (1995),

[14] For a very intuitive discussion of the common elements and the differences between NTT and NEG see Krugman (1999).

[15] There are by now various surveys of the NEG, as well as textbooks that cover the contents of the original contributions. For surveys see Ottaviano/Puga (1998), or Puga (2002). The standard models are furthermore covered in the textbook of Brakman/Garretsen/Marrewijk (2001) on an introductory level. The most recent contribution of Baldwin et.al. (2003) presents a very comprehensive overview of different NEG-models in great technical detail. A broader survey that also discusses NEG in relation to other approaches in urban and regional economics is Fujita/Thisse (1996).

only with the difference that there has been no explicit input-output-structure between upstream and downstream firms. The (monopolistically competitive) manufacturing sector in that model is rather assumed to produce differentiated commodities under increasing returns that are used both as final consumption goods and as intermediate inputs for other manufacturing firms.

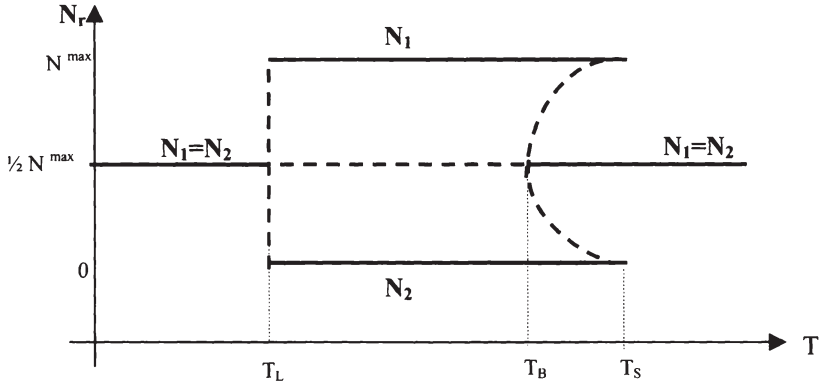
The results and the underlying logic of the two approaches of Venables (1996) and Krugman/Venables (1995) are, however, qualitatively quite similar to each other. It is important to note that there are 'forward' and 'backward' linkages between vertically linked firms from the manufacturing sector. Again, these linkages show up in the form of pecuniary externalities. The basic intuition can be described as follows: with high transportation costs each region will be essentially self-sufficient and produce both manufacturing and agricultural goods locally to satisfy the regional consumption demand. As trade costs fall, there will be increasing two-way or intra-industrial trade of manufacturing intermediates and final goods. But there will at first be no change in the relative sectoral specialization of either region. However, at some point when trade costs have fallen further, the involved agglomeration economies make it profitable to spatially concentrate manufacturing production and satisfy peripheral consumption demand through (costly) exports. This leads both to backward linkages, as manufacturing firms located in the agglomeration area have easier access to intermediate inputs. And there is a forward linkage, as more intermediates also imply declining costs for producing the goods destined to final consumption. These linkage effects together dominate over the centrifugal force, which again stems from the fact that final goods need to be shipped also to the de-industrialized region.

There is thus also a critical level of transportation costs below which there will be industrial concentration in one region. Since this process can not be triggered through migration of workers from the other region, the industrial core must draw on workers formerly employed in the agricultural sector. This increase in labour demand will put incipient upward pressure on real wages in the industrial core. Similarly, real wages in the de-industrialized periphery tend to drop, since manufactured goods must now be imported and transportation costs need to be paid. A real wage (or: real income) gap opens up between the two regions as the industrial sector concentrates in only one location. But if transportation costs continue to fall, the market access considerations, the importance for firms to be close to each other, gradually vanishes. Eventually factor price consideration will again come to dominate the location decision of industrial firms, which have an incentive to relocate their plants to the low-wage peripheral area.

Krugman/Venables (1995) see as the main appeal of this model that initial spatial divergence and subsequent convergence between the two regions are attributed to the same cause, the long-term decline in transportation costs. The exact shape of the bifurcation diagram depends on several details that shall not be discussed

here.[16] The essence of the Krugman/Venables-model can, however, be highlighted in figure D4).

Figure D4) Location of the industrial sector in the Krugman/Venables-model



The figure depicts the number of manufacturing firms in region r as a function of transportation costs. As argued above, there is complete regional symmetry if trade costs are very high. Once trade costs fall short of the critical level T_B, there is a catastrophic tomahawk bifurcation and a multiple equilibrium range between T_B and T_S. This time, however, the "catastrophic" event refers to the complete de-industrialization of region 2, and the complete specialization of region 1 to manufacturing production. What is new compared to the Krugman-model is that below some other critical transport cost level T_L, there is a re-emergence of symmetry. Below T_L, factor price consideration outweigh market access considerations. The core-periphery structure, that gives rise to an interregional real wage gap, is only stable for intermediate levels of transportation costs.

The regional real wage levels as functions of T can also be highlighted in a stylised manner. This is done in figure D5).[17]

As long as the location of industries is symmetrical, there is no interregional real wage gap. With the catastrophic transition at T_B, there is a spontaneous rise in the real wage of region 1, and a spontaneous drop in ω_2. The agglomeration rents, however, eventually fade away as the trade cost decline evolves. Finally, when symmetry re-emerges, there is again regional wage equalization.

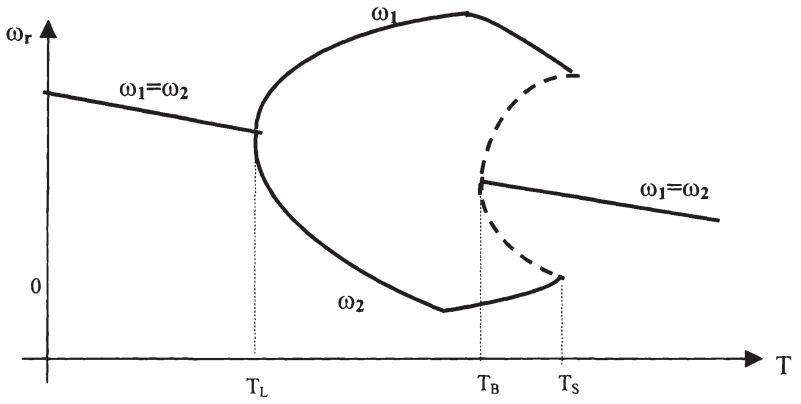
---

[16] See F/K/V, ch.16 ff. and Venables (1996) for an extensive discussion.

[17] Again, figure D4b) is only meant as a stylised representation of the main ideas of the class of models introduced by Krugman/Venables (1995) and Venables (1996). It must not be understood as the bifurcation diagram of one particular model version.

All in all, one can look at the Venables-model as a way to explain the emergence of core-periphery patterns also for environments in which geographical labour mobility plays a minor role. It has been shown that agglomeration forces can also arise because of inherent linkages only between vertically linked firms. The dispersion force that has been used was exactly identical to that in Krugman (1991a). Another main theoretical contribution in our view is the possibility to account for the existence of a lower bound of transportation costs, where regional disparities do no longer prevail. It is a representation of a plausible economic consideration, namely that factor price differentials will become an increasingly important factor for firms when market access considerations loose importance.

Figure D5) Real wages in the Krugman/Venables-model



This "Ω-shaped" relationship between transport costs and real wage disparities highlighted in figure D5) is studied further by Puga (1999). He uses an unified framework in which both interregional migration and input-output linkages may drive agglomeration. Essentially, the framework proposed by Puga combines the approaches of Krugman (1991a) and Krugman/Venables (1995). In Puga's model, either both agglomeration forces can be at work, or only one of them. He confirms that the Krugman-type of agglomeration (as described in section D5) is going to prevail when labour is regionally mobile, and he shows that the centripetal forces are even magnified if one additionally allows for input/output-linkages between firms. If labour is immobile, however, industrial agglomeration can only be expected for intermediate levels of transportation costs. Since spatial concentration

raises local wages, symmetrical equilibria will reconvene as trade costs approach low levels. This logic is essentially identical to the model of Venables (1996).
The Puga-model has some attractive features. This concerns the fact that it allows for agglomeration patterns that seem more realistic than the results of Krugman (1991a), Krugman/Venables (1995) or Venables (1996). His model is able to replicate the emergence of stable equilibria in which the manufacturing sector is neither completely agglomerated in one region, nor symmetrically distributed across space. It rather exists the possibility of an intermediate equilibrium in which one of the two regions has a larger manufacturing share than the other, but the "periphery" is not completely deserted from all industrial activity. And secondly, the Puga-model can go beyond the logic of "catastrophic" or discontinuous agglomeration. It leaves the possibility of gradual and continuous changes in the economic landscape.[18] The bifurcation results are surely fascinating from a theoretical point of view. Whether this logic is applicable to real world scenarios, is, however, disputable.
It is worth noting that the type of NEG-models in spirit of Krugman/Venables (1995), Venables (1996) and Puga (1999), which rely on input/output-linkages within the production sector, again have an important precursor in the NTT. This is the trade model of Ethier (1982). That model must be seen in direct relation to the NTT approach of Krugman (1980). Ethier's point was to construct a model where the welfare gains of international trade do not develop because of a "love for variety"-effect for the consumers. He rather saw an increase in the variety of intermediate industrial inputs as the main effect of international trade. The positive welfare effects of trade in the Ethier-model are caused by the greater differentiability of the production process. However, he did not develop a "spatial" model. He assumed free tradability of industrial intermediates. His model was thus not equipped to study the possibility of endogenous agglomeration, even though he also worked with internal increasing returns to scale, since location questions do not matter if there are no transportation costs involved. But there is a fundamental similarity between Ethier (1982) and the NEG-models discussed in this section insofar, as both focus on linkage effects and increasing returns that occur solely within the production sector of an economy.

D6.2.) Housing scarcity: Helpman (1998)
The NEG-model of Helpman departs in a more fundamental way from Krugman (1991a). It has been argued that the seminal model of NEG uses one particular dispersion force, namely the immobility of demand in combination with transpor-

[18] The analytical conditions that need to hold in order for these "attractive" results to prevail in the Puga-model are quite complicated and can not easily be phrased in an intuitive form. We therefore leave out the discussion of the respective conditions at this point.

tation costs. But it was also argued that there are other, maybe more relevant centrifugal mechanisms in the real world.

Helpman therefore uses housing prices as the dispersion force opposing the technologically given tendency to agglomerate. He replaces the traditional constant returns sector ("agriculture") from the Krugman-model with an immobile housing stock that is assumed to be equally owned by all individuals in the economy. The manufacturing sector in the Helpman-model is identical to the standard NEG-model, including the ´iceberg´-costs for interregional transportation. Labour is again assumed to mobile across regions.

With this model, it is quite straightforward to see the competing impacts on real wages that drive the location decision of workers. Given the increasing returns to scale technology alone, workers like to concentrate in only one region. But the higher housing prices oppose this tendency. The critical parameter that determines the relative strength of agglomeration and dispersion forces is again the level of transportation costs.

If transportation costs are quite low, the location of industrial production does not matter very much. The workers location decision is not driven by agglomeration wage differentials, but rather by housing cost considerations. The workforce will split evenly across regions, since there is little to be gained from the concentration in one region. The balance, however, shifts with higher transportation costs. Since interregional trade becomes more costly, workers have an incentive to concentrate. This is so for two reasons. Firstly, in order to enjoy lower goods prices, and secondly to exploit the localised agglomeration economies. Thus, the higher are transport costs, the smaller is the impact of housing prices on the workers' location decision. The fundamental diagram that shows the population fraction living in either region as a function of transportation costs looks like figure D6:

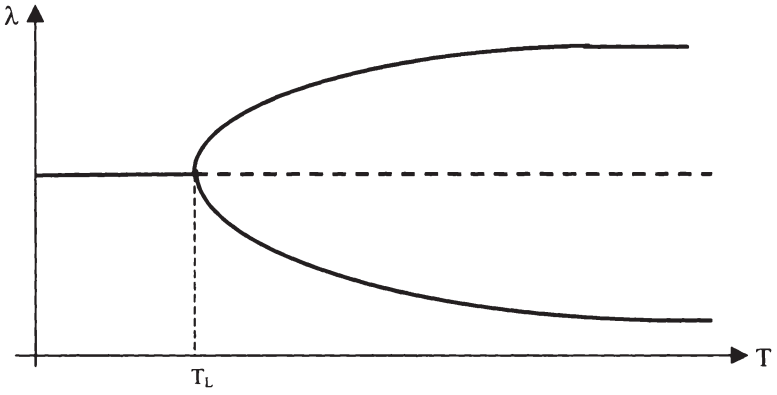
Below some critical level T_L, the two-region economy is characterised by complete symmetry, as considerations on housing prices drive the location decisions. The higher are transportation costs, the higher is the degree of population concentration in one of the two regions (which one is again indeterminate). Note that there is no catastrophic bifurcation in this model, but rather a smooth transition from symmetry to a core-periphery structure. There is the possibility of intermediate equilibria beyond full concentration and complete symmetry.[19]

These are surely attractive features that improve on the realism of NEG. However, one should note that the final results of the Helpman-model completely reverse the logic of the Krugman-approach. Helpman finds that agglomeration is more likely the higher are transportation costs. This fundamentally contradicts the notion that is central to much of the NEG-literature, namely that secularly falling transportation costs result at least initially in regional divergence and the emer-

[19] This property is the reason why the Helpman-model is often used for empirical research on NEG, see e.g. Brakman/Garretsen/Schramm (2001)

gence of a c-p-pattern.[20] The implication that industrial agglomeration occurs when trade impediments are very high, seems implausible in the medium and long run. However, in section D7 we will present an own NEG-approach that somehow combines the Krugman- and the Helpman-approach and that preserves the desirable features from either of them.

Figure D6) The Helpman-model



The Helpman-model constitutes probably the most prominent example of a NEG-model that works with a different centrifugal force than the Krugman-model. However, it has been pointed out already in section D4 that yet other centrifugal and location forces exist. Of course there are also NEG-models that have incorporated such arguments to the analysis. Urban congestion costs in NEG-models have e.g. been studied by Junius (1999). Intrinsic regional preferences that prevent individuals from instantaneously migrating are covered in the model of Ludema/Wooton (2000).

[20] Indirect empirical evidence seems to support the implications of the original Krugman-model over the Helpman-model. Faini (1983) has argued that the infrastructure improvements, vulgo: falling transportation costs, between Northern and Southern Italy in the 1950s have led to a deindustrialisation of the Mezzogiorno, as many firms found it profitable to shift production to the more efficient northern regions. More recently, Combes/Lafourcade (2001) report a similar finding for France: the reduction in spatial transaction costs that was estimated to amount to 38% between 1978 and 1993 led to a higher concentration of production and employment.

D6.3) Analytically tractable models

Another direction of departure from the Krugman-model was motivated by the fact that the solution method of NEG-models has to rely on numerical techniques, because the involved non-linearities prevent an analytical solution. Various authors such as Forslid (1999), Pflueger (2001), Ludema/Wooton (2000) or Ottaviano/Tabuchi/Thisse (2002) have proposed different ways how to cope with this problem.

The changes that have been undertaken usually have something to do with the form of the utility function that is assumed. In the standard Krugman-model, individuals have Cobb-Douglas preferences with a nested CES-function that represents preferences for the single manufacturing commodities. If one rather works e.g. with a quasi-linear upper-tier utility function, or with a quadratic utility function for manufacturing varieties, one is in general able to present closed-form analytical solutions. The details of the different approaches shall not be discussed here. It is worth mentioning, however, that the basic insights of the Krugman-model remain all in all stable and seem quite robust to changes in the modelling strategy.

D6.4) More sectors, more regions

A different limitation of the standard model from section D5 is its "2 x 2 x 2"-character. There are only two regions, two sectors and two types of workers. F/K/V admit that this set-up is quite restrictive, and further research should aim to generalize the findings in more realistic set-ups.

F/K/V contribute to this task themselves, first by introducing a three-region case and finally by working with a model with continuous space (F/K/V, ch. 6). They conclude that most of the original insights from the case with only two regions survive. The underlying logic of centrifugal vs. centripetal remains unchanged, and core-periphery structures emerge endogenously as a function of transportation costs. By introducing more than two regions one can observe a more complex picture in which, starting from the agglomeration of industry in a subset of regions, industrialisation might e.g. spread in a series of waves from country to country (Ottaviano/Puga, 1998). But fundamentally, the logic of NEG-models prevails also in these multi-region settings. The same is true for the two-region, three-sector model of de Vaal/van den Berg (1999).

This research direction has probably been pushed furthest by Forslid/Haaland/Midelfart- Knarvik (2002). They have generalized the standard NEG-approach to a setting with 14 sectors and 10 regions and calibrated a computable general equilibrium model against actual European data. They find that the relation between trade liberalization (falling transportation costs) and economic agglomeration can be approximated by an inverted U-shape, just as the Krugman-model implies.

D6.5.) Dynamic models

Attempts to marry NEG-models with endogenous growth models have been provided by Baldwin (1999), Baldwin/Martin/Ottaviano (2001) and Martin/Ottaviano (2001). The intuition of these models is close to the mechanisms that have been described in section D3d). The main result can be summarized that growth and agglomeration can be viewed at as mutually reinforcing processes.

For example, Martin/Ottaviano (2001) use a two-region model of endogenous growth, similar to quality ladder model of Grossman/Helpman (1991). They show that industrial centres reveal a higher growth rate than peripheral areas due to their high innovative activity. This higher growth enables the centre to invest even more heavily in R&D and to reinforce and strengthen its advantaged position. The consequence of this cumulative logic is a divergent trend of regional per capita GDP. Empirical support for this view comes from Quah (1997), who has shown that there is a positive correlation between the degree of agglomeration and the aggregate growth rate among European cohesion countries. According to his study, countries like Portugal and Spain that exhibited the fastest aggregate growth on the national level at the same time experienced a divergent trend of regional income levels over the recent decade. Greece on the other hand had roughly stable relative income levels of its single regions, but also a smaller aggregate growth rate.

D6.6.) Empirics, politics and other unsettled issues

Last but not least we want to highlight two further directions that research in economic geography has taken in the aftermath of the Krugman-model. It has been pointed out by Neary (2001) that the next steps that should be taken in NEG is to gather empirical evidence in support of the central propositions. In this respect, the contributions of Hanson (1997, 2001), Brakman/Garretsen/Schramm (2002), Dumais/Ellison/Glaeser (2002), Ciccone (2002) and Ciccone/Hall (1996) shall at least be mentioned as examples of a currently very active area of research. In general, these studies indicate that geography and agglomeration economies indeed matter, although the studies are mostly not decisive which agglomeration force is most relevant and which NEG-model is supposedly most appropriate.

Finally, policy issues have so far played little role in NEG. Again, various directions of research have been pursued already. In general it should be noted that policy of course plays a role in shaping economic landscapes. Some jurisdictions might pursue better policies than others, which can attract mobile economic agents. Analysis of the locational impacts of public policy has been provided e.g. by Ludema/Wooton (2000), Baldwin/Krugman (2002), Haufler/Wooton (2001) or Ross (2002). These papers subscribe to the idea that single regional governments use policy instruments in a strategic way in order to attract mobile economic activity, since this bestows rents on the local voting population.

A different way how policy can enter the spheres of NEG is by asking what the positive theoretical models imply for regional policy, e.g. for the structural policy pursued by the European Commission. This issue has been analysed e.g. by Martin (1998, 1999, 2000), Puga (2002), Boldrin/Canova (2001), Ottaviano/Thisse (2002) and Suedekum (2002a). These two areas, empirics and politics, will probably be the main directions of further research in the future of NEG.

All in all, our view is that NEG is a useful strand of theory that has caused a revival of spatial and regional questions to mainstream economics. Arguably, the main contribution of the NEG was not so much the introduction of ideas that were completely new to economists. Its main achievement rather has been the formulation of a rigorous general equilibrium model of a spatial economy with increasing returns as a distinguishing feature.[21] Nevertheless, we want to single out two shortcomings of the current "state of art" in NEG and geographical economics on which we want to develop further in this book.

Firstly, all standard approaches of NEG assume full employment and abstract from the analysis of unemployment. This is surely problematic, as regional unemployment disparities are at least as pronounced as income disparities, surely in the European Union (see chapter A) but supposedly also elsewhere. This deficiency of the theoretical literature is one motivation to look for an integration of agglomeration theory with unemployment theory, since this has not yet been done. The next chapter is devoted to this issue.

In the remainder of this chapter, however, we want to point to a different problem of the current NEG-literature: the issue of regional cost-of-living indices.

D7) Regional costs-of-living: an extension of the Krugman-model

The standard NEG-model of Krugman (1991a) predicts that the *overall costs-of-living* are higher in the periphery and lower in the centre. The reason is that the overall CPI in that model is identified with the composite manufacturing price index that is lower in the agglomeration area. This fact has been discussed above in section D5.4.) and is a so-called forward linkage that further increases the sustainability of a c-p-structure once it is established. But this result is at odds with real world experiences. Certainly, *some* products are cheaper in metropolitan core areas than in remote rural places, partly because producers are located nearby and therefore save on transportation costs, partly because competition is stiffer in many sectors. However, overall costs-of-living are definitely higher in metropolitan centres like New York or London than in remote rural areas, most importantly because of differentials in housing costs. Put differently, the real wage premium that centres need to pay in order for a c-p-structure to sustain in the long run is not

[21] For a critical debate on what has been added by NEG to the profession's state of knowledge, see e.g. forum debate in the first issue of the *Journal of Economic Geography*, Vol. 1 (1).

complemented by lower central CPIs in the real world, but rests exclusively on a nominal wage premium.

This artificial and (as we shall see) counterfactual result is due to the fact that the limited availability of land and housing is neglected as a force putting boundaries on regional concentration in the Krugman-model. The same is true for other standard NEG-models like Venables (1996) or Puga (1999), which similarly rest on local immobility of demand in combination with transportation costs as the centrifugal force. The only model in spirit of NEG, where housing prices are explicitly used as the dispersion force opposing the technologically given tendency to agglomerate is Helpman (1998). But we have argued above that his model leads to quite unconvincing results that sharply contrast central insights of the NEG-literature. In the remainder of this chapter, we therefore aim to build a model that preserves the main features of the seminal approach by Krugman (1991a), but that explicitly takes into consideration the important role of housing scarcity.

We present a standard NEG-approach where the agricultural sector will not be eliminated in exchange for a housing stock as in Helpman (1998). We rather add a home goods sector and extend the analysis to a three-sector case. The term ´home goods sector´ is more or less synonymous to housing, and only slightly more general. The purpose of its use is only to point out the two basic properties that the new good in our model must have: it must be non-tradable, and it must be non-reproducible. Herewith we show that the endogenous emergence of a c-p-structure in spirit of the Krugman-model can prevail. Not surprisingly, agglomeration is less likely the more important is the additional centrifugal force. Our second and more important point is to show that c-p-structures can develop and persist in which the core exhibits a higher overall CPI than the periphery. For this type of equilibrium to occur, not only transportation costs, but also the expenditure share for home goods must be on an intermediate level. If this so, agglomeration advantages in form of higher nominal wages and lower manufacturing prices dominate over the higher home goods prices in the centre. But even though rents do not overcompensate both centripetal forces, they are substantial enough to make the centre the more expensive area on balance.

Our contention is that this approach is somehow more complete than the standard c-p-model of Krugman (1991a). It takes account for the view that housing (home goods) prices play an important role for shaping the geographical structure of an economy, but it does not end up in a reversal of the compelling "u-shape" logic with respect to the spatial effects of falling transportation costs. Most parts of the standard NEG-model remain unaltered, but we can account for the plausible and empirically relevant case that c-p-structures emerge in which the centre has higher overall costs-of-living than the periphery. Before we lay out the formal model, we present some descriptive evidence about the relation of agglomeration and regional costs-of-living indices in section D7.1.).

D7.1.) Regional costs-of-living: the evidence

Unfortunately it is not possible to present evidence on regional price indices for the European Union or only for Germany, simply because this data is not available. Regional price level data do not even exist on the basis of US states. However, there exist comparative costs-of-living (COLI) data for US metropolitan areas compiled by the private sector company Dowden & Co. and used extensively for salary surveys of the American Chamber of Commerce. The dataset entails local costs-of-living for 399 metropolitan areas relative to the US average in 2000. We will relate this data with variables that reflect the degree of economic agglomeration within a spatial unit. The basic NEG-model implies that an economic core region is characterised by a larger population size and a larger workforce than a peripheral area. Furthermore, the core regions pay higher nominal wages and hence workers have higher nominal incomes.

Operational measures for these agglomeration indicators can be obtained by the regional accounts information provided by the Bureau of Economic Analysis (BEA) of the US Department of Commerce. Available data from this source are the size of the local population (POPULATION) and the number of jobs (JOBS) in a metropolitan statistical areas (MSA), the average nominal wage per job (WAGE) and the average personal income per capita relative to the US average (INCOME), all for 2000. Our full dataset contains 208 MSAs in the US for which all data, the COLI and the data provided by the BEA, were available. Table D1 provides an overview by showing characteristics of the ten most expensive metropolitan areas in the US as well as of those ten MSAs with the lowest COLI in 2000.

As can be seen, among the expensive MSAs are the classical US metropolis areas like New York, San Francisco, Boston, Washington D.C. etc. with huge populations, numerous jobs and relatively high wage and income levels. Areas at the other end of the scale reveal exactly the opposite characteristics. The group of relatively cheap MSAs consists of more rural parts of the USA, with low size and relatively low earnings and incomes.

In table D2 we show how the regional characteristics are correlated in the full dataset with all 208 observations. Our main interest is the relation between the regional COLI-level and the various measures for the degree of agglomeration. The correlation coefficients reported in the first line indicate that there is a strong positive relation between the COLI and each of the agglomeration indicators. All coefficients are significantly positive at the 1%-confidence level. The MSA's nominal earnings indicators (WAGE, INCOME) seem to be somehow stronger correlated with the COLI-level than the regional size measures (POPULATION, JOBS). In particular, relatively expensive MSAs tend to have high average wages per job.

Table D1: Income, size and cost-of-living in US metropolitan areas

Area (n=208)	Jobs (thousands)	average wage per job ($)	rel. pers. income	population (thousands)	rel. COLI
New York, NY (PMSA)	4500,156	56,434	133	9321,82	199,000
Honolulu, HI (MSA)	476,349	31,682	102	875,67	155,700
San Francisco, CA (PMSA)	1185,213	59,077	195	1731,716	151,400
San Jose, CA (PMSA)	1099,966	74,374	187	1683,908	139,700
Boston, MA	3410,183	44,395	132	6067,51	127,400
Philadelphia, PA-NJ (PMSA)	2534,87	38,648	114	5104,291	124,500
Newark, NJ (PMSA)	1048,783	47,651	136	2035,127	124,300
Hartford, CT (NECMA)	655,624	41,673	123	1150,619	123,000
Santa Rosa, CA (PMSA)	205,975	35,148	118	460,268	120,700
Washington, DC-MD-VA-WV (PMSA)	3009,72	45,129	136	4948,213	120,300
[...]					
Lincoln, NE (MSA)	158,191	28,389	98	251,008	90,800
Amarillo, TX (MSA)	103,147	26,357	83	218,321	90,30
Hattiesburg, MS (MSA)	54,465	23,366	72	112,105	90,000
Little Rock-North Little Rock, AR(MSA)	338,296	29,541	93	585,228	89,600
Anniston, AL (MSA)	52,462	24,764	72	111,35	89,50
Danville, VA (MSA)	49,441	24,65	71	110,05	89,500
Montgomery, AL (MSA)	178,367	28,245	87	333,479	89,400
Joplin, MO (MSA)	82,314	24,76	75	157,667	89,300
Clarksville-Hopkinsville, TN-KY (MSA)	99,761	26,296	76	207,613	88,900
Jonesboro, AR (MSA)	43,647	24,44	74	82,436	88,700
United States, total	**139552**	**34,652**	**100**	**282124,631**	**100**

MSA = metropolitan statistical area. PMSA = Primary metropolitan statistical area. NECMA= New England metropolitan statistical area.

This descriptive evidence supports the view outlined in the introduction that large agglomeration areas with high nominal wage and income levels tend to have higher costs-of-living than corresponding rural places. The model implication of Krugman (1991a) and F/K/V (1999) that core regions have lower COLI-levels is strongly rejected by the data.

A plausible hypothesis seems to be that the higher central COLI-levels are due to housing costs differentials. A casual empirical finding that supports this belief is that New York City reveals housing prices 325% above the US average, even

though the overall COLI is only 99% above average (American Chamber of Commerce).

Table D2: Correlation matrix (all 208 observations)

	COLI	POPULATION	JOBS	WAGE	INCOME
COLI	1	0.53 **	0.52 **	0.66 **	0.57 **
POPULATION		1	0.99 **	0.61 **	0.46 **
JOBS			1	0.65 **	0.52 **
WAGE				1	0.89 **
INCOME					1

** = significant at the 1 % level.

Note: For information about the definitions of MSAs and a full list of all areas used in this classification scheme consult http://www.bea.gov/bea/regional/docs/msalist.htm. For detailed information about the regional accounts data see http://www.bea.gov/bea/regional/reis/. More information about the COLI-data can be found under www.datamasters.com. The full list of all 208 metropolitan areas for which both COLI- and BEA-data exist is available upon request from the author.

All in all, we feel that the empirical relation described in this section is robust, plausible, and supposedly quite similar for the European Union, although we can not show this because of data limitations. Starting from this empirical motivation we now come to the model that aims to explain the endogenous emergence of c-p-structures in spirit of the traditional NEG-literature, but with the important difference that the core region in our model can reveal the higher regional COLI in the long-run equilibrium.

D7.2) The basic structure of the extended model

The economy under consideration is structurally very similar to the standard NEG-model, and we closely follow the notation of section D5). The representative consumer in each region still maximizes a Cobb-Douglas utility function

$$U_r = M_r^{\mu} H_r^{\gamma} A_r^{1-\mu-\gamma} \qquad \text{(D.19)}$$

where H_r are the units and γ is the expenditure share of the "new" non-tradable home good. The budget constraint now reads as

$$p_r^A A_r + p_r^H H_r + \left(n_r p_r m_{rr}(i) + n_s T\, p_s m_{sr}(i)\right) = Y_r \quad , \qquad \text{(D.20)}$$

where p_r^H is the price of the home good in region r. The demand for the home good is simply $H_r = \gamma Y_r / p_r^H$. The price for the agricultural good A is still the numeraire and in equal to unity in both regions. Demand for the agricultural good is

simply $A_r = (1-\mu-\gamma)Y_r$, and for the consumption aggregate $M_r = \mu Y_r / G_r$. The composite manufacturing price index G_r , as well as the demand and supply functions for the single manufacturing commodities are completely identical to the standard model, i.e. the respective equations are given by (D.4)-(D.10). We can directly turn to the equilibrium conditions, analogously to section D5.3).

Most equilibrium condition also remain unchanged with respect to section D5.3). The market for the agricultural good still clears automatically at the numeraire price $p^A=1$. The regional supplies of home goods are fixed at $H_1 = H_2 = \gamma/2$. In equilibrium, home goods prices must adjust such that supply equals demand for home goods, $H_r = \gamma Y_r / p_r^H$. For market clearing, home goods prices and wages must thus be directly proportional to the respective regional income.

$$p_r^H = w_r^H = 2Y_r \tag{D.21}$$

The equilibrium prices and wages in the manufacturing sector are still given by

$$p_r = w_r = [Y_r G_r^{\sigma-1} + Y_s G_s^{\sigma-1} T^{1-\sigma}]^{1/\sigma} \tag{D.22}$$

But regional income Y_r is now

$$Y_r = \lambda_r \mu w_r + \frac{\gamma}{2} w_r^H + \frac{1-\mu-\gamma}{2}. \tag{D.23}$$

The composite manufacturing price index remains unchanged. It is given by

$$G_r = [\lambda_r (w_r)^{1-\sigma} + (1-\lambda_r)(w_s T)^{1-\sigma}]^{\frac{1}{1-\sigma}} \tag{D.24}$$

Substituting (D.23) and (D.24) into (D.21) and (D.22), we can find the set of nominal wages w_r, w_r^H at which there is market clearing in all sectors and regions. These equilibrium values only depend on the parameters T, μ, γ, σ and on the endogenous variable λ_r. Of course an analytical solution is "even less" feasible in this three-sector case. But one can still single out some effects and provide the respective economic intuition.

As λ_r increases, more individuals compete for a given supply of home goods. Formally, we know from (D.21) that w_r^H increases with regional income Y_r. Y_r is clearly an increasing function of λ_r, since nominal wages w_r can never drop so sharply to overcompensate the effect stemming from the entry of more manufacturing workers. Thus, we find the plausible result that w_r^H (p_r^H) rises with λ_r.

The effects of an entry of workers on regional manufacturing wages have been extensively discussed above. To briefly repeat the intuition: an immigration of new manufacturing workers increases competition on the regional goods markets, which puts incipient downward pressure on nominal manufacturing wages. At the same time, there are linkage effects operating in the other direction. The only effect that is new compared to the standard model in section D5) is a demand linkage effect that applies to the home goods producers: since their income rises with λ_r, they will demand more manufacturing commodities and thus also perpetuate the upward tendency of nominal manufacturing wages.

Let us turn to real manufacturing wages ω_r. Contrary to the above model, the regional cost-of-living index is no longer identical with the composite manufacturing price index G_r, since now also the regional price level of home goods must be included. Let ψ_r denote the regional CPI. Real wages are then given by

$$\omega_r = \frac{w_r}{\psi_r} = \frac{w_r}{(G_r)^{\mu}(p_r^H)^{\gamma}} \tag{D.25}$$

Similar to the nominal wage, also the real wage ω_r endogenously only depends on λ_r. Considering real wages, there are two additional effects to consider. The first is the well known backward linkage from above. Since G_r is decreasing in λ_r, the establishment of new firms in the region implies an increase in real wages by lowering the living expenses for manufacturing commodities. However, this is not the end of the story here. As more manufacturing workers pool in one location, home goods prices are driven up and put real wages under strain. Thus, the overall impact of an increase in λ_r on real wages ω_r is ambiguous and heavily depends on the exogenous parameters.

In this extended model there are also only two possible structures that can prevail in the long run. The economy will either be characterised by a complete agglomeration of all manufacturing labour in one region, or the two regions are completely symmetrical. The analysis will proceed completely analogous to the standard model above. First we check sustainability of an initially existing core-periphery-structure. Afterwards we determine if symmetry breaking can occur around the steady state with $\lambda = ½$. Note that Krugman-model is entailed in our analysis as a special case where $\gamma = 0$. Therefore we will consider this benchmark case repeatedly.

D7.3) Sustainability and stability

When all manufacturing labour is pooled in region 1 ($\lambda=1$), the regional income levels in the core (Y_1) and in the periphery (Y_2) are

$$Y_1 = \mu w_1 + \frac{\gamma}{2} w_1^H + \frac{1-\mu-\gamma}{2}$$

$$Y_2 = \frac{\gamma}{2} w_2^H + \frac{1-\mu-\gamma}{2} \tag{D.26}$$

Since the expenditure on home goods γY_r in each region must in equilibrium equal the income from home goods $(\gamma/2)\, w_r^H$, we find that

$$p_1^H = w_1^H = \frac{2\mu}{1-\gamma} w_1 + \frac{1-\mu-\gamma}{1-\gamma}$$

$$p_2^H = w_2^H = \frac{1-\mu-\gamma}{1-\gamma} < 1. \tag{D.27}$$

Note that the equation(s) (D.27) can only be derived if γ is strictly positive, since otherwise there is no home goods sector and hence no home goods prices. To determine the manufacturing wage w_1 we look at total national income for the case with λ=1:

$$Y_1 + Y_2 = \mu w_1 + \frac{\gamma}{2}(w_1^H + w_2^H) + (1-\mu-\gamma) = w_1\left(\frac{\mu}{1-\gamma}\right) + \frac{1-\mu-\gamma}{1-\gamma}$$

Since a fraction μ of total income is spent on manufacturing, and this expenditure needs to equal manufacturing earnings μw_1, it is still the case that

$$w_1 = 1 \tag{D.28}$$

Using (D.28) in (D.24), we can see that regional manufacturing price indices in this situation are still given by G_1=1 and $G_2 = T > 1$ respectively. Thus, the *manufacturing price index* is indeed lower in the core than in the periphery, unlike the prices for home goods. Using (D.28) in (D.27), the closed form solution for home goods prices in the core is given by $p_1^H = w_1^H = (1+\mu-\gamma)/(1-\gamma) > 1$. This implies that the *aggregate costs-of-living* in the two regions are

$$\psi_1 = \left(\frac{1+\mu-\gamma}{1-\gamma}\right)^{\gamma}$$

and

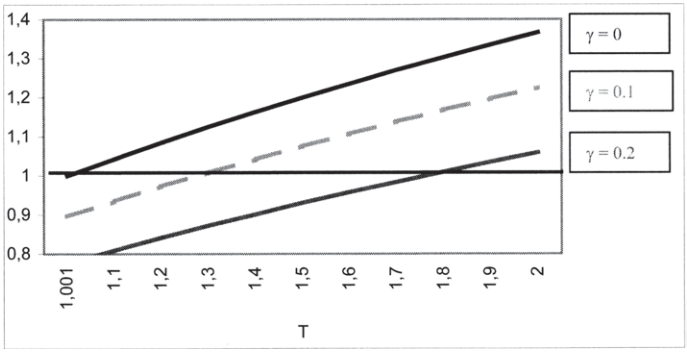
$$\psi_2 = T^{\mu}\left(\frac{1-\mu-\gamma}{1-\gamma}\right)^{\gamma}. \tag{D.29}$$

From (D.29) one can easily see that it is possible that the periphery has a lower CPI than the core in this completely agglomerated situation.

Proposition D2. *The overall CPI is lower in the periphery ($\psi_2 < \psi_1$) if transportation costs are below a critical level: $T < T^{CPI}$, with* $T^{CPI} = \left((1+\mu-\gamma)/(1-\mu-\gamma)\right)^{\gamma/\mu}$.

Figure D7) plots the relative CPI of region 2 as a function of T and γ. Overall living expenses in the periphery are lower if the curve ψ_2/ψ_1 runs below the thick horizontal line. The critical level T^{CPI} is at the crossing point.

Figure D7) Relative CPI ψ_2/ψ_1 as a function of T and γ
ρ=0,7; μ=0.5



Obviously, the relative CPI of the periphery is monotonously increasing with T, since it needs to pay more and more for its manufacturing imports. T^{CPI} is shifted to the right the higher is γ. The range of T for which living expenses are cheaper in the periphery is larger, the more important is the home goods sector in which it has a price advantage. Moreover, T^{CPI} is also increasing in μ. For a given level of γ, nominal manufacturing wages and home goods demand in the core rise rapidly in the core if the manufacturing share μ is increased. Home goods prices are driven up, and the periphery is the cheaper area for a larger parameter range of T. The parameter ρ on the other hand does not affect the relative CPI. Note that the critical level T^{CPI} is equal to one with γ=0. The condition $T < T^{CPI}$ can by definition not hold in this case, and therefore the core can never have higher costs-of-living than the periphery in the basic Krugman-model.
Using (D.28) and (D.29), the real manufacturing wage in the core region 1 can be computed as

$$\omega_1 = \left(\frac{1-\gamma}{1+\mu-\gamma} \right)^{\gamma} \tag{D.30}$$

We can furthermore derive, in a similar way as above, the *theoretical value* of the nominal wage w_2 from equation (D.22), and deflate this expression with the regional CPI ψ_2 from (D.29) to obtain the theoretical real wage ω_2 that one single manufacturing worker could earn if he decides to move from region 1 to region 2. The nominal wage w_2 is given by

$$w_2 = \left[\frac{1+\mu-\gamma}{2(1-\gamma)} T^{1-\sigma} + \frac{1-\mu-\gamma}{2(1-\gamma)} T^{\sigma-1} \right]^{1/\sigma}, \tag{D.31}$$

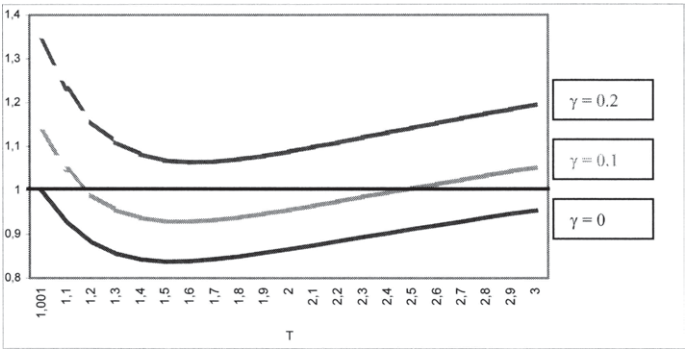
and the corresponding real wage is $\omega_2 = w_2 / \psi_2$. Using (D.29), (D.30), and (D.31) the relative real wage in the initial situation is given by

$$\frac{\omega_2}{\omega_1} = \left(\frac{1+\mu-\gamma}{1-\mu-\gamma} \right)^{\gamma} T^{-\mu} \left[\frac{1+\mu-\gamma}{2(1-\gamma)} T^{1-\sigma} + \frac{1-\mu-\gamma}{2(1-\gamma)} T^{\sigma-1} \right]^{\frac{1}{\sigma}} \tag{D.32}$$

The existing core-periphery-structure is sustainable if $\omega_2/\omega_1 < 1$. Figures D8) plots the real wage quotient (D.32) as a function of T for different constellations of exogenous parameters μ, γ and σ.

Figure D8) Real wage quotient ω_2/ω_1 as a function of γ

ρ=0.7 ; μ=0.5



As can be seen, the function ω_2/ω_1 has a "u-shape" with respect to T. For the case with $\gamma = 0$ in figure D8), the Krugman-model, sustainability is warranted for low and intermediate values of T. In cases with a home goods sector, the core-periphery-structure is sustainable only for intermediate levels of T. The reason why an asymmetric geographical structure can not be maintained with high transportation costs is well known from the standard model: it is too costly to satisfy the manufacturing demand of the immobile workers in the periphery through exports from the core. But in our extended approach, sustainability in our model also fades away for low values of T. The reason is that nominal manufacturing wages do not differ markedly in core and periphery if T is close to one, nor do manufacturing price indices. But home goods are significantly cheaper in region 2, leading to higher real wages ω_2 that make it worthwhile for a worker to leave the core.
However, in our model there can also be sustainable core-periphery-equilibria if transportation costs are on an intermediate level. The core can exploit the increasing returns in production and consequently pay higher nominal wages. Moreover, individuals in the core face a lower manufacturing price index G_1. In opposite to these forward and backward linkages there are the higher home goods prices in region 1. The greater is γ, i.e. the more value people put on home goods, the more pronounced is this centrifugal force. The u-shaped curve is shifted upwards in figure D8) as γ increases, and beyond some critical point γ^S the home goods sector is so important that a core-periphery-structure can never be sustainable. But if the home goods share is not too large, i.e. if $\gamma < \gamma^S$, sustainability of the c-p-equilibrium is possible.
The two other parameters affecting the equilibrium are μ and ρ (i.e. σ). Ceteris paribus, sustainability is more likely the lower is ρ, i.e. the more important are increasing returns in manufacturing, and the higher is μ, i.e. the more important is the manufacturing sector.
The important question is now, whether there can be sustainable c-p-structures in which the core reveals a higher overall price index ψ_1 than the periphery. In proposition D2 we have shown that ψ_2 is lower than ψ_1 if T is below some critical level T^{CPI}. Therefore, a c-p-structure must be sustainable for values of $T < T^{CPI}$ in order to match both desired properties $\omega_2/\omega_1 < 1$ and $\psi_2/\psi_1 < 1$.

Proposition D3. *c-p-equilibria can be sustainable with $\psi_2 < \psi_1$, if: i) the home goods share γ is on intermediate levels, but below γ^S; ii) transportation costs T are on an intermediate level, but below T^{CPI}. Such an equilibrium is more likely iii) the larger is μ, and iv) the smaller is ρ.*

Sustainability of the pre-existing c-p-structure requires intermediate levels of T and is more likely the lower is γ. On the other hand, the property $\psi_2 < \psi_1$ requires that T is low enough to fall short of T^{CPI}, which is more likely to hold the higher is γ. For the parameters T and γ there is thus a trade-off: Transportation costs have to

be substantial enough in order for the core to pay sufficiently higher wages, but they must not be higher than T^{CPI}. The home goods sector must be important enough, so that the periphery is in total the cheaper place. On the other hand, real wages in the core must not be put under too much strain from home goods prices, so that individuals are better off moving to the periphery. Such a trade-off does not exist for the parameters μ and ρ. A large value of μ contributes both to sustainability and lower overall prices in the periphery for reasons discussed above. In the same vein, ρ leaves the relative CPIs unchanged, but sustainability is strengthened the lower is ρ. Taken together, these comparative statics imply proposition D3.

The final thing to do is to show that a core-periphery structure with lower central CPI is consistent with the property of symmetry breaking. The analytical expression for dω/dλ as a function of exogenous variables can be obtained in an identical way as in section D5.5). Details of the complicated derivation are given in the appendix of this chapter.

$$\frac{d\omega}{d\lambda} = 2G^{-\mu}Z\left[\frac{(1-\mu Z-\frac{\gamma\mu}{1-\gamma})(\mu - Z(1-\gamma))}{\sigma(1-\gamma)-\mu Z - Z^2(\sigma-1)(1-\gamma)} - \frac{\mu}{Z}\left(\frac{Z}{1-\sigma}+\frac{\gamma}{1-\gamma}\right)\right]. \quad \text{(D.33)}$$

Figure D9) graphs (D.33) as a function of T and γ, for given values of ρ and μ. The symmetrical equilibrium is unstable if (D.33) is positive, i.e. if the curve runs above the thick horizontal line. This is the case for low and intermediate transportation costs if γ = 0, as known from above. If γ is positive but not too high, symmetry breaking occurs for intermediate values of T.[22] Analogously to sustainability, symmetry breaking is less likely the higher is γ.[23] In fact, if γ becomes too large, it turns out that a c-p-equilibrium can never emerge endogenously. However, the existence of a home goods sector per se does not rule out that a symmetrical equilibrium breaks up and a c-p-equilibrium occurs.[24]

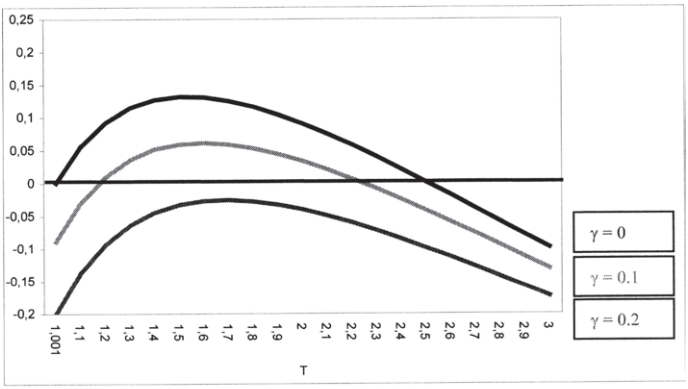
---

[22] At some critical value of T, the property of symmetry breaking will vanish. It can be shown that this critical level, the break point, is strictly lower than the sustain point, i.e. the level of T at which sustainability would vanish. This 'locking-in-effect' that gives rise to a tomahawk bifurcation and multiple equilibria is extensively discussed at other places in context with the seminal Krugman-model.

[23] Changes in ρ and μ produce qualitatively similar results as for sustainability: symmetry breaking is more likely the lower is ρ and the higher is μ.

[24] This is different from a model where agricultural transportation costs are added (see F/K/V, ch.7). It can be shown that under such a model it is also possible to establish sustainable c-p-equilibria with a lower CPI in the periphery. However, F/K/V, 103 show (albeit without any refer-

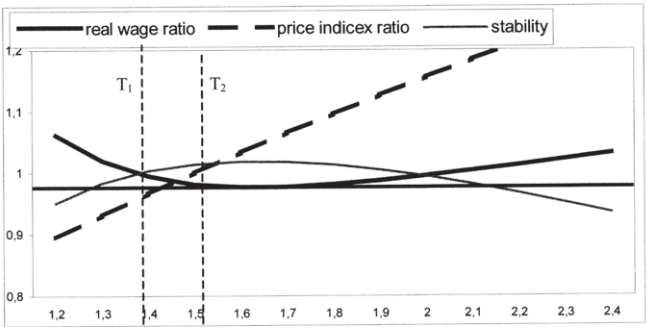
Figure D9) Stability of the symmetric equilibrium
$\rho = 0.7$; $\mu = 0.5$



Importantly, there is no contradiction between the property of symmetry breaking and the condition that the CPI is higher in the centre in the resulting c-p-equilibrium. In figures D10a)-D10c) we show for some examples that there are parameter constellations where the condition of symmetry breaking is satisfied for values of T below T^{CPI}. Therefore, the parameter ranges T_1-T^{CPI} indicate those levels of transportation costs where for given values of γ, μ and ρ an asymmetric c-p-structure with a lower CPI in the periphery emerges and persists.

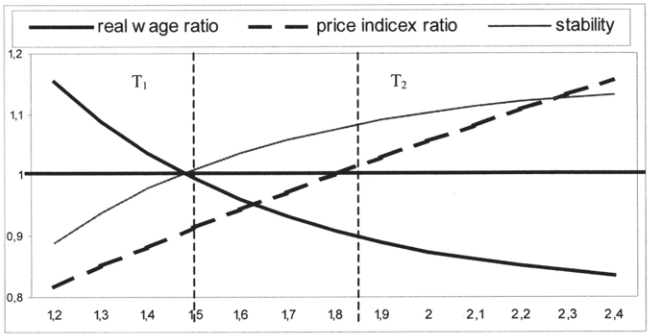
Figure 10) Different scenarios of symmetry breaking

a) $\rho = 0.75$; $\mu = 0.5$; $\gamma = 0.15$

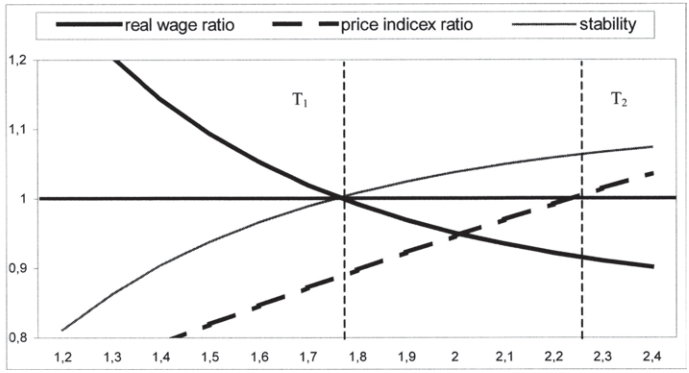


ence to the issue of regional CPIs) that with homogenous agricultural products symmetry will never break.

b) $\rho = 0.6$; $\mu = 0.5$; $\gamma = 0.2$



c) $\rho = 0.6$; $\mu = 0.5$; $\gamma = 0.25$

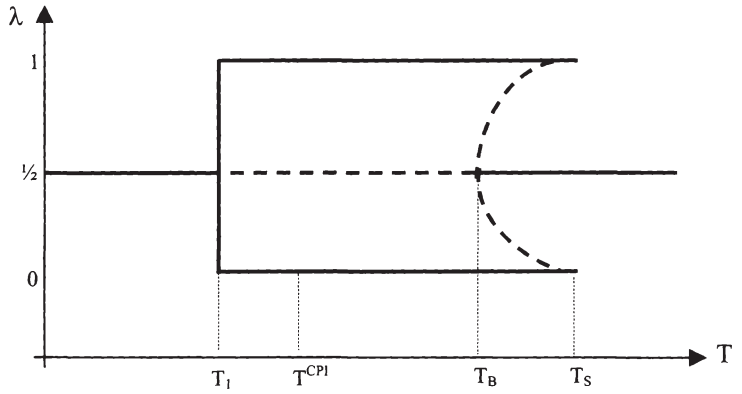


Again it is instructive to look at some comparative statics. Lower values of ρ increase the probability of symmetry breaking, but do not influence the relative price index locus ψ_2/ψ_1. Similarly, a higher manufacturing share μ increases both T^{CPI} and the chance of symmetry breaking. On the other hand, higher values of γ increase T^{CPI}, but they lower the feasible range of T where symmetry breaking can occur. Thus, the qualitative impacts of parameter changes as well as the underlying economic intuition are the same as described in proposition D3.

Lastly, we can graphically illustrate our results with a bifurcation diagram to get a clear picture in what respect our approach differs from those of F/K/V and Helpman (1998). This is done in figure D11). As argued above, our model predicts the emergence of a c-p-equilibrium only for intermediate levels of transportation costs, namely for the range of T between T_I and T_B. The level T^{CPI} lies in between the two. The tomahawk bifurcation at the transition from high to intermediate trade costs (between T_B and T_S) as well as the inevitably symmetrical equilibrium outcome for trade cost levels higher than T_S occur for the same reasons as in the

Krugman-model. But, in our model there is a critical level T_1 below which the c-p-structure can not be sustained and symmetry re-emerges. In this parameter range ($1 < T < T_1$), the agglomeration advantages are too low to compensate for the high prices in the home goods sector. This is essentially the same mechanism that also applies to the range of low trade costs in the Helpman-model. The final result, that agglomeration necessarily requires intermediate levels of transportation costs, is similar to findings from Venables (1996) or Puga (1999). But the underlying reasons in our model are quite different.

Figure D11) The bifurcation diagram of our NEG-model



D7.4) Conclusion of our model approach

With our modified approach, we show that c-p-structures can endogenously emerge and persist in spirit of the original NEG-model by Krugman (1991a). But the parameter range for which agglomeration can occur is lower: ceteris paribus, mobile manufacturing workers will only be completely concentrated in the core if transportation costs are on an intermediate level and the home goods share is not too important. We show that within this range of agglomeration equilibria, there are parameter constellations for which the core reveals a higher aggregate price level (CPI) than the periphery. Since the original Krugman-model is entailed in our analysis as a special case, the traditional result can also follow in our approach that actually the core is the low-price area. This result holds if the home goods share γ is small. Housing scarcity is then not a binding force for regional agglomeration.

However, our analysis is more general than the basic NEG-model, and it leaves room for the possibility that c-p-structures emerge and persist in which the core

exhibits higher costs-of-living than the periphery. As argued above, this case is empirically more relevant. In the US, it tends to be the case that local agglomeration and local COLI-levels are positively correlated. This is an empirical finding, supposedly robust to the European Union, that the standard NEG-model is unable to account for. Thus, we feel that our extended approach is somehow more complete than the model of Krugman (1991a), since it shows explicitly that agglomeration areas can pay a real wage premium despite of having a higher COLI-level. Let us now turn to the second deficiency of the current NEG-literature that we have spelled out in section D6.6.): the neglect of unemployment.

Appendix D: Derivation of equation (D.33)

For the derivation of dω/dλ we know that around symmetric equilibrium the following values of endogenous variables hold: λ= ½ , w = 1, w^H=p^H=1, Y= ½ , $G^{1-\sigma} = (1+T^{1-\sigma})/2$. We further know that the real wage can be written as

$$\omega = wG^{-\mu}\ p_H^{-\gamma} \tag{A1}$$

Total differentiation yields

$$\frac{G^{\mu} p_H^{\gamma}}{w} d\omega = \frac{dw}{w} - \mu \frac{dG}{G} - \gamma \frac{dp_H}{p_H}. \tag{A2}$$

Hence, we can write

$$G^{\mu} d\omega = dw - \mu \frac{dG}{G} - \gamma\, dp_H \tag{A3}$$

Now we express the endogenous variables on the right hand side of (A3) in terms of dω and dλ, which is done by totally differentiating equations (D.21) - (D.24). This yields

$$dp^H = 2\, dY \tag{A4}$$

$$dw = \frac{2Z}{\sigma} dY + \frac{(\sigma-1)Z}{\sigma} \frac{dG}{G} \tag{A5}$$

$$dY = \frac{\mu}{2(1-\gamma)} dw + \frac{\mu}{(1-\gamma)} d\lambda \tag{A6}$$

and

$$\frac{dG}{G} = \frac{2Z}{1-\sigma} d\lambda + Z dw \tag{A7}$$

where $Z \equiv \dfrac{\left(1-T^{1-\sigma}\right)}{(1+T^{1-\sigma})} = \dfrac{\left(1-T^{1-\sigma}\right)}{2G^{1-\sigma}}$

The term Z is defined for notational convenience as an index of trade barriers which takes the value of 0 when there are no transportation costs (T=1) and the value 1 if T→ ∞. After several lengthy but straightforward substitutions of (A4)-(A7) into (A3) we reach the messy equation (D.33).

Chapter E)
Regional agglomeration and regional unemployment

E1) Introduction

During the last two chapters we have essentially laid out the theoretical foundations necessary for the analysis of regional unemployment in relation to regional agglomeration. In chapter C we have introduced the wage curve literature as a regional labour market theory. But we have argued that the product market structure in these models is problematic from a theoretical point of view, because regional disparities can only exist if they are exogenously assumed. Regional asymmetries can not develop endogenously.

In chapter D we have therefore introduced theoretical approaches that aim to explain the endogenous emergence of regional agglomeration and core-periphery patterns of economic activity. This literature descended from the insights of Marshall (1890), and is represented in modern economic theory through the models of NTT and NEG. However, this strand of theory abstracts from unemployment and focuses on the explanation of regional disparities in production and income. It can thus not be used to analyse spatial unemployment differences.

The aim of this chapter is to combine elements from both strings in the literature and propose a theoretical model that allows to analyse regional unemployment and regional agglomeration in a unified framework. There is very little that has been done in this "boarder region" of different fields within economic theory. Still, very little is more than nothing. The existing work on the relation of trade and location theory with unemployment theory will be discussed in section E2. We then turn to the introduction of our model approach in section E3. In section E4 we critically review our approach from a theoretical point of view, and against the available empirical evidence from the EU-15. Section E5 concludes and raises some further issues for the upcoming chapter.

E2) The existing literature

Our theoretical analysis is related closest with the NTT-model of Matusz (1996). Another contributions that is worth mentioning is the paper of Peeters/Garretsen (2000).

E2.1.) The model of Matusz (1996)

Matusz builds a NTT-model in spirit of Ethier (1982), where final goods are assembled under the use of intermediate inputs. The production of intermediates involves internal increasing returns to scale, whereas a greater variety of industrial inputs induces a pecuniary externality on the unit costs in the final goods sector. Matusz works with a two-country model without labour mobility. In autarky, each country can assemble final goods only with local intermediates. The more inter-

mediates can be produced in a nation, the more efficient is the final goods production and (via a zero profit condition) the higher is the real wage consistent with product market equilibrium. Unemployment enters in his model through a shirking approach and yields a labour market equilibrium condition that is analogous to a wage curve.

Matusz moves from an autarky setting to an analysis of international trade, but he does so in a quite extreme way: he assumes that once trade barriers fall, there is costless trade in both the final good and the intermediates sector across countries. The introduction of trade widens the scope of available intermediates and thereby reduces unit costs in the final goods sector for both countries. Since firms make zero profits, this higher efficiency is absorbed by higher wages and lower unemployment rates that necessarily must be equalized across countries. Since trade is costless, no element exists that could discriminate between the factor income in the different spatial units. The final result of Matusz (1996) is thus that international trade increases wages and (via the shape of the wage curve) must reduce unemployment.

The model of Matusz is discussed here, because we will use elements of his paper also for our approach. But it has already become clear that substantial changes are necessary in order to make the model structure applicable for regional issues. This firstly concerns the fact that there are no regional disparities in wages and unemployment rates across nations (regions). This artefact stems from Matusz's assumption of perfect and costless trade. But since we are precisely interested in the nature of regional disparities, we generalize his model and introduce transportation costs. And secondly, a regional analysis must also derive implications with respect to labour mobility.

With the appropriate model extensions and modifications, our generalized version of the model comes much closer in spirit to the theories of regional agglomeration and NEG, and it provides then a useful framework to study the interrelation of agglomeration and unemployment on a regional basis.

E2.2.) The model of Peeters/Garretsen (2000)

The contribution from Peeters/Garretsen (2000) is noteworthy, since it is to our knowledge the only existing attempt where unemployment is integrated into a standard and full fledged NEG-model.

The research focus of Peeters/Garretsen is the impact of globalisation (falling transportation costs) on the labour market position of unskilled workers in advanced countries. They mean by that the skilled/unskilled-wage gap that might change as globalisation proceeds. But they also analyse the impact on unemployment if unskilled labour is subject to an exogenously given minimum wage. They formulate a NEG model with two countries (labelled as *home* and *foreign*), mobile skilled and immobile unskilled labour. They adopt a more complicated production setting compared to the standard Krugman-model. They consider a two-step pro-

duction process of manufacturing goods. Firstly, there are so called "unfinished" goods that are produced by unskilled labour only. To these raw goods, producer services of high skilled workers must be added in order to make the goods marketable. Both unfinished goods and services exhibit scale economies in the production, and both are tradable across regions subject to transportation costs. The standard agricultural sector with constant returns that employs low skilled labour only is also part of their model. In principle, they derive standard NEG-implications with respect to the impact of globalisation: the effect of falling transportation costs on the labour market position of low-skilled workers is ambiguous and depends on the range of transportation costs. They are able to make some further qualifications, as they can distinguish what type of transportation costs declines (for raw goods, for producer services, both). But we want to focus on their implications with respect to unskilled unemployment.

They consider different minimum wage regimes for the low-skilled workers in the manufacturing sector of the home country. All other labour markets, including that for low-skilled manufacturing workers in the foreign country, must clear for the analysis to be internally consistent. The prevalence of a minimum wage induces unemployment for the unskilled in the manufacturing sector of the home country. In view of falling transportation costs, parts of the globalisation impacts on low skilled workers do not occur through changes in the relative skilled/unskilled wage, but rather through changes in the unemployment exposure of the low skilled workers. Of course these conclusions can not be obtained as closed-form analytical solutions, but Peeters/Garretsen (2000) have to rely on numerical solution techniques.

For the purpose to analyse regional unemployment disparities within the EU-15, the approach of Peeters/Garretsen (2000) is hardly applicable. Recall that there will be unemployment only in the home country, whereas all foreign labour markets clear. For an analysis of spatial disparities, there should be unemployment in both regions. Furthermore, Peeters/Garretsen rely on institutional differences, since the minimum wage legislation can exist only in one country, but not in the other. For our purposes, however, we would like to see regions with identical labour market institutions in order to analyse if disparities in unemployment rates can still arise. Furthermore, Peeters/Garretsen do not incorporate a wage curve, or a labour market equilibrium relation between wages and unemployment, but rather impose a specific minimum wage. Hence, the model of Peeters/Garretsen, that should really be seen as pioneering work, can serve for us as a related piece of literature. But it is designed to address a totally different economic problem, not to analyse regional unemployment disparities. There is thus still enough room for us to formulate an own theoretical approach.

E3) Regional agglomeration and the wage curve: The model

The exposition of our model goes in three steps: first we present a closed-economy setting where there is neither trade in intermediate inputs, nor factor mobility. This setting is identical to the autarky model of Matusz (1996). We then allow for trade in intermediate inputs. But we generalize the Matusz-model to account explicitly for geographical factors by assuming the presence of iceberg costs for interregional transportation of intermediate inputs. Yet, we keep at first the assumption of immobile workers. The last step is to relax also this assumption and analyse the impacts of (imperfect) labour mobility.

E3.1.) The closed economy setting

Consider an economy that produces (without using labour) a final consumption good Y under the use of a large variety of N single intermediate inputs X_i. The production function of the final product Y is given by the symmetrical CES function

$$Y = \left(\sum_{i=1}^{N} X_i^{\theta} \right)^{1/\theta} \qquad 0 < \theta < 1. \qquad \text{(E.1)}$$

The parameter θ is a measure of the differentiability of single intermediate inputs. If θ is close to one, inputs are nearly perfect substitutes. The elasticity of substitution between the single intermediates is given by $\sigma = 1/(1-\theta)$.

The minimum cost function of producing Y can be obtained by minimizing total consumption expenditure subject to (E.1). This yields

$$C(p_1, \ldots, p_N, Y) = GY \qquad \text{where } G = \left(\sum_{i=1}^{N} p_i^{\theta/(\theta-1)} \right)^{\frac{\theta-1}{\theta}} \qquad \text{(E.2)}$$

The term G in (E.2) can be understood as a minimum unit cost function for the final good Y. We apply the standard assumption of NTT that all intermediate inputs enter symmetrically into the production function. This implies that the production function simplifies to $Y = \left(N(X_i)^{\theta} \right)^{1/\theta}$, and furthermore implies that the minimum unit cost function becomes

$$G = \left(N(p)^{\theta/(\theta-1)} \right)^{(\theta-1)/\theta} \qquad \text{(E.3)}$$

where p is simply the price of one of the symmetrical intermediates produced in that economy. As can be seen, the minimum unit costs decrease with N. Accord-

ing to Ethier (1982) and Matusz (1996), this intends to capture the famous "pin factory"-idea of Adam Smith. It is an advancement for an economy to have a deeper division of labour, i.e. more narrowly defined sub-steps in which a specific production task (Y) is partitioned.
An important assumption made by Matusz (1996), that will also be used here, is perfect competition in the Y sector, which ensures that profits must always be equal to zero. We furthermore use the price of the final good p^Y as the numeraire and normalize it to one. By the condition $\pi^Y = 1$. $Y - G.Y = 0$, we can easily derive the first product market equilibrium condition of the model. It is given by the requirement that minimum unit costs need to equal the product price

$$G = \left(N(p)^{\theta/(\theta-1)}\right)^{(\theta-1)/\theta} = 1.$$

Since Y is the only commodity in this economy that is directly consumed, also the consumer price index is given by the value $p^Y=1$ at any time. This assumption is thus simplifying in another respect, as it allows us to abstract from the distinction between nominal and real wages. In equilibrium, prices of the intermediates will have to adjust such that this condition is satisfied.
Each of the N single intermediates X_i is produced by using labour only. The production function in the X-sector is virtually identical to that described for the manufacturing sector in section D5.2.). There are N single firms, each producing one distinct (but symmetrical) intermediate under increasing returns to scale and within a monopolistically competitive market. The labour requirement ℓ_i necessary to produce the quantity X_i is given by

$$\ell_i = \alpha + \beta X_i \qquad with\ \alpha>0,\ \beta>0$$

Each firm sets prices as a constant mark-up over marginal costs[1].

$$p = \frac{\beta}{\theta} w \tag{E.4}$$

Despite of the assumption of monopolistic competition, profits for every single intermediate good are driven down to zero by the entry of potential competitors. This again implies that all X-firms operate at a unique scale of output, and employ a well defined number of workers, respectively given by

[1] In the formulation with σ, which is also the perceived elasticity of demand, the pricing rule takes the form $p(1-1/\sigma) = \beta w$. This form might look more familiar.

$$X = \left(\frac{\alpha}{\beta}\right)\left(\frac{\theta}{1-\theta}\right)$$

and

$$\ell = \frac{\alpha}{1-\theta} \tag{E.5}$$

a) The model with full employment

The determination of the equilibrium price and the equilibrium number of intermediate inputs can most easily been done for the case with full employment. We know that the equilibrium number of workers that each firm employs is given by (E.5). Let the size of the total labour force be given by L. The equilibrium number of firms N is then simply given by

$$N = \frac{L}{\ell} = \frac{L(1-\theta)}{\alpha}. \tag{E.6}$$

This condition states that the equilibrium number of firms is higher, the larger is the local labour force L. But recall that the number of firms N will also affect the production costs in the Y-sector. We have already established the equilibrium condition to be $G = N^{(\theta-1)/\theta} p = 1$. Using (E.4) and (E.6) in this equilibrium condition, we can derive the equilibrium wage per worker (w) as a function of N:

$$w = \frac{\theta}{\beta}\left(\frac{(1-\theta)}{\alpha} L\right)^{\frac{1-\theta}{\theta}}. \tag{E.7}$$

The individual wage is higher, the larger is the local labour force and the higher is the equilibrium number of firms N. What is the intuition of this result? We have seen that production costs for the final product decrease in the number of available intermediate inputs. But the price p^Y is the numeraire and always equal to one. Suppose N increases. On instance, there will be positive profits in the Y sector, because costs have decreased at constant sales prices. With perfect competition, new entrepreneurs will enter the market for Y production and compete profits down to zero. This must be done by paying more for intermediate products. By (E.4) and the assumption of zero profits in the X sector, these price increases will be absorbed by higher wages. In other words: If more intermediates can be produced, the increasing returns to scale can better be exploited. This will lead to higher wages.

b) The model with unemployment

Matusz (1996) now adds unemployment to the closed-economy model through a standard efficiency wage approach in the vein of Shapiro/Stiglitz (1984). His set-up is in principal the same as described in section C3.1a). We shall therefore present the essence of this model here only very briefly. We stick to the notation of chapter C and apply our version of a shirking model rather than the set-up of Matusz (1996).

Workers living in the closed economy are risk-neutral and derive utility from wage income w, but disutility from work-effort e, which is simply a technologically fixed number. The work utility is

$$V = w - e. \tag{E.8}$$

"Shirking" individuals spend zero effort (e=0), but run a risk ($1-\gamma_r$) of being detected and then fired. By using the unemployment rate $U = 1 - N\ell/L$, we can derive the utility levels of an unemployed individual (V_u), a non-shirking employed worker (V_{en}) and a shirker (V_{es}) respectively.

$$V_u = \alpha(w - e)$$
$$V_{en} = w - e$$
$$V_{es} = \gamma w + (1-\gamma)(\alpha(w - e)),$$

where α is again the outflow probability from unemployment that negatively depends on the unemployment rate U (see section C3.5.). The firms pay efficiency wages such that the "non-shirking condition" $V_{es} = V_{en}$ holds. The regional wage curve, or aggregate non-shirking condition, is given by the following expression

$$w = e + \frac{\gamma e}{(1-\gamma)(1-\alpha(U))} \tag{E.9}$$

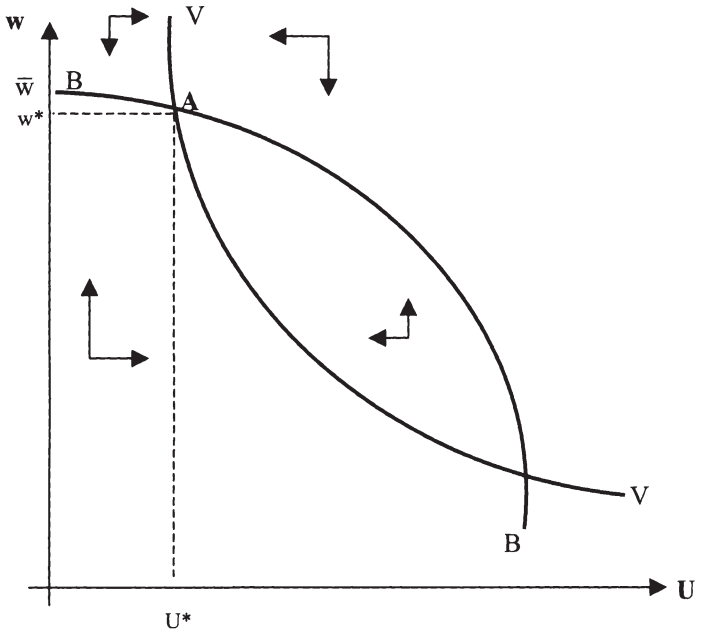
Of course the required efficiency wage is again lower the higher is the regional unemployment rate U. Equation (E.9) is the labour market equilibrium curve, and can be viewed at as the "first half" of full general equilibrium.

The product market equilibrium condition has also been derived above. It only needs to be slightly modified if we allow for unemployment. The maximum number of varieties N that can be produced in this economy is no longer given by labour supply. Since labour supply and employment can now differ, it is by definition now given by the latter. The equilibrium condition (E.7) becomes

$$w = \frac{\theta}{\beta}\left(\frac{1-\theta}{\alpha}(1-U)L\right)^{\frac{1-\theta}{\theta}} \tag{E.10}$$

The two equilibrium relations (E.9) and (E.10) can be illustrated within the same graph, given in figure E1.

Figure E1: Equilibrium in the closed economy



The locus VV represents the familiar wage curve locus. It is the graphical illustration of all combinations of wages and unemployment rates where shirking is just prevented for workers in the intermediate good sector. For all points to the right of VV, unemployment is too high for any given wage. Consequently firms can hire new workers and trust that they do not shirk. Hence, equilibrium unemployment must fall. This determines the phase arrows in the horizontal direction. The locus BB represents all combinations of w and U where all goods markets (Y and all X) are in equilibrium: there are no profits for either firm, there is cost minimization in the Y-sector and profit maximization for the X-sector. The curve is downward-sloping and convex, because of the involved increasing returns in the technology.

For all points below BB, wages are too low for a given unemployment rate. Profits arise in the Y-sector than attract new producers. Labour demand rises and consequently wages have to go up.
Full equilibrium in this closed economy is given at an intersection of VV and BB. The phase arrows indicate that only one equilibrium is stable. This one is at point A, with equilibrium unemployment U* and equilibrium wages w*. Note that the equilibrium unemployment is strictly involuntary. Workers would in principle be willing to work at going wages w. But employers know that further recruiting would lead the incumbent workers to shirk at going wages. Therefore they do not hire additional workers. Put differently, the combination with full employment (U=0) and wages equal to $\overline{w}$ might be desirable, but it is not feasible, since rational individuals would respond to this constellation with shirking behaviour.
Let us briefly look at some comparative statics. No parameter change affects both curves. The wage curve VV shifts with the disutility of effort e, or the shirking detection rate γ. As argued in chapter C, these parameters might reflect structural characteristics of the respective labour market and could (through appropriate model extensions) be explicitly modelled as contingent on public policy, like e.g. the welfare state regime or employment laws.
The goods market curve BB shifts out to the right top as α, β, or θ decrease. Lower values of α or β reflect lower production costs at given output prices, i.e. an increase in profits. Due to perfect competition, this must lead to higher real wages stemming from the product market equilibrium condition. A decrease in θ reflects a higher degree of increasing returns to scale. By a similar reasoning this advantage for the production sector must be absorbed by higher equilibrium wages.

c) The introduction of perfect trade

The final reason why the BB-locus could shift out to the top right is an increase in the exogenously given size of the labour force L. Since the VV-curve is not affected by this parameter change, the resulting new equilibrium would be one with higher equilibrium wages w and a lower equilibrium unemployment rate U. The intuition is straightforward. An increase in L implies that it becomes feasible for the economy to produce a larger number of intermediate goods, which leads to a decline in production costs for one unit of Y. This intuition is also at the core of the analysis of Matusz (1996), who moves from a closed-economy setting to a two-country case with perfect (=costless) trade in final goods and intermediate products.
Suppose that the closed economy described in this section now starts to trade with a structurally identical foreign country. The final good Y as well as all intermediates X can be freely shipped across space, and all intermediates will be used symmetrically in the Y-production of either country. In such a setting, the home country can draw on imported inputs from the foreign country as if the intermedi-

ates were produces locally. In other words, there is no substantial difference between the introduction of trade and an increase in the size of the domestic labour force in the Matusz-model. The unit costs for the home country now read as

$$G = \left(N(p)^{\theta/(\theta-1)} + N^*(p^*)^{\theta/(\theta-1)} \right)^{(\theta-1)/\theta} = 1 \quad , \tag{E.11}$$

where an asterisk denotes variables of the foreign country. An analogous equation like (E.11) applies to the foreign country. Since there is free tradability and perfect competition in the Y-sector, the equilibrium condition is still given by the requirement that unit costs G (and G*) need to equal the numeraire price $p^Y=1$. Due to free trade in the intermediates sector, there must also be price equalization for the symmetrical industrial inputs ($p = p^*$). The equilibrium condition for the product market in both countries is thus given by

$$w = w^* = \frac{\theta}{\beta}(N + N^*)^{(1-\theta)/\theta} \tag{E.12}$$

If both countries face structurally identical labour markets, i.e. the same wage curve locus, (E.12) inevitably implies that there must be equalization of real wages and unemployment rates across the two countries. The real wage is higher and the unemployment rate is lower in either country than it has been the case in autarky. The basic implication of the model is thus that the introduction of free trade per see has positive welfare effects by increasing wages and lowering unemployment.

.

E3.2.) The two-region case with imperfect trade

As argued above, this basic version of the Matusz-model is ill-equipped to study regional disparities and agglomeration We will therefore extend and generalize the model of Matusz (1996) in this section. In chapter D we have shown that there are basically three core elements of the regional agglomeration theories in the vein of NEG. The first, increasing returns to scale, is already a feature of the model approach by Matusz. The second, transportation costs, will now be introduced. It will open up the possibility of regional disparities, since now one can explicitly discriminate the economic variables of the different spatial units. The third core element, the presence of an endogenous mechanism that pushes for spatial concentration, will be introduced in the next section through (imperfect) labour mobility.

a) The product market

We now consider a nation consisting of two regions r and s with identical technology and preferences (if preferred, one can use the original terminology of Matusz with a home and a foreign country). Both regions produce the final consumption good Y under the use of symmetrical intermediate inputs which are partly manufactured in the respective region itself (X_{rr}), and partly imported from the other region (X_{sr}). Transportation of intermediates across regions now imposes 'iceberg' costs $\tau>1$. For each unit X_{sr} dispatched, only $1/\tau$ units arrive. The standard CES production function for region r is now given by

$$Y_r = \left(N_r X_{rr}^{\ \theta} + N_s \left[\frac{X_{sr}}{\tau} \right]^{\theta} \right)^{1/\theta} \qquad \text{(E.13)}$$

The minimum unit cost function G_r becomes

$$G_r = \left(N_r (p_r)^{\frac{\theta}{\theta-1}} + N_s (\tau p_s)^{\frac{\theta}{\theta-1}} \right)^{\frac{\theta-1}{\theta}}, \qquad \text{(E.14)}$$

and the demand functions for intermediates X_{rr} and X_{sr} can be written as

$$X_{rr} = p_r^{\frac{1}{\theta-1}} G_r^{\frac{\theta}{1-\theta}} Y_r \qquad \text{and} \qquad X_{sr} = (\tau p_s)^{\frac{1}{\theta-1}} G_r^{\frac{\theta}{1-\theta}} Y_r . \qquad \text{(E.15)}$$

where p_r is the mill price of an symmetrical intermediate from region r. G_r is decreasing in both N_r and N_s: A larger array of intermediate inputs reduces costs in the Y-sector in both regions. But due to the transportation costs, the decline is larger in the region where the increase in the number of varieties has occurred. Put differently, an increase in N_r has a stronger negative effect on G_r than on G_s.

With respect to the final goods sector, we maintain the assumptions from the last section. The Y-sector is perfectly competitive, and the final consumption good can be freely traded across space. This implies that there is price equalization on the market for the final consumption good. Y-producers in both regions have to take the price p^Y as given, potentially as determined on world markets outside the nation. Without loss of generality we keep $p^Y=1$ as the numeraire. At first sight the assumption of perfect tradability in the consumption goods sector, but imperfect tradability in the sector of industrial inputs might seem peculiar. But apart from offering analytical convenience, its use can be justified on intuitive grounds. Production in the Y-sector is just the assembly of intermediates without use of labour.

Therefore, the function Y_r simply reflects with what level of technological sophistication the region r can produce a given final output bundle. The advantage of sophisticated (large) regions is the high availability of locally produced intermediates, whereas small peripheral regions have higher costs of producing the same level of output.
Zero profits and efficient production in the Y-sector imply that unit costs need to equal one in both regions. Using (E.14), the respective equilibrium condition is

$$1 = N_r (p_r)^{\frac{\theta}{\theta-1}} + N_s (\tau p_s)^{\frac{\theta}{\theta-1}} \qquad \text{for region r}$$

and

$$1 = N_r (\tau p_r)^{\frac{\theta}{\theta-1}} + N_s (p_s)^{\frac{\theta}{\theta-1}} \qquad \text{for region s.} \qquad \text{(E.16)}$$

The production of the single intermediates is unchanged compared to the closed-economy case. Therefore, the number of locally produced intermediates is restricted by regional employment:

$$N_r = (1-U_r)\frac{L_r}{\ell}$$

and

$$N_s = (1-U_s)\frac{L_s}{\ell} \qquad \text{(E.17)}$$

where L_r and L_s denote the exogenously given sizes of the regional labour forces. The regional wage levels w_r and w_s are determined in the same way as in (E.4), i.e. prices are constant mark-ups β/θ over marginal costs.
Contrary to the case with free tradability of intermediates, there is no longer price and wage equalization in the X-sector if the two regions differ in size and $\tau > 1$. If one region is larger than the other one, it has an advantage since it can produce more intermediates locally. We substitute (E.17) into (E.16) and set $\beta=\theta$ for notational simplicity. We obtain

$$w_r = \left(\frac{(1-U_r)\frac{L_r}{\ell}}{1-(1-U_s)\frac{L_s}{\ell}(\tau w_s)^{\frac{\theta}{\theta-1}}} \right)^{\frac{1-\theta}{\theta}} \qquad \text{(E.18)}$$

The nominal (=real) regional wage w_r derived from the condition for product market equilibrium is increasing in employment in both regions, but decreases

with higher wages in region s and tranportation costs τ. An analogous equation applies to region s. Solving for w_r and w_s, we can obtain closed-form solutions for the regional equilibrium wages

$$w_r = \left(T\,(1-U_r)\frac{L_r}{\ell} \right)^{\frac{1-\theta}{\theta}},$$

$$w_s = \left(T\,(1-U_s)\frac{L_s}{\ell} \right)^{\frac{1-\theta}{\theta}} \qquad \text{(E.19)}$$

where $T = \dfrac{1-\tau^{\frac{2\theta}{\theta-1}}}{1-\tau^{\frac{\theta}{\theta-1}}} \geq 1$

At these wage levels, there is profit maximization and zero profits in both sectors and both regions. As can be seen, w_r and w_s only depend (positively) on employment in the respective region itself. An increase in L_s e.g. in equilibrium has only positive effects on the wage in region s, but not on the wage of region r. The analytical reason is the symmetrical use of all intermediates in both regions. It can be seen in (E.18): an increase in L_s lowers the denominator and ceteris paribus has positive spillover effects in region r. But once the endogenous effect on w_s is taken into account, the impact will cancel out. Economically, equation (E.19) implies that the model incorporates a purely regional scale externality. Despite the openness, there are no interregional effects of an increase e.g. in the regional labour force L_r. But within region r, a larger labour force induces positive pecuniary spillovers on all workers in the form of higher equilibrium wages.

The variable T can be understood as an inverse measure of the resource waste from shipping. If trade is prohibitively costly ($\tau \to \infty$), T takes on the value T=1 and (E.19) turns into the autarkic expression (E.10). The closer the economy comes to free trade ($\tau \to 1$), the value of T approaches T=2.

The regional wages in (E.19) are consistent with efficient production, but we still have to show that the consumption good market and the market for intermediate products clear. Only then will the wages be the true equilibrium wages in this economy.

Proposition E1: *Regional wages in (E.19) imply not only efficient production, but also market clearing in both sectors and regions. They are the true equilibrium values.*

Proof: Each firm in the intermediates-sector supplies an regionally invariant quantity given by $X=\alpha/(1-\theta)$ (Equation (E.5) with $\beta=\theta$). Equilibrium requires that X

equals total sales to both regions $X_{rr}+ \tau X_{rs}$. Using (E.15) together with the equilibrium condition $G_r=1$ we can write this as

$$X = (w_r)^{\frac{1}{\theta-1}} \left(Y_r + \tau^{\frac{\theta}{\theta-1}} Y_s \right) \tag{E.20}$$

Using (E.19) and solving for Y_r yields

$$Y_r = \frac{X}{1-\tau^{\frac{2\theta}{\theta-1}}} \left((TN_r)^{\frac{1}{\theta}} - \tau^{\frac{\theta}{\theta-1}} (TN_s)^{\frac{1}{\theta}} \right) \tag{E.21}$$

Equation (E.21) determines the regional production level Y_r at which markets for intermediates clear. The total national production of Y is

$$Y_r + Y_s = \frac{X}{T} \left((TN_r)^{\frac{1}{\theta}} + (TN_s)^{\frac{1}{\theta}} \right) \tag{E.22}$$

Since Y is freely tradable at $p^Y=1$, (E.22) needs to equal total national income and consumption expenditure, which is given by

$$w_r(1-U_r)L_r + w_s(1-U_s)L_s \tag{E.23}$$

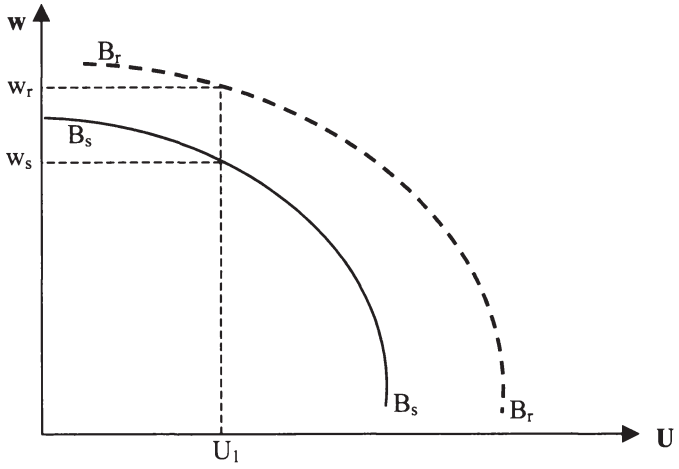
By using (E.19) in (E.23) and rewriting (E.22), one can show that both expression are equivalent to

$$\left(\frac{T}{X} \right)^{\frac{1-\theta}{\theta}} \left[\left((1-U_r)L_r \right)^{1/\theta} + \left((1-U_s)L_s \right)^{1/\theta} \right], \tag{E.24}$$

which proofs proposition E1. Markets for Y and all intermediates X clear with wages and prices given by (E.19). Equation (E.24) is the gross national product of this two-region economy: it is increasing in employment, and decreasing in transportation costs.

Let us consider the intuition of the expression (E.19). It is telling that for identical unemployment rates $U_r = U_s = U_1$, the region with the larger labour force pays the higher wage. Graphically, this can be seen in figure E2). The two downward sloping curves B_rB_r and B_sB_s illustrate those combinations of unemployment rates and wages consistent with product market equilibrium in both regions. The labour force in region r is assumed to be larger than in region s.

Figure E2: Product market equilibrium in both regions



The intuition for this scale effect is straightforward. At any given unemployment rate, region r produces more local intermediate goods than region s. Since all intermediates from both regions are used symmetrically in the Y-production, the larger region saves on transportation costs. It must consequently pay higher wages for the zero profit conditions in the Y- and the X-sector to hold. The higher is the difference between L_r and L_s, the further apart are B_rB_r and B_sB_s. An increase in τ shifts both curves downwards and to the left, because the dead weight loss of resources wasted in transportation increases. The same shift occurs upon an increase in α, β, or θ.

b) The labour market

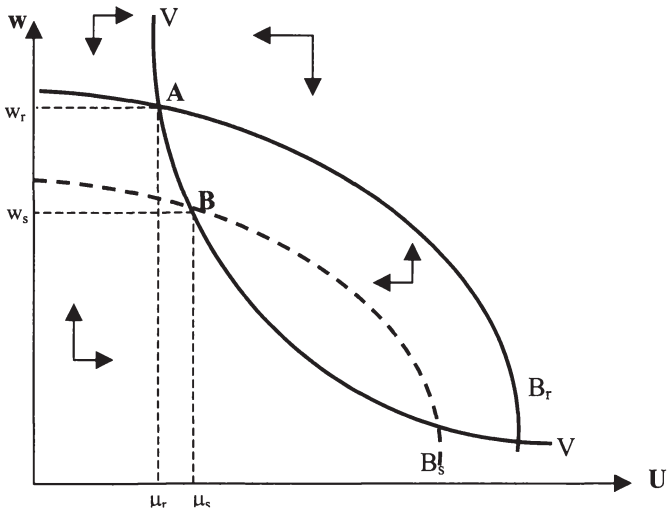
We have explicitly described the product market in this economy and derived a negatively sloped curve for each region that represents the equilibrium relation of wages and unemployment consistent with efficient production and market clearing. To close the model, the regional unemployment rates U_r and U_s need to be determined. This will of course be a matter of the wage curve that constitutes the second equilibrium relation in this model.

The derivation of the wage curve remains unchanged compared to the closed economy setting. The wage curve in both regions r and s is thus given by (E.9). There are no structural differences between the labour market in region r and region s. Both regions face the same downward sloping labour market equilibrium curve in the (w,U)-space that already showed up as the VV-locus in figure E1.

c) Equilibrium in the two-region economy with trade and immobile agents

Full equilibrium is obtained if both product and labour markets are jointly in equilibrium. Both regions face the same wage curve, represented by the curve VV in figure E3). The two BB-loci are the product market equilibrium curves from (E.19) that were graphically represented in fig. E2. The stable regional equilibrium points are at A and B respectively. As can be seen, region r has both the higher equilibrium wage and a lower unemployment rate.

Figure E3: Equilibrium in the two-region economy with immobile agents



Recall that we have assumed that region r is larger than region s, and can therefore better exploit the scale economies entailed in the adopted technology. For any given wage rate, it has a higher labour demand. This drives down unemployment at first instance and simultaneously increases the necessity to pay efficiency wages in order to prevent individuals from shirking. In equilibrium, the larger region ("the core") is advantaged over the smaller one along two dimensions: by having higher wages and a lower unemployment rate.

The fact that larger regions pay higher wages in presence of increasing returns in combination with transportation costs has already been noted in the NTT model of Krugman (1980) that has been discussed in the last chapter. Yet, the existence of unemployment exacerbates this agglomeration wage premium. It is not only because of technological factors and the better exploitation of scale economies. With efficiency wage based unemployment, we can see that the larger region must pay an additional wage premium to deter shirking in view of the low unemployment rate.

E3.3.) The impact of labour mobility

So far he have taken the sizes of the regional labour forces as exogenously given, which is equivalent to assuming that there is no labour mobility. But spatial differences as indicated in fig. E3) will induce workers from region s to migrate to region r. This conclusion seems less clear for Europe than for the US, since geographical mobility in Europe is much lower than in the US e.g. because of cultural and language barriers (Puhani, 1999). But nevertheless, for an analysis on the basis of regions, factor mobility must be taken into account. Furthermore, also in Europe there is some degree of responsiveness of workers to interregional wage and unemployment differentials (Decressin/Fatas, 1995; Pissarides/McMaster, 1990). A more general discussion about the causes and consequences of interregional labour migration is presented in the next chapter.

a) Perfect labour mobility

Labour migration is *not an equilibrating force* in our model framework, it rather perpetuates and accentuates regional disparities. In fact, if workers were *perfectly* mobile across regions, there would always be full concentration of all economic activity in one region regardless of the level of transportation costs.

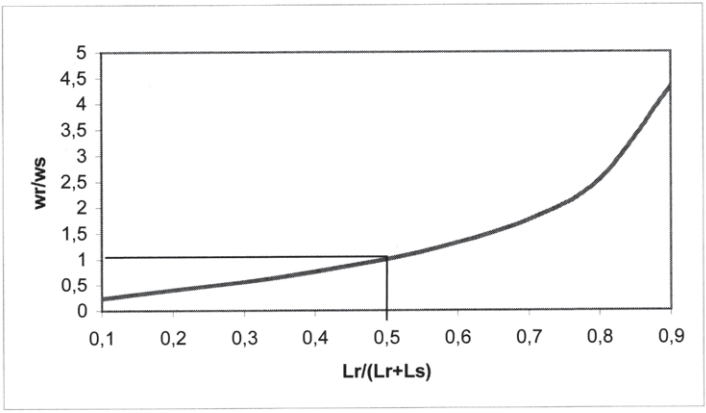
For notational convenience, let us normalize the size of the total national labour force to $\tilde{L} = L_r + L_s$. A fraction λ lives and works in region r, $L_r = \lambda \tilde{L}$. The remaining fraction $(1-\lambda)\tilde{L}$ is located in region s. From (E.19), the relative regional wage $\hat{w} = w_r / w_s$ is given by

$$\hat{w} = \frac{w_r}{w_s} = \left(\frac{(1-U_r)\lambda}{(1-U_s)(1-\lambda)} \right)^{\frac{1-\theta}{\theta}} \qquad \text{(E.25)}$$

As shown in figure E4), the relative wage $\hat{w}$ is greater than one if $\lambda > ½$. This is so for two reasons. Firstly, the better exploitation of increasing returns warrants a higher production wage. And secondly, unemployment is lower in region r. Thus, higher efficiency wages need to be paid. Suppose that we are initially in a symmetrical situation with $\lambda=1/2$. Initially, there will thus be identical wage levels and unemployment rates across regions.

But workers have an unambiguous incentive to concentrate in one region, say region r, since the relative real wage rate is steadily increasing in the degree of spatial concentration. All workers would thus migrate up to the point with $\lambda=1$ under perfect mobility. This concentration process is also socially optimal, as can be seen from the expression of total national output (E.22), or gross national income (E.23) in this economy. Both expressions, which have been proofed to be equivalent to (E.24), reveal larger values under an asymmetric distribution of the total labour force across space than under a symmetrical one with $L_r=L_s$.

Figure E4: Total agglomeration with perfectly mobile labour



The relative wage function is non-linear and strictly upward sloping for all values of $\tau > 1$ and $0 < \theta < 1$. In other words, there is no break on the feasibility of regional agglomeration in the model with perfect mobility; there is no "centrifugal" force. We will come back to the discussion of this issue below, when we critically assess our approach and compare it with a "typical" NEG-model.

b) The degree of agglomeration under imperfect labour mobility

For now, we will introduce an "ad-hoc" centrifugal force by assuming imperfect labour mobility. Two reasons for this construction can be put forward at this point: Firstly, the assumption of perfect mobility is neither appealing per se for most countries, nor is the result of complete agglomeration very reasonable. And secondly, recall from chapter C that the collapse of the wage curve relation in the model of B/O (1996) is only prevented by the presence of "ad-hoc" intrinsic regional preferences. By adopting a comparable set-up in our approach, we will be able to highlight in the clearest possible way in what respects our approach differs from that of B/O.

We add an individual specific discount factor $\kappa_i > 0$ that captures the relative attractiveness of region r perceived by individual i. The value κ_i is equal to one if an individual has no intrinsic preference for either region. Values $\kappa_i > 1$ indicate preference for region r. Vice versa, $\kappa_i < 1$ indicates a discount for living in region r. We assume that in the initial situation that half of the total labour force that is located in region r also has an intrinsic preference for this location (i.e. $\kappa_i > 1$). The other half of $\tilde{L}$, which is located in region s, instead dislikes region r.

Consider all $\tilde{L}$ workers in the nation rowed up in a line from i=0 up to i=$\tilde{L}$. We assume that the distribution of the individual preference parameter κ_i over the whole working population will be described by the following function

$$\kappa_i = \left(\frac{\tilde{L} - i}{i} \right)^{1/d} \qquad with\ d \geq 0. \tag{E.26}$$

The worker i=$\tilde{L}/2$ is the only one who has no intrinsic preferences for either region (κ_i=1). All workers in the range from λ=0 to λ=1/2, i.e. those who are originally located in region r, also have a preference for that location. Since these workers are indexed in the range $i \in \left(0\,;\,\tilde{L}/2\right)$, their value of κ_i is greater than one. On the contrary, all workers indexed in the range $i \in \left(\tilde{L}/2\,;\,\tilde{L}\right)$, i.e. those who initially live in region s, discount to live in region r. The individuals at the extremes, i=0 and i=$\tilde{L}$, have values κ_i=∞ and κ_i=0 respectively, and are thus inevitably tied to one specific region. The parameter d is a measure of how strong the overall labour force is biased towards living in the home location. The lower is d, the more pronounced are the preferences for one specific region. In the extreme case with d=0, all individuals are completely tied to one specific place. For d$\rightarrow\infty$, all individuals are intrinsically indifferent between locations.[2]

Let us consider the workers living in region s, who face the following trade-off: [3] they would receive an "economic" benefit if they were to move to region r, but they are intrinsically attached to their home location s. Thus, these workers must receive a premium in order to actually move. They move if and only if the relative wage of region r is greater than the individual dislike of that location. The analytical condition for a move of individual i to occur is $\kappa_i \hat{w} > 1$, or

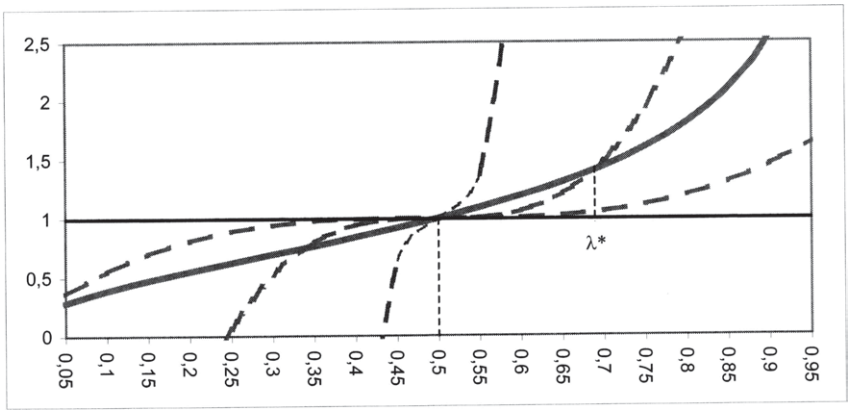
$$\hat{w} > \frac{1}{\kappa_i}.$$

[2] Note that the modelling of preferences κ_i only captures intrinsic preferences for the region r that are independent of the actual degree of spatial concentration in that location. The term κ_i does not represent a dislike of congestion etc.. However, given that the construction of regional preferences is "ad-hoc" anyway, one could easily extent the preference formation and introduce a term $\kappa(\lambda)$ that depends on the overall degree of spatial concentration. Total individual preferences might e.g. look like $\kappa(i,\lambda)=\kappa_i + \kappa(\lambda)$, where the former term captures purely individual factors and the latter the preference for or against congestion.

[3] Again, the analogous case where region s is the agglomeration area also applies. But we consider here the migration direction from s to r.

To obtain a better intuitive feel for this trade-off, it is graphically illustrated in figure E5. The thick solid line represents the relative regional wage $\hat{w}$ that was introduced in figure E4. The dotted lines represent three different examples of the inverted regional preferences $1/\kappa_i$. The steeper are the $1/\kappa_i$–curves, the more biased are individuals towards their home region, i.e. the lower is the parameter d. People move only if the curve $\hat{w}$ runs above curve $1/\kappa_i$. For the case with high regional preferences (the steepest curve in fig.E5), the final equilibrium will be the symmetrical initial situation, because the higher relative wage (the technological advantages from concentration) can never compensate for the individual dislike of region r. The opposite happens with low attachment to specific regions, the flattest dotted line in fig. E5). Here, agglomeration will be almost complete, as the location of workers is driven mainly by regional relative wages. With intermediate regional preferences, the equilibrium distribution of the workforce λ^* is given by the intersection point of the two curves.

Figure E5: Distribution of the labour force with regional preferences



The regional preferences do thus not change the fundamental result that the larger region is better off than the periphery with respect to both w and U. The existence of κ_i introduces parameter constellations, in which migration does not occur due to very pronounced attachment to home locations. Importantly, the construction with regional preferences allows for less-than-full agglomeration. There are interior equilibria $\frac{1}{2} < \lambda^* < 1$, where wages and regional preferences just compensate for each other. This enables us to draw implications from the *degree of agglomeration* (λ) for the magnitude of regional disparities.

Proposition E2: *Regional disparities are more pronounced the higher is λ^* in the range $\lambda \in [1/2;1]$. λ^* is larger, i) the stronger are the increasing returns to scale θ; ii) the less important is the overall home bias d.*

We have shown that regional disparities are highest, the closer λ is to the boundary case $\lambda=1$. Regional disparities are nil if $\lambda=1/2$. In between the two extremes the relation between λ^* and the magnitude of regional disparities is smooth.

E4) Critical discussion of our model approach

The model presented in section E3 aimed to integrate a wage curve into a the general equilibrium framework where the technology allows for agglomeration forces similar to the literature in NTT of NEG. The partial equilibrium foundation of the wage curve remained unchanged compared to existing work. But the product market structure in our model has been changed. We have formulated a generalized and extended version of the NTT model of Matusz (1996). Our framework allows us to explicitly take up the issue of regional unemployment disparities in relation to regional agglomeration. In this section, we want to first summarize the crucial mechanisms and results of our model. We then compare our approach with the core models of wage curve theory and NTT/NEG. This discussion, that will highlight in what respects our approach improves on the current state of art in the literature, will be followed by an assessment of the empirical implications of our model. These shall be contrasted with the available empirical evidence from chapter A. Finally we will highlight some further issues that motivate the contents of the next chapters.

E4.1.) Discussion of the model from a theoretical point of view

The central mechanisms in our model can be summarized as follows: Larger regions are advantaged over smaller regions, because they can better take advantage of the adopted technology with localised increasing returns to scale. This advantageous scale effect leads to higher regional wages in the larger region. This result has been known since Krugman (1980). If unemployment is explicitly introduced through a wage curve, one can show that the large region is advantaged also through a lower unemployment rate, and it needs to pay an additional efficiency wage premium. This insight can be seen as the fundamental contribution of the model approach from section E3.

Stated differently, it implies that unemployment rates will exhibit the same core-periphery structure as wages and output. Regional unemployment disparities are driven by the degree of regional agglomeration, as large ("core") regions exhibit low unemployment rates and high wages, whereas small (peripheral) regions have high unemployment rates and low wages.

a) Comparison with the wage curve theory

It is now worth discussing how our approach compares to the basic wage curve models from chapter C. We have named some fundamental criticisms with respect to the models of B/O (1996) and Blien (2001), which we took as the motivation to formulate an own approach. These critical points were:

- Regional disparities in the two model can not develop endogenously, but are due to exogenous assumptions, e.g. on sectoral specialization patterns.
- Regions were essentially identified with industries, and regional disparities were implicitly explained solely by sectoral factors.
- The existing models do not take into account regional agglomeration, which is one of the most salient features of economic landscapes and should thus not be neglected in regional approaches.
- Labour mobility leads to an erosion of the wage curve, and the foundation of the wage curve as a long-run equilibrium relation must rely on restrictive ad-hoc assumptions.

Our model approach leads to fundamentally different conclusions with respect to all these four points. In our model with a technology incorporating increasing returns to scale, regional asymmetries can develop completely endogenously. There is an endogenous tendency towards regional divergence transmitted through the market mechanisms alone. Furthermore, sectoral specialization patterns play no critical role in our model. In our two-region model, both locations are engaged in production activities within the same sector. Differences in the production structure exist only insofar that the larger region can produce a higher number of industrial intermediates. But all regional differentiation, and all interregional trade is of an intra-industry type. Through the working with localised scale economies, our model is able to account for forces that drive an economy towards spatial agglomeration. We have taken one of the classical arguments from Marshall (1890) that were discussed in chapter D (market linkages), and incorporated it explicitly in the analysis. It has been argued that such an approach in general seems much more plausible on a regional level than a conventional approach with constant returns to scale and perfect competition.

What about labour mobility? In Blien (2001), it is explicitly said that labour migration leads to the gradual erosion of the wage curve, i.e. to a gradual equilibration of regional wage and unemployment levels. For the model of B/O (1996), principally the same result holds. They assume the existence of regional disparities through an assignment of single regions to the production of specific distinct goods. The region that produces the "better" good ultimately has the higher equi-

librium wage rate and the lower unemployment rate. Workers from the region that produces the "bad" commodity therefore have an incentive to emigrate. But B/O assume that intrinsic regional preferences prevent migration from one region to the other, because individuals in the region with the unfavourable economic values are compensated by regional amenities, e.g. the lack of congestion.

In our model, we also have adopted "ad-hoc" regional preferences, which per se seem quite plausible. In one way, the construction is thus quite similar in spirit to B/O: The existence of intrinsic preferences slows down or even excludes migration in response to interregional wage and unemployment disparities. However, their role in our model is still fundamentally different. In B/O, the lack of migration prevents the erosion of the wage curve, or the convergence of regional wage levels and unemployment rates. In our model, the opposite is true. Here, the sluggish migration prevents further regional divergence. In absence of intrinsic regional preferences, the B/O-model would imply total convergence of regions, whereas our model (see section E3.3b) would imply complete concentration of all economic activity in only one location, i.e. "complete divergence".

This conclusion becomes most obvious when comparing the two central diagrams of the B/O-model (fig. C2) and of our approach (fig. E3). As can be seen, the wage curve locus in both diagrams is identical, and was similarly rationalized on the basis of the shirking approach of efficiency wage theory (Shapiro/Stiglitz, 1984). The product market equilibrium curves in the B/O-model under constant returns to scale are just horizontal lines on an exogenously given level in the wage/unemployment-space. It has been shown in chapter C that this shape is the outcome of a limitational production function under constant returns to scale.[4] The product market equilibrium curves in our model are instead downward sloping and non-linear. This reflects the scale effect that is incorporated in our technology. The lower is unemployment, the better can the scale economies be exploited and the higher is the equilibrium wage rate that is consistent with product market equilibrium. The figures C2 and E3 are constructed for given sizes of the regional labour forces. Introducing labour mobility would shift the two curves in fig.C2 closer together as workers from region 2 migrate to region 1. In fig. E3 on the other hand, the two curves would shift further apart. In other words, in the B/O-model the wage curve-type relation has an inherent tendency of vanishing. The implication of our model is fundamentally different: The wage curve is a stable relation that is even strengthened through labour mobility.

[4] If a straight neo-classical production function with diminishing marginal returns would be assumed, the respective product market equilibrium curve would be upward sloping in the diagram. With low unemployment, the marginal productivity (=wages) falls. The higher is unemployment, the scarcer is labour and the higher is the MPL and the wage. For a further discussion see section B3.3.), where this case has been discussed in the vein of the PS curve of the ELMM.

Hence, if one works with an increasing returns technology, the theoretical case for the existence of a wage curve is even stronger than it has been argued by B/O themselves. They are preoccupied to view the wage curve as a stable equilibrium relation over time. With their constant returns to scale technology, they can only rationalize this perception with the "ad-hoc" construction of intrinsic preferences. The use of an increasing returns to scale-technology, however, quite naturally implies the long-run stability of the wage curve as a basic outcome. The "ad hoc" preferences in our model put boundaries on the occurrence of regional divergence, but their use is not critically needed to prevent the erosion of the wage curve. All in all, the use of an increasing returns considerably improves the general equilibrium wage curve theory. Mechanisms for the emergence of regional disparities (which after all are of central interest also to the wage curve literature) develop endogenously and do not need to be exogenously assumed. Agglomeration is explicitly accounted for, and the wage curve is truly a long-run equilibrium relation due only to deep technological factors.

b) Comparison with NEG-models

How does our model relate to the NTT- and NEG-literature? Our approach was build on one specific NTT-model by Matusz (1996). We have argued that this framework is useful, because it combines a labour market equilibrium curve that can be interpreted as a wage curve with a production technology exhibiting increasing returns to scale.

We have extended the Matusz-model and introduced transportation costs and labour mobility. Thereby the analysis was brought closer in spirit to the NEG-literature. Nevertheless, the approach from section E3 differs in some important respects from a "typical" NEG-model. Recall from the discussion in chapter D that the core question of NEG is how the landscape of an economic area shapes as a function of opposing centripetal and centrifugal forces. Usually, the relative strength of either force is determined by exogenous parameters like the level of transportation costs.

This is different in our approach. Technologically there are only centripetal forces. With perfectly mobile labour, there would be complete agglomeration in order to fully exploit the increasing returns in production. This agglomeration would be socially optimal as it maximizes total national income and minimizes national unemployment. As a quasi-centrifugal force we have added intrinsic regional preferences (an "attachment to home" parameter) that led to imperfectly mobile labour. In the terminology of section D4, this type of mechanism should be labelled a "status-quo bias" rather than a "pure" centrifugal force.

Therefore, one can criticise our approach as not offering a fully *endogenous* explanation for the emergence of a geographical equilibrium structure. Our first reaction to this reproach would be to admit that our model is indeed not a "typical"

NEG-model. But we feel that there are good reasons for having taken our simpler approach. We like to comment now on these reasons.

We have argued in the introduction of this chapter that the motivation of our own model approach is twofold. The first objective was to formulate a general equilibrium model of the wage curve with a technology that allows for agglomeration forces. As argued in E4.1a), we feel that this objective has been achieved. We have formulated our model such that the analogy with the approach of B/O (1996) becomes as clear as possible, which enabled us to point also as clear as possible to differences and common features. As the second objective we have formulated that unemployment should be integrated as an element to the new regional agglomeration theories. This class of models is wider than just "typical" NEG-models in spirit of Krugman (1991a) or Venables (1996). Given that our model from section E3 also explicitly accounts for the role of space and for regional agglomeration, it surely belongs to this broad class of models in the literature, without being a NEG-model in the strictest meaning of the word.

But why did we not use a "typical" NEG-model? This is mostly due to the intractability of these models. Recall from chapter D that NEG, even without the analysis of unemployment, usually has to rely on numerical solution techniques. With unemployment added, these problems will be exacerbated, and instructive and intuitive analytical solutions will be even more out of reach. After all, the important question is whether working with a "typical" NEG-model would have offered additional insights that are not available with our relatively more simple approach. To assess this question, it is important to again consider the basic mechanisms e.g. of the seminal Krugman-model. The centripetal force in that model is structurally very similar to our approach.[5] The main difference is that Krugman's model entails an endogenous centrifugal force. This stem from the fact that he assumes local immobility of the agricultural workers. In our model, the centrifugal tendency is absent because we do not assume *prima facie* immobility of parts of the workforce. This immobility of consumption demand in combination with transportation costs drives the results of the Krugman-model and leads to the non-linear interplay between agglomeration and dispersion forces. But recall that in the long-run equilibrium, the economy is either operating in complete symmetry, or parameters are such that a process of regional divergence is triggered at the end of which the whole mobile labour force is concentrated in one region.

We have abstracted from the particular centrifugal force of the Krugman model, because our main interest is not to derive the equilibrium geographical structure of the economy as a function of parameters, but rather to draw implications from the *degree* of agglomeration for the magnitude of regional unemployment and wage

[5] The only (not crucially important) difference here is that the endogenous linkage effects in the Krugman model are between the firm and the consumption sector, whereas in our model they accrue on the production side alone.

disparities. If the underlying model can only predict either complete symmetry or complete asymmetry, the analysis of regional disparities becomes somehow meaningless. With our set-up, we can specifically look at the impact of the variable λ (that reflects the degree of spatial concentration) on differences between region r and region s. With the Krugman-model, such an analysis of gradual differences would not have been possible.

There are of course NEG-models that also allow for imperfect labour mobility and can thus predict intermediate degrees of agglomeration (e.g. Ludema/Wooton, 2000). In their model, there are parameter ranges where complete symmetry prevails. But then eventually, spatial asymmetries emerge as transport costs fall. But the change in the economic landscape is smooth and intermediate equilibria are feasible. The model of Ludema/Wooton would thus be applicable also for our purposes, since one could try to introduce unemployment in their model and see what the degree of agglomeration implies for spatial unemployment differences. But one has to acknowledge that using their (much more complicated) set-up would not add significant new insights compared to our approach. As long as transport costs are such that the economy operates under symmetry, by definition there can be no regional unemployment disparities. Once centripetal forces prevail, the model of Ludema/Wooton would likewise predict that the degree of spatial unemployment disparities is larger, the higher is the concentration of the increasing returns sector in one region. In our simpler approach, there is no parameter range of transportation costs for which there is symmetry. But for the "interesting" case with regional disparities, our model qualitatively behaves in the same way as a more complicated NEG-model would do.

Hence, we feel confident that our model set-up, where technologically only centripetal tendencies exist and quasi-centrifugal forces are entered "ad-hoc", is sufficient to make our basic point that regional unemployment rates depend (negatively) on regional agglomeration. We now want to turn to a discussion of our theoretical result in view of the available empirical evidence.

E4.2.) Discussion of the model from an empirical point of view

In section A3.1.) we have shown that regional unemployment rates in the European Union (EU-15) closely resemble the core-periphery-structure of regional GDP per capita. Low unemployment is centred around the "European Banana", medium income regions usually belong to the group with intermediate unemployment rates, whereas the poor "objective 1"-regions mostly have very high unemployment rates. We have also shown that the relation between regional GDP per capita and the regional unemployment rate is not "one-by-one".[6] But all in all,

[6] In section A3.1.) we have stated the following empirical regularity: belonging to the 'objective 1'-group is a necessary, but not a sufficient condition for exhibiting extreme regional unemployment rates of above 15 per cent or so.

the spatial distribution of unemployment rates closely follows the spatial agglomeration pattern.

a) The theoretical model and the stylised empirical facts from the EU

The theoretical model from section E3) is consistent with this observed stylised fact. The model predicts that unemployment is low in regions with high real wages (=real income levels) and high output levels (equation (E.21)). In the model, scale effects play a crucial role. The output and income level of a core region is high, because a large number of workers is concentrated in this location. The evidence presented in section A4.2.) is also consistent with this theoretical result. The rich regions from the "European Banana" with low unemployment rates are also characterised by a high population density, which is a good proxy for saying that these regions are large.[7] The population density (the "size") in "objective 1"-regions is considerably lower. Low unemployment and high income levels in the EU-15 thus coincide with a high spatial concentration of population (and workers). This stylised fact is broadly consistent with the prevalence of technological scale effects.

Empirical support for the negative effect of economic agglomeration on the unemployment rate is provided by Blien et. al. (2002). The authors analyse the determinants of employment growth across East German regions for the time period 1993-2000. The partial effect of agglomeration is found to be significantly positive for explaining local employment growth. In view of the strong correlation between employment growth and falling unemployment rates, this empirical finding can also be interpreted such that economic agglomeration in any region has a negative impact on the local unemployment rate.

The next stylised fact we have presented in section A4.2.) is that "objective 1"-regions have tended to loose population to the rich core regions through migration. Our model implies that this will lead to further regional divergence. And again, the evidence presented in section A2.2.) is at least not inconsistent with this claim. There has been regional divergence on a regional level in the second half of the 1990s, and the regions that faced emigration during this period (e.g. Southern Italy) also tended to fall back with respect to relative income figures.

[7] Of course one has to use population densities rather than population sizes in absolute terms, because the surface of NUTS2-regions is utterly unequal. If the territorial size of a region is not taken into account, one can come to quite misleading conclusions. E.g. Andalucia (ES) is one of Europe's largest NUTS2-regions, which in the vein of our model would imply that the regional unemployment rate should be quite low. Andalucia, however, has one of Europe's highest unemployment rates (about 26%). Yet, Andalucia is also one of Europe's largest regions in terms of surface, and a view at map A6 indicates that the population density of Andalucia ranges somewhat below the European average. Hence, our theoretical results look much more reasonable again if a meaningful measure of the "size" of the respective regional labour force is used.

When looking at the correlation between regional agglomeration and regional unemployment rates, there are of course single counter-examples that contradict our general conclusion: Portuguese regions e.g. tend to be small and sparsely populated, GDP per capita and particularly GDP per employed worker is low. Still, the unemployment rates in these regions are also below the European average. However, Portugal constitutes an exception to a rule. All in all, this single case is insufficient to falsify the general spatial pattern that emerges from the maps in chapter A, which can be reasonably well explained by our theoretical model.

Based on these considerations, we conclude that the approach presented in this chapter is well consistent with the empirical evidence on the regional dimension of economic activity in the EU-15. It is therefore not only a contribution from a theoretical point of view, as it combines and extends previously unrelated strings of literature. The model also yields empirically relevant conclusions.

b) Intra- versus international disparities and the role of institutions

It is important to note that regional unemployment rates differ in our theoretical model, even though there are no differences in the underlying structural or institutional characteristics of the regional labour markets. We have argued above that the two parameters e and γ in principle can be thought of as reflecting structural characteristics of the labour markets in the two locations. But purposely we have assumed that e and γ do not differ, i.e. we have assumed that both regions face the same labour market institutions. Still, unemployment disparities can arise. But they are solely driven by other factors, namely the agglomeration economies that drive the location of labour demand, which in a secondary step determines the geographical configuration of unemployment.

In other words, our model neglects institutional differences as an explanation for spatial unemployment disparities. In this respect, our approach offers a perspective about the phenomenon of unemployment that is quite different from the conventional viewpoint in economic research, which usually attributes the observed cross-sectional variation of unemployment rates across different administrative units to the impact of institutional arrangements, most notably to various forms of labour market rigidities (e.g. Nickel, 1997). This conventional type of analysis is widely known. Put bluntly, it views the unemployment rate of any given jurisdiction as a measure of the malfunctioning of the labour market, somehow in the vein of Friedman's *natural rate of unemployment* (see section B2.3.). Arguments in this spirit are normally concerned with unemployment in its national dimension. There is of course some degree of institutional heterogeneity across the single national labour markets. E.g., the Danish and the Dutch labour market institutions are quite different from the French and German ones. Typically, these institutional differences are argued to explain fairly well the differences between the national unemployment rates (Nickell, 1997).

However, when looking at the evidence presented in chapter A, one can have serious doubts whether it is really useful to predominantly think about the phenomenon of unemployment along national boarders. We have shown in quite some detail that dramatic spatial unemployment disparities exist within almost each EU member country. On an intra-national basis, however, institutional differences are almost negligible. In Germany e.g., the institutional rigidities that are frequently argued to have a significant adverse impact on the performance of the labour market apply to the entire nation without any regional differentiation. This concerns labour laws, the tax regime, the welfare state arrangements, the high degree of union power etc.[8] The same set of institutions can bring about utterly different unemployment rate outcomes on a regional level, ranging from around 3.5 per cent in Oberbayern to roughly 20 per cent in Halle. The case of East Germany might be subject to very special influence factors due to the particular historical circumstances of division and reunification. But think of Italy or Spain. The "classical" labour market rigidities are also valid nationwide. Yet, unemployment differences of more than 20 percentage points exist and show no tendency to vanish. In practically all EU countries, except for the Netherlands, regional unemployment disparities are present, which can not be attributed to differences in labour market institutions across the single regions. Moreover, we have argued in chapter A – following Overman/Puga (2002) – that there exists a more meaningful grouping scheme for regional unemployment rates other than the pure assignment of regions to different member countries. There exist trans-national unemployment clusters that do not obey to national boarders. The spatial configuration of unemployment rates across the EU as a whole was shown to follow a quite distinct spatial pattern (European banana, intermediates, objective 1), in which national boarders have played a minor role.

What do these empirical facts imply for the role of labour market institutions on unemployment? First of all it is clear that it is far too simple to interpret a regional unemployment rate as being determined at any point in time by the prevailing institutional characteristics of the local labour market. It is inappropriate to work with a simple function

$$U_r = f(V_r) \tag{E.27}$$

that describes the unemployment rate of region r only as a function of institutional variables V_r. With a relation like (E.27), there could be no intra-national unemployment disparities, since there is no notable intra-national variation in labour market institutions.

[8] Regional institutional differences in Germany exist at best with respect to the impact of collective bargaining, because East Germany is covered to a far lesser extent by union minimum wages. But unemployment in the East is far above Western levels.

But it would be precipitate to conclude that the presence of intra-national unemployment disparities implies that institutions are entirely irrelevant. This point can be illustrated by going back to the theoretical analysis from section E3. Let us consider two cases. Firstly suppose that the two regions r and s belong to the same nation, in which case the two institutional parameters e and γ (by definition) are identical in both regions. And secondly, assume that the two regions belong to different nations, and the parameters can be different in r and s. The first case has been analysed in section E3. It was shown that the assumption of no variation in the parameters e and γ implies that both regions face the same wage curve locus VV. Full equilibrium in figure E3 is obtained at the intersection points with the respective product market equilibrium curve BB. Due to the agglomeration economies, it is possible that spatial unemployment and wage disparities open up. They are entirely driven by the two BB-curves.

In the second case with institutional variation, the wage curve loci VV do not necessarily have to be identical in the two locations. Suppose that the work effort parameter e is higher in region r than in region s ($e_r > e_s$). Recall that the parameter e reflects the perceived disutility from working, and it can be understood as a real reservation wage. Maybe this effective wage floor is higher in region r because of a more generous welfare state regime. What are the consequences of this institutional difference? The wage curves of region r and s now run strictly parallel, and the curve for region r is located further apart from the origin. With the BB curves as given in figure E3, we might reach a final equilibrium where the unemployment rate in region r is lower (and wages higher) than in region s. But compared to the benchmark case with no institutional variation, the spatial disparities are now smaller.

In other words, institutions do matter in our approach insofar as they determine the labour market equilibrium locus VV, which is "one half" of the full equilibrium. But institutions are not the only influence factor for determining the regional unemployment rates. The second half, that is emphasised in our approach and (as we see it) often neglected in the literature, is the influence from the product market, with the agglomeration economies and the location of labour demand.

With these considerations in mind, we can more specifically demonstrate the empirical relevance of our model. In the version with no institutional variation, the model is able to explain why regional unemployment rates can be so utterly different within the same nation, notwithstanding the prevalence of the same institutions across regions. But the model is relevant also in a wider sense. The main motivation for our theoretical approach was to explain the geographical structure of unemployment in the EU as a whole, i.e. also across the single nations. The geography of unemployment closely resembles the overall core-periphery structure of economic activity. We have argued above that our model predicts precisely this resemblance, and is thus well suited to explain the regional dimension of unemployment in the EU. But we have also mentioned a few outliners. For example,

unemployment rates are very low in Portugal even though all Portuguese regions are located in the economic periphery. Supposedly this is due to the particular labour market institutions that render a wage curve in the Portuguese regions that is located closer to the origin as in other European countries. In other words, it might be the case that the VV curves are different in the single European regions, depending on national affiliation. The regional unemployment rate and the real wage level in equilibrium are determined jointly by the wage curve and the product market equilibrium curve, i.e. by institutions and agglomeration together.
Institutional rigidities are by no means irrelevant for the determination of unemployment rates. But all in all, the influence of international differences in labour market institutions on unemployment can not be all to large. The geographical structure of unemployment on the basis of regions in the EU corresponds very little to national boarders. If institutional differences were very relevant, country boarders should be much more visible in a map like A3. Since this is not the case, we conclude that the spatial configuration of unemployment is mainly driven by product market agglomeration.

E5) Further issues

In section E4 we have tried to convince the reader that the theoretical model presented in this chapter is both an innovation from the point of view of economic theory, and an empirically relevant model. In the final section of this chapter we want to raise two additional issues that shall be further discussed in the next chapter.
Firstly, we think that a comprehensive discussion about the role of labour mobility is needed. We have argued that any regional model must take into account geographical factor mobility and must say something about the causes and the consequences of it. Labour mobility is a crucial element in the interregional general equilibrium of the wage curve literature, it is essentially the distinguishing feature between NTT and NEG, and it also plays a very prominent role in our theoretical model. The next chapter F will therefore be devoted to a systematic assessment about the role of labour mobility in view of the different theoretical models we have discussed so far. The general issue at stake is if labour migration is an equilibrating force that ceteris paribus leads to an erosion of regional disparities, or if labour migration rather perpetuates regional asymmetries.
During this discussion, we will highlight that the type of the underlying technology is an important determinant of the spatial effects of labour migration. But we will also raise another issue that is crucial for the consequences of internal labour migration, but that has not been explicitly spelled out yet. Namely, we will explicitly distinguish between different types of labour, skilled and unskilled labour, and analyse the impacts of selective labour migration. We show that migration will not lead to regional convergence, but rather to a strengthening of regional disparities, particularly if only the high skilled workers are regionally mobile. This is surely

the case if the underlying technology is subject to agglomeration economies, but even under a very neoclassical set-up with constant returns to scale.

The theoretical analysis in the next chapter, however, will also be interesting from another perspective. In the approach of this chapter, we have introduced unemployment through a wage curve, which has been grounded on efficiency wages. In the model in the next chapter, unemployment will occur, and regional unemployment disparities will arise, because of union wage setting. Since unionisation is one of the most salient features of European labour markets, this alternative approach seems interesting for its own sake already. But further insights will be gained by that model. We will consider one particular form of wage setting that seems very relevant for continental European economies. Namely, we will take the case where a national union sets national minimum wages irrespective of regional productivity differences. This already makes clear that the models derived in the next chapter will be specifically concerned with the phenomenon of *intra-national* unemployment disparities, not with the trans-national unemployment clusters that have been discussed in this chapter. The models we derive in the next chapter are no wage curve models, as will be discussed in detail below. But we reach conclusions that are consistent with the wage curve literature. The insights from the model approach of this chapter are thus complemented and extended.

Chapter F) Internal migration and regional disparities

F1) Introduction

The first aim of this chapter is to look more specifically at the issue of geographical labour mobility. For this task, it is extremely important to distinguish sharply between the *causes for* labour migration between regions, and the *consequences of* labour mobility for the sending and the receiving locations.

The former issue, the causes, will only briefly be discussed in section F2, since it is relatively uncontroversial. The consequences and economic effects of labour mobility on the other hand are heavily disputed in the literature. Is labour mobility an adjustment force that gradually leads to an erosion of existing spatial disparities? Or do differences even get larger and more pronounced when workers are mobile across regions?

Controversies about this issue have a long tradition in the history of economic thought. And this old discussion is also alive in this book. The theoretical models that have been discussed in the preceding chapters have drawn utterly different conclusions about the consequences of labour migration. In section F3 we illustrate the conventional result from neoclassical economics that labour mobility leads to regional factor price convergence. But it will become apparent that this result hinges on quite restrictive assumptions, and empirical evidence does not seem to support the neoclassical convergence hypothesis all too strongly.

In the following two sections we will therefore put forward two distinct arguments why labour mobility might lead to spatial divergence instead of convergence. The first reason, presented in section F4.1.), is the presence of scale and agglomeration economies. This issue has already shown up in the last chapters, but it will be restated here with a special emphasis on the effects of migration on the regional labour markets. The second reason for "non-convergence" is self-selectivity within the group of migrants. We show in section F4.2.) that there is a great deal of evidence in support of the view that skilled workers tend to be geographically more mobile than unskilled workers. This fact can be explained by straightforward economic considerations.

The consequences of selective labour migration are then derived in detail in section F5 where we explicitly distinguish between two types of workers, skilled and unskilled ones. In order to exclusively focus on the selectivity argument for "non-convergence", we will work with a very simple product market structure with constant returns to scale in this section. We will consider both a full employment version of the model, and a version where immobile unskilled labour is subject to unemployment due to a union minimum wage. The latter model variant is interesting also in another respect, as it specifically addresses the role of national union wage setting for regional unemployment outcomes.

In section F6 we come back to the issue of regional agglomeration by constructing a model that features both localised increasing returns to scale and the selectivity of internal migrants. Within this set-up, we will likewise demonstrate the "non-convergence"-character of labour migration, and analyse the impact of undifferentiated national collective bargaining on regional unemployment rates and effective earnings. In section F7 we compare the theoretical approaches from this chapter with that of chapter E, and we assess again the empirical applicability and relevance of the models from this chapter.

F2) The causes of internal migration

The location decision of any individual is an inherently economic decision. Any individual calculates a gross utility for remaining in the region of residence, and one for moving to the best alternative region of his or her choice. A moves takes place if the cost of migration does not exceed the gross gains from moving (Pissarides/McMaster, 1990). There is of course a large array of individual-specific characteristics and political impact factors that influence the mobility choice of a single person. Individual characteristics that influence the mobility behaviour are e.g. the age, family backgrounds, language capabilities, regional preferences etc.[1] Migration might also be driven by jurisdictional policy differences. Certain regions might attract immigrants because of an attractive tax and welfare state regime (Razin/Sadka, 2001; Wildasin, 2000). Supposedly these political factors are less relevant in the context of internal, i.e. intra-national, than in the context of international migration.

But we abstract from these influence factors and focus on two particular important forces that seem most relevant for explaining regional migration patterns. Economic considerations suggest that regions should all in all loose working population the lower is the regional wage level relative to other locations, and the higher is the relative regional unemployment rate.[2] In all models in the preceding chapters, interregional migration was always driven by regional wage and unemployment disparities.

A look back at figure A8 casually supports this theoretical construction also from an empirical point of view: European NUTS2-regions that were subject to population decline between 1993 and 1998 shared on average some common characteristics. Among these were an above-average unemployment rate, and a below-average per capita income (which should be highly correlated with wages). Further pieces of evidence can be obtained from section A5, where we have presented

[1] In a later section of this chapter we will take into account the individual skill level as an important impact factor for the personal mobility decision. For an older, but still very instructive survey about the micro-motives for migration, see Greenwood (1975).

[2] One might also label a high regional unemployment rate a "push factor" that spurs emigration, and high regional wages a "pull factor" that attracts immigrants from other areas.

stylised facts on regional disparities within West Germany. It was shown that the southern part of West Germany reveals higher effective earnings and lower unemployment rates than the northern part. Internal migration flows consequently went from the north to the south.

The plausible hypotheses that internal migration is driven by regional wage and unemployment differences has been verified by numerous econometric studies. Pissarides/McMaster (1990) e.g. estimate a net internal migration equation for nine regions in the UK for the time period from 1961-1982. They find that regional unemployment rates and regional wages are significant explanatory variables (with the expected sign) for regional net migration rates. The result is verified for the case of West Germany by Suedekum (2003), who estimates a similar net migration equation for ten west Bundeslaender (without Berlin) covering the time period from 1988-1999. Consistent empirical findings for the case of West Germany come e.g. from Alecke/Untied (2000), Buettner (1999) or Decressin (1994).

Hence, there seems to be considerable consistency in the empirical literature about the *causes* of internal migration. But does the well established empirical fact that workers move out of regions with high unemployment rates and low wages imply anything about the consequences of migration?

F3) The consequences of internal migration: The neoclassical view

Some authors, e.g. Pissarides/McMaster (1990) or Buettner (1999) claim that the empirically observed direction of internal migration flows indeed implies the existence of a long-run equilibrium where regional wage and unemployment differences have vanished, or persist only in form of compensating differentials á la Harris/Todaro (1970) and Hall (1972). It is admitted that migration might be a quite weak equilibrating force. Hence, a very long time period might be needed in order for regional disparities to disappear. But eventually, regional convergence will occur as the result of emigration out of the areas with poor economic conditions. These inferences about the consequences of migration are straightforward derivations from the neoclassical theory of migration. This can easily be demonstrated by a simple theoretical framework that we will introduce in this section.

Suppose there is a static economy consisting only of two regions r={1,2}. Production in either region is described by a Cobb-Douglas production function, where output Y_r is produced by combining the input factors physical capital K_r and (homogenous) labour L_r. The parameter $0 < \upsilon < 1$ is the usual capital share. The capital stock is assumed to be fixed for either region. There is furthermore a term A_r capturing the total factor productivity of region r.

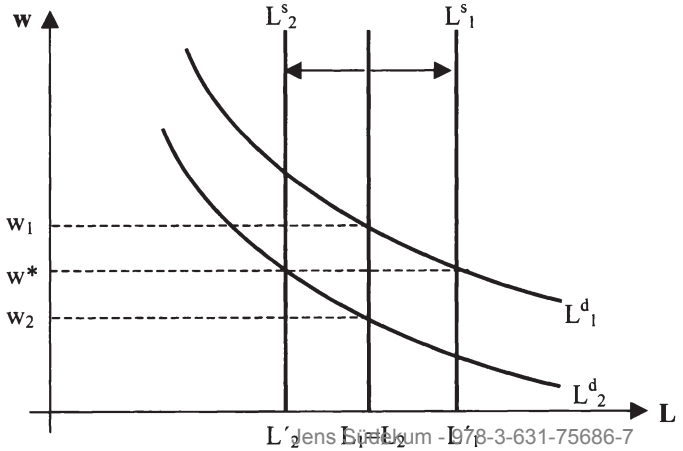
$$Y_r = A_r(K_r)^{\nu}(L_r)^{1-\nu} \tag{F.1}$$

Product and factor markets work perfectly, i.e. we consider at first instance a full-employment model. The good Y_r is freely tradable, and the price p^Y is normalized to unity for both regions. We know that under these conditions the equilibrium wage in region r, w_r^*, is given by

$$w_r^* = (1-\nu)A_r\left(\frac{K_r}{L_r}\right)^{\nu} \tag{F.2}$$

The equilibrium wage rate in region r depends on the capital/labour-ration in the respective location. Ceteris paribus, it decreases when more labour enters the market. Now consider the following thought experiment: suppose that the labour force and the total factor productivity in region 1 and 2 are initially identical ($L_1 = L_2$, $A_1 = A_2$), but the fixed capital stock of region 1 is larger than in region 2 ($K_1 > K_2$). From (F.2) it immediately follows that the equilibrium wage of region 1 is above the level in region 2, because labour is the relatively scarcer factor. With mobile labour, this will cause migration from region 2 to region 1 and put equilibrium wages in r=1 under strain, as employment increases and the marginal product of labour declines. The migration process will continue up to the point of wage equalization, which in our example is equivalent with the point where the capital/labour-ratios are equalized across regions. Migration thus leads to wage convergence, since it changes the relative factor intensity in either region. This logic can be illustrated in figure F1, which shows the labour markets of both regions jointly in one diagram.

Figure F1: Migration in a neoclassical two-region model



In the initial situation, labour supply (which we assume to be perfectly inelastic) is identical in both regions. But due to the larger fixed capital stock, labour demand is higher for any given wage rate in region 1. The equilibrium wage w_1 is therefore larger than w_2. The subsequent labour migration shifts the labour supply schedule of the receiving region 1 to the right, labour supply in the sending region 2 is shifted to the left. The labour demand schedules are not affected by the mobility of input factors. Final equilibrium is obtained when there is interregional wage equalization at w*.

Introducing unemployment complicates the analysis to some extend, also depending on what is modelled as the source of unemployment. But intuitively, migration will also lead to convergence of regional unemployment rates if neoclassical technological conditions prevail. This point can be made also in a very simple way. Suppose that the initial low wage region 2 is subject to unemployment, whereas there is full flexibility and thus full employment in region 1. If the unemployed of region 2 can freely emigrate, the unemployment rate of region 2 will decrease and converge to the level of zero that prevails in region 1. Similarly, the equilibrium wage in region 1 decreases and converges to the level of region 2, because additional labour is added to production, which causes diminishing returns.

All in all, one can see even from this rudimentary static model what is essentially behind the hypothesis of factor price convergence: it is the prevalence of neoclassical assumptions on the production function, i.e. constant returns to scale and diminishing marginal returns in case of partial factor variations. This leads to the well known proposition that factor prices are determined by factor proportions alone, as the relative scarcity implies the marginal factor productivity that in perfectly competitive factor markets is equal to the real factor price. Of course all factors in the neoclassical production function are assumed to be homogenous, which specifically means that all workers in either region are perfectly substitutable input factors in the production process (F.1).

F4) Internal migration and regional divergence: alternative views

But does the empirically observed direction of migration flows from section F2 necessarily imply the long-run equalization of regional disparities like the small neoclassical model from section F3 is telling? In this section we will propose two distinct arguments why this might not be the case.

But before doing so, it is worth noting that there is considerable empirical dissatisfaction with the neoclassical implication that labour mobility will lead to regional convergence. The available empirical evidence on this matter is mixed at best (Walz, 2001). It is highly questionable from an empirical point of view if migration actually spurs β-convergence in the context of international growth regressions (Braun, 1993). The persistence of regional disparities in the EU (section A2-A4), or in West Germany (section A5) alone suggest that migration seems to be a quite poor adjustment force. A very pronounced conclusion can be found in Fass-

mann/Mausburger (1997:190), who point out that "*The basic question of regional economics, whether migration contributes to the adjustment of regional [...] disparities, has to be answered with 'no' from a short- and a medium-run perspective*".

The purpose of this section is thus to provide theoretical rationale for the possible "non-convergence" character of internal labour migration. We want to point out how sensitive the neoclassical convergence hypothesis is with respect to the underlying technological assumptions. Even slight departures from the straight neoclassical world will lead to fundamentally different conclusions about the spatial consequences of internal labour mobility. The first argument for "non-convergence", presented in section F4.1), concerns the presence of increasing returns to scale in production that prevent the marginal productivity of labour from falling

F4.1) Increasing returns to scale

Can the real world actually be described reasonably well by a neoclassical production function with its restrictive underlying properties? This issue is ultimately a matter of empirical research.[3] But, as pointed out in chapter D, at least on a regional level one can have serious doubts whether a production function with a scale elasticity of one is an appropriate approximation of reality. In this section we show what happens to the impact of labour migration if production on a regional level does not exhibit constant returns to scale, but rather is subject to scale economies.

In the neoclassical model in section F3, the equilibrium factor prices in either region were determined by relative factor intensities alone. During the process of migration only the labour supply functions shifted in both regions, whereas the labour demand schedules remained constant. But labour migration might also have labour demand effects, which can moderate or even offset the supply effects. Recall in this respect the evidence that was presented in section A4.1.). There it was shown that regions with declining unemployment rates in the EU faced immigration of workers (i.e. an increase in labour supply). But at the same time, these regions experienced an even stronger increase in employment growth (i.e. in labour demand). It might be, that this increase in employment growth is causally due to the increase in labour supply.

Effects on labour demand can result from various channels. Let us first consider the case where labour demand is affected because of a technological externality. Suppose the production function (F.1) is rewritten in the following way

[3] For a brief assessment whether neoclassical production functions are an appropriate approximation of reality see Hamermesh (1993).

$$Y_r = A(L_r)\left[(K_r)^{\nu}(L_r)^{1-\nu}\right] \tag{F.3}$$

with $\partial A_r/\partial L_r > 0$. The term that captures the total factor productivity in region r is no longer given exogenously, but now depends endogenously on the size of the regional labour force L_r. An intuitive reasons why a larger labour force might lead to a higher total factor productivity is e.g. the presence of intra-regional knowledge spillovers. The concentration of more workers in one region leads to a more vital exchange of work experience etc.. Other plausible stories have been presented also sections D2 and D3b), where we have pointed out that this modelling strategy was very popular in the first wave of the endogenous growth literature. One major technical advantage of a formulation like (F.3) is that the parametric introduction of the aggregate externality is consistent with perfect competition. Each single firm maximizes profits while taking the aggregate total factor productivity as given. The equilibrium wage in either region is now given by

$$w_r^* = (1-\nu)A(L_r)\left(\frac{K_r}{L_r}\right)^{\nu} + \frac{\partial A_r}{\partial L_r}\left[K_r{}^{\nu}L_r{}^{1-\nu}\right] \tag{F.4}$$

Compared to (F.2), an additional term enters the expression of the equilibrium wage. This term captures the aggregate externality and is increasing in the labour force size L_r. Equation (F.4) implies that during the process of internal migration not only labour supply, but also labour demand curves are affected. This is illustrated in figure F2. Additional to the labour supply shifts, there are now also secondary shifts in the labour demand schedules of both regions.

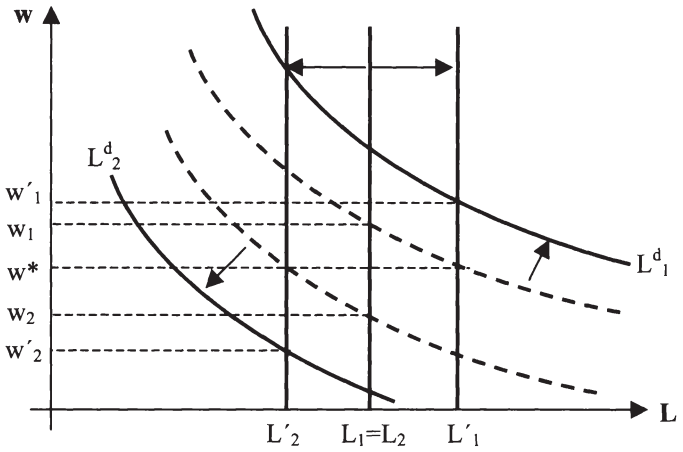
As migration proceeds from region 2 to region 1, the labour demand schedule of region 1 is shifted to the right due to the external scale economies, the opposite happens in region 2. Depending on the strength of the externality, i.e. on the exact shape of the function $A(L_r)$, the migration of labour might not lead to a closing interregional wage differential, but rather to a widening. Labour mobility in this example thus leads to further regional divergence rather than to regional convergence.[4]

An argument in this spirit has e.g. been made by Burda/Wyplosz (1992). They also model a two-region economy, where regionally mobile workers embody human capital that enters the production function externally in a way that is loosely

[4] If nothing else is added, and labour is perfectly mobile, all workers will move out of region 2 and concentrate in region 1, in order to fully exploit the external economies of scale. Note also that if the externality is not too strong, the shift in labour supply might be stronger than the shift in labour demand. In this case, the regional equilibrium wages would convergence as internal migration proceeds, but the convergence speed is reduced through the aggregate externality.

comparable to (F.3). They also conclude that labour migration might not be an equilibrating force, but might rather lead to a "Mezzogiorno" syndrome for the sending region.

Figure F2: Migration in a two-region model with an aggregate externality



Labour migration of course also leads to regional divergence if the scale economies do not enter parametrically through a technological externality, but if the scale economies are internal and transmitted through a pecuniary externality. The difference between the two concepts has been explained in depth in section D2. A good example is our approach from section E3. We have derived in great length that labour migration in our approach does not lead to regional convergence, but that regional disparities get sharper if workers move out of the disadvantaged areas. In the light of the discussion from this section, we can illustrate the divergence result from section E3 also in term of labour supply and labour demand. Consider the product market equilibrium condition for the regions r and s, equation (E.19). Let us abstract for the moment from unemployment and consider the easier full employment case. Manipulating the simplified version of expression (E.19) yields the following upward-sloping labour demand curve for region r

$$L = \left(\ell / T\right)\left(w\right)^{\frac{\theta}{1-\theta}} \quad . \tag{F.5}$$

An analogous expression applies to region s. The single variables are defined as in chapter E. The intuition for the positive slope of the labour demand curve (F.5) has been extensively discussed in the last chapter. In short, the underlying technology is subject to an internal scale effect. The expansion of the labour force leads to decreasing production costs, and via a zero profit condition to increasing equilibrium wages.

With an upward-sloping labour demand curve, immigration to area r will not lead to falling equilibrium wages. It will rather lead to increasing equilibrium wages, because of a better exploitation of the scale economies. An upward-sloping labour demand curve is an essential feature of NEG- and NTT-models. It implies that labour migration perpetuates rather than cures regional wage disparities. From our approach in section E3 it can be concluded that the non-convergence character of labour mobility also applies to regional unemployment rates.

All in all, the message of this section is that the neoclassical hypothesis, according to which labour migration leads to regional convergence of factor prices (and unemployment rates), does not hold when the technology exhibits sufficiently strong increasing returns to scale. These can be externally or internally transmitted. Of course people migrate from low-wage to high-wage areas. But, as shown in this section, this does not imply that the interregional wage differences must vanish.

F4.2.) Selective labour migration

We now want to turn to a different argument why migration might lead to regional divergence: the issue of self-selectivity of migrants. So far we have assumed that all labour is homogenous. This means, that all workers are assumed to be identical, with no differences in skill, education etc. This of course bares very little real world relevance. It has therefore become common in economic theory to at least distinguish between two types of labour, skilled and unskilled, where members of the former group are e.g. thought of as having completed upper secondary or tertiary education. Increasingly, individual skill is even modelled as a continuous function. This modelling strategy will be adopted in the model of section F6. In the model of section F5 we will at first only distinguish between "skilled" and "unskilled" labour.

a) Evidence on selective migration

Distinguishing between different types of labour is not only important for its own sake. Additionally, there are also drawbacks between the skill level and the propensity of any single individual to be geographically mobile. In the economics literature, it has been acknowledged since decades (and is repeated over and over) that young workers tend to be more mobile than old workers, and that skilled la-

bour is more mobile across regions than unskilled labour.[5] If there is internal migration within an economy, the migrants on average tend to be educated and young individuals.

It is quite obvious that the emigration of such workers from any given jurisdiction does not relief potential local problems by relaxing competition on the labour supply side. Almost surely there are secondary, (labour) demand side oriented effects that are more pervasive. Otherwise it would be very difficult to understand e.g. the continuous moaning about the large waves of emigration from East Germany (e.g. Wenz, 2002). From a straightforward neoclassical point of view, policymakers who are interested in a convergence process of East German wages and incomes to western levels should be happy about every emigrant, because the marginal productivity and thus the factor price of labour should increase for the remaining Easterners. Yet, arguments in this spirit are hardly ever heard in the current discussion. One major reason seems to be that emigrants from East Germany in particular tend to be young and well educated individuals.

Hunt (2000) has undertaken a large scale micro-econometric study about migration from East to West Germany. She finds strong evidence for selective migration. She concludes: "*Emigrants are much younger than stayers, and conditional on age are more skilled, [...]. This youth and brain-drain suggests that emigration from the east could be a legitimate concern for policy-makers anxious about the economic viability of the Eastern region.*" (p. 28). But the skill bias is by no means restricted to the case of East Germany. Mauro/Spilimbergo (1998) present a comparable empirical results for Spain. They analyse regional mobility across 50 Spanish provinces over the time period from 1964-1992. They distinguish between 5 skill groups (from illiterate to college-educated), and find strong evidence for differences in the adjustment behaviour of skill groups to sudden labour market shocks: "*The high-skilled are found to migrate very promptly in response to a decline in local labour demand, whereas the low-skilled drop out of the labour force or stay unemployed for a long time.*" This finding is consistent with the view that an individual's willingness to be geographically mobile is a strictly increasing function of his of her skill level. Complementary evidence for Spain comes from Antolin/Bover (1997). Evidence for the case of Italy is provided by Gianetti (2001). She looks at migration patterns across Italian regions for the time period from 1980-1992, and similarly concludes that there are "*huge differences in the mobility rate of educated workers and less educated ones*" (p.24).

b) Reasons for selectivity

The skill bias in the group of internal migrants can be viewed at as a well established empirical fact. In this sub-section, we will present some theoretical reasons

[5] See amongst others Greenwood (1975:406 f.); Hughes/McCormick (1985:123); Martin (1997:245); Ritsilä/Tervo (1999:174); OECD (2000:33).

why skilled workers tend to be more mobile than unskilled workers. The intuition why younger people have a higher propensity to migrate is quite straightforward: they have a longer income stream of discounted earnings to pay-off the mobility costs.

General theoretical rationale for self-selection of migrants with respect to the skill level has been developed by Borjas (1987) and Borjas et.al. (1992). Within the framework of the Roy-Model, this literature shows that the relative income inequality in source and destination area is an important determinant for the composition of migrants from the source region. Skilled workers seek to migrate to regions with higher income inequality, whereas low skilled workers prefer to migrate to low-inequality jurisdictions.

Theoretical explanations that explicitly address the higher mobility rates of high-skilled workers usually rely on the fact that mobility gains are larger for skilled than for unskilled workers, whereas mobility costs are more or less constant across skill groups. The empirical result of Mauro/Spilimbergo (1998) indicates that skilled and unskilled workers react very differently to sudden and unexpected unemployment spells. The skilled people move to other areas, whereas the unskilled often drop out of the labour force or remain unemployed. A plausible reason is that the value of employment is relatively higher than the value of unemployment for skilled workers than for unskilled labour. The opportunity costs of being unemployed increase with the individual skill level. High-skilled workers have thus a higher incentive to change regions and take up jobs in other areas. The incentives for geographical mobility are relatively lower for unskilled workers, since the material gains from employment are lower and might not even compensate the mobility costs.

This argument has been extended and applied to a context with regional agglomeration. Moeller/Haas (2003), Moeller (2002) and Gianetti (2001) have shown that there is a skill bonus in the agglomeration wage premium. In other words, the earnings gap for observationally equivalent individuals between core regions and rural, peripheral regions is larger for skilled than for unskilled workers. Let mobility costs again be approximately skill invariant. High-skilled workers will then reveal a higher spatial mobility, since they have a higher incentive to concentrate in the industrial core regions in order to realize the agglomeration wage premium.

For most continental European countries, with their heavily regulated and unionised labour markets, also institutional factors are relevant for explaining low and selective mobility rates (Bertola, 2000). Take Germany as a typical example. As argued above in chapter C, there is virtually nil regional differentiation in contracted wages. Collective bargaining mostly applies to workers with low and medium formal skill level. Therefore, these skill groups have little gains from geographical mobility, since spatial wage differences are artificially compressed. Regional effective earnings differentials might not be substantial enough to compensate for mobility costs. Once unemployed, workers in Germany enjoy relatively

generous benefits. This also will have negative drawbacks on the individual willingness to be geographically mobile, since the pecuniary difference between employment and unemployment is institutionally compressed. This argument applies to a lesser extend to skilled workers. Not surprisingly, unemployment is mainly centred around people from the low and medium skill categories, since the rigid wage setting schemes and the generous welfare state arrangements not only rule out labour market clearing, but also distort geographical mobility.

c) The spatial effects of selective labour migration

What happens to the convergence hypothesis from above if there is selectivity among the group of internal migrants?

We will address this issue in depth in the following sections, where we derive illustrative theoretical models. But in this section, we want to point briefly to some parts of the older literature that is also concerned with the economic effects of selective migration. It is often seen at the root of a cumulative causation logic where migration amounts to a "vicious cycle". This idea actually dates back to old theorists like Myrdal (1957), Hirshman (1958) or Kaldor (1970). Myrdal (1957:27) writes: "*The localities and regions where economic activity is expanding will attract net immigration from other parts of the country. As migration is always selctive, at least with respect to the migrant's age, this movement by itself tends to favor the rapidly growing communities and disfavor the others.*" Economic geographers are inclined by this viewpoint until today. Fassmann/Mausburger (1997:190) e.g. point out that: "*Internal migration leads to the social erosion in the regions of origin and not to the automatic adjustment of region's endowment with factors of production. The origin areas lose human capital to the destination areas. This very uneven process of internal migration in terms of qualification between centre and periphery reinforces the regional economic disparities*".

However, this reasoning, which is fundamentally different from the simple neoclassical model presented in section F3, still has a hard time entering the spheres of mainstream economics. But also neoclassical models increasingly began to acknowledge the importance of heterogeneous labour. This mostly applies to the neoclassical growth theories. In the seminal growth model of Solow (1956), labour is still modelled as a homogenous input factor, and growth of per capita income is seen as an accumulation process of physical capital. In the Solow-model, an increase in the growth rate of the native labour force and immigration of workers from other countries is analytically equivalent. It has negative effects on the steady-state level of income per capita and increases the speed of convergence to the long-run equilibrium.[6] Since modelling labour as a homogenous input factor

[6] For an extensive discussion of this well known standard model see e.g. Barro/Sala-i-Martin (1999).

was increasingly seen as too simplistic,[7] the Solow-model was extended such that it explicitly accounted for the role of human capital as an important source of growth (e.g. Mankiw/Romer/Weill, 1992). In a model with human capital, there is a fundamental difference between immigration and "normal" labour force growth: immigrants can already carry human capital, whereas young natives still need to accumulate skills. This import of human capital opposes the "Solowian" effects of immigration. The more human capital is brought along by the immigrants, the less adverse is the effect on the steady state income in the destination area. If the immigrants' level of human capital is large enough, the impact can even become positive and the convergence speed is reduced (see also Walz, 2001; Barro/Sala-i-Martin 1999:285ff.). Hence, this short appraisal of growth theory shows that neoclassical models also draw more differentiated conclusions when a distinction is made between different types of labour.

F5) Selective migration in a two-region model

But let us now come back to the usual static two-region framework from this chapter. In this section we want to illustrate – under the use of a theoretical model – that the convergence character of labour mobility does not hold if only high skilled labour is moving across regions. In order to focus exclusively on the selectivity argument, we abstract for the moment from increasing returns or agglomeration forces as a source for regional divergence. Regional agglomeration, which after all is one of the main themes of this book, will re-enter the theoretical analysis in the next section, where we present a unified framework that jointly analyses agglomeration economies, selective migration and national union wage setting.

In this section, we rather consider an economic area consisting of two regions, both subject to a Cobb-Douglas technology. The only essential difference compared to section F3 is that we now consider two types of labour, skilled and unskilled workers. We first look at a model with full employment. Afterwards, we introduce unskilled unemployment that results because of regionally undifferentiated union wage setting.

F5.1.) The full employment case

Consider a nation consisting of two regions ($r = 1,2$) and two factors of production: Skilled labour H_r and unskilled labour L_r are both supplied perfectly inelastic. Each region again produces some perfectly tradable good Y_r, for which the price

[7] A summary about the major developments in growth theory, and a discussion for the reasons to depart from the Solow-model is provided by Romer (1994).

is normalized to unity. For simplicity we abstract from physical capital.[8] The production function of region r = {1,2} now reads as

$$Y_r = A_r(H_r)^{\nu}(L_r)^{1-\nu}. \tag{F.6}$$

Unskilled labour L_r is regionally immobile and is distributed equally across regions. Unskilled labour supply in either region is given by $\bar{L}_1 = \bar{L}_2 = \bar{L}$. Skilled labour on the other hand is perfectly mobile, and the total number of skilled workers is equal to $H_1 + H_2 = \bar{H}$. In the first variant of the model we assume that all labour markets work perfectly. Both types of workers are fully employed and paid according to their marginal products. The regional equilibrium wages for skilled and unskilled labour, w_r^H and w_r^L, can be written as

$$w_r^H = \nu A_r(H_r)^{\nu-1}(L_r)^{1-\nu} \tag{F.7}$$

$$w_r^L = (1-\nu)A_r(H_r)^{\nu}(L_r)^{-\nu} \tag{F.8}$$

The factor prices are fully determined by relative factor proportions and the exogenous parameters ν and A_r. Suppose that both regions initially have the same TFP-level (A_1=A_2), and human capital is distributed equally across the two regions (H_1=H_2). From (F.7) and (F.8) it follows that there is full equalization of both factor prices across the two regions. There is no incentive for high skilled workers to migrate.

Now suppose that the economy is hit by a permanent exogenous shock that only has negative impacts on the TFP-level in region 2, i.e. A_1>A_2. As an immediate effect of this asymmetric shock, factor prices for both the skilled and the unskilled decrease in region 2, since they are both directly affected by TFP. As a result, the mobile skilled labour will flow out of region 2 and towards region 1.

This has drawbacks on unskilled wages in both regions: Additional to the negative effect from the TFP-shock, the wage w_2^L decreases further as skilled labour moves out of the region. Vice versa, the wage for unskilled labour in region 1 increases as skilled workers immigrate. The reason is that unskilled labour becomes relatively more abundant in region 2, and relatively more scarce in region 1. One can also think about this impact on w_r^L in the following way: the Cobb-Douglas function, albeit being a neoclassical production function with substitutability of input factors, also features a complementary relation of input factors. An increase in the

[8] As long as the capital stock is assumed to be fixed, it only enters the production function as an exogenously given number anyway.

availability of one factor (skilled labour) also shifts up the productivity of every other factor.[9] Hence, an increase of skilled labour in region 1 produces a positive pecuniary externality for the local unskilled labour force. The opposite happens in region 2, where the number of skilled workers decreases.

For given values of ν, A_1 and A_2, the regional difference in unskilled wages is thus entirely determined by the spatial distribution of skilled labour. This spatial distribution of H_r can be derived explicitly. After the exogenous TFP-shock has occurred, high skilled workers who live in region 2 face the following trade-off: due to the higher TFP-level, it is attractive to migrate to region 1. But on the other hand, a concentration of human capital in region 1 will induce diminishing marginal productivity, which puts w_1^H under strain.[10] In equilibrium, these two forces just compensate for each other. Perfect mobility of skilled labour ensures that wages must be equal in both regions.

$$\nu A_1 (H_1)^{\nu-1} (\bar{L})^{1-\nu} = \nu A_2 \left(\bar{H} - H_1\right)^{\nu-1} (\bar{L})^{1-\nu} \tag{F.9}$$

Rearranging terms yields the equilibrium fraction of high skilled labour that concentrates in region 1 for any given magnitude of the interregional TFP-difference.

$$\frac{H_1}{\bar{H}} = \frac{1}{1+\left(A_1/A_2\right)^{1/(\nu-1)}} < 1 \tag{F.10}$$

The larger is the technological shock, the stronger is the concentration of skilled labour is region 1. The stronger is the concentration of human capital, the higher is the regional wage disparity for the low skilled workers. The relative low skilled wage of region 1 is given by

$$\frac{w_1^L}{w_2^L} = \frac{A_1}{A_2}\left(\frac{H_1}{H_2}\right)^{\nu}, \tag{F.11}$$

which in equilibrium is equal to $w_1^L / w_2^L = \left(A_1/A_2\right)^{1/1-\upsilon} > 1$. Equation (F.11) shows that the equilibrium regional disparity of low skilled wages is compounded of a

[9] More technically, the cross-partial derivative $\partial^2 Y_r / \partial L_r \partial H_r$ is positive.

[10] Note that this would not be the case in a model with increasing returns to scale, where concentration increases marginal productivity and equilibrium wages. But under neoclassical conditions, such agglomeration or pooling advantages do not exist even for high skilled labour.

direct effect from the asymmetric technological shock, and a secondary effect that stems from the relocation of high skilled workers. If all labour were immobile, the low skilled wage disparity would only be given by the technological component, i.e. $w_1^L / w_2^L = (A_1 / A_2)$. The mobility of the high skilled workforce thus leads to a stronger regional disparity of low skilled wages.

What does the simple model from this section tell about the spatial effects of internal migration? Above all it implies that it is much too simple to claim that labour mobility leads to factor price convergence, even with free trade of goods. We have shown (within a straightforward neoclassical model framework) that regional mobility of the mobile parts of the workforce, the high-skilled, leads to a regional divergence of earnings for the immobile parts, the low-skilled. The convergence hypothesis that we have discussed in section F3 is only valid for the mobile fraction of the labour force. At least in Europe, however, large parts of the labour force (the unskilled) can indeed be seen as immobile factors. For them, more migration of the high skilled from the blurring to the blooming regions does not imply factor price convergence. Those who are tied to the disadvantaged locations instead loose out from the emigration of mobile factors.

F5.2.) Union wage setting and unskilled unemployment

We now want to extend the model and introduce unemployment. But we pursue a different path than in chapter C by focussing now on the role of national union wage setting for regional unemployment outcomes. The merits of this alternative theoretical approach are discussed in greater detail in section F7.

Suppose that the economy under consideration is subject to high union power, and assume that there is no regional dispersion in contracted wages. We have argued in section C3.1.) that this reflects fairly well an essential institutional feature of the German labour market that is common also in other European countries. Introducing minimum-wage unemployment does not change the basic implication from the last section that mobility of skilled workers causes a wage divergence for the immobile unskilled workers. But furthermore, with our extended model, we can show that selective migration also leads to regional unemployment divergence.

The production function for either region is still of a Cobb-Douglas-type. The only change of the function (F.12) compared to (F.6) is that the employment of unskilled workers in region r (L_r)no longer needs to be identical with labour supply.

$$Y_r = A_r (H_r)^{\nu} (L_r)^{1-\nu}. \qquad \text{(F.12)}$$

This is so because of a contracted union minimum wage, which equally applies to both regions ($w_1^L = w_2^L = \overline{w}$). If the union wage is not binding, unskilled employ-

ment is simply equal to labour supply $\bar{L}$. Otherwise, if it is set on a non-competitive level, employment is determined by labour demand L_r^d.

$$L_r = L_r^d = \left(\frac{(1-\nu)A_r}{\bar{w}} \right)^{\frac{1}{\nu}} H_r \tag{F.13}$$

The mobile high skilled workers on the other hand are still paid competitively according to

$$w_r^H = \nu A_r (H_r)^{\nu-1} (L_r)^{1-\nu}, \tag{F.14}$$

and are therefore never unemployed. For every fixed wage $\bar{w}$ above the market clearing level, the unemployment rate in region r is given by $U_r = (\bar{L} - L_r)/(\bar{L} + H_r)$, and the disparity between region 1 and region 2 is

$$U_1 - U_2 = \frac{\bar{L} - \left((1-\nu)A_1/\bar{w}\right)^{1/\nu} H_1}{\bar{L} + H_1} - \frac{\bar{L} - \left((1-\nu)A_2/\bar{w}\right)^{1/\nu} H_2}{\bar{L} + H_2} \tag{F.15}$$

An increase in H_r lowers the unemployment rate in the respective region for two reasons: Firstly, as a pure matter of accounting, since the fully employed skilled workers are counted within the region's labour force. And secondly, the involvement of more skilled workers in the production of region r adds to the productivity of the local unskilled workers. This pecuniary external effect has already been discussed above. For any given $\bar{w}$, firms are more willing to hire unskilled labour because of the enhanced productivity.

Suppose again that both regions initially have the same TFP-level, and human capital is equally distributed across the two regions. Furthermore, suppose that there is unemployment because of a non-competitive level of $\bar{w}$. From (F.13) and (F.15) it follows that unskilled labour demand, unskilled employment, as well as unemployment rates will be identical across regions. And from (F.14) it is easy to see that this initial situation is an equilibrium, since there is no incentive for high skilled workers to change locations.

Now suppose that the economy is hit by a comparable asymmetric exogenous shock like in the last section. Only region 2 is affected negatively, i.e. $A_1 > A_2$. For the case of Germany, one might think of this shock as representing the oil crises of the 1970s, which surely affected all German areas, but the North (region 2) with its traditional industries like shipbuilding more adversely. As an immediate

effect of this shock, both unskilled labour demand and high skilled wages in region 2 decrease. But it follows from (F.14) that additional to the direct impact, w_2^H is also negatively affected by the decrease in the unskilled employment of region 2. As a result, human capital will flow out of the region 2. Unemployment will further increase thereby, whereas unemployment drops in the region 1.
Importantly, as long as there is unskilled unemployment in region 1 at the given union wage $\overline{w}$, skilled workers are not subject to diminishing marginal returns as they migrate to region 1. This can be seen by substituting (F.13) into (F.14). As long as unskilled labour is not fully employed, i.e. as long as employment is given by labour demand (F.13), the human capital remuneration can be written simply as a function of exogenous variables.

$$w_r^H = \nu\left(A_r\right)^{1/\nu}\left(\frac{1-\nu}{\overline{w}}\right)^{(1-\nu)/\nu} \tag{F.16}$$

Thus, as long as unskilled employment is given by (F.13), high skilled workers move to region 1 that has the higher TFP-level. As human capital concentrates, the marginal productivity of unskilled labour and thereby labour demand increases due to the pecuniary externality inherent to the Cobb-Douglas technology. Put differently, an increase of H_1 is not a partial, but a total factor variation as long as there are unemployed unskilled workers available who can enter the newly created jobs.
Note that we have just described a cumulative causation mechanism in this extremely simple model: some skilled workers change regions and thereby "create jobs" for unskilled workers in region 1 and destroy some in region 2. This in turn has positive drawbacks on productivity and wages of skilled workers in r = 1, negative ones in r = 2. And effectively, the motivation for human capital to concentrate in the advantaged region is even stronger than before. A small technological shock can have pronounced effects, since it starts off a vicious cycle for the disadvantaged region. To which extremity this type of circular logic is taken depends on the magnitude of the asymmetric shock, and on the prevailing wage level $\overline{w}$.

Proposition F1. *If* $\overline{w}$ *is set on a level equal or higher than* $\hat{w} = (1-\nu)A_1\left(\overline{H}/\overline{L}\right)^{\nu}$, *high skilled workers will completely concentrate in the South. If the union wage is only* $\overline{w} = \gamma\,\hat{w}$ *, with* $\gamma < 1$, *the pooling of human capital might not be complete, but at least sufficient to restore full employment in the South.*

In the case where $\overline{w} \geq \hat{w}$, the consequences of an asymmetric shock – no matter how small it may be – are most extreme. Since enough unskilled workers are unemployed to begin with, the move of human capital to region 1 is never associated

with diminishing returns, and all skilled workers move to region 1 that has the higher TFP-level. If $\bar{w} = \hat{w}$, there would be full concentration of human capital ($H_1 = \bar{H}$) and full employment for the unskilled in region 1. With $\bar{w} > \hat{w}$, there remains some unemployment in region 1 despite of full human capital concentration.

If the union wage is fixed on a lower non-competitive level $\bar{w} = \gamma\hat{w}$, human capital will flow into the South at least until full employment is reached.[11] This is the case if $L_1 = \left((1-\nu)A_1/\gamma(1-\nu)A_1(\bar{H}/\bar{L})^{\nu}\right)^{1/\nu} H_1 = \bar{L}$, which can be rewritten as

$$H_1 = \gamma^{1/\nu} \bar{H} . \tag{F.17}$$

Up to the pooling level (F.17), unskilled labour is not a binding factor. But once H_1 is as large as (F.17), skilled workers who are still located in region 2 face the trade-off that is already known from the full-employment model in the last section: Due to the higher TFP-level A_1, it is attractive to move to region 1. But since there is now full employment of unskilled workers, i.e. L_1 is given by $\bar{L}$ instead of L_1^d, human capital now faces diminishing returns.

But one can show that the human capital concentration in region 1 will be stronger in equilibrium than described by (F.17). Consider the equilibrium condition $w_1^H = w_2^H$ for the case where $L_1 = \bar{L}$. Using (F.14), this yields after some manipulations

$$\frac{H_1}{\bar{H} - H_1} = \left(\frac{A_1}{A_2}\right)^{\frac{1}{1-\nu}} \frac{\bar{L}}{L_2^d} \quad ,$$

or

$$H_1 = \frac{\left(\frac{A_1}{A_2}\right)^{\frac{1}{1-\nu}} \bar{L}}{\left(\frac{(1-\nu)A_2}{\bar{w}}\right)^{1/\nu}} .$$

We know that the prevailing wage is $\bar{w} = \gamma\hat{w}$, and we can therefore rewrite this expression as

[11] The parameter γ might reflect the union or insider power or the degree of labour market imperfection in the economy. Note further that the union can not set $\gamma < (½)^{\nu}$, because this would be the market wage with equal distribution of human capital.

$$H_1 = \left(\frac{A_1}{A_2}\right)^{\frac{1}{1-\nu}+\frac{1}{\nu}} \gamma^{1/\nu} \bar{H} \quad \leq \bar{H} \tag{F.18}$$

The higher is the difference in TFP, and the higher is the union power γ, the greater is the human capital concentration $H_1/\bar{H}$, and the wider are the regional unemployment disparities. Since $A_1 > A_2$, the equilibrium pooling level (F.18) is stronger than the full employment threshold for unskilled workers in region 1, equation (F.17). Therefore, there is excess demand for unskilled labour in region 1 at the going wage rate $\bar{w} = \gamma \hat{w}$. This can be verified by plugging the equilibrium pooling level of human capital, (F.18), in the demand function for unskilled labour, (F.13).

$$L_1^d = \left(\frac{A_1}{A_2}\right)^{\frac{1}{1-\nu}+\frac{1}{\nu}} \bar{L} > \bar{L} \tag{F.19}$$

The excess demand can push up the effective wages of unskilled labour in region 1. This has no effect on the equilibrium pooling level of high skilled workers, since this depends on the actual level of employment $\bar{L}$, not on labour demand for unskilled workers. An upward tendency of w_1^L above $\bar{w}$ is to expect, because collective bargaining wages are wage floors with allowance for an upward wage drift. The wage for unskilled labour w_1^L can lie in the range

$$\bar{w} \leq w_1^L \leq \bar{w}\left(\frac{A_1}{A_2}\right)^{\frac{1}{1-\nu}}, \tag{F.20}$$

without having any effects on unskilled employment or wages for the skilled workers. If w_1^L is equal to the upper bound of (F.20), labour demand will exactly match labour supply $\bar{L}$. But where exactly in this range the actual wage will be, is a matter of bargaining and can not be answered by this model.

F5.3.) Conclusion of the model with constant returns to scale

Let us summarize the model variant from section F5.2.). We have shown that selective internal migration results in regional divergence with respect to both wages and unemployment rates. Starting from a symmetrical initial situation, an exogenous asymmetric shock induces human capital to migrate from region 2 to region 1. As a consequence of this factor movement, jobs are created and unemployment

is decreasing for the low skilled workers in the receiving region, the opposite happens in the sending region. This is due to the mutually beneficial relationship, or complementarity of input factors within a substitutional production function. Higher unskilled employment in region 1 increases the incentives for high skilled workers to leave region 2 even more. At the end of this cumulative causation mechanism, the model predicts that human capital is either completely or partially concentrated in one region, the unemployment rate in that location is lower, and unskilled workers have higher effective earnings. The latter results holds despite of undifferentiated union wages, because excess demand for unskilled labour gives rise to a positive wage drift of effective over contracted salaries.

The fundamental insight from the model of section F5 is that labour mobility does not lead to regional convergence if migration is a selective process. Even within a very simple neoclassical model, that did not make use of any agglomeration or scale effect, migration of high-skilled labour leads to divergence of regional unemployment rates and wages for the unskilled immobile workers. There are thus two truly distinct arguments for the "non-convergence"- character of labour mobility. Firstly, convergence might be ruled out either because the underlying technology exhibits (internal or external) scale economies, so that labour does not face diminishing marginal returns. This has been demonstrated in chapter E as well as in section F4 for a situation where labour was assumed to be a homogeneous input factor. And secondly, convergence does not occur – even under a very conventional neoclassical technology – because internal migrants tend to be high-skilled workers whose emigration does not help but rather hurt the sending region.

The two arguments might be substitutes or complements. Labour mobility can result in spatial divergence either because one of the two described forces is at work, or because they are jointly operating. Various authors have put different emphasis on either of the two sources for non-convergence. Krugman (1991a) e.g. relies on increasing returns to account for the divergent trend brought about by labour mobility, but he neglects the heterogeneity of labour. Other authors, like e.g. Myrdal (1957), are more inclined with the selectivity argument. In the remainder of this chapter, we want to construct a theoretical model that finally pulls all pieces together. We will formulate a model that explicitly entails regional agglomeration and worker heterogeneity in a unified framework.

F6) Agglomeration, worker heterogeneity and national union wage setting

The model we are going to develop in this section will be suited to explain the emergence of regional disparities *endogenously*. In the model with constant returns to scale, spatial differences were due to exogenous TFP-shocks, the origin of which has essentially remained an open issue. With endogenous agglomeration economies, such exogenous ad-hoc constructions are no longer needed. In this respect, the framework from this section is again similar to the agglomeration model from chapter E, in particular the production side of the economy. Yet, some

extensions and modifications are needed in comparison to the framework of chapter E, since we want to explicitly introduce worker heterogeneity.

We keep on considering an integrated economic area consisting of two regions labelled r = {1,2}. Each region is populated now with two generations of heterogeneous agents, i.e. we construct an *overlapping generations (OLG)-model*. For simplicity we abstract both from output and population growth, and we assume that in each period a new generation L_r is born. Young and old individuals in both regions are endowed with one unit of non-leisure time and endogenously decide on the time fraction used for education. Human capital investments pay off in the old age period by expanding the available effective time budget that is then solely used for working. Heterogeneity enters insofar, as learning productivity differs across individuals depending on endowed personal ability characteristics denoted by η^i. We therefore do not only distinguish between two groups of workers, skilled and unskilled. We rather (realistically) model skill/education as a continuous variable.

Individuals are tied to their original location of birth during young age. They can, however, emigrate after the first period and spend their second lifetime period in the other region. The incentive for emigration comes from the fact that the technology is characterised by localised increasing returns to scale, i.e. there is an economic benefit from spatial concentration of workers in only one location. But labour mobility is assumed to be costly. More specifically, the migration of an individual from one region to the other imposes costs equal to m, which accrue independently of the agent's skill level.[12]

Individuals in this model thus make two endogenous choices at the beginning of their lifetime: they decide on their individual education demand, and they decide on their old age residence region. At first (in F6.1.) we will consider a full-employment version of the extended model, where the labour markets in both regions are flexible and will always clear. In section F6.2.), the issue of unemployment will again be approached in a way comparable to section F5. For both versions of the model we are basically establishing two propositions. Firstly, we show that the individual propensity to be regionally mobile increases with the personal level of education. And secondly, we will show that this selective labour migration, in combination with the increasing returns technology, will lead to a divergence process with respect to regional wages and regional unemployment rates. Labour mobility does thus not cure regional disparities, but rather makes them worse.

[12] For a discussion of this assumption see the discussion from section F4.2.)

F6.1.) The full employment case

a) Consumer behaviour

The consumer side of this model is similar to the OLG-model of Haque/Kim (1995). Let us consider an individual i who is born in region 1, and who derives utility from consumption. To keep the analysis tractable, we work with a logarithmic utility function with a time discount rate $0<\beta<1$. The regional superscript $s=\{1,2\}$ denotes the residence region of the individual at old age. If s = 1, the individual remains in her original area of birth. Otherwise (if s = 2) she emigrates after the first period. Lifetime utility $U^{i,rs}$ is given by

$$U^{i,1s} = \log c_{t,t}^{i,1s} + \beta \log c_{t,t+1}^{i,1s} \quad , \tag{F.21}$$

where $c_{t,t+1}^{i,1s}$ denotes consumption of some individual i born at time t in region 1 and residing in region $s=\{1,2\}$ at time t+1. Analogous equations apply to individuals born in region 2. $\ell_t^{i,1s}$ is the time fraction devoted to education during young age. There are no direct costs of education, but only opportunity costs for foregone earnings. There are also no financial markets and hence no savings in the model, i.e. the education choice is the only means for consumption smoothing. The budget constraints can be written as

$$c_{t,t}^{i,1s} = w_{1,t}(1-\ell^{i,1s}) \tag{F.22}$$

$$c_{t,t+1}^{i,1s} = w_{s,t+1}(1+\eta^i \ell^{i,1s}) - m_{1s} \tag{F.23}$$

The variables $w_{1,t}$ and $w_{s,t+1}$ denote the wage per effective labour unit devoted to work in the respective region and time period. Mobility costs m_{1s} arise only for individuals who choose to leave region 1 after the first period of life.

$$m_{1s} = \begin{cases} 0 & if\ s=1 \\ m>0 & if\ s=2 \end{cases} \tag{F.24}$$

Utility maximization with respect to the education choice $\ell^{i,1s}$ yields the following first-order-condition

$$\frac{c_{t,t+1}^{i,1s}}{c_{t,t}^{i,1s}} = \beta\eta^i \frac{w_{s,t+1}}{w_{1,t}} \quad . \tag{F.25}$$

The optimal consumption path is thus simply proportional to the individual's income stream. Together with the intertemporal budget constraint from (F.22) and (F.23), the optimal education choice can be computed as

$$\ell^{i*,1s} = \frac{\beta}{1+\beta} - \frac{1-(m_{1s}/w_{s,t+1})}{\eta^i(1+\beta)}, \qquad \text{(F.26)}$$

Proposition F2. *$\ell^{i*,1s}$ increases with η^i and m. It decreases with $w_{s,t+1}$, and it is greater if s=2 than if s =1.*

An evaluation of (F.26) shows that more able people spend more time on education than individuals with a low learning capability η^i. Interestingly, individuals who plan to emigrate after period t (m_{1s}=m) ceteris paribus demand more education than do people who are going to remain in region 1 also in t+1 (m_{1s}=0). The anticipation of future emigration already induces stronger educational attainment today, which is an argument close to Stark/Helmenstein/Prskawetz (1997). The effects of $w_{s,t+1}$ on the optimal learning choice of emigrants represents an income effect in the intertemporal smoothing of the lifetime earnings profile.[13]
By substituting (F.26) into the budget constraints, we can compute the optimal consumption path for given residence choices. Provided that the individual will remain in her home location r = 1 during t+1, she will reveal the following consumption profile

$$c_{t,t}^{i,11} = w_{1,t}\left(\frac{1+\eta^i}{\eta^i(1+\beta)}\right)$$

$$c_{t,t+1}^{i,11} = \frac{\beta}{1+\beta} w_{1,t+1}\left(1+\eta^i\right)$$

If she spends her second lifetime period in region 2, the consumption path is

$$c_{t,t}^{i,12} = w_{1,t}\left(\frac{1+\eta^i - m/w_{2,t+1}}{\eta^i(1+\beta)}\right)$$

[13] The negative impact of $w_{2,t+1}$ on $\ell^{i*,1s}$ actually requires that $(m/w_{2,t+1}) < 1$.

and $$c_{t,t+1}^{i,12} = \frac{\beta}{1+\beta} w_{2,t+1} \left(1 + \eta^i - m / w_{2,t+1}\right),$$

Not only education activity, but also consumption in both periods differs depending on the old age residence choice that is anticipated in the first period of lifetime. By inserting these consumption levels in the utility function (F.21), we can compute individual i's utility levels for the case that she remains in her original location ($U^{i,11}$), and for the case of emigration ($U^{i,12}$).

$$U^{i,11} = \log\left(\left[1+\eta^i\right]^{1+\beta} w_{1,t+1}{}^{\beta} \, \mathrm{K}^i\right) \tag{F.27}$$

$$U^{i,12} = \log\left(\left[1+\eta^i - m / w_{2,t+1}\right]^{1+\beta} w_{2,t+1}{}^{\beta} \, \mathrm{K}^i\right)$$

where $\mathrm{K}^i = w_{1,t}\left(\frac{1}{\eta^i(1+\beta)}\right)\left(\frac{\beta}{(1+\beta)}\right)^{\beta}$.

An individual i will reside in that region during old age that offers the higher utility level for given unit wage rates $w_{1,t+1}$ and $w_{2,t+1}$. By equating $U^{i,11}$ and $U^{i,12}$ we find after some manipulation the level of personal ability η^i at which an individual is indifferent between migrating and remaining in region r *for given unit wages*

$$\tilde{\eta}^i = \left(\frac{m / w_{2,t+1}}{1-\omega_1} - 1\right) \tag{F.28}$$

with $\omega_1 = \left(w_{1,t+1} / w_{2,t+1}\right)^{\frac{\beta}{1+\beta}}$

ω_1 measures the unit wage of region 1 relative to region 2. It can be shown that individuals with personal skills below $\tilde{\eta}^i$ derive higher utility from remaining in region 1 ($U^{i,11} > U^{i,12}$), whereas individuals with skills larger than $\tilde{\eta}^i$ are better of spending their second lifetime period in region 2. Thus, (F.28) can be understood as a *cut-off ability level* beyond which migrating to region 2 is more attractive than staying in region 1.

Individuals of course only consider to emigrate if the unit wage in region 2 is higher than in region 1, i.e. (F.28) is only a meaningful condition as long as $\omega_1 < 1$. We will assume that this condition is satisfied initially. More specifically, we take

the unit wage of region 2 as exogenously given at first instance, and assume that it is higher than the initial value of $w_{1,t+1}$. This renders an initial migration incentive. Below we will fully endogenise the unit wage of region 2. The important message for the time being, however, is that *if* region 2 offers a higher unit wage level, only the high skilled workers from region 1 will migrate.

The intuition for this result is central to this model. Note that the agglomeration wage premium is higher for single individuals the higher is the number of embodies labour units. This is so, because $w_{s,t+1}$ denotes a *unit wage*. Skilled workers therefore have a greater incentive to work in the central region than unskilled workers. The skill bonus in the agglomeration wage premium is a well established empirical fact, as it was argued above in section F4.2b). Together with the skill-invariant mobility costs, it is clear that the gains from migrating only overcompensate the costs for individuals with a high number of embodied labour units. Low skilled workers with low levels of η^i would also like to move from region 1 to region 2 in order to increase their income level. But for them, the wage premium for living in the "core" region 2 is not high enough to compensate for the mobility costs.

What fraction of each young generation L_1 has learning abilities larger than $\tilde{\eta}^i$ is a matter of the distribution of learning skills across the young population. Suppose that η^i is uniformly distributed across the L_1 individuals in the range [1;d], i.e. the least talented individual (indexed i=0) can not expand her effective labour units through education, whereas the average learning efficiency is 1+d/2. With this distribution of skills, the fraction μ of the population L_1 that is going to remain in region r is given by

$$\mu = \frac{1}{d}\left(\tilde{\eta}^i - 1\right) \tag{F.29}$$

From (F.28) and (F.29) it can be seen that emigration is attractive to a smaller fraction of the population (i.e. μ is larger), the higher is the regional unit wage $w_{1,t+1}$ relative to the (given) level in region 2 ($\overline{w}_{2,t+1}$), and the higher are mobility costs m.

b) The production side

We now turn to the production side of this economy, which is structurally almost identical to the model from chapter E. There is a single final consumption good Y_r which is produced in both regions $r = \{1,2\}$ without direct use of labour by assembling a large number of symmetrical intermediate inputs X_r. We keep the assumption of perfect competition and perfect tradability in the Y-sector, which renders the numeraire price p^Y=1 that is valid in both regions.

The interregional transport of intermediate inputs X_r on the other hand imposes the usual 'iceberg'-costs. Let the production function for Y_r in each region again be given by a symmetrical CES function that is described in equation (E.13.). Due to symmetry, we can again write the minimum unit cost function in region r = 1 as

$$G_1 = \left(N_1 (p_1)^{\frac{\theta}{\theta-1}} + N_2 (\tau p_2)^{\frac{\theta}{\theta-1}} \right)^{\frac{\theta-1}{\theta}} \qquad \text{(F.30)}$$

An analogous equation applies to region r = 2. The production of the single intermediate inputs in either region is still done by small, monopolistically competitive firms that use labour only. But here a change is needed compared to the construction introduced in section E.3.1.), since we no longer work with homogeneous workers. We assume that firms now *demand effective units of labour* ς rather than a certain number of working people. Recall from above that single individuals do now embody different numbers of labour units, depending on endowed personal ability characteristics and on the individual educational choice. The requirement of labour units necessary to produce the quantity X of an intermediate good is given by

$$\varsigma = a + bX \qquad \text{with } a>0,\ b>0 \qquad \text{(F.31)}$$

Every single intermediate will be produced by only one firm and N_1 indicates the number of active firms in region r = 1. Profit maximizing prices are a constant mark-up over marginal costs, $p_{1,t} = (b/\theta)\, w_{1,t}$, and firm's profits are driven down to zero by the entry of potential competitors, i.e. $\pi_1 = p_{1,t} X - w_{1,t} (a + bX) = 0$. Solely for notational convenience we again choose units such that $b=\theta$. By using (F.31), we can rewrite the zero profit condition in the X-sector as

$$X = \varsigma = \frac{a}{1-\theta} \qquad \text{(F.32)}$$

Equation (F.32) is analogous to (E.5.), i.e. it shows the equilibrium scale of each single firm regardless in which region it is operating. The output scale is identical with the effective demand for labour units per firm. Note, however, that firm sizes can very well differ with respect to the number of employed persons, as a firm does not care if it employs one worker with ς embodied labour units or ς workers with one labour unit each.

As before, the maximum number of intermediates that can be produced in region r is restricted by effective regional labour supply. Let $S_{r,t}$ denote labour supply in region r at time t. The number of firms and varieties is then simply

$$N_{r,t} = \frac{S_{r,t}}{X} = \frac{(1-\theta)}{a} S_{r,t}. \tag{F.33}$$

It is now straightforward to compute the equilibrium remuneration per labour unit $w_{r,t}$. This is done in the same way as in section E3.2.). The equilibrium condition of this model is that the unit costs G_r must be equal to one in both regions. Using (F.30) this yields the familiar result that regional unit wages are an increasing function of regional labour supply alone.

$$w_{r,t} = \left(T N_{r,t}\right)^{\frac{1-\theta}{\theta}} = \left(\frac{T}{X} S_{r,t}\right)^{\frac{1-\theta}{\theta}} \tag{F.34}$$

where the economic interpretation of the parameter $T = (1-\tau^{\frac{2\theta}{\theta-1}}) \Big/ (1-\tau^{\frac{\theta}{\theta-1}})$ is as in the preceding chapter E.

In this benchmark version of the model, labour supply is equal to employment. This implies that the unit wage consistent with product market equilibrium is an upward sloping function of employment. This is intuitive, as more labour supply in region 1 implies a higher number of locally produced intermediate inputs. Unit costs G_r decline, while the price p^Y remains unchanged. Temporary profits arise in the Y-sector that induce producers to enter the market. Prices for intermediates X_r are competed up, and these higher prices must completely be absorbed by higher unit labour remunerations via the zero profit condition in the X-sector.

Our assumption from above, that the unit wage in region 2 is initially higher than in region 1, can thus be rationalized by assuming the each new born generation in region 2 is larger than in region 1. In other words, r = 2 is assumed to be the larger region, which translated via the embedded scale effect in a higher regional unit wage level.

c) Effective labour supply

Due to the increasing returns technology, we have again embedded a centripetal tendency that has been intensively discussed in section E4.1.b). Hence, given the technology alone, all workers regardless of their skill level would want to concentrate in only one region. Due to the initial advantage, this pooling process will

occur in region 2.[14] However, not all workers born in r = 1 will actually change regions after the first period of lifetime, since we have also introduced an opposing "quasi-centrifugal" force in this approach. This is no longer the prevalence of ad-hoc regional preferences as in chapter E, but rather the presence of mobility costs. Since interregional migration is costly, only such individuals will migrate for whom the agglomeration wage premium at least compensates for the mobility costs. As shown in (F.28) and (F.29), these are the high skilled individuals with abilities greater than the critical level $\tilde{\eta}^i$. In other words, the workers who remain in region 1 (i.e. those who do not participate in the regional pooling process) are now the low skilled workers for whom migration does not pay off. It are no longer such individuals who feel intrinsically attached to the peripheral location as in section E3.3b).

We can now explicitly derive an expression for the effective labour supply in region 1 at time t. This variable $S_{r,t}$ does not only depend on the population size in region 1, but also on the education and migration decisions of the individuals. Labour supply at time t consists of the number of labour units that the two generations offer. For the young generation with size L_1, this is the amount of time that they do not devote to education. The old generation only has the size μL_1, since the $(1-\mu)L_1$ most talented workers spend their old age in region 2. Recall that members of the young generation reveal different education demand depending on their old age residence choice. Effective labour supply in region 1 can be written as

$$S_{1,t} = \int_{i=0}^{\mu L_1} (1-\ell_t^{i*,11}) + \int_{i=\mu L_1}^{L_1} (1-\ell_t^{i*,12}) + \int_{i=0}^{\mu L_1} (1+\eta^i \ell_{t-1}^{i*,11})$$

or

$$S_{1,t} = (1+\mu)L_1 + \int_{i=0}^{\mu L_1} \ell^{i*,11}(\eta^i - 1) - \int_{i=\mu L_1}^{L_1} \ell^{i*,12} \qquad \text{(F.35)}$$

The first term in the second row of (F.35) represents the pure population size that is constant in the steady state when μ is at its equilibrium level. The second term are the net returns to education of those who remain in region 1 also in the second period. The third term indicates the costs for region 1 that arise because later emigrants do not use their entire time budget for working. From (F.35) it can be seen that $S_{1,t}$ increases with μ for several reasons. Firstly, because the pure population

[14] One can show that in our model a "core-periphery" structure would also develop endogenously when starting from a completely symmetrical initial situation. Which region is going to be the core, and which one the periphery would be indeterminate. But the centripetal tendencies, the gains from concentration that re to exploit, are also valid if both regions are ex-ante identical, i.e. the symmetrical equilibrium is unstable.

mass is larger the fewer people emigrate to region 2. Secondly, because more people realize the returns to education in region 1. And thirdly, because fewer opportunity costs arise in region 1 for educating people whose private and social returns will be realized elsewhere.
It also becomes clear that the linkage effect that runs from labour supply to equilibrium remunerations in (F.34) can represent both a pure scale effect and a human capital externality: $S_{1,t}$ and thereby $w_{1,t}$ can be high either because many people are around ("agglomeration wage premium"), or because they embody a high number of labour units.

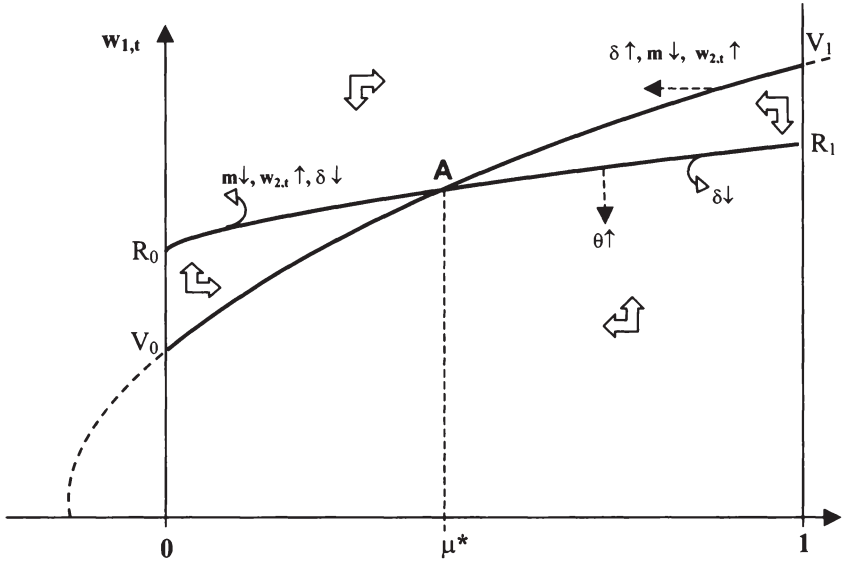
d) Equilibrium

With the model that has been outlined in this section, we have established a circular, or cumulative causation mechanism. We have derived the fraction $(1-\mu)$ of each generation L_1 that derives a higher lifetime utility by leaving region 1 after the first period of lifetime for given regional unit wages. This fraction is larger, the lower is the unit wage rate in region 1 relative to region 2. On the other hand, we have shown that the equilibrium wage in region 1 decreases, the lower is labour supply. Put differently, people leave if wages are low, and wages are low if people leave.
This circular logic in particular applies to individuals with strong learning capabilities η^i. Their emigration has a stronger bearing on region 1, firstly because they have demanded a high amount of education during young age. At time t+1, when the investment pays off both privately and socially, the high skilled workers leave the small region, which consequently foregoes the positive linkages that are associated with their human capital. Put differently, high skilled workers embody a higher amount of effective labour supply in our model. And since the individual earnings of each individual positively depend on the overall labour supply in the region of residence, the emigration of high skilled workers out of region 1 is particularly harmful for all workers who remain in that location.
The purpose of this section is to derive the equilibrium level of emigration $(1-\mu^*)$ out of region 1 for given exogenous parameter values m, θ and $w_{2,t}$. Note that we will continue to treat the wage in the core region r = 2 as an exogenous parameter in this section, and we assume that it does not change with μ.[15] An exact analytical expression of μ^* can be obtained by plugging (F.34) and (F.35) into (F.28) and (F.29). This yields a rather complicated expression that in principle can be solved

[15] With the adopted technology, the effective labour supply in region 2, $S_{2,t}$, and thereby the regional unit wage $w_{2,t}$ would endogenously increase as workers immigrate from the small region 1. We come back to this issue in below.

for μ^*.[16] For expositional purposes, however, we have chosen to pursue an instructive graphical approach, which is presented in figure F3.

Figure F3: The determination of μ^*



The locus V_0V_1 is derived from (F.29), the optimal residence choice based on consumers' utility maximization. It shows the fraction μ of the generation L_1 that chooses to remain in region 1 as a function of $w_{1,t}$ and for given parameter values. The positive slope represents the fact that μ is increasing in ω. The locus R_0R_1 represents the technological relation (F.34) and depicts equilibrium unit wages $w_{1,t}$ as a function of labour supply $S_{1,t}$, which is endogenously increasing in μ.
Within the feasible range $\mu \in \{0;1\}$, the adjustment mechanisms in this system work as follows: for points above (below) the R_0R_1 schedule, the wage $w_{1,t}$ is too high (low) for any given value of μ. Using the zero profit condition described above, the wage must realign such that it is consistent with the equilibrium locus

[16] The (highly non-linear) expression that has to be solved for μ is given by

$$\mu = \frac{1}{d\Phi_1}\left((m/\overline{w}_{2,t+1}) \Big/ \left[1 - \left[\frac{1-\theta}{a} S_{1,t}(\mu) \right]^{\frac{1-\theta}{\theta}} / \overline{w}_{2,t+1} \right] - 1 \right) - \frac{1}{d}$$

R_0R_1. This determines the phase arrows in the vertical direction. Similarly, for points to the right (left) of V_0V_1, μ is too high (low) for any given wage $w_{1,t}$. Individuals can still increase lifetime utility through changing locations, and migration will occur until μ is consistent with V_0V_1.

As long as V_0V_1 is steeper than R_0R_1, which will be the only case we consider throughout, there is a unique and stable equilibrium at point A with a spatial equilibrium configuration μ^*.[17] This level μ^* is consistent both with efficient production and with optimal residence choice.

It is instructive to look at some comparative statics and analyse changes in the exogenous parameter m, $w_{2,t}$ and θ. Changes in θ are easiest to study, since only the R_0R_1 locus is affected. The parameter θ reflects the differentiability of the single intermediate inputs in region 1 and can be understood as an inverse measure of the degree of increasing returns. The higher is θ, the lower is the equilibrium wage $w_{1,t}$ for any given value of μ. The curve R_0R_1 shifts down as θ increases, which implies that μ^* is a decreasing function of θ. There is thus more population drain from region 1 the less important are the localized increasing returns.

A change in the (exogenous) wage $w_{2,t}$ affects both curves in figure F3. The impact on V_0V_1 is obvious: if the attainable wage in region 2 increases, the incentive to leave home after the first period is larger for given values of $w_{1,t}$ and m. The curve V_0V_1 is shifted to the left. The curve R_0R_1 is also affected, however, because education demand and thereby labour supply change. This can be seen best by considering the following: the point R_1 shows the equilibrium wage $w_{1,t}$ if nobody of the young generation L_1 will emigrate after the first lifetime period. This wage can be computed as

$$w_{1,t} = \left(\frac{1}{X} \left[2L_1 + \int_{i=0}^{L_1} \ell^{i*,1} (\eta^i - 1) \right] \right)^{(1-\theta)/\theta},$$

which is independent of $w_{2,t}$. Yet, at all other points along the R_0R_1 schedule, any given fraction of later emigrants $(1-\mu)L_1$ will spend less time on education as $w_{2,t}$ increases. This consequently increases labour supply of later emigrants during their young age in region 1 and thus has positive impacts on wages $w_{1,t}$ for any given value of μ. Graphically, an increase in $w_{2,t}$ implies a clockwise rotation of R_0R_1 around the point R_1. The net effect of an increase in $w_{2,t}$ on μ^* is thus theoretically ambiguous.

A similar point applies to changes in the parameter m, the level of mobility costs. V_0V_1 shifts to the left as migration barriers are removed, because emigration is

[17] In the other case with R_0R_1 steeper than V_0V_1 the system is characterised by dynamic instability of μ^*, and will in general be driven towards a corner solution.

more attractive for given values of $w_{1,t}$ and $\overline{w}_{2,t}$. But again, a decrease in m implies a reduction in the education demand of later emigrants, thereby an increase in labour supply and thus a clockwise rotation of R_0R_1 around R_1.
Supposedly (given some numerical simulations of parameter changes in both $w_{2,t}$ and m) the "direct" effects on V_0V_1 will dominate over the effects on R_0R_1 that originates in the individuals' intertemporal substitution, but theoretically the other possibility can not be excluded.

e) The effects of labour mobility

After having described the complete model structure, we can now draw conclusions about the impact of labour mobility. Recall that the model in this section features both worker heterogeneity and thus the element of selective labour migration, as well as endogenous agglomeration economies through the presence of localised increasing returns to scale.
Not surprisingly, it is easy to see that labour mobility is a device for regional divergence. The higher is the equilibrium fraction of emigrants from region 1, $(1-\mu^*)$, the lower is the regional equilibrium wage per effective labour unit, i.e. the lower are the per capita incomes in region 1. Contrarily to chapter E, it are specifically the high skilled workers who emigrate from the periphery to the core in this chapter. This is so, because the agglomeration wage premium is greater than the mobility costs only for such workers whose ability level is greater than some critical level. But in principle, due to the localised increasing returns to scale, everyone has an incentive to move to the core region 2.
The "unfortunate" consequences of labour mobility out of the already poorer area 1 become even more pronounced when we generalize our approach and explicitly model the wage formation in the core region 2. Consumer behaviour and goods production in region 2 is structurally identical to region 1 as described above. This specifically means that the final output Y_2 is manufactured under the use of N_2 symmetrical local intermediates X_2, and that the number of firms N_2, as well as the equilibrium producer wage for each effective labour unit $w_{2,t}$ are then functions of *regional* labour supply in region 2. By applying (F.34) to region 2 we find that

$$w_{2,t} = \left(\frac{T}{X} \, S_{2,t} \right)^{\frac{1-\theta}{\theta}}$$

We have assumed that labour supply in region 2 is initially larger, e.g. because each new born generation in region 2 is larger than in region 1. This translates to a higher unit wage in region 2, and thus implies that there is a migration incentive only from region 1 to region 2. If this is warranted, we can apply the consumer

problem also to individuals from region 2 and compute the optimal education choice $\ell^{i*,22}$ as

$$\ell^{i,22} = \frac{\beta}{1+\beta} - \frac{1}{\eta^i(1+\beta)} \quad .$$

The overall labour supply in region 2, $S_{2,t}$, is given by

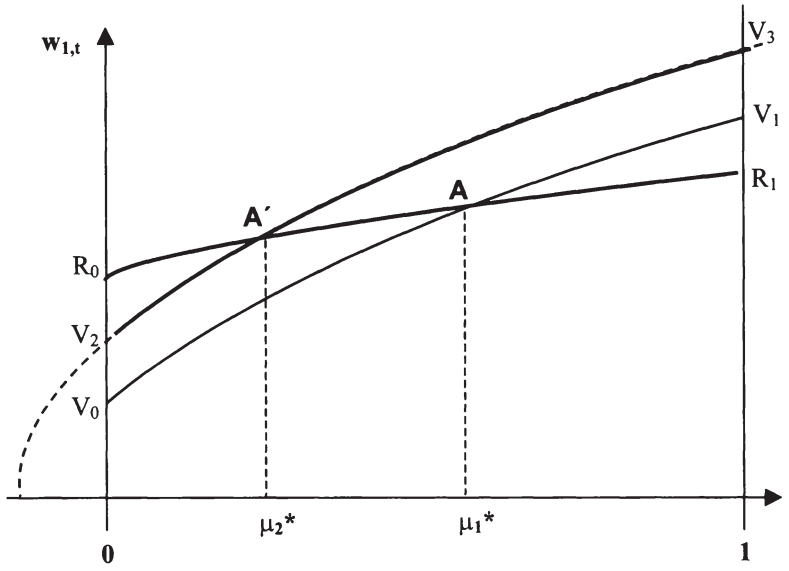
$$S_{2,t} = 2L_2 + \int_{i=0}^{L_2} \ell^{i*,22}(\eta^i - 1) + (1-\mu)L_1 + \int_{i=\mu L_1}^{L_1} \eta^i \ell^{i*,12} \,, \qquad \text{(F.36)}$$

which is an increasing function of the immigrant population $(1-\mu)L_1$. Hence, $w_{2,t}$ is not independent of μ, but it rather endogenously increases the more people emigrate from region 1. This has drawbacks on the optimal location decision of individuals from region 1 as described by (F.29). Since $w_{2,t}$ is larger the lower is μ, the actual cut-off ability level beyond which emigration starts off is actually lower than implied by figure F3, where the endogenous impact on the $w_{2,t}$ has been neglected. This effect can again be graphically illustrated in figure F4.

The graphical relation that describes utility equalization across the two regions is not given by the V_0V_1-curve. It is rather given by some curve V_2V_3 that runs strictly to the left of the V_0V_1-schedule. For any given value of $w_{1,t}$, the corresponding value of μ consistent with interregional utility equalization is strictly lower, if individuals take into account the endogenous effects of migration on the wage in region 2. Equilibrium is thus reached at A´ rather than at A. The fraction μ of the generation L_1 that remains in region 1 during t+1 is only μ_2* rather than μ_1*. If one acknowledges that the R_0R_1-schedule will rotate around R_1 in the clockwise direction (see the discussion above), the additional push for emigration out of region 1 is moderated to some extend. But the actual equilibrium value of μ* will be lower than μ_1*, and the "true" equilibrium will lie somewhere between A and A´.

The cumulative causation spiral described above is thus accentuated if we endogenise the wage formation of the core region. Any emigration out of region 1 puts the relative wage ω under strain from two sides. The true amount of brain drain has been understated by μ_1*. But other than that, the central insights from our theoretical approach remain qualitatively unchanged. Labour mobility leads to regional divergence of wages and income levels, firstly because of the underlying increasing returns technology. And secondly, this divergence is magnified by the fact that the emigrants are the high skilled, most talented workers.

Figure F4: The determination of μ^* with endogenous $w_{2,t}$



F6.2.) The case with unemployment

The final thing to do is to let unemployment re-enter the theoretical analysis. This will be done in a very simple (and admittedly artificial) way. We will also discuss this model version only briefly and mostly verbally, since it is supposedly already clear that labour migration in this approach will likewise induce regional unemployment divergence (in analogy again to section F5).

For analytical simplicity, we assume that there is a national union, which does not set a minimum *unit wage*, but rather a minimum income level per employed person. This minimum income standard $\overline{W}$ is regionally undifferentiated and applies to individuals from both regions. Firms have to pay at least this minimum income to each individual that they wish to employ. In our view it is not unrealistic to assume this type of union policy, since in the real world unions often negotiate about the growth rate of monthly incomes rather than e.g. about hourly wages (which can be seen as a proxy for unit wages in our model). The reasons why we do not study a minimum unit wage, however, are mainly technical.[18]

[18] A minimum unit wage would imply that firms in the X-sector in both regions will charge product prices p_r equal to the union unit wage. However, the zero-profit conditions and thus the product market equilibrium condition might be violated at the predetermined union level of p_r. But producers can not restore product market equilibrium by lowering the level of employment, since this

The consequences of this type of minimum income policy are straightforward. Firms in region r = {1,2} will only employ such workers who, for a given regional unit wage $w_{r,t}$, embody enough effective labour units to render a profitable employment. The other workers, who do not possess enough human capital, will have to remain unemployed because of the union's minimum standards. Unemployment will thus be centred around the low skilled individuals. More specifically, only those young individuals will be employed for which the following condition holds

$$w_{r,t}(1-\ell^{i^*,rs}) \geq \bar{W} \tag{F.37}$$

Old individuals are also not fully employed, but only if

$$w_{r,t}(1+\eta^{i}\ell^{i^*,rs}) \geq \bar{W} \tag{F.38}$$

The left-hand sides of the two inequalities represent the market income level of individual i, which is given by the regional unit wage times the number of effective labour units that individual i can offer. Individuals who do not cross this threshold have to stay (unwillingly) unemployed.

We could assume that the young unemployed individuals can use their "free time" to invest in education in order to increase the employment opportunities for the old age period, but we abstract from this mechanism again for analytical simplicity.

Recall that the unit wage in the core region 2 is higher than in the peripheral region 1. It is then clear that the periphery suffers more from this regionally undifferentiated union policy, because it is more difficult for individuals from region 1 to match the conditions (F.37) or (F.38) respectively. In other words, the peripheral location 1 will face a higher regional unemployment rate. Furthermore, the minimum income reduces employment and thus the effective labour supply in both regions (and particularly in region 1). This has, via (F.34), negative drawbacks on the equilibrium unit wage and thus on the income level of all workers, also the employed ones. In other words, the union policy not only invokes unemployment, it also depresses the earnings of employed workers because the agglomeration economies are only sub-optimally exploited.

---

would reduce the equilibrium level of the unit wage and bring the union wage claim even more out of range. In other words, equilibrium might not be reached in a set-up where the union fixes the unit wage. Problems like these do not arise, however, if we assume that the union fixes a minimum income level.

If now migration from region 1 to the relatively advantaged region 2 starts off, labour supply is reduced further in the peripheral location. The opposite happens in the central region 2. Thereby unemployment will increase further in r = 1, because the unit wage $w_{1,t}$ decreases and even fewer individuals will cross the employment threshold (F.37) or (F.38). In region 2 on the other hand, immigration will lead to an increase of $w_{2,t}$, and more individuals will become employed.
To sum up, labour migration does – not surprisingly – also lead to a divergence trend with respect to regional unemployment rates. Unemployment is entered into the model of this section in an admittedly very simplistic way. But still it is sufficient to see the consequences of regionally undifferentiated union policy for spatial unemployment disparities, and to examine the impacts of geographical labour mobility.

F7) Discussion of the theoretical models and their empirical relevance
In this final section, we want to discuss how the models from section F5 and F6 relate to the approaches that have been presented in the preceding chapters, in particular to the wage curve/agglomeration-model from chapter E. Afterwards we want to assess the empirical relevance of the models from this chapter.

F7.1.) Discussion from a theoretical point of view
The first contribution of the models from chapter F is to provide a systematic analysis about the impact of labour mobility on relative regional income levels and unemployment rates. We have shown that the neoclassical convergence hypothesis only holds under very special circumstances. Under an increasing returns technology, or with selective labour migration (or a combination of the two), internal labour migration will lead to regional divergence rather than convergence with respect to both income levels and unemployment rates. This point has been discussed at length above and shall not be repeated here.
The second objective of this chapter has been to provide an alternative theoretical analysis about regional unemployment disparities that explicitly addresses the role of *national* labour market institutions for unemployment outcomes on the *regional* level. The approach is different from chapter E insofar, as the models from this chapter are no longer wage curve models. What are the reasons for the departure from the wage curve as an analytical tool? Recall that a wage curve shows how the *local* unemployment rate has an impact on the *local* wage level. In the vein of the efficiency wage model that we have used as the micro-foundation, the wage curve relation was understood as a regional non-shirking condition: The lower is unemployment in region r, the more efficiency wages had to be paid in order to restore labour market equilibrium in that location.
In section F5, there is no downward sloping labour market equilibrium curve that would constitute an inherent link between *local* labour market variables. The bottom line message is rather that a regionally undifferentiated union wage, which

does not take into account regional productivity differences, can lead to differentiated unemployment outcomes on an intra-national level. Hence, union wage setting matters for spatial unemployment disparities. But the channel through which it matters is different from the wage curve reasoning. We have argued in section C3.1.) that in principle a wage curve relation could also be rationalized on the basis of a collective bargaining model. The intuition of such a model would be that the bargaining power of a *regional* union is a decreasing function of the *regional* unemployment rate. As a consequence, regional wage claims are lower the higher is the regional rate of joblessness. An approach like this, however, that inherently relates local bargaining power with local labour market conditions is inappropriate given the institutional structures of most European labour markets and completely misses the point. Union wage setting affects regional unemployment *precisely because* there is so little intra-national differentiation in collective bargaining wages.

This point has been made e.g. by Faini (1999). He also builds a two-region model with mobile skilled and immobile unskilled labour. Unemployment is not a feature of his model. But union wage setting plays a role, because he distinguishes two sectors, agriculture and manufacturing. Unions set wages for the unskilled in the latter sector. Faini (1999) shows that unions, which fix national wages irrespective of local productivity conditions, depress labour demand particularly in the backward region. Due to the national union wage, unskilled workers particularly from the backward location are driven to the less well paid agricultural sector. A greater acknowledgement of regional circumstances through a regionalization of the bargaining process would moderate this process. Essentially the same point can be made for our regional unemployment models from this chapter. Unemployment disparities could be smaller if the unions would comply more to regional productivity levels.

F7.2.) Discussion from an empirical point of view

Even though the models from sections F5 and F6 are no wage curve-models from the theoretical point of view, their final results and implications are somehow consistent with the predictions of the wage curve literature.

In the models from this section, the low skilled workers who live in the economically backward region are disadvantaged along two dimensions. Firstly, the unemployment rate is higher than in the blooming location. And secondly, unskilled workers in the advantaged region have higher effective earnings. This was so in the model from section F5, because there is a positive wage drift of effective earnings over contracted salaries. And similarly in section F6, the wages in the advantaged region are higher because of a better exploitation of the agglomeration economies. All in all, high regional wages and low regional unemployment rates coincide in the theoretical models from this chapter. Such a negative correlation between wage levels and unemployment rates on a regional level is the basic pre-

diction of the wage curve literature. The subsequent migration flows worsen the regional disparities with respect to both wages and unemployment rates. The "wage curve" is thus again a seemingly stable socio-economic relation that is not eroded through labour mobility.

These theoretical results are broadly consistent with the stylised facts on the spatial structure of unemployment rates and effective earnings in West Germany as described in section A5. We have shown that there are large and persistent regional unemployment disparities between the North and the South. Workers in the low-unemployment region, the South, also have higher effective earnings, even though union wages do not differ. The migration flows from the North to the South (which have been going on since decades) have apparently not led to an erosion of the existing disparities. Hence, the theoretical models from this chapter seem to be applicable to explain the regional disparities in West Germany.

These theoretical predictions are broadly consistent with the stylised facts on the relevance of agglomeration and the wage curve that have been discussed in section E4.2.), and the models are thus relevant also in a wider sense. Yet, one important qualification is needed with respect to the empirical applicability. Unemployment results because a national union fixes minimum wages on a national level without any regional differentiation. Therefore the models from this chapter are really only applicable to explain *intra-national* regional disparities. This is so because collective bargaining is often coordinated within single countries, but not yet on an international scale. It seems fair to say that contracted wages are the same in the northern and in the southern part of West Germany. But this is not true e.g. for the case of Germany vs. France, or Spain vs. Portugal. Up to now, trade unions have not spent a truly notable effort to coordinate wage claims across the single EU-countries. In principle, however, a development in this direction is very well conceivable in the future when European integration proceeds. The slogan "equal pay for equal work" that is behind the low regional differentiation of contracted wages e.g. in Germany could in principle also be extended to an European scale. The theoretical models from this section suggest that severe economic problems would be associated with such a policy that pays low attention to spatial productivity differences. However, a "Europeanization" of collective bargaining has so far not become reality.

Hence, the models from chapter F are not suited to address the trans-national unemployment clusters in the EU-15 as a whole that we have talked about in the chapters A and E. But the models are applicable to the wide and persistent intra-national unemployment and wage disparities that exist in virtually all EU countries. The fact that the main theoretical predictions from this chapter (the negative correlation of wages and unemployment rates, the influence of agglomeration, the divergence character of labour mobility) are compatible with the results from chapter E indicates, that the two broad theoretical approaches which have been introduced in this book are complementing and reinforcing each other.

Concluding remarks

In this book we have tried to understand the spatial structure of economic activity in the European Union (EU-15). We were specifically concerned with regional differences in unemployment rates that exist in the EU both within and across member countries. We have shown that the spatial configuration of unemployment rates cross European regions follows a quite distinct spatial pattern and closely resembles the core-periphery pattern of overall economic activity. On average, unemployment rates are low in the "European Banana", high in the poor, peripheral "objective 1"-areas and on an intermediate level in the regions with intermediate income per capita levels.

Theoretical research about the interrelation of regional unemployment and regional economic agglomeration is astonishingly underdeveloped, given the real world relevance of these phenomena. The class of modern agglomeration theories, like 'new economic geography' or the 'new trade theory', are mostly silent on the issue of unemployment. Regional labour market theories like the wage curve-approach on the other hand leave no room to account for regional agglomeration.

We have therefore developed own theoretical approaches. In chapter E we have formulated a model that combines the idea of a wage curve with a technology in spirit of the new agglomeration theories. In this model, there is an endogenous tendency for spatial concentration due to the presence of localised increasing returns to scale in combination with transportation costs. This technology spurs divergence of real income levels across regions. If combined with a wage curve based on efficiency wages, also regional unemployment disparities will open up. Central agglomeration areas will face a lower unemployment rate than backward peripheral locations. The unemployment disparities in turn lead to a magnification of real wage differences, since the low-unemployment area needs to pay a higher efficiency wage premium.

The model approach proposed in chapter E is able to account for the empirical evidence on the regional dimension of economic activity in the EU-15, because it predicts a core-periphery pattern of regional unemployment rates that resembles the spatial structure of overall economic agglomeration. The model highlights that labour mobility is *not* an adjustment force that leads to a gradual erosion of spatial asymmetries, like conventional neoclassical theory is implying. If the underlying production technique exhibits scale economies, internal migration leads to a perpetuation of spatial differences.

Apart from this, we identify selectivity as an additional source why labour mobility can lead to regional divergence instead of convergence. If only high skilled labour is geographically mobile, migration perpetuates rather than cures regional disparities. This is so even in a very conventional neoclassical setting with a Cobb-Douglas production function.

The constant returns to scale model that is presented in chapter F is used to analyse the implications of union wage setting for regional unemployment. We have argued that continental European labour markets are characterised by collective bargaining at the sectoral level without notable regional differentiation of union wages. Regional wage differentiation occurs through a wage drift of effective over contracted earnings. The uniformity of national union wages irrespective of regional productivity differentials, however, has implications for the spatial unemployment structure. If a single region is hit by an adverse asymmetric shock, and wages can not adjust to reflect the regional productivity level, higher unemployment results in the negatively affected region. Bad news become even worse as high unskilled unemployment spurs emigration of mobile high skilled labour out of the disadvantaged area. The blooming region on the other hand benefits from the immigration of human capital through a decrease in unemployment and an increase in effective wages for the local unskilled workforce.

Divergence trends become even more pronounced if the model structure allows both for selective labour migration and for endogenous agglomeration economies. A theoretical approach in this spirit is developed in section F6. The results of this framework can be summarized such that regional disparities get larger as migration proceeds, in particular because the group of migrants tends to include mainly high skilled workers. If a national union wage is introduced, one can show that divergence occurs not only with respect to regional wages, but also with respect to regional unemployment rates. In equilibrium, the blooming region is advantaged over the blurring one along two dimensions: it has a lower regional unemployment rate, and higher regional wages.

The model approaches from chapter F imply a spatial structure of key economic variables that is compatible with the approach from chapter E: Regional unemployment rates and wage levels are negatively correlated in equilibrium. The two theoretical models are thus complementing each other. Even though they differ to some extent with respect to their underlying assumptions and rely on different theoretical mechanisms and channels, they reach comparable conclusions, which are consistent with the observed empirical evidence. However, we have argued that the framework from chapter E is more concerned with the trans-national unemployment clusters across the European Union as a whole, whereas the models from chapter F specifically address intra-national unemployment disparities.

Throughout this book, we have largely abstained from any policy discussion and adopted a purely positive perspective. But regional economic disparities are also a great concern for policymakers. It is the declared political will that the level of territorial inequities within the EU-area shall be reduced (Art.2, Treaty of the European Community). Therefore, a vast amount of resources is spent each year by the European Commission to achieve economic and social cohesion in the EU by means of regional policy. This figure will almost surely not decrease over the

and bring along with them a dramatic increase of regional inequities within the European community.

The theoretical analysis in this book somehow supports the anxiety of policymakers that free market mechanisms alone might lead to regional divergence and more spatial inequity. Further economic integration, which e.g. leads to more mobility of labour across space, might not be an equilibrating force, but rather perpetuate existing regional disparities. But what does this purely positive statement imply for the conduct of regional policy?

We have argued at other places (Suedekum, 2002a,b) that policymakers should not draw the conclusion that centripetal agglomeration tendencies must necessarily be countervailed by means of policy. If localised increasing returns are at work in reality, there might rather be a case for policymakers to sponsor spatial economic concentration. This is so, because aggregate agglomeration advantages like those discussed in chapter D can only be realized under an asymmetric distribution of economic activity.

Furthermore, regional policy might not even be able to effectively oppose private agglomeration tendencies. As shown by Suedekum (2002b) and Martin (2000,1999), regional policies can, through secondary market adjustments, lead to actual effects that are exactly opposite to the political intentions. That is, various types of regional policy might actually lead to more instead of less territorial inequity.

The implications for public policy shall not be discussed further at this point. It should, however, be noted that one must sharply distinguish between positive and normative questions also in the context of geographical economics. This book – for good reasons, we believe – was concerned only with the positive issues.

List of References

Adams, J. (1985) "Permanent Differences in Unemployment and Permanent Wage Differentials", *Quarterly Journal of Economics* 100: 29-56.

Akerlof, G. (1982) "Labor Contracts as Partial Gift Exchange", *Quarterly Journal of Economics* 97: 543-569.

Akerlof, G., J. Yellen (1990) "The Fair Wage-Effort Hypothesis and Unemployment", *Quarterly Journal of Economics* 105: 255-283.

Alecke, B., G. Untied (2000) "Determinanten der Binnenwanderung in Deutschland seit der Wiedervereinigung", *(mimeo)*, University of Muenster.

Antolin, P., O. Bover (1997) "Regional Migration in Spain: The Effect of Personal Characteristics and of Unemployment, Wages and House Price Differentials Using Pooled Cross-Sections", *Oxford Bulletin of Economics and Statistics* 2: 215-235.

Appelbaum, E., R. Schettkat (1995) "Employment and Productivity in Industrialized Economies", *International Labor Review* 134: 605-623.

Armstrong, H., J. Taylor (1993), *Regional Economics and Policy*, London: Harvester Wheatsheaf.

Arrow, K. (1962) "The Economic Implications of Learning by Doing", *Review of Economic Studies* 29: 155-173.

Audretsch, D., M. Feldman (1996) "R&D Spillovers and the Geography of Innovation and Production", *American Economic Review* 86: 630-640.

Baldwin, R. (1999) "Agglomeration and Endogenous Capital", *European Economic Review* 43: 253-280.

Baldwin, R., R. Forslid, P. Martin, G. Ottaviano, F. Robert-Nicoud (2003), *Economic Geography and Public Policy*, Princeton: Princeton University Press.

Baldwin, R., P. Krugman (2002) "Agglomeration, Integration and Tax Harmonization", *NBER Working Paper* 9290, Cambridge (Mass.).

Baldwin, R., P. Martin, G. Ottaviano (2001) "Global Income Divergence, Trade and Industrialization: The Geography of Growth Take-Offs", *Journal of Economic Growth* 6: 5-37.

Barro, R. (1991) "Economic Growth in a Cross-Section of Countries", *Quarterly Journal of Economics* 106: 407 - 443.

Barro, R., X. Sala-i-Martin (1999), *Economic Growth*, Cambridge (Mass.): MIT Press.

Barro, R., X. Sala-i-Martin (1992) "Convergence", *Journal of Political Economy* 100: 223-251.

Bean, C. (1994) "European Unemployment: A Survey", *Journal of Economic Literature* 32: 573-619.

Bertola, G. (2000) "Labour Markets in the European Union.", *CESifo Forum* 1: 3-20.

Bils, M. (1987) "The Cyclical Behaviour of Marginal Cost and Price", *American Economic Review* 77: 838-855.

Bils, M. (1989) "Pricing in a Customer Market", *Quarterly Journal of Economics* 104: 699-718.

Bispink, R. (1999) "Tarifpolitik und Lohnbildung in Deutschland am Beispiel ausgewählter Wirtschaftszweige", *WSI Discussion Papers* 80, Duesseldorf.

Blanchard, O., S. Fisher (1989), *Lectures in Macroeconomics*, Cambridge (Mass.): MIT Press.

Blanchard, O., L. Katz (1997) " What We Know and Do not Know about the Natural Rate of Unemployment", *Journal of Economic Perspectives* 11: 51-72.

Blanchflower, D., A. Oswald (1996), *The Wage Curve*, Cambrige (Mass.): MIT Press.

Blanchflower, D., A. Oswald (1994) "An Introduction to the Wage Curve", *(mimeo)*, University College London.

Blanchflower, D., A. Oswald (1990) "The Wage Curve", *Scandinavian Journal of Economics* 92: 215-235.

Blien, U. (2001), *Arbeitslosigkeit und Entlohnung auf regionalen Arbeitsmärkten*, Heidelberg, New York: Physica.

Blien, U., K. Wolf, E. Mayerhofer, D.Vollkommer (2002) "Determinanten der regionalen Beschäftigungsentwicklung", *IAB Kurzberichte* 12: 1-8.

Boldrin, M. F. Canova (2001) "Inequality and Convergence in Europe's regions: Reconsidering European Regional Policies", *Economic Policy*, no. 32: 207-245.

Borjas, G. (1987) "Self-Selection and the Earnings of Immigrants", *American Economic Review* 77: 531-553

Borjas, G., S. Bronars, S. Trejo (1992) "Self-Selection and Internal Migration in the United States", *Journal of Urban Economics* 32: 159-185.

Brakman, S., H. Garretsen, C. van Marrewijk (2001), *An Introduction to Geographical Economics*, Cambridge: Cambridge University Press.

Brakman, S., H. Garretsen, M. Schramm (2002) " New Economic Geography in Germany: Testing the Helpman-Hanson Model", *HWWA Discussion Papers* 172, Hamburg.

Braun, J. (1993), *Essays on Economic Growth and Migration*, PhD Thesis, Harvard University.

Buettner, T. (1999), *Agglomeration, Growth, and Adjustment*, ZEW Economic Studies, Heidelberg, New York: Physica.

Burda, M., C. Wyplosz (1992) "Human Capital, Investment and Migration in an Integrated Europe", *European Economic Review* 36: 677-684.

Calmfors, L., J. Driffill (1988) "Bargaining Structure, Corporatism and Macroeconomic Performance", *Economic Policy*, no. 6: 13-61.

Canova, F., A. Marcet (1995) "The Poor Stay Poor: Non-Convergence across Countries and Regions", *CEPR Discussion Paper* 1265, London.

Card, D. (1995) "The Wage Curve: A Review", *Journal of Economic Literature* 33: 785-799.

Carlin, W., D. Soskice (1990), *Macroeconomics and the Wage Bargain. A Modern Approach to Employment, Inflation and the Exchange Rate*, Oxford: Oxford University Press.

CER (1998) "Il lavaro negli anii dell´Euro", *Rapporto CER* 3, Milano.

Chamberlin, E. (1950) "Product Heterogeneity and Public Policy", *American Economic Review* 40: 85-92.

Chipman, J. (1970) "External Economies of Scale and Competitive Equilibrium", *Quarterly Journal of Economics* 84: 347-385.

Christaller, W. (1966), *Central Places in Southern Germany*, London: Prentice Hall.

Ciccone, A. (2002) "Agglomeration Effects in Europe", *European Economic Review* 46: 213-227.

Ciccone, A., R. Hall (1996) "Productivity and the Density of Economic Activity", *American Economic Review* 86: 54-70.

Clower, R. (1965) "The Keynesian Counter-Revolution: A Theoretical Approach", in: Hahn, F., F. Brechling (eds.), *The Theory of Interest Rates*, London: MacMillan.

Combes, P., M. Lafourcade (2001) "Transport costs decline and regional inequalities: Evidence from France", *CEPR Discussion Paper* 2894, London.

Decressin, J. (1994) "Internal Migration in West Germany and Implications for East-West Salary Convergence", *Weltwirtschaftliches Archiv* 130: 231-257.

Decressin, J., A. Fatas (1995) "Regional Labor Market Dynamics in Europe", *European Economic Review* 39: 1627-1655.

De Vaal, A., M. van den Berg (1999) " Producer Services, Economic Geography, and Service Tradability", Journal of Regional Science 39: 539-572.

Dixit, A., J. Stiglitz (1977) "Monopolistic Competition and Optimum Product Diversity", *American Economic Review* 67: 297-308.

Dixon, R., A. Thirlwall (1975), *Regional Growth and Unemployment in the United Kingdom*, London: Macmillan.

Dohmen, T. (2000) "Housing, Mobility and Unemployment", *IZA Working Paper* 210, Bonn.

Dréze, J. (1975) "Existence of an Exchange Equilibrium under Price Rigidities", *International Economic Review* 16: 301-320.

Dumais, G., G. Ellison, E. Glaeser (2002) "Geographic Concentration as a Dynamic Process.", *Review of Economics and Statistics* 84: 193-204..

Elhorst, J.P. (2000) "The Mystery of Regional Unemployment Differentials: A Survey of Theoretical and Empirical Explanations", *University of Groningen Working Papers*.

Esteban, J. (2000) "Regional Convergence in Europe and the Industry Mix: A Shift-Share Analysis", *Regional Science and Urban Economics* 30:353-364.

Ethier, W. (1982) "National and International Returns to Scale in the Modern Theory of International Trade", *American Economic Review* 72: 389-405.

European Commission (2001), *Second Report on Economic and Social Cohesion*, Brussels. *available under:*
http://europa.eu.int/comm/regional_policy/sources/docoffic/official/reports/contentpdf_en.htm

Fagerberg, J., B.Verspagen (1996) "Heading for Divergence? Regional Growth in Europe Reconsidered", *Journal of Common Market Studies* 34: 431 - 448.

Faini, R. (1999) "Trade Unions and Regional Development", *European Economic Review* 43: 457-474.

Faini, R. (1983) "Cumulative Process of Deindustrialisation in an Open Region: The Case of Southern Italy 1951-1973", *Journal of Development Economics* 12: 277-301.

Fassmann, H., P. Meusburger (1997), *Arbeitsmarktgeographie. Erwerbstätigkeit und Arbeitslosigkeit im räumlichen Kontext*, Stuttgart: Teubner Verlag.

Fehr, E., U. Fischbacher (2002) "Why Social Preferences Matter - The Impact of Non-Selfish Motives on Competition, Cooperation and Incentives", *Economic Journal* 112: 1-33.

Felderer, B., S. Homburg (1994), *Makrooekonomik und Neue Makrooekonomik*, 6th ed., Berlin: Springer.

Forslid, R. (1999) "Agglomeration with Human and Physical Capital: An Analytically Solvable Case", *CEPR Discussion Papers* 2102, London.

Forslid, R., J. Haaland, K. Midelfart-Knarvik (2002) "A U-shaped Europe? A Simulation Study of Industrial Location", Journal of International Economics 57: 273-297.

Friedman, M. (1968) "The Role of Monetary Policy", *American Economic Review* 58: 1-17.

Fuhita, M., P. Krugman, A. Venables (1999), *The Spatial Economy: Cities, Regions, and International Trade*, Cambridge (Mass.): MIT Press.

Fujita, M., J. Thisse (1996) "Economics of Agglomeration", *Journal of Japanese and International Economics* 10: 339-379.

Giannetti, M. (2002) "The Effect of Integration of Regional Disparities: Covergence, Divergence, or Both?", *European Economic Review* 46: 539-567.

Giannetti, M. (2001) "Skill Complementarities and Migration Decisions", *Labour* 15: 1-31.

Gordon, R. (1990) "What is New-Keynesian Economics?", *Journal of Economic Literature* 28: 1115-1171.

Greenwood, M. (1975) "Research on Internal Migration in the United States: A Survey", *Journal of Economic Literature* 13: 397-433.

Grossman, G., E. Helpman (1994) "Endogenous Innovation in the Theory of Growth", *Journal of Economic Perspectives* 8: 23-44.

Grossman, G., E. Helpman (1991), *Innovation and Growth in the Global Economy*, Cambridge (Mass.): MIT Press.

Hall, R. (1970) "Why is the Unemployment Rate so High at Full Employment?", *Brookings Papers on Economic Activity* 3: 369-402.

Hall, R. (1972) "Turnover in the Labor Force", *Brookings Papers on Economic Activity* 3: 709-756.

Hamermesh, D. (1993), *Labor Demand*, Princeton: Princeton University Press.

Hanson, G. (2001) "Scale Economies and the Geographic Concentration of Industry", *Journal of Economic Geography* 1: 255-276.

Hanson, G. (1997) "Increasing Returns, Trade and the Regional Structure of Wages", *Economic Journal* 107: 113-133.

Haque, N., S.-J. Kim (1995) "Human Capital Flight: Impact of Migration on Income and Growth", *IMF Staff Papers* 42: 577-607.

Haufler, A., I. Wooton (2001) "Regional Tax Coordination and Foreign Direct Investment", *CeGE Discussion Paper* 14, Goettingen.

Harris, J., M. Todaro (1970) "Migration, Unemployment and Development: A Two-Sector Analysis", *American Economic Review* 60: 126-142.

Hayek, F.A. (1933), *Monetary Theory and the Trade Cycle*, London: Jonathan Cape.

Helpman, E. (1998) "The Size of Regions", in: Pines, D., E. Sadka, I. Zilcha (eds.), *Topics in Public Economics. Theoretical and Applied Analysis*. Cambridge: University Press.

Hicks, J. (1937) "Mr. Keynes and the Classics: A Suggested Interpretation", *Econometrica* 5: 147-159.

Hirschman, A. (1958), *The Strategy of Economic Development*, New Haven: Yale University Press.

Hughes, G., B. McCormick (1985) "Migration Intentions in the U.K. – Which Households Want to Migrate and Which Succeed?" *Economic Journal* 95:113-123, Conference Papers Supplement

Hunt, J. (2000) "Why do People Still Live in East Germany?", *NBER Working Paper* 7564, Cambridge (Mass.).

Isard, W. (1956), *Location and Space Economy*, Cambridge(Mass.): MIT Press.

iwd (2002) "Viel über einen Kamm: Regionale Lohndifferenzierung", *Informationsdienst der Deutschen Wirtschaft* 28, no. 40: 6-7.

Jarchow, H.-J. (1998), *Theorie und Politik des Geldes I*, 10th ed., Goettingen: Vandenhoeck & Ruprecht.

Junius, K. (1999), *The Economic Geography of Production, Trade, and Development*, Kieler Studien, Tübingen: Mohr & Siebeck.

Kaldor, N. (1970) "The Case for Regional Policies", *Scottish Journal of Political Economy* 17: 337 - 348

Keynes, J.M. (1936), *The General Theory of Employment, Interest and Money*, London: Macmillan.

Kirsch, J., R. Bispink (2002) "Allgemeinverbindliche Tarifverträge", *WSI Informationen zur Tarifpolitik*, Duesseldorf.

Koopmans, T. (1957), *Three Essays on the State of Economic Science*, New York: McGraw – Hill.

Kromphardt, J. (1991), *Konzeptionen und Analysen des Kapitalismus*, 3rd ed., Goettingen: Vandenhoeck & Ruprecht.

Krugman, P. (1999) "Was it all in Ohlin?" (mimeo), Massachusetts Institute of Technology.

Krugman, P. (1998) "Space: The Final Frontier", *Journal of Economic Perspectives* 12: 161-174.

Krugman, P. (1991a) "Increasing Returns and Economic Geography", *Journal of Political Economy* 99: 483-499.

Krugman, P. (1991b), *Geography and Trade*, Cambridge(Mass.): MIT Press.

Krugman, P. (1980) "Scale Economics, Product Differentiation, and the Pattern of Trade", *American Economic Review* 70: 950-959.

Krugman, P., A. Venables (1995) "Globalization and the Inequality of Nations", *Quarterly Journal of Economics* 60: 857-880.

Kydland, F., E. Prescott (1982) "Time to Build and Aggregate Fluctuations", *Econometrica* 50: 1345-1369.

Layard, R., S. Nickell, R. Jackman (1991), *Unemployment: Macroeconomic Performance and the Labor Market*, Oxford: Oxford University Press.

Layard, R., S. Nickell (1986) "Unemployment in Britain", *Economica* 53: 121-170.

Leijonhufvud, A. (1967) "Keynes and the Keynesians: A Suggested Interpretation", *American Economic Review* 57: 401-410.

Leijonhufvud, A. (1968), *On Keynesian Economics and the Economics of Keynes*, Oxford: Oxford University Press.

Lindbeck, A. (1992) "Macroeconomic Theory and the Labour Market", *European Economic Review* 36: 209-235.

Lindbeck, A., D. Snower (1986) "Wage Setting, Unemployment, and Insider-Outsider Relations", *American Economic Review* 76: 235-239.

Lipsey, R.G. (1960) "The Relation Between Unemployment and the Rate of Change of Money Wages in the United Kingdom 1862-1957: A Analysis", *Economica* 27: 1-31.

Loesch, A. (1954), *The Economics of Location*, New Haven: Yale University Press.

Lopez-Baso, E., Vaya, E., Mora, A., Surinach, J. (1999) "Regional Economic Dynamics and Convergence in the European Union", *Annals of Regional Science* 33: 343-370.

Lucas, R. (1988) "On the Mechanics of Economic Development", *Journal of Monetary Economics* 22: 3-42.

Lucas, R. (1980) "Methods and Problems in Business Cycle Research", *Journal of Money, Credit, and Banking* 12: 231-245.

Lucas, R. (1972) "Expectations and the Neutrality of Money", *Journal of Economic Theory* 4: 103-124.

Ludema, R., I. Wooton (2000) "Economic Geography and the Fiscal Effects of Regional Integration", *Journal of International Economics* 52: 331-357.

Magrini, S. (1999) "The Evolution of Income Dispersion among the Regions of the European Union", *Regional Science and Urban Economics* 29: 257-281.

Malinveaud, E. (1977), *The Theory of Unemployment Reconsidered*, Oxford: Basil Blackwell.

Mankiw, G., D. Romer, D. Weil (1992) "A Contribution to the Empirics of Economic Growth", *Quarterly Journal of Economics* 107: 407-438.

Marshall, A. (1890), *Principles of Economics*, London: Macmillan.

Marston, S. (1985) "Two Views of the Geographic Distribution of Unemployment", *Quarterly Journal of Economics* 100: 57-79

Martin, P. (2000) "The Role of Public Policy in the Process of Regional Convergence", *European Investment Bank Papers* 5: 69-79.

Martin, P. (1999) "Public policies, regional inequalities and growth", *Journal of Public Economics* 73: 85-105.

Martin, P. (1998) "Can Regional Policies affect Growth and Geography in Europe?", *World Economy* 21: 757-774.

Martin, P., G. Ottaviano (2001) "Growth and Agglomeration", *International Economic Review* 42: 947-968.

Martin, R. (2001) "EMU Versus the Regions? Regional Convergence and Divergence in Euroland", *Journal of Economic Geography* 1: 51-80.

Martin, R. (1997) "Regional Unemployment Disparities and their Dynamics", *Regional Studies* 31: 237-252.

Martin, R., P. Tyler (2000) "Regional Employment Evolutions in the European Union: A Preliminary Analysis", *Regional Studies* 34: 601-616.

Matsuyama, K. (1995) "Complementarities and Cumulative Processes in Models of Monopolistic Competition", *Journal of Economic Literature* 33: 701-729.

Matusz, S. (1996) "International trade, the division of labour, and unemployment", *International Economic Review* 37: 71-84.

Mauro, P., A. Spilimbergo (1999) "How Do the Skilled and the Unskilled Respond to Regional Shocks? The Case of Spain", *IMF Staff Papers* 46: 1-17.

McDonald, I., R. Solow (1981) "Wage Bargaining and Employment", *American Economic Review* 71: 896-908.

Mincer, J. (1974), *Schooling, Experience, and Earnings*, New York: Columbia University Press.

Modigliani, F. (1977) "The Monetarist Controversy or, Should we Forsake Stabilization Policies?", *American Economic Review* 67: 1-19.

Moeller, J., A. Haas (2003) "Die Entwicklung der räumlichen Lohnstruktur - Empirische Befunde für Westdeutschland im Zeitraum 1984-1997", *Jahrbuch fuer Regionalwissenschaft* 23: 55-89.

Moeller, J. (2002) "The Spatial Allocation of Skills and the Agglomeration Differential", *(mimeo),* University of Regensburg.

Myrdal, G. (1957), *Economic Theory and Under-Developed Regions*, New York: Duckworth.

Neary, P. (2001) "Of Hype and Hyperbola: Introducing 'New Economic Geography'", *Journal of Economic Literature* 39: 536-561.

Neven D., C. Gouyette (1995) "Regional Convergence in the European Community", *Journal of Common Market Studies* 33: 47-65.

Nickell, S. (1997) "Unemployment and Labor Market Rigidities: Europe versus North America", *Journal of Economic Perspectives* 11: 55-74.

Ochel, W. (2000) "Immigration Policies: Competing for Skills", *Cesifo Forum* 1 (4): 27-33.

OECD (2000) "Disparities in Regional Labour Markets", *OECD Employment Outlook* Paris, p. 31-78.

Ottaviano, G., D. Puga (1998) "Agglomeration in the Global Economy: A Survey of the 'New Economic Geography'", *World Economy* 21: 707-731.

Ottaviano, G., T. Tabuchi, J. Thisse (2002) "Agglomeration and Trade Revisited", *International Economic Review* 43: 409-435..

Ottaviano, G., J. Thisse (2002) "Integration, Agglomeration, and the Political Economics of Factor Mobility", *Journal of Public Economics* 83: 429-456.

Ottaviano, G., J. Thisse (2001) "On Economic Geography in Economic Theory: Increasing Returns and Pecuniary Externalities", *Journal of Economic Geography* 1: 153-179.

Overman, H., D. Puga (2002) "Unemployment Clusters across European Regions and Countries", *Economic Policy,* no.34:115-147.

Partridge, M., D. Rickman (1997) "Has the Wage Curve Nullified the Harris-Todaro Model? Further US Evidence", *Economics Letters* 54: 277-282.

Peeters, J., H. Garretsen (2000) "Globalisation, Wages and Unemployment: An Economic Geography Perspective", *CESifo Working Paper* 256, Munich.

Pflueger, M. (2001) "A Simple, Analytically Solvable Chamberlinian Agglomeration Model", *IZA Discussion Paper* 359, Bonn.

Phillips, A. (1958) "The Relation between Unemployment and the Rate of Change of Money Wages in the United Kingdom 1886-1957", *Economica* 25: 283-299.

Pissarides, C., I. McMaster (1990) "Regional Migration, Wages and Unemployment: Empirical Evidence and Implications for Policy", *Oxford Economic Papers* 42: 812-831.

Pred, A. (1966), *The Spatial Dynamics of U.S. Urban-Industrial Growth*, Cambridge (Mass.): MIT Press.

Puga, D. (2002) "European Regional Policies in Light of Recent Location Theories", *Journal of Economic Geography* 2: 302-334.

Puga, D. (1999) "The Rise and Fall of Regional Inequalities", *European Economic Review* 43: 302-334.

Puhani, P. (1999) "Labour Mobility - An Adjustment Mechanism for Euroland?", *ZEW Discussion Papers* 47, Mannheim.

Quah, D. (1997) "Regional Cohesion from Local Isolated Actions: Historical Outcomes", *CEPR Discussion Papers* 378, London.

Razin, A., E. Sadka (2001), *Labor, Capital, and Finance: International Flows*, Cambridge: Cambridge University Press.

Reichlin, P., A. Rustichini (1998) "Diverging Patterns with Endogenous Labor Migration", *Journal of Economic Dynamics and Control* 22: 703-728.

Reza, A. (1978) "Geographical Differences in Earnings and Unemployment Rates", *Review of Economics and Statistics* 60: 201-208.

Ritsilä, J., H.Tervo (1999) "Regional Differences and the Role of Migration in Labour Market Adjustment: The Case of Finland" in: G. Crampton (ed.), *Regional Unemployment, Job Matching, and Migration*, London: Pion, pp. 166-182.

Roback, J. (1987) "Determinants of the Local Unemployment Rate", *Southern Economic Journal* 53: 735-750.

Romer, P. (1994) "The Origins of Economic Growth", *Journal of Economic Perspectives* 8: 3-22.

Romer, P. (1990) "Endogenous Technological Change", *Journal of Political Economy* 98: 71-102.

Romer, P. (1986) "Increasing Returns and Long-Run Growth", *Journal of Political Economy* 94: 1002-1037.

Ross, M. (2002) "The Impact of Optimal Tariffs and Taxes on Agglomeration", *(mimeo)* University of Hamburg

Sala-i-Martin, X. (1996) "Regional Cohesion: Evidence and Theories of Regional Growth and Convergence", *European Economic Review* 40: 1325-1352.

Salop, S. (1979) "A Model of the Natural Rate of Unemployment", *American Economic Review* 69: 111-125.

Samuelson, P. (1952) "The Transfer Problem and Transport Costs: The Terms of Trade when Impediments are Absent", *Economic Journal* 62: 278-304.

Samuelson, P., R. Solow (1960) "Analytical Aspects of Anti-Inflation Policy", *American Economic Review* 50: 177-194.

Schlicht, E. (1978) "Labor Turnover, Wage Structure, and Natural Unemployment", *Zeitschrift für die gesamte Staatswissenschaft* 134: 337-346.

Schnabel, C. (1995) "Übertarifliche Entlohnung: Einige Erkenntnisse auf Basis betrieblicher Effektivlohnstatistiken", in: Gerlach, K., R. Schettkat (eds.), *Determinanten der Lohnbildung*, Berlin: Ed. Sigma.

Scitovsky, T. (1954) "Two Concepts of External Economies", *Journal of Political Economy* 62: 143-151.

Shapiro, C., J. Stiglitz (1984) "Equilibrium Unemployment as a Worker Discipline Device", *American Economic Review* 73: 433-444.

Solow, R. (1990), *The Labor Market as a Social Institution*, London: Basic Blackwell.

Solow, R. (1956) "A Contribution to the Theory of Economic Growth", *Quarterly Journal of Economics* 70: 65-94.

Stark, O., C. Helmenstein, A. Prskawetz (1997) "A Brain Drain with a Brain Gain", *Economics Letters* 55: 227-234.

Starrett, D. (1978) "Market Allocations and Location Choice in a Model with Free Mobility", *Journal of Economic Theory* 17: 21-37.

Stiglitz, J. (2002) "Employment, social justice and societal well-being", *International Labour Review* 141: 9-29.

Suedekum, J. (2003) "Selective Migration, Union Wage Setting and Unemployment Disparities in West Germany", forthcoming in: *International Economic Journal*.

Suedekum, J. (2002a)"Wie sinnvoll ist die Regionalpolitik der Europäischen Union?", *Zeitschrift für Wirtschaftspolitik* 51:121-141.

Suedekum, J. (2002b) "Subsidizing Education in the Economic Periphery: Another Pitfall of Regional Policies?", *HWWA Discussion Papers* 209, HWWA Hamburg.

Taylor, J., S. Bradley (1983) "Spatial Variations in the Unemployment Rate: A Case Study of North-West England", *Regional Studies* 17: 113-124.

Thuenen, H.v. (1826), *Der isolierte Staat in Beziehung auf Landwirthschaft und Nationaloekonomie : oder Untersuchungen ueber den Einfluss, den die Getreidepreise, der Reichthum des Bodens und die Abgaben auf den Ackerbau ausueben*, Hamburg: Perthes.

Tobin, J. (1977) "How Dead is Keynes?", *Economic Inquiry* 16: 459-468.

Venables, A. (1996) "Equilibrium Locations of Vertically Linked Industries", *International Economic Review* 37: 341-359.

Walz, U. (2001) "Migration, Humankapital und Wachstum", in: Franz, W., H.-J.Ramser, M. Stadler (eds.), *Wirtschaftspolitische Herausforderungen an der Jahrtausendwende*, Tübingen: Mohr & Siebeck, pp. 163-179.

Walz, U. (1995) "Growth (Rate) Effects of Migration", *Zeitschrift fuer Wirtschafts- und Sozialwissenschaft* 115: 199-221.

Weber, A. (1909), *Standort der Industrien*, Tübingen: Mohr & Siebeck.

Wenz, D. (2002) "Wieder ist die Rede vom Rübermachen: Die jungen Leute verlassen Mecklenburg-Vorpommern", *Frankfurter Allgemeine Zeitung*, Issue 162, 16/07/2002.

Wildasin, D. (2000) "Factor Mobility and Fiscal Policy in the EU: Policy Issues and Analytical Approaches", *Economic Policy,* no. 31: 337-378.

Yellen, J. (1984) "Efficiency Wage Models of Unemployment", *American Economic Review* 74: 200-205.

CeGE-Schriften

Das CeGE – Center for Globalization and Europeanization of the Economy – wurde 1999 von der Wirtschaftswissenschaftlichen Fakultät der Georg-August-Universität Göttingen gegründet. Das Zentrum dient als Forum zur internationalen und interdisziplinären Zusammenarbeit bei der Analyse ökonomischer Fragestellungen von europäischer oder globaler Bedeutung. In den CeGE-Schriften werden Forschungsergebnisse aus Dissertationen, Habilitationen und anderen Forschungsprojekten des Zentrums veröffentlicht.

Band 1 Axel Gerloff: Wechselkurspolitik in Mittel- und Osteuropa. Eine theoretische Analyse unter besonderer Berücksichtigung der Erfahrungen der zehn Kandidaten für eine EU-Osterweiterung. 2001.

Band 2 Ingo Konrad: Zur Integration ausgewählter mittel- und osteuropäischer Länder in die währungspolitische Ordnung Europas. 2002.

Band 3 Axel Hennighausen: Wegekostenfinanzierung und Lenkung im deregulierten europäischen Verkehrsmarkt. 2002.

Band 4 Wolfgang Münch: Effects of EU Enlargement to the Central European Countries on Agricultural Markets. 2002.

Band 5 Holger Niermann: Welchen Finanzausgleich braucht Europa? Ein konkreter Vorschlag zur Neugestaltung der europäischen Finanzbeziehungen mit Simulationsmodell. 2002.

Band 6 Jens Südekum: Agglomeration and Regional Unemployment Disparities. A Theoretical Analysis with Reference to the European Union. 2003.

Peter Lang · Europäischer Verlag der Wissenschaften

Wolfgang Münch

Effects of EU Enlargement to the Central European Countries on Agricultural Markets

Frankfurt/M., Berlin, Bern, Bruxelles, New York, Oxford, Wien, 2002.
242 pp., num. tab. and graph.
CeGE-Schriften. Editor in charge: Stefan Tangermann. Vol. 4
ISBN 3-631-39939-1 / US-ISBN 0-8204-6033-8 · pb. € 37.80*

Accession of the Central and Eastern European Countries is one of the biggest political projects for the European Union. As in all previous enlargements the agricultural chapter proves to be one of the most sensitive fields of the negotiations. This book examines the effects of enlargement on agricultural markets and government expenditure in the candidate countries as well as possible effects of different choices in the politically sensitive areas of direct payments and production quotas. The ten candidate countries have quite diverse economic and agricultural characteristics. Membership in the European Union, therefore, leads to varying macroeconomic effects. Two chapters specifically deal with these aspects and their effects on griculture as well as likely country-specific developments in Hungary and in Slovenia. The quantitative tools developed and used for this analysis have also been more widely used by governments and institutions in analyses of questions related to enlargement. The book describes and documents these quantitative tools.

Contents: Agricultural Markets in Central and Eastern Europe · The CAP of 1992 · The Agenda 2000 policies · Quotas and direct payments · Land retirement · Welfare effects for producers and consumers

Frankfurt/M · Berlin · Bern · Bruxelles · New York · Oxford · Wien
Distribution: Verlag Peter Lang AG
Moosstr. 1, CH-2542 Pieterlen
Telefax 00 41 (0) 32 / 376 17 27

*The €-price includes German tax rate
Prices are subject to change without notice

Homepage http://www.peterlang.de