Patterns, Processes and Causes of Economic Growth

Analyses at the Level of Firms and Regions

Kumulative Dissertation

zur Erlangung des Doktorgrades
der Naturwissenschaften (Dr. rer. nat.)

am Fachbereich Geographie (FB 19)
der Philipps-Universität Marburg

vorgelegt von:

Diplom-Geograph

Matthias Duschl

aus Ingolstadt

Marburg/Lahn 2014

Vom Fachbereich Geographie der Philipps-Universität Marburg am 26.11.2014 angenommen.

Erstgutachter:
Prof. Dr. Dr. Thomas Brenner
(Fachbereich Geographie, Philipps-Universität Marburg)

Zweitgutachter:
Prof. Dr. Uwe Cantner
(Wirtschaftswissenschaftliche Fakultät, Friedrich-Schiller-Universität Jena)

Die Disputation erfolgte am 13.05.2015 am Fachbereich Geographie der Philipps-Universität Marburg.

Hochschulkennziffer: 1180

To Valentin Hong-Jun

Acknowledgements

I would like to express the deepest appreciation to my supervisor Prof. Dr. Dr. Thomas Brenner. He contributed with ingenuous ideas, but always gave me enough freedom that I could grow intellectually. With his honest, fair- and open-minded attitude regarding scholarship and personal matters he conveyed an inspiring role model for my personal and professional development. Without his supervision, constant help and confidence in me this dissertation would not have been possible.

Writing a doctoral thesis is mostly a solitary task. The more important is a pleasant and friendly atmosphere among the closest colleagues, which was definitely the case with Dr. Anja Dettmann, Gesa Pflitsch, Dr. Sidonia von Proff and Dr. Charlotte Schlump. Thank you for your open doors and open ears. Particular thanks go to Gesa Pflitsch for proofreading my thesis and to Robert Csiscics and Daniel Lee for helping me with all my questions concerning programming and servers.

A young researcher grows on scientific guidance of the seniors. I feel fortunate to have met Prof. Giulio Bottazzi, Prof. Tom Brökel, Dr. Alex Coad and Prof. Alessio Moneta. Many ideas of this thesis are closely connected with their names. I should also like to thank the co-authors of the papers, Dr. Dennis Luxen, Prof. Shi-shu Peng, Dr. Falk Raschke, Dr. Antje Schimke and Tobias Scholl, for the uncomplicated and prolific collaborations.

I will always keep in mind the helpful and friendly colleagues which I met during my visiting research stays at the Scuola Superiore Sant'Anna, Pisa and at the Institute of Economics at Academia Sinica, Taipei. My special thanks are due to Prof. Giulio Bottazzi and Prof. Shin-Kun Peng, who hosted me at their institutes. Not to forget all other researchers who I met during the many conferences and seminars. I have much appreciated the inspiring conversations and your feedbacks on my work.

Much credit is also due to the Studienstiftung des Deutschen Volkes. With the scholarship, I gained much independence and many opportunities to meet other doctoral candidates.

Prof. Uwe Cantner has written the second report of this thesis, for which he deserves my sincerest thanks.

I close the acknowledgements with the most important people in my life - my family, my friends and my wonderful wife Ling-yu. Thank you so much for your support, sympathy, and of course, endless patience. It's great to have you all!

Matthias Duschl Marburg, October 2014

Preface

The doctoral thesis is structured as an anthology, consisting of nine chapters. The introductory chapter outlines the motivation and research questions. Moreover, it discusses the theoretical and methodological background and introduces the data used throughout the thesis. The following seven chapters consist of papers representing research that I have done on my own or with colleagues as a doctoral student. An overview on the publications, including the share of my own contribution, is provided in Table V1. The final chapter summarizes the main findings, discusses some limitations, draws conclusions for policy and science and provides an outlook on future research extensions.

Table V1: The seven papers of the cumulative dissertation

	Title of publication	Journal	Year	Co-Authors (with contribution)	Own contribution
I	Characteristics of Regional Industry-Specific Employment Growth Rates' Distributions	Papers in Regional Science	2013	Thomas BRENNER (50%)	Main author (50%)
II	Growth dynamics in regional systems of technological activities – A SVAR approach	WP[1] (submitted to: Annals of Regional Science)	2013	Thomas BRENNER (5%)	Main author (95%)
III	Firm growth and the Spatial Impact of Geolocated External factors	Jahrbücher für Nationalökonomie und Statistik	2014	Antje SCHIMKE (20%), Thomas BRENNER (20%), Dennis LUXEN(5%)	Main author (55%)
IV	Industry-specific firm growth and agglomeration	Regional Studies	2014	Tobias SCHOLL (20%), Thomas BRENNER (20%), Dennis LUXEN (3%), Falk RASCHKE (2%)	Main author (55%)
V	Firm dynamics and regional resilience: an empirical and evolutionary perspective	WP[1] (submitted to: Industrial and Corporate Change	2014	-	Single author (100%)
VI	Chinese firm dynamics and the role of ownership type	WP[2] (submitted to: Industrial and Corporate Change)	2014	Shi-shu PENG (20%)	Main author (80%)
VII	Modelling Firm and Market Dynamics – A Flexible Model Reproducing Existing Stylized Facts	WP[1] (submitted to: Journal of Industrial Economics)	2014	Thomas BRENNER (75%)	Co-author (25%)

[1] Working Papers on Innovation and Space
[2] Papers on Economics & Evolution

Abstract

This thesis studies the patterns, processes and causes of economic growth at the level of firms and regions. How are growth rates of regional economies distributed and how do these distributions emerge? What are the implications of firm growth processes on the dynamics of regional economies? Which factors external to the firm, but internal to the region drive firm growth and what is the spatial dimension of their growth effects? And how are these factors causally related to each other at the regional level? In seven research papers, these questions are investigated on basis of comprehensive empirical data and innovative statistical approaches, taking especially into account the heterogeneity of firms, industries and space.

A research gap on the stochastic characteristics of growth rates at the intermediate level between firms and countries is filled by fitting the Asymmetric Exponential Power (AEP) distribution to regional industry-specific employment data for Germany. Exploiting this additional information residing in the non-normality of the growth rates, the causal structure of economic, research, innovation and educational activities is uncovered within a structural vector autoregression framework, drawing on a unique industry-specific and multi-dimensional regional panel dataset. Introducing a conditional estimation approach of the AEP distribution in the context of the analysis of Chinese firm data, advantages from both ordinary regression and distributional approaches are combined. The role of firm dynamics for the aggregate economy is investigates by means of an empirically validated simulation models and conceptual work, in which firm dynamics assume a central role in the empirical implementation of the evolutionary perspective of regional resilience. Finally, geocoded micro-data, travel times and a flexible distance decay function approach provide more robust findings on the spatial extent of industrial clusters, unaffected from the Modifiable Areal Unit Problem, ultimately allowing for a more realistic assessment of the heterogeneous and firm-specific impacts of different kinds of agglomerations.

The results of this thesis prove empirically the importance of taking into account the firm level dynamics and the heterogeneity across industries and space for the analysis of regional economic growth. Especially the frequency of extreme growth events, both at the level of firms and regions, and their role for regional development stand out. Besides, knowledge-generating activities are an important factor for growth, although their effects strongly depend on the industry and properties of the firm. Industrial clusters, which can take quite different spatial extensions, are not necessarily beneficial for the incumbent firms populating such a cluster. Finally, many stylized facts on the statistical properties of economic growth rates found in the literature survive at different aggregation levels, like regional economies, and national contexts, like in China.

Zusammenfassung

Die Promotionsschrift untersucht die Muster, Prozesse und Ursachen von Wirtschaftswachstum auf Ebene der Firmen und Regionen. Wie sind die Wachstumsraten regionaler Ökonomien verteilt und wie entstehen diese Verteilungen? Welche Implikationen besitzen Firmenwachstumsprozesse für die Dynamiken regionaler Ökonomien? Welche Firmen-externe, aber regionale Faktoren führen zu Firmenwachstum und welche räumliche Dimension besitzen ihre Wachstumseffekte? Und wie sind diese Faktoren auf regionaler Ebene kausal miteinander verknüpft? Diese Fragen werden in sieben Forschungsbeiträgen, basierend auf umfassenden empirischen Daten und innovativen statistischen Ansätzen, näher untersucht. Hierbei wird insbesondere die Heterogenität von Firmen, Industrien und des Raums berücksichtigt.

Eine Forschungslücke zu den stochastischen Eigenschaften von Wachstumsraten auf der Meta-Ebene zwischen Firmen und Nationen wird gefüllt, indem die *Asymmetric Exponential Power* (AEP) Verteilung für regionale und industrie-spezifische Beschäftigtendaten für Deutschland geschätzt wird. Die zusätzlichen Informationen, die einer Nicht-Normalität von Wachstumsraten innewohnen, ausnutzend, wird die kausale Struktur von Wirtschafts-, Forschungs-, Innovations- und Ausbildungsaktivitäten im Rahmen eines Strukturmodells identifiziert, wobei ein einmaliger industrie-spezifischer und multi-dimensionaler regionaler Paneldatensatz zugrunde liegt. Ein konditionaler Ansatz zur Schätzung der AEP-Verteilung wird im Rahmen der Analyse chinesischer Firmendaten eingeführt, um die Vorteile gewöhnlicher Regressionsansätze mit den Vorteilen von Verteilungsschätzungen zu kombinieren. Die Rolle von Firmendynamiken für die aggregierte Ökonomie wird anhand eines empirisch validierten Simulationsmodells und einer konzeptuellen Arbeit untersucht, in welcher Firmendynamiken eine zentrale Rolle in der empirische Implementation der evolutorischen Perspektive der regionalen Resilienz einnehmen. Schließlich ermöglichen geokodierte Mikrodaten, Fahrzeiten sowie eine flexible Distanzfunktion robuste Rückschlüsse auf die räumliche Reichweite von industriellen Clustern, die nicht durch das *Modifiable Areal Unit Problem* betroffen sind und somit eine realistischere Einschätzung des heterogenen und Firmen-spezifischen Einflusses verschiedenartiger Agglomerationen erlauben.

Die Ergebnisse zeigen empirisch die Bedeutung der Berücksichtigung von Firmendynamiken und der Heterogenität von Industrien und des Raums für Analyse von regionalem Wirtschaftswachstum. Insbesondere stechen die Häufigkeit extremer Wachstumsereignisse, auf Ebene der Firmen sowie Regionen, und ihre Rolle für die regionale Entwicklung hervor. Zudem sind Wissenserzeugende Aktivitäten ein wichtiger Wachstumsfaktor, obwohl ihre Effekte stark von der Industrie und den Firmeneigenschaften abhängen. Industrielle Cluster, welche ganz unterschiedliche räumliche Ausmaße annehmen können, sind nicht unbedingt förderlich für die ein solches Cluster bewohnenden Firmen. Zuletzt überleben viele der in der Literatur bekannten *stylized facts* über die statistischen Eigenschaften wirtschaftlicher Wachstumsraten verschiedene Aggregationsebenen, wie regionale Ökonomien, und verschiedene nationale Kontexte, wie in China.

List of contents

1 Introduction ... 15

1.1 Motivation and research questions ... 15
1.1.1 Motivation ... 15
1.1.2 Research questions ... 17

1.2 The science philosophical background ... 20
1.2.1 Looking at the facts – a tribute to Herbert Simon ... 20
1.2.2 Implications from an evolutionary economics perspective ... 21

1.3 Conceptual aspects ... 27
1.3.1 The concept of growth and its empirical operationalization ... 28
1.3.2 The concept of regional technological activities ... 30
1.3.3 The concept of the firm ... 31
1.3.4 The concept of industrial clusters and agglomerations ... 33
1.3.5 The concept of space ... 35

1.4 Methodological aspects ... 36
1.4.1 Distributional analysis of growth rates ... 36
1.4.2 Distributional characteristics and regression models ... 39
1.4.3 Distance decay of agglomeration economies ... 41

1.5 Summary on the literature of growth rate distributions ... 42

1.6 Data ... 44
1.6.1 Firm data ... 44
1.6.2 Regional data ... 47
1.6.3 Travel time distances ... 50

1.7 Overview on the content of the papers ... 51

2 Characteristics of regional industry-specific employment growth rates' distributions ... 53

2.1 Introduction ... 54

2.2 Theory and predictions ... 55
2.2.1 Gibrat's law and scaling relationship ... 55
2.2.2 Firm-like behaviour of regional economies ... 57
2.2.3 Relationship between type of industry and stochastic characteristics of regional growth ... 58

2.3 Method of fitting the best theoretical distribution ... 59

2.4 Empirical data ... 60

2.5 Empirical results ... 62
2.5.1 Testing for Gibrat's law and the scaling relationship ... 62
2.5.2 Fitting of the theoretical distribution to regional industry-specific growth rates ... 65
2.5.3 Relation between type of industry and stochastic characteristics of regional growth ... 69

2.6 Conclusions ... 74

3 Growth dynamics in regional systems of technological activities – A SVAR approach ... 76

3.1 Introduction ... 77

3.2 Literature ... 78
3.2.1 Some general notes on causal discovery ... 78
3.2.2 Vector autoregression and related empirical applications ... 79
3.2.3 Causal relations in the dynamics of regional systems of technological activities ... 82

3.3 Empirical method ... 85

3.4 Data ... 89

3.5 Results ... 91

3.6 Conclusions ... 96

4 Firm Growth and the Spatial Impact of Geolocated External Factors ... 98

4.1 Introduction ... 99

4.2 Literature and hypothesis ... 100
4.2.1 Firm growth and its growth related external factors ... 100
4.2.2 The spatial dimension of external growth factors ... 104

4.3 Empirical method ... 105

4.4 Data ... 107
4.4.1 Dependent variable ... 108
4.4.2 Independent variables ... 109
4.4.3 Driving distances ... 112

4.5 Results ... 112
4.5.1 Control variables ... 112
4.5.2 Impact of firm-specific location variables ... 114

4.6 Conclusions ... 117

5 Industry-specific firm growth and agglomeration ... 118

5.1 Introduction ... 119

5.2 Literature ... 120
5.2.1 The impact of agglomeration on firm growth ... 120
5.2.2 Spatial matters of firm growth and agglomeration ... 122

5.3 Data ... 124
5.3.1 Definition of industries ... 124
5.3.2 Dependent variable ... 125
5.3.3 Independent variables ... 126

5.4 Model and estimation ... 128
5.4.1 Construction of firm-specific location variables ... 128
5.4.2 Estimation using quantile regression techniques ... 128
5.4.3 Identification of decay function parameters ... 129

5.5 Results ... 131
5.5.1 Distance decay function specification ... 131

5.5.2 Control variables ... 133
5.5.3 Impact of own-industry employment and publications ... 134
5.6 Conclusions ... 140
6 Firm dynamics and regional resilience: an empirical and evolutionary perspective ... 142
6.1 Introduction ... 143
6.2 Literature ... 144
6.2.1 Regional resilience – insights from growth processes at the level of firms ... 144
6.2.2 The distribution of firm growth rates – a regional perspective ... 146
6.2.3 Estimation of the AEP density – specificities of employment growth ... 147
6.2.4 Main propositions ... 149
6.3 Data ... 150
6.3.1 Firm level data ... 150
6.3.2 Regional level data ... 151
6.4 Methodology ... 153
6.4.1 AEP density ... 154
6.4.2 Contemporaneous left and right tail estimation ... 154
6.4.3 Regression model ... 156
6.4.4 Limitations ... 157
6.5 Results ... 158
6.5.1 Inter-regional heterogeneity in the firm growth rate distributions ... 158
6.5.2 Regional factors accounting for the fat tails ... 158
6.6 Conclusions ... 162
7 Chinese firm dynamics and the role of ownership type: A conditional estimation approach of the Asymmetric Exponential Power density ... 164
7.1 Introduction ... 165
7.2 Literature ... 166
7.2.1 A brief historical account on the ownership types in China ... 167
7.2.2 The distribution of firm growth rates ... 169
7.3 Data ... 170
7.4 Methodology ... 172
7.5 Results ... 174
7.6 Discussion ... 177
7.6.1 The variance-scaling relationship – another Chinese growth puzzle ... 177
7.6.2 The growth rate distributions of Chinese firms – the type of ownership matters ... 179
7.7 Conclusions ... 182
8 Modelling Firm and Market Dynamics – A Flexible Model Reproducing Existing Stylized Facts ... 184
8.1 Introduction ... 185
8.2 Simulation approach ... 185

8.2.1 Fundamental considerations ... 185
8.2.2 Application to firm growth ... 187

8.3 Literature on firm growth ... 188
8.3.1 Modelling firms ... 188
8.3.2 Characteristics of firm growth ... 190

8.4 Simulation model ... 191
8.4.1 Basic considerations ... 191
8.4.2 Market dynamics ... 192
8.4.3 Firm dynamics ... 194

8.5 Model calibration and implications ... 197
8.5.1 Simulation procedure ... 197
8.5.2 Model calibration ... 197
8.5.3 Identified models ... 199

8.6 Conclusions ... 202

9 Conclusions ... 204

9.1 Main findings ... 204
9.1.1 The stochastic patterns of growth processes ... 205
9.1.2 Regional growth factors ... 208
9.1.3 Spatial aspects of growth and agglomeration ... 210

9.2 Limitations of the approaches used ... 211
9.2.1 Focus on growth rates instead of growth paths ... 211
9.2.2 The lack of institutions, networks and entrepreneurship ... 212
9.2.3 An incomplete picture of heterogeneity ... 212
9.2.4 The one-dimensional conception of space ... 213
9.2.5 Endogeneity of firm growth and distances ... 214
9.2.6 Neglecting the panel structure in the quantile regression setup ... 214
9.2.7 Some limitations of the SVAR approach ... 215
9.2.8 The unconditional analysis of distributions ... 215
9.2.9 Computational issues ... 216

9.3 Implications for policy and research ... 217
9.3.1 Policy implications ... 217
9.3.2 Research implications ... 220

9.4 Outlook on future research ... 223
9.4.1 Distributional analysis of growth rates: three possible extensions ... 223
9.4.2 Breaking down heterogeneity ... 229

References ... 230

Appendix ... 255

List of figures

Figure 1: The main research questions and topics of this thesis 19
Figure 2: The concept of regional systems of technological activities 31
Figure 3: The AEP distribution for different shape parameters b_l and b_r 37
Figure 4: Growth rate densities for regions with above (black line) and below (grey line) average number of employees 39
Figure 5: The log-logistic distance decay function 41
Figure 6: Age distribution of the firms in Amadeus 45
Figure 7: Spatial distribution of firm locations 46
Figure 8: Spatial distribution of employees, R&D-related employees, patents and university graduates for the machine tool industry in 2007 according to the IAB labour market regions 48
Figure 9: Regional delimitation systems 49
Figure 10: Road distance (black line) versus beeline (grey line) from Berlin to Griesbach im Rottal 51
Figure 11: Levels of investigation 61
Figure 12: Frequency distributions of the estimated parameters b_1 and b_2 of the extended Gibrat growth equation, which is estimated separately for each industry and consecutive year pair. On the upper right corner of each histogram the percentages of significant estimates (5% significance level) are reported 62
Figure 13: Scaling relationship for two exemplary industries within a bin scaling approach (left panel) and frequency distributions of the estimated industry-specific scaling exponents β_i within a LAD scaling approach (right panel) 64
Figure 14: Frequency distribution of rescaled $g_{r,i,t}$ pooled over all years and industries 66
Figure 15: Density distributions of the estimated parameters a_l, a_r, b_l and b_r of the AEP. 67
Figure 16: Evolution of the shape parameters b_l and b_r with increased time lags for $g_{i,r,t}$ 68
Figure 17: Relationship between the average (log) number of firms per region and the shape parameters b_l and b_r for all industries 70
Figure 18: Expected causal relations in the dynamics of the activities of regional technological systems (if the causal relations are expected especially in some industries, these industries are given in brackets) 83
Figure 19: Summary of the identification algorithm 88
Figure 20: Data matching and aggregation 91
Figure 21: The contemporaneous causal and the lagged effects. Positive relationships are indicated by a solid line, negative ones by a dashed line. 94
Figure 22: Log-logistic decay function for inflection point r = 90min and different degrees of steepness s 130
Figure 23: Examples of decay parameters optimization procedure with one minimum (left) and two minima scenario (right). Minima are shown as white points, confidence interval borders are shown as white curves. 131
Figure 24: Estimated distance decay function parameters r (inflection point) and s (degree of steepness) with confidence intervals. Significant results (at 5%) are coloured black, a possible second optimum indicated by a triangle symbol. 132
Figure 25: Scatterplot of coefficients β_1 (*AGGL*) and β_2 (*KNOW*) for different quantiles. Significant results (at 5%) are indicated by the symbols. 138

Figure 26: Comparison of estimation procedures 156
Figure 27: Spatial distribution of the values for b_l and b_r 160
Figure 28: Growth rate distributions for employees 176
Figure 29: Growth rate distributions for sales 176
Figure 30: Variance-scaling relationship in China and US economies with limiting cases from tree-based hierarchical model 178
Figure 31: First step: Simulating the logical link from the model sphere to the implication sphere 186
Figure 32: Second step: Inducing from knowledge on the implication level (empirical realisations) on the set of realistic model specifications in the model sphere..... 187
Figure 33: Empirical densities of firm growth rates (black line) and pooled industry-specific regional growth rates (grey line) for the yearly growth rates between 2008 and 2010. ... 207
Figure 34: Domains of research on economic growth rates 221
Figure 35: Possible extensions of the PhD thesis..... 223

List of tables

Table 1: Literature overview on empirical studies on (unconditional) growth rate distributions.. 43
Table 2: Overview on the research questions, concepts, methods and data of the research papers ... 52
Table 3: Identification of the best fitting theoretical distribution for $g_{r,i,t}$ ... 65
Table 4: Conditional median values of the residuals and significant differences (Kruskal-Wallis test) ... 71
Table 5: Conditional median values of the residuals and significant differences (Kruskal-Wallis test) ... 72
Table 6: Conditional median values of the residuals and significant differences (Kruskal-Wallis test) ... 72
Table 7: Validity of the three identifying conditions ... 92
Table 8: Estimation results from SVAR model ... 93
Table 9: Sample composition and distributional properties of the growth rates ... 108
Table 10: Overview of independent variables and data sources ... 111
Table 11: Regression results for demographic and regional control variables ... 113
Table 12: Regression results for firm-specific location variables ... 116
Table 13: Regression results for firm- and region-specific control variables ... 116
Table 14: Literature overview ... 123
Table 15: Overview on analysed industries ... 126
Table 16: Coefficients for *AGGL* at different quantiles ... 136
Table 17: Coefficients for *KNOW* at different quantiles ... 137
Table 18: Pearson's correlation coefficients of β_1 (*AGGL*) and β_2 (*KNOW*) with the industries' age. ... 139
Table 19: Summary statistics for estimated AEP parameters ... 158
Table 20: Regression results from OLS ... 161
Table 21: Descriptive statistics ... 171
Table 22: The estimated variance-scaling parameter β ... 174
Table 23: Results from conditional AEP estimation ... 175
Table 24: Selection criteria for empirical model validation ... 199
Table 25: Realistic parameter sets ... 200
Table 26: Overview on the main findings from the papers ... 205

1 Introduction

1.1 Motivation and research questions

1.1.1 Motivation

> *"It is a fact that if you arrange the cities […] of the United States in the order of their population in 1940, the population of each city will be inversely proportional to its rank in the list. […] It is a fact that if you arrange the words that occur in James Joyce's Ulysses in the order of their frequency of occurrence in that book, the frequency of each word will be inversely proportional to its rank in the list".*

With these words Nobel-laureate Herbert SIMON introduced his little-noticed article "On judging the plausibility of theories" (1968). This article can be regarded as a plea for a scientific practice that starts its inquiry with the facts observable instead of relying on abstract assumptions that do not conform to reality. In light of the recent economic and financial crises, which also have become a crisis for economics as an academic discipline (KIRMAN 2010, STIGLITZ 2011), taking one step backward by looking at the data first would mean an important way forward.

Leaving aside all the assumptions of general market equilibria, perfectly informed and rational agents, maximising behaviour or perfect mobility in space, what do we know about economic growth of, say, regions? Are there any robust "facts" that can be told about the dynamics of regional economies? There are. It is a fact that the economic performance of regions varies more than the performance of countries (OECD 2009). It is a fact that regional growth is increasingly driven by innovation and the generation of new knowledge (FELDMAN 1994). It is a fact that growth-relevant knowledge is more easily generated in spatial agglomerations of related economic activities, that means in industrial clusters (PORTER 2003a). It is also a fact that regional growth is not equally distributed across space, implying the rise and fall of agglomerations (FUJITA 2002).

In the last two decades, the interest in regional growth and the role of geographical space for economic growth has resurged (GARRETSEN et al. 2013). This development, confirmed by influential policy reports like OECD (2009) or WORLD BANK (2009), can be directly attributed to the facts observed above, namely the political necessity to address issues such as increased regional inequalities as well as the scientific interest in the genesis of economic growth. Here, it is increasingly acknowledged that the key mechanisms seem to work at the level of regions (BREINLICH et al. 2014). This holds especially true in the modern world, where markets integrate and competition has become globalized. Firms and their regional environment now compete against other firms or regions, with the nation state and the domestic market loosing importance (FRATESI/SENN 2009).

Hence, STORPER (2011: 334) addresses both economists and geographers when he phrases the following questions: "why do regions grow? Why do some decline? What differentiates regions that are able to sustain growth from those that are not? […] What are the principal regularities in urban and regional growth?" A few paragraphs later, he

concludes that the current economists' regional growth models are based on unrealistic assumptions "because they are driven by requirements of theoretical consistency rather than from what occurs in the real world" (STORPER 2011: 335). Instead of a single canonical model of regional growth, which is "neither likely nor desirable" (BREINLICH et al. 2014: 684), the need for more realism is observed in the literature as a *mega-trend* by CAPELLO and NIJKAMP (2009). And more realism inevitably implies to look at the data first.

A closer examination of the reasons why many growth models have failed to predict the recent economic crises can be particularly revealing in this context. The focus of these models lies on the expected outcome, that is, the prediction of the average growth rate, while considering results far off the mean as random noise. When empirically implemented, these models are based on normal distributed fluctuations. With the normal distribution, which is deprived of any probability mass in the tails, extreme experiences are simply ruled out by assumption (BEINHOCKER 2007, TALEB 2012). At least at the level of firms and national economies, vast empirical evidence exists that the normal distribution provides an invalid description of growth processes (e.g. STANLEY et al. 1996, AMARAL et al. 2001, BOTTAZZI et al. 2001, FAGIOLO et al. 2008). If we are able to reject the normal distribution, a natural starting point is to look at more realistic distributional models. For firms and national economies it is found that the Asymmetric Exponential Power (AEP) distribution can describe the empirical growth rates adequately (BOTTAZZI/SECCHI 2011).

This doctoral thesis aims to fill the research gap at the intermediate level of regional economies. Moreover, a precise description of the stochastic patterns of growth rates enables us to better understand the processes and causes of regional growth. Here, two streams of literature have dominated the debate in science and policy in recent years: on the one hand, the role of agglomerations and industrial clusters (FRENKEN et al. 2014), and on the other hand, especially in the light of the recent macroeconomic and financial crises, the concept of regional resilience (MARTIN/SUNLEY 2014). This thesis seeks to contribute to both of them by acknowledging the empirically observed growth patterns in space, and in particular by studying firm growth as the main ingredient for the aggregate regional performance. The analysis of the underlying micro-sources of growth has become a recent focus in economic research (ACEMOGLU et al. 2012, AGHION et al. 2013, GABAIX 2011, HARRIS 2011, OTTAVIANO 2011). Therefore, this thesis finally sketches a new empirically validated simulation model on the emergence of the growth rate distributions which is rooted in basic mechanisms and processes at the level of firms.

This thesis relies on a data-driven approach to establish stylized facts on the growth of firms and regions. Therefore, the chosen methodology is motivated by new access to data and computational power. Phrasing the euphoric words of MILLER (2010: 181), "We have access to an unprecedented amount of fine-grained data on cities, transportation, economies, and societies, much of these data referenced in geo-space and time. There is a tremendous opportunity to discover new knowledge about spatial economies that can inform theory and modelling in regional science". For this undertaking, methods from disciplines like econophysics or machine learning are employed to explore the "rich statistical structure in the dynamics of business firms and industries" (DOSI et al. 2010: 1872).

The lesson to be drawn from this is that first of all the patterns, processes and causes of how firms and regions grow must be understood. Only then we might try to explain the underlying mechanisms of why they grow (HUGGINS/THOMPSON 2014). The main research questions are outlined in the following section in more details.

1.1.2 Research questions

This thesis asks *how* firms or regions grow. In more precise terms, *how* do economic activities, organized in regionally embedded firms, unfold over time, and *how* do they depend on other activities, like education and research in universities or public research institutes, or on the economic activities within other private firms? And if all these activities, which can always be traced back to specific locations in space, matter, *how* does their impact on growth decay with geographical distance? *How* is the way firms grow related to the performance and development of the aggregate region of their location? And finally, *how* do the observed patterns of growth emerge?

> *By providing answers to these 'how' questions, new light is shed on the patterns, processes and causes of economic growth.*

A preceding analysis of the patterns of growth, mainly the analysis of the stochastic characteristics of the growth rates, opens the door towards a deeper understanding of growth processes. What is the likelihood of a region or a firm within that region to grow or decline at a specific rate? Which distributional model is able to adequately describe the empirically observed growth rates, and how do the growth rate distributions depend on the kind of industry or regional factors, like the regions' knowledge base or industrial structure? The literature still lacks of a systematic account on the distributional characteristics of regional growth rates, both of the aggregate regional economy as well as of the elementary processes at the level of firms within a region. This is despite the widespread evidence that some regions, or firms therein, perform exceptionally well whereas others do very poorly, thus discrediting the normal distribution as an adequate model for describing economic growth processes independent of the spatial level.

The question on the stochastic patterns of regional growth processes aims at establishing new stylized facts on how economic activities unfold in space over time. Beyond this, it is interesting to ask how these activities unfold in relation to other activities that take place within the region, like research, innovation or education activities. This question becomes a search for the causal effects among the activities, which represents a main research challenge in the literature on regional economic growth (HARRIS 2011). Empirical research within the framework of growth econometrics (DURLAUF et al. 2005) or growth accounting (BARRO 1999) has had difficulties in going beyond mere correlations by revealing the direction of causality. Natural or controlled experiments are in the social sciences in general, and in the context of economic growth in particular, rarely feasible (HOYER et al. 2009). Some profound knowledge on the stochastic patterns of growth rates in

combination with new causal discovery methods, however, can help to uncover causal relationships even from uncontrolled data (BROCK 1999, HOYER et al. 2009).

The study of the patterns, processes and causes of how firms or regions grow is strongly interwoven with the role of agglomerations and industrial clusters (BREINLICH et al. 2014, FRENKEN et al. 2014). To address this issue in more depth, this thesis focuses on research questions which directly follow from the critique that the corresponding literature is often confronted with. The critique concerns the inconclusiveness and robustness of the findings on the effects of agglomerations and industrial clusters on regional economic performance (for a recent discussion see, amongst others, VANOORT 2014). Most of these studies show ambiguities in the level of aggregation, the measurement of clusters, or in the use of the empirical methodology (MAINE et al. 2010, BEAUDRY/SCHIFFAUEROVA 2009). Studies that take into account the heterogeneity of firms, industries and space are rare, but, as pointed out by VANOORT (2014: 9), are "paramount for future research in this field", because "different firms may be influenced differently by different dimensions of agglomeration". Strongly related hereto is the issue of the geographical dimension of industrial clusters and the often only implicitly assumed underlying mechanisms. Many authors (e.g. BALDWIN et al. 2010, BASILE/USAI 2012, DRUCKER 2012, MCCANN/SHEFER 2005, MCCANN/VANOORT 2009) consider the spatial extent and decay of agglomeration economies like knowledge spillovers as a central current research question for regional science and economic geography. As PORTER (1998: 202) already noted, "The strength of these 'spillovers' and their importance to productivity and innovation determine the ultimate boundaries" of an industrial cluster, making its delimitation in space an empirical question.

Hence, this thesis asks: how is the growth of heterogeneous firms affected by the various sources of agglomeration economies, like the presence of related firms or research activities in universities and public research institutes? Are industrial clusters more beneficial to smaller or larger firms or to firms from certain kinds of industries? How do the effects of agglomerations on growth decay with distance, that is, which distance function can describe their spatial decay most adequately? To study the impact of agglomerations and industrial clusters on firm performance, also the stochastic patterns of firm growth rates should be taken into account. Put differently, how are the best and worst performing firms residing in the tails of the growth rate distribution affected by their spatial surrounding in comparison to the average growing firm?

Regional resilience is another important aspect of the regional economies' growth dynamics which has attracted increasing interest by scientists and policy makers mainly due to the recent financial and economic crises (MARTIN 2012). The literature on regional resilience is subject to criticism from a similar perspective as pointed out above in case of industrial clusters. The regional aggregation level has turned out to be inappropriate, and the few quantitative approaches at this level (e.g. FINGLETON et al. 2012) are only capable of measuring the resistance or recovery dimension of regional economic systems. Rather, structural change and adaptive processes are behind the core idea of evolutionary resilience (REGGIANI et al. 2002, SIMMIE/MARTIN 2010, BOSCHMA 2014). These processes, however, work at the level of firms (BOSCHMA/FRENKEN 2006). Theoretical (Metcalfe/Foster 2010) and empirical evidence (BRAVO-BIOSCA et al. 2013, COAD et al. 2014) strongly suggest that firm level dynamics and turbulences are an indicator for structural change and adaptation, possessing a significant impact on the long-run performance of regional

economies. Hence, the way how the regions' firms grow is an important component of a more fundamental and evolutionary understanding of regional resilience. In the words of HOLM and OSTERGAARD (2013: 14): "Future studies should focus on identifying regional sources of resilience, looking at the micro-dynamics within regions to determine the adaptability of the firms and analysing why some regions fail to be resilient."

To gain a deeper understanding of the elementary growth processes at the level of firms, it is first of all important to learn which factors shape the observed growth rate distributions. Are there significant differences in the patterns of their growth rates across regions and to which regional aspects can these differences be attributed? Are there important firm-specific aspects, like firm size, innovation intensity, ownership type or financial constraints, as has been shown by BOTTAZZI et al. (2014), which have sizeable influence on the distribution's shape? This raises also methodological questions of how to explicitly account for the heterogeneity of firms in a distributional analysis of their growth dynamics.

Ultimately, there stands the question of how the growth rate distributions of regional economies emerge. It seems obvious to state that regional growth results from elementary processes at the level of firms. The existing theoretical or mathematical models, which mainly can be found in the discipline of New Economic Geography, however, do not take fully into account the knowledge about patterns and processes of firm growth. What does this knowledge, represented in well-established stylized facts, reveal about the role of industrial and regional characteristics for the emergence of regional growth rate distributions? How do firm growth rates interact and aggregate up in specific regional environments to yield the observed regional growth rate distribution? One way to answer this question is to calibrate a simulation model that builds upon basic and general assumptions about the processes and mechanisms at the firm level using the empirical knowledge about firm growth (BRENNER/WERKER 2007).

The various research questions of this thesis, which are all concerned with patterns, processes and causes of economic growth, are summarized and depicted in Figure 1. They encompass the topics of regional growth rate distribution, causal effects in growth dynamics, aspects like regional resilience and agglomerations, and finally, the emergence of growth rate distributions. These topics are all addressed by asking the "how" question.

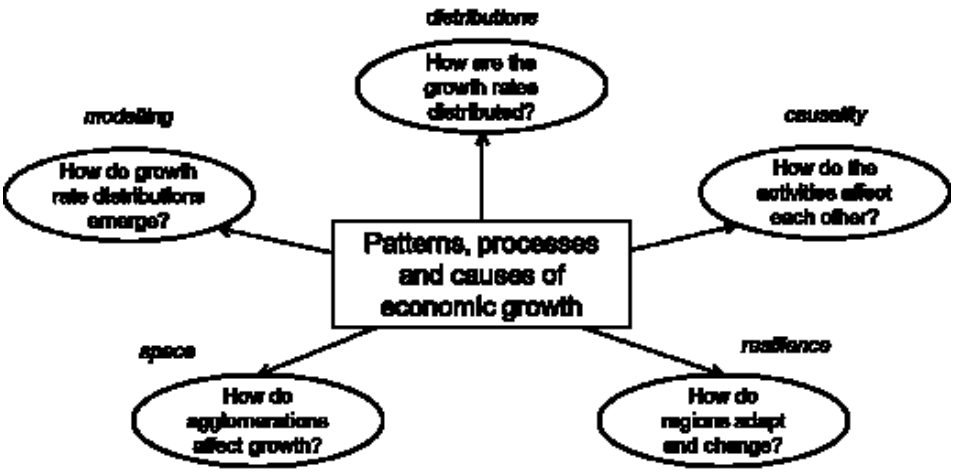


Figure 1: The main research questions and topics of this thesis

1.2 The science philosophical background

This section sheds more light on the science philosophical background framing the thesis. It argues that the research questions outlined above should be approached by *looking at the facts* first. This corresponds to an evolutionary economics research agenda, from which four implications central to the thesis are derived – the importance of disaggregating the economy, the advantages of a distributional analysis, the role of knowledge in the economy, and the concept of relatedness.

1.2.1 Looking at the facts – a tribute to Herbert Simon

Herbert SIMON (1968) proposed a strategy for scientific inquiry which can be summarized as follows. First, simple generalizations should be identified that approximately describe the facts. Secondly, limiting conditions should be constructed under which deviations of facts from generalizations decrease. Finally, the researcher should provide explanations why the generalizations might fit the facts.

This strategy implies that one should "start from data and carefully investigate their regularities while, at the same time, abstaining himself, as far as possible, from introducing untested theoretical hypothesis" BOTTAZZI (2008: 7). This was already anticipated long ago by VEBLEN (1900: 252-253), who wrote in an extensive study on the preconceptions of economic sciences: "But, since a strict uniformity is nowhere to be observed at the first hand in the phenomena with which the investigator is occupied, it has to be found by a laborious interpretation of the phenomena and a diligent abstraction and allowance for disturbing circumstances, whatever may be the meaning of a disturbing circumstance where causal continuity is denied". This perspective assumes that "all things ultimately tend towards (even if they are temporarily disturbed from) ends of patterns that the common sense of any era holds to be valuable or worthy" (LAWSON 2013: 964).

Such patterns or regularities in the phenomena of interest, which can be carved out from the data as stylized facts, are then useful for guiding and disciplining theory formation, because they provide information on the underlying data generation process (BROCK 1999). Hence, formulating explanatory theories becomes an attempt to explain the stylized facts that emerge in a complex world (COAD 2009). This means to explain empirical regularities by discovering "a set of simple mechanisms that would produce the former in any system governed by the latter" (SIMON 1968: 445). Usually, "striking features of the data […] provide for a simple generalization that summarizes them – approximately" (SIMON 1968: 457). Hypotheses can then, of course, be derived from these generalizations, which make predictions in other domains and, in the tradition of POPPER (1973), can be tested by falsification. Put differently, SIMON (1968) proposes a close interaction between hypotheses generation and data exploration for building and testing theories.

Pursuing this strategy is especially recommended in the context of economic growth (BOTTAZZI 2008, COAD 2009). It is hard to explain why firms, regions or countries grow without sound knowledge of how they grow. If they grow in different ways, the reasons

leading to growth and the outcomes thereof will also be quite different for heterogeneous entities (DELMAR et al. 2003). Hence, starting from pure theory is not helpful here. Various theories seem to be applicable and growth theories are known to be open ended, such that the truth of one theory does not automatically contradict the truth of another one (DURLAUF 2001). However, studying the patterns in the data can only be the starting point of a scientific inquiry on economic matters. From the identified regularities alone it is difficult to infer on the mechanisms and causalities (BROCK 1999, GABAIX 2009). By conditioning the patterns on other variables, or by comparing systematically specific circumstances in which these patterns materialize, more can be learnt on the underlying mechanisms. Finally, new methods have been developed in machine learning (e.g. SHIMIZU et al. 2006) to uncover from observable patterns also causal relationships.

This strategy of scientific inquiry departs from the neoclassical approach of modelling, which is rooted in methods of mathematical modelling based on theoretical assumptions. However, as LAWSON (2013: 953) criticizes, "the sorts of conditions under which the modelling methods economists have employed would be useful are found to be rather uncommon, and indeed unlikely, occurrences in the social realm. Alternatively put, the ontological presupposition of the heavy emphasis on mathematical modelling does not match the nature of the 'stuff' of the social realm". For LAWSON (2013) this mainly explains the explanatory failings of the discipline, particularly in the context of the recent economic turbulences. Instead of presupposing regularities at the level of agents or events, the regularities first must be scrutinised and stylized facts established. These can then be used to uncover the causes and mechanisms and to approach a better understanding of the phenomena of interest, for example with means of empirically calibrated simulation models (BRENNER/WERKER 2007).

In this vein, the thesis starts by looking at the facts. Chapter 2 investigates the regularities of regional economic growth processes. How are growth rates distributed, and how do they depend on the entities' size? Which differences in the growth rate distributions are observed for specific industries? Chapter 3 uses the established knowledge on the growth rate distributions to identify the causal structure in regional growth processes. Chapter 4 and 5 continue by exploring spatial aspects of economic growth processes. Chapter 6 goes one step further by identifying the factors that make extreme growth events more likely to occur in a region. With help of theoretical reasoning, it tries to establish economic meaning on some aspects of the observed distributional regularities. Chapter 7 studies firm dynamics in China and develops a new conditional estimation approach of the growth rate distribution. Finally, chapter 8 develops a simulation model in which the observed empirical patterns emerge from elementary processes at the level of firms.

1.2.2 Implications from an evolutionary economics perspective

Economic growth, conceptualized as dynamic change over time, by definition is evolutionary (FRENKEN/BOSCHMA 2007; for a more extensive discussion of the conceptualization of growth see section 1.3.1). The discipline of evolutionary economics has emerged as an alternative paradigm to mainstream neoclassical economics, covering a variety of theories and topics in economics (WITT 2013). Within its Neo-Schumpeterian

conception, rooted in the work of NELSON and WINTER (1982), evolutionary economics focuses inter alia on technological change and economic dynamics (DOSI/NELSON 1994). Here, economic development is not conceived as the path to equilibrium, but as the result and the actual cause of these out-of-equilibrium dynamics (CANTNER/HANUSCH 2002). The manifold topics and approaches attributed as "evolutionary" all have in common an investigation of change as driven by endogenous forces, and not their marginalization as exogenous shocks (WITT 2013). More recently, also a spatial account of evolutionary economics was proposed (BOSCHMA/LAMBOOY 1999) and further systematized (e.g. BOSCHMA/FRENKEN 2006). It gained increasing momentum especially in the discipline of economic geography. Evolutionary economic geography seeks to explain the emergence and incessant changes in the economic landscapes by the underlying industrial dynamics of firms (BOSCHMA/FRENKEN 2011, HASSNIK et al. 2014).

CANTNER/HANUSCH (2002), following WITT (1987), characterize all evolutionary theory as being built on four basic features. First, it focuses on dynamic changes in the course of time, whereby time is, secondly, conceptualized as being historic, i.e. irreversible. Thirdly, it tries to explain the capacity to create novelty, the "ultimate cause of endogenous change" (WITT 2008: 551), and the implications of such innovations on economic development, for which, finally, heterogeneity is a condition *sine qua non*. Heterogeneity concerns both the actors, i.e. the organizational and technological aspects of the firms in an economy, as well as their activities, i.e. the economic behaviour or performance. From this definition, four major implications are deduced for the thesis. First, regional economic growth should be analysed from a disaggregated perspective, in which, secondly, a focus on the entire distributions accounts for the heterogeneity. Thirdly, growth is conceived as innovation-driven, fundamentally based on the generation of new knowledge, which, finally, often originates from related knowledge sources external to the firm. In the following, these implications are discussed in more details.

First implication: the importance of disaggregation

Heterogeneity, which can be found "wherever one looks" (DOSI et al. 2010: 1868), is a main ingredient for evolutionary analysis. Three layers of heterogeneity are particularly relevant for this work: the heterogeneity of firms, industries and space. All these layers strongly shape how economic actors engage in and organize their activities of different technological nature.

Evolutionary growth is driven by elementary processes at the level of firms. Ultimately, all economic outcomes, like the spatial patterns of industry location in economic geography, arise from historic growth processes at the level of firms (FRENKEN/BOSCHMA 2007). As ANDERSEN and HOLM (2014: 298) put it, the "primary issue of evolutionary analysis is not the aggregate growth of the population but its structural change due to differential growth of members". Thus, the differential rates of growth at the micro-level of firms are strongly related to the development at the aggregate level of regional economies. The origin of growth at the aggregate level is best conceived as an emergent property of complex micro-level processes and activities (METCALFE/FOSTER 2010). However, processes and mechanisms of change might be hidden by spatial or industrial aggregation. For example,

new technologies or business models emerge within firms, but are reflected only decades later in the rather static industrial classification scheme. A regional economy might show a rather stable aggregate performance, even though "turbulences underneath the big calm" (DOSI et al. 2012) would be an indicator for important long-term changes in the structure and functions of the regional economic system. Hence, the firms' individual decision rules and selection mechanisms on the market have been elected by NELSON and WINTER (1974) as the central units for investigation. Disaggregating the economy into its basic constituents is a key to understand its growth dynamics. Firms can be perceived as economic actors that perform certain activities. These activities are not only shaped by various demographic characteristics, like their size or age, but also to a considerable degree by the technologies on which they rely on. The function of firms as "agents of change" (FRENKEN/BOSCHMA 2007: 636), therefore, can be understood only by fully taking into account the heterogeneity in their technological assets and resources (CANTNER/HANUSCH 2002). These are intrinsically related to the firms' spatial location and industry.

Industries, or more precisely, the firms affiliated to them, differ systematically. Industries can be distinguished according to their innovation intensity and innovation patterns, the underlying technology, the possibilities for appropriation (PAVITT 1984, WINTER 1984, BRESCHI et al. 2000, MALERBA 2002), their stage in the industrial life cycle (KLEPPER 1997) or the composition of their knowledge base (ASHEIM/GERTLER 2005). As GRILICHES und MAIRESSE (1995: 23) noted, "something like Mandelbrot's fractal phenomenon seem to be at work here also: the observed variability-heterogeneity does not really decline as we cut our data finer and finer. There is a sense in which different bakeries are just as much different from each other, as the steel industry is from the machinery industry". Besides, it is often empirically observed that the size structure and the stochastic growth patterns of the firms (e.g. BOTTAZZI/SECCHI 2003) as well as the geographical concentration tendencies (FORNAHL/BRENNER 2009) differ across industries. In the context of industrial clusters, various studies (e.g. ANESLIN et al. 2000, BEAUDRY/SWANN 2009, MAINE et al. 2010, RASPE/VANOORT 2011) show that agglomeration effects strongly depend on the industry. For instance, DÖRING and SCHNELLENBACH (2006) conclude from a literature survey that knowledge spillovers are particularly relevant for younger industries. This finding has motivated some recent studies (e.g. NEFFKE et al. 2011, POTTER/WATTS 2011) which try to explain the strength and kind of agglomeration externalities focusing on the life cycle of an industry. A related stream of literature (e.g. ROSENTHAL/STRANGE 2001, ERIKSSON 2011, DRUCKER 2012) shows that also the spatial extension of the observed externalities differs across industries.

Finally, space itself is heterogeneous. The conditions and quality of the socio-economic landscape vary not only between regions, but also within them, independently of the geographic scale used for regional delimitations. For instance, only a few cities in a state host a university and only a few locations within the city are suitable for an eased regular interaction. This intra-regional variability makes it impossible to delimit regions as consistent aggregates (HARRIS 2010, PINSKE/SLADE 2010). Hence, FRENKEN et al. (2011: 21) conclude that "theorizing about clusters should start from theory of the firm that highlights firm heterogeneity and role of external learning".

From these remarks follows that the three levels of heterogeneity must be the guiding principle for the empirical analysis of this thesis. More precisely, they imply that growth processes should be analysed at various levels and dimensions of disaggregation. Following the evolutionary maxim, "the more we aggregate the more we hide the evidence for and causes of economic evolution" (METCALFE/FOSTER 2010: 68), the economy is disaggregated from both a spatial and sectoral perspective. The growth of a national economy is the sum of the growth of its heterogeneous regions, which ultimately results from the growth of the regions' firms. An analogous logic can be applied to the disaggregation of the economy into heterogeneous industries and the firms that are more or less technologically related to each other. This dual disaggregation is indispensable as different regions are characterized by different industry-mixes (ROBERTS/SETTERFIELD 2010). Hence, regional growth is strongly interwoven with industrial dynamics (FRENKEN/BOSCHMA 2007, FRENKEN et al. 2014), implying that the evolution of a regional economy is to be viewed from a disaggregated industry-specific perspective.

For the analysis of agglomerations, industrial clusters and knowledge spillovers within the context of regional growth, regions are often chosen in the literature as a natural candidate for the observational unit. Taking the discussion above seriously, the micro-level of firms seems to be more appropriate. At the aggregate regional level, agency is usually absent (MASKELL 2001). Rather, it is the firms within the regions that (inter-)act, learn and adapt new technologies (NELSON 1998, CHESHIRE/MALECKI 2004). Therefore, the constitutive mechanisms underlying agglomerations or industrial clusters are difficult to assess at the regional level alone. Firms differ in their ability to benefit from external factors and to absorb knowledge spillovers (MCCANN/FOLTA 2008, ERIKSSON 2011), depending on their size, age, organizational structure and many other, often immeasurable, factors (RIGBY/BROWN 2013). Hence, factors affecting the growth of a region as an aggregate do not necessarily affect all its firms in a similar way. In other words, studying the regional level alone tells us little about the nature and concrete mechanisms of agglomeration effects like knowledge spillovers (ROBERTS/SETTERFIELD 2010). This becomes clear as the theories that underlie agglomeration economies are microeconomic in nature. From this follows that the impact of agglomerations and industrial clusters on regional growth should be studies using a micro-level approach (FESER 2002, RASPE/VANOORT 2008), "in which firms are the main actors and their interactions with other firms [and organizations like universities] in specific environments determine each firm's performance" (BEUGELSDIJK 2007: 196).

The micro-level has become a research agenda in many fields traditionally concerned with the growth dynamics of aggregate systems, like New Economic Geography, the Endogenous Growth Theory or the literature on Industrial Clusters (e.g. HARRIS 2011, OTTAVIANO 2011, BASILE/USAI 2012, AGHION et al. 2013), as a mean to better understand the actual mechanisms and causes of evolutionary change (METCALFE/FOSTER 2010) or increasing returns (AUDRETSCH/DOHSE 2007).

Second implication: distributions of growth rates as explanandum and explanas

Another implication that immediately arises from the omnipresent heterogeneity of economic actors and activities is that economic growth dynamics are more

comprehensively understood through the lenses of a distributional analysis of growth rates. This allows for differences in the performance of the members of a population. Such an implicitly assumed population perspective is a recommended feature of each empirical evolutionary analysis (CANTNER/HANUSCH 2002). "The evolutionary question is 'Why do rates of growth differ across activities and over time?' and not the question 'Why are they uniform and stable?'" (METCALFE/FOSTER 2010: 68).

Looking at the average growth rate of a population is insufficient to understand the rate and direction of evolutionary change in that population, as evolutionary change is characterized and driven by diverse, non-representative behaviour (METCALFE/FOSTER 2010). CAPASSO et al. (2011: 23) state that "many 'average' relations retrieved in literature […] are driven by the few dynamic firms in the sample rather than by the many stables one". In particular, such off-the-mean growth events are crucial for evolutionary processes (RICE 2004, ANDERSEN/HOLM 2014). Hence, the heterogeneity of growth rates, i.e. the entire distribution, becomes the relevant *explanandum* (METCALFE/FOSTER 2010).

Having described and eventually explained the emergence of the observed growth rate distributions, valuable insights on the underlying growth processes can be gained. These insights might serve as incentives to improve the quality and scope of economic models as well as to re-formulate policy instruments. Particularly the latter demands also a better understanding of the economic meaning of the distributional characteristics, say, the fatness of the tails, on the performance and evolutionary dynamics of the corresponding aggregate system. Hence, the distribution of growth rates should also become an *explanans*. For example, it is generally agreed upon that "evolutionary phenomena tend to undergo non-gaussian lives influenced by persistent positive or negative interactions among agents within and across relevant populations" (DOSI et al. 2010: 1982). However, it is less well understood why the population dynamics deviate from normality and what this means for the performance and evolution of the entire population.

To sum up, the distributional analysis of growth rates of each member of the population is deduced as an empirical evolutionary approach to understand more comprehensively the complex dynamics of the heterogeneous actors and activities. As BOTTAZZI and SECCHI (2003) argue, the growth rate distributions, which tend to be less affected by historical events, often carry more relevant information concerning the nature of the underlying economic processes than the corresponding size distributions. Moreover, a distributional analysis of growth rates can be regarded as a diplomatic approach between traditional economists and geographers. Whereas the former look at average trends, while ignoring cases that are far off the mean as random noise, the latter prefer to look at all individual cases as complex, contest-specific development processes (STORPER 2011). In this sense, distributions acknowledge the entire population, while at the same time allowing for some degree of generalization.

Third implication: innovation-driven growth and the role of (external) knowledge

ROMER's (2007) dictum that "economic growth occurs whenever people take resources and rearrange them in ways that are more valuable" can be regarded as a growth fundamental. This is very much an evolutionary process of discovering novel ways of doing

things. Following WITT (2008: 552), evolution is a "process of self-transformation whose basic elements are endogenous generation of novelty and its contingent dissemination". The "invasion of economic novelty" (METCALFE/FOSTER 2010: 65) is inseparably related to the generation of new knowledge (e.g. WITT et al. 2012) and to the question of how it *spills over* across the actors of an economic system. This perspective on the role of knowledge also suggests that much growth-relevant knowledge is not exclusively generated within the boundaries of the firm, but is sourced and absorbed from outside. It is widely observed that firms increasingly depend on external sources of knowledge (FRATESI/SENN 2009, HARRIS 2011). From a co-evolutionary perspective, it is even suggested that firms need to co-evolve with their environment in order to grow (CLARKE et al. 2014).

Therefore, concepts from economic geography, like regional innovation systems or industrial clusters (e.g. ASHEIM/GERTLER 2006 and ASHEIM et.al. 2006), have recently received much attention. These concepts acknowledge local environments and interactions as central sources of new variety (HODGSON 2013), making innovation-driven growth possible in the first place. In the words of GORDON (2013: 699), "economic growth is all about combinations and we are best at it when we work with and among others". Hence, many important research questions on regional economic growth directly relate to the role of the spatial distribution and agglomeration of knowledge sources external to the firm. This is to say that economic geography, i.e. the observable spatial patterns of economic activities, not only emerges from historic growth processes, but also affects the subsequent patterns of growth (FRENKEN/BOSCHMA 2007). The literature focuses here, amongst others, on the following questions: which sources of external knowledge, like public research institutes, universities or other firms, affect firm growth? What are the underlying mechanisms of the mostly implicitly assumed knowledge transfer, like unintended knowledge spillovers, labour flows or trade? How does geography in the form of geographical proximity impact and mediate the observed externalities? Furthermore, one might ask how the various knowledge-related activities, like innovation, research or higher education, evolve and influence each other within a regional economy.

Fourth implication: relatedness as a unifying evolutionary framework for agglomeration studies

Traditionally, the literature on agglomerations and external factors has circulated around the debate of whether specialization (e.g. HENDERSON 2003) versus diversification (e.g. GLAESER et al. 1992) matters for the performance of firms or regions. This debate implied that co-located actors or activities were distinguished according to their similarity in a dichotomous fashion, that is whether they are identical or not in some measurable quality like their affiliation to an industry group. However, taking sufficiently detailed qualities into account, not two identical firms or technologies exist in practice. Not surprisingly, the main conclusion which is drawn in several meta-studies and literature overviews on that debate (e.g. BEAUDRY/SCHIFFAUEROVA 2009, DEGROOT et al. 2009, MELO et al. 2009) is that of empirical inconclusiveness. Even though much of the ambiguity can be ascribed to measurement errors, differences in the empirical set-up and issues related to aggregation (VANOORT et al. 2012), a more fundamental reason must be seen in the dichotomous distinction, which is somewhat artificial and little helpful as it ignores the intermediate part

between the two poles of similarity versus dissimilarity. In contrast hereto, relatedness as a universal property makes the notion of specialization and diversification redundant.

From an evolutionary perspective, regional growth and industrial dynamics can be viewed as a branching process, in which technological activities emerge out of existing, related activities (FRENKEN/BOSCHMA 2007). Therefore, it is a matter of the degree of relatedness: too little variety does not provide any heterogeneity, the main ingredient of evolutionary change, and too much variety results in a system which parts are too distant and disconnected for interactions. Hence, evolutionary economic geography suggests that intermediate degrees of relatedness might matter most. NOOTEBOOM (2000) applied this evolutionary logic in the context of knowledge, which for an optimal absorption should be complementary to some degree, but not too similar. On the one hand, cognitive distance might hinder effective communication and learning, while on the other hand, only little opportunities for combining different pieces of knowledge in new ways is provided. From this follows that related variety can serve as a unifying framework for agglomeration studies. Directly deducible from evolutionary processes, it moves the debate beyond the simple distinction between Marshall versus Jacobs (BOSCHMA/FRENKEN 2011). Moreover, heterogeneity as discussed above suggests that the strength of agglomeration effects and the kind of external factors are not important for all firms equally, but depend on their internal characteristics. And these, as WITT (2000) has observed, change tremendously during their life-time and growth periods.

A brief conclusion

Evolutionary thinking implicitly underlies the research questions, theoretical reasoning and empirical approaches of all chapters of this thesis. Only in chapter 6, an explicit evolutionary perspective is taken in the interpretation of the observed empirical regularities. More precisely, this chapter develops an empirical framework to operationalize regional resilience from an evolutionary perspective. The thesis mainly focuses on the exploration of the growth dynamics within a population of firms, industries or regions. To fully explain how and why the distributions have emerged and evolved one would be required, as recently claimed by WITT (2013), to take into account the mechanisms at work in certain historical periods and to develop a theory of human economic behaviour from an evolutionary perspective. This is beyond the scope of the thesis. Only in chapter 8, a simulation model based on some fundamental theoretical considerations on the competition and innovation behaviour of firms with the aim to replicate many of the empirically observed regularities is devised.

1.3 Conceptual aspects

After the general science-philosophical background has been set out, more specific and conceptual aspects of the thesis' theoretical framework are to be discussed. First of all, growth must be defined and empirically operationalized (see section 1.3.1). Here, growth is defined as the change in an economic activity over time. This activity perspective is more

broadly framed in the subsequent section 1.3.2. Firms are the most important actor engaging in these activities and are conceptualized separately in section 1.3.3. Positioning the chosen resource-based view of firm performance into a spatial perspective, the concepts of industrial clusters and agglomerations turn out to be intrinsically related hereto (see section 1.3.4). Finally, the concept of space itself has to be defined (see section 1.3.5) – it will depend on the specific research question about the economic actors and activities.

1.3.1 The concept of growth and its empirical operationalization

From an evolutionary definition, growth is the change over time, the differential outcome between at least two points in time (McKelvie/Wiklund 2010). Growth can mean either positive or negative changes, i.e. both the expansion and decline of some measurable economic activities. This simple definition rules out the perception of growth as convergence or divergence over time (see the voluminous literature on regional convergence, rooting in the seminal work of Barro and Sala-i-Martin 1992), as a long-term development path, which refers to a more qualitative and comprehensive literature on regional development (Malecki 1997), or as the growth mode, representing the ultimate historical reasons of growth, like the transition to an innovation-driven growth in current times (Etzkowitz/Leydesdorff 2000).

This definition opens the questions of growth of who (which actor or entity) and what (which activity) and how it is measured. In this thesis, the growth rates *g* of either firms or regional economies, depending on the research question, are calculated as the log difference of size *S* between time *t* and *t-1*:

$$g = log(S_t) - log(S_{t-1}) \quad (1)$$

To capture economic activities, a variety of alternative size measures can be found in the literature, like sales, productivity, profits, market shares, income, etc. (Delmar et al. 2003). Being a multi-dimensional phenomenon, no universal best size indicator exists for economic growth (Gilbert et al. 2006, McKelvie/Widklund 2010). Many of the distributional characteristics of the growth rates, however, are observed for the various measures (Erlingsson et al. 2013). Only in chapter 7, different size measures are systematically compared. In all other chapters, employees are chosen as the primary measure of size on the ground of the following reasoning. From an empirical point of view, employment growth is not sensitive to inflation or currency exchange rates (Barbosa/Vasco 2011). Besides, it is identified to be causal for many other measures, like sales or profits (Coad 2010a). From a theoretical point of view, employees as the most important asset of the firms are closely related to the Penrosian theory of the firm (McKelvie/Widklund 2010). As such they indicate also more directly organizational complexity (Delmar et al. 2003). From a normative point of view, employment growth is the main variable of interest for policy makers who are concerned with the creation and loss of jobs (Raspe/vanOort 2008).

Although employees are the preferred variable both at the level of firms and regions, two important limitations have to be addressed. First, many mechanisms work only indirectly on employment (Blien et al. 2006). For instance, knowledge directly affects innovation

propensity. In cases of product innovations, positive employment effects can be expected due to growth opportunities as a result of new products, whereas in case of process innovations, negative effects may arise from rationalisation measures (BUERGER et al. 2012). Secondly, firms often make lumpy adjustments to their employment level, which tend to be, due to non-convexities and irreversibility either large or nil (CABALLERO et al. 1997, COAD 2012). HALTIWANGER (1997) observed that plants spend a large fraction of time within plus or minus 30% of their desired employment level and COAD and HÖLZL (2009) even report that 65% of small establishments listed in the Austrian Social Security files do not display any changes in employment from one year to the next. This zero-growth problem introduces an artificial discreteness into the growth processes, which often leads to empirical flaws. Therefore, in chapter 6 an advanced estimation approach is developed to obtain less biased estimates of the continuous growth rate distribution. Besides, quantile regression approaches allow for a systematic focus on the actually growing firms.

After having defined and empirically operationalized the growth rates, the growth processes are to be conceptualized. In the literature, GIBRAT's law is often used as an entry point and benchmark for empirical studies (BOTTAZZI et al. 2011). It models growth as a simple stochastic process, in which the size *S* at time *t* is the result of a proportional effect ε_t and *S* at *t-1*:

$$S_t = (1 + \varepsilon_t)S_{t-1} \tag{2}$$

Starting from the initial size, and after taking the natural log and assuming that for small effects, log(1 + ε) converges to ε_t, one gets:

$$log(S_t) = log(S_0) + \sum_{k=1}^{t} \varepsilon_k \tag{3}$$

which would result in an approximately log-normal size distribution as *t* goes to infinity. The empirical validity of the assumptions and implications of GIBRAT's law are, occasionally controversially, discussed in a huge body of literature, on which Sutton (1997) or SANTARELLI et al. (2006) provide comprehensive overviews. Besides, GIBRAT's law provides the logic for defining the growth rates as the differences of the natural log of sizes and it has motivated particularily in the firm growth literature the use of a GIBRAT-like growth regression, in which the growth rates are regressed on the size and further variables of interest (AUDRETSCH/LEHMANN 2005). This regression framework is also used in chapter 4 and 5 to study firm growth.

In chapter 2, GIBRAT's law is explicitly tested at the regional level. It is rejected on basis of all of its assumptions (for an overview on the assumptions see GABAIX 2009). The average and the variance of the growth rates are not independent of size, growth rates are correlated over time and they are far from being normally distributed. In light of the more complex nature of regional growth, it is obvious that also more sophisticated models are required. This is taken into account in the empirical model of chapter 3, while the simulation approach in chapter 8 lays the foundation of a more realistic model.

1.3.2 The concept of regional technological activities

Growth is defined in this thesis as the change over time in some economic activities. Economic activities measure human actions that aim to generate economic value. In a modern market economy, these activities are usually carried out within the organizational context of private firms (COASE 1937). This organizational form is discussed separately in section 1.3.3. A simple proxy of the extent of activities carried out within a firm, i.e. the "size" of a firm, is its number of employees, as already proposed in section 1.3.1. From a spatial perspective, economic activities, as intrinsically related to humans, can be always traced back to a specific location in space. To anticipate the conceptualization of space in section 1.3.5, this allows to count the number of activities taking place within a spatial unit (space as a container) or to measure the distance to activities taking place in other locations (space as distance). However, economic activities do not unfold in isolation. They depend on, are influenced by and co-evolve with other activities as well as structural aspects that determine how the activities are organized within the contexts of firms or regions.

The other activities most relevant in modern, knowledge-based economies are research activities, innovation activities and educational activities (BRENNER/BRÖKEL 2011). Research activities represent the systematic search process for finding new opportunities, which might result in different kinds of economic activities. Innovation activities are concerned with the challenge of how and what kind of new economic activities are implemented. The creation of novelty is often regarded as the only source of endogenous growth of economic activities (WITT 2003). It should be noted that innovation not necessarily follows from research activities but often is the result of learning processes from economic activities. Finally, educational activities feed the other activities with humans possessing the knowledge and ability to carry out all these activities.

The structural aspects that determine how the activities are organized refer, at the level of firms, to their internal characteristics, namely, their organisational forms and routines. In the literature, these characteristics are often approximated by the size and age of a firm (COAD 2009). At the level of regions, in which the economic activities of the firms are embedded and carried out simultaneously and interdependently with the other activities, these structural aspects refer to the physical and socio-economic landscape of a region and are much more complex and difficult to assess. The physical landscape, together with the built infrastructure, determines the accessibility of other activities, which are distributed in space while often showing an agglomeration tendency. The socio-economic landscape, in the various meanings of institutions, culture, demography, or policies, influences how and which of these activities are carried out. It also mediates the region-specific interplay between the various activities, for example, the way how innovations are transformed into more economic activities (BRENNER/BRÖKEL 2011). RODRÍGUEZ-POSE (1999) summarizes this socio-economic landscape as a "social filter", which can be approximated only roughly by measures like the population density or unemployment rate.

In chapter 3, the activity perspective underlying this thesis is conceptualized more systematically (see also Figure 2). Here, a regional economy consists of the four different activities as mentioned above. It is assumed that the growth dynamics of the activities are interdependent and evolve endogenously over time. Because both the nature of the

activities and the dependencies among them strongly depend on the technological field, different technologies are studied separately by assuming an industry-specific perspective.

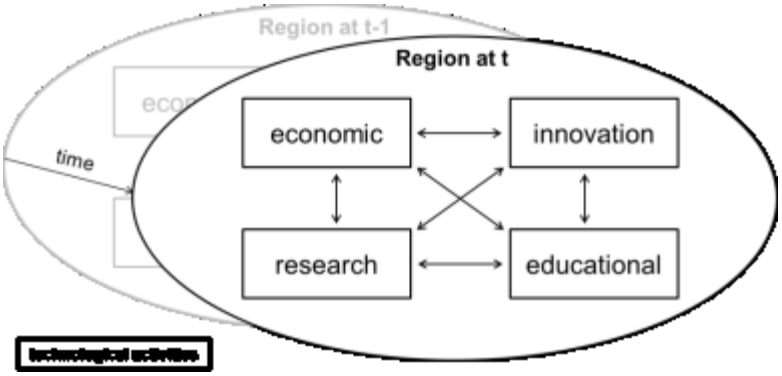


Figure 2: The concept of regional systems of technological activities

The depicted concept of regional systems of technological activities shares many basic features with the innovation system literature (e.g. FREEMAN 1995, COOKE et al. 1997), foremost the systemic and multi-dimensional view on a regional economy. The most important difference, however, lies in the exclusive focus on activities, while regional innovation systems are centred on the actors. Hence, the two concepts represent two different entryways to study such regional systems. This distinction is best exemplified with innovation or research activities, which can be carried out by different actors, like private firms, universities or public research institutes. Moreover, the list of activities could be further extended, for example by entrepreneurship activities, which might be useful to differentiate among economic activities within existing firms and opposed to economic activities within new firms.

Other important aspects of regional economies, as highlighted by the innovation system literature, are institutions and networks. Owing to the empirical orientation of the proposed concept of regional systems of technological activities, these aspects are doomed to recede because of well-known measurement issues. At least, institutional aspects can be considered, according to the "social filter" approach, as exogenous control variables. Knowledge on social networks would be essential to operationalize the links among the activities more concretely. Until now, they are only implicitly assumed as the channels for the mechanisms leading to the endogenous and interdependent growth dynamics.

1.3.3 The concept of the firm

In modern market economies, firms are the central actors engaging in economic activities. COASE (1937) has initiated with his seminal work on the nature of the firm an ongoing debate on the boundaries (e.g. DOSI/TEECE 1998) and the theoretical conception of firms (see COAD 2009 for a more extensive discussion). As this thesis focuses on the role of the

external environment to firm performance, a resource-based view of firm growth seems to be most appropriate on the ground of the following reasoning.

According to PENROSE (1959), firms are heterogeneous bundles of resources. Within this framework, KOGUT and ZANDER (1992) stressed the critical role of knowledge for competitive advantages, i.e. the exploitation of non-tangible productive assets (HYMER 1976) that might ultimately translate into better performance. TEECE et al. (1997) introduced a dynamic perspective, arguing that the key to performance is the ability to create and release resources, and by doing so, to adapt and reconfigure the existing resource portfolio – the dynamic capabilities. However, not the entire set of relevant knowledge for the production process can be provided and generated internally by the firm. Firm performance will also depend on the ability to absorb new knowledge and other resources from outside its boundaries and to combine it with its internal resources – the absorptive capacity (COHEN/LEVINTHAL 1990). External knowledge, acquired through market transactions and unintended or intended knowledge spillovers, always originates from other actors that can be located in space. Therefore, the resource-based view on firm performance is strongly related to the literature on spatial externalities.

From this resource-based view on firm growth, several research questions and implications arise. The following ones are explicitly addressed by the thesis:

- Which external sources provide growth-relevant knowledge to the firms? Among the most relevant external knowledge generating activities seem to be public research, university education and other economic activities that are more or less technologically related to the firm's own activities.
- Which internal factors determine the necessity of relying on externally provided knowledge as well as the capacity to absorb this knowledge? Generally, firm size is regarded as the most important variable measuring internal resource constraints (ALMEIDA et al. 2003, MAINE et al. 2010). However, also other aspects, like age, internal research activities or export activities, which all have been repeatedly studied in the economic literature (COAD 2009), might become important control variables.
- How do the firms' characteristics interact with the type of external knowledge sources that are most relevant to the firm? Put differently, which kind of firms benefit from a specific external activity?
- How is the absorption of external knowledge moderated by the distance between the firm and the location of the activity, where knowledge is generated? The concept of distance, or proximity as its reciprocal, is multi-dimensional, hence various forms of distances might matter, like geographical distance or technological distance (BOSCHMA 2005). Again, the characteristics of the firms as well as the type of knowledge source might shape the moderating role of distance.
- How do firms adapt their resources in face of sudden changes in their external environment, like new business opportunities or negative shocks? Do firms immediately adapt their resources, or are they sometimes reluctant to grow or shrink if they are able to rely on redundant resources? Hence, the resource-based view requires one to think also about the growth processes which might ultimately results in specific growth patterns.

Linking the resource-based view on firm performance to regional growth provides a micro-funded view on agglomerations and industrial clusters, as will be argued in the subsequent section.

1.3.4 The concept of industrial clusters and agglomerations

Approaching economic growth from a spatial perspective, agglomerations are perceived as an important factor. At the regional level, especially the notion of industrial clusters has attracted high interest both in science and politics (PORTER 2000, PORTER 2003b). The minimum definition of an industrial cluster requires a group of firms from the same or related industries located geographically near to each other (PORTER 2003b, MAINE et al. 2010). Often, such spatial agglomerations are conceptually extended to include associated institutes like universities. Although the unequal, and agglomerated, distribution of economic activities across space is an interesting research topic by its own (BRENNER 2005, FORNAHL/BRENNER 2009), it is the long-term implication of local industrial clusters to regional development why this concept has become so popular (BRENNER/GILDNER 2006). From this follows that some agglomeration economies must exist explaining the better (or worse) performance of regions hosting industrial clusters. Such externalities can be said to occur "if an innovation or growth improvement implemented by a certain enterprise increases performance of other enterprises without requiring the benefiting enterprise to pay (full) compensation" (VANOORT et al. 2012: 469).

The literature tends two draw two further conclusions: first, according to the endogenous growth theory (e.g. ROMER 1986) and in the light of the high mobility of human and physical capital, at least at the regional level, these externalities should result primarily from knowledge spillovers. Secondly, as agglomeration economies are observed to work to a large degree within regions, the underlining mechanisms such as knowledge spillovers must be limited in their effects by some spatial range (FELDMAN 1999). Therefore, it is argued in the literature that knowledge embedded in individuals and firms tends to be sticky (GERTLER 2003). This is due to the high costs of transmitting knowledge over longer distances, especially economic valuable knowledge, which is often complex, unstructured and tacit (SORENSEN et al. 2006).

Knowledge-based externalities can be differentiated along several aspects. BRESCHI and LISSONI (2001) distinguish between pecuniary, that is market-based, versus technological spillovers, whereas MAGGIONI et al. (2007) differentiate between unintended knowledge spillovers versus intentional knowledge barter exchange. Besides, various mechanisms are discussed in the literature: knowledge is transmitted by the trade of goods, by licencing of patented technology, by mobile labour, by shared research projects or information exchange in formal or informal social networks (DÖRING/SCHNELLENBACH 2006). Finally, knowledge spillovers can be distinguished by their sources of origin. Growth-relevant knowledge might originate from R&D activities within universities or public research institutes, from economic activities of other related firms, or from education activities resulting in qualified graduates (RASPE/VANOORT 2011). However, the mere presence of potential sources for knowledge spillovers also increases the competition for other scarce resources, like land or labour, as well as the probability of unwanted knowledge disclosure

by imitation. Hence, agglomeration economies can be both positive and negative, and the net effect remains ultimately an empirical question (MAINE et al. 2010).

Not surprisingly, the literature on industrial clusters is subject to criticism, which is mostly related to the conceptualization and operationalization of the underlying spatial externalities. For instance, the boundaries of a cluster, the underlying mechanisms and the relevance of several sources of externalities often remain unclear (MARTIN/SUNLEY 2003). Things become more complicated in reality, as the different mechanisms operate at different geographical scales and different firms benefit differently from the various sources of externalities (MARTIN/SUNLEY 2003, RIGBY/BROWN 2013). Hence, even PORTER (2003b) concludes that the strength of knowledge spillovers and their importance to the performance should determine the ultimate boundaries of an industrial cluster. To take into account these issues seriously and to disentangle the variegated effects of industrial clusters, this thesis studies the impact of external knowledge-related activities on the growth prospects of the heterogeneous firms. Hereby, it focuses primarily on the spatial range of the externalities. Hence, "potential externalities" (GORDON 2013: 680) are conceived as the revealed spatial impact of geographically proximate activities. This ultimately implies a convergence of the concept of industrial clusters with the accessibility approach as based on WEIBULL (1980). The location in space determines which activities are accessible, and a higher accessibility should therefore translate into a higher potential for spatial externalities (ANDERSSON/GRASJÖ 2009: 164). The underlying idea is that especially new and complex, hence growth-relevant knowledge is not easily accessible at every point in space, because it is highly geographically concentrated and mostly embodied in people and generated within firms or other institutes (ADERSSON/KARLSSON 2007).

This thesis contributes to the debate on localization versus urbanization (for a literature overview see BEAUDRY/SCHIFFAUEROVA 2009). Various sources of externalities can be distinguished according to their relatedness with firms, and different degrees of relatedness, from more similar to more diverse activities, can be compared. The chosen approach is also in line with the research trajectory of firm growth and external factors. AUDRETSCH and LEHMAN (2005), to name one example, study the impact of access to university-based knowledge spillovers on firm growth. Moreover, the spatial extent of agglomeration economies is still a central and open research question (see among others BALDWIN et al. 2010, BASILE/USAI 2012, DRUCKER 2012, MCCANN/SHEFER 2005 or MCCANN/VANOORT 2009). This research contributes also to this question by empirically operationalizing the accessibility approach from a micro-geographical perspective in which both firms as well as various sources of externalities are geo-localized into space. Instead of having to rely on the appropriateness of the spatial scale chosen *ex ante*, the spatial extent of the externalities can be directly inferred from the data – separately for each knowledge source, each kind of firm, each industry or each other dimension of heterogeneity. By doing so, also sub- or supra-regional dimensions of externalities can be accounted for (ERIKSON 2011). The following section will cover the discussion on the concept of space in a more systematic way.

1.3.5 The concept of space

Geographic space can be conceptualized in various ways. It is widely agreed that the actual conception used should depend on the precise research question (ROBERTS/SETTERFIELD 2010, AGNEW 2013). Therefore, it is all the more important to be explicit on the definition of space, as different streams of literature tend to work with different spatial scales – New Economic Geography studies the city level, the regional science literature usually includes the regional hinterland, and the cluster literature even prefers to be silent on the spatial extent of agglomerations altogether (DRUCKER 2012).

Generally speaking, space can be conceptualized either as distance or as place (LORENZEN et al. 2012). The former concept focuses on geographic proximity, on space as a physical barrier, which can be operationalized by measures like physical distance or transportation costs. The place-based definition of space refers in the context of regional economic growth to uniform-abstract, geographically delimited subunits of a national economy. Put differently, the geographic space of a country is divided into more or less homogenous regions. Here, space is reduced to a mere physical container and performs a purely passive role in economic growth (CAPELLO 2009). CAPELLO (2009) adds to this traditional distinction a third definition, the diversified-relational space. This more complex conception of space allows for heterogeneity of the economic activities within a region. This demands to focus on the individual economic actors and also to define regions from the perspective of these actors. For instance, how and to which spatial extent does the individual firm benefit from other firms or institutes like universities located nearby? Here, space becomes an economic resource, an independent production factor, which generates static or dynamic advantages for firms in forms of agglomeration economies.

For the analysis of regional growth processes, the thesis begins with the definition of space as a container in which economic activities take place, expand or decrease in intensity over time. Although straightforward, this definition still requires the imposition of some boundaries on a continuous economic space à la PERROUX (1950). It is often argued that functional labour market regions, as ideally self-contained zones of economic activity, are more meaningful than administrative regions (ECKEY et al. 2006). In chapter 2, two existing delimitation systems for labour market regions in Germany are compared vis-à-vis the administrative level of districts. However, BATTY (2008) observes that innovation and knowledge generation activities are more strongly concentrated in space than, for instance, the population. Besides, the transfer of knowledge is mostly embedded in people who are regarded as imperfectly mobile (DÖRING/SCHNELLENBACH 2006). Hence, the use of regional containers is insufficient for the analysis of the economic impact of agglomerations or industrial clusters. Without knowing the spatial extent of the underlying mechanisms, it is difficult to choose the appropriate definition of regions. By varying the spatial scale of the analysis, BUERGER et al. (2010) as well as WENNBERG and LINDQVIST (2010) show empirically that even the sign of agglomeration effects might change depending on the level of spatial aggregation and the delimitation of regions, a problem that was first described by OPENSHAW (1984) and coined with the term Modifiable Areal Unit Problem (MAUP). In practice, the regions' firms are merely homogenous, but each of them is affected differently and at a different spatial scale by the various external factors.

To avoid border biases and arbitrary spatial units, the effects of agglomerations are best studied by focusing on the micro-level and the locations of the relevant actors in a continuous space (DURANTON 2008, OVERMAN 2010).[1] Theoretically, all actors and activities can be traced back to concrete locations in space. This is obvious for firms or universities, but also for people, who work, live and consume in specific places. Using distance-based measures, which count each activity nearby for each individual firm, agglomerations can be considered at all spatial scales simultaneously (MARCON/PUECH 2012). This approach also relates to the diversified-relational conception of space as proposed by CAPELLO (2009), as the firm-specific relations to external activities can be explicitly acknowledged, for instance by assessing the technological relatedness among the activities and by modelling the specific way of how these factors affect each firm individually. Other factors, like the social and structural condition of the firms' surroundings[2], can still be summarized at the territorial, i.e. regional, level (RODRÍGUEZ-POSE/CRESCENZI 2008).

1.4 Methodological aspects

The outlined research questions and theoretical concepts demand a specific toolset of empirical methods. This thesis either uses methods that can be regarded in the literature as non-standard or it introduces new methodological innovations. Therefore, the three central methodological approaches of this thesis are briefly introduced in the following. Section 1.4.1 describes the facets and particularities of a distributional analysis. The implications and opportunities from such a distributional perspective on regression models are reflected in section 1.4.2. In the final section 1.4.3, an empirical framework of how to operationalize the distance sensitivity of agglomeration externalities is briefly elaborated. The simulation model approach used in chapter 8 is extensively described in BRENNER and WERKER (2007).

1.4.1 Distributional analysis of growth rates

The distribution of growth rates serves in this thesis both as an *explanandum* and *explanas* (see section 1.2.2). On the one hand, focusing on other moments than the mean value is a requirement directly stemming from the heterogeneity of economic actors and activities. In the words of DOSI et al. (2010: 1982) one should not ignore the "rich statistical structure in dynamics of firms and industries, which goes well beyond the ones identified simply focusing on the average relations between performance and characteristics". On the other hand, knowing the characteristics of the entire distributions provides insights into the

[1] An increasingly proposed alternative to continuous space modelling is multilevel modelling, in which cross-level interaction effects can be addressed (VANOORT 2014, HUNDT/STERNBERG 2014).

[2] Even the functioning of institutions might be broken down spatially, because they are intrinsically related to the activities of people. (Formal) institutions are implemented by policy activities, enforced in courts, and surveilled by the police. It can be safely assumed, at least in more developed countries, that these functions are uniformly provided across space.

dynamic changes and processes of economic systems from an explicit evolutionary perspective.

The main empirical and methodological challenge resides in determining the shape and specification of the distribution function. In the earlier studies (e.g. AMARAL et al. 1997, BOTTAZZI et al. 2001), a tent-shaped Laplace distribution was proposed, thus refuting a normal distribution as would follow from a Gibrat-like growth process (GABAIX 2009, HALVARSSON 2012). Later studies found that "the Laplace distribution of growth rates cannot be considered as universal property" (BOTTAZZI et al. 2011: 104), as the growth rates tend to be asymmetrically distributed with tails even fatter than the Laplace. Therefore, more flexible distribution models are suggested. FU et al. (2005) and SCHWARZKOPF et al. (2010), for instance, model the central part as Laplace, but allow the tails of the distributions to decay as a power law. This approach, however, requires two different distributional models and a parameter which decides between both of them.

An alternative approach is the Asymmetric Exponential Power (AEP) distribution, as described by BOTTAZZI and SECCHI (2011), which resembles a flexible family of fat tailed distributions. FAGIOLO et al. (2008) also prove empirically that it is the preferred theoretical distribution in the context of economic growth as compared to alternative models, like the Student-t, Cauchy or Levy-Stable distribution, because it is especially flexible in the tail behaviour. In accordance with a huge body of empirical literature, the AEP distribution function F_{AEP} is used in this thesis. Having five parameters, it is highly flexible and can cope with both asymmetry and leptokurtosis.

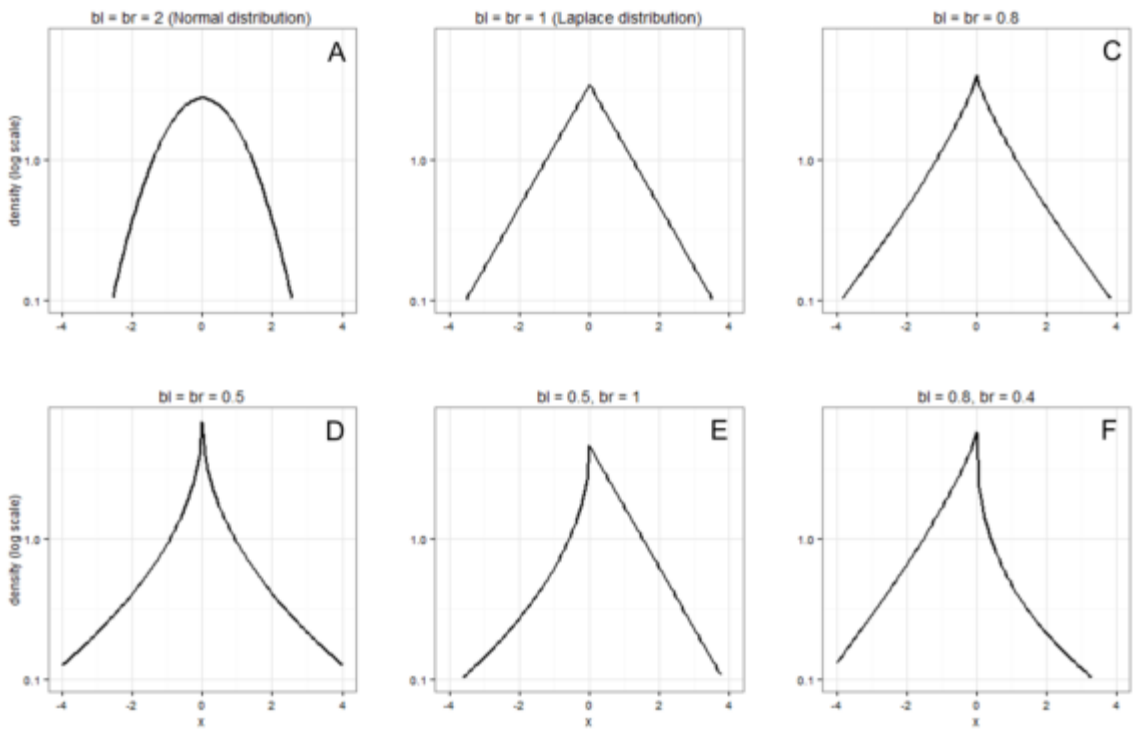


Figure 3: The AEP distribution for different shape parameters b_l and b_r

Figure 3 illustrates the AEP distribution for different parameter sets (mathematical details are discussed in chapter 2). The scale and location parameters are held constant, while the shape parameters b_l and b_r are varied. The smaller the shape parameters, the fatter are the tails at the respective side of the mode (see plots A to D in the figure). It is also called a distribution family, because both the shapes of the normal (plot A) and Laplace (plot B) distribution are special cases and it allows for a continuous variation from non-normality to normality. If b_l equals b_r, it is reduced to a symmetric version (again, plot A to D). As chapter 2 will show, the less general distributions of this family, like the normal distribution, the Laplace distribution or the symmetric Exponential Power distribution, are outperformed due to the presence of fat tails and asymmetry. Plot E and F illustrate two parameter sets which are more representative for the empirical data.

In this thesis, the parameters of the AEP distribution are obtained by solving the Maximum Likelihood (ML) function (see BOTTAZZI/SECCHI 2011 and ZHU/ZINDE-WALSH 2009 for a discussion on the asymptotic properties of the ML estimators). As suggested in earlier works (e.g. MINEO 2003), the ML framework is preferred over Moments estimators. An advanced estimation scheme, drawing on BOTTAZZI (2012), is developed in chapter 6. In this chapter, the AEP distribution is estimated while leaving out the central part around zero on the ground of particularities in the employment growth of firms, which is characterised by a high frequency of zero-growth events. Furthermore, a conditional estimation approach of the AEP distribution is introduced in chapter 7. The distributional function of the conditional AEP for the growth rates *g* reads:

$$\begin{aligned} F_{AEP}(g;\ b_l &= b_{l,0} + \beta_{bl}\boldsymbol{X}, \\ b_r &= b_{r,0} + \beta_{br}\boldsymbol{X}, \\ a_l &= a_{l,0} + \beta_{al}\boldsymbol{X}, \\ a_r &= a_{r,0} + \beta_{ar}\boldsymbol{X}, \\ m &= m_0 + \beta_m\boldsymbol{X}) \end{aligned} \tag{4}$$

where **X** is a matrix containing the conditioning variables and $\{\beta_{bl}, \beta_{br}, \beta_{al}, \beta_{ar}, \beta_m\}$ a set of coefficients containing the magnitude of the effects of *X* on the distributional parameters $\{b_{l,0}, b_{r,0}, a_{l,0}, a_{r,0}, m_0\}$. This extension can be used to assess the impact of **X** on the entire shape of the distribution, i.e. the *simultaneous* impact on the central location, the fluctuations and the tails. In the subsequent section 1.4.2 it is argued that this approach combines advantages from an unconditional distributional analysis with insights usually gained from regression models.

Before continuing with the discussion of the regression methods used in this thesis, the analysis of growth rate distributions additionally has to deal with the variance-scaling issue, which is declared to be a universal law in the growth of all complex organisations (STANLEY et al. 1996, AMARAL et al. 2001). The variance-scaling law describes the relationship between the size of an entity and the variance of its growth rate. To illustrate this law, Figure 4 separates the regional growth rate densities in the electronic industry into two groups: regions which show above average number of employees in this industry are depicted by a black line and regions with below average number of employees by a grey line. It

becomes immediately evident that the variance of the latter group is higher than of the variance of former group.

Simply stated, larger firms or regions fluctuate less than their smaller counterparts, usually by a factor that follows a power-law. Therefore, chapter 2 tests the variance-scaling relationship for regional economies using a parametric approach as suggested by BOTTAZZI et al. (2014), which also takes into account that the functional form of the relationship might be non-linear. In chapter 7, the variance-scaling is tested among Chinese firms and deviations from Western economies are explained by referring to existing theoretical models. To finally be able to pool the growth rates of differently sized firms or regions together as a result of the same underlying stochastic process, the growth rates are rescaled in the other chapters by using the empirically observed scaling relationship.

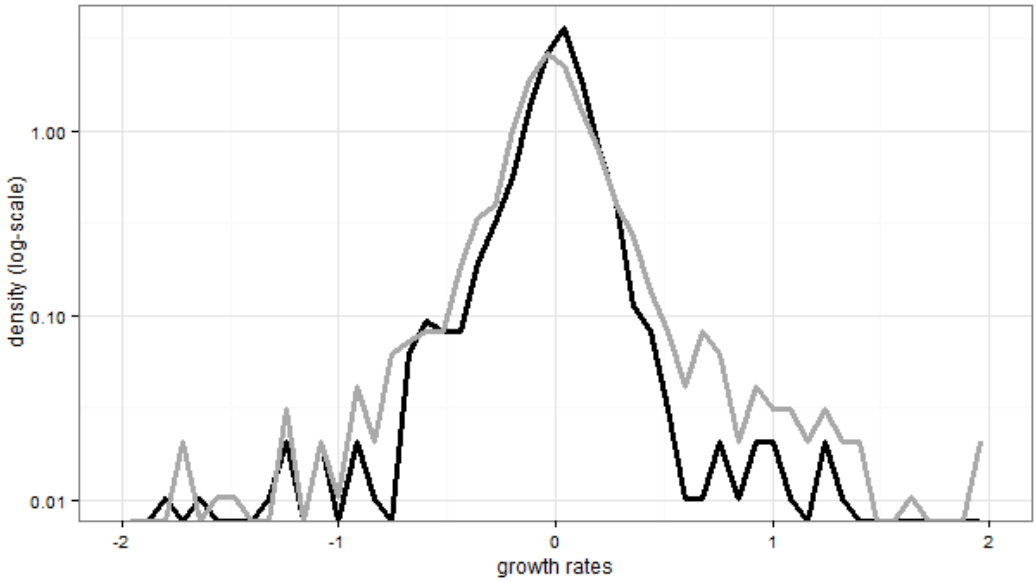


Figure 4: Growth rate densities for regions with above (black line) and below (grey line) average number of employees

1.4.2 Distributional characteristics and regression models

A (unconditional) distribution analysis misses the insights on the effects of further variables usually gained from a regression model. However, as already noted by MOSTELLER und TUKEY (1977: 266), a “regression often gives a rather incomplete picture. Just as the mean gives an incomplete picture of a single distribution”. To take into account the entire distribution of the growth rates, two estimation strategies are proposed in this thesis.

First, the meanwhile well-established quantile regression (e.g. KOENKER 2005) is a useful tool to analyse specific quantiles of the conditional distribution of the dependent variable. From a conceptual point of view, the economic relationships at the highest and lowest quantiles, i.e. of the best and worst performing firms, are of special interest, because the average growing firm tends to grow closely at rate zero and thus does not contribute much

to the overall turbulences in market share or employment (COAD 2007). In the literature on high-growth firms it is increasingly recognized that these firms disproportionately contribute to labour market dynamics (e.g. STOREY 1994, DELMAR et al. 2003 or COAD et al. 2014). From an econometric point of view, quantile regression estimators are also more efficient than OLS estimators when the error term is non-normal (BUCHINSKY 1998). In case of a Laplacian distributed error term, the least absolute deviation estimator becomes the ML estimator. In all other cases, the estimator is less affected by observations residing in the tails (DASGUPTA/MISHRA 2004). Hence, quantile regression is used in chapter 4 and 5 for studying the relationship between firm growth and its spatial environment.

However, ordinary regression models are designed to capture the effects on the average value or some given quantiles, or in other terms, the "location-shift effect […] in the conditional distribution of the dependent variable" (BOTTAZZI/SECCHI 2013: 2). Following a distributional perspective it becomes clear that "the variables may continue to impact on other distributional characteristics" (MAASOUMI et al. 2007: 449). Taking into account the fundamental heterogeneity of economic processes and, as a consequence thereof, the expected heterogeneity in the responses of economic entities to specific independent variables, a focus solely on the location-shift effect draws a picture that is much too narrow. The conditional estimation approach of the AEP distribution, as already introduced in the previous section 1.4.1, combines the advantages from both ordinary regression models and unconditional distributional analyses. Compared to regression approaches, it accounts for the possibility that the variables might impact on the entire shape of the distribution *simultaneously*. But it also goes beyond the insights from an unconditional distributional analysis. For instance, studying the factors leading to shifts in the distributional mass towards the tails is of high interest given the importance of the fastest growing firms with respect to the dynamics of the aggregate economy. Moreover, high-growth firms, which populate the tails of the distribution, might now be studied without the requirement of delimiting a sub-population *ex ante* (on the difficulties of measuring high-growth firms, for example by choosing some growth rate threshold see DELMAR et al. 2003 or more recently COAD et al. 2014).

Both approaches, quantile regression and conditional distributional analysis, however, cannot address the issue of causality. Causal discovery in regression models aims to go beyond simple correlations and to uncover the causal structure in the variables. This is crucial to understand the *causes* behind economic growth processes, especially in the context of regional growth dynamics, in which the variables are often highly endogenous (PACK 1994) and ordinary regression methods fail to capture the subtleties of causal feedback (GARNSEY et al. 2006). Therefore, chapter 3 introduces a method which exploits additional information from the growth rate distributions to identify the causal structure in vector autoregression models. From a methodological point of view, this method is particularily interesting as it requires the condition of non-normality – and therefore it is fundamentally different from classical statistical methods including regression models that rely on the assumption of normality (HYVÄRINEN 2013).

1.4.3 Distance decay of agglomeration economies

TOBLER's (1970) first law of geography states that everything is related to everything else but near things even more so. Consequently, the impact of agglomeration economies should decay with (geographical) distance. However, not much is known about the functional form and spatial range of this decay. Theories, say, the New Economic Geography or cluster literature, do not provide any predictions on the spatial dimension of externalities (HOOGSTRA/VANDIJK 2004, GRAHAM et al. 2010). Hence, empirical studies have used a wide range of different functional specifications, ranging from binary distance circles (e.g. HOOGSTRA/VAN DIJK 2004, BALDWIN et al. 2008, ERIKSSON 2011), inverse linear weights (e.g. SORENSEN/AUDIA 2000, GRAHAM 2009, MAINE et al. 2010) to more complex power or exponential decay functions (e.g. RICE et al. 2006, ANDERSSON/KARLSSON 2007, DRUCKER/FESER 2012). Besides, it is argued that the form of decay and the spatial extent might vary by industry or source of the agglomeration economies. Hence, the specification of the distance decay function ultimately remains an empirical question (GRAHAM et al. 2010). For instance, LYCHAGIN et al. (2010) propose a semi-parametric estimation of the functional form in the spatial weight matrix.

The approach used in this thesis is to employ a flexible decay function, derived from behavioural assumptions, and to determine its parameters from the data. The function proposed is the S-shaped log-logistic decay function. This function can be described by two parameters, *s* for the steepness of the decay and *r* for the spatial range of the inflexion point (DEVRIES et al. 2009). In all plots of Figure 5, *r* is set to the distance of 100. In the left plot, the log-logistic decay function is drawn for an intermediate steepness of *s* = 5. At *s* = 1 (middle plot), the function turns into a linear decay function. The binary decay function is the solution in which *s* approaches infinity (right plot). Hence, the generality of the log-logistic distance decay function let discriminate empirically among the more simple specifications often found in the literature.

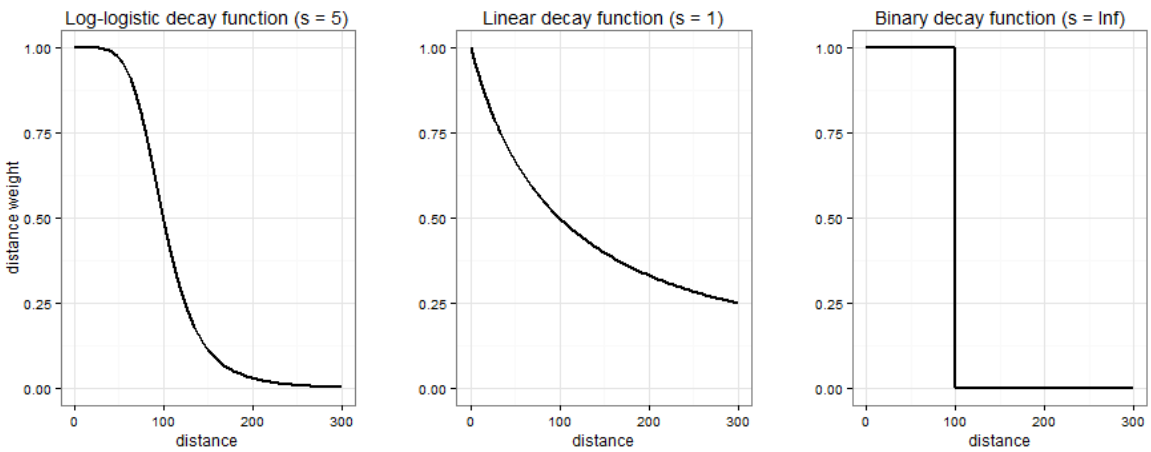


Figure 5: The log-logistic distance decay function

Theoretically, the general shape of this functional form can be grounded on THORSEN et al. (1999), who argue that within short distances random interactions take place, whereas interactions over long distances are increasingly governed by the minimum cost principle. This non-linear influence of distance, with the highest sensitivity at intermediate distances, was first observed in the Swedish literature on commuting behaviour (e.g. JOHANSSON et al. 2003) and described as an S-shaped curve of willingness to commute. Further empirical evidence on a non-linear decay also exists in the recent literature on agglomeration effects (MAINE et al. 2010, LYCHAGIN et al. 2010).

By including the distance decay function into a linear model with spatially discounted explanatory variables, as proposed by ANDERSSON and GRASJÖ (2009), the decay function parameters can be directly estimated from the data. As further suggested by GRAHAM et al. (2010), that parameter combination is chosen which maximizes the explanatory power of the model, i.e. which contributes most to the explanation of firm growth. Of course, the spatial decay can be separately studied and compared for different kinds of firms, industries or spatial sources of externalities, making this approach particularly suited against the background of heterogeneity.

1.5 Summary on the literature of growth rate distributions

One of the main emphases of this thesis lies in the distributional analysis of the growth rates of economic entities. Therefore, this section summarizes the literature on the distributional characteristics of growth rates. The stochastic patterns of economic growth can be studied at different levels of aggregation, starting from the basic building block of the firm, to sectoral aggregations like industries, or to spatial aggregations like regional or national economies. Besides, the distribution of an entire population of entities can be analysed in a cross-sectional manner, or sub-samples of the same type of observational unit can be compared regarding their spatial location, sectoral affiliation or temporal dimension. Such systematic comparisons by means of sub-samples according to a chosen reference system provide first empirically grounded insights into the generating mechanisms of the distributions. Table 1 classifies the literature into a taxonomy which distinguishes the different aggregation levels of the observational unit and the reference system for an optional comparison.

Table 1: Literature overview on empirical studies on (unconditional) growth rate distributions

Unit of observation	*Reference system*			
	Cross-sectional	Spatial comparison	Sectoral comparison	Temporal comparison
Firms	STANLEY et al. (1996) AMARAL et al. (1997) BOTTAZZI et al. (2001)	DUSCHL (2014) (for regions)	BOTTAZZI/SECCHI (2003) REICHSTEIN/JENSEN (2005) BOTTAZZI et al. (2007) BOTTAZZI et al. (2011)	ERLINGSSON et al. (2013)
Industries	SAPIO/THOMA (2006)	CASTALDI/DOSI (2006) CASTALDI/SAPIO (2009) (for countries)		
Regional economies			DUSCHL/BRENNER (2013)	
National economies	CANNING et al. (1998) LEE et al. (1998) AMARAL et al. (2001) CASTALDI/DOSI (2009)			FAGIOLO et al. (2008)

From the literature overview in Table 1 it becomes clear that the distributions of growth rates are well studied at the levels of firms (e.g. STANLEY et al. 1996) and countries (e.g. CANNING et al. 1998). Subsequently to these early studies, the literature was extended to take into account the sectoral dimension, either by directly looking at industries as an economic entity by its own (SAPIO/THOMA 2006), or by comparing firm growth rates of different industries (e.g. BOTTAZZI/SECCHI 2003).

The aim of this thesis is to fill the research gap at the intermediate spatial level of aggregation, namely regional economies. Therefore, chapter 2 investigates the characteristics of regional growth rate distributions, already assuming an industry-specific perspective. In contrast hereto, chapter 6 takes the regional level as a reference system to analyse the growth rate distributions of firms located in the regions. The latter perspective is especially interesting as it provides numerous estimated values of the distributional parameters, which in a next step can be systematically compared and related to other regional variables. Comparable studies that exploit the richness of inter-regional variations in the distributional manifestations are BARBOSA and VASCO (2011) testing for GIBRAT's law of firm growth for 19 Portuguese regions, or OKUBO and TOMIURA (2014) investigating the shape of the plant productivity distribution across regions and linking them to economic geography variables.

Finally, one should also mention a third dimension of comparison, the temporal dimension, which was firstly explored by FAGIOLO et al. (2008), who study the evolution of national growth rate distributions across time. This dimension would be also highly interesting at the level of regions or firms, especially if the growth rate distributions are related to specific historical contexts. For instance, by comparing the shape of the distributions before, during and after the recent financial and economic crisis, a more complete picture on the impact of macroeconomic recessions might be provided. Thus, the literature overview shows that there are still promising research gaps to be filled by analysing the units of observations with different reference systems.

From the analysis of growth rate distributions a tent-shaped distribution was found independently of the levels of aggregation. Hence, it was claimed that the growth process of complex economic organizations follows some universal mechanisms, which are independent of the particular details of the system (AMARAL et al. 2001), or more simply stated, the "evolution of organizations with complex structures is governed by similar growth mechanisms" (LEE et al. 1998: 3275). However, the spatial and sectoral splitting of the population into disaggregated regions and industries suggest that the parameters of the common distributional shape differ to a considerable degree. By relating the parameters to regional or industrial variables, like population density or knowledge intensity, some hints on the specific manifestations of the underlying mechanisms at the respective level can be gained. Finally, in chapter 7 other variables are directly included in a conditional estimation approach of the distribution. By theorizing on the underlying mechanisms, it shows how and why the ownership type leads to shifts in the distributional mass of Chinese firms.

1.6 Data

This thesis relies on three different kinds of data: firm data, regional data as well as data on travel time distances between the locations of the activities. In this section, the variables, data sources and related issues are discussed in more details.

1.6.1 Firm data

Chapter 4, 5 and 6 investigate the growth dynamics of German firms. Firm data is obtained from the Bureau von Dijk's Markus (chapter 4) as well as Amadeus databases (chapter 5 and 6). These databases provide information on an extensive sample of firms that are located within Germany. Besides the number of employees, other variables like turnover, export and import shares, industry affiliation or founding date can be retrieved. These information are available for several years, however the coverage of the data entries abruptly decreases after about five years from the access year into the past.

Although the population of German firms is covered to a great extent, some drawbacks exist. They are of general nature and concern most of the firm databases used in the literature. On the one hand, very small firms with just a handful of employees, sometimes denominated as micro-firms, tend to be underrepresented. This also holds for very young firms (COAD 2009). Figure 6 shows the age distribution of the sample's firms. The expected exponential decay (COAD 2010b) seems to be given only after around ten years.[3] As clearly visible, the data lacks entries of very young firms. However, a large share of the very young

[3] The highest peak at around 20 years is an artefact of the German re-unification and the two bumps result from the two world wars. The histogram is cut at the age of 120 years, which only affects 167 firms. The oldest firm, the Dierig Holding AG, was founded in 1805.

firms exit the market again, as the high correlation between entry and exit rates indicates (DUNNE et al. 1988).

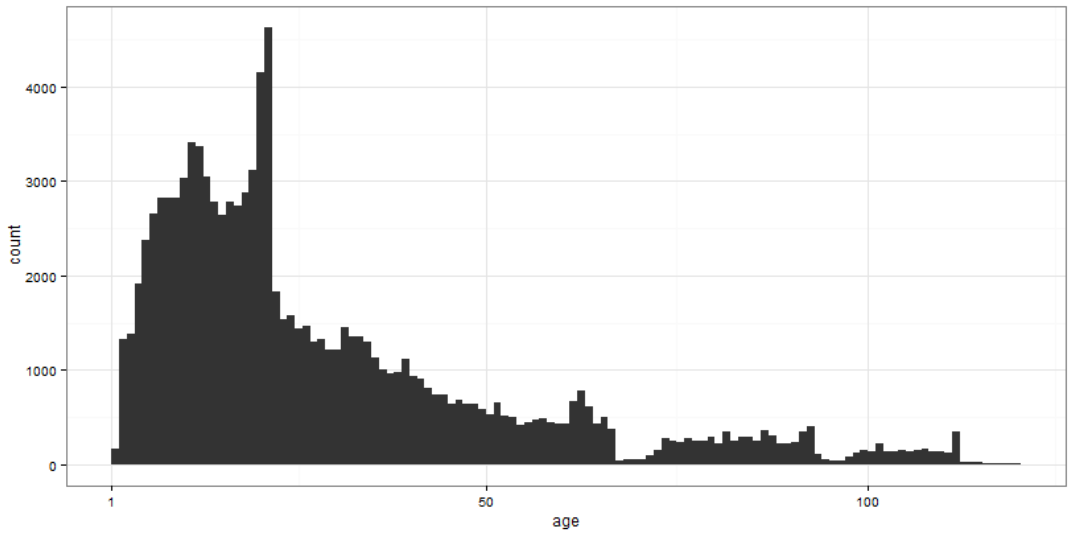


Figure 6: Age distribution of the firms in Amadeus

On the other hand, firms that have exited the market do not appear in the database. This implies that a sample selection bias might be present because the characteristics of the exiting firms cannot be taken into account, for example by using Heckman's two-step estimator (HECKMAN 1979). The interpretations of the results are, hence, constrained to the group of surviving and still existing firms. Furthermore, owed to the information procurement of the database provider and the lack of a compulsory reporting requirement of the number of employees for firms not listed at the stock market, no balanced panel is available for all firms. It was decided to use an unbalanced panel, as this guarantees the most extensive coverage of the diversity of firms and avoids introducing additional biases. Besides, an unknown share of the database entries for one year is simply taken as a repeated value of the previous year. Chapter 7 provides an approach to control for this data quality problem in the estimation of growth rate distributions.

Finally, the firms are geo-located by using the address of their headquarters. The spatial distribution of the firms' locations is visualized in Figure 7. Instantly, well-known centres of economic activities become visible. They mainly resemble the metropolitan areas like Berlin, Hamburg, Munich or Stuttgart as well as the Rhine-Ruhr-Main area in Western Germany.

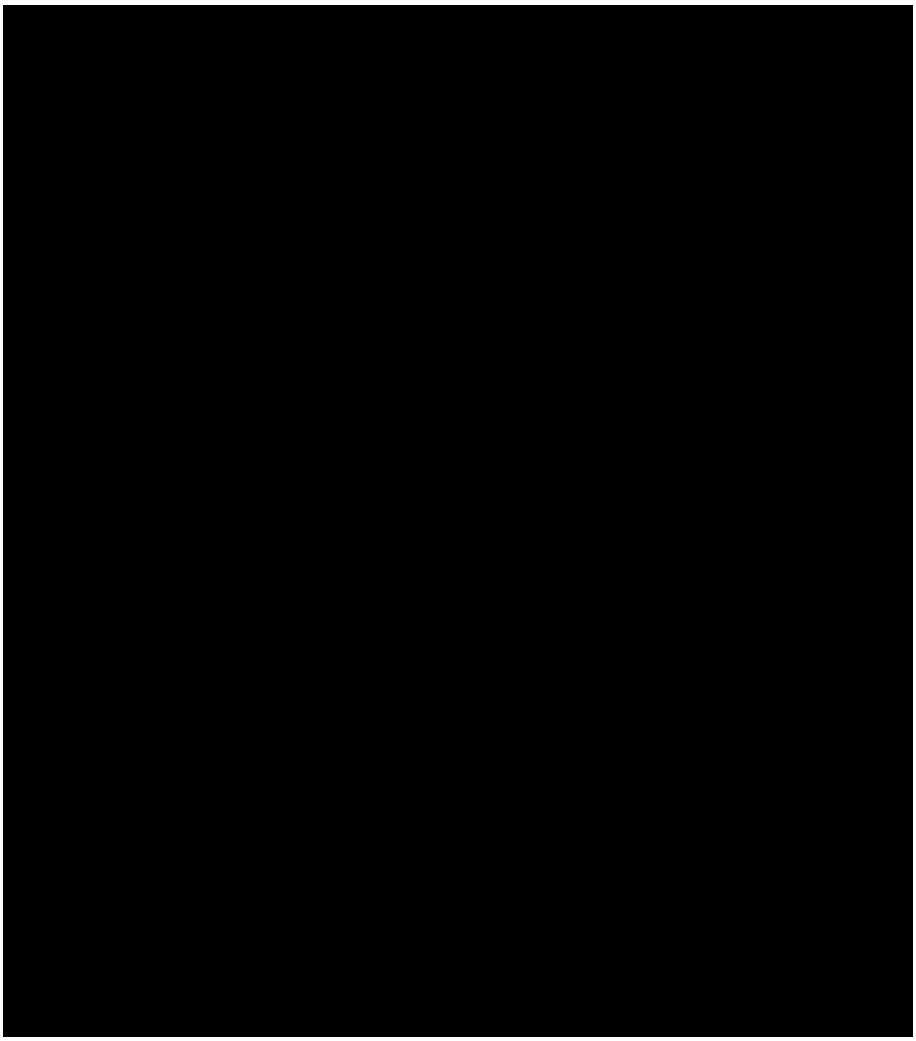


Figure 7: Spatial distribution of firm locations

However, many firms disperse their activities in increasingly independent plants across the country and even beyond (IAMMARINO/MCCANN 2013). This holds especially true for larger firms, for which the results on the external factors thus should be interpreted with more care. Nonetheless, we follow BEAUDRY and SWANN (2009) in assuming that it is the regional environments of a firm's headquarter which is most decisive in affecting the growth prospects and the decision of the organization on taking up employees. Besides, from an empirical point of view, there is an almost perfect correlation between the number of firms from the Amadeus database and the number of plants within the labour market regions (the PEARSON's correlation coefficient is slightly above 0.98). This also indicates that the spatial sampling of the firms from Amadeus is representative, which is in particular important in the light of merging firm data with regional data.

Chapter 7 extends and compares the existing evidence on firm dynamics in Western economies to the firm dynamics in China. In emerging economies, issues like the data quality and the role of the ownership type become particularly relevant. These issues are discussed in the corresponding chapter at length.

1.6.2 Regional data

Depending on the level of analysis, regional data is either used for deriving the variables of interest (chapter 2 and 3) or for constructing explanatory variables for firm performance (chapter 4, 5 and 6). Data on various aspects of the spatial economic landscape is obtained: the number of employees (from the German Institute for Employment Research IAB), the number of publications (from ISI Web of Science), the number of patents (from the European Patent Organization's Worldwide Statistical Patent Database), and the number of university graduates (from the German statistical office destatis). Further regional level control variables, say, the population density or the unemployment rate (both from destatis), are also considered.

The spatial distributions of the four main technological activities (see section 1.3.2) of a regional economy are mapped in Figure 8, representing one specific industry and year, namely machine tools in 2007. In general, the more innovation and knowledge-related activities, measured by R&D-related employees, patents or university graduates (with Gini coefficients of 63.4, 66.7 and 82.5, respectively), happen to be more concentrated than general employees (with Gini coefficient of 56.7) or the population (with Gini coefficient of 51.3). This is in line with findings and theoretical models from the literature, especially with BATTY (2008, 2013) and what is called by him the 'New Science of Cities'. The machine tools industry is in particular strongly represented in the southern parts of Germany. The eastern part of Germany tends to lack behind, especially regarding patent activities. The larger metropolitan regions naturally dominate in absolute terms. Whereas the spatial distributions of employees, R&D-related employees and patents highly correlate (with a Spearman correlation coefficient always higher than 0.75), university graduates show a different picture: their correlation coefficient with the other three activities is always lower than 0.5. This is partly driven by some regions, like Marburg, which host universities but hardly show any economic activities in the corresponding industry.

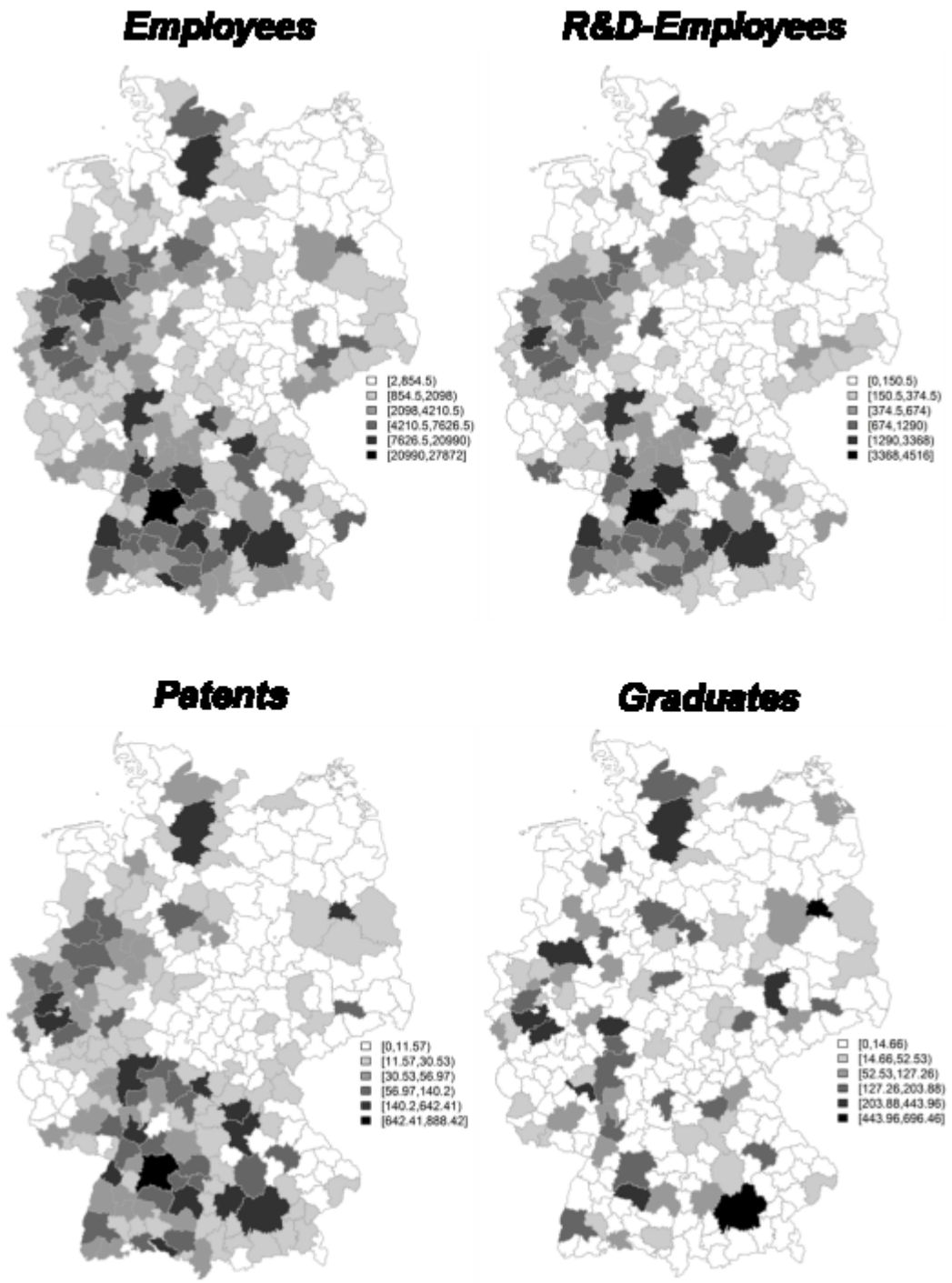


Figure 8: Spatial distribution of employees, R&D-related employees, patents and university graduates for the machine tool industry in 2007 according to the IAB labour market regions

Regarding the use of regional data, mainly two issues arise. First, how can the variables which rely on different classification schemes, like the patent and the industry classification, be matched such that a disaggregated industry-specific analysis becomes possible? Secondly, which is the appropriate spatial aggregation level for analysing the research questions with this kind of data?

The matching issue primarily concerns chapter 3, in which the industry-specific growth dynamics among the regional variables are studied. As will be described in this chapter in more details, data on graduates and patents are matched by counting all patent applicants with a professor title and by relating the study field of their university faculty to the patent classes. Subsequently, data based on the International Patent Classification is matched with data based on the standard industry classification according to an existing concordance matrix for Germany (SCHMOCH et al. 2003). This matching procedure results in a unique database, which embraces multi-dimensional aspects for clearly separated industries.

The aggregation issue concerns all chapters together. In chapter 4 and 5, in which space is conceptualized as distance, this issue is explicitly addressed. Here, all data is provided at the level of municipalities, the lowest level of aggregation in Germany. DURANTON and OVERMAN (2005) refer to this already as micro-geographic data. For the other chapters, space is conceptualized as a container, in which economic activities take place, expand or decrease in intensity over time. Depending on the research question and the nature of the variables, different regional delimitation systems are used. Functional labour market regions are the natural delimitation system for economic activities. For Germany, there exist two commonly used definitions for functional labour markets, the definition by ECKEY et al. (2006) and the more official definition by the IAB (see Figure 9). Both definitions aggregate the administrative districts using data on commuter flows but by applying different methods. A more extensive discussion on the differences can be found in ECKEY et al. (2007). For other socio-economic aspects, like the population density, administrative districts are more appropriate.

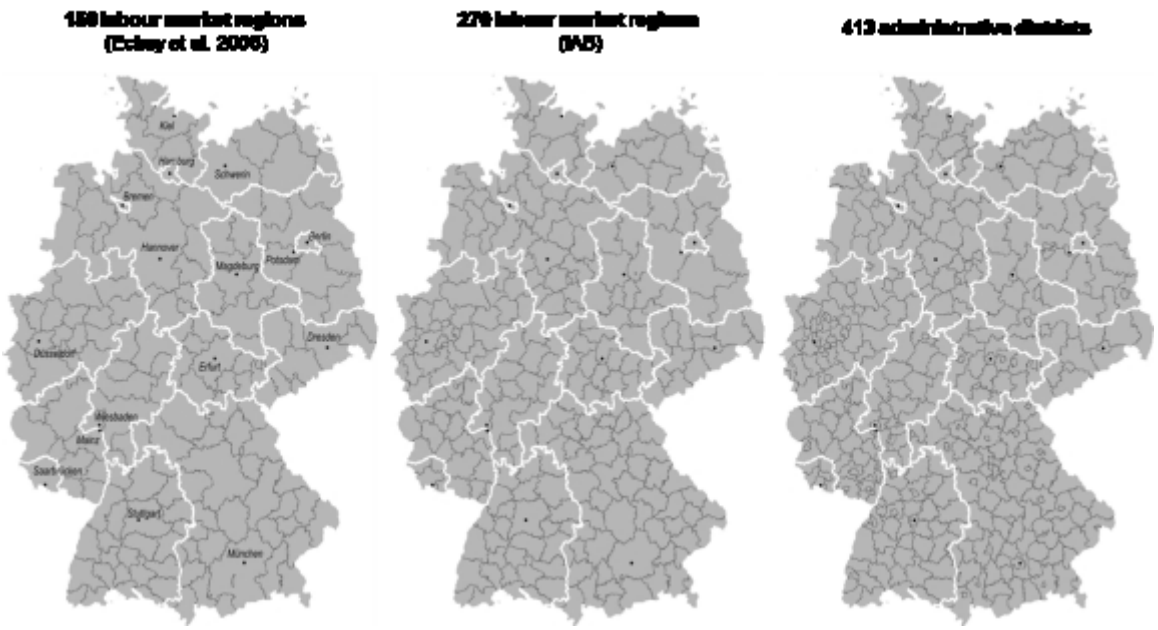


Figure 9: Regional delimitation systems

1.6.3 Travel time distances

Data on distances is required in chapter 4 and 5 to model the firms' accessibility to the knowledge generating activities that are unequally distributed in space. Compared to the geographic distance, the driving distance or travel time is related more directly to the frequency of interactions (ANDERSSON/GRASJÖ 2009). Travel time not only depends on the geographical distance, which still remains the frame for the interactions (RODRIGUEZ-POSE 2011), but also on the physical infrastructure (ANDERSSON/KARLSSON 2007).

The bilateral travel time distances between the locations of the firms to the locations of the activities are calculated by exploiting results from graph theory and data on the German road network from the OpenStreetMap project. The route planning algorithm is more extensively described in GEISBERGER et al. (2010). It allows to efficiently calculate travel times for many locations. As described in the previous section, municipalities are chosen here to represent the locations both for the firms and other activities. Because of the traditionally high computational costs implied by route planning, which increase exponentially with the number of nodes in the interaction network, travel time data has not been used before in an extensive manner or only for incomparable small regions (for instance DAUTEL/WALTHER 2013 for Luxembourg).

The Mantel test (MANTEL 1967) can be used to compare the travel time distance matrix with a geographic distance matrix based on Great circle distances. The Mantel test is a non-parametric approach to compute the correlation between two distance matrices and its test statistic can be interpreted like the Pearson coefficient. In case of the travel time matrix and the geographic distance matrix it is 0.89. Hence, the correlation between both matrices is high, however, not perfect. For example, from Berlin to Griesbach im Rottal (Lower Bavaria) it takes around 1.5 as much in travel time as from the mere geographic distance between both municipalities could have been expected (see Figure 10). The correlation coefficient decreases even further to 0.82 when only distances below 120 minutes are considered. This is also the spatial scale within which growth relevant externalities are expected to work. Therefore, the use of travel times becomes even more crucial.

Figure 10: Road distance (black line) versus beeline (grey line) from Berlin to Griesbach im Rottal

Although more realistic than geographic distances, travel time data from route planning is still imperfect, as it considers only the average travel time for specific road types. In the present case, no information on alternative travel modes, like public transportation or flight transportation, is available. Against the former it can be argued that public transportation, like trains, roughly correlate with private cars in time and costs of traveling, and regarding the latter that flight traveling is generally less relevant at this level of analysis, which mostly encompasses within labour market interactions.

1.7 Overview on the content of the papers

In this introductory chapter, the research questions, the underlying theoretical frameworks and concepts, methodological issues as well as the data sources of the thesis were discussed and the relevant literature was summarized. The introductory chapter concludes by providing an overview on the content of the following seven research papers, indexed by their chapter number (see

Table 2). In chapter 2, the stochastic patterns of regional growth processes are analysed. In chapter 3, the causality of the regional growth processes is explored by applying an identification method that exploits the information in the stochastic patterns of the growth rate distributions. In chapter 4, the focus is shifted to the spatial matters of growth factors, that is why the micro-level of firms has to be assumed. Chapter 5 extends this analysis of the relationship between agglomerations and growth by explicitly taking into account an industry-specific perspective. In chapter 6, the differences of the firm-growth rate

distributions among regions are investigated and discussed in the light of the concept of regional resilience. In chapter 7, existing evidence on the stylized facts in firm dynamics is extended to China, and a conditional estimation approach of the growth rate distribution is introduced to account for the role of the ownership type. Finally, chapter 8, which performs an empirical calibrated simulation model, represents a synthesis in two senses: first, it explains the emergence of the growth rate distributions, as they are observed in the other chapters, and it does by providing a potential framework for linking regional growth processes to the elementary firm dynamics.

Table 2: Overview on the research questions, concepts, methods and data of the research papers

C.	Research questions	Theoretical concepts	Level of analysis	Definition of regions	Main methods	Data
2	How do the distributions of regional growth rates look like? Are there systematic differences between industries or the time period?	Gibrat's law, stochastic growth models	Regional (macro)	Functional and administrative	AEP estimation, LR-test	*Regions*: Employees
3	How do endogenous regional activities evolve and cause each other to grow? Are there differences across industries?	Regional systems of technological activities	Regional (macro)	Functional	SVAR, ICA	*Regions*: Employees, graduates, patents
4	How and over which distance do external knowledge sources impact on firm growth?	Resource-based view of the firm, absorptive capacity, knowledge spillovers	Firms (micro)	Firm-specific	Quantile regression, Distance decay functions	*Firms*: employees *Regions*: Employees, Universities' budget, third-party funding, graduates
5	How do agglomerations of economic and knowledge generating activities affect firm growth?	Industrial clusters, agglomerations	Firms (micro)	Firm-specific	Quantile Regression , distance decay functions	*Firms*: employees, *Regions*: employees, publications
6	How do firm growth rates differ across regions? Which regional factors affect the fatness of the tails?	Regional resilience	Firms (Micro), Regions as reference system	Functional	Advanced AEP estimation	*Firms*: employees *Regions*: various variables
7	How are firm growth rates distributed in China? What is role of the ownership type?	stochastic growth models, emerging economies	Firms (Micro)	a-spatial	Conditional AEP estimation	*Firms*: employees, sales, productivity
8	How do the stylized facts of growth rate distributions emerge?	stochastic growth models, resource-based view of the firm	Firms (Micro)	a-spatial	Simulation, empirical calibration	-

2 Characteristics of regional industry-specific employment growth rates' distributions

This chapter is a reprint[4] of:

DUSCHL, M. & BRENNER, T. (2013): Characteristics of regional industry-specific employment growth rates' distributions. In: Papers in Regional Science 92: 249-270.

Abstract: Regional growth dynamics deviate from a normal distribution. Using industry-specific employment data for German regions, we find that the asymmetric exponential power distribution best accounts for the high frequency of extreme positive and especially negative growth events. This research confirms previous studies on growth rates of firms, industries and countries and fills the research gap at the level of regions. Furthermore, we show that even in the case of prolonged time lags, regional industry-specific employment growth rates are far from being normally distributed, move towards a Laplacian shape and that knowledge intensive industries increase the regional economies' risk of being affected by extreme events of growth and decline.

Keywords: Regional employment growth, Gibrat's law, scaling, asymmetric exponential, power distribution, extreme events

JEL Classification: C46, C50, R11

[4] Due to better readability, numbering of sections, tables, figures and formulas has been changed. References and Appendix can be found at the end of the thesis.

2.1 Introduction

The analysis of stochastic characteristics of economic phenomena has played a prominent role in economic research, both in the past and in the present. The seminal work of GIBRAT (1931) paved the way for comprehensive investigations on statistical properties regarding the size distribution of firms (e.g. HART/ PRAIS 1956, SIMON/BONINI 1958, LUCAS 1978, AXTELL 2001) and its relation to firm dynamics in terms of an autoregressive stochastic process (e.g. MANSFIELD 1962, EVANS 1987, HALL 1988). Later studies have begun to focus on the shape of the distribution of firm growth rates (e.g. STANLEY et al. 1996; AMARAL et al. 1997; BOTTAZZI/SECCHI 2003). Recent empirical evidence on the basis of firm level data from several countries (e.g. BOTTAZZI et al. 2002 for Italy, BOTTAZZI et al. 2011 for France, DUSCHL et al. 2014a for Germany, or REICHSTEIN/ JENSEN 2005 for Denmark) as well as at the disaggregated level of industries (BOTTAZZI et al. 2001 for the pharmaceutical industry) shows that the expectation of normally distributed growth rates stemming from GIBRAT's framework is consistently rejected (AMARAL et al. 1997: 5). Rather than the bell-shape of a normal curve, an exponential tent-like shaped distribution is observed, with tails that are significantly fatter than the ones of a normal distribution. In other words, growth events occur with a higher probability at the extremes (BOTTAZZI/ SECCHI 2006a). Similar findings have been subsequently reported for countries, at a much higher level of economic aggregation. Whilst QUAH (1996 and 1997) and JONES (1997) amongst others studied the income distributions of national economies, more recently the focus has shifted to the countries' income growth rates (e.g. LEE et al. 1998; CANNING et al. 1998, AMARAL et al. 2001, MAASOUMI et al. 2007, CASTALDI/ DOSI 2009).

Tent-shaped and fat-tailed distributions of growth rates are a robust feature for both firms and countries. In addition, they show a much higher regularity and homogeneity than the corresponding size distributions. Accepted as a stylized fact, these distributions have been further extended to and confirmed for the growth of whole industries within a country (SAPIO/THOMA 2006; CASTALDI/SAPIO 2008). Complementing this sectoral perspective, a logical next step is to look at the spatial intermediate level between firms and countries, namely at the level of regions. Whilst the distribution of economic activities across regions and spatial clustering tendencies of different industries are well studied (e.g. ELLISON/GLAESER 1997; DUMAIS et al. 2002; BRENNER 2004 and 2006; BOTTAZZI et al. 2008), the distributional characteristics of the regions' growth rates still remain an important research gap – neither the growth processes of regional economies as a whole, nor of the regions' industries have been investigated yet. Therefore, the aim of this paper is to explore and explain the stochastic properties of regional economic growth from an explicitly industry-disaggregated perspective.

Exploring the stochastic characteristics of regional industry-specific employment growth could reveal underlying mechanisms which govern economic growth. More precisely, we focus on the following research questions: do the average and variance of regional employment growth rates depend on the regions' number of employees? Are the growth rates correlated over time? And above all, are regional growth rate distributions fat-tailed, similar to firms and countries, and how does their distributional shape evolve when longer time lags are considered? *Explaining* the emergence of the observed stochastic patterns by considering industrial specificities could improve our understanding of contemporary

regional economic growth processes. In particular, we ask which types of industries fluctuate more in their regional employment and are more prone to experience extreme growth or decline.

Using a time panel of industry-specific employment data for German regions from 1999 to 2007, we first estimate the basic independence assumptions of the growth process within GIBRAT's framework. The revealed variance scaling of industry-specific regional employment growth rates will be removed by introducing an explicit heteroskedasticity term into the growth regression. This allows us to subsequently identify the best fitting theoretical distribution model for describing the rescaled growth rates. Starting from a two-parameter Laplace distribution, we then increase the number of free parameters and test whether the more general symmetric as well as asymmetric exponential power distribution fits the empirical observations significantly better. We find that the asymmetric exponential power distribution most adequately describes regional industry-specific growth rates, as it accounts for fat tails which are a prominent feature of the employment growth process – regional economies are heavily prone to experience extreme positive and especially negative industry growth events. Even if longer than one year time lags are considered, the mechanisms leading to fat tails do not vanish and the distributions approach a Laplacian shape. Finally, in an attempt to explain the emergence of the stochastic properties, we compare the distributional characteristics of regional growth rates of different types of industries. The most striking finding in this analysis lies in the observation that industries for which knowledge and innovation matter are strongly exposed to extreme events, both regarding expansion and decline. Hence, the industrial specialization structure of regional economies is a key to understand their growth dynamics, stability and resilience.

The paper is structured as follows. In the second section (2.2), the theoretical framework is outlined and predictions for the stochastic properties of regional industry-specific employment growth processes are presented. The methodology to identify the best fitting theoretical distribution is developed in the third section (2.3), and data issues are discussed in the fourth (2.4). In section five (2.5), the methodology is applied to German regions and economic explanations for the observed patterns are provided. Section six (2.6) summarizes and gives some tentative conclusions for both economic theory and modelling.

2.2 Theory and predictions

2.2.1 Gibrat's law and scaling relationship

GIBRAT's law of Proportionate Effect is a widely used starting point in the firm growth literature. Although economically meaningless, it provides a useful benchmark model to test the basic independence assumption of growth processes of economic entities (BOTTAZZI et al. 2011). For the weak version to apply it requires that growth rates are independent from the initial entities' size. Additionally, it is based on the assumption that successive growth rates are uncorrelated in time. Empirical findings on GIBRAT's law in case of firms are rather mixed, at times contradictory and mostly conditional on other

aspects like firms' age, size or the specific growth rates quantile under consideration (SUTTON 1997, or more recently COAD 2009 and LOTTI et al. 2009 provide an overview on the literature). Hence, we begin by assuming for regional economies:

> **Hypothesis 1a (Gibrat's law for the mean growth rate)**: *Regional employment growth rates are independent from the regions' number of employees and the regions' past growth experience.*

Beyond its' relevance in and of itself, *hypothesis 1a* also allows us to decide whether or not growth rates from different regions (in terms of number of employees) as well as years can be considered as random draws from the same population and, thus, can be pooled together for a subsequent analysis of the growth rates. For the strong version of GIBRAT's law to hold, two additional requirements must be met. First, the variance of the growth rates must be independent of the entities' size and secondly, growth rates must converge to a normal distribution. The former dates back to HYMER and PASHIGIAN (1962), who first discovered that the variance of the firms' growth rates is inversely related to their size, and by this violating GIBRAT's law for variance (GABAIX 2009). STANLEY et al. (1996) formalize this scaling relationship as a power law:

$$\sigma(S) = cS^{\beta} \qquad (5)$$
$$log\, \sigma(S) = c + \beta S$$

with S being the entities' size, $\sigma(S)$ the standard deviation of their growth rates conditional on S, c a constant and β the scaling exponent. Empirical studies for firms in different industries (e.g. AMARAL et al. 1997) and for national economies (e.g. LEE et al. 1998) estimate that β mostly ranges between -0.15 and -0.20. This robust power-law scaling regularity is declared to be a universal feature of the growth of complex economic organizations involving large numbers of interacting subunits (STANLEY et al. 1996, AMARAL et al. 2001). Two explanations for the observed scaling behaviour can be put forward. The first is supported by CANNING et al. (1998) and others, who consider the scaling exponent β as an indicator for the strength of micro-economic interactions between the subunits of a system, here the firms of a region.[5] If β is 0, a perfect correlation between the subunits exists, since there is no dependence of the dispersion of growth rates on the number of employees. If β is -0.5, any correlation between the subunits is absent, meaning that the volatility of a system falls with the square root of its size. Finally, if β lies between the two limiting cases, the variance of the growth rates decays as a power law, but not as fast as the square root, here indicating the presence of some considerable positive interactions between the firms within a region. BOTTAZZI and SECCHI (2006b) argue that the observed scaling relationship can be explained as a diversification effect, since larger economic entities tend to operate in a higher number of sub-markets of an industry due to scope economies of diversification. Up-scaling this argumentation from the level of firms to

[5] Scaling law models rely on the assumption of an identical size of the subunits, which strictly speaking does not hold in the reality of regional economies. This restricts to some degree the explanatory power of the scaling exponent as an indicator of the strength of the interactions between the firms within a region.

regions, the dispersion of overall growth may be reduced over-proportionally for regions with a higher number of employees. Both explanations lead us to:

> **Hypothesis 1b (scaling relationship)**: *The higher the number of regional employees, the smaller the variance of the conditional distribution of the respective growth rates.*

Finally, growth is assumed in GIBRAT's framework to follow a random walk process leading to log-normal growth rates distributions. In this paper, the main emphasis is on the characteristics of these distributions in case of regional industry-specific employment growth. Theoretical expectations will be deduced in the next section.

2.2.2 Firm-like behaviour of regional economies

A significant departure from normality suggests that some other mechanisms than the central limit theorem are at work (CANNING et al. 1998: 340). Confronted with the empirically observed firm growth rate distributions, BOTTAZZI & SECCHI (2006a) developed a stochastic model of firm growth. In contrast to Gibrat, they assume that economic opportunities, which sum up to a firm's growth rate in a given time period, are limited in their amount available to the firm and are not independent: a kind of increasing returns mechanism induces a cumulative and self-reinforcing process in the assignment of economic opportunities.[6] Successful firms which have realized more of those opportunities in the past exhibit a higher probability of taking up new ones. Due to the limitedness of economic opportunities, competing firms directly affect each other – if one firm's market share grows the market shares of its competitors shrink. Tent-shaped distributions with fat tails can emerge due to the increased probability that a large fraction of growth opportunities is assigned to a few firms only, whereas most other firms do not receive any at all.

Employment in a region is the sum of firms' employees located within this region. Hence, regions compete for limited economic growth opportunities at the level of firms. If regions, which have received such firm-level growth opportunities, are exposed to higher probabilities to receive even further opportunities, the stochastic model of firm growth would apply *mutatis mutandi* to regional economies and the distribution of their growth rates should consequently depart from normality. However, CASTALDI and DOSI (2009: 489) argue that inter-regional competition differs in many aspects from inter-firm competition. On the one hand, it is reasonable to consider territorial competition as a competition for mobile production factors and firm locations, which may cause dynamic and turbulent processes (CHESHIRE/MALECKI 2004: 260) and are in line with the basic idea of this model. On the other hand, limited growth opportunities may be bypassed by collective efforts in the regional policy arena, for example by trying to achieve higher growth rates in the overall regional system through cooperation and coordination or by diversifying the regional production portfolio into new areas of activity. Due to the latter issue, it is important to focus on the regional growth dynamics of single industries because industry-specific markets are

[6] Economic opportunities as well as the resulting growth rates may be either positive or negative by nature.

the very locus of competition for economic opportunities. Moreover, the drivers of positive and negative feedback loops in the context of opportunity accumulation are also industry-specific (CASTALDI/DOSI 2009, CASTALDI/ SAPIO 2008). Referring to the language and concepts of the New Economic Geography (e.g. KRUGMAN 1991), localization economies, which are by definition industry-specific, might induce increasing returns and imply an inter-regional dependency of the growth rates. Already MARSHALL (1890) observes that firms might benefit from the localisation of new or existing firms in the same region and industry due to the trinity of mechanisms of labour pooling, economies of scales in intermediate inputs and knowledge spillovers. The exit of a firm from the region or market might decrease the chance of survival for other co-located firms of that industry analogously. Finally and in accordance with endogenous growth theories (e.g. ROMER 1990), the diffusion of growth opportunities stemming from innovation throughout the economy has both a geographical and technological dimension (see DÖRING/SCHNELLENBACH 2006 for an overview on the issue of knowledge spillovers): the resulting feedback loops are bounded to some degree to work within regions and the capacity to absorb knowledge spillovers depends amongst others on the regions' industrial structure, making the disaggregated level of industries an adequate observational unit for the analysis of regional growth processes.

Briefly stated, we argue that regions compete over a finite set of industry-specific growth opportunities and that the regions' firms interact and learn at the industrial level, leading to self-reinforcing dynamics in opportunity catching. This reasoning leads us to:

> **Hypothesis 2a**: *Self-reinforcing mechanisms still prevail at the regional level. Consequently, fat tails as signs of a higher probability of extreme positive and negative growth events emerge as a key feature of regional industry-specific employment growth.*

However, it could be argued that the influence of the mechanisms responsible for fat tails may fade if longer than one year time lags are considered. Growth events become more independent over time and a progressive normalization of the growth rates' distribution should take place as the central limit theorem comes into force (BOTTAZZI/SECCHI 2006a). Hence, we should expect:

> **Hypothesis 2b**: *The longer the time lag for which growth rates are calculated, the closer their distribution is to the normal one.*

2.2.3 Relationship between type of industry and stochastic characteristics of regional growth

Two arguments make a finer break-down of the regional economy into single industries necessary. First, stochastic properties at a higher level of aggregation may be a sheer outcome of the aggregation process (BOTTAZZI/SECCHI 2003: 218, BROCK 1999: 431). As argued above, true economic mechanisms of regional economic growth operate at the level of specific industries. Hence, it has to be tested whether the findings survive at different levels of aggregation. Second, the stochastic properties may also differ across heterogeneous industries. A systematic comparison of different types of industries could

provide new insights for the explanation of the characteristics and mechanisms of regional growth dynamics.

Three lines of distinction seem to be relevant. First, industries differ in their knowledge intensity in production (SMITH 2002). Using data on German labour market regions, SCHLUMP and BRENNER (2010) prove that the effects of universities and public research and development on regional employment growth differ significantly across industries. FORNAHL and BRENNER (2009) point out that especially industries which rely on a scientific knowledge base, show a higher geographical concentration of innovation activities. As innovations are strongly connected to growth and their diffusion is bounded in space, it is expected that these industries are more prone to experience extreme regional growth events. Secondly, patterns of innovation per se vary across industries. PAVITT (1984) demonstrates the differences in underlying production technologies, competition processes and learning modes. The extent of opportunities for innovations and thus growth obviously depends on the particular technological regime of an industry (DOSI 1988). Finally, we are the first in literature to explicitly include service industries in the study of the stochastic characteristics of growth. Thus, it is natural to contrast the growth dynamics of service and manufacturing industries. This results in the following hypothesis:

> **Hypothesis 3**: *The stochastic characteristics of regional employment growth depend on the type of industry under consideration.*

2.3 Method of fitting the best theoretical distribution

In search of a more general and flexible distributional model that describes the empirical distribution of growth rates *g*, the exponential power (EP) distribution family was introduced into economics by BOTTAZZI et al. (2002):

$$f(g;b,a,m) = \frac{1}{2ab^{\frac{1}{b}}\Gamma(1+\frac{1}{b})} exp\left(-\frac{1}{b}\left|\frac{g-m}{a}\right|^{b}\right) \tag{6}$$

with Γ*(.)* standing for the gamma function. Three parameters define the distribution: the location parameter m, which indicates the existence of a general trend in the data, the scale parameter *a*, which determines the spread or dispersion of the distribution, and the shape parameter *b*. Both the normal (*b* = 2) and Laplace (*b* = 1) distribution are particular cases of the EP family of probability densities. This family allows for a continuous variation from non-normality to normality, with a smaller shape parameter *b* representing fatter tails of the corresponding density. Furthermore, it can be extended to a five-parameter family of distributions, which is able to cope with asymmetries in the data. In addition to *m*, the asymmetric exponential power (AEP) distribution possesses two scale parameters a_l and aright for the values below and above m and two shape parameters b_l and b_r describing the tail behaviour on the left and right side of the distribution:

$$f(g;\ b_l, b_r, a_l, a_r, m) = \tag{7}$$

$$= \frac{1}{C} exp\left(-\left[\frac{1}{b_l}\left|\frac{g-m}{a_l}\right|^{b_l}\theta(m-g) + \frac{1}{b_r}\left|\frac{g-m}{a_r}\right|^{b_r}\theta(g-m)\right]\right)$$

where $\theta(g)$ is the Heaviside theta function and $C = a_l {b_l}^{1/b_l - 1}\Gamma(1/b_l) + a_r {b_r}^{1/b_r - 1}\Gamma(1/b_r)$ a normalization constant (BOTTAZZI/SECCHI 2011).

The aim of the study conducted here is to find out which theoretical distribution fits best to reality. To anticipate, several parametric tests robustly reject the normal distribution. This confirms the findings in literature for our case and consequently, we restrict the analysis to the Laplace, EP and AEP. We apply a similar procedure as described in BRENNER (2006). First, a maximum likelihood estimation of the respective theoretical distributions is performed using the regional employment growth rates. This is done for each single industry separately. The Laplace distribution contains two parameters, the EP three and the AEP five. Furthermore, the Laplace is a special case of the EP, and the EP can be extended to the AEP. Therefore, a likelihood ratio test is appropriate to check whether the more general theoretical distribution with more free parameters, which describes reality always better than the more restricted one, so significantly. This test is applied to each pair of theoretical distributions, such that the best distributional choice, which counterbalances goodness of fit versus loss in degrees of freedom, can be reported.

This test procedure answers the question of which theoretical distribution describes the empirical data better. However, it does not answer the question of whether the respective distribution describes the empirical data adequately. To test whether the empirical distribution and the AEP, the most general theoretical distribution of the analysed distributional portfolio and consequently the distribution with the highest fit, deviate from each other significantly, the Kolmogorov-Smirnov (KS) and Anderson-Darling (AD) goodness-of-fit statistics are used. Although the former is the most popular test, the latter has the advantage of giving more weight to the tails of the distributions (CIRILLO/HÜSKER 2009) and of possessing more power compared to the former (STEPHENS 1974). This implies a higher rejection rate of the null hypothesis of no statistical difference between the fitted theoretical and empirical distribution function.[7]

2.4 Empirical data

The data used in this approach was collected by the German Federal Institute of Labour (IAB) on the 30th of June each year. The data set contains a time panel of the entire population of employees and firm establishments for each industry and each region in Germany from 1999 to 2007.[8] Industries are denoted by *i*, regions by *r*. To check the robustness of the results regarding the choice of level of industrial aggregation and the

[7] Critical values at 5% significance level of the test statistics are approximated by Monte Carlo simulations with 5000 replications, because parameters of the hypothesized AEP are not known ex ante but are estimates themselves (CAPASSO et al. 2009).

[8] Industries are classified according to the WZ-03 classification, which was the standard classification of industries in Germany for the analysed time period.

definition of regions, we perform the analysis separately for each 4-digit, 3-digit and 2-digit industry as well as for three different definitions of regions: labour market regions as defined by the IAB (BINDER/SCHWENGLER 2006), labour market regions as defined by ECKEY et al. (2006) and administrative districts. For the latter, results are expected to be more blurred, because delimitations follow political and historical reasons. Not taking inter-regional economic dependencies fully into account might influence the stochastic properties of regional growth rates. In contrast, the two functional definitions are less affected by construction. Although both adjacent districts amalgamate on the basis of commuting flows as a proxy for economic interactions, ECKEY's definition uses more sophisticated statistical tools and ends up with fewer regions. However, the definition by IAB remains popular within policy analysis. Altogether, nine combinations of industries and regions result as levels of investigation (see Figure 11).

Definition of regions	Level of industrial aggregation					
	LMR-IAB	**2-digit**	**LMR-IAB**	**3-digit**	**LMR-IAB**	**4-digit**
	270 regions	59 industries	270 regions	212 industries	270 regions	459 industries
	LMR-Eckey	**2-digit**	**LMR-Eckey**	**3-digit**	**LMR-Eckey**	**4-digit**
	150 regions	59 industries	150 regions	212 industries	150 regions	459 industries
	Districts	**2-digit**	**Districts**	**3-digit**	**Districts**	**4-digit**
	413 regions	59 industries	413 regions	212 industries	413 regions	459 industries

Figure 11: Levels of investigation

As an effect of integers, the growth of small quantities can manifest itself only in a limited number of different growth rates. Regions in which employees of certain industries are few in numbers potentially distort the well-behaved growth rates distributions. Truncating the data at a minimum of ten industry-specific employees working in a region, we calculate industry-specific regional employment growth rates $g_{i,r,t}$ by taking the differences of the natural logarithms of the regional employment stocks E in an industry between two successive years t:[9]

$$g_{r,i,t} = log(E_{r,i,t+1}) - log(E_{r,i,t}) \tag{8}$$

[9] This specific size threshold for the truncation is chosen so that the overall shape of the growth rates distributions is not affected, but their visible noise strongly reduced. In order to ensure statistical reliability, industries with less than a total of 100 observations summed over eight years are removed entirely from the sample.

2.5 Empirical results

2.5.1 Testing for Gibrat's law and the scaling relationship

A simple regression framework, which is based upon GIBRAT's growth equation that is extended by an autocorrelation term, allows to test for the basic independence assumptions of regional growth processes:

$$g_{r,i,t} = b_0 + b_1\, log\left(E_{r,i,t}\right) + b_2\, g_{r,i,t-1} + \varepsilon_{r,i,t} \tag{9}$$

If b_1 = 0, the region's growth rate is independent from its (log) number of employees and if b_2 = 0, it is also independent from its past year growth rate. We estimate equation (9) separately for each industry and each consecutive year pair using least absolute deviations (LAD), as this technique delivers more robust and less biased estimators compared to OLS in case of fat-tailed error terms $\varepsilon_{r,i,t}$, which we anticipate to hold true in the present case.

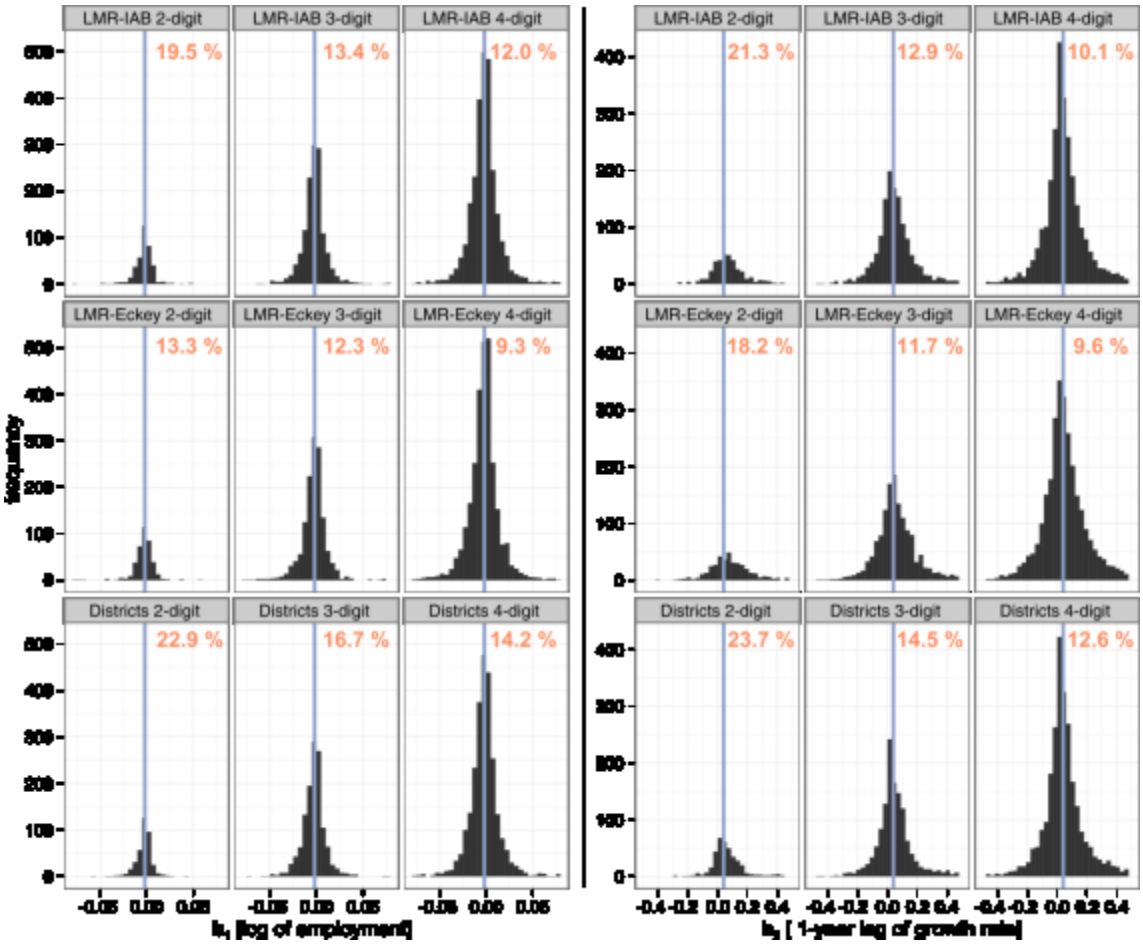


Figure 12: Frequency distributions of the estimated parameters b_1 and b_2 of the extended Gibrat growth equation, which is estimated separately for each industry and consecutive year pair. On the upper right corner of each histogram the percentages of significant estimates (5% significance level) are reported

As can be seen in Figure 12, the estimates of b_1 concentrate around zero and have a slightly negative median value (vertical lines). This implies that regions with fewer employees tend to grow faster. The estimates of b_1 indicate the presence of a positive temporal autocorrelation. However, both relationships are not very strong and only significant in a small fraction of the cases. Hence, the weak form of GIBRAT's law, as expressed in *hypothesis 1a*, cannot be rejected in the majority of cases. So we decided to pool together growth rates from different years and from regions with different numbers of employees.

A further important issue regarding the possibility of pooling together data of different regions is the dependence of variance of the growth rates on the regions' number of employees – this relationship was expected to be negative in *hypothesis 1b*. One empirical approach to assess variance scaling, as made popular by STANLEY et al. (1996), is to allocate the entities, ordered by a size measure, into equipopulated bins. Within each bin, the average size, here the numbers of regional employees, as well as the standard deviations of the associated growth rates are calculated. Plotted on a log-log scale, a linear relationship indicates that the variance of *g* scales as a power law of the number of employees. Bin scaling has the advantage that the relationship can be displayed visually, as shown in the left panel of Figure 13 for two exemplary industries using 20 bins. Likewise to firms and countries, a power-law scaling can be observed. However, enough observations must be available to get reliable estimates for the variance conditional on size. An alternative approach, suggested by BOTTAZZI et al. (2014), is to directly model the scaling relationship by introducing a heteroskedasticity term into the stochastic process of the growth rates:

$$g_{r,i,t} = \alpha_i + u_{r,i,t} \tag{10}$$

where α_i is a constant term that converges to the average growth rate. The error term can be written as $u_{r,i,t} = \exp(\beta_i(e_{r,i,t} - \bar{e}_i))\, \varepsilon_{r,i,t}$, with $e_{r,i,t} = \log(E_{r,i,t})$ and $\bar{e}_i$ is the corresponding industry-specific arithmetic mean over regions and time. This expression takes into account that the functional form of heteroskedasticity might be non-linear, as recently observed in the firm growth literature (e.g. BOTTAZZI et al. 2011) and here already visible for the two industries of the bin-scaling example. Replacing the heteroskedastic error term by equation (10) and solving for $\varepsilon_{r,i,t}$, we get:

$$\varepsilon_{r,i,t} = \frac{g_{r,i,t} - \alpha_i}{exp(\beta_i(e_{r,i,t} - \bar{e}_i))} \tag{11}$$

The last equation yields the rescaled growth rates, which are cleaned from heteroskedasticity and the average growth trend. Again anticipating the shape of the

growth rates distributions, we assume that the error term is not normally distributed and stay in the LAD regression framework to estimate the two unknown parameters:[10]

$$\{\beta_i; \alpha_i\} = argmin_{\beta,\alpha} \sum_{r,t} \left| \frac{g_{r,i,t} - \alpha_i}{e^{\beta_i(empl_{r,i,t} - \overline{empl_i})}} \right| \tag{12}$$

The frequency distributions of the estimated industry-specific scaling exponents β_i are illustrated on the right panel of Figure 13. Virtually all estimates of β_i vary between the two limiting cases $\beta_i = 0$ and $\beta_i = -0.5$. In the majority of cases, β_i concentrates around the value of -0.2, hence confirming the findings for firms and countries and indicating the presence of some considerable positive interactions between the regions' firms, and in addition to or in lieu of this, a diversification effect.

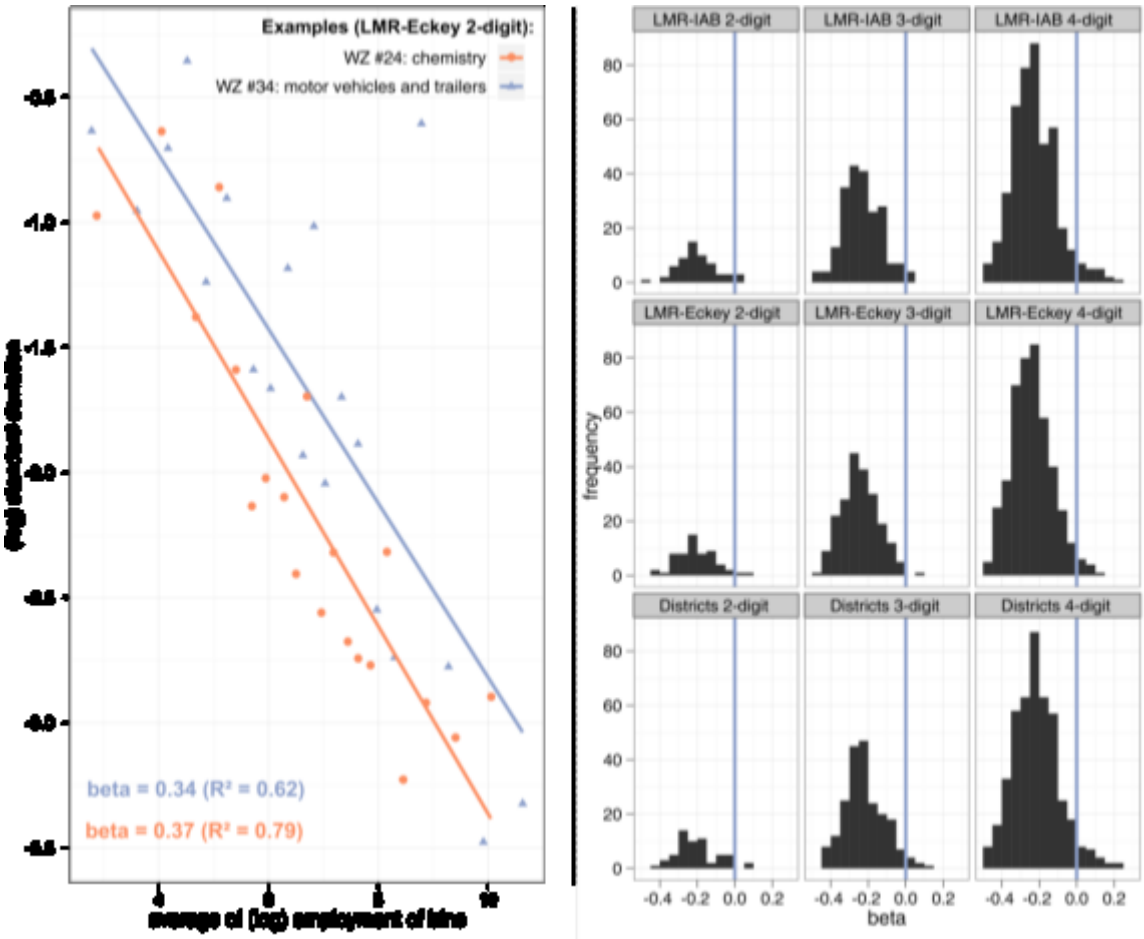


Figure 13: Scaling relationship for two exemplary industries within a bin scaling approach (left panel) and frequency distributions of the estimated industry-specific scaling exponents β_i within a LAD scaling approach (right panel)

[10] Minimizing the absolute deviations is equivalent to the log-likelihood function in the case of a Laplace distribution.

$\tilde{g}_{r,i,t} := \varepsilon_{r,i,t}$ can only be interpreted as different realizations of the same underlying stochastic process after rescaling. In the remainder of the empirical part, the focus lies on the distributional characteristics of the industry-specific growth rates.

2.5.2 Fitting of the theoretical distribution to regional industry-specific growth rates

As a prefacing visual exercise, we pool together the rescaled $\tilde{g}_{r,i,t}$ over all years and industries and discretize them by constructing small bins. Plotting the frequency distribution on a log-log scale, a tent-like shape emerges that is well described by the Laplace distribution at the centre (see Figure 14). However, the tails on both sides are much fatter in comparison to the normal and the Laplace distribution. As predicted in *hypothesis 2a*, growth events at the extremes are more likely to occur than they could have been expected according to the normal or Laplace distribution. Furthermore, it is already visible that the left tails are even more pronounced.

A normal distribution of the growth rates as the underlying theoretical distribution is rejected by several standard parametric tests for virtually all industry-specific growth rates, independently of the regional definition and industrial aggregation level.[11] Therefore, we restrict the analysis to the Laplace, EP and AEP distribution. The method unfolded in section 2.3 allows for each industry to separately identify the best distribution out of the potential candidates. The left part of Table 3 reports the (relative) numbers of occurrences of the best fitting distribution identified by the likelihood ratio test procedure.

Table 3: Identification of the best fitting theoretical distribution for $\tilde{g}_{r,i,t}$

	Laplace		EP		AEP		KS-Test H0 rejected	AD-Test H0 rejected	N[1]
LMR-IAB 2-digit	1	(2%)	2	(3%)	54	(95%)	44%	61%	57
LMR-IAB 3-digit	7	(3%)	24	(12%)	170	(85%)	46%	50%	201
LMR-IAB 4-digit	8	(2%)	40	(10%)	351	(88%)	45%	45%	399
LMR-Eckey 2-digit	5	(9%)	6	(10%)	47	(81%)	7%	18%	58
LMR-Eckey 3-digit	4	(2%)	41	(20%)	162	(78%)	14%	19%	207
LMR-Eckey 4-digit	13	(3%)	81	(19%)	328	(78%)	15%	16%	422
Districts 2-digit	0	(0%)	4	(7%)	52	(93%)	68%	79%	56
Districts 3-digit	1	(1%)	15	(8%)	176	(91%)	69%	70%	192
Districts 4-digit	1	(0%)	42	(11%)	334	(89%)	63%	59%	377

Note: [1] Number of industries analyzed. Industries with less than 100 observations are excluded from the sample in order to ensure statistical reliability. Hence, N varies between different definitions of regions: the more detailed the regional delimitations are, the more missing values across the regions naturally occur.

[11] In addition to the KS- and AD-tests, we employ the Jarque-Bera test for a deviation from normality in terms of kurtosis or skewness or both (JARQUE & BERA 1980), and the Anscombe-Glynn test for a deviation from normality in terms of an excess kurtosis (ANSCOMBE & GLYNN 1983). Results are reported in the Appendix X.1.

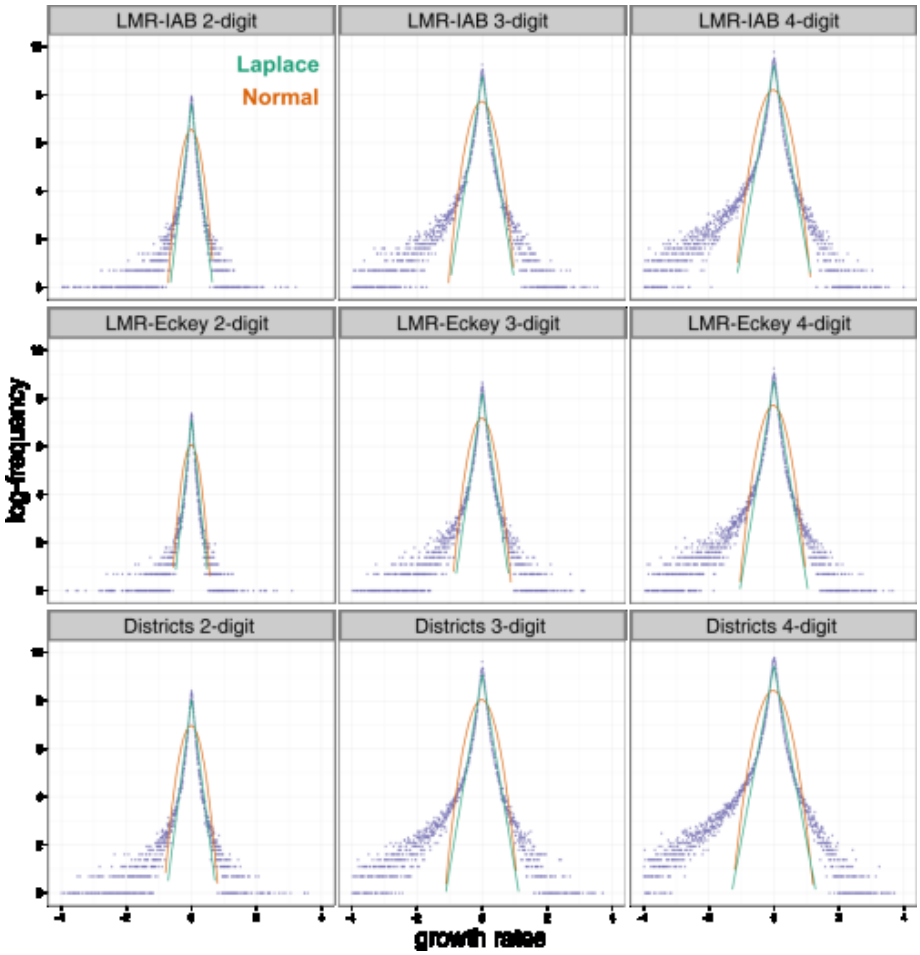


Figure 14: Frequency distribution of rescaled $\tilde{g}_{r,i,t}$ pooled over all years and industries[12]

The AEP is the best choice out of the theoretical distributions for describing regional industry-specific growth in more than four out of five cases. Its symmetric counterpart is sufficient at best in one fifth of the cases and the Laplace is nearly always outperformed by the more general models. In order to assess the adequacy of the AEP to describe the empirical growth rates distributions, the relative rejection frequencies of the KS- and AD-tests at 5% significance level are reported. Since the number of observations (i.e., number of spatial units with employees multiplied by eight years) is relatively high, small deviations already suffice to reject H0. Acknowledging this, a considerable improvement of the fit to the empirical data compared with the normal distribution, which is rejected consistently, can be achieved. This holds especially true in the case of labour market regions as defined by ECKEY et al. (2006), for which the AEP is never rejected in more than one fifth of the industries, whereas both other regional definitions show higher rejection rates. The difference between the three definitions of regions can be driven, on the one hand, by a higher number of observations for districts (413 spatial units) and IAB labour market regions (270) in comparison to ECKEY (150), which increases the statistical power and

[12] The plots are truncated on the x-axis at -4 and 4 for visual reasons. The plotted interval still includes more than 99.9% of all observations.

consequently the probability of a null rejection. On the other hand, however, ECKEY's definition might be the most adequate one in terms of describing intra-regional feedback loop mechanisms which lead to fat tails. Although confronted with less distinctive results, the same argumentation could serve to explain the lower rejection rates at a higher industrial resolution: the AEP is rejected less often by the AD-test at the 4-digit level, whereas no differences are found concerning the KS-test.

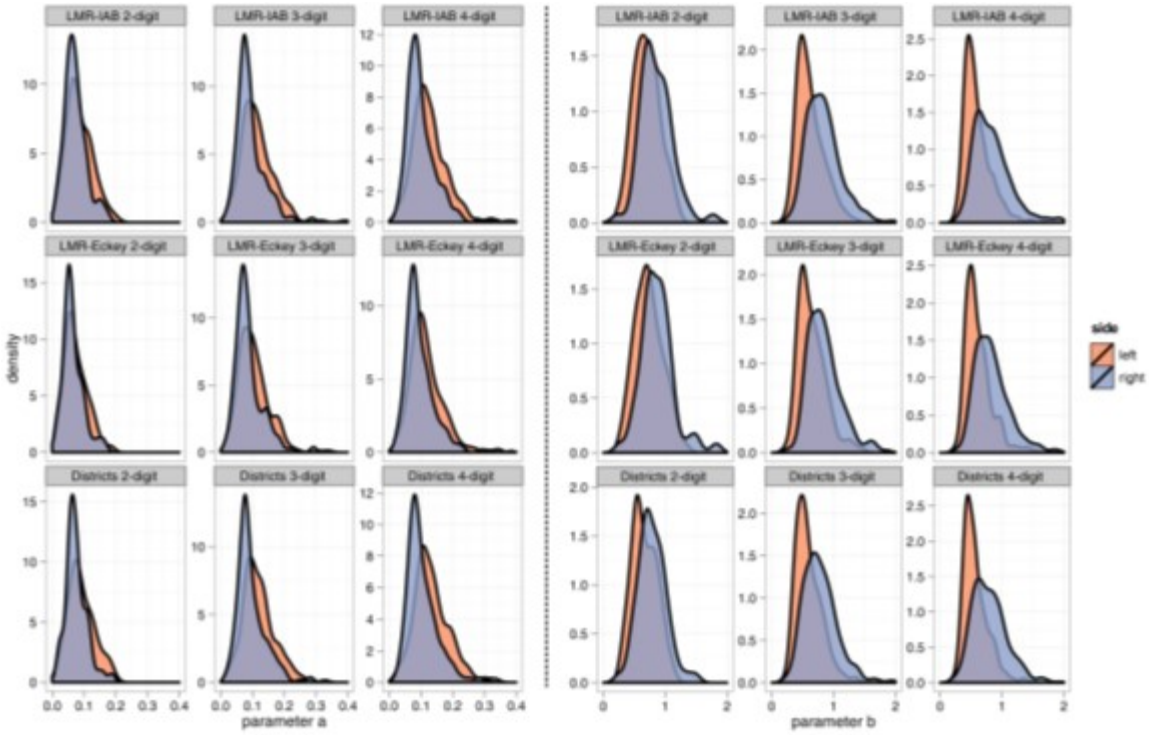


Figure 15: Density distributions of the estimated parameters a_l, a_r, b_l and b_r of the AEP.

A closer look at the estimated parameters for the scale as well as the shape of the AEP reveals further insights into the stochastic properties of regional industry-specific growth processes. The left panel of Figure 15 contrasts with the respective right and left side scale parameters a_l and a_r, which are estimated separately for all industries. Growth rates that are below the general trend m tend to have a slightly higher dispersion than growth rates above m. A test for stochastic dominance, as introduced by FLIGNER and POLICELLO (1981), confirms a significant difference regarding eight out of nine combinations of industrial levels and regional definitions. Only in the case of LMR-Eckey (2-digit) is it not found that a randomly drawn industry-specific scale parameter a_l has a probability significantly higher than 50% to be larger than a randomly drawn a_r. Differences between both sides of the distribution are even more pronounced in the case of shape parameters (right panel of Figure 15): b_l is always significantly dominated by b_r, meaning that regions are more prone to experience extreme negative industry growth events than extreme positive events. But both the tails on the left and the right side are fatter than the respective

tails of the normal distribution ($b_l = b_r = 2$) and, in most industries, even than the tails of the Laplace distribution ($b_l = b_r = 1$).

To test *hypothesis* 2b, whether increasing the time lag leads to a progressive normalization, we calculate all possible growth rates for a one year up to an eight year time lag, always starting with the number of employees in 1999. Figure 16 shows that the shape parameter b for both sides increases with time, but still remains far from two, the value expected for a normal distribution. Rather the median values of b_l and b_r approach the Laplace distribution. In summary, extreme growth events are a prominent feature of regional industry-specific growth, independently of time scales, industrial aggregation levels or regional definitions.

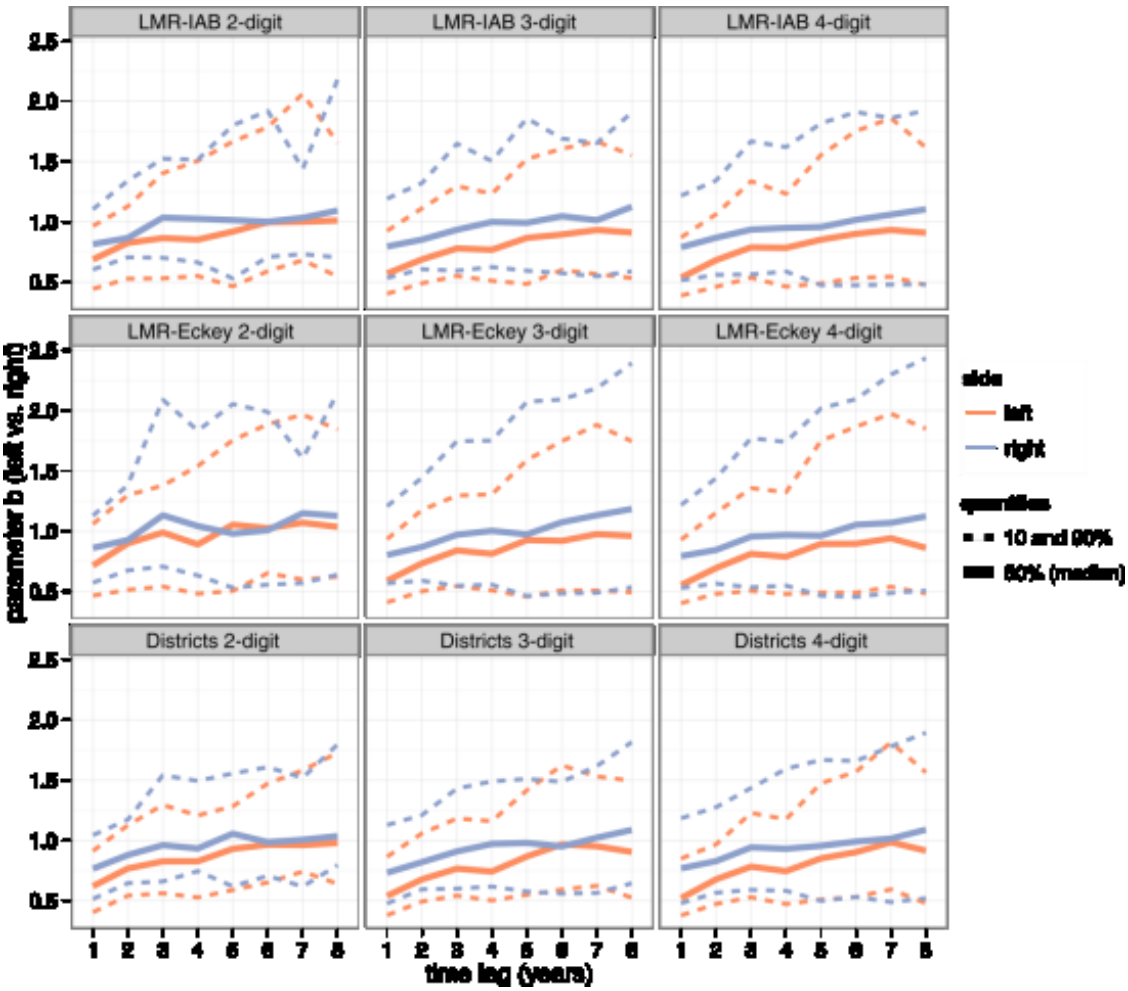


Figure 16: Evolution of the shape parameters b_l and b_r with increased time lags for $g_{i,r,t}$

2.5.3 Relation between type of industry and stochastic characteristics of regional growth

To assess the relationship between the stochastic characteristics of regional employment growth and the type of industry, we group all industries according to the following classifications:

- *Economic sector*: the secondary or manufacturing sector (industries from WZ 15 to 45) can be distinguished from the tertiary or service sector (WZ 50 to 99). The primary sector is excluded because it does not contain enough observations.
- *Knowledge intensity*: all industries can be classified into knowledge intensive or non-knowledge intensive industries. The assignment here is based on LEGLER and FRIETSCH (2006) with 3-digit industries as the highest level of disaggregation.
- *Pavitt taxonomy*: four categories can be differentiated on the 2-digit level, namely science-based, specialised suppliers, scale intensive and supplier dominated industries. We use the revised Pavitt taxonomy by BOGLIACINO and PIANTA (2010), as service industries are addressed by it as well (up to WZ 74).

Using these classification schemes, we then follow a two-step procedure. First, we regress the estimated distributional parameters b_l, b_r, a_l and a_r of the AEP with the average (log) number of firms per region. It is a necessary step as this number is both correlated with the types of industries and the distributional parameters (see Figure 17 for the relationships between the average (log) number of firms per region and b_l, b_r).

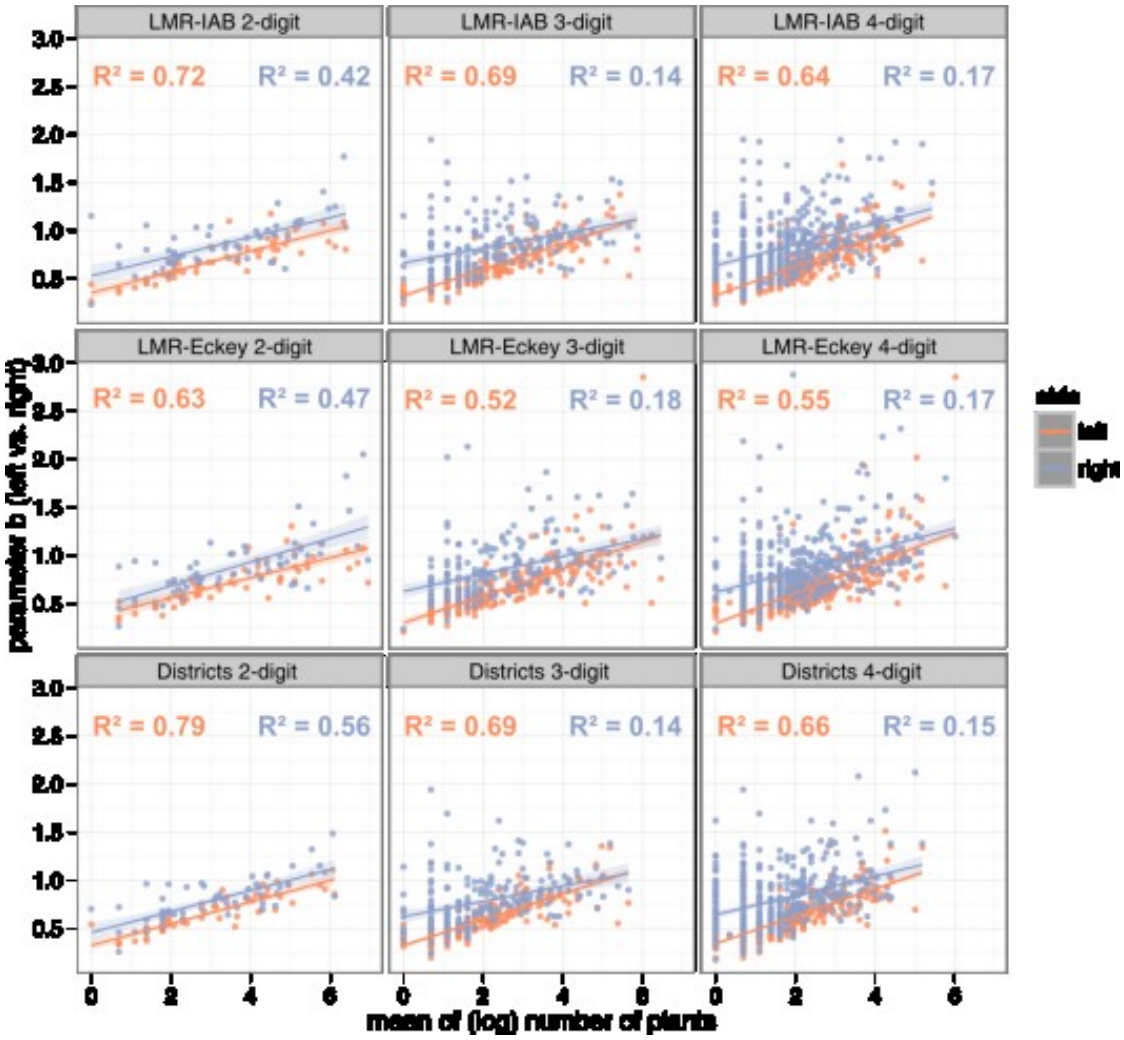


Figure 17: Relationship between the average (log) number of firms per region and the shape parameters b_l and b_r for all industries

The explained variance (R^2) from a simple OLS regression model is especially high concerning the left tails of the distributions. The revealed relationship between the tail behaviour and the industries' average number of firms within a region is an interesting finding by itself, which deserves more attention in future research.[13] Here, we are not

[13] This finding supports the idea that observed patterns at higher levels of aggregation can be explained to a considerable degree by the micro-level of firms (e.g. GABAIX 2011; ACEMOGLU et al. 2012). If the region has only a few firms that change their size each year, the characteristics of regional growth rates directly originate from the firm growth rates, because regional growth is essentially the sum of aggregation of all firm growth rates located in a region and the latter are well known to be fat tailed distributed. However, it can be observed that even b_l and b_r stay visibly below the value of a normal distribution regarding most industries with more than a hundred or hundreds of firms per region. Further investigation is needed to systematically disentangle the two effects causing the persistence of fat tails: they may emerge due to correlating mechanisms working purely at the level of regions, or firm level growth rates aggregate non-trivially: instead of compensating out in the aggregation process, they may amplify and produce extreme regional growth events (FU et al. 2005, CASTALDI/SAPIO 2008: 525).

interested in the standard results of the regressions but in the residuals for each industry. The residuals provide information about whether each industry has higher or lower parameter values b_l, b_r, a_l and a_r than the average. In a second step, we test whether the groups of the various industrial classification schemes show significant deviations in the parameters from the whole economy's average. To this end, we perform the Kruskal-Wallis rank sum test of the null hypothesis that the median values are equal in each group (KRUSKAL/WALLIS 1952). This test can be considered as a non-parametric alternative to the single-factor analysis of variance (BORTZ et al. 2008: 222) and is suitable here, because the residuals are in most cases significantly rejected as normally distributed.[14] Conditional median values of the residuals and significant differences are reported in Table 4 for the manufacturing and service industries, in Table 5 for the knowledge intensity classification scheme, and in Table 6 for the revised Pavitt taxonomy.

In section 2.5.2 we show that even though the AEP best describes the empirical growth rate distributions, this distributional model can still be rejected by goodness-of-fit statistics for certain industries. First, it has to be evaluated whether the rejection rate of the KS- and the AD-test varies across the industrial classification schemes. The respective χ^2-statistics never signal any systematic differences. Secondly, as an additional robustness check, we repeat the analysis above considering only those industries for which the AEP was neither significantly rejected by the KS- nor the AD-test. A reduced sample size inevitably implies the loss of some former weakly significant results. Beyond that, a noteworthy deviation cannot be found.[15] Therefore, we consider the subsequent interpretation of the distributional parameters of the full sample of industries as meaningful.

Table 4: Conditional median values of the residuals and significant differences (Kruskal-Wallis test)

	Economic sector	~ a_l	~ a_r	~ b_l	~ b_r
LMR-IAB 2-digit	manufacturing	-0.012	-0.009	-0.013	-0.047
	service	0.001	-0.001	-0.004	0.010
LMR-IAB 3-digit	manufacturing	-0.018 **	-0.021 ***	-0.018	-0.113 **
	service	-0.001	0.001	-0.014	-0.017
LMR-IAB 4-digit	manufacturing	-0.020 ***	-0.020 ***	-0.009	-0.113
	service	0.001	0.000	-0.027	-0.048
LMR-Eckey 2-digit	manufacturing	-0.013	-0.011	-0.012	-0.054
	service	-0.001	-0.005	-0.015	-0.051
LMR-Eckey 3-digit	manufacturing	-0.017 **	-0.019 ***	-0.012	-0.085
	service	-0.001	-0.004	-0.033	-0.056
LMR-Eckey 4-digit	manufacturing	-0.017 ***	-0.018 ***	-0.006 *	-0.086
	service	-0.002	-0.003	-0.042	-0.052
Districts 2-digit	manufacturing	-0.013	-0.010	-0.003	-0.044
	service	0.004	-0.001	0.021	-0.015
Districts 3-digit	manufacturing	-0.019 ***	-0.025 ***	-0.012	-0.107 *
	service	0.003	0.004	-0.007	-0.028
Districts 4-digit	manufacturing	-0.023 ***	-0.022 ***	-0.005	-0.105
	service	0.004	0.004	-0.021	-0.057

Note: *** p<0.001, ** p<0.01 and * p<0.05

[14] Results of the KS-test against normality at 5%-significance level are available upon request.
[15] Again, detailed results of the robustness analysis are available upon request.

Table 5: Conditional median values of the residuals and significant differences (Kruskal-Wallis test)

	Knowledge intensity	~ a_l	~ a_r	~ b_l	~ b_r
LMR-IAB 2-digit	non knowledge intensive	-0.001	-0.007	0.011	-0.022
	knowledge intensive	-0.001	-0.003	-0.031	-0.032
LMR-IAB 3-digit	not knowledge intensive	-0.008	-0.013	-0.004 **	-0.040 *
	knowledge intensive	-0.015	-0.012	-0.043	-0.122
LMR-IAB 4-digit	not knowledge intensive	-0.008	-0.012	-0.003 ***	-0.049 ***
	knowledge intensive	-0.023	-0.013	-0.050	-0.151
LMR-Eckey 2-digit	not knowledge intensive	-0.002	-0.008	0.012 **	-0.027
	knowledge intensive	-0.004	-0.008	-0.071	-0.078
LMR-Eckey 3-digit	not knowledge intensive	-0.007	-0.011	0.001 *	-0.029 *
	knowledge intensive	-0.016	-0.012	-0.050	-0.100
LMR-Eckey 4-digit	not knowledge intensive	-0.009	-0.011	-0.007 ***	-0.039 ***
	knowledge intensive	-0.021	-0.015	-0.053	-0.138
Districts 2-digit	not knowledge intensive	-0.001	-0.008	0.016	0.006
	knowledge intensive	-0.001	-0.003	-0.022	-0.049
Districts 3-digit	not knowledge intensive	-0.008	-0.018	0.002 **	-0.042 **
	knowledge intensive	-0.015	-0.010	-0.032	-0.134
Districts 4-digit	not knowledge intensive	-0.009	-0.013	-0.005 **	-0.053 **
	knowledge intensive	-0.020	-0.013	-0.039	-0.148

Note: *** p<0.001, ** p<0.01 and * p<0.05

Table 6: Conditional median values of the residuals and significant differences (Kruskal-Wallis test)

	Revised Pavitt taxonomy	~ a_l	~ a_r	~ b_l	~ b_r
LMR-IAB 2-digit	supplier dominated	-0.002 **	-0.006 *	-0.010	0.040
	scale intensive	-0.027	-0.026	0.006	-0.057
	specialized suppliers	0.008	-0.008	-0.086	-0.169
	science based	0.031	0.027	-0.045	-0.018
LMR-IAB 3-digit	supplier dominated	-0.001 **	-0.013 **	-0.006	-0.036
	scale intensive	-0.031	-0.027	-0.017	-0.085
	specialized suppliers	-0.014	-0.011	-0.027	-0.127
	science based	-0.003	0.002	-0.043	-0.118
LMR-IAB4-digit	supplier dominated	-0.003 **	-0.012 *	-0.018	-0.034 **
	scale intensive	-0.028	-0.019	-0.002	-0.126
	specialized suppliers	-0.010	-0.007	-0.038	-0.131
	science based	-0.007	-0.011	-0.031	-0.116
LMR-Eckey 2-digit	supplier dominated	0.004 **	-0.007 **	0.071 *	-0.012
	scale intensive	-0.023	-0.018	-0.027	-0.077
	specialized suppliers	0.009	0.001	-0.050	-0.106
	science based	0.034	0.033	-0.084	0.047
LMR-Eckey 3-digit	supplier dominated	-0.003 **	-0.013 ***	-0.008	-0.017
	scale intensive	-0.029	-0.028	-0.009	-0.122
	specialized suppliers	-0.007	-0.011	-0.012	-0.109
	science based	0.008	0.011	-0.063	-0.084
LMR-Eckey 4-digit	supplier dominated	-0.004 ***	-0.011 **	-0.014 *	-0.028 **
	scale intensive	-0.028	-0.018	0.007	-0.141
	specialized suppliers	-0.009	-0.008	-0.033	-0.115
	science based	0.002	0.008	-0.039	-0.102
Districts 2-digit	supplier dominated	0.004 **	-0.005 *	0.014	0.051
	scale intensive	-0.024	-0.019	0.013	-0.056
	specialized suppliers	0.005	-0.009	-0.028	-0.082
	science based	0.033	0.023	-0.035	-0.031

Districts 3-digit	supplier dominated	-0.006 **	-0.017 **	-0.004	-0.032 *
	scale intensive	-0.030	-0.028	-0.014	-0.132
	specialized suppliers	-0.013	-0.014	-0.017	-0.110
	science based	0.033	0.010	-0.039	-0.098
Districts 4-digit	supplier dominated	-0.004 *	-0.012 *	-0.006	-0.036 **
	scale intensive	-0.032	-0.024	-0.001	-0.085
	specialized suppliers	-0.014	-0.005	-0.028	-0.125
	science based	-0.004	-0.008	-0.054	-0.155

Note: *** p<0.001, ** p<0.01 and * p<0.05

Hypothesis 3 can be confirmed partially – some stochastic characteristics of regional employment growth depend on the type of industry. The general dispersion of the growth rates, as measured by the parameters a_l and a_r, is significantly higher for service industries compared to manufacturing. In contrast, the knowledge intensity of an industry neither bears consequence on the above nor below the average growth rates dispersion. Grouped according to Pavitt's taxonomy, a higher heterogeneity in the growth rate fluctuation patterns is observed: science based industries show the highest dispersion parameters and scale intensive industries the lowest dispersion parameters. Again, this applies to growth as well as decline. In short, we find that foremost science based and service industries show higher fluctuations in their regional growth dynamics.

Further interesting characteristics of the growth process reside in the probability of a regional economy to be affected by extreme growth events. Manufacturing and service industries general show no significant differences in their tail behaviour, although extreme positive growth events are slightly more frequent in manufacturing. Results for Pavitt's taxonomy are mixed. Extreme regional employment decline tends to be more frequent in science based industries, followed by specialized suppliers, and less frequent in supplier dominated and scale intensive industries. Differences concerning positive growth show a higher significance: here fat tails are more pronounced in specialized suppliers and scale intensive industries and less pronounced in supplier dominated industries. Hence, especially supplier dominated industries consistently indicate a tendency to be less affected by extreme events, irrespective of growth or decline. Finally, there is strong and robust evidence that knowledge-intensive industries have fatter tails than non-knowledge intensive industries. In other words, industries in which knowledge and innovation assume an important role in the production process are more exposed to experience extreme negative as well as extreme positive growth events.

Relating these findings to the mechanisms that potentially lead to fat tails and are discussed in section 2.2.2, we draw the following preliminary conclusions. The strong relationship between the knowledge intensity and the shape parameters b_l and b_r indicates that spatially bounded knowledge spillovers, which are naturally more relevant for knowledge intensive industries, are a candidate mechanism leading to the emergence of fat tails. Pavitt's taxonomy reveals the idea that besides knowledge spillovers, competition effects might also play a role. This becomes apparent if we agree upon the fact that supplier dominated industries are less exposed to fierce inter-regional competition as compared to specialized suppliers or science based industries. At the same time, the higher relevance of intra-regional localization economies for manufacturing industries, as it can be found in the agglomeration literature (e.g. GRAHAM 2009), might explain the

reason why these industries have fatter positive tails compared to service industries. In the concluding section, we will present some further variables that might allow future research to analyse the mechanisms of the stochastic growth model of BOTTAZZI and SECCHI (2006a) at the levels of regions more directly.

2.6 Conclusions

In this paper we find that the AEP is the theoretical distribution that best describes the industry-specific employment growth of regional economies. The estimated distributional parameters show that the growth rate distribution is tent-like shaped with tails significantly fatter than the normal and Laplace distribution. Despite the aggregate nature of regions, extreme growth events occur with a higher probability than could be expected under a *normal* data-generating process. A comparison of both tails of the distribution reveals that particularly extreme negative events are a prominent feature of regional employment growth. This "rich statistical structure in dynamics" (DOSI et al. 2010: 1872) of economic entities like regions has important implications for both theory and modelling.

Economic theory should go beyond simply focusing on and explaining the average dynamics of regional growth. In an evolutionary context, persistent interactions between firms as the actual economic agents may lead to observable non-normal behaviour at higher levels of the economy (DOSI et al. 2010). Complexity thinking (e.g. ARTHUR 1999, BEINHOCKER 2007) and especially recent economic turbulences should have shifted our attention away from the equilibrium towards extreme events, which potentially have pervasive and long-lasting impact on the regional economy, industrial structure and society. In a regional economy with zero average growth and a reasonable fluctuation of 0.1 (i.e., the standard deviation of normality), the probability of doubling, or analogously of declining by at least half the number of regional employees is, according to the normal distribution an infinitesimal 1.5e-8%. This probability increases to an already noticeable 0.5% for the Laplace, and reaches a remarkable 1.6% and 5.2% for the fat-tailed EP with a shape parameter *b* of 0.8 and 0.5, respectively. This research shows that the latter probabilities are the most plausible ones in the reality of regional economies. The apparent necessity of shifting the focus to the tail events is reflected by recent advances in the regional development literature, where concepts like regional vulnerability and resilience have been introduced (PIKE et al. 2010). Acknowledging the high frequency of extreme employment growth events provides an entry point into this new stream of literature. The damage to a car that crashes ten times against a wall with the speed of 10 kmh compared to a car that hits the wall only once but with 100 kmh will not be the same. Assuming that the impact of an industry employment shock to a regional economy is likewise nonlinear, extreme growth events may have long-lasting and path-breaking effects on regional development, as they gain the potential to cross regional thresholds of robustness and trigger cascades of further changes in related industries (MCGLADE et al. 2006). If resilience is the aim of regional policy, it is of high relevance to understand the ability of regional economies to withstand and resist extreme events (*engineering resilience*), to absorb and accommodate them without undergoing any major structural or functional changes (*ecological resilience*), and to adapt and respond to them in a way that allows

favourable structural change and increases the long-term growth prospect (*evolutionary view on resilience*) (SIMMIE/MARTIN 2010). Analysing the stochastic characteristics and causes of extreme events is a first step in this direction.

Economic models should take these stylized facts seriously. Statistical inference from models based on the assumption that the observations are drawn from a normal distribution may deliver invalid implications (FAGIOLO et al. 2008). Three alternative econometric approaches seem to be promising in terms of modelling and explaining regional growth. First, new insights into the dynamics of growth can be gained by looking at the entire shape of the distribution via quantile regression techniques, and thus acknowledging the fact that extreme growth events are not just mere outliers but a fundamental phenomenon of regional growth (e.g. COAD 2007, REICHSTEIN et al. 2010). Secondly, the estimation of the models can be performed directly on the assumption of fat-tailed errors, as suggested by MINEO (2003) or FAGIOLO et al. (2008), and first applied by SCHLUMP and BRENNER (2010) using EP distributed errors. Finally, recent advances in the machine learning community (e.g. SHIMIZU et al. 2006) show that the information which resides in the non-normally distributed data can be fruitfully exploited to establish causal relationships within a structural vector autoregression (SVAR) framework. MONETA et al. (2013) provide a first econometric application in the case of firm growth.

In an attempt to explain the emergence of the stochastic properties of regional industry-specific employment growth, the distributional parameters of the industry-specific growth rates are compared across various industrial classification schemes. The most clear-cut result is found for the shape parameter *b* – knowledge-intensive industries are significantly more exposed to experience extreme negative as well as extreme positive growth events. Many further potential economic mechanisms may be able to explain the emergence of fat-tailed distributions at the regional level (ALFARANO/MILAKOVIC 2008: 274). More light could be shed on the determinants of the distributional characteristics by explicitly taking into account phenomena like industry life cycles, demand shocks, global trade linkages, the relative frequency of lumpy growth events like the entry of new or the exit of existing firms, or the degree of importance of innovative and therefore inherently disturbing activities. Besides these industry-related factors, spatial measures such as Moran's I, which serve as direct proxies for correlating mechanisms at the level of regions, may have explanatory power with respect to the distributional parameters. The strength and configuration of spatial interactions between regions, the structural characteristics of regions (for instance, urban versus rural areas), or the extent of the spatial concentration of innovative activities (FORNAHL/BRENNER 2009) are expected to influence the dispersion of regional industry-specific employment growth as well as the exposure of regions to experience extreme events of growth and decline.

3 Growth dynamics in regional systems of technological activities – A SVAR approach

Abstract: This paper analyses the causal relationships in regional technological systems within a structural vector autoregression (SVAR) framework. Applying a data-driven identification strategy based on Independent Component Analysis, it shows how the regional growth dynamics of economic, research, innovation and educational activities affect each other instantaneously and over time. By matching differently classified data of employees, patents and graduates, the analysis is based on a unique database which embraces multi-dimensional aspects for five clearly separated industries. The findings on how industry-specific growth processes unfold are explained by referring to the type of industry and its knowledge base. For instance, it is found that more engineering-oriented industries like machine tools show a success-driven pattern, in which innovations result in a stronger subsequent engagement in research activities, whereas a linear model from research to innovation is observed for chemistry. Such knowledge on the causal relations among the various activities in regional technological systems is of utmost relevance to the design and implementation of policy instruments.

Keywords: regional growth, SVAR, non-normality, innovations, universities, R&D, regional technological systems

JEL Classification: C33, O30, R11

3.1 Introduction

The development of regional economies is closely related to the various technological activities taking place, such as economic, research, innovation, or educational activities. These activities, which compose a regional technological system, evolve and unfold in an endogenous and interdependent way. The interrelated dynamics are of utmost interest for regional scientists and policy makers who aim to foster regional development and to spur economic growth. Since policy instruments are usually tailored to specifically support one of the above activities, knowledge about the causal structure of the dynamic interdependencies in regional technological systems is essential to efficiently design, focus and implement policies. Does an increase in research automatically lead to more innovations and economic growth, as the traditional linear thinking would suggest, or do successful innovations generate economic opportunities and set incentives to perform further research? Are educational and innovation activities direct generators of economic growth? Are there industries in which economic growth (for instance, due to the expansion into export markets) lead to further innovation and workforce skilling? By ignoring questions like these policies often fail to reach their goal of contributing to regional development.

In the literature, the relationships between the activities are mostly studied in isolation. For instance, a huge bulk of literature exists that investigates the link between regional innovation and economic growth (e.g. RODRÍGUEZ-POSE/CRESCENZI, 2008). However, the simultaneous and endogenous development with other interdependent activities, such as research efforts or university education, tends to remain unaddressed. They cannot be neglected if the causal structure in the growth dynamics of regional economies as coherent, multi-dimensional systems should be uncovered. Besides, only few studies exist which take an explicit industry-specific perspective (e.g. BUERGER et al. 2012). Therefore, the study at hand uses a structural vector autoregressive (SVAR) framework to address the issue of causality in systems of endogenous and interdependent variables. The models, which are estimated separately for five industries, go beyond simple temporal associations by recovering the instantaneous causal effects, which would be missed if only the lagged effects are considered (COAD et al. 2012).

For identifying SVAR models, the literature usually refers to theory to derive the necessary identification restrictions. Because theory is not always unambiguous and assumptions like market clearance or perfect factor mobility are rather controversial, the insights from the results are often put into question. Only recently, data-driven identification techniques, which do not rely on any *a priori* theoretical assumptions, have been developed in the machine-learning community (HYVÄRINEN et al. 2010). This paper uses a strategy based on Independent Component Analysis (ICA), as introduced by MONETA et al. (2013) into the economic literature. By exploiting the non-normality of residuals' distributions to establish causal relationships, this strategy is particularly suited for the analysis of systems governed by economic shocks – the existence of fat tails in economic growth rate distributions has emerged as a stylized fact from a huge body of literature (see COAD 2013 for an overview).

The paper is structured as follows. The first part of section 3.2 describes the vector autoregression methodology and discusses identification issues of the structural version. Furthermore, related empirical applications in the literature are summarized. The second

part of section 3.2 sketches a conceptual model of regional systems of technological activities and derives hypothesis on the industry-specific causal linkages among these activities. Section 3.3 presents the identification algorithm based on ICA in detail and discusses further empirical aspects concerning the estimation. Having provided an overview on the data in section 3.4, in the subsequent Section 3.5 the results on the causal relationships are presented and discussed. Section 3.6 concludes by drawing policy recommendations.

3.2 Literature

3.2.1 Some general notes on causal discovery

Discovery processes, with scientific causal discovery being no exception, are based on the systematic exploitation of information on the object in question. Causal discovery in regression models, for instance, aim to go beyond simple correlations and to uncover the causal structure in the variables. This is especially important in the context of regional growth dynamics, because the variables are often highly endogenous (PACK 1994, BREINLICH et al. 2014). This is why ordinary cross-sectional regression methods fail to capture the subtleties of causal feedback (GARNSEY et al. 2006). As STORPER and SCOTT (2009: 157) put it, "causalities are in practice both multidirectional and diachronic".

In the literature, several sources of information necessary for causal discovery can be distinguished. The first candidate is economic theory, which is able to provide a causal interpretation of the observed correlations by judging on the plausibility of alternative causal explanations. The econometrics literature (e.g. ANGRIST/PISCHKE 2008) increasingly advocates experimental designs, like quasi-natural or controlled experiments, as the gold standard for establishing causal evidence. Here, the information arises from the study design. If only observational data, i.e. non-experimental data, is available, methods such as matching estimators or regression discontinuity designs can mimicking the experimental designs, for example by studying some specific historical events (NICHOLS 2007). Instrumental variables are a further source of information often used in the econometrics literature.

Information theory suggests that the scope and reliability of the causal discovery strongly depends on the quality of the information used. In the context of growth, various theories seem to be applicable and growth theories are known to be open ended, meaning that the truth of one theory does not automatically contradict the truth of another (DURLAUF 2001). Controlled experiments are rarely feasible in social sciences, because they are often unethical, too expensive or technically impossible (HOYER et al. 2009). The few examples in the literature on regional performance and agglomeration, like BUENSTORF and GUENTHER (2011), FALCK et al. (2013), or KLINE and MORETTI (2014), exploit some specific historical events. However, the occurrences of such binary settings, which are required for such quasi-experimental designs, are rather the exception than the rule in the realm of endogenous growth processes. Finally, the literature has struggled to find suitable and strong instruments.

An alternative approach relies on additional information stemming directly from the sample data. Again referring to information theory, a system that unfolds according to a normal distribution implies that it is in a condition of maximum entropy. Hence, no additional information resides within the system's state. On contrary, deviations from normality, i.e. from a purely random process, can represent a source of this information. Because one of the few things robustly known about growth rate distributions is their deviation from normality, this regularity can be exploited for causal discovery. By applying the Linear Non-Gaussian Acyclic Model (LiNGAM) algorithm, which was introduced in the machine-learning community by SHIMIZU et al. (2006), the paper at hand tries to disentangle the causal links within structural vector autoregression (SVAR) framework. In the econometrics literature, the possibilities of an identification of linear models with non-normal distributed errors have already been discussed much earlier (REIERSOL 1950).

3.2.2 Vector autoregression and related empirical applications

A brief introduction into vector autoregression

Vector autoregression (VAR) was advocated by Nobel-laureate Christopher SIMS (1990) to analyse fluctuations of interrelated variables in macroeconomic systems. Meanwhile, this multiple equation approach is widely used in empirical economics to investigate co-evolutionary dynamics of systems of variables (MONETA et al. 2013, BUERGER et al., 2012). By treating all variables as endogenous (LÜTKEPOHL 2003), it is especially suited to analyse how (regional) economic systems adjust to exogenous shocks (ALECKE et al. 2010, PATRIDGE/ RICKMAN, 2003).

In the basic model, a system of *K* variables $\mathrm{y}_1, \dots, \mathrm{y}_\mathrm{K}$ is driven by a linear combination of the contemporaneous values of the other variables at time *t*, the past values (up to lag *p*) of all variables, and a vector of random disturbances:

$$y_t = \boldsymbol{B} y_t + \boldsymbol{\Gamma}_1 y_{t-1} + \cdots + \boldsymbol{\Gamma}_p y_{p-1} + \varepsilon_t \tag{13}$$

with **B** and Γ_j $(\mathrm{j} = 1, \dots, \mathrm{p})$ being *K*K* coefficient matrices and ε_t a *K*1* vector representing a zero-mean white noise error process (MONETA et al. 2013). By definition, the diagonal of **B** is set to zero. In the structural setup, $\boldsymbol{\Gamma}_\mathrm{j}$ contain the autoregressive effects, while **B** contains the contemporaneous effects which occur within one time period *t*. Equation (13) is often transformed, by denoting $\boldsymbol{\Gamma}_0 = \mathbf{I} - \mathbf{B}$, into the standard SVAR form:

$$\boldsymbol{\Gamma}_0 y_t = \boldsymbol{\Gamma}_1 y_{t-1} + \cdots + \boldsymbol{\Gamma}_p y_{p-1} + \varepsilon_t \tag{14}$$

Because equation (13) and (14) are endogenous, and hence, the direct estimation of the model biased, a reduced-form VAR model is often derived (MONETA et al. 2013):

$$\begin{aligned} y_t &= \boldsymbol{\Gamma}_0^{-1} \boldsymbol{\Gamma}_1 y_{t-1} + \cdots + \boldsymbol{\Gamma}_0^{-1} \boldsymbol{\Gamma}_p y_{p-1} + \boldsymbol{\Gamma}_0^{-1} \varepsilon_t \\ &= \boldsymbol{A}_1 y_{t-1} + \cdots + \boldsymbol{A}_p y_{p-1} + u_t \end{aligned} \tag{15}$$

Equation (15) corresponds to *K* individual regressions of the variables y_t on the past values of their own and the other variables (STOCK/WATSON 2001, GOTTSCHALK 2001). $\mathrm{u_t} = \mathbf{\Gamma}_0^{-1}\varepsilon_\mathrm{t}$ denotes a vector of random disturbances. In contrast to the structural version, this system of equations, which simply describes inter-temporal relations, can be directly estimated. Hereby, usually a constant term or further control variables are included. However, the reduced-from does not provide any information on the instantaneous effects. Rather, $\mathbf{A}_p$ mixes up **B** with the lagged effects $\mathbf{\Gamma}_\mathrm{p}$. This implies that only the temporal co-evolution of a system can be observed (i.e. correlation), but not which variable directly drives the others within-the-period (i.e. causality) (COAD et al. 2012).

Therefore, the reduced-form is an explorative tool, which is useful for time-series forecasting but not for policy analysis (PARTRIDGE/RICKMAN 2003). Knowledge on the causal relationships is required to assess the expected effects of an exogenous policy stimulus to one variable on the other variables (MONETA et al. 2013, BUERGER et al. 2012). Compared to the reduced-form VAR model, the structural equivalent allows to observe how random shocks unfold and propagate throughout the system (GOTTSCHALK 2001, MONETA et al. 2013). This is achieved by disentangling the contemporaneous effects, which occur when measurements have a lower time resolution than the causal mechanisms, from the lagged effects, which occur when the time resolution is higher (HYVÄRINEN et al., 2010). Reversed causality, which often is prevalent in dynamic and endogenous systems, is not an issue, "because by tracing out the dynamics of the system to an unexpected shock the causality is pinned down and runs unambiguously from the shock to the other variables in the model" (GOTTSCHALK 2001: 27). Thus, GOTTSCHALK (2001) and KUERSTEINER (2008) conclude that the shock analysis of the SVAR is the best alternative for situations where the variables are endogenously determined and where controlled experiments are not feasible.

The main challenge lies in the identification of the SVAR model. The VAR parameters, which are directly estimable, are not sufficient to recover the parameters in **B** and $\mathbf{\Gamma}_\mathrm{p}$, which are much more in numbers (MONETA et al. 2013). Without any further assumptions, the same reduced-form model could support infinitely different SVAR models (GOTTSCHALK 2001). In the literature, two general approaches exist to identify the structural information. On the one hand, economic theory provides non-sample information that can be used as external identifying restrictions (STOCK/WATSON 2001, LÜTKEPOHL 2003). However, this approach faces several problems. Already SIMS (1980) noted that only few powerful *a priori* restrictions can be derived from theory. Often, competing theories exist, or one theory even allows alternative causal orderings of the variables (DEMIRALP/HOOVER, 2003, MONETA et al. 2013). This problem becomes aggravated in larger systems of variables, as the number of necessary restrictions grows exponentially with *K*. On the other hand, data-driven approaches try to overcome the difficult search for credible restrictions by letting data speak (DEMIRALP/HOOVER 2003). The approach used here is based on a Linear Non-Gaussian Acyclic Model (LiNGAM), which was introduced in the machine-learning community by SHIMIZU et al. (2006) and extended by HYVÄRINEN et al. (2010) to the SVAR framework. The identification becomes possible by making three basic assumptions (MONETA et al. 2013). First, the model is acyclical, i.e. there are no cycles or feedback loops in the causal relations within one time period. Second, the SVAR errors ε_t are

independent. Finally, ε_t are not normally distributed. Taken together, the three assumptions suffice to fully recover the contemporaneous causal relations by using Independent Component Analysis. More precisely, the information of non-normality helps to identify a set of independent latent shocks. Subsequently, the estimates of **B** and $\mathbf{\Gamma}_p$ are recovered by searching for an acyclical causal ordering of the variables (MONETA et al. 2013). The exploitation of the information on the distributional characteristics of the error terms makes this approach particularly attractive for the analysis of economic systems. From a huge body of literature it has emerged as a stylized fact that their growth dynamics, independent of the level of aggregation, are not normally distributed but can be characterized by fat tails (see, for example, BOTTAZZI et al. 2011 for firms, DUSCHL/BRENNER 2013 for regions, or FAGIOLO et al. 2008 for countries).

Empirical applications of VAR/SVAR methodology in related fields

The reduced-form VAR methodology has been introduced in regional science by BLANCHARD and KATZ (1992), who analyse the effects of demand shocks on employment, unemployment and wages. With ALECKE et al. (2010) and VEGA and ELHORST (2013) we highlight two recent studies which were inspired by this seminal contribution. Whereas the former investigates the linkages between regional labour market variables and internal migration flows among German states within a panel VAR setup, the latter extends the BLANCHARD and KATZ model using a dynamic spatial panel data approach to assess both the temporal and spatial propagation of labour demand shocks. Focusing more on the innovation aspects of regional development, BUERGER et al. (2012) analyse the co-evolution of patents, R&D and employment of German labour market regions separately for different industries.

To fully grasp the causal relations among the variables, soon the structural version has become popular. For example, FRITSCHE and LOGEAY (2002) study unemployment dynamics of the German national economy, while PATRIDGE and RICKMAN (2003) tackle the famous question in urban economics whether people follow jobs or jobs follow people. Both studies are based on simple labour-market models to deduce identification restrictions. The latter, for instance, assumes the equality of labour-demand supply, constant returns to scale, perfect labour and capital mobility in the long run.

To provide an alternative to the theory-guided view, MONETA et al. (2013) has recently introduced a data-driven identification strategy, named VAR-LiNGAM, into the empirical economic literature. In two cases studies, the authors apply this methodology to understand the dynamics and causal relationships of different aspects of firm growth and the impacts of changes in monetary policy on macroeconomic variables. COAD et al. (2012) confirm the usefulness of this approach in a more comprehensive empirical contribution which seeks to understand how growth processes of high-growth firms unfold. The algorithm's popularity is also increasing in other fields, like welfare economics (COAD/BINDER 2014) or within sociology, psychology, neurosciences, genomics, epidemiology, genetics, engineering and chemistry (for an overview on recent publications see SHIMIZU 2014).

However, a data-driven identification strategy, which does not rely on often untenable theoretical assumptions, has not yet been applied at the level of regional economies. Hence, we position the present study within the existing body of literature as an advancement of BUERGER et al. (2012) by introducing the VAR-LiNGAM identification strategy of a structural VAR model into the literature of regional science and economic geography.

3.2.3 Causal relations in the dynamics of regional systems of technological activities

The framework of regional system of technological activities

We conceptualize a regional technological system as consisting of four activity categories: economic, research, innovation and educational activities. The dynamics in economic activities are generated by the growth, exit or entry of firms. Research is measured by R&D conducted in firms, which expand or reduce their engagement in (private) research activities. Furthermore, a regional economy is often assessed by its innovative capacity or the innovation activities taking place in firms or the public research institutes. Finally, education assumes a key role in the modern knowledge-based economy.

These activities are strongly interrelated and their dynamics evolve highly endogenously. The growth of one activity impacts on the growth of the others, which in turn propagates further throughout the system.[16] Therefore, this section aims to derive hypotheses on the expected causal linkages among the four interdependent activities of a regional technological system by elucidating the underlying mechanisms. Besides, these linkages depend on the industries under consideration. Already PAVITT (1984) observed that industries differ in their innovation patterns. Likewise, the literature on technological regimes (WINTER 1984) and subsequent work on sectoral innovation systems (BRESCHI et al. 2000, MALERBA 2002) have highlighted several dimensions along which industries systematically differ. More recently, the literature stresses industry-specific knowledge bases, which impact on innovation and learning processes. Here, ASHEIM and co-authors (e.g. ASHEIM/GERTLER 2006) mainly distinguish between the analytical and synthetic knowledge base. Industries, in which the analytical knowledge base predominates, tend to rely more on scientific knowledge generated both internally and externally. Compared hereto, industries that are characterized by a synthetic knowledge base generate innovation through application and novel combinations of existing knowledge in the context of daily problem-solving situations (ASHEIM 2007).

This paper analyses and compares five industries. First, the chemical industry (CHEM) is a typical representative of science-based manufacturing. Mainly based on analytical knowledge, it strongly relies on systematic R&D for generating innovations (ASHEIM 2007).

---

[16] Because of this endogeneity, STORPER and SCOTT (2009 157) note that the cross-section regression analysis of regional growth is problematic: “causalities are in practice both multidirectional and diachronic”. For a more extensive discussion on the use of structural models in the light of endogeneity to analyse regional growth, see BREINLICH et al. (2014).

Additionally, patents play an important role to appropriate returns from innovations (BRESCHI et al. 2000). Secondly, the machine tool industry (MACH) can be classified as specialised supplier manufacturing (see PAVITT 1984, or revised taxonomies such as BOGLIACINO and PIANTA, 2010). Foremost in Germany this industry consists of rather small and relatively specialized firms, which tend to rely on a synthetic knowledge base. Here, other appropriation methods than patents play a more important role. The electronics industry (ELEC) can be regarded as an industry that shares both the properties of science-based and specialised supplier manufacturing and that combines elements from a synthetic and analytic knowledge base. In contrast hereto, scale-intensive manufacturing, like in the metal (META) or transportation (TRAN) industry, is characterized by a high concentration of rather large firms (BOGLIACINO/PIANTA 2010). Yet the two mentioned industries differ in their knowledge base. Whereas synthetic knowledge prevails in TRAN (ASHEIM 2007), META is considered to share many characteristics of the analytical knowledge base of CHEM.

Implications of the industries' properties are considered in the subsequent discussion of the expected causal linkages among the different activities in regional technological systems.

Expectations on the causal relations in the dynamics of the activities

From the literature, expectations on bivariate causal linkages among the four different activities can be derived. Figure 18 summarizes the causal relations that below are discussed in detail.

Affected variables \ Causing variables	Economic	Research	Innovation	Educational
Economic			+ / -	+ (MACH)
Research	+		+ (MACH, TRAN)	
Innovation		+ (CHEM, META)		+ (CHEM, ELEC)
Educational				

Figure 18: Expected causal relations in the dynamics of the activities of regional technological systems (if the causal relations are expected especially in some industries, these industries are given in brackets)

The link between research and innovation activities seems to be clear at a first glance. Private R&D is usually regarded as an important innovation generator both at the organisational and regional level (BRENNER/BRÖKEL 2011, or COOKE et al. 1997 in the context of innovation systems). Hence, we expect that an increase in research activities leads to more innovations, especially so in science-based industries. However, the linear thinking has recently been challenged. Instead, a success-breeds-success phenomenon often is observed empirically (BUERGER et al. 2012). We argue that this is mainly an "engineering phenomenon" and can be observed in industries based on synthetic knowledge, such as MACH and TRAN. Here, innovations are often generated through learning-by-doing and interactively with suppliers or customers. Once innovations promise economic success, investments in systematic R&D are intensified. In short, a positive causal effect can be inferred from theory into both directions, probably depending on the industry.

The link between innovation and economic activities is less unambiguous. Extending the linear model, innovations which result from research are often seen as an important driver of economic growth. Focusing on the effects of employment, which is the preferred indicator of economic activities in this paper, an additional distinction between product and process innovation becomes necessary. Whereas positive employment effects are expected due to growth opportunities as a result of new products, negative effects may arise from rationalisation measures as a result of process innovations (see BUERGER et al. 2012 for a literature overview). Hence, the net effect remains an empirical question.

A direct link is also expected for research and economic activities. In the empirical literature on firm growth strong evidence exists that growth in employment is associated with subsequent growth in R&D (COAD/RAO 2010). As firms expand, they also tend to reinforce investments in R&D. We argue that the same rationale holds at the aggregate level of regions. This stands in contrast to the main idea of endogenous growth theories which assume that economic growth is driven by accumulation of new knowledge through intentional R&D (e.g. ROMER 1990). This link rather works indirectly (via innovations) and with some time lag.

Considering educational activities, a clear relationship seems to exist with innovation activities. "Man- (or, better, brain-) power […] are needed for creating innovations" (BRENNER/BROEKEL 2011: 12). For instance, graduates from higher educational institutes bring up-to-date knowledge into firms to enable innovation processes (BRENNER/SCHLUMP 2013). Although being the most mobile group in society, a large fraction of graduates stay in their region of education after graduation (MOHR 2002). We expect that the impact of higher education on innovations is strongest in industries which rely on new scientific knowledge, such as CHEM and ELEC.

Furthermore, educational activities can be assumed to directly impact economic activities. The role of human capital as a main driver of economic growth (LUCAS 1998) is widely acknowledged at the regional level (e.g. CRESCENZI 2005). In this context it has been shown empirically that especially strongly expanding firms rely on the local presence of qualified graduates (DUSCHL et al., 2014a). BRENNER and SCHLUMP (2013) argue that regional growth is fostered by universities only if regional labour markets offer job opportunities. Otherwise, graduates tend to leave their region of education. Hence, we

expect that the impact of educational activities on economic activities is strongest in industries where the supply of qualified labour is relatively scarce, which foremost should be the case in MACH.

No direct link in either ways is expected to be found between educational activities and research activities.

3.3 Empirical method

This section introduces the VAR-LiNGAM identification algorithm using ICA. Both the ICA and the underlying assumptions for identification are described in more details. Finally, further aspects concerning the empirical implementation are discussed.

The VAR-LiNGAM identification algorithm using ICA

Once having estimated the reduced-form VAR regressions of equation (15), the goal is to recover $\mathbf{\Gamma}_0$ and thus **B**, which contains the contemporaneous causal relations. The strategy chosen here is to relate the VAR errors to the SVAR shocks by the expression $\mathrm{u_t} = \mathbf{\Gamma}_0^{-1}\varepsilon_\mathrm{t}$ (MONETA et al. 2013). $\mathbf{\Gamma}_0$ can be uniquely identified using ICA given three general assumptions: the shocks ε_t are independent and not normally distributed and the contemporaneous causal structure among y_t is acyclic. With the help of ICA, latent components, which are not directly observable, can be recovered from the observed random variables, provided that these components are statistically independent and not normally distributed (HYVÄRINEN/OJA 2000). In other terms, the original independent components, representing economic shocks, are recovered by exploiting the information of non-normality in the error terms. However, the components are only found up to permutation, sign and scaling, meaning that they still allow various specifications of $\mathbf{\Gamma_0}$ (MONETA et al. 2013). To finally establish a unique correspondence between the identified components and the variables, an acyclical causal structure is assumed. This implies that $\mathbf{B} = \mathbf{I} - \mathbf{\Gamma_0}$ should be lower triangular if the variables are ordered accordingly, i.e. the entries above the diagonal become zero (HYVÄRINEN et al. 2010). Hence, by arranging the variables and by finding a permutation of **B** for which the elements in the upper triangle are as close to zero as possible, the unique contemporaneous causal structure is identified. In the following section, we want to provide the intuition behind ICA and to discuss the three assumptions underlying the identification strategy.

Independent Component Analysis

ICA represents the basic building block of the SVAR identification strategy. As a probabilistic method it seeks to recover the original signals or processes, which cannot be directly observed, from measured signal mixtures. Because the latent, unobservable signals, also called independent components, usually contain important information, this method has become popular in various fields like digital image processing, biomedical

signal processing, telecommunication, neurology, or finance (HYVÄRINEN/OJA 2000, HYVÄRINEN 2013). The most prominent example is the cocktail-party problem, which assumes two simultaneously speaking individuals and two microphones. The microphones can only record mixtures (the observed random variables) of both voices (the independent components). Because the microphones are located in different parts of the room, the voices are present with different weights in the recorded signal mixtures. By linearly transforming the random variables, ICA aims to find those components that are as statistically independent as possible. The central limit theorem implies that the sum of independent random variables moves closer to normality than any of its original variables (HYVÄRINEN/OJA 2000). In other words, the additive mixture distribution is closer to normality than any of its independent components, which, consequently, can be found by maximizing some measure of non-normality. Various measures like kurtosis or negentropy exist which can handle both leptokurtic and platykurtic distributions, that means distributions with tails fatter or thinner than the normal one. Negentropy is more robust than kurtosis-based measures and from statistical theory the optimal estimator for non-normality, however it is difficult to compute as the densities of the independent components must be known (HYVÄRINEN/OJA 2000). With FastICA, a fixed point, semi-parametric ICA algorithm, the negentropy can be efficiently approximated (HYVÄRINEN 1999).

The three assumptions for full identification

As mentioned above, the SVAR model can be uniquely identified given three general assumptions: the shocks ε_t are independent and not normally distributed and the contemporaneous causal structure among y_t is acyclical.

Statistical independence means that information of a value of one variable does not give any information on the value of the other variable.[17] In economic terms, ε_t are regarded as structural innovations or "primitive" exogenous forces, which affect the system at each time period and cause its dynamics and oscillations, but do not depend on the shocks before (GOTTSCHALK 2001; MONETA et al. 2013). However, only mixtures of the structural economic shocks are observed. Therefore, testing for statistical independence only provides information on the independence structure of the decomposed components by ICA. Whether or not the true economic shocks are independent ultimately has to be judged on background knowledge on economic processes (MONETA et al. 2013). Thus, it is not clear whether this assumption is satisfied and we will empirically check it below.

The independent components are recovered by maximizing their non-normality. However, this does not work if the components, here the economic shocks ε_t, are normally distributed. In that case, their joint density would be completely symmetric, thus providing insufficient information for the decomposition of the observed signal mixtures (HYVÄRINEN/OJA 2000). This assumption, fundamentally empirical in nature (HYVÄRINEN et al. 2010), is supported by a huge body of literature which argues that economic shocks tend to deviate from normality (e.g. MCKELVEY/ANDRIANI 2005). Non-normality usually

[17] In case of non-normal distributed random variables, this requirement is stronger than linear uncorrelatedness (MONETA et al. 2013).

remains even after controlling for other variables (MAASOUMI et al. 2007) or the lag structure (BOTTAZZI et al. 2014). Nevertheless, it is suggested to test for non-normality. Any form of non-normality is allowed (MONETA et al. 2013) and if only one independent component is normally distributed, ICA still is possible (HYVÄRINEN/OJA 2000). To sum up, if the independent components deviate from normality, the information on their distribution can be exploited to identify the SVAR model using ICA (MONETA et al. 2013). The assumption of non-normality of economic shocks is supported by and will be additionally tested for our case.

The full identification in form of a unique contemporaneous causal ordering of the variables in y_t, however, becomes only possible if acyclicity is assumed. The implied recursive structure restricts two variables from being their mutual cause in one time period, giving rise to self-reinforcing feedback loops (COAD et al. 2012, MONETA et al. 2013): if *a* causes *b*, then *b* cannot cause *a* (HYVÄRINEN 2013). With this additional assumption the permutation, sign and scaling indeterminacies of ICA can be fixed and the SVAR shocks ε_t uniquely connected to the components of u_t in a one-to-one relationship. Therefore, **B** is arranged and permuted in such a way that the major causal directions receive more weight, whereas the relatively minor causal directions are minimized towards zero (MONETA et al. 2013, COAD et al. 2012). This leads to a causal ordering of the variables so that the shock to y_1 feeds into $y_2, \ldots, y_K$ within the same time period, the shock to y_2 into $y_3, \ldots, y_K$, and so forth (DEMIRALP/HOOVER 2003). Even though acyclicity is a common assumption in the literature (MONETA et al. 2013), a cyclical alternative was proposed by LACERDA et al. (2008), which might be of interest in future research. Our theoretical discussion above has shown that only in the case of one pair of variables (research and innovation) we have theoretical arguments that lead us to expect a mutual causal relationship. Hence, this assumption seems to be not far from reality.

Summary of the algorithm

Before we highlight further aspects concerning its empirical implementation in the context of regional economic growth, Figure 19 summarizes the algorithm for identifying a structural VAR model. For a detailed mathematical description we refer to MONETA et al. (2013), SHIMIZU et al. (2006) or HYVÄRINEN et al. (2010).[18]

[18] The named authors also provide an implementation of the algorithm for R, which was adapted for the use in this paper (http://www.cs.helsinki.fi/u/entner/VARLiNGAM).

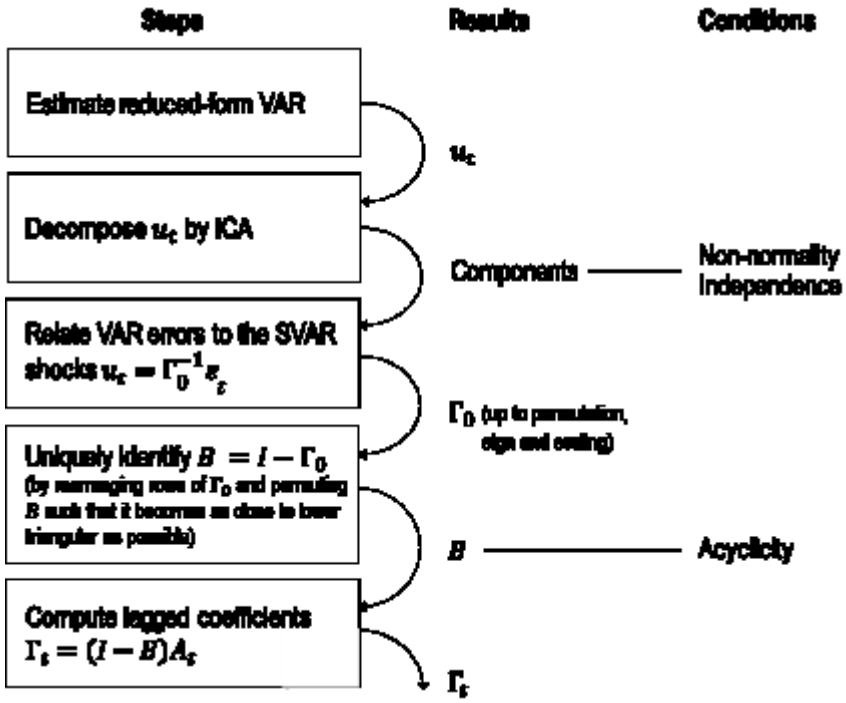


Figure 19: Summary of the identification algorithm

Further empirical aspects concerning the implementation

To remove the time-invariant region-specific effects (MONETA et al. 2013), growth rates $g_{i,t}$ instead of levels $S_{i,t}$ are used for the variables $y_{i,t}$:

$$g_{r,i,t} = log(S_{r,i,t+1}) - log(S_{r,i,t}) \tag{16}$$

with *r* denoting the region. Next, the growth rates are rescaled to control for the inverse relationship between their size and variance, a universal feature in the growth of complex economic organisations (AMARAL et al. 2001). BOTTAZZI et al. (2014) model the scaling relationship directly by introducing a heteroskedasticity term into the stochastic growth process:

$$g_{r,i,t} = \alpha_i + v_{r,i,t} \tag{17}$$

where α_i is a constant term that converges to the average growth rate. The error term can be written as $v_{r,i,t} = \exp(\beta_i(s_{r,i,t} - \bar{s}_i))\, e_{r,i,t}$, with $s_{r,i,t} = \log(S_{r,i,t})$ and $\bar{s}_i$ is the corresponding industry-specific arithmetic mean over regions and time. This expression takes into account that the functional form of heteroskedasticity might be non-linear, as recently observed in the firm growth literature (e.g. BOTTAZZI et al. 2011). Replacing the heteroskedastic error term and solving for $e_{r,i,t}$, we get:

$$e_{r,i,t} = \frac{g_{r,i,t} - \alpha_i}{exp(\beta_i(s_{r,i,t} - \bar{s}_i))} \tag{18}$$

The last equation yields the rescaled and normalized growth rates $\tilde{g}_{r,i,t} := e_{r,i,t}$. Rescaling cleans the data from heteroskedasticity and normalization from the average growth trend. The latter controls for common macroeconomic shocks and cyclical effects (PARTRIDGE/RICKMAN 2003) and shifts the focus on the stochastic part, which usually contains the "behavioural relations" (LÜTKEPOHL 2003). Anticipating non-normality, equation (19) with two unknown parameters is estimated by minimizing the absolute deviations (LAD):

$$\{\beta_{i,t}; \alpha_{i,t}\} = argmin_{\beta,\alpha} \sum_r \left| \frac{g_{r,i,t} - \alpha_i}{exp(\beta_i(s_{r,i,t} - \bar{s}_i))} \right| \tag{19}$$

The rescaling step is performed for each industry, year and variable separately to account for possible differences in the variance-size relationship.

Being more than a pre-requisite for the VAR-LiNGAM identification strategy, non-normality also concerns the estimation of the single VAR equations. In this case, OLS may provide unreliable results. When errors follow a double exponential distribution, minimizing the absolute deviations is equivalent to the log-likelihood function. In all other cases, tail events affect less the estimators compared to OLS, in which the residuals are squared. Hence, LAD provides more reliable results than OLS in particular for fat-tailed error distributions (DASGUPTA/MISHRA 2004, MAASOUMI et al. 2007).

The selection of the number of lags *p* is based on various statistics, like the Akaike Information, the Hannan-Quinn or the Schwarz Criterion (LÜTKEPOHL 2003). Here, all criteria advocate a 1-lag model, which is driven by the disproportionate loss of information that is not counterbalanced by additional explanatory power from the inclusion of further lags. Because the selection of lag lengths might statistically collide with the determination of the causal ordering (DEMIRALP/HOOVER 2003), we checked whether the latter stays robust when increasing the number of lags. No changes in the causal ordering are observed and the estimates remain very similar in a 2-lag model.

Finally, p-values are estimated by bootstrapping with 500 replications. The bootstrap analysis is also used to assess the stability of the causal ordering of the variables. This is important because the effects of a shock might depend on the way the variables are arranged (LÜTKEPOHL 2003).

3.4 Data

To measure the co-evolving activities of a regional technological system, namely economic, research, innovation, and educational activities, the following variables are constructed.

Regional economic activities can be measured by the number of employees (*Empl*). The German Institute for Employment Research (IAB) provides industry-specific employment data. Private research activities are approximated by R&D employees (BRENNER/BRÖKEL 2011). These are defined as the occupational groups of engineers, chemists and natural scientists (BADE 1987) and retrieved from the same database as *Empl*. [19] Education activities can be measured by the number of graduates from higher educational institutes (*Grad*). Data on graduates stems from the German Federal Statistical Office (destatis) and encompasses universities in a narrower sense as well as universities of applied science, that is, all graduates with a technical, diploma, bachelor and master degree. Innovation activities are reflected by patents (*Pat*). Although this indicator is not without problems, it is the most widely used in the literature (see SMITH 2005 for an extensive discussion). All patents with at least one inventor address in Germany are taken from the European Patent Organization's (EPO) Worldwide Statistical Patent Database, the so-called PATSTAT database. To assign the patents to the regions, the inventor's addresses are used as they most closely represent the places where the innovations took place (BRENNER/SCHLUMP 2013).

Further control variables can be included into the reduced-form VAR model (equation (15). Here, two variables are chosen that account for the general regional socio-economic environment. First, the population density (*Pop*) represents a catch-up variable of several unobserved region-specific factors (FRITSCH/SLAVTCHEV 2011). For instance, it measures urbanization economies, which are rather independent from the surrounding industrial structure (BUERGER et al. 2012). Second, the unemployment rate (*UR*) reflects the vitality of the regions' socio-economic conditions. In the special case of Germany it also accounts for structural differences along the east-west and north-south divide. Data on both variables is obtained from destatis. Because of their asymmetry, they were first normalized by division through the mean value and then made symmetric by the transformation $\tilde{x} = (x - 1)/(x + 1)$.

The spatial unit of analysis are 270 labour market regions as defined by the IAB (BINDER/SCHWENGLER 2006), which have been also used in the related study of BUERGER et al. (2012).[20] All variables span the years 1999 to 2008. Furthermore, the four main variables are measured separately for various industries. The industries are represented by five broad groups, which result from aggregating up 19 distinct technological fields (BROEKEL 2007). The technological fields are used as an intermediary to match both data based on the International Patent Classification (IPC), like *Pat*, and data based on the standard industry classification (NACE Rev. 1.1), like *Empl* and *RD*, according to a current version of the concordance developed by SCHMOCH et al. (2003).[21] *Grad*, distinguishable by their field of study, is assigned to IPC based on a professor-patent matching. Therefore, all German patent applicants with a professor title are identified and, if possible, affiliated to a university faculty. From this, the graduates can be assigned to the technologies

[19] For statistical reasons, the number of R&D employees is subtracted from *Empl*, because they build their own variable (*RD*).

[20] Using Moran's I test statistics, spatial autocorrelation in the growth rates variables were not found to be present.

[21] The current version is an update of the concordance originally published by SCHMOCH et al. (2003) and was obtained directly from the author. For a full list see BRENNER and SCHLUMP (2013).

according to the resulting contribution shares for each study field to the patents classes (BRENNER/SCHLUMP 2013). Figure 20 summarizes the matching and aggregation procedure of differently classified data, resulting in a unique database, which embraces multi-dimensional aspects for five clearly separated industries.[22] However, it is obvious to note that the variables are mere proxies for the technological activities. Often, the variables may reflect different underlying quantities. Hence, the economic interpretations of the findings must be treated with caution.

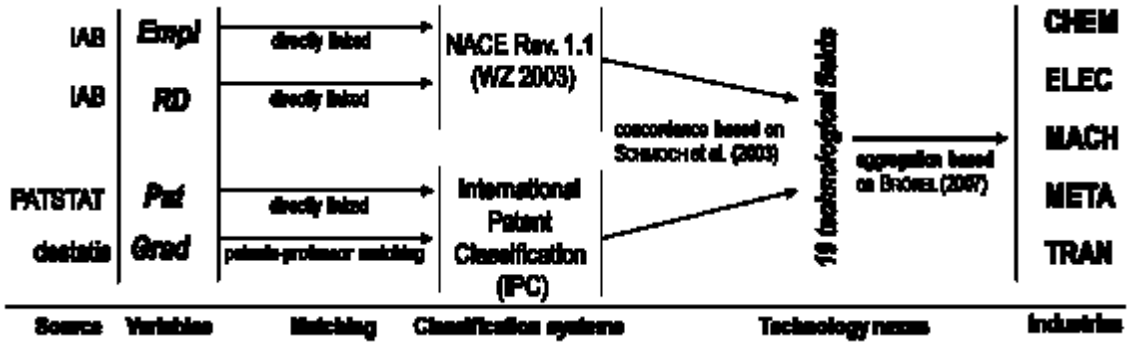


Figure 20: Data matching and aggregation

Hence, industry-specific growth rates are constructed for all activity variables.[23] In the previous section it is argued that the growth rates should be corrected for the general growth trend and heteroskedasticity. Optimizing equation (19) yields an industry- and year-specific normalization $\alpha_{i,t}$ and rescaling $\beta_{i,t}$ parameter. Whereas the general growth trend varies for the variables and years, the rescaling parameter mostly ranges between the values of -0.2 and -0.3, which confirms the literature (AMARAL et al. 2001, DUSCHL/BRENNER 2013).

3.5 Results

Testing the three identifying assumptions

The non-normality assumption of ε_t is assessed by three test statistics: the Shapiro-Wilk (SW) normality test, which is generally known to have a high test power; the Kolmogorov-Smirnov (KS) test against normality, which is more robust against outliers; and the Jarque-Bera (JB) test, which focuses on the skewness and kurtosis. Finally, the flexible 5-

[22] The bivariate correlation coefficients between the growth variables $\tilde{g}_{r,i,t}$ are all below the threshold of 0.70 and, except for the pair *Empl* and *RD*, tend to be rather small. Hence, multicollinearity is not an issue and the variables seem to reflect different facets of regional economic development.

[23] To ensure statistical reliability, growth events which are based on less than five units of their corresponding level variable are excluded from the analysis. An unbalanced panel is possible, unless in one single region-year the values for all variables are available (COAD et al. 2012).

parameter Asymmetric Exponential Power (AEP) distribution (BOTTAZZI/SECCHI 2011), often used in the literature to describe fat-tailed economic growth processes, is fitted to the four recovered components. Because the normal distribution is a special case of the AEP-density family, the likelihood-ratio (LR) test can be used to assess whether the normal distribution is significantly outperformed in terms of asymmetry and/or tail behaviour. Table 7 shows the number of components for which the normality tests do not reject the null hypothesis. Above it is noted that ICA is still possible if only one independent component is normally distributed (HYVÄRINEN/OJA 2000). This condition holds consistently for each industry except for META if the KS-test is used. Hence, our theoretical expectation that non-normality is prevalent in most cases is confirmed empirically.

Table 7: Validity of the three identifying conditions

Condition	Non-normality				Independence	Acyclicity
Test	SW-test	KS-test	JB-test	LR-test	GG-test	Bootstrapping
n	4	4	4	4	6	500
CHEM	1 (25%)	1 (25%)	1 (25%)	1 (25%)	1 (17%)	194 (39%)
ELEC	0 (0%)	1 (25%)	0 (0%)	0 (0%)	1 (17%)	232 (47%)
MACH	1 (25%)	1 (25%)	0 (0%)	1(25%)	1 (17%)	218 (44%)
META	0 (0%)	2 (50%)	0 (0%)	1 (25%)	0 (0%)	253 (51%)
TRAN	0 (0%)	1 (25%)	1 (25%)	0 (0%)	0 (0%)	264 (53%)

The independence assumption is assessed by a test statistic, as recently developed by GRETTON and GYÖRFI (2010), which is based on kernel estimates of the distance between the joint densities and the product of the marginal densities. Here, each of the six possible pairings of the recovered components is tested. In none of the industries more than one pair of components is identified to be statistically dependent at a 5%-significance level (see Table 7). Because ICA has proven to be robust against some degree of dependence (HYVÄRINEN 2013), this condition can be said to hold.

The acyclicity assumption cannot be tested directly. Therefore, its validity is assessed by bootstrapping in order to see whether the causal ordering of the variables remain robust. The analysis reveals that in around half of the 500 replications the same causal ordering of the variables results. Considering that the probability that one specific ordering occurs by chance is 1/4! = 1/24 = 0.042%, the results provide some evidence against randomness in the ordering.[24] The frequency distribution of all causal orderings (see Appendix X.2), resulting from the bootstrapping analysis, shows that there exist at maximum one or two alternative orderings. Hence, as expected, cyclicity plays a minor role but cannot be excluded completely.

[24] We thank Alex COAD for pointing this out.

Results

All assumed conditions being validated, the SVAR model can be recovered from the reduced-form VAR model. The estimated coefficients for the five industries are depicted in Table 8, while the contemporaneous and lagged effects among the variables are visualised in Figure 21. Significant relationships at the 5%-level are connected by an arrow indicating the causal direction of the impact.

Table 8: Estimation results from SVAR model

		B				Γ_1				
		Empl	*RD*	*Pat*	*Grad*	*Empl*	*RD*	*Pat*	*Grad*	N
CHEM	*Empl*	-				.080	-.054	.015	-.015	511
	RD	.695*	-		-.065	-.054	.083	.004	.024	
	Pat	-.049	.114*	-	.172*	-.016	-.022	-.308*	.413*	
	Grad	.021			-	-.012	.032	.003	-.230*	
ELEC	*Empl*	-			.021	.016	.040*	.003	-.008	1352
	RD	.989*	-		-.018	.042	.059*	.012	.023	
	Pat	.109	.024	-	.117*	-.060	.038	-.212*	.117	
	Grad				-	.017	-.025	-.009	-.134*	
MACH	*Empl*	-				.075*	-.003	.018*	.019	913
	RD	1.128*	-		.001	.016	-.006	-.007	-.014	
	Pat	.091	-.039	-	.016	.032	-.042	-.251*	-.080	
	Grad	.093*			-	.015	-.015	.015	-.186*	
META	*Empl*	-				.048	.019	-.011	-.027	266
	RD	.939*	-		.009	.131	.003	.029	.025	
	Pat	-.418	.197*	-	.187	-.157	-.011	-.238*	-.095	
	Grad	.015			-	.108	-.029	.029	-.268*	
TRAN	*Empl*	-	.460*		.060	.014	.019	-.001	-.019	425
	RD		-		-.089	-.052	.069	.052*	-.008	
	Pat	.050	-.043	-	.127	.338*	-.146	-.168*	.143	
	Grad				-	.020	-.018	-.005	-.249*	

p-values: * < 0.05

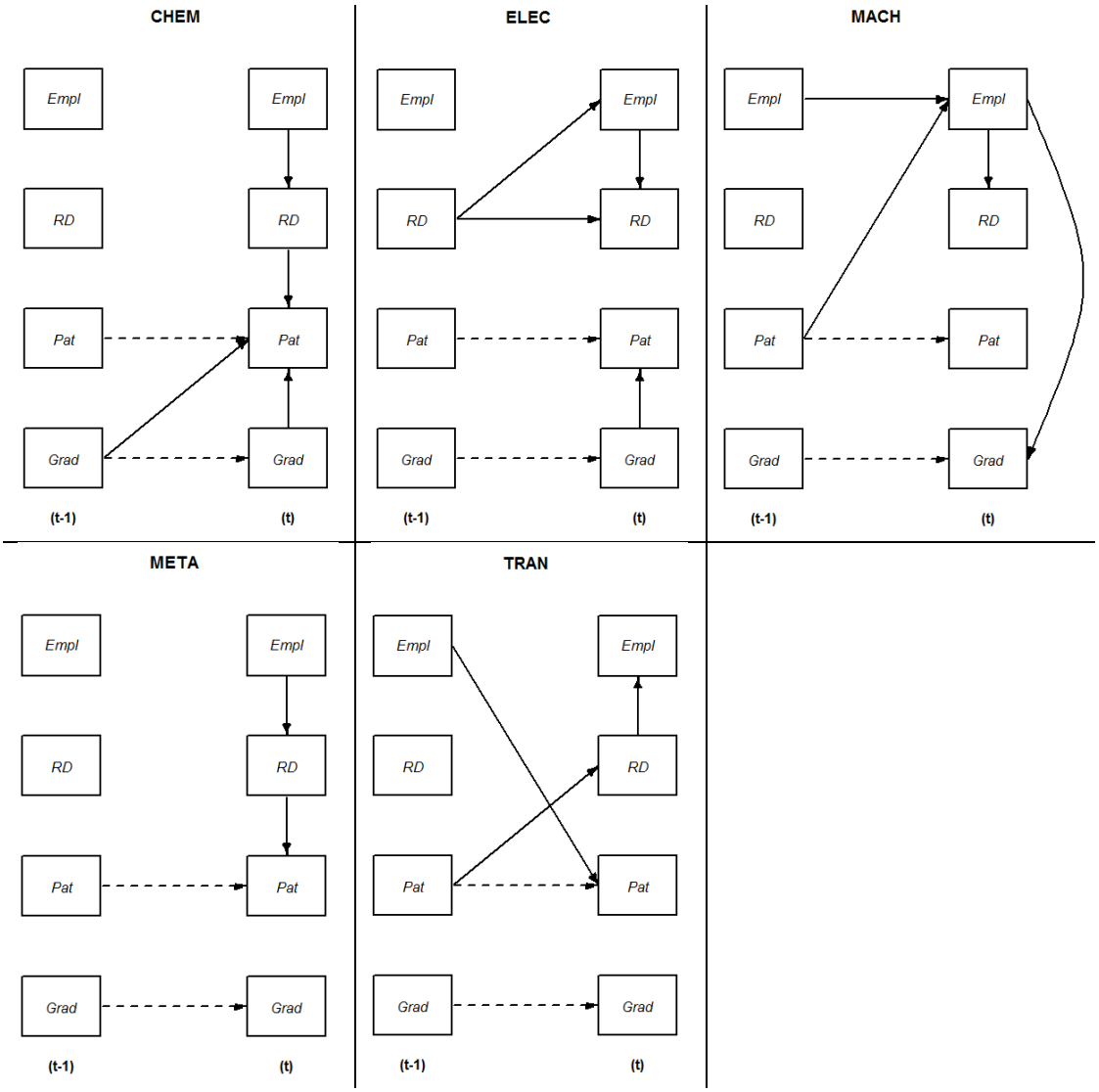


Figure 21: The contemporaneous causal and the lagged effects. Positive relationships are indicated by a solid line, negative ones by a dashed line.

The purely horizontal connections from *t-1* to *t* (or the diagonal entries for Γ_1 in Table 8) represent the temporal autocorrelations of the variables. Whilst *Grad* and *Pat* show a negative autocorrelation, *Empl* and *RD* tend to show a positive, however, less often significant autocorrelation. This reflects the different nature of the variables. Growth of the latter means changes in the regular and research-related stock of employees. The positive autocorrelations indicate the existence of some sort of self-reinforcing mechanisms. In contrast hereto, growth in patents and graduates stems from differences in the yearly realizations of stochastic outcome variables. The negative autocorrelations indicate the existence of year-to-year fluctuations that are much larger than the year-to-year changes due to the overall development. Although the temporal autocorrelation patterns provide interesting insights, the focus of this research is on the causal relationships in the dynamics among the variables. Here, differences between the industries immediately become evident.

Increasing *RD* leads to more *Pat* in CHEM and META. For both industries, a traditional linear thinking seems to apply, which argues that systematic research leads to innovations. For ELEC, the link is still positive, yet insignificant. In contrast hereto, no positive contribution of *RD* on *Pat* is found within-the-period or across time for MACH and TRAN. It might be argued that the temporal resolution of one year is too low to fully capture the effects of research on innovation. In that case our model systematically underestimates the effects. But interestingly, it can be observed within this time frame that more *Pat* leads to more *RD* in the next period for TRAN, and indirectly via *Empl* for MACH. Hence, the more engineering-oriented industries seem to be driven by success: innovations, not necessarily resulting from systematic R&D, provide incentives to increase subsequent research efforts.

In MACH, innovations are also found to be an important driver of economic growth. *Pat* translates into more *Empl* in the subsequent time period. Here, the positive effects of new growth opportunities due to product innovations clearly surpass the negative ones of rationalization endeavours as a result of process innovations. An indirect contribution of *Pat* to *Empl* is observed in TRAN via *RD*. The negative forces tend to prevail in META, however, not significantly. In the remaining industries, the two opposite forces seem to be neutralized. Again it must be noted that especially the positive effects of innovations take more time to become materialized and are hence not fully captured by our model.

The growth of research and economic activities goes strongly hand in hand. In most industries, first the stock of regular employees grows (*Empl*), followed by a change into the same direction of research-related employees (*RD*). This result reproduces robust findings from the empirical literature on firm growth. Only for TRAN, a reversed relationship is observed, resembling the traditional assumption from the endogenous growth theories that economic growth is driven by accumulating new knowledge through intentional research activities. An alternative view on the relationship between *Empl* and *RD* can be put forward by arguing that research-related employees tend to be more high-skilled than regular employees. Hence, the positive relationship found between both variables provides some evidence that skilled labour complements rather than substitutes unskilled labour.

The instantaneous causal effect of *Grad* on *Pat* is positive for all industries, however significantly only for CHEM and ELEC. This confirms our expectations that educational activities are an important factor for innovation activities, especially so in industries which rely on new scientific knowledge. In CHEM, the link even remains significant over time. We have to highlight in this context that *Grad* is our only variable that is connected to university activities. Hence, we cannot be sure whether this variable works (only) as a direct operationalization of university education or (also) as an indirect operationalization of university research (which is highly correlated with university education in space). The strong link to the patent activity suggests that university research is the more important part in our analysis.

This is further supported by the fact that *Grad* is found to influence neither *Empl* nor *RD*. Whilst the lack of a significant relationship with *RD* is not surprising, the lack of a significant relationship with *Empl* is unable to support our expectation that expanding firms rely on the local presence of qualified graduates. Hence, firms either source their highly-qualified employees supra-regionally, or employment growth generally depend less on university

graduates (see, for example, DUTZ et al. (2011), who find that even innovation-driven employment growth mainly depends on the unskilled workforce). For MACH, even a reversed causal link is observed. Here, fluctuations in economic activities and demand in labour might send signals both to students for timing their graduation and to universities for aligning their study and course program.

To summarize, the following general model of how regional technological systems develop can be sketched for CHEM and META: increasing economic activities (*Empl*) imply expanding research activities (*RD*), which then translate into more innovations (*Pat*). For the science-based industry CHEM, educational activities (*Grad*) become an additional factor regarding innovations. A different, rather success-driven pattern is observed for MACH and TRAN: past innovations generate opportunities to expand economic activities, which also facilitate the widening of research activities (MACH), or they set incentives to perform further research, which also demands hiring more regular employees (TRAN). The new employees, in turn, increase research activities in the following time period. ELEC, finally, stands in-between. On the one hand, the expansion of innovation activities relies on educational activities. On the other hand, more research activities in the past increase both research and economic activities in the present. Because economic activities also increase research activities within-the-period, a highly dynamic process, similar to the notion of circular and cumulative causation (MYRDAL 1959), might unfold in this industry.

3.6 Conclusions

This paper applies a new identification strategy for SVAR models to the analysis of regional systems of technological activities. In contrast to the approaches used so far, which often make untenable prior theoretical assumptions on the existence of causal effects, the identification of causal relationships is achieved empirically by exploiting the information of non-normality (HYVÄRINEN et al. 2010). This data-driven approach is particularly useful for analysing dynamics in systems like regional economies, which are well-known to show fat-tailed growth rate distributions. More precisely, non-normality enables the recovering of the independent components by ICA from the errors of a reduced-form VAR model. Assuming acyclicity in the contemporaneous effects, a causal ordering of the variables is possible. Finally, the lagged effects are corrected by taking into account the contemporaneous effects (HYVÄRINEN et al. 2010, MONETA et al. 2013).

The literature agrees upon the usefulness of SVAR models as a tool for policy analysis in situations where controlled experiments are not feasible (GOTTSCHALK 2001, PARTRIDGE/RICKMAN 2003). COAD et al. (2012) suggest that policy interventions to be efficient need to focus on the causal mechanisms: “a well-placed intervention will target one particular variable to have predictable effects on other variables as the shock propagates throughout the system. Without knowledge on causal relations, a misplaced intervention might have no effect (or even perverse effects) if the variable targeted has no causal effect or unexpected effects) on the other variables”. Curing the symptoms does not necessarily cure the diseases (PEARL 2009). Analysing the dynamics in a regional system of technological activities, this papers draws the following policy conclusions.

Our variable *Grad* is found to foster patenting especially in science-based industries, like CHEM or ELEC. Hence, in these industries policy makers that seek to increase innovation activities should stimulate higher university activities. It can be assumed that fitting university education to the regional research activities is especially relevant. If innovation, and ultimately structural change (e.g. DOSI 1988), is the aim, one should facilitate economic and research activities. This, however, is only true in industries which follow a linear model, like META or CHEM. In the rather engineering-oriented and success-driven industries like TRAN or MACH, policy makers are advised to directly stimulate innovation activities and the transfer of innovations, which then would lead to economic growth. Finally, by stimulating research activities in ELEC, a highly dynamic economic growth process is expected to unfold. Briefly stated, this paper suggests that sound policies should be based on the industry-specific causal relationships among the regional technological activities.

Besides helping to specify industry-specific policy instruments, the results might also have implications for national industrial policies. Which industries are most suitable to pick to become large export-oriented national champions? This strongly depends on how much a policy stimulus, which propagates throughout the whole system of technological activities, translates into longer-term success. For ELEC, the highly dynamic feedback loop between research and economic activities was already highlighted. But also MACH and TRAN seem to be promising candidates. In MACH, innovations lead to subsequent economic growth, which in turn instantaneously contributes to the skilling of the workforce and to increase research activities, while economic growth by itself shows a self-reinforcing dynamic. In TRAN, innovations set incentives to further research, instantaneously translating into economic growth, which in the next period feeds back to more innovations. Policy stimuli for innovations might contribute that these industries to become large export-oriented national champions.

To yield the desired outcomes, (regional) policies need to be designed and implemented on basis of a profound understanding of the underlying mechanisms. This paper might guide further research in uncovering the mechanisms that led to the observed causal effects.

4 Firm Growth and the Spatial Impact of Geolocated External Factors

This chapter is a reprint[25] of:

DUSCHL, M., SCHIMKE, A., BRENNER, T. & LUXEN, D. (2014): Firm Growth and the Spatial Impact of Geolocated External Factors. In: Jahrbücher für Nationalökonomie und Statistik 234: 234-256.

Reprint with kind permission of Lucius & Lucius Verlagsgesellschaft.

Abstract: This paper studies the relationship between firm growth and external factors. Externalities from related economic, public research and higher educational activities are traced back to specific locations in space. The spatial characteristics of their impact are examined within a distance-based, micro-founded approach. Applying quantile regression techniques on a large sample of German firms, we empirically disentangle the complex interplay between internal factors (firm size), external factors and their spatial extent. In particular, we find that the larger firms are, the more diverse are the activities they benefit from and that the geographical meaning of "nearby" depends on the kind of activity.

Keywords: Firm growth, external factors, distances, quantile regression, relatedness, universities, public research, graduates

JEL Classification: C31, D92, L25, R11

[25] Due to better readability, numbering of sections, tables, figures and formulas has been changed. References can be found at the end of the thesis.

4.1 Introduction

Firms are a central unit of analysis in economics. Understanding their activities and development is fundamental to understand economies and economic growth. Nevertheless, the dynamics of firms have been neglected for a long time within economics. Neoclassical theory has focused on static analyses of competition scenarios and aggregated production functions, ignoring the individual developments of firms. Empirical studies have argued that firm growth is a purely random process.

NELSON and WINTER (1974) made firms the central actors of a dynamic perspective on economies, explaining their developments on the basis of individual decision rules and selection mechanisms on the market. Since their path-breaking work evolutionary economics as well as the study of firm growth has advanced. Firm growth has been studied empirically in many works, mainly focussing on the characteristics and the internal determinants of growth. Evolutionary economics has extended its scope tremendously with topics reaching from market competition and industry evolution to game theory and consumption behaviour (see, e.g. WITT 2008).

These topics also comprise the role of knowledge for economic activities (see, e.g. Witt et. al. 2012) and the study of innovation processes as well as their geographic context. Regional innovation systems and clusters are concepts in this field that have received much attention in recent years (see, e.g. ASHEIM/GERTLER 2006 and ASHEIM et.al. 2006). The literature agrees that knowledge and local interaction play an important role for economic growth. This implies that they should be also important determinants for firm growth. Yet another theme of evolutionary economics is the heterogeneity of the economies' constituents. Hence, the question arises which firms are affected by which kinds of activities in their local environment? Unfortunately, empirical studies that examine the influence of local knowledge sources and industrial agglomeration on the growth of firms are rare. Detailed knowledge about these relationships would provide the link between the literature on regional innovation systems and clusters and the original idea by NELSON and WINTER to explain economic growth on the level of firms.

Given this missing link, the aim of this paper is to examine the impact of spatial externalities, which can be traced back to specific locations in space, on the growth of firms. By doing so, new light is shed on the complex interplay between factors internal and external to the firms. This is achieved by calculating firm-specific location variables, which measure the firms' access (in travel time) to related economic, public research and higher educational activities. Using a quantile regression framework, the impact of these activities is compared for employment growth of small, medium and large firms in Germany. Owing to the lack of *a priori* knowledge on the impacts' spatial extent, this paper endogenizes distances. The findings suggest that firm size, an indicator for the firms' necessity and capacity to absorb and implement external knowledge, critically shapes how external activities affect firm growth. Firms benefit from public research and higher educational activities at a geographical scale much smaller than usually assumed as "regional", whereas related economic activities tend to transcend predefined regional boundaries. Only for large firms, the negative effect of technologically narrowly related and geographically nearby activities dominate.

The paper is structured as follows. In the second section (4.2), expectations on the spatial impact of external factors on firm growth are derived from the existing literature. The empirical method is outlined in the third section (4.3), while data issues are discussed in the fourth section (4.4). Section five (4.5) discusses the results and section six (4.6) concludes.

4.2 Literature and hypothesis

According to the resource-based view of firm growth (PENROSE 1959), the ability of generating, combining, and ultimately translating new knowledge into economic opportunities depends on factors internal and external to the firm. Internal factors, like size, age, industry affiliation or R&D, have been repeatedly studied in the economic literature (for an overview see COAD 2009). External factors can be addressed within the conceptual framework of knowledge externalities or spillovers (RASPE/VANOORT 2008). In this paper, we focus on external knowledge as a driver of firm growth, and distinguish between different kinds of knowledge sources. These are naturally more abundant in agglomerations, yet some technological relatedness is necessary for knowledge spillovers to effectively take place (FRENKEN et al. 2007). Here we ask, what is the right degree of relatedness between the firms' activities and the surrounding economic activities? Furthermore, do public research and higher educational activities, with the explicit aim to transfer new knowledge into the economy, unequivocally foster firm growth? By taking a less deterministic perspective, we expect that it depends on both the firms' necessity and capacity to absorb and apply external knowledge, which in turn are related to the firms' internal resources and growth dynamics. Finally, the spatial dimension of the external factors has to be addressed. Although the literature claims that knowledge externalities are spatially bounded, no consensus exists on their spatial extent (DÖRING/SCHNELLENBACH 2006).

By summarising the literature, we derive the hypotheses that aim to disentangle the complex interrelation between firm growth, external factors and internal characteristics. Regarding the spatial dimension of the external factors, we will conclude that this question has to be determined empirically.

4.2.1 Firm growth and its growth related external factors

The impact of related economic activities

Already MARSHALL (1890) stated that firms might perform better when located within agglomerations. In respect to the economic geography literature, traditionally two kinds of externalities are distinguished. Whilst positive effects of localization economies occur through specialization of similar industries, the positive effects of urbanization economies arise from the variety of different industries. After many decades of intensive research, the literature on regional agglomeration remains rather indecisive about the real effect of specialization versus diversification at the regional level (BEAUDRY/SCHIFFAUEROVA 2009).

Reasons for this inconclusiveness are manifold, here we highlight what seems to us most relevant.

Studies following this stream of literature mostly focus on regional economies. However, the aggregate regional outcome is composed of the foundation of new firms and the growth and survival of existing firms, which might be affected differently by agglomeration economies. Whilst a positive effect on start-ups (e.g. SORENSEN/AUDIA 2000) and survival (e.g. RENSKI 2011) is well documented, "the main gap in our empirical understanding concerns the effect of localization economies on firm performance, which some may even consider the key question in economic geography at large" (FRENKEN et al. 2011). Mostly analysing the agglomeration-productivity (e.g. RIGBY/ESSLETZBICHLER 2002, BALDWIN et al. 2008, DRUCKER/FESER 2012) or agglomeration-innovation relationship (e.g. FELDMAN 2000, BEUGELSDIJK 2007), only few exceptions exist that link agglomeration to growth at the micro-level of firms (e.g. AUDRETSCH/DOHSE 2007). By referring to the resource-based view of firm performance, in which KOGUT and ZANDER (1992) emphasize the critical role of knowledge (RIGBY/BROWN 2013), several mechanisms of agglomeration economies can be expected to have a positive impact on firm growth. Knowledge is exchanged and diffuses within an agglomeration among competing and cooperating firms, either without any direct interaction through constant mutual monitoring (MALMBERG/MASKELL 2002) or as a result of direct interactions and learning processes in formal and particularly informal social networks (SINGH 2005). Furthermore, the mobility of individuals (BRESCHI/LISSONI 2009) and the exchange of intermediate goods (DÖRING/SCHNELLENBACH 2006) cause specialized knowledge embodied in human and physical capital to circulate. The more economic activities agglomerate around a firm, the more readily available growth-relevant knowledge should be. Hence, we simply expect:

> **Hypothesis 1:** *Agglomerations (of other economic activities) increase the firms' growth prospects.*

However, recent studies show that not all firms benefit in the same way from co-location (RIGBY/BROWN 2013, MCCANN/FOLTA 2011, KNOBEN et al. 2010). Besides the kind of activities a firm carries out (see BEADURY/SWANN 2009 or DUSCHL et al. 2014b for the industry-specific impact of agglomerations), also its internal capabilities to capture different forms of externalities matter. From a resource-based view, it is often argued that firm size plays a critical role. Furthermore, the organisational structure of firms changes tremendously during their life-time and growth (WITT 2000). VANOORT et al. (2012) report that the relationship between firm performance and agglomeration is strongest for medium-sized firms. Neither small firms nor large firms seem to be fully able to internalize externally available resources or knowledge. The former tend to lack absorptive capacities (COHEN/LEVINTHAL 1990), whereas the latter reduce their openness towards the local environment, as they become organizationally complex and hence inflexible (VANOORT et al. 2012). These results are supported by RIGBY and BROWN (2013), who find in their analysis of manufacturing plant performance in Canada that small plants do not benefit from the local density of upstream suppliers.

The issue of the moderating effect of firm size becomes more complex, as it additionally depends on the kind and composition of activities surrounding a firm. On the one hand, NOOTEBOOM (2000) presumes for knowledge spillovers to take place that external knowledge should be neither too similar nor too different to the own knowledge base. If it is too similar, no new contribution will be made, and if it is too different, the cognitive distance, which constitutes a barrier for effective communication, will cause problems in absorption and implementation (ERIKSSON 2011). The literature on related variety (e.g. FRENKEN et al. 2007) concludes that externalities like knowledge spillovers arise from technologically related industries. On the other hand, it is a well-known empirical regularity that firms diversify into new fields of activity with increased size (BOTTAZZI/SECCHI 2006). Hence, the right degree of relatedness depends on firm size, as firms can realize potential benefits from locating in agglomerations "only to the extent that they are capable of using knowledge from other, co-located firms in combination with their own knowledge assets to create value" (KNOBEN et al. 2011: 8). Smaller firms, which are more specialized in their activities, are expected to benefit most from a more narrowly defined relatedness. Activities that are too different do not match with the internal knowledge base (NOOTEBOOM 2000). As the size of the firms increases, their scope of activities broadens, and the meaning of relatedness expands. KNOBEN et al. (2011) show empirically that medium sized firms in general benefit from being located in proximity to rather dissimilar, yet related firms. Other activities that for small firms are complementary to their own knowledge base tend to become increasingly internalized by larger firms. Ultimately, such activities, which cannot provide new pieces of knowledge, are reduced to mere sources of rivalry, leading to adverse effects of agglomeration.

Briefly stated, we investigate the theoretical interaction between firm size and the degree of relatedness of agglomerated economic activities regarding their impact on firm growth. Conditioning on firm size, we hypothesize:

> **Hypothesis 1a:** *With increasing firm size, firms benefit most from more and more broadly related external activities.*
>
> **Hypothesis 1b:** *For larger firms, the presence of closely related external activities may even hamper their growth prospects, as these activities become a source of rivalry instead of providing complementary knowledge.*

The impact of public research activities

The explanation of why firms might benefit from industrial clusters has more recently shifted to phenomena like innovation, learning and knowledge spillovers (MALMBERG et al. 2000). Growth-relevant knowledge is generated not only by economic activities of competing and cooperating firms, but also by explicit research activities in universities or public research institutes. In this case, the literature that focus on "regions rich in knowledge resources" (AUDRETSCH/DOHSE 2007) or "knowledge-intensive environments" (RASPE/VANOORT 2011) is less ambiguous. Many studies show that public research in universities and research institutes support the innovativeness of nearby firms (e.g. JAFFE 1989, ANSELIN et al. 1997) and ultimately their growth performance (e.g. AUDRETSCH/LEHMANN 2005, CASSIA et al. 2009 or RASPE/VANOORT 2011). Scientific

knowledge is transferred by various mechanisms, for example directly by public-private research collaborations, or indirectly by "scientific publications, seminars, workshops and informal relationships" (FRITSCH/SLAVTCHEV 2007). Usually, these studies do not condition on firm size. Nevertheless, two opposing arguments can be found in the literature: One based on the absorptive capacities of firms, the other based on the dependence of firms on external knowledge inputs. On the one hand, BRENNER and SCHLUMP (2013) argue that larger firms, which are also more likely to conduct internal R&D, receive more often and benefit more from knowledge generated by public research activities. Larger firms benefit more because of their higher absorptive capacities (COHEN/LEVINTHAL 1990), or in the words of LAURSEN and SALTER (2004): "The argument contained in previous research is that larger firms are more likely to have the capability to exploit external knowledge sources and to manage interactions with universities." On the other hand, small firms depend more on research conducted outside the firm: "We find substantial evidence that corporate R&D is a relatively more important source for generating innovations in large firms, while spillovers from university research laboratories are more important in producing innovative activity in small firms" (ACS et.al. 1994). While large firms are able to conduct substantial research themselves and are able to interact with public research over large distances, small firms depend more on nearby public research (TORRE 2008). It is unclear which of these two mechanisms dominates, so that we formulate two contradicting hypotheses:

> **Hypothesis 2a:** *Larger firms benefit from public research activities, whereas no growth effects exist for small ones.*
>
> **Hypothesis 2b:** *Small firms benefit from nearby public research activities, whereas the proximity of public research activities has no effect on larger firms.*

The impact of higher educational activities

Studies addressing the impact of universities usually consider the presence (CASSIA et al. 2009; RASPE/VAN Oort 2011) or proximity to the next university (AUDRETSCH/LEHMANN 2005). Not much is known about the impact of the different mechanisms of universities on firm growth. Besides academic research, which in the previous section is subsumed under public research activities, universities' function is to perform education (FRITSCH/SLAVTCHEV 2007). Both functions differ substantially in their underlying mechanisms. For instance, qualified graduates bring up-to-date knowledge into firms (BRENNER/SCHLUMP 2013). Besides, the mere availability of graduates is a prerequisite for firms that want to expand in their number employees, as graduates contribute to the pool of available and highly qualified workers. The faster firms grow the more workers they need to hire. This holds particularly true for larger firms, as the same growth rate implies a different absolute meaning. Hence, we expect:

> **Hypothesis 3:** *The more the firms grow in quantitative terms (i.e., larger firms and high-growth firms), the more relevant higher educational activities become.*

4.2.2 The spatial dimension of external growth factors

The various sources of externalities discussed above can all be traced back to specific locations in space. Related economic activities take place within co-located firms, new scientific knowledge is published by scientists who work in public research institutes or universities, and from the latter students become graduated who might decide to work in one of the firms. These activities are unequally distributed across space, hence the locations of firms relational to these sources matter (ANDERSSON/KARLSSON 2007). Unfortunately, it is virtually impossible to measure the real individual impact of each single source of externality on the growth of each firm. For example, the assumed knowledge flows are intangible (KOO 2005), and data on the entire collaboration network or on the migration patterns of graduates is lacking. Yet the potential strength of the externalities can still be analysed by assessing their expected impact on the growth prospects of firms, to which they are accessible. KARLSSON and MANDUCHI (2001) argue that the accessibility approach, based on early ideas of WEIBULL (1980), makes the general concept of geographical proximity operational in the first place. A high accessibility means a high potential for interaction, increasing the probability of knowledge spillovers, labour flows or other mechanisms alike. Therefore, we calculate the potential strength of externalities, which is set equal to the accessibility of the firms' locations, as their revealed impact on firm growth.

The discussion (implicitly) assumes that agglomeration economies are bounded in space (FRENKEN et al. 2011). This holds especially true for knowledge spillovers. The inherent properties of the nature of knowledge, like the degree of tacitness or complexity (Sorensen et al. 2006), increase the cost of transmitting knowledge over longer distances. Transferring complex, that means often unstructured, but economically valuable knowledge, demands personal contacts. Because this kind of knowledge is mostly embedded in people, knowledge spillovers are a function of people's mobility and interactions (ANDERSSON/KARLSSON 2007). In lieu of recent improvements in ICT (SONN/STORPER 2008), there are still strong empirical findings that social interactions decrease with geographical distance (e.g. HOEKMAN et al. 2010 on the research collaboration between firms).

Despite the general acceptance of the distance-sensitivity of agglomeration economies, empirical studies lack a consensus on their spatial extent (DÖRING/SCHNELLENBACH 2006). Distances as diverse as 10 km (BALDWIN et al. 2008), 120 km (ANSELIN et al. 1997) or 300 km (BOTTAZZI/PERI 2003) are reported. FRENKEN et al. (2011) concludes that the findings depend on the type of mechanisms and external activities investigated as well as on the composition of the sample firms. CAINELLI and LUPI (2008), for instance, show that the effect of similar activities is greatest at very small distances, whereas the effect of variety becomes positive only for larger distances between 10 and 30 kilometres. Considering institutional and geographical proximity simultaneously, ERIKSSON (2011) argues that greater distances between plants require that their technological knowledge base is more similar, so that differences in the firms' local institutional contexts, another cause of communication problems, can be overcome. Regarding public research activities, the spatial extent of externalities depends amongst others on the spatial distribution and effectiveness of formal and informal science-industry collaborations. As this kind of knowledge transfer often requires social interactions and face-to-face contacts, it might be

a narrow local phenomenon. Regarding graduates, we are aware that they are among the most mobile group in society (MOHR 2002). Nonetheless, a study for Germany shows that 70% of them stay in their region of education ten years after graduation (BRENNER/SCHLUMP 2013). The likelihood of outmigration decreases if growing firms are able to provide new job opportunities (BUSCH 2007). Because the literature does not provide a conclusive picture on the spatial extent of the various externalities, this issue has to be addressed empirically.

4.3 Empirical method

Distance-based methods, which take into account the bilateral distances between the locations of all activities, are suited to operationalize the effect of externalities as a matter of their accessibility (ANDERSSON/GRASJÖ 2009).[26] A firm-specific location variable *X* for each time period *t* can be calculated by summing up all corresponding external activities *x* (here, related economic, public research or higher educational activities), which are discounted by a distance decay function $f(d_{im})$, where d_{im} indicates the distance between the location of firm *i* and the location of activity *m*:

$$X_{i,t} = \sum_m f(d_{im})\, x_{m,t} \tag{20}$$

This approach requires that two choices are made: the specification of $f(d_{im})$ and the way how distances are measured. Regarding the former, many approaches exist. Mostly, simple linear (e.g. AUDRETSCH/LEHMAN 2005) or exponential decay functions (e.g. DRUCKER/FESER 2012) are used. The literature on commuting behaviour (e.g. ANDERSSON/KARLSSON 2007) shows that the negative distance sensitivity is not linear in space, but varies between different geographical scales: within a narrow local context, interactions are primarily governed by randomness, because they can take place at short notice (THORSEN et al. 1999). Thus, within agglomerations interactions are only marginally affected by distance. At some threshold distance, however, the minimal cost principle predominates and consequently, the frequency and contribution of growth-relevant economic interactions become highly distance-sensitive and may decrease rapidly. DEVRIES et al. (2009) model these spatial interactions by using a sigmoidal log-logistic decay function. DUSCHL et al. (2014a) find that this rather flexible S-shaped curve often converges to a step-wise decay function, which is similar to distance bands, i.e. summing up the activities within specific radii around the firms (e.g. ROSENTHAL/STRANGE 2003). Hence, this paper employs such a binary distance function $f(d_{im})$, which becomes one if the value is below the threshold distance d_{im}. As the literature does not reveal much about an adequate threshold distance (see section 4.2.2), d_{im} is endogenized. In doing so, not

[26] Distance-based methods are an alternative to methods which rely on regional boundaries. The latter try to explain firm performance by means of characteristics of the region the firm is located in. However, results are affected by the arbitrariness of regional boundaries and moreover by the chosen level of aggregation, also known as the Modifiable Areal Unit Problem (MAUP) (OPENSHAW 1984). By varying the spatial scale of analysis, BUERGER et al. (2010) show empirically that the MAUP is highly relevant for agglomeration economies.

only information regarding the magnitude of the impact of external factors is obtained, but also regarding the spatial extent.

Furthermore, results may be sensitive to the way how distances are measured. Agglomeration economies are assumed to arise from low transportation costs or the convenience of face-to-face contacts. Thus, the firms' access to external factors not only depends on the location pattern of the corresponding activities, but also on the physical infrastructure (ANDERSSON/KARLSSON 2007). Whereas physical distance is still the frame, in which relevant interactions occur (RODRIGUEZ-POSE 2011), it is driving distance or travel time that is directly related to the frequency of interactions (ANDERSSON/GRASJÖ 2009). The vast majority of distance-based investigations uses orthodromic distances (e.g. km or miles). One of the few exceptions is the work of AUDRETSCH and LEHMANN (2005), where the growth of firms is analysed with respect to the firms' driving distance to the their closest university. Owing to high computational costs of route planning, studies in which driving distances are computed to thousands of locations are still very rare.

The firm-specific location variables $X_{i,t}$, which result from equation (21), can then be included in a simple linear model (for spatial econometric issues regarding this kind of spatially discounted variables we refer to ANDERSSON/GRASJÖ 2009).[27] Because of the flexible specification of $f(d_{im})$, a normalization procedure is first applied to make the corresponding regression coefficients comparable:

$$\bar{X}_{i,t} = \mu X_{i,t} \tag{21}$$

where

$$\mu = \frac{\sum_m x_{m,t}}{\sum_i \sum_m f(d_{im})\, x_{m,t}/N} \tag{22}$$

with N denoting the number of firms. This normalization allows for an interpretation of the coefficients as the impact of an additional activity at a distance with an average impact on firm growth. The regression equation reads

$$g_{i,t} = \alpha + \sum_j \beta_j \bar{X}_{i,t,j} + \sum_k \beta_k U_{i,t,k} + \sum_l \beta_l T_{i,t,l} + \varepsilon_{i,t} \tag{23}$$

with α and β representing the coefficients to be estimated, U the various control variables, and T the dummies for the years t of observation. The error term is denoted by $\varepsilon_{i,t}$.

Furthermore, we assume that the variables which are related to public research and higher educational activities have only a positive impact: firms either benefit from these activities or they don't. Considering this, a negative coefficient simply would capture some structural effects, which are rather related to general characteristics of areas which host universities and public research institutes and less to the mechanisms of growth-relevant knowledge

[27] Performing an extensive Monte Carlo analysis, these authors confirm that this approach captures substantive spatial dependence in the dependent variable and accounts for both local and global spillovers.

transfers. Therefore, the corresponding coefficients are submitted to non-negative constraints.[28] Furthermore, we assume that a driving distance of two hours represent a reasonable upper range even for larger metropolitan areas, so that d_{im} is constrained at maximum 120 minutes.

The model in equation (23) is estimated by using quantile regression techniques. Quantile regression seems to be the adequate technique on the grounds of the following reasons. Firstly, the stochastic analysis of section 4.4.1 reveals that firm growth rates are not normally distributed, but show fat tails. In contrast to OLS, quantile regression is free from any distributional assumption in the error term (BUCHINSKY 1998). Secondly, it is not sensitive to outliers on the dependent variable. This is especially relevant here due to the high frequency of extreme growth events, which also might stem from data problems. Thirdly, the specific conditional quantiles of strongly expanding ($\theta_{0.75}$) and declining ($\theta_{0.25}$) firms can be analysed in addition to the median growing firm ($\theta_{0.5}$). These firms significantly contribute to the overall economic development and hence are of interest in their own right (COAD/RAO 2008). Our intuition is that high growth events, a dominant feature of firm growth, rely differently on internal as well as external factors. Focusing exclusively on the average firm obscures these relationships (COAD/RAO 2008).

Technical details are described amongst others in KOENKER (2005).[29] Here we only point out that, likewise to OLS, the coefficient estimates can be interpreted as partial derivatives, meaning the impact of a one-unit change of an independent variable on the firms' growth rate at the θ_{th} quantile holding all other variables fixed. To recall, the binary distance functions are endogenized and the threshold distances d_{im} are identified by minimizing the log-likelihood value of the model. One should note that this optimization is done for each conditional quantile θ separately.

4.4 Data

The BvD-Amadeus database encompasses firms from both manufacturing and service industries. Regarding the firms' location, the addresses of the headquarters are disclosed. Operational and strategic decisions are often made within this organizational unit. Although R&D intensive foreign direct investments have recently become more important (UNCTAD 2005), many of the firms' R&D activities still remain located close to the headquarters (GUIMÓN 2009). Even for multinational enterprises a home bias for innovation activities is evident (COHEN et al. 2009, DUNNING/LUNDAN 2009). Therefore, we follow BEAUDRY and SWANN (2009) in assuming that the regional environment of the firms' headquarters is most decisive in affecting their growth prospects.

This rationale, of course, breaks down for very large firms, which tend to be less focused on their headquarters, but disperse activities in many increasingly independent

[28] Indeed, the introduction of constraints was necessary, as for certain distances the coefficients turned out to be negative.

[29] Standard errors are estimated using bootstrapping techniques. In line with KOENKER and HALLOCK (2001), we only detected negligible small discrepancies between various available methods.

establishments across the country and even beyond. Hence, the analysis is restricted to firms with no more than an annual average of 1000 employees.[30] Also very small firms with less than ten employees, which growth processes are known to be rather erratic, are omitted (COAD 2009). Because in section 4.2.1 it is argued that the size of the firm determines its growth logic and the impact of externalities, the remaining firms are split into the three different size classes small [10-50), medium [50-250) and large [250-1000) according to the European Commission (2003). The composition of the subsamples is presented in Table 9.

Table 9: Sample composition and distributional properties of the growth rates

	Sample composition		AEP estimates	
	N firms	N $g_{i,t}$	b_l	b_r
Small [10-50)	33062	78653	0.864	0.381
Medium [50-250)	25199	77147	0.470	0.608
Large [250-1000)	7423	26011	0.424	0.473

4.4.1 Dependent variable

Firm growth, which is a multi-dimensional process, can be addressed by a variety of measures like employees, turnover, sales or productivity. No universally best indicator exists and the pros and cons of the different measures are discussed in the literature at length (see COAD 2009 for an overview). RASPE and VANOORT (2008) argue that the employment measure is most adequate from the resource-based view of the firm, because employees represent a firm's most important asset. Besides, employment growth primarily should concern regional policy makers and thus can be regarded as a suitable measure to assess the impact of spatial externalities. Growth rates are calculated by taking the difference of the natural logarithms of the employees *Empl* of firm *i* between two successive years *t*:

$$g_{i,t} = log(Empl_{i,t+1}) - log(Empl_{i,t}) \tag{24}$$

Confronted with an unbalanced panel from 2004 to 2010, yearly growth rates are pooled together. Meanwhile, it counts as a stylized fact that the distribution of the growth rates shows an exponential tent-like shape, similar to the one of the Laplace distribution, implying that the tails are much fatter than the normal distribution would suggest. More recently, an asymmetric shape is found with extreme negative growth events being particularly predominant (e.g. BOTTAZZI et al. 2011). Fitting the flexible five-parameter Asymmetric Exponential Power (AEP) distribution (BOTTAZZI/SECCHI 2011) to our data, we

[30] We follow COAD (2007) and use for all size thresholds the annual average to avoid a regression fallacy originating from the difficulty of sorting growing entities into size classes (FRIEDMAN 1992).

find that the shape parameter *b* takes values clearly smaller than one, which would be the case of a Laplacian shape (see Table 9). For medium and large firms, especially negative fat tails are much more pronounced, as indicated by a smaller b_l. Only for small firms, the asymmetry is reversed. Firms of this subsample are more likely to realize large positive growth jumps, whereas strong decline events are less often observed, maybe due to the fact that exits are not considered in the analysis. Hidden by an adverse shock, smaller firms tend to be prone to disappear from the market altogether (DUNNE et al. 1988). Exit as well as entry events are not comprehended as growth events. We are not able to include them in the analysis for two reasons: the growth rate as defined in equation (24) is not defined for such events and our data only contains information on firms that exist at the end of the observation period. This implies that inference can only be made on the growth of surviving firms. This causes a bias in our data. The exit of a firm represents an extreme decline event in terms of firm growth. Hence, these extreme decline events are underrepresented in our sample. Exits of large firms are rare, but in the case of small firms we have to be aware of the fact that we analyse a biased sample. The most extreme decline events are missing. This impacts mainly the results for the quantile $\theta_{0.25}$ and slightly the results for the quantile $\theta_{0.5}$ in the case of small firms. The results for the quantile $\theta_{0.75}$ should not be influenced. Hence, we interpret especially quantile $\theta_{0.25}$ regressions for small firms with care. If these results do not fit the overall picture, we do not trust them.

A stochastic analysis shows that standard OLS is expected to perform poorly, because errors will not be distributed normally (MAASOUMI et al. 2007). Also the discussion on high-growth firms becomes apparent: most firms do not grow (or only slightly), whilst a small, but non-negligible part of firms experiences very rapid growth or decline. As argued in section 4.3, quantile regression techniques are more adequate for our purpose.

4.4.2 Independent variables

Firms' potential to benefit from externalities is specific to characteristics of the firms as well as of the corresponding regions (BEUGELSDIJK 2007). Therefore, the independent variables consist of three different kinds (see Table 10): First, we control for relevant demographic properties of the firms. Second, we include measures of the general environment of the region the firm is located in. Third, the focus of this paper lies on firm-specific location variables reflecting related economic, educational and scientific activities.

Control variables: demographic and regional variables

Building upon the literature on firm growth, which mostly extends a Gibrat-like growth regression (COAD 2009), we control for four demographic variables: the logarithm of employees (*SIZE*), years passed since founding date (*AGE*), a dummy indicating whether or not it is a subsidiary firm (*d_SUBS*), and a dummy assigning one to firms from knowledge intensive industries (*d_KNOW*).[31] In the literature of industrial dynamics, it

[31] GEHRKE et al. (2010) provide a classification for knowledge intensive industries at the 3-digit level.

counts as a stylized fact that firm growth is negatively related to both size and age (EVANS 1987). Firms that are formally identified as a subsidiary are by definition more or less dependent on their mother institution, with an unknown growth impact. The knowledge intensity dummy should proxy for internal research activities as well as the absorptive capacity, which are expected to increase with the knowledge intensity of the activities a firm is engaged with (KOO 2005).

In addition to the firm-specific demographic variables, two variables are chosen to control for the general regional environment. Population density (*POP*) measures urbanization economies *per se*, which are rather independent from the surrounding industrial structure (BUERGER et al. 2012). Its impact might be either positive or negative: on the one hand, *POP* is known to foster innovation activities and ultimately growth (FELDMAN 2000), but on the other hand, negative agglomeration economies, mostly due to higher wages, could prevent firms of hiring employees. Additionally, FRITSCH and SLAVTCHEV (2011) argue that this variable catches up other types of unobserved region-specific influences. Unemployment rate (*UR*) reflects the vitality of the regions' socio-economic conditions. In the special case of Germany it also accounts for structural differences along the east-west and north-south divide. Data for both variables is obtained from the German Federal Statistical Office (destatis). *POP* is most meaningfully measured at district level, whereas *UR* at the level of functionally defined regional labour markets (ECKEY et al. 2006).

Moreover, also the absolute location within Germany might influence the potential magnitude of externalities. Cross-border effects cannot be considered, discriminating firms located close to the border. Owing to historical reasons, two dummies are constructed: one for the location in border regions with the New Member States of the EU (*d_EUnew*) and one for all other border regions (*d_other*). Finally, general macroeconomic conditions, foremost the global recession 2008-10, might systematically affect the firms' growth prospects. Therefore, year dummies are included.

Firm-specific location variables: related employment, graduates and publications

In contrast to the regional control variables, which account for a rather diffuse socio-economic environment (or "social filter", as denominated by RODRÍGUEZ-POSE/CRESCENZI 2008), other activities as sources of externalities can be traced back to concrete localizations in space: firms compete, cooperate, and learn from each other, new scientific knowledge is published at universities and research institutes, and students get hired after graduation at often nearby universities. The firms' accessibility to these activities and their impact on the growth prospects might be mediated by the general regional environment. For example, the regional environment influences the effectiveness of university-industry knowledge-related linkages (VARGA/PARAG 2009), and graduates tend to prefer a diverse and open urban atmosphere (FLORIDA 2005). Therefore, the firm-specific location variables, which are distance-based and micro-founded measures of the sources of externalities, are complemented by regional level control variables, as presented above.

Related economic activities can be approximated by the accessibility of other employees in the firms' industries. The issue of relatedness is tackled by a hierarchical approach: all employees which belong to the same 2-, 3- or 4-digit industry and are located within a

certain distance are taken into account. In case of 3-digit industries, the number of employees of the same 4-digit level is excluded to avoid double counting. Analogously, employees of the 2-digit industry are adjusted by subtracting the same 3-digit employees. The Federal Employment Agency (BA) provides data of the location of industry-specific employment. Three firm-specific agglomeration variables result (AGGL_2, AGGL_3 and AGGL_4, respectively).

Public research activities are reflected in the number of scientific publications. Although firms engage in publishing, the vast majority originates from universities and research institutes. Hence, publications tend to measure public research activities. Data on publications was collected from the ISI Web of Science and can be geolocated on basis of the authors' addresses. We anticipate that the fields of study most relevant are science, technology, engineering, and mathematics, commonly known as the STEM fields. By counting only those publishing activities in STEM that are accessible to the firms, a firm-specific publication variable can be calculated (PUBL).

Higher educational activities can be measured by the number of graduates. Applying the same argument, graduates are restricted to the STEM fields. Data on graduates was taken from destatis and encompasses universities in a narrower sense as well as universities of applied science, that is, all graduates with a technical, diploma, bachelor and master degree. Again, by counting those higher educational activities that are accessible to the firms, a firm-specific graduate variable can be calculated (GRAD).

The next section discusses how distances between the firms and related employees, graduates and publications are measured.

Table 10: Overview of independent variables and data sources

Variable type	Variable name	Description	Data source
Demographic variables	SIZE	Logarithm of employees	BvD Amadeus
	AGE	Years since founding date	BvD Amadeus
	d_SUBS	Subsidiary status (dummy)	BvD Amadeus
	d_KNOW	Sectoral affiliation to a knowledge intensive industry (dummy)	BvD Amadeus
Regional variables	POP	Population density of the firms' district	destatis
	UR	Unemployment rate of the firms' labour market region (ECKEY et al. 2006)	destatis
	d_EUnew	Firms' district shares border with one of the new EU member states (dummy)	
	d_other	Firms' district shares border with one of the other countries (dummy)	
Firm-specific location variables	EMPL_2	Accessible employees of the same 2-digit industry	BA
	EMPL_3	Accessible employees of the same 3-digit industry	BA
	EMPL_4	Accessible employees of the same 4-digit industry	BA
	PUBL	Accessible publications in STEM	Web of Science
	GRAD	Accessible graduates in STEM	destatis

4.4.3 Driving distances

The calculation of driving distances is done by exploiting results from graph theory: the road network is modelled as a directed graph with travel time metric as edge weights.[32] KNOPP et al. (2007) introduced an algorithm to compute large-scale distance matrices without naively computing a quadratic number of distances. The small search spaces of a speedup technique to DIJKSTRAS seminal algorithm are precomputed and intersected to produce the matrix. This method only needs a linear number of shortest computations and therefore is several orders of magnitude faster than the naive algorithm. In addition, an algorithm of GEISBERGER et al. (2008) is used that exploits the natural hierarchy of road networks, called Contraction Hierarchies (CH). The method preprocesses a road network and produces a linear sized amount of auxiliary data that is used to speed up any subsequent queries. CH have the benefit of small search space, i.e. a query has to look at only a few hundred nodes in the graph. Combined with the previous algorithm we can compute distance matrices of 10,000 by 10,000 nodes within a matter of several seconds.

To investigate externalities on firm growth from a realistic spatial perspective, the aggregation level of the data should be as low as possible. For Germany, this is given by municipalities, currently with a number of 11249 and an average size of 31.6 km^2, which DURANTON and OVERMAN (2005) would already classify as micro data. Thus, the locations of the firms and all other activities are approximated by using the geocentroids of the corresponding municipalities. In doing so, a new issue arises. If one firm is located in the same municipality as, for instance, a university, it would be inappropriate to set the distance to zero. As a substitute, the existence of a general intra-municipality friction can be assumed. To obtain its value, a random sample of 1000 pairs of firms' address locations is drawn, each belonging to the same municipality, and all bilateral distances are measured. The mean of all intra-municipal travel times is 5.01 minutes.

4.5 Results

4.5.1 Control variables

The estimated coefficients of the control variables are largely in line with the current literature on firm growth and industrial dynamics (see Table 11). Yearly growth rates generally tend to correlate negatively with *SIZE* and *AGE*, even viewed within narrower size classes. The revealed relationships become even more pronounced for highly growing firms at the conditional quantile $\theta_{0.75}$, implying that growth jumps are less likely the larger or older a firm becomes. *SIZE* is positively related to growth only at $\theta_{0.25}$ for small firms, the case that might be influenced by the exclusion of exits. Due to the possible bias in this case we do not trust the result and ignore it. Regarding *AGE*, we observe an clear overall picture. However, younger firms have higher exit rates, so that the bias leads to more

[32] Data on the German road network was taken from the OpenStreetMap project as of July 22nd, 2011 and consists of 8,226,112 nodes and 15,501,574 edges.

negative estimates than without such a bias. Hence, the negative relationship of AGE with firm growth at $\theta_{0.25}$ for small firms might be a consequence of the data bias. This confirms the above statement that especially extreme growth events are negatively related to the size of firms, while we are not able to make a concluding statement for the extreme decline events.

Table 11: Regression results for demographic and regional control variables

	small [5, 50)			medium [50, 250)			large [250, 1000)		
	$\theta_{0.25}$	$\theta_{0.50}$	$\theta_{0.75}$	$\theta_{0.25}$	$\theta_{0.50}$	$\theta_{0.75}$	$\theta_{0.25}$	$\theta_{0.50}$	$\theta_{0.75}$
SIZE	.0270 ***	-.0760 ***	-.1112 ***	-.0181 ***	-.0209 ***	-.0454 ***	-.0286 ***	-.0185 ***	-.0475 ***
AGE	-.0001 ***	-.0001 ***	-.0007 ***	-.0001 ***	-.0002 ***	-.0005 ***	.0000	-.0001 ***	-.0003 ***
d_SUBS	.0097 ***	.0023 ***	.0183 ***	-.0003	-.0031 ***	-.0070 ***	-.0005	-.0083 ***	-.0115 ***
d_KNOW	.0017 '	.0015 ***	.0181 ***	.0044 ***	.0051 ***	.0089 ***	.0025 *	.0037 ***	.0048 **
Pop	-.0000 **	-.0000	.0000	-.0000 ***	-.0000 ***	-.0000	-.0000 **	-.0000 *	.0000
UR	.0007	-.0024 '	-.0448 ***	.0020	-.0090 *	-.0275 **	-.0098	-.0065	-.0313 *
d_EUnew	-.0016	-.0002	-.0018	-.0004	.0010	.0027	-.0012	.0028 '	.0049 '
d_other	-.0014	.0002	.0048 **	.0001	.0015 **	.0020	-.0004	.0012	.0008

p-values: ‘ < 0.1, * < 0.05, ** < 0.01, *** < 0.001

Two further demographic control variables, *d_SUBS* and *d_KNOW*, are included. Interestingly, the status as subsidiary is beneficial for small firms, but hampers the growth prospects of medium and large firms. This result suggests that the formal reliance on a mother institution might facilitate access to growth-relevant resources and at the same time might provide a shelter against adverse events, because also negative growth events at $\theta_{0.25}$ become less likely. With increasing firm size, this dependence turns into a constraint, whereby especially high growth events at $\theta_{0.75}$ are affected. The results for the affiliation to a knowledge intensive industry are quite straightforward. These firms possess greater growth prospects, in particular at higher conditional quantiles.

The second group of control variables relate to the general socio-economic environment of the firms' region. *POP* tends to come along with lower average growth rates. This finding confirms other studies of Germany (e.g. FORNAHL/OTTO 2008) and suggests that cost aspects due to congestion in densely populated places dominate when agglomeration effects of related employees and of proximate publications and graduates are directly taken into account. Put differently, higher wages prevent firms of hiring employees. During phases of high growth, for which no significant relationships are found, the price competition seems to become less relevant. As expected, *UR* shows a slightly negative correlation with firm growth. Foremost high growth events are hampered in structurally less-favoured regional economies.

Finally, the two border-region dummies, which control for a potential underestimation of agglomeration economies across national boundaries, are merely significant, but if so, they show up to be positively correlated with firm growth. Not surprisingly, growth rates were significantly reduced during the years of macroeconomic recession, particularly 2009 (estimates for the yearly dummies are not reported).

4.5.2 Impact of firm-specific location variables

Having controlled for several demographic and region-specific variables, the impact of the firm-specific location variables can be discussed (see Table 12).

Related employees

In general, the agglomeration of related employees increases the firms' growth prospects. Estimates tend to be positive and in the majority of cases significant. This supports *hypothesis 1*. Furthermore, the estimates are larger and more often significant at $\theta_{0.25}$ and $\theta_{0.75}$, confirming previous literature insofar as agglomeration effects are more relevant for fast growing firms (e.g. FORNAHL/OTTO 2008).

Nonetheless, differences among the size groups and the degree of relatedness exist. Small firms benefit most from being located in proximity to employees of the same 4-digit industries, medium sized firms to employees of the same 3-digit industries, and finally large firms to employees of the same 2-digit industries. This clearly underlines the argument that the relevant degree of relatedness is conditional on firm size; *hypothesis 1a* can be confirmed – the larger the firms, the more diverse the related activities they benefit from. However, in contrast to the expectation that small firms lack the necessary absorptive capacity to benefit from activities that are rather broadly related, still a significantly positive sign is observed for the 2-digit industries. As the effect becomes more pronounced for $\theta_{0.25}$, one could conclude that the main role of these rather diverse activities is to reduce the risk of adverse negative growth shocks (by the portfolio effect of diversity); high growth events are not fostered neither for small nor medium sized firms.

As predicted in *hypothesis 1b*, large firms are hampered by activities of the same 4-digit industries. An agglomeration of very similar activities cannot provide complementary knowledge, but rather tends to become a source of rivalry. This does not hold at $\theta_{0.25}$, for which a significant positive impact is found. An explanation can be provided by taking into account the spatial dimension of the externalities.

To recall, distances are endogenized in this paper. Table 13 contains the optimal distances of the corresponding estimates, printed in bold if these are significant. In 13 out 21 cases, in which the impact of related employees is significantly positive, the optimal distance threshold surpasses 30 minutes, and in eleven cases even 90 minutes. This finding reveals that the impact of externalities from related employees tend to have a larger extent, which sometimes even transcend traditional regional boundaries, and which could not been captured by methods purely relying on regions. Interestingly, employees of the same 4-

digit industry are most beneficial at larger distances. Referring to ERIKSON (2011), the negative effects of technologically too narrowly related activities can be traded off by variety in the local institutional knowledge, manifested in sharing the same heuristics and routines, which increases by geographical distance. In the two cases of large firms, in which a negative effect at the 4-digit level is found, the corresponding distances are 13 and 14 minutes, confirming the same argument: if both the cognitive and geographic distance is too close, the negative effect dominates. In contrast hereto, large firms, at $\theta_{0.25}$, are able to benefit from activities of the same 4-digit industry which are located within 51 minutes. These shrinking firms, for which the opportunity reducing effect of rivalry seems to become less relevant compared to expanding firms, are able to successfully trade off the technological overlap by other forms of variety that increase with geographical distance.

Publications and graduates

Regarding the variables *PUBL* and *GRAD*, the issue of multicollinearity has to be addressed first. Graduates originate from universities as does a major share of publications; the larger the universities, the higher the output in both measures. This makes the variables strongly correlated. Although the optimal distances d_{im} for both mechanisms have been determined endogenously, it is not possible to guarantee that their growth effect can be separated entirely. Therefore, besides the simultaneous model with both variables included we estimate additionally the same model but with either *PUBL* or *GRAD* excluded and report the significance level in brackets (Table 12).

The growth rates of small firms are not significantly correlated with *PUBL* or *GRAD.* Although it is not possible to conclude that such relationships are not given, this result tends to confirm the expectation that small firms, because of the lack of absorptive capacities, are not or only marginally able to benefit from public research and higher educational activities. The impact of *PUBL* becomes significant for medium sized firms at $\theta_{0.5}$ and for large firms at $\theta_{0.25}$ and $\theta_{0.5}$ in the joint model (together with *GRAD*), and significant for all conditional quantiles in the individual model. No patterns for differences along the quantiles become apparent. Put differently, after a certain size threshold public research activities are beneficial for all kinds of growth levels. Hence, *hypothesis 2a* is confirmed, while *hypothesis 2b* is rejected. The medium-sized and large firms are those that benefit from nearby public research activities.

Similar findings can be reported for *GRAD*, with the exception that the estimates clearly increase at higher quantiles (it becomes even slightly significant for small firms at $\theta_{0.75}$), and that for declining firms at $\theta_{0.25}$ they are not significantly different from zero. This confirms *hypothesis 3*, which states that especially larger firms and high-growth firms benefit from higher educational activities. Firms require an adequate pool of available and qualified workers to expand, which in absolute terms holds even more so for large, fast-growing firms. In contrast to *PUBL*, which helps to decrease the likelihood of negative growth events, no such effect is observed for *GRAD*, which supports the idea that this variable measures the contribution of universities to the local labour market.

Comparing the effects of *GRAD* and *PUBL* in the joint model, it turns out that *PUBL* can keep its significance in three cases, and *GRAD* only in one case. Higher educational activities, which seem to directly increase the firms' growth opportunities by their contribution to the local pool of qualified graduates, nonetheless cannot compete with public research activities in explaining firm growth. New scientific knowledge from public research activities, which first has to be transformed by the firms into economic opportunities, seems to be at least equally relevant.

Finally, the estimated distances provide some information on the spatial dimension of the growth impact of public research and higher educational activities. In contrast to related employees, considerable smaller distances of a few minutes result for both *GRAD* and *PUBL*. Here, we expect that small firms cooperate only with local universities or public research institutes due to travel costs, and that larger firms either cooperate with local ones or with the best global alternatives. If public research activities are performed nearby, firms benefit most as short distances facilitate social interactions and face-to-face contacts. Also for graduates, small distances predominate. To conclude, externalities that originate from public research and higher educational activities occur to a large part at a geographical scale much smaller than usually assumed as "regional". Hence, distance-based and micro-founded methods help to avoid underestimating their effect. Agglomeration effects from related economic activities do not take halt at predefined regional boundaries, which in many studies coincide with administrative territories.

Table 12: Regression results for firm-specific location variables

	small [5, 50)			medium [50, 250)			large [250, 1000)		
	$\theta_{0.25}$	$\theta_{0.50}$	$\theta_{0.75}$	$\theta_{0.25}$	$\theta_{0.50}$	$\theta_{0.75}$	$\theta_{0.25}$	$\theta_{0.50}$	$\theta_{0.75}$
EMPL_2	.0004 **	.0001 *	-.0002	.0002 **	.0004 ***	-.0003	.0013 ***	.0009 ***	.0016 **
EMPL_3	.0001 '	.0001 *	.0007 **	.0004 ***	.0002 *	.0009 ***	.0004 **	.0002 *	.0003 *
EMPL_4	.0005 ***	.0001 '	.0010 **	.0003 ***	.0001	.0002 *	.0005 ***	-.0001 *	-.0006 ***
PUBL	.0004	.0001	.0003	.0002 (*)	.0006 * (***)	.0005 (*)	.0014 **	.0010 * (***)	.0017 (**)
GRAD	.0003	.0000	.0012 ' (')	.0001	.0003 (**)	.0006 (*)	.0000	.0008 * (***)	.0012 (**)

p-values: ' < 0.1, * < 0.05, ** < 0.01, *** < 0.001

Table 13: Regression results for firm- and region-specific control variables

	small [5, 50)			medium [50, 250)			large [250, 1000)		
	$\theta_{0.25}$	$\theta_{0.50}$	$\theta_{0.75}$	$\theta_{0.25}$	$\theta_{0.50}$	$\theta_{0.75}$	$\theta_{0.25}$	$\theta_{0.50}$	$\theta_{0.75}$
EMPL_2	105	94	18	8	119	24	119	120	120
EMPL_3	12	26	10	99	16	25	16	15	14
EMPL_4	120	120	120	116	32	32	51	13	14
PUBL	9	8	11	9	17	12	5	16	16
GRAD	119	6	24	21	6	6	49	9	15

Numbers in bold if p-value of corresponding variable is significant

4.6 Conclusions

FRENKEN et al. (2011) have suggested for future research that one of the main challenges "lies in settling contradictory empirical findings. In particular […] the main gap in our empirical understanding concerns the effect of localization economies on firm performance, which some may even consider the key question in economic geography at large". In line with these authors, this paper argues that many contradictory empirical findings are closely related to the heterogeneity of firms, the heterogeneity of the sources of externalities, and the heterogeneity of spatial economic landscapes, which are a product of the location of both the firms and the sources.

This paper takes the call for a finer resolution seriously. The effect of different sources of externalities on firm growth is compared for small, medium and large firms. Quantile regression techniques additionally shed light on the relationships for highly growing and declining firms. This paper finds that the impact of related economic activities, measured by related employees, depends on both the size of the firm and the degree of relatedness: the larger the firm, the more diverse nearby activities should be. To put differently, specialised agglomerations rather hamper the growth of large firms, but might stimulate the growth of small firms. For public research and higher educational activities we observe a clear size threshold: only medium and large firms are able to benefit from nearby publications and graduates in the STEM fields.

The geographical meaning of "nearby" differ for the analysed sources. By endogenizing distances, it turns out that some externalities, mostly related economic activities, transcend traditional regional boundaries, whereas externalities from public research and higher educational activities are very local phenomena of few driving minutes. This implies that the locations of universities and research institutions are crucial for firms and, hence, regional economic development. It is important that their location matches the quite local economic activity in order to realise to potential economic benefit.

The above findings highlight the advantages of using a distance-based, data-driven approach, which does not rely on predefined regional delimitations. The spatial reach of various relationships between variables differs. More studies in this direction are necessary to obtain a better understanding of the spatial aspect in economic dependencies.

Finally, it becomes clear that not only firm internal factors can drive the growth of firms, nor is it exclusively stimulated by firm external factors. Instead, it is the complex interplay between internal factors and external factors, an interplay that depends on the kind (i.e., related employment, graduates and publications) and spatial extent of externalities. Like VANOORT et al. (2012) or RIGBY and BROWN (2013), this paper focuses on differences along firm size. However, the heterogeneity of firms – one of the few invariables of industrial dynamics (DOSI et al. 2010) – also concerns other aspects than size. For instance, firms' age (e.g. NEFFKE et al. 2012) or industry affiliation (e.g. DUSCHL et al. 2014a, BEAUDRY/SWANN 2009) mediate the growth impact of externalities. Hence, we would motivate for a stronger integration of the various dimensions of heterogeneity in a comprehensive empirical framework. Furthermore, one could explore more systematically the reasons behind the firm-specific differences in the effects of externalities, for example by disentangling the mechanisms in comparative case studies.

5 Industry-specific firm growth and agglomeration

This chapter is a reprint[33] of:

DUSCHL, M., SCHOLL, T., BRENNER, T.. LUXEN, D. & RASCHKE, F. (2014): Industry-specific firm growth and agglomeration. In: Regional Studies DOI: 10.1080/00343404.2013.861059.

Abstract: This paper analyses the industry-specific relationship between industrial clustering and firm growth. Micro-geographically defined agglomeration measures based on travel time distances and a flexible log-logistic decay function framework are used to study the spatial impacts of related economic and knowledge generating activities in 23 industries. We find that firms' growth prospects tend to be generally hampered by the agglomeration of own-industry employment, whereas the impact of proximate scientific activities systematically depends on the kind and age of industry. Furthermore, the optimal specifications of decay function that measures agglomeration effects considerably vary both between the industries and variables.

Keywords: Firm growth, industrial clusters, agglomeration, MAUP, distance decay function, quantile regression

JEL Classification: C31, D92, L25, R11

[33] Due to better readability, numbering of sections, tables, figures and formulas has been changed. References can be found at the end of the thesis.

5.1 Introduction

The geographical location has for a long time been "a neglected determinant of firm growth" (AUDRETSCH/DOHSE 2007), but recently an increasing bulk of literature examines the impact of being located within agglomerations and industrial clusters or in proximity to universities on the performance of firms. However, empirical findings are still contradictory. This is not surprising, as one of the few invariables of industrial dynamics is the heterogeneity of firms (DOSI et al. 2010). From the existing literature, four key issues are known as a potential source of contradicting results: First, there are strong differences between industries. For instance, agglomeration economies differ between manufacturing and service industries as well as at finer levels of disaggregation (e.g. BEAUDRY/SWANN 2009). Second, processes and mechanisms differ with the industry's age and particularly its stage in the industrial life-cycle. Empirical investigations suggest that agglomeration of economic and knowledge generating activities is more important in the initial phase of an industry (e.g. POTTER/WATTS 2011). Third, agglomeration can be measured with different statistical methods and fourth, the spatial dimension used within the methods matters. Results might depend on the chosen regional level, i.e. if the investigation is done on the city-level, zip-codes, or other functional definitions of regions. Some investigations even report contradicting results using the same dataset when changing from one aggregation level to another (e.g. BUERGER et al. 2010).

Recently, much effort has been undertaken to improve the understanding of spatial dependencies by shifting from aggregated large-scale investigations to more micro-geographic data driven approaches (e.g. DURANTON/OVERMAN 2005). Beside the issue of availability, micro-geographic data also requires new techniques in order to integrate them into econometric models. Given the few existing publications that deal with micro-geographic data, the literature remains unclear on how the methodological challenges can be met. While micro-geographic data enhances the validity of research findings on spatial matters of firm growth, investigations have yet mostly neglected the physical nature of firm's geographic location. As regards the access to growth relevant sources of agglomeration economies, its location relative to the road network plays an important role.

Given these yet unresolved questions, the aim of this paper is to re-examine the effects of external factors to the growth prospects of incumbent firms. The literature provides clear evidence for the fact that agglomerations or the proximity of universities impacts on firm growth, but the details of these impacts are not well studied so far. More detailed empirical knowledge about these impacts will provide hints about the underlying mechanisms. The theoretical literature puts forward various potential mechanisms, but little can be said about their (relative) relevance. We go into detail in two aspects: First, we study several industries separately in order to examine the role of industrial characteristics, such as maturity. Second, we take a detailed look on spatial aspects in order to examine the spatial range of relationships, which provides hints about the mechanisms behind the detected relationships. This is achieved by geolocating firms into space and by calculating travel time distances to all related economic and knowledge generating activities. Firms' access to these activities is spatially discounted by a flexible log-logistic distance decay function, which can be deduced from behavioural assumptions and which is further specified based on empirical data. Using a quantile regression framework, the impact of nearby economic

and knowledge generating activities on employment growth of German firms is compared across 23 industries, that is groups of related industries as defined by the EU Cluster Observatory.

The findings suggest that being located in agglomerations of own-industry employment does not increase but rather reduces firms' growth prospects. In contrast hereto, the impact of being located in proximity to knowledge generating activities depends more systematically on the type and age of the industry – in less mature, more knowledge-intensive industries the relationship with firm growth tends to be positive. Also the spatial scale of what proximity means varies across industries. In order to account for the heterogeneity of the analysed industries, results from three representative cases are finally discussed more in-depth.

The paper is structured as follows. In section 5.2, expectations on the industry-specific impacts of nearby economic and knowledge generating activities on firm growth are deduced from the existing literature. Data issues are described in section 5.3, while the methodology of a data-driven distance decay function specification and a quantile regression framework with spatially discounted variables is outlined in the fourth section 5.4. Section 5.5 presents and discusses the results. Section 5.6 concludes.

5.2 Literature

5.2.1 The impact of agglomeration on firm growth

A rich body of academic work exists that can be labelled with the term 'agglomeration theory', among whom the theory of industrial clusters (PORTER 2000), which comprises being located in spatial proximity to similar firms and associated institutes like universities, has attracted high interest both in science and politics. The impact of agglomeration economies has been studied from many different points of views. Whilst the effect on start-ups (e.g. BUENSTORF/KLEPPER 2009), firm survival (e.g. NEFFKE et al. 2012) or the innovative (e.g. FORNAHL et al. 2011) and productive performance (RIGBY/ESSLETZBICHLER 2002) is well documented, FRENKEN et al. (2011) identify in their survey on *industrial dynamics and economic geography* a gap in the empirical understanding of the effects on the growth performance of incumbent firms.

PORTER and his co-authors find in several papers that within clusters wages, innovativeness and entrepreneurship are higher (PORTER 2003a, DELGADO et al. 2012). Benefiting from the co-location with other similar firms is one of the major arguments within cluster theory (PORTER 2000). This implies that firms that are located within agglomerations of their industry should have higher growth potentials. The comprehensive study by BEAUDRY and SWANN (2009) on 56 two-digit industries in the UK supports this argument: In about half of the industries, being located within agglomerations of workers in the same sector has a positive impact on firms' growth prospects.

In contrast, other researchers do not find higher survival or growth rates for firms located in clusters (BUENSTORF/KLEPPER 2009). While still acknowledging the positive influence of

clusters on spin-offs, FRENKEN et al. (2011: 2) conclude that "there is little evidence that clusters enhance firm growth and survival". Various empirical investigations show that industrial clusters contribute to firm growth only under certain circumstances and that it is of high importance which constituent parts (agglomeration of similar firms, research institutions, etc.) are observed. However, further evidence is missing and various mechanisms are claimed to be responsible for this benefit. Therefore, a more distinguishing perspective might be helpful.

One explanation of why firms might benefit from industrial clusters is based on phenomena like innovation, learning and knowledge spillovers (MALMBERG et al. 2000). Growth relevant knowledge is generated by competing and cooperating firms, but also by research activities in universities or public research institutes. In the latter case, the literature is less ambiguous. Many studies show a positive link between the presence of universities and firms' innovation (e.g. JAFFE 1989) and growth performance (e.g. AUDRETSCH/LEHMANN 2005, CASSIA et al. 2009 or RASPE/ VANOORT 2011). This relates to research on the relationship between regional knowledge intensity and firm performance, which (indirectly) assumes spatially bounded knowledge spillovers to be one of the main mechanisms of agglomeration economies (FRENKEN et al. 2011). Again, an industry-specific perspective is suggested: knowledge spillovers from external research activities should matter most for firms of science-based or knowledge-intensive industries, which also can be expected to engage more in internal research activities as well as to provide the required absorptive capacities (KOO 2005).

Besides the kind of industry, the industry's age and its life-cycle stage can help to understand the multifaceted influences of industrial clusters on firm growth (FELDMAN 1999). This is also reflected by the recent focus on the dynamics of agglomeration economies during industrial life-cycles (e.g. NEFFKE et al. 2011). In early life-cycle phases, which often coincidence with the emergence of new industrial clusters, the rates of start-ups and spin-offs tend to be high. Local conditions, like the presence of related industries, public research institutes or universities, play a major role in the initial development of clusters. At more mature stages, however, market growth slows down and a kind of equilibrium is reached (BRENNER/SCHLUMP 2011). Empirical evidence suggests that under these circumstances, firms do not benefit any more from being located in industrial clusters (AUDRETSCH/FELDMAN 1996), and their growth prospects might be even hampered due to increasingly prevailing negative agglomeration economies such as intensified competition (POTTER/WATTS 2011).

To conclude, we expect that industrial clusters have manifold effects on firm growth: effects of the agglomeration of related economic activities are ambiguous, depending on the kind of industry and its stage in the industry life-cycle. In contrast, the effects of proximate knowledge generating activities should be rather positive, especially so in less mature, more knowledge-intensive industries.

5.2.2 Spatial matters of firm growth and agglomeration

Although the impact of industrial clusters on firm growth is highlighted in many studies, the literature has remained quite silent on the spatial range of their influence and the actual definition of space. For most of the papers, the definition of space arises from the used dataset, i.e. the spatial aggregation level of the data such as cities or regions. For quantitative driven investigations, mostly regression models, this spatial definition concerns both the dependent and the independent variable.

Table 14 provides a non-exhaustive overview on the different approaches that can be found in the literature. Starting with the dependent variable, two general concepts can be separated: The macro approach investigates growth effects on a regional level while the micro approach deals with growth effects on single firms. Empirical evidence exists that the clustering of industries exerts a positive impact on regional economic performance, both for the entire regional economy (e.g. DELGADO et al. 2010) as well as at the disaggregate level of industries within the region (e.g. DELGADO et al. 2012). This evidence is refined by the insight that the underlying mechanisms like knowledge spillovers are most pronounced in regions where at the same time the variety and the relatedness of the agglomerated industries are highest (FRENKEN et al. 2007). Spatially more sophisticated approaches deal with neighbourhood effects – how neighbouring regions influence the growth of a specific region. Recently, however, both the growing access to firm data and to computational power, have led to a focus on micro approaches. Investigating growth on the firm level allows for a more sophisticated testing as firm-specific variables can be included and intra-regional heterogeneity can be observed. While the dependent variable of the micro approach is the growth rate of a single firm and therefore not aggregated, most studies explain growth processes by means of characteristics of the region the firm is located in. Here, the literature has brought forward manifold regional measures for concentration, diversity or competition, ranging from simple counts, relative measures like the LQ to more complex, derivative measures. Because imperfect competition and heterogeneous firms are defining characteristics of the economic landscape, regions as consistent and homogenous aggregates do not exist in reality. As a consequence, regionalization, an ex-post abstraction of the continuous landscape, would imply a huge loss of information (for an extensive discussion on this issue we refer to PINSKE/SLADE 2010 or HARRIS 2011). Thus, analogue to the macro approaches, results are affected by the arbitrariness of regional boundaries and moreover by the chosen level of aggregation. This issue of zoning and scaling was first described by OPENSHAW (1984) and coined with the term Modifiable Areal Unit Problem (MAUP). By varying the spatial scale of analysis, BUERGER et al. (2010) as well as WENNBERG and LINDQVIST (2010) show empirically that the MAUP is highly relevant for agglomeration economies.

Avoiding the MAUP requires two methodological aspects: First, the aggregation level of the data should be as low as possible. DURANTON and OVERMAN (2005) refer to this as micro-geographic data, which we obtain in our case by computing the easting and northing of each firm's municipality. Municipalities represent the lowest aggregation level in Germany, currently with a number of 11249 and an average size of 31.6 km^2. Second, distance-based methods have to be applied for the calculation of the independent variables. One method is using distance bands, i.e. counting the observance of firms at specific radii (e.g. ROSENTHAL/STRANGE 2003). Another approach is the use of distance

decay functions that build proxy values of agglomeration by summing up localizable activities multiplied by inverted distances. Various specifications of both the distance bands and the decay functions exist in the literature. Concerning the latter approach, mostly simple linear (e.g. AUDRETSCH/LEHMAN 2005) or exponential decay functions (e.g. DRUCKER/FESER 2012) are used, although DEVRIES et al. (2009) have shown that a log-logistic function is best suited for modelling spatial interactions, in their case, the effect of transportation costs on commuting flows. This function, derivable from behavioural assumptions, represents a rather flexible approach, to which the exponential decay and even the distance bands are only special cases. Because agglomeration economies are reported for a wide range of different distances, from a narrow local to supra-regional scale, in our approach the best fitting decay function will be identified based on empirical data and for each industry separately.

Table 14: Literature overview

Approach	Observation	Industry-focus	MAUP	Examples
Macro (regions)	Intra- region effects	No	Yes	FRENKEN et al. (2007)
		Comparison		GLAESER et al. (1992); HENDERSON et al. (1995); PORTER (2003a), SPENCER et al. (2010)
	Neighbourhood effects	No		KUBIS et al. (2009); DELGADO et al. (2010); ARTIS et al. (2011)
		Comparison		DELGADO et al. (2012)
Micro (firms or plants)	Intra-region effects	No	Yes	GUISO/SCHIVARDI (2007); CAINELLI (2008); BOSCHMA et al. (2009); CASSIA et al. (2009); ANDERSSON/LÖÖF (2011)
		Comparison		RIGBY/ESSLETZBICHLER (2002); HENDERSON (2003); AUDRETSCH/DOHSE (2007); BEAUDRY/SWANN (2009); WENNBERG/LINDQVIST (2010)
	Distance bands	No	No	HOOGSTRA/VANDIJK (2004); BALDWIN et al. (2008); ERIKSSON (2011)
		Comparison		ROSENTHAL/STRANGE (2003); BALDWIN et al. (2010); GRAHAM (2009)
	Distance decay	No		AUDRETSCH/LEHMANN (2005); LYCHAGIN et al. (2010)
		Comparison		VANSOEST et al. (2006); GRAHAM et al. (2010); DRUCKER/FESER (2012); DRUCKER (2012)

Beside the choice of the distance model, results may also depend on the way how distance between firms is computed. The vast majority of distance-based investigations uses orthodromic distances (e.g. km or miles), although this might cause errors if the 'economic' distance between firms deviates from the orthodromic distance, for instance, if firms are located in mountainous or less well connected regions (DURANTON/OVERMAN 2005).

Obviously, driving distance or travel time are more appropriate, given that agglomeration economies are assumed to arise from low transportation costs or the convenience of face-to-face contacts. One of the few exceptions using travel times is the work of AUDRETSCH and LEHMANN (2005), where the growth of firms is investigated with respect to the firms' driving distance to their closest university. However, studies where driving distances are computed to thousands of locations are, due to the high computational costs of route planning, very rare. Using an efficient many-to-many route planning algorithm, introduced by KNOPP et al. (2007), we compute travel times between all German municipalities, allowing us to investigate agglomeration effects on firm growth from a more realistic spatial perspective.

With respect to the discussed literature, the paper at hand belongs to the group of micro approaches as it observes the growth rates of each individual firm. It uses a flexible distance decay function and compares agglomeration effects resulting from related economic and scientific knowledge generating activities across disaggregated industries. From a methodological point of view, our paper differs from the existing literature regarding two aspects: First, instead of spherical distances travel time in minutes is used. Secondly, we do not anticipate a specific distance decay function but include its identification into our analysis in order to detect possible differences among industries.

5.3 Data

5.3.1 Definition of industries

The 23 industries used for the current analyses were taken from the EU Cluster Observatory and can be seen as a standard definition for industry-related policy programmes on regional development in Europe. The definition goes back to a US cluster mapping project undertaken by PORTER in the early 2000s and is based on the distinction between local and natural-resource-driven industries on the one hand, and export-oriented traded industries on the other (PORTER 2003a). The latter industries were grouped at the 4-digit level of the standard industrial classification (NACE Rev. 2) according to co-location patterns within economic activities from data across the US and led to groups of related industries. PORTER's analysis concludes that the regional presence of those activities that tend to cluster in space can be seen as a driver for regional economic performance and the positive development of embedded firms (PORTER 2003a, WENNBERG/LINDQVIST 2010). Therefore, the definition of industries as groups of co-located and hence, related activities, is suitable to survey the relationship between industrial clustering in space and firm growth.

5.3.2 Dependent variable

The BvD Amadeus database discloses the address of the firms' headquarter location. Operational and strategic decisions are often made within this organizational unit. Although R&D intensive foreign direct investments have recently become more important (UNCTAD 2005), many of the firms' R&D activities still remain located close to the headquarters (GUIMÓN 2009). Even for multinational enterprises a home bias for innovation activities is evident (COHEN et al. 2009, DUNNING/LUNDAN 2009). Therefore, we follow BEAUDRY and SWANN (2009) in assuming that it is the regional environments of a firm's headquarter which is most decisive in affecting the growth prospects and the decision of the organization on taking up employees. This rationale breaks down for larger firms, which tend to be less focused on their headquarters, but disperse activities in many increasingly independent establishments across the country and even beyond. Therefore, the analysis is restricted to firms with no more than an annual average of 1000 employees.[34] Also very small firms with less than 5 employees, which growth processes are known to be rather erratic, are excluded (COAD 2009).

Growth rates are calculated by taking the difference of the natural logarithms of the size *S* (measured by the entire stock of employees) of firm *i* between two successive years *t*:

$$g_{i,t} = log(S_{i,t+1}) - log(S_{i,t}) \tag{25}$$

Confronted with an unbalanced panel from 2004 to 2010, yearly growth rates are pooled together.[35] In the course of one year, firms essentially face three options: they may expand, shrink or remain at the previous level. Zero-growth events, usually quite abundant for employment, make up 44.5% of the original data. A considerable but unknown share of these events can be attributed to a lack of regular updating of database entries, which are simply extrapolated from previous years. Including these events would bias the assessment of the impact of agglomeration on firm growth. Besides this data issue, firms that opt for not to grow might be distinct from actually changing (expanding or shrinking) firms. Although being an economically rational choice in the absence of any changes in business opportunities, this option is often preferred even in cases when opportunities have changed. To name just a few examples, firms might be reluctant to expand because the inclusion of new employees is costly as it implies re-organisation of internal tasks and management functions, or the fear of losing control might frighten some managers (COAD 2009). In a similar vein, firms can be reluctant to shrink despite reduced business opportunities. Firms invest in building up redundancies in difficult times instead of immediately dissolving existing working contracts, or managers might be not fully aware of the necessary down-sizing. Because from a technical point of view it is impossible to distinguish between data problems and the various other reasons why firms do not grow, we analyse only growth events in which the size of firms actually changes.

---

[34] This size restriction implies that 2.2% of total firms in the database are dropped from the sample. Also a different threshold of 500 employees (with the loss 4.5%) was tested, however the results remain robust.

[35] By repeating the analysis on a year-by-year basis (for industries with enough observations) it can be shown that the results remain robust.

Industries with less than 1000 yearly growth events are omitted due to robustness issues. The remaining 23 industries are listed in Table 15, together with their number of pooled growth events $g_{i,t}$, number of firms, and the average age of these firms.

Table 15: Overview on analysed industries

ID	Name	N ($g_{i,t}$)	N (firms)	Age
1	Agricultural products	1077	688	19
2	Automotive	1721	632	20
3	Building fixtures & equip.	2529	1115	23
4	Business services	5057	2417	13
5	Chemical products	1385	504	24
6	Construction	6278	3057	24
7	Distribution	5108	2488	22
8	Entertainment	1179	558	14
9	Financial services	1108	540	14
10	Heavy Machinery	1041	391	21
11	Instruments	1336	539	25
12	IT	2668	1161	15
13	Media & publishing	3038	1654	23
14	Medical devices	1068	573	19
15	Metal manufacturing	7189	3265	25
16	Paper products	2159	970	24
17	Plastics	1861	748	24
18	Processed food	4652	2222	26
19	Production technology	5273	2072	24
20	Telecom	1406	524	21
21	Textiles	1097	488	26
22	Tourism & hospitality	2100	1472	18
23	Transportation & logistics	2406	1013	17

5.3.3 Independent variables

Firms' potential to benefit from industrial clusters is specific to characteristics of the firms as well as of the corresponding regions (BEUGELSDIJK 2007, ERIKSSON 2011). Therefore, the independent variables consist of three different kinds: First, we control for relevant demographic properties of the firms. Second, we include measures of the general environment of the region the firm is located in. Third, the focus of this paper lies on firm-specific location variables reflecting agglomerations of related economic and scientific activities.

Control variables: demographic and regional variables

Building upon the literature on firm growth, which mostly extends a GIBRAT-like growth regression (see COAD 2009 for an overview), we control for the logarithm of size, age, and whether or not it is a subsidiary firm. It counts as a stylized fact in industrial dynamics that firm growth is negatively related to both size and age. In addition, two variables are chosen to control for the general regional environment. Urbanization economies *per se*, which are rather independent from the surrounding industrial structure (BUERGER et al. 2012) and which might be both positive or negative, can be measured by the population density of the corresponding district, wherein a firm is located (ERIKSSON 2011).[36] The unemployment rate of the firm's regional labour market (as defined by ECKEY et al. 2006) reflects the vitality of the regions' socio-economic conditions. In the special case of Germany it also accounts for structural differences along the east-west and north-south divide. Data for both variables is obtained from the German Federal Statistical Office. The global macroeconomic recession 2008-10 systematically lowered the firms' growth prospects. Therefore, a dummy variable for the crisis years is constructed. Finally, the absolute location within Germany might influence the magnitude of agglomeration economies. Potential cross-border effects cannot be considered, discriminating firms located close to the border. Due to historical reasons, two dummies are constructed: one for the location in border regions with the New Member States of the EU and one for all other border regions.

Firm-specific location variables: own-industry employment and publications

In contrast to the regional control variables that account for a rather diffuse socio-economic environment (or "social filter", as denominated by RODRÍGUEZ-POSE and CRESCENZI 2008), other economic and knowledge generating activities can be traced back to concrete localizations in space: firms compete, cooperate, and learn from each other, and new scientific knowledge originates from universities and research institutes. These related economic and knowledge generating activities can be approximated by the number of employees in the same group of related activities (hereafter, industry) and scientific publications, respectively. The Federal Employment Agency provides data on industry-specific employment for municipalities, the lowest aggregation level in Germany. Data on scientific publications were collected from the ISI Web of Science and assigned to municipalities on the basis of authors' addresses. Since the two variables will be the focus of our paper, the subsequent section explains in more details how firm-specific location variables can be constructed by discounting these geo-localized activities with their distances to the firm's location. Therefore, bilateral travel times are calculated by exploiting results from graph theory and data on the German road network from the OpenStreetMap project; the algorithms are described in DUSCHL et al. (2014a) and more extensively in GEISBERGER et al. (2010). Intra-municipality distances are set to 5.01 minutes, the average bilateral travel time between 1000 randomly drawn pairs of firms' address locations, each belonging to the same municipality.

[36] A quadratic term to control for nonlinear relationships was initially included, but never found to be significant.

5.4 Model and estimation

5.4.1 Construction of firm-specific location variables

Employees in the same industry and scientific publications are discounted by an industry-specific distance decay function *f(d$_{im}$)* based on travel time distances *d$_{im}$* between the places of firms *i* and the municipalities *m*. The firm-specific variables for agglomeration of related economic (*AGGL*) and knowledge generating activities (*KNOW*), after normalizing with $\mu_{AGGL} = \frac{\sum_m empl_{m,t}}{\sum_i \sum_m f(d_{im}) empl_{m,t}}$ and $\mu_{KNOW} = \frac{\sum_m publ_{m,t}}{\sum_i \sum_m f(d_{im}) publ_{m,t}}$, read:

$$\begin{aligned} AGGL_{i,t} &= \mu_{AGGL} \sum_m f(d_{im})\, empl_{m,t} \\ KNOW_{i,t} &= \mu_{KNOW} \sum_m f(d_{im})\, publ_{m,t} \end{aligned} \tag{26}$$

These spatially discounted variables can be included in a simple linear model (for spatial econometric issues we refer to ANDERSSON/GRASJÖ 2009)[374]:

$$g_{i,t} = \alpha + \beta_1 AGGL_{i,t} + \beta_2 KNOW_{i,t} + \sum_{j=3}^{10} \beta_j x_{i,t,j} + \varepsilon_{i,t} \tag{27}$$

with α and β representing the coefficients to be estimated and *x* the seven firm- and region-specific control variables. The error term is denoted by $\varepsilon_{i,t}$. The applied normalization procedure allows for an interpretation of the corresponding regression coefficients as the impact of one additional employee or publication on the growth of a firm with a given distance. After briefly introducing quantile regression techniques as an adequate estimation method in the context of firm growth, the still outstanding specification of the distance decay function *f(d)* will be discussed in the following sections.

5.4.2 Estimation using quantile regression techniques

It is one of the stylized facts of industrial dynamics that firm growth rates are not normally distributed, but show fat tails (for an overview on empirical studies for different countries see COAD *forthcoming*). Therefore, quantile regression techniques, which are robust to outliers in the dependent variable and free from any distributional assumption in the error term (BUCHINSKY 1998), are more appropriate. Besides, the specific conditional quantiles of strongly expanding ($\theta_{0.75}$) and declining ($\theta_{0.25}$) firms can be analysed in addition to the median growing firm ($\theta_{0.5}$). Our intuition is that high growth events, a dominant feature of firm growth, rely differently on internal as well as external factors. Technical details are described in KOENKER (2005). Here we only point out that, similar to OLS regression, the coefficient estimates can be interpreted as partial derivatives, meaning the impact of a one-unit change of an independent variable on the firms' growth rate at the θ_{th} quantile holding all other variables fixed. Standard errors are obtained by bootstrapping based on

[37] Performing an extensive Monte Carlo analysis, these authors show that this approach captures substantive spatial dependence in the dependent variable and accounts for both local and global spillovers.

200 replications to alleviate the problem of underestimation in presence of heteroscedastic error distributions (e.g. GOULD 1992).

5.4.3 Identification of decay function parameters

Social interactions are fundamental to all mechanisms that underlie agglomeration economies, like labour market pooling, contracting with suppliers and customers, transfer of knowledge, but even local competition. From simple transaction cost reasoning, the frequency of interactions should decay with distance. Moreover, the literature on commuting behaviour (JOHANSSON et al. 2003, ANDERSSON/KARLSSON 2007) shows that the negative travel time sensitivity is not linear in space, but varies between different geographical scales: within a narrow local context, interactions can take place at short notice and are primarily governed by randomness (THORSEN et al. 1999). Thus, within agglomerations interactions are only marginally affected by distance. At some threshold distance, however, the minimal cost principle predominates and consequently, the frequency and contribution of growth relevant economic interactions become highly distance-sensitive and may decrease rapidly. This threshold can be said to define the range of the region from a firms' perspective. For very long distances, geography ceases to matter once again. Mathematically, these behavioural assumptions can be expressed as a *S*-shaped and downward sloping log-logistic decay function of travel time *d*:

$$f_{r,s}(d) = 1/(1 + r^{-s} * exp(s * log(d))) = 1/(1 + (d/r)^{-s}) \quad (28)$$

with *r* and *s* representing two parameters that describe the shape of the curve (see DEVRIES et al. 2009 for technical details). Parameter *r* determines the location of the curve's inflection point, and parameter *s* its degree of steepness. The curve starts rather flat with the value of 1, becomes steeper, and then gradually flattens again to approach 0. If *s* becomes 1, the curve takes the shape of a negative exponential function. If *s* tends towards infinity, the function resembles a binary distance circle, with values of 1 for distances below *r*, and 0 for distances above *r*. Keeping *r* constantly at 90 minutes, Figure 22 depicts five curves for different values of *s*.

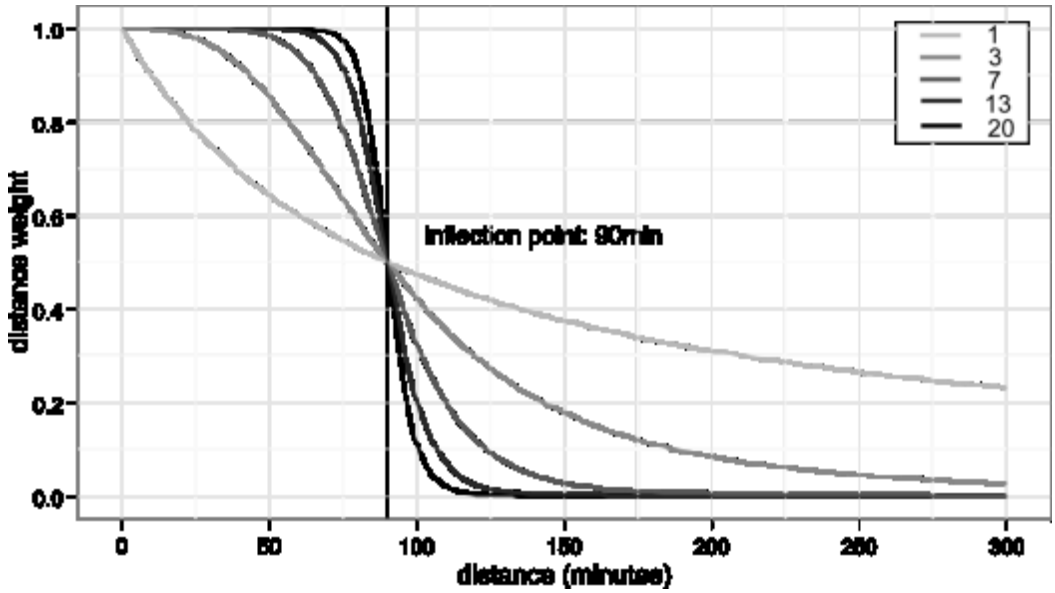


Figure 22: Log-logistic decay function for inflection point r = 90min and different degrees of steepness s

To identify the best specification of $f_{r,s}(d)$, the two variables *AGGL* and *KNOW* are regressed on the growth rates for each industry, for each quantile θ, as well as for each possible integer combination of the parameters r and s within the intervals [5, 300] and [1, 20], respectively. This means, r is allowed to vary between the minimal distance of 5 minutes, which corresponds to the intra-municipality distances, and 300 minutes, which is half the maximal travel distance found in Germany. The lower bound for s is mathematically given by 1, whereas the upper bound is set to 20, in which case the properties of a binary distance circle are approached. The smallest log-likelihood value gives the best fitting combination of r and s. Furthermore, the confidence intervals around these parameters, which contain plausible alternative specifications of $f_{r,s}(d)$, are determined using the likelihood ratio test. Besides a straightforward single optimum scenario like in production technology for the variable *KNOW* at $\theta_{0.5}$, also multiple optima are feasible (see Figure 23): firm growth in financial services correlate with related employees at a narrow local scale, expressed by a decay function with r = 5 minutes and a sharply declining shape of s = 20, and simultaneously at a wider spatial scale with r = 161 and s = 20. In cases such as the latter, all significantly distinguishable optima will be included into the regression model as separate decay function specifications. However, never more than two optima are identified in any single case.

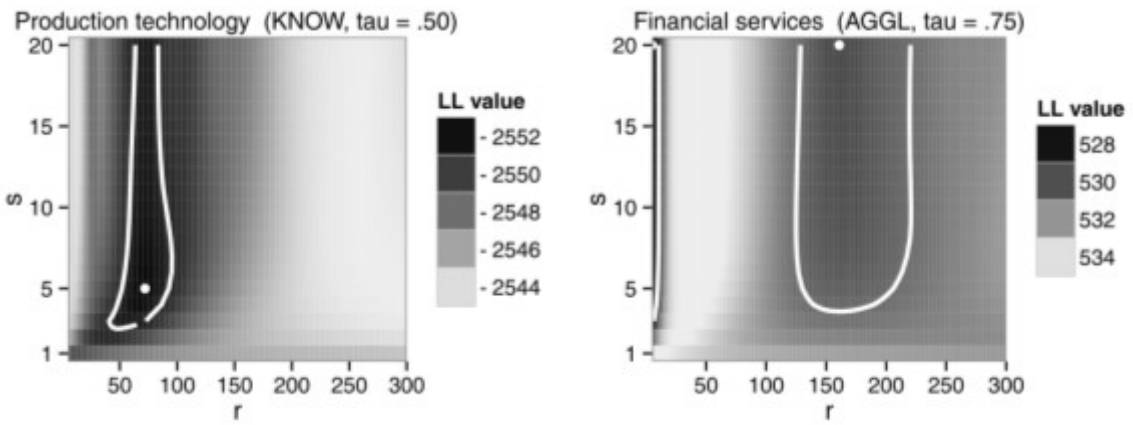


Figure 23: Examples of decay parameters optimization procedure with one minimum (left) and two minima scenario (right). Minima are shown as white points, confidence interval borders are shown as white curves.

5.5 Results

5.5.1 Distance decay function specification

Figure 24 provides an overview on the estimated parameters of the distance decay function for each industry. The optimization procedure tends to converge to the following decay function specifications: in most cases, the specifications are rather similar to distance bands, with s = 20, and sometimes even significantly outperform the exponential decay function, which is often used in the literature. Here, the effects may abruptly decay at very short distances of few minutes, which indicate the predominance of municipality-level and intra-regional effects, like in processed food (for *AGGL*) or heavy machinery (for *PUBL*), or even at distances beyond traditionally defined regional boundaries, like in medical devices or textiles, for which publications within the driving distance of two or three hours matter. In other cases, the impact of the location-specific variables on firm growth does not abruptly cease at any outstanding boundary, but decays exponentially, like in tourism & hospitality (for *AGGL*) or production technology (for *PUBL*). Furthermore, in some industries, like financial services (for *AGGL*) or metal manufacturing (for *PUBL*), different spatial scales may matter at the same time. In the remaining cases, the confidence intervals suggest that all specifications are feasible. This mostly holds true in industries, for which the subsequent regression analysis does not reveal any significant impact of own-industry employment and publications on firm growth. Here, cognitive, organizational, social or institutional proximity might be more relevant (BOSCHMA 2005). To conclude, the apparent industry-specific heterogeneity highlights the importance of a sound and flexible distance decay function specification to assess the impact of industrial clusters, which otherwise would be biased by the MAUP.

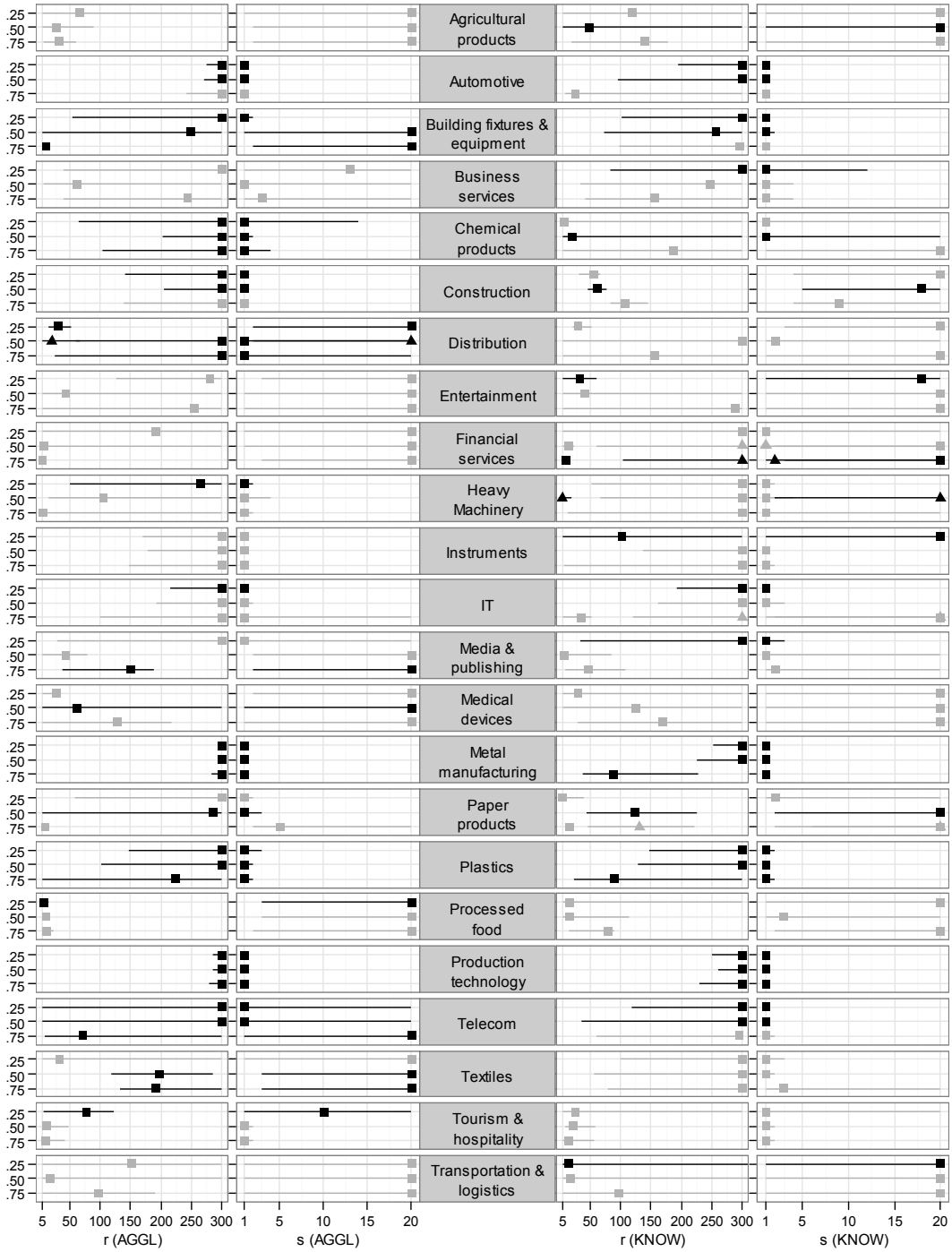


Figure 24: Estimated distance decay function parameters *r* (inflection point) and *s* (degree of steepness) with confidence intervals. Significant results (at 5%) are coloured black, a possible second optimum indicated by a triangle symbol.

5.5.2 Control variables

The estimated coefficients of the control variables are in line with the current literature on firm growth and industrial dynamics. For the sake of brevity, the main findings are only summarized.[38]

First, yearly growth rates always negatively correlate with the firms' size and in most industries also with age. This confirms the literature (e.g. EVANS 1987) rejecting GIBRAT's law which assumes that growth is independent of size (and age). The revealed relationships become even more pronounced for highly growing firms at the conditional quantile $\theta_{0.75}$. Only for shrinking firms, at $\theta_{0.25}$, the growth relationship vanishes for age and becomes even positively significant for size in most industries. These patterns simply imply that the growth of larger and older firms is less volatile: they are less likely to realize large growth jumps, and at the same time they are less prone to strong negative growth events. Since HYMER and PASHIGIAN (1962), this negative relationship between growth rate variance and firm size is well studied.

Second, in one third of the analysed industries, population density, a general measure of urbanization economies, comes along with lower average growth rates. The only exception is found in entertainment, where being located in high density districts means higher growth prospects. At the lower and higher quantile, its influence diminishes and remains significant in a handful of industries only. This finding confirms other studies of Germany (e.g. FORNAHL/OTTO 2008) and suggests that cost aspects due to congestion in densely populated places dominate when agglomeration effects of own-industry employment and proximate publications are directly taken into account. Similar findings can be reported for the unemployment rate, measuring the general structural and economic conditions, for which the correlation also tends to be negative in most industries, with the exceptions of construction and plastics.

Third, the two border-region dummies, which control for a potential underestimation of agglomeration economies across national boundaries, are merely significant. However, growth rates are, not surprisingly, significantly reduced during the years of the financial crisis. Only firms operating in agricultural products and paper products have shown to be resistant to the macroeconomic recession. Finally, being a subsidiary firm primarily matters for shrinking firms: with the support of a parent company, strong negative growth impulses seem to be cushioned more easily.

[38] The detailed regression results of the control variables are available on request.

5.5.3 Impact of own-industry employment and publications

Having controlled for various firm-specific and region-specific variables, the impact of the spatially discounted location variables can be discussed. Before taking a closer look at certain peculiarities at the industry-level, emerging general patterns are highlighted.

Being located in proximity to employees of the same industry (*AGGL*) does reduce the firms' growth prospect in many industries (see Table 16). At the lower quantile $\theta_{0.25}$ this relationship is significantly negative in seven, at the medium quantile $\theta_{0.5}$ in five and at the higher quantile $\theta_{0.75}$ in 12 out of 23 analysed industries. A significant positive relationship is found at each quantile only in three cases.

The prevailing negative agglomeration economies seem to contradict the usual belief that firms benefit from being located in a cluster. This finding may be explained as follows: Firms in clusters benefit from the surrounding in terms of higher innovativeness and competitiveness, but they have to pay higher wages for their employees and higher prices for real estate due to intensive competition (KETELS 2013). As a consequence, they do not show higher profits and growth rates. Particularly high growth events (at $\theta_{0.75}$) become difficult to realize for incumbent firms in industrial clusters, as many relatively expensive and often highly qualified workers have to be hired. Simpler business activities might even be moved outside of cluster places, where wages are lower, so that the number of employees might even shrink. Besides, firms also become more dependent on the development of surrounding firms that increases vulnerability to industry-specific problems, as the negative effects at $\theta_{0.25}$ indicate.

Exceptions to this pattern exist for construction, distribution, media & publishing and transportation & logistics, for which agglomeration of own-industry employment significantly stimulates the growth performance of firms. This finding can be explained by industry-specific characteristics. In these industries, which are composed to a major part of service activities, value creation strongly depends on frequent interactions with suppliers and customers. Despite negative agglomeration economies in industrial clusters, it seems that benefits from geographic proximity cannot be substituted in these industries by other dimensions of proximity (BOSCHMA 2005).

Table 17 contains the results for nearby scientific publications (*KNOW*). In eight of the analysed industries, *KNOW* decreases the growth prospects of firms in the lowest quantile ($\theta_{0.25}$). This negative impact, however, diminishing at higher quantiles, with four significant negative cases found at $\theta_{0.5}$ and only two at $\theta_{0.75}$. In contrast hereto, four industries benefit from *KNOW* foremost at $\theta_{0.25}$, whereas different four industries benefit at $\theta_{0.75}$.

These findings might be explained as follows. Growth, stimulated by distance-sensitive knowledge spillovers and learning processes between public research institutes and universities on the one hand, and private firms on the other hand, is innovation-driven. By providing new market opportunities, product innovations often unleash great potentials for expansion. For process innovations, which imply that the same amount of output can be produced with less input factors, a negative effect on employment growth is often observed in the empirical literature (see BUERGER et al. 2012 for an overview). Hence, innovation-driven growth results in turbulences, as both negative and positive growth events at the tails of the conditional growth rate distribution become more likely.

The positive effect of nearby publications on the growth prospects of firms is found in industries which tend to be more knowledge intensive: financial services, IT, medical devices, and telecom. In these industries, firms both rely on input from science and provide the necessary absorptive capacity to implement external knowledge. Besides, *KNOW* reduces the likelihood of negative growth events in industries like heavy machinery, plastics or textiles. For these industries, less science-based and innovation-driven in nature, the variable might capture other mechanisms of public research institutes and universities, like graduates, which protect these firms from declining.

Bridging the two measures for economic and knowledge generating activities by conjointly plotting the estimated coefficients in Figure 25 shows that the effects of *KNOW* and *AGGL* are less pronounced at $\theta_{0.5}$, where the estimates concentrate more closely to the origin, compared to the lower and upper quantiles. This means that the average growing firm is less dependent on economic and knowledge generating activities in its proximate surrounding, whereas the location matters most for firms in the tails of the conditional growth rate distributions. Besides, a significant negative relationship emerges between the industry-specific estimates of *KNOW* and *AGGL* at $\theta_{0.75}$, with a Pearson's correlation coefficient of -0.411 (p-value = 0.046). The counter-balancing tendencies of the two external factors for strongly expanding firms support the idea that industrial clustering is not an infinite self-reinforcing process.

Table 16: Coefficients for *AGGL* at different quantiles

AGGL

ID	Name	$\theta_{0.25}$	$\theta_{0.50}$	$\theta_{0.75}$
1	Agricultural products	-0.00015	-0.00004	-0.00002
2	Automotive	-0.00001 *	0.00000	0.00004
3	Building fixtures & equipment	0.00006	-0.00003	-0.00011 **
4	Business services	-0.00015 * -0.00003	-0.00004	-0.00011
5	Chemical products	-0.00004	-0.00000 -0.00010 *	-0.00013 **
6	Construction	0.00025 ***	0.00005 *	0.00008
7	Distribution	0.00014 *	0.00007 *	0.00069 *
8	Entertainment	0.00022	-0.00014	-0.00037 *
9	Financial services	0.00002	-0.00002	-0.00001 -0.00004 *
10	Heavy Machinery	-0.00013 *	0.00005	-0.00014 *
11	Instruments	0.00008	0.00007	-0.00002
12	IT	-0.00026 *	-0.00020 ‘	0.00014 ‘
13	Media & publishing	0.00033 *	-0.00004	0.00026 *
14	Medical devices	-0.00003	-0.00021	-0.00029 *
15	Metal manufacturing	-0.00013 **	-0.00007 *	-0.00009 *
16	Paper products	0.00012	0.00000 -0.0001	-0.00001
17	Plastics	0.00019	-0.00014 **	-0.00006 *
18	Processed food	-0.00003 *	0.00001	0.00000
19	Production technology	-0.00001	-0.00016 ‘	-0.00020 *
20	Telecom	-0.00002	-0.00009 *	-0.00012 **
21	Textiles	0.00010 ‘	-0.00007	-0.00050 *
22	Tourism & hospitality	-0.00023 **	-0.00082 **	-0.00077 *
23	Transportation & logistics	-0.00002	0.00001 *	0.00012 *
Significantly positive cases (at 5%)		3	3	3
Significantly negative cases (at 5%)		7	5	12

p-values: ‘ < 0.1, * < 0.05, ** < 0.01, *** < 0.001

Table 17: Coefficients for *KNOW* at different quantiles

KNOW

ID	Name	$\theta_{0.25}$	$\theta_{0.50}$	$\theta_{0.75}$
1	Agricultural products	-0.00016	-0.00013	-0.00069 ‘
2	Automotive	-0.00031 **	-0.00028 *	-0.00019
3	Building fixtures & equipment	-0.00026 *	-0.00060	-0.00024
4	Business services	0.00039	-0.00069 0.00031	-0.00055 ‘
5	Chemical products	-0.00031 **	-0.00001	0.00002
6	Construction	-0.00095 ***	-0.00099 ***	-0.00120
7	Distribution	0.00015	0.00050	-0.00137
8	Entertainment	0.00064	-0.00001	0.00007
9	Financial services	0.00038	0.00021	0.00061 *
10	Heavy Machinery	0.00011 ***	0.00007 **	-0.00013 *
11	Instruments	0.00045	0.00058	-0.00007
12	IT	0.00061 *	0.00045 **	0.00037 **
13	Media & publishing	0.00022	-0.00008	-0.00076 ‘
14	Medical devices	-0.00048 **	0.00019	0.00033 **
15	Metal manufacturing	-0.00139	-0.00142 **	-0.00140 -0.00133
16	Paper products	-0.00008	0.00003	0.00006 0.00027
17	Plastics	0.00018 *** -0.00203	-0.00220	-0.00006
18	Processed food	0.00077	0.00011	-0.00031
19	Production technology	-0.00064 *	-0.00080 *	-0.00155 *
20	Telecom	-0.00006	0.00018	0.00039 *
21	Textiles	0.00017 ** -0.00127 ***	-0.00072 ‘	-0.00046
22	Tourism & hospitality	-0.00017	0.00114	0.00196
23	Transportation & logistics	-0.00031 *	0.00002	-0.00069
Significantly positive cases (at 5%)		4	2	4
Significantly negative cases (at 5%)		8	4	2

p-values: ‘ < 0.1, * < 0.05, ** < 0.01, *** < 0.001

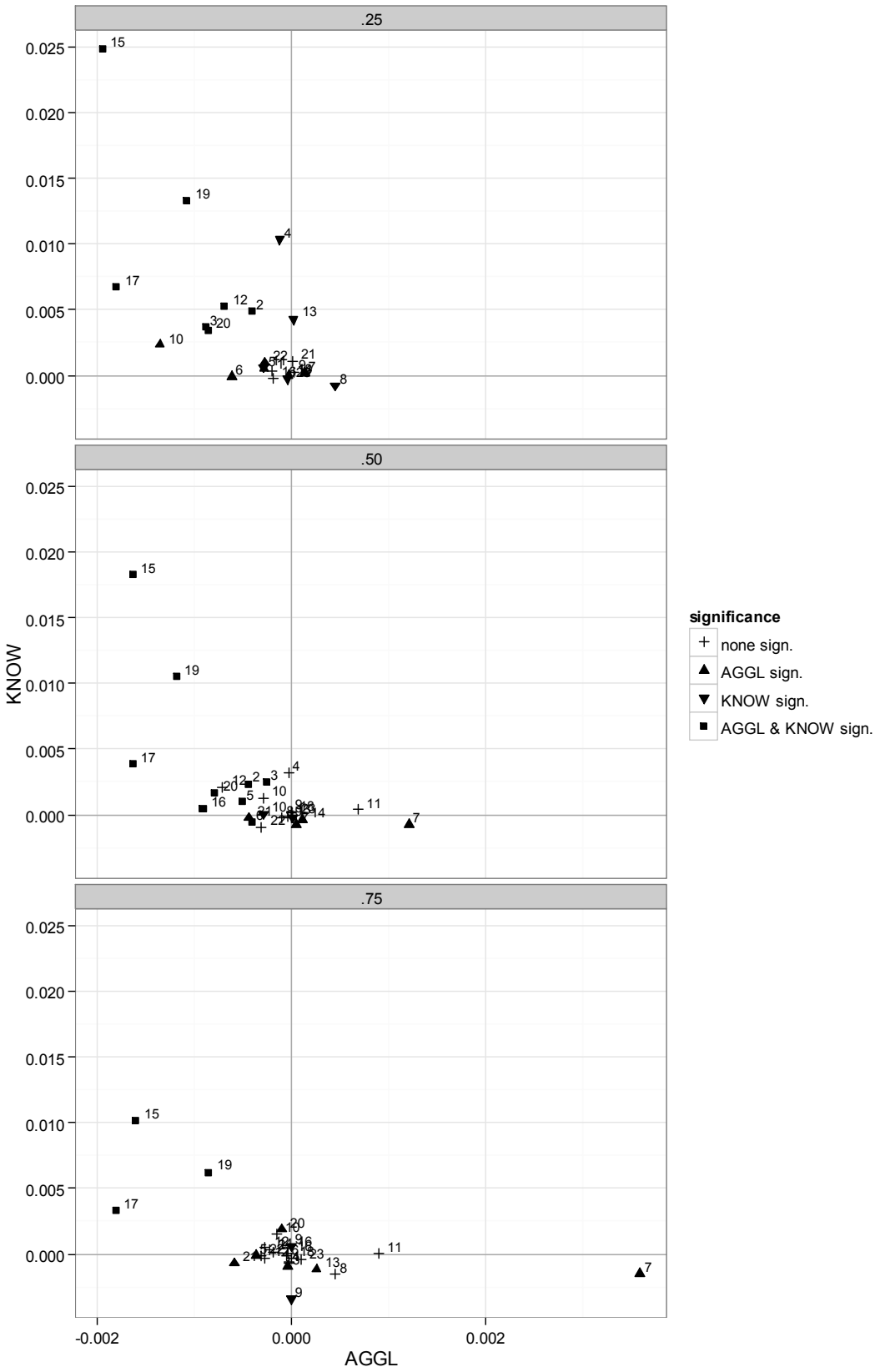


Figure 25: Scatterplot of coefficients β_1 (*AGGL*) and β_2 (*KNOW*) for different quantiles. Significant results (at 5%) are indicated by the symbols.

In section 5.2, it is hypothesized that the actual impact of *AGGL* and *KNOW* on firm growth should depend on the current stage of the industrial life-cycle, which can be roughly approximated by the average age of the industries' firms. Correlating this age measure with the corresponding regression coefficients reveals that the older the industry, the less relevant becomes *KNOW* (see Table 18). One the one hand, this might be explained by a decreasing importance of inputs from public research institutes and universities in more mature industries (BRENNER/SCHLUMP 2011), as firms tend to rely more on internal research activities. On the other hand, these firms are increasingly able to invest in and to manage cooperations with the best possible partners, irrespective of the geographic location. Hence, firms of science-based, yet more mature industries, like chemistry, metal manufacturing or production technology, do not benefit from nearby publications. In contrast to our expectation, no significant relationship is observed for the industry's age and *AGGL*.

Table 18: Pearson's correlation coefficients of β_1 (*AGGL*) and β_2 (*KNOW*) with the industries' age.

	$\theta_{0.25}$	$\theta_{0.50}$	$\theta_{0.75}$
β_1 (AGGL)	0.158 (0.461)	0.292 (0.156)	0.046 (0.831)
β_2 (KNOW)	-0.400 (0.041)	-0.361 (0.082)	-0.3876 (0.056)

p-values in parentheses

Although some general patterns are identified, a high heterogeneity apparently exists at the industry-level: firms of the various industries are not affected by the economic and scientific landscape in the same way. For illustrative reasons, three industries are analysed more in-depth: medical devices, distribution, and textiles.

The medical devices industry is primarily composed of knowledge-intensive activities. Not being a mature industry yet, locations in proximity to scientific publications increase the firms' likelihood of high growth events. The downside of innovation-driven growth becomes visible at $\theta_{0.25}$, as these firms are simultaneously more prone to decline. The impact of *KNOW* is most pronounced at a spatial scale from 1 to 2 hours, indicating that these effects would have been underestimated by considering regional boundaries that are too narrowly defined. *AGGL* becomes significantly negative at $\theta_{0.75}$ – the upward-pressure on wages due to stronger competition in industrial clusters makes it for incumbent firms in the medical devices industry more difficult to strongly grow by hiring many new employees.

Firms belonging to the distribution industry, which is widely composed of wholesale activities, are not affected by *KNOW*. Similar to other service industries, like business services, entertainment, media & publishing or tourism & hospitality, firms do not rely on scientific input. However, *AGGL* matters all the more: irrespective of the conditional growth rate quantile, firms from distribution strongly benefit from a high agglomeration of own-industry employment. Whilst the impact is most distinctive at short distances of 30 and 21 minutes for $\theta_{0.25}$ and $\theta_{0.5}$, respectively, highly positive growth events, by contrast, are visible at a larger spatial scale of around two hours.

Finally, the textiles industry provides an interesting example in which more than one spatial scale matters at the same time. Looking at the lower quantile of $\theta_{0.25}$, *KNOW* has a positive impact at a distance of 29 minutes, but a negative one at around three hours, implying that different mechanisms might work simultaneously at different spatial scales. Translating this finding into the managers' perspective: optimal location choices require complex spatial multi-level decisions.

5.6 Conclusions

FRENKEN et al. (2011) have suggested for future research that one of the main challenges "lies in settling contradictory empirical findings. In particular […] the main gap in our empirical understanding concerns the effect of localization economies on firm performance, which some may even consider the key question in economic geography at large". In line with these authors, this paper argues that contradictory empirical findings are closely related to the heterogeneity of firms and industry they belong to, and of the spatial economic landscape they are located in.

This paper takes the call for a finer resolution seriously. Several methodological choices are made to account for the omnipresent heterogeneity. First, the approach is micro-geographic in nature, as both, the firms and sources of agglomeration economies are geolocated in space. The unevenly distributed infrastructure which determines the accessibility to these growth relevant external sources is modelled via travel times in the road network, and behavioural assumptions of spatial interactions are reflected by the log-logistic distance decay function. Moreover, distance-based methods at the micro-geographical level make choices regarding the definition (of the existence) and spatial boundaries of industrial clusters obsolete. As the spatial scale and configuration differ strongly from cluster to cluster (for instance, see SCHOLL/BRENNER 2012 on the spatial clustering of the microsystem technology firms in Germany), we argue that the cluster concept is most meaningfully implemented in cross-sectional, comparative studies by methods which do not rely on a pre-defined regional (mostly administrative) aggregation level. Second, our cluster concept distinguishes related economic activities from knowledge generating activities by measuring own-industry employment as well as scientific publications in each municipality. Here, the identification of the best fitting decay function specification is performed for both variables separately, as the "relevant spatial level and spatial decay may well be different for different mechanisms underlying localization externalities" (FRENKEN et al. 2011: 21). Results show that the spatial impact of agglomeration effects are in some cases a sub-regional phenomena, whilst in other cases they transcend traditionally defined regional boundaries. Third, quantile regression techniques shed light on differences in the relationship between highly growing and declining firms and agglomeration economies. As theorized by HOOGSTRA and VANDIJK (2004), the spatial surrounding is more influential for firms at the tails of the conditional growth rate distribution. Finally, the disaggregated level of industries accounts for heterogeneity in the underlining technologies and differences along the life-cycle stages. Both aspects have increasingly gained attention in the recent literature. Our results confirm the existence of differences and particularities when comparing agglomeration economies

systematically across industries and support the idea that especially the relevance of external scientific knowledge depends on the industry's age and, hence, on its stage in the life-cycle.

Despite the high flexibility of the modelling assumptions, a rather coherent picture of the effects of industrial clusters on the growth performance of incumbent firms emerges. Firstly, being located in agglomerations of related economic activities does not stimulate or even hamper the firms' growth prospects in most industries. This finding is in line with the literature that acknowledges positive cluster effects on start-ups (e.g. SORENSEN/AUDIA 2000), yet negating the benefits for growth of incumbent firms (FRENKEN et al. 2011). Secondly, proximate knowledge generating activities, which more directly reflect the specific mechanism of knowledge spillovers, tend to be positively related to firm growth in industries which rely on external scientific knowledge, i.e. knowledge-intensive industries in the early phases of their life-cycles. Exceptions of these general patterns exist, and often can be explained by taking a closer look at properties and particularities of the respective industries.

The results have important implications for both policy makers and firm managers. The growth dynamics of incumbent firms is generally seen as a key driver for the creation and elimination of job opportunities in regional labour markets. In most industries, policy measures that foster further clustering of economic activities are confronted with prevailing diseconomies of agglomeration. Depending on kind of industrial specialization, scarce resources might be more effectively invested into activities of universities or public research institutes. However, due to increased competition, wages and, thus, higher value added and taxes, it is the region and the people that benefit most from the agglomeration of economic activities, making industrial clusters a relevant policy issue. But also firm managers might improve their growth prospects and the related market expansion objectives by taking the economic environment (at different spatial scales) into account when deciding on the headquarter location.

Like BEAUDRY and SWANN (2009), this paper focuses on the differences of agglomeration economies among industries. In future research, one could explore more systematically the reasons behind the industry-specific differences, for instance by employing measures on the knowledge properties of the industries or by disentangling the effects and mechanisms in comparative case studies. Complementary hereto, an emerging body of literature investigates the differences which can be observed among firm-specific categories, like their size or age (e.g. DUSCHL et al. 2014a, NEFFKE et al. 2012, RIGBY/BROWN 2013). Hence, we would motivate for a stronger integration of these related streams of literature.

6 Firm dynamics and regional resilience: an empirical and evolutionary perspective

Abstract: This paper breaks down the distributional analysis of firm growth rates to the domain of regions. Extreme growth events, i.e. turbulences at the level of firms, are conceptualized as an indicator of competitive regional environments which enable processes like structural adaptation or technological re-orientation. This provides a micro-funded empirical framework to operationalize the evolutionary dimension of regional resilience. Therefore, the flexible Asymmetric Exponential Power (AEP) density is fitted to firm data for each German region during the years of economic downturn (2008-2010). Results show that firm level turbulences are more pronounced in regions with a higher aggregate growth performance, with a highly qualified workforce and with more unrelated variety in the industrial portfolio.

Keywords: regional resilience, firm growth, growth rates distributions, fat tails, asymmetric exponential power, evolutionary perspective

JEL Classification: R11, L16, C46

6.1 Introduction

The concept of regional resilience, which has become popular among scholars and politicians who are concerned with the development of regions, seeks to understand how regional economies are able to cope with disturbances like the recent economic crisis (MARTIN 2012). Empirical attempts to measure regional resilience usually look at changes in some regional indicator, say, unemployment rate, employment or income level, to assess the impact of an external shock on a regional economy's growth path. Besides a vast amount of case studies of individual regions, only few systematic cross-sectional studies exist (e.g. CHAPPLE/LESTER 2010, FINGLETON et al. 2012). In most of these studies, an equilibrium-oriented neoclassical notion is prevalent, in which the immediate effect on the growth path of a region or the time to recovery is analysed. On the contrary, the evolutionary perspective on regional resilience is strongly related to processes like adaptation and structural re-orientation. These processes work at the level of firms, which FRENKEN & BOSCHMA (2007) call the true agents of change. For instance, new technological trajectories have to be explored by the region's firms, leading to "turbulence underneath the big calm" (DOSI et al. 2012). Hence, firm dynamics are intrinsically related to regional resilience. In particular, extremely growing as well as extremely shrinking firms indicate a region's potential to de-lock from its old paths and to re-invent itself. Hence, the questions arise how the way firms grow differ across regions and which region-specific factors are related especially to the occurrence of extreme growth events and, consequently, might contribute to regional resilience from an evolutionary point of view.

To answer these questions, this paper breaks down the stochastic analysis of firm growth rates to the domain of regions. Extreme growth events are more likely to occur than a normal distribution would suggest. Meanwhile, it counts as a stylized fact that growth rate distributions show fat tails and often an asymmetric shape. Therefore, the flexible Asymmetric Exponential Power density (AEP) is fitted to firm data for each German region during the years of economic downturn (2008-2010). Peculiarities of employment growth are explicitly taken into account to improve the quality of estimation. The estimated parameters measuring the tails' fatness are then related to various region-specific factors that are discussed in the literature on regional resilience. This approach provides a framework to operationalize the evolutionary perspective of regional resilience.

Results show that firm growth rate distributions, albeit remaining asymmetric and fat tailed at the spatially disaggregated level, markedly differ across regions in their shape. Extreme growth events are found to be more likely to occur in regions which show a stronger aggregate performance and have a higher number of employees with university degree. The latter confirms GLAESER et al. (2011) who ascribe the workforce skill a key role in making regions resilient. Furthermore, the fatness of the tails depends on the regions' industrial structure. In line with evolutionary economic geography, variety seems to be crucial that new technological trajectories can unfold and old ones decline (CASTALDI et al. 2014, BOSCHMA 2014).

The remaining paper is structured as follows. Section 6.2 bridges three different streams of literature. Firstly, it shows how the theoretical concept of regional resilience can be enriched by considering the heterogeneous responses and dynamics at the level of firms. Secondly, it extends the literature of firm growth rate distributions to the domain of regions,

allowing for new insights by systematically comparing other moments than the distributions' mean as it is the case in ordinary regression approaches. Thirdly, it argues that additional measures have to be taken into account when fitting the AEP distribution function for employment growth, which is, strictly speaking, discrete in nature. Finally, the main propositions to be investigated are presented. Section 6.3 discusses both the firm level and regional level data, while section 6.4 introduces into the AEP distribution function and the estimation procedure. After setting up a regression model to explain regional differences in the distributional parameters, section 6.5 discusses the results. Section 6.6 concludes.

6.2 Literature

6.2.1 Regional resilience – insights from growth processes at the level of firms

The theoretical concept of resilience, which originates from studies of ecological dynamics and socio-ecological systems (WALKER/SALT 2006), is increasingly applied to the domain of regional economic development, in particular to analyse responses of regional economies to major recessionary shocks. Being a rather broad and multifaceted concept, MARTIN (2012) identifies four dimensions of regional resilience: resilience as resistance, recovery, structural re-orientation and renewal or resumption of a growth path. The first two roughly correspond to the concepts of engineering resilience, which focuses on the resistance of a system to disturbances and the speed it takes to return to its pre-shock state, and ecological resilience, which analyse the magnitude of shocks that can be absorbed before the system changes form, function or position (HUDSON 2010). The two latter dimensions provide an evolutionary perspective. Resilience as a dynamic process can mean the ability of a regional economy to reconfigure and adapt its structure in order to maintain an acceptable growth path or the ability to create novel variety in response to external shocks (MARTIN 2012). The creation of novelty, the "ultimate cause of endogenous change" (WITT 2008), is ascribed a prominent role. However, the ability of a region to permanently re-invent itself might be hampered by the regional socio-economic conditions (GILBERT 2012). In this context, disturbances, which are often reinforced by recessions, can have positive effects by releasing potential for structural adaptation: "A deep recession may sweep away outmoded and unproductive activities, the removal of which opens up opportunities for the development of new sectors and a new phase of growth" (MARTIN 2012: 11).

Already REGGIANI et al. (2002: 215) distinguished between the robustness and changeability of a system: "resilience points out the 'possibility to change', while stability emphasises the 'impossibility to change'". Empirical studies (e.g. DAVIES 2011, DIODATO/WETERINGS 2012, or FINGLETON et al. 2012), on contrary, find it rather difficult to go beyond resistance and recovery: an evolutionary view can be hardly conceptualized by remaining at the level of regions and industries. Evolutionary processes work at the level of firms, the true agents of change (FRENKEN/BOSCHMA 2007). These processes are hidden by spatial or industrial aggregation. For example, new technologies or business models emerge within firms, but are reflected only decades later in the static industrial

classification scheme. Following the Schumpeterian notion of innovation, successful firms are often accompanied by a comparable or even higher share of firms that stagnate or are on their way to demise, but are averaged out in the spatial or industrial aggregation process. As MARTIN (2012) states, a key to understand regional resilience is the analysis of reactions and adjustments of firms, which ultimately drive the development of regional economies. The (un)successful adaptation of firms translate into different economic performances, resulting in firm-level turbulences, which are often reinforced by external conditions and might even push the whole regional economic system beyond thresholds of bifurcation points to new stability domains. Essentially, such turbulences are required to de-lock from path dependencies (SIMMIE/MARTIN 2010). In the context of the recent crisis, ARCHIBUGI et al. (2013) uncovered that this crisis led to a concentration of innovative activities within a small group of fast growing new firms and those firms already highly innovative before the crisis. They argue that during phases of economic uncertainty about technological trajectories, demand conditions and new market opportunities, the exploration of new products and markets become even more important. Firms engaging in this risky strategy will be more exposed to either success or failure. In other words, turbulences, manifested by extreme positive and extreme negative growth events at the level of firms, are the driver of evolutionary development processes and the related adjustment to recessions. As SETTERFIELD (2010: 8) put it, "extreme experiences that propel a system sufficiently far from its current state are thought to result in structural change". Focusing on the frequency distribution of firm growth rates, fat tails can be conceptualized as an indicator of highly dynamic, vibrant and re-inventing regional economies.

Arriving at the level of firms, one has to ask which role the firms' location plays in shaping their dynamics. BARBOSA and EIRIZ (2011) distinguish between two ways of how the spatial dimension matters, namely by general environmental factors, which are external to firms, unevenly distributed across space and imperfectly mobile, and by the role of proximity. The latter has recently become popular in studies analysing the effects of industrial agglomerations, although its roots can be traced back to MARSHALL's (1890) trinity of external economies. Mainly focusing on spatially bounded knowledge spillovers, empirical studies have analysed the impact of agglomerations, industrial clusters, access to universities, or regional innovation systems, with the notion of interactive learning and innovation, on the growth of firms (see FRENKEN et al. 2011 for a survey on the empirical literature). These studies also indicate that the regional dimension is especially relevant for highly shrinking and expanding firms (e.g. FORNAHL/OTTO 2008, DUSCHL et al. 2014a). Briefly stated, the way how firms grow is often shaped by factors external to them, but internal to regions.

As noted above, the dynamics at the level of firms are crucial to understand processes and outcomes at the level of regions. However, the responses of firms to economic crises, which might be mediated or constrained by regional factors (HOOGSTRA/VANDIJK 2004), are far from straightforward: "Some firms prosper in recessions while others fare very badly" (GEROSKI/GREGG 1996: 551). What these authors call selectivity is a ubiquitous and persistent heterogeneity of firms and their responses, letting DOSI et al. (2010) make a plea for considering entire distributions instead of averages for assessing the relationship between the micro and the macro. This is confirmed by HIGSON et al. (2004) or HOLLY et

al. (2013), who find that firm growth rate distributions change systematically over business cycles and even contribute to shaping macroeconomic fluctuations. An example may illustrate what firm dynamics reveal about the resilience of a regional system. Two regional economies A and B, both with an unchanging number of total employees, might show quite different dynamics at the level of firms. In A, not any single firm is growing, whereas B is confronted with turbulences due to many shrinking and expanding firms. Even though the short-term effect on regional growth is equal, the long-term outcomes of both regions probably will differ: B seems to be more able to reconfigure its structure and to adapt to changing environments, hence ultimately exploiting new technological opportunities. I argue that the evolutionary dimension of regional resilience is hidden in the dynamics of firms and thus can be uncovered by analysing their distribution of growth rates. Hereby, the average growth rate has little to tell, because it obscures turbulences, as the example has demonstrated. Besides, most firms are able to withstand external forces and even remain unaffected at all by macroeconomic recessions. Only after exceeding a certain threshold, some firms are badly hit, while others might benefit strongly from such path-breaking crisis. Put differently, the secret lies in the tails of the distribution.

To conclude, this paper argues that turbulences at the level of firms allow for a first assessment of a region's long-term ability to adapt its structure and to re-invent itself, key aspects of evolutionary regional resilience, especially so but not exclusively during times of economic crises.

6.2.2 The distribution of firm growth rates – a regional perspective

The previous section argues that the entire distribution provides a more complete picture than single moments like the arithmetic mean. Although knowing the distributional form of a specific phenomenon is a valuable insight by itself, further information is revealed by comparing it to a reference distribution. As such, three options are conceivable.

First, empirically estimated distributions can be used to verify expectations derived from theoretical models. GIBRAT's (1931) "law of proportionate effect", meanwhile a common starting point in the literature of industrial organizations, requires in its strong version normal distributed growth rates (AMARAL et al. 1997). But instead of a bell-shaped normal curve, a tent-like shaped exponential one, also known as the Laplace distribution, is observed on basis of firm level data from several countries (e.g. STANLEY et al. 1996 for US, BOTTAZZI et al. 2002 for Italy) and at the disaggregated level of industries (BOTTAZZI et al. 2001 for the pharmaceutical industry). To account for this stylized fact, stochastic models of firm growth have been developed (e.g. FU et al. 2005, BOTTAZZI/SECCHI 2006, COAD 2012). Subsequently, empirical evidence was also reported for the income growth rates of countries, that is, at a much higher level of economic aggregation (e.g. LEE et al. 1998, AMARAL et al. 2001, MAASOUMI et al. 2007, FAGIOLO et al. 2008). Only recently, a research gap at the intermediate level of industries (CASTALDI/SAPIO 2008) and regions (DUSCHL/BRENNER 2013) was filled, indicating that these models can be generalized to hold for the growth of all complex economic organizations irrespective of the level of aggregation. Besides, empirical evidence is emerging that these distributions show tails significantly fatter than the Laplacian ones and often an asymmetric shape (e.g. BOTTAZZI

et al. 2011, DUSCHL/BRENNER 2013, REICHSTEIN/JENSEN 2005), motivating BOTTAZZI/SECCHI (2011) to introduce the even more flexible five-parameter AEP. This refined evidence still waits to be explained by new stochastic models more accurately.

Secondly, distributions can be conditioned on further variables. Ordinary regression models represent a specific case of this general prediction problem by focus on one specific point of the conditional distribution, like the mean or any other quantile (VARIAN 2014). By comparing the shape of the conditional and unconditional distribution, a more complete picture of the impact of the conditioning variables on the distribution can be drawn. For instance, this exercise is found in BOTTAZZI et al. (2014) or MAASOUMI et al. (2007: 499). The latter authors note already that in the residual growth rates "control is only achieved on the mean of the growth rates, and the variables may continue to impact on other distributional characteristics". A first approach to capture the complex shifts in the entire growth rate distributional mass is introduced by DUSCHL and PENG (2014), who simultaneously condition all five parameters of the AEP density.

Thirdly, distributions of the same type of observational unit can be compared by taking empirical snapshots in various contexts. Growth rate distributions of firms belonging to different industries (e.g. BOTTAZZI/SECCHI 2003) are used to confirm that predictions from stochastic models survive at different levels of industrial disaggregation. Introducing the time dimension, DOSI et al. (2012) point out the heterogeneous impact of the Euro adoption on the performance of Italian manufacturing firms. At the regional level, BARBOSA and EIRIZ (2011) investigate the way how firms grow by visually comparing the evolution of the firm size densities of 19 Portuguese regions. The paper at hand aims to uncover the region-specific factors leading to differences in the distributional characteristics of the regions' firm growth rates by relating them to the estimated parameters. It focuses particularly on explaining the tails of the distribution, which are a measure of internal turbulences and, as such, an indicator of resilience from an evolutionary perspective.

Each of the three strategies has its own advantages in highlighting specific aspects of a complex phenomenon. The latter approach of systematically comparing estimates of distributional parameters is particularly insightful in situations in which a high number of comparable samples, say, the regions of a country, are available. By letting the data speak, it is also a response to difficulties of theoretical models and simulation studies, which are confronted by the existence of a hypothetically unlimited number of economic mechanisms that may be able to explain the emergence of fat-tailed distributions (ALFARANO/MILAKOVIC 2008).

6.2.3 Estimation of the AEP density – specificities of employment growth

Up to here, nothing has been said about the various measures of firm growth, like sales, turnover, productivity or employment, all governed by distinct mechanisms (COAD 2009). Whereas the first three primarily concern business managers, employment growth is of utmost relevance for regional policy makers, even more so during times of economic crisis (MARTIN 2012). All measures share the common property that the underlying change events are not, strictly speaking, continuous. An inherent discreteness becomes

particularly apparent for employees, which are by nature indivisible. COAD (2012: 17) takes this reasoning seriously by putting the reactions of firms to growth stimuli at the heart of his stochastic model: "The lumpy nature of resources within a firm implies that firm expansion is characterized by non-constant marginal costs that depend on the degree of utilization of the firm's resources". Consequently, fat tailed distributions of growth rates emerge as firms tend towards a critical state of full utilization of resources: if resources are already more or less fully employed, then growth will only be possible with addition of extra resources, while the "interdependent nature of discrete resources may lead to triggering off of a series of additions to a firm's resources". The resulting growth process might show non-linearities as firms add indivisible resources to arrive at efficient levels of production.

The incentive to exploit unused resources provides an intentional perspective on growth, in contrast to GIBRAT's law and island models (e.g. IJIRI/SIMON 1997, SUTTON 1998, BOTTAZZI/SECCHI 2006), in which growth opportunities are passively absorbed and accumulated (COAD 2009). Taking both perspectives into account, a conceptual two-step firm growth model is proposed, which disentangles the outcome of a change in the number of employees from the actual growth processes. Put simply, in a first step each of the N firms is confronted with the options to grow or not grow based on its internal resource composition and the external business opportunities. Both options can be modelled as a binomial process with probability p, in which $p*N$ firms change the number of employees and $(1-p)*N$ firms remain at the previous level. In a second step, all of those firms which experience such kind of growth impulse due to the mismatch between opportunities and their level of resources try to respond by integrating new resources or by releasing existing ones. This ultimately leads to h expanding and k shrinking firms. The remaining $p*N - (h+k)$ firms are those which would grow but are not able to, and thus delaying their growth momentum.

Leaving aside the question whether or not both steps represent analytically distinguishable phenomena, they are appropriate from a stochastic point of view. Assuming a continuous probability distribution function, the realization of a specific empirical value occurs in the limit with zero probability. This clearly contrasts with the observed occurrences of zero-growth events in typical databases – in Bureau van Dijk they amount for up to 50% of all events, and COAD and HÖLZL (2009) even report that 65% of small establishments listed in the Austrian Social Security files do not display any changes in employment from one year to the next. This abundance of zero-growth events calls for an explanation. Firstly, following the proposed model, firms simply may prefer to remain at the previous level of employees. This can be an economically rational choice in absence of any changes in business opportunities, but it can also be the preferred choice in cases when opportunities have changed. To name just a few examples, firms might be reluctant to expand because the inclusion of new employees is costly as it implies re-organisation of internal tasks and management functions, the labour market might only insufficiently meet the demand for (qualified) workers, or the fear of losing control might frighten some managers (COAD 2009). In a similar vein, firms can be reluctant to shrink despite reduced business opportunities. Firms invest in building up redundancies in difficult times instead of immediately dissolving existing working contracts, or managers might be not fully aware of the necessary down-sizing. Secondly, there are those firms which would grow but are not able to due to the discrete nature of employees, which inhibits these firms from marginally

increasing or decreasing their size by, say, a quarter of employee. Instead, these firms tend to respond by re-organizing tasks internally. Although it is still an issue for actually growing firms, it becomes statistically less and less relevant as the number of employees to change increases. Around zero, however, the discreteness is fully noticeable with respect to the firms' ability to grow. Thirdly, firm databases sometimes lack a regular updating of their entries, resulting in many zero-growth events as simple extrapolations. To sum up, zero-growth events arise due to the choice of avoiding growth, the inability to grow and data problems. The latter makes it impossible to recover p, and hence restricting the analysis to the actual growing firms. However, an entire exclusion of zero-growth rates would bias the estimation of a continuous probability distribution function because those firms not able to grow due to the discrete nature of employees would be dismissed. Therefore, this paper deals with a new method for estimating the AEP to account for possible biases resulting from the discrete nature of employees.

6.2.4 Main propositions

Placing dynamics of firms' employees at the heart of regional resilience, two general propositions can be stated.

> **Proposition 1**: *A region's firm growth rate distribution is asymmetric and fat tailed, but its exact shape, especially its tail behaviour, differs across regions.*

The first proposition arises from the empirical observations which show that the way how firms grow depend on factors and conditions specific to the region they are located in (see section 2.1). Although the stylized facts on the general shape of firm growth rate distributions are expected to be manifest already with regions as the reference system, important differences across regions might be observed, especially for the fatness of tails and the degree of asymmetric. On the one hand, aggregate shocks, which tend to be differently pronounced across regions, affect the entire distributional shape in a complex way. For instance, HOLLY et al. (2013) find that especially the left hand side of the distribution is responsive to economic shocks. On the other hand, the mechanisms leading to the fat tails, like the increasing returns in the model of BOTTAZZI and SECCHI (2006), can vary in their extent across regions. This is strongly related to the idea of localization economies (e.g. KRUGMAN 1991), which might foster this self-reinforcing accumulation of growth opportunities of the region's firms, and thus contribute to the emergence of fat tails. However, it is only one thing to study regional differences in the distributional parameters, but quite another thing is to ask about the economic meaning of these differences:

> **Proposition 2**: *The tails' fatness depends on regional factors, which provides a first assessment of the potential for evolutionary resilience.*

This proposition needs to be further elaborated. The underlying firm dynamics of a regional economy reveal whether it is resilient from an evolutionary view. As CAPASSO et al. (2013: 612) state it: "A major economic implications of heavy tails is that fast-growing and shrinking firms account for a non-negligible share of an industry population and significantly affect the industry dynamics". Positive fat tails indicate the ability of a regional system to adapt its structure: extreme positive growth events result from the exploration of

new technologies or business models, which creates new opportunities to spur growth. The other side of the coin is that evolution driven by the creation of new variety implies that existing modes of activities become outdated. But only in competitive regional environments, firms unable or unwilling to adapt will perform worse, thus increasing the likelihood of extreme negative growth events. In short, a regional economy with fat tails on both sides of the firm growth rate distribution is assumed to have a higher adaptive capacity. This is defined as *resilient* from an evolutionary perspective.

The number of positive and negative extreme growth events that a region encounters might not necessarily be balanced. If the asymmetric distribution is skewed towards the left than the negative events outweigh the positive ones. In this case a regional economy is considered to be more *vulnerable*, as its firms are more sensitive to the negative effects of a shock, without having an equally large fraction of firms that prosper.

6.3 Data

6.3.1 Firm level data

This paper analyses firm dynamics and compares their distributional properties across different regions. Firm level data are retrieved from the BvD Amadeus database in early 2012. It provides the most comprehensive data entries for the time period from 2007 to 2010, which roughly concurs with the years of macroeconomic downturn. However, it is not free of data problems. For instance, zero growth rates make up 44.5% of all entries. Although excluded from further consideration, they still could bias the results insofar as growth rates in the subsequent year must be artificially higher. Several heuristics are applied to identify zero growth events stemming from data inconsistencies based on extrapolation, and the subsequent non-zero growth rates are eliminated.[39]

The Amadeus database discloses the address of the firms' headquarter location. As operational and strategic decisions are often made within this organizational unit, their regional environment will be most decisive in affecting their growth prospects (BEAUDRY/SWANN 2009). This rationale breaks down for larger firms, which tend to be less focused on their headquarters, but disperse their activities in many establishments across the country and beyond. Therefore, the analysis is restricted to firms with no more than an annual average of 1000 employees. Also very small firms with less than 5 employees, which growth processes are known to be rather erratic and which have limited abilities to generate jobs, are excluded (COAD 2009). Furthermore, industries are affected differently by macroeconomic recessions. Following PORTER (2003), traded cluster industries can be distinguished from local cluster, resource-based cluster and non-cluster industries. This paper focuses exclusively on firms from traded cluster industries, e.g. firms from traded industries that tend to co-locate. For Europe, these industries are defined within the EU

[39] For example, zero growth in both employment and turnover in the same year indicate that data were simply adopted from the past year, probably due to the lack of updated information. In total, about half of the zero growth events are identified as extrapolations.

Cluster Observatory Project. The motivation for focusing on this subset of firms is twofold. On the one hand, traded-cluster industries are more exposed to the global economy just as they depend on their regional environment, and on the other hand, they are expected to be more influential in shaping long-term technological trends within their home region. Although they account for less than half of the employment, they "register much higher wages, far higher rates of innovation and influence local wages" (PORTER 2003: 549).

In total, 37403 growth events, different from zero, are analysed. These growth events stem from 20962 firms, which are spatially distributed across German labour market regions.

6.3.2 Regional level data

Labour market regions as defined by ECKEY et al. (2006) serve as the regional reference space for firm locations. Combining insights from empirical studies of firm growth and regional resilience, several region-specific variables that might be related to the underlying micro-dynamics are identified. These variables can be classified into four broader categories: the region's 1) general socio-economic conditions, 2) innovation conditions, 3) workforce qualification, and 4) industrial structure.

Being an obstacle to the ability to adapt, unfavourable general socio-economic conditions are expected to reduce regional resilience. These are approximated by the population density (*PopDensity*), the unemployment rate (*UnemplRate*) and the aggregate regional growth performance (*RegGrowth*). *PopDensity*, being rather independent from the surrounding industrial structure, reflects urbanization economies (BUERGER et al. 2012). The *UnemplRate* indicates the vitality of the regional labour markets. In the special case of Germany, it also accounts for structural differences along the east-west and north-south divide. Data for both variables is obtained from the German Federal Statistical Office (destatis). Finally, *RegGrowth*, the (logarithmic) change of regional employment in the study period from 2007 to 2010, measures the aggregate growth performance of a region's economy during the time of macroeconomic recession. Better performing regions are expected to be more able to reconcile turbulent processes at the level of firms. Like all subsequent variables based on employees, data is retrieved from the German Institute of Employment Research (IAB).

The innovation conditions might directly measure a region's "ability to replace declining or uncompetitive activities with new, dynamic and competitive ones" (FINGLETON et al. 2012). Innovativeness is measured by the university third party research funding (*ResFunding*). Here data is obtained from destatis. Alternative measures, like universities' budget, patents or employees in R&D-related occupations, were tested beforehand but showed an inferior fit compared to *ResFunding*. Due to multicollinearity issues, they are omitted from further analysis.

Among researchers (e.g. CHAPPLE/LESTER 2010, MARTIN 2012 and HILL/ATKINS 2012) it is widely acknowledged that the region's workforce skills are a key factor for regional resilience: "human capital and urban reinvention" are strongly connected, making skills "particularly valuable in places that are hit with adverse shocks" (GLAESER et al. 2011: 4).

The regional qualification level is measured by the number of employees with university degree (*EmplUniv*).

The recent macroeconomic recession has also revealed that the region's industrial structure matters: different industries were affected differently (GROOT et al. 2011, DAVIES 2012). Two variables control for the share of observations which are associated to the *Manufacturing* and *Construction* industries. The region's share of manufacturing is also a proxy for export orientation, and hence measures the exposure to global markets (CHAPPLE/LESTER 2010). Besides, it is often observed that localization economies are more relevant among manufacturing industries (BEAUDRY/SWANN 2009), which might lead to fatter tails according to the model of BOTTAZZI and SECCHI (2006). During the last recession, especially the construction industry was targeted with fiscal stimuli and hence might show different growth dynamics.

However, it is not only the concentration of certain industries that is expected to matter, but also how the economic activities are technologically related to each other (see BOSCHMA 2014 for an extensive discussion of the role of industrial variety and regional resilience). A high degree of relatedness means the existence of many inter-industry technological linkages and interdependencies (BOSCHMA/IAMMARINO 2009). If the region's industries are too similar, i.e. the regional economy is diversified only to a small degree without showing much variety, few recombinatory options are available, which are essential to develop new growth paths (BOSCHMA 2014). In regions, in which strong industrial clusters have been established, the ability of a transition to new technological forms might be constrained by a higher inertia and myopia of its actors and institutions (GILBERT 2012). On the contrary, variety may enhance the region's adaptability by increasing the potential to make new recombinations (BOSCHMA 2014). Drawing on recent empirical evidence based on patent data (CASTALDI et al. 2014), it is argued here that especially unrelated variety, reflecting the presence of very different activities, increases the number of potential sources for technological breakthroughs, which often translate into the tails of the firm growth rate distributions. Thus, unrelated variety is expected to enhance the region's long term ability to renew its growth path. This is different from the often discussed portfolio effect of unrelated variety (FRENKEN et al. 2007), which is more a matter of immediate vulnerability than a matter of adaptability.

The following measures are adopted from FRENKEN et al. (2007), BOSCHMA and IAMMARINO (2009) and ERIKSSON (2011), who also provide a more detailed discussion. Regional similarity is constructed by inverting the entropy at the four-digit industry level, with $\mathrm{p}_{\mathrm{k}}^{\mathrm{IV}}$ the regional share of employment within four-digit industry *k* and N^{IV} the number of different four-digit industry classes:

$$Similarity = 1/\sum_{k=1}^{N^{IV}} p_k^{IV} log_2(\frac{1}{p_k^{IV}}) \tag{29}$$

The higher the similarity measure, the higher is the concentration of similar industries within the region. Assuming a hierarchical understanding of relatedness, related variety is defined as the weighted sum of the entropy measure at the four-digit level within each two-digit class, with $\mathrm{p}_{\mathrm{l}}^{\mathrm{II}} = \sum_{\mathrm{k} \in \mathrm{S}_{\mathrm{l}}^{\mathrm{II}}} \mathrm{p}_{\mathrm{k}}^{\mathrm{IV}}$ the respective regional share of employment within two-

digit industry *l* and N^{II} the number of different two-digit industry classes S_l^{II}, which nest the respective four-digit industries *k:*

$$RelVar = \sum_{l=1}^{N^{II}} p_l^{II} \left(\sum_{k \in S_l^{II}} \frac{p_k^{IV}}{p_l^{II}} log_2 \left(\frac{1}{p_k^{IV} / p_l^{II}} \right) \right) \tag{30}$$

The higher the related variety measure, the more technological complementarities in the activities exist within the region. Finally, unrelated variety is measured as the entropy at the one-digit level, with p_j^I the regional share of employment within one-digit industry *j* and N^I the number of different one-digit industry classes:

$$UnrelVar = \sum_{j=1}^{N^I} p_j^I log_2 \left(\frac{1}{p_j^I} \right) \tag{31}$$

The higher the unrelated variety measure, the more the region is diversified into technologically different activities.

All region-specific variables are calculated for the base year of 2007. The highly asymmetric distributed variables of *EmplUniv* and *ResFunding* are first normalized by division through their mean value and then made symmetric by the transformation $\tilde{x} = (x - 1)/(x + 1)$. Descriptive statistics and correlations are reported in the Appendix X.3.

6.4 Methodology

Growth rates are calculated by taking the difference of the natural logarithm of the firm size *S* between two successive time periods *t*:

$$g_{i,m,t} = log(S_{i,m,t+1}) - log(S_{i,m,t}) \tag{32}$$

where the subscript *i* indicates the respective firm and *m* the region in which the firm is located. The growth rates are then rescaled to control for the inverse relationship between their size and variance, a universal feature of the growth of complex economic organisations (AMARAL et al. 2001). Here, a similar rescaling procedure is used as in BOTTAZZI et al. (2014), which takes into account that the functional form of the relationship might be non-linear, as recently observed in the firm growth literature (BOTTAZZI et al. 2011). Because the scaling relationship might differ across regions, this step is performed for each region separately. Only after rescaling, growth rates can be interpreted as different realizations of the same underlying stochastic process. Its specification by a distributional model is the aim of this paper.

6.4.1 AEP density

In search for a more general and flexible distributional model that describes the empirical distribution of (rescaled) growth rates, the exponential power (EP) distribution family was introduced into economics by BOTTAZZI et al. (2002). Its density *f(x)* reads

$$f_{EP}(\tilde{g}; b, a, m) = \frac{1}{2ab^{\frac{1}{b}}\Gamma(1+\frac{1}{b})} exp\left(-\frac{1}{b}\left|\frac{g-m}{a}\right|^{b}\right) \tag{33}$$

with Γ(.) standing for the gamma function. Three parameters define the distribution: the location parameter *m*, which indicates the general trend in the data, the scale parameter *a*, which determines the spread or dispersion of the distribution, and the shape parameter *b*. Both the normal (*b* = 2) and Laplace (*b* = 1) are particular cases of the EP family of probability densities. It allows for a continuous variation from non-normality to normality, with a smaller shape parameter *b* representing fatter tails of the corresponding density. Furthermore, it can be extended to a five-parameter family of distributions, which is able to cope with asymmetries in the data. In addition to *m*, the asymmetric exponential power (AEP) distribution possesses two scale parameters a_l and a_r for the values below and above *m* and two shape parameters b_l and b_r describing the tail behaviour on the left and right side of the distribution:

$$f_{AEP}(g;\ b_l, b_r, a_l, a_r, m) = \frac{1}{C}\exp\left(-\left[\frac{1}{b_l}\left|\frac{g-m}{a_l}\right|^{b_l}\theta(m-g) + \frac{1}{b_r}\left|\frac{g-m}{a_r}\right|^{b_r}\theta(g-m)\right]\right) \tag{34}$$

where $\theta(g)$ is the Heaviside theta function and $C = a_l b_l^{1/b_l - 1}\Gamma(1/b_l) + a_r b_r^{1/b_r - 1}\Gamma(1/b_r)$ a normalization constant. This new class of AEP and related ML inference problems are discussed in details in BOTTAZZI and SECCHI (2011). By applying numerical simulations, they show that the bias of ML estimators can be safely ignored if N > 100, except for m, which in case of asymmetry, that means $b_l \neq b_r$, is often observed to be biased, even for very large samples (N > 5000). For the study at hand this implies that regions with less than 100 firm growth events are dropped, leaving 100 labour market regions out of to the initial 150. Besides, the potentially biased estimates for m are ignored, which anyway are not the focal point, unlike the tails of the distribution.

6.4.2 Contemporaneous left and right tail estimation

HILL (1975) has shown that in some cases it can be useful to make inference about certain parts of the distribution, in his case the tail, without assuming any global form of the distribution function. By exploiting properties of spacings of exponential order statistics, BOTTAZZI (2012) generalizes HILL's analysis to any continuous distribution. In the present paper, the AEP distribution is estimated blinding out the central part, which is delimited by the lower and upper threshold values $\underline{d}$ and $\overline{d}$. Based upon the empirical data, these values are set to -0.006 and 0.006, respectively, in order to maximize the gap around zero, while guaranteeing that all non-zero growth events x_u are still included. Ordering the empirical

observations of the sample by increasing size, only the *k* smallest (with $x_u < \underline{d}$) and *h* largest realizations (with $x_u > \overline{d}$) are considered. Conditioned on these threshold values, the likelihood function for estimating contemporaneously the upper and lower tails reads

$$L^{Tails} = \frac{N!}{(N-h-k)!}(F_{AEP}(\overline{d}) - F_{AEP}(\underline{d}))^{N-h-k} * \prod_u f_{AEP}(x_u) \tag{35}$$

where the AEP probability function F_{AEP} is integrable from the density f_{AEP} (BOTTAZZI/SECCHI 2011). The log-likelihood function can be deduced:

$$LL^{Tails} = \log\left(\frac{N!}{(N-h-k)!}\right) + (N-h-k)\log\left(F_{AEP}(\overline{d}) - F_{AEP}(\underline{d})\right) + \sum_u \log * f_{AEP}(x_u) \tag{36}$$

To account for those firms which prefer to grow but are not able to grow due to the discreteness of employees, *N* (here, *N:=p*N*) is additionally endogenized. Finally, expression (37) is to be minimized:[40]

$$\{b_l, b_r, a_l, a_r, m, N\} = argmin_{b_l, b_r, a_l, a_r, m, N}(LL^{Tails}) \tag{37}$$

Endogenizing *N* not remains without any consequences. Obviously, it reduces the estimation bias stemming from the discrete nature of changes in the number of employees. It does so by raising the competition of the countervailing forces of the scale and shape parameters, which both simultaneously try to account for (extreme) positive and negative events. Leaving out the central part, this flexibility regarding asymmetry increases, hence implying that higher peaks might be reached and distributional mass shifted from the variance to the tails. Based on empirical data from one arbitrarily chosen region, Figure 26 displays the main aspects and implications of this refined estimation procedure: in a) the empirical firm growth rate distribution is plotted. The points represent the midpoints of the frequency bins, using a log-scale on the y-axis. It is instantly visible that zero-growth events are dramatically over-represented. In b) all zero-growth events are removed. The parameters of the AEP distribution (light-blue line) fitted to the zero-cleaned data are biased, because some zero-growth events result from firms that would have grown if employees were not discrete in nature. To reduce this bias, in c) the same distribution (dark-blue line) is estimated by leaving out the central part around zero, which is coloured in black, and by making the number of observations lying within this part endogenous. This endogenization increases the number of actually growing firms *h* and *k* by around 4%.

[40] This formula is optimized using DEoptim in the R environment. Global optimization by differential evolution is especially "useful in situations in which the objective function is stochastic, noisy or difficult to differentiate" (MULLEN et al. 2011).

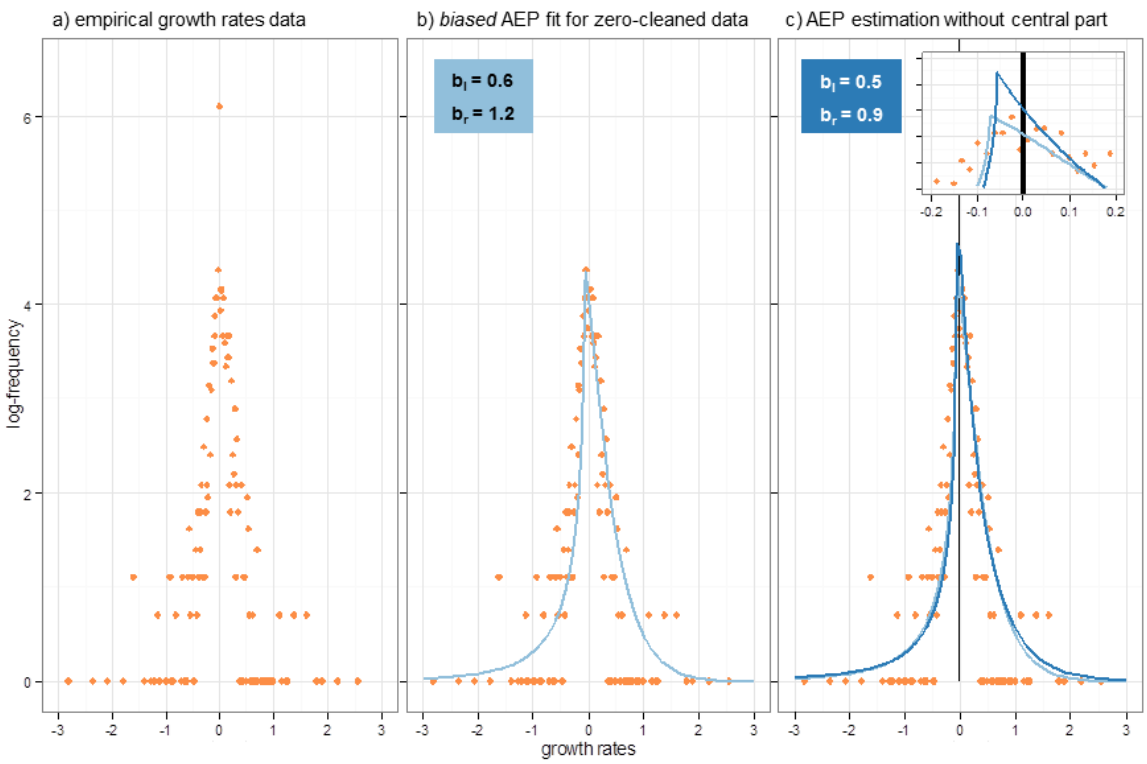


Figure 26: Comparison of estimation procedures

6.4.3 Regression model

In a next step, the distributional parameters, which are estimated for each region *m*, are related to regional factors in a regression model estimated by OLS. Turbulences arise through both positive and negative extreme growth events. Instead of explaining the fatness of the tails for both sides of the distribution separately, the two-dimensional space of the shape parameters is rotated such that the sum of b_r and b_l is finally explained. This sum represents overall turbulences that are expected to accompany processes like adaptation and structural re-orientation. Recall that smaller shape parameters mean fatter tails. Hence, the smaller the value of the sum of b_r and b_l, the more likely extreme events are to occur in a regional economic system. The other dimension in the rotated space, b_r minus b_l, measures the asymmetry of the distribution and indicates a kind of vulnerability: this value is positive for $b_r > b_l$, implying that extreme negative growth events are more likely to occur than positive ones.

Two models are estimated, one for *resilience* ($b_r + b_l$) and one for *vulnerability* ($b_r - b_l$). Each model contains two control variables: the number of firms in the sample and the respective opposite dimension in the rotated space. The former accounts for the issue that fat tails might be sensitive to extreme events in the case of just a handful of observations. The latter should capture distortions from the possible relationship between the *resilience* and *vulnerability* measures, say, if more vulnerable regions are at the same time more resilient, as it is often argued regarding the recovery dimension of resilience (e.g. MARTIN/SUNLEY 2014). The resulting models read:

Model 1: *resilience*

$$(b_{r,m} + b_{l,m}) = \alpha + \beta_1(b_{r,m} - b_{l,m}) + \beta_2 N_{firms} + \sum_u \beta_u x_u + \varepsilon_m \quad (38)$$

Model 2: *vulnerability*

$$(b_{r,m} - b_{l,m}) = \alpha + \beta_1(b_{r,m} + b_{l,m}) + \beta_2 N_{firms} + \sum_u \beta_u x_u + \varepsilon_m \quad (39)$$

with α and β representing the coefficients to be estimated, *u* indexing the regional variables *x*, and ε_m standing for a normal distributed error term.

6.4.4 Limitations

However, this empirical perspective on regional resilience, based on firm growth rate distributions, is not without any limitations. Especially three points deserve a further discussion.

Firstly, firms that enter and exit are omitted from the analysis, because they are qualitatively different from growth processes of existing and surviving firms. Growth rates become infinite when the size changes towards zero. However, entry and exit events, which are known to show different regional dynamics and determinants (COMBES et al. 2004), are an important aspect regarding all dimensions of structural change and regional resilience (NEFFKE et al. 2014, BOSCHMA 2014), but cannot be tackled within the framework of growth rate distributions. Yet the author beliefs that a huge bulk of the processes of adaptation and re-orientation occurs within existing firms and are thus revealed by their growth performance. This is confirmed by studies like BERGEK et al. (2013), arguing that the ability of new entrants to destroy and disrupt established industries is often overestimated, while the ability of incumbents to absorb and integrate new technologies with their existing capabilities is often underestimated. Besides, METCALFE and FOSTER (2010) argue that the effects of exit and entry on aggregate dynamics, by inducing structural change, operate at time periods much larger than analysed in this paper.

Secondly, instead of being a longitudinal approach, a cross-sectional snapshot of growth rates is analysed. Here, the data stems from the years of macroeconomic recession. However, a systematic comparison to pre- and post-crisis growth processes might be particularly interesting at the level of firms, as the temporal auto-correlation patterns of firm growth rates tend to be especially pronounced at the tails (COAD 2007).

This relates to the third point, as short term approaches do not allow for a direct analysis of the evolutionary dimension of regional resilience (DAVIES 2011). From an evolutionary perspective, regional resilience is not bound to describe the immediate reaction to shocks, but it is understood as an on-going process of adaptation and structural re-orientation. However, SIMMIE and MARTIN (2010: 34) argue that resilience "depends both on longer term, region-wide processes and on shorter term microscale processes and on how these interact". The latter can have permanent effects on the potential long term output (CROSS

et al. 2010). By focusing on yearly growth rates, at least a first assessment of the potential for long-term adaptability and structural re-orientation can be provided.

6.5 Results

6.5.1 Inter-regional heterogeneity in the firm growth rate distributions

Summary statistics of the estimated AEP parameters are reported in

Table 19. In average, the tails are to a considerable degree fatter compared to the normal and even Laplace distribution, accompanied by a significant asymmetric shape towards the left: extreme negative growth events are much more likely to occur than their corresponding positive ones. These findings confirm the recent literature on firm growth rate distributions within national economies (see section 6.2.2). However, the high variances of both shape parameters b_l and b_r points to a high inter-regional heterogeneity.

Table 19: Summary statistics for estimated AEP parameters

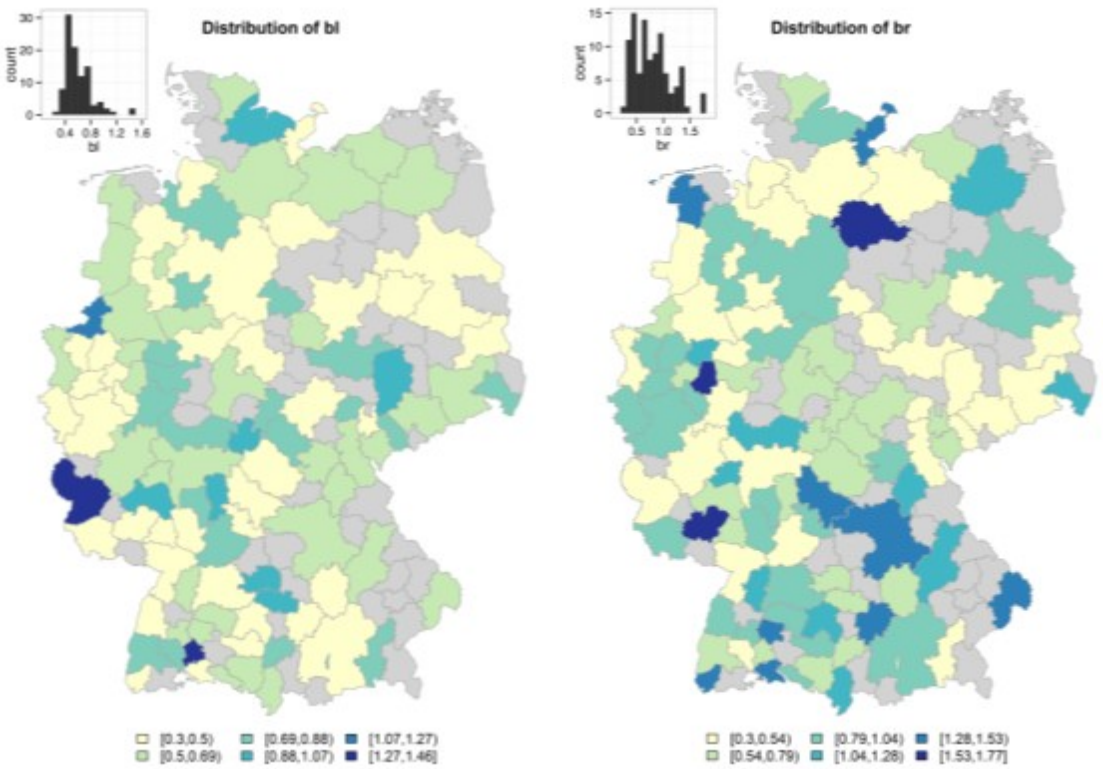
Estimation technique	$N_{regions}$		b_l	b_r	a_l	a_r	m
Without central part	**100**	**mean**	**0.601**	**0.791**	**0.111**	**0.117**	**-0.007**
		sd	**0.213**	**0.347**	**0.025**	**0.038**	**0.059**
Without zero data	100	mean	0.734	0.932	0.131	0.132	-0.001
		sd	0.277	0.391	0.026	0.042	0.067

Furthermore, this table compares the estimates resulting from the extended estimation technique, which leaves out the central part around zero, with the ones resulting from the conventional approach of optimizing the log-likelihood of f_{AEP}. For the latter, all zero growth events are excluded. The new technique makes the distributional mass shifting from the center to the tails: in average, both b_l and b_r become smaller as compared to the conventional technique without zero data. Simultaneously, the variance decreases. Despite one additional degree of freedom, asymmetry is reduced by the new technique. The absolute difference between b_l and b_r slightly decreases, in average, from 0.493 to 0.415 and between a_l and a_r from 0.042 to 0.029.

6.5.2 Regional factors accounting for the fat tails

The spatial distribution of the values for b_l and b_r as well as their sum $(b_r + b_l)$ and their difference $(b_r - b_l)$ are mapped in Figure 27. The high regional heterogeneity of firm growth rate distributions already suggests that they might reveal more about the underlying dynamics of regional economies. It was argued that turbulences at the level of firms allow for a first assessment of a region's long-term ability to adapt its structure and to re-invent itself, key aspects of evolutionary regional resilience. This leads to the question which region-specific factors are related to a higher potential for regional resilience. Results from the regression models are summarized in Table 20.

The two base models solely control for the number of observations and the perpendicular dimension. The models 1 to 4 include further regional variables. For each independent variable two models are devised, as the variables of similarity and related variety are strongly correlated to unrelated variety, which to some extent resemble two different sides of the same coin. Regression diagnostics do not reveal any problems regarding normality of the residuals, multicollinearity or spatial autocorrelation.[41] Only the null hypothesis of the Breusch-Pagan tests is rejected, that is why White's heteroskedasticity consistent standard errors are reported.



[41] The Moran's I test statistics is also robust to different weight matrix specifications.

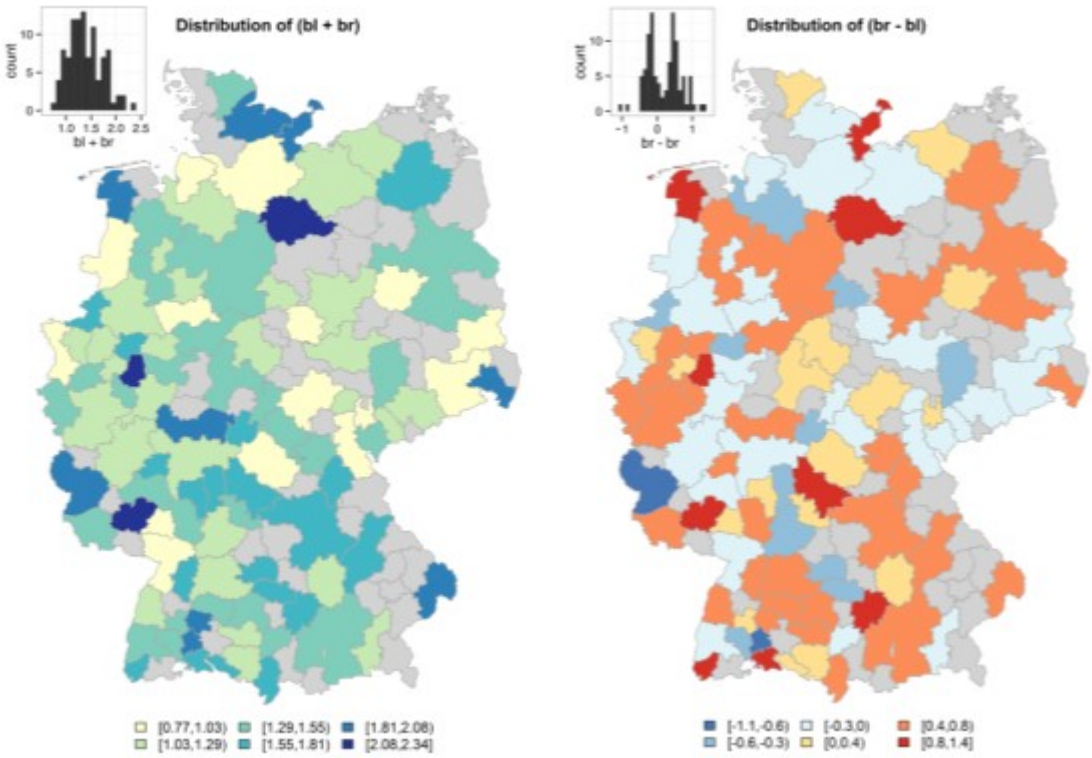


Figure 27: Spatial distribution of the values for b_l and b_r

Table 20: Regression results from OLS

	$(b_r + b_l)$			$(b_r - b_l)$		
	base model	model 1	model 2	base model	model 3	model 4
$(b_r - b_l)$ / $(b_r + b_l)$	-0.352 0.000 ***	-0.300 0.000 ***	-0.310 0.000 ***	0.702 0.000 ***	0.737 0.000 ***	0.742 0.000 ***
N_{firms}	-0.000 0.000 ***	0.000 0.849	-0.000 0.759	-0.000 0.004 **	0.000 0.492	0.000 0.318
PopDensity		-0.000 0.681	-0.000 0. 693		0.000 0.019 *	0.001 0.009 **
UnemplRate		-0.355 0.540	-0.558 0.520		0.397 0.590	-0.313 0.833
RegGrowth		-2.766 0.003 **	-2.312 0.007 **		1.840 0.139	1.696 0.150
ResFunding		0.277 0.006 **	0.290 0.004 **		-0.010 0.475	-0.076 0.559
EmplUniv		-0.536 0.027 *	-0.487 0.042 *		-0.203 0.498	-0.263 0.408
Manufacturing		-0.746 0.039 *	-1.154 0.014 *		-0.191 0.642	-0.213 0.777
Construction		-0.616 0.560	-0.826 0.448		-0.500 0.768	-0.260 0.882
Similarity		28.143 0.042 *			-11.188 0.724	
RelVariety		0.755 0.095 ‘			-0.286 0.695	
UnrelVariety			-0.559 0.032 *			0.409 0.554
Adj. R^2	0.269	0.390	0.386	0.233	0.212	0.227
BP-test (*p-value*)	0.002	0.012	0.046	0.001	0.005	0.013
KS-test (*p-value*)	0.677	0.828	0.799	0.103	0.349	0.343
Moran’s I (*p-value*)	0.438	0.944	0.921	0.688	0.747	0.719
vif	1.001	3.758	4.510	1.052	3.965	4.718

By relating the tail measure $(b_r + b_l)$ to various regional variables, the explained variance (R^2) of model 1 and 2 increases compared to the base model from 27% to almost 40%. Regarding the general socio-economic conditions, neither *PopDensity* nor *UnemplRate* turn out to correlate with the tails’ fatness, meaning that no evidence is found that the stochastic properties of firm growth rate distributions differ along the urban-rural as well as east-west or north-south divides. However, *RegGrowth* is highly relevant. The better the aggregate regional growth performance, the lower are the shape parameters and hence, the fatter the tails. This result shows that firm-level turbulences are predominantly a phenomenon of better performing regions. Yet it remains an open question whether more resilient regions manage to grow faster or whether aggregate regional growth, by providing new opportunities for the firms, but also by requiring them to change and to adapt, is the cause for regional resilience.

The regions’ innovation conditions, approximated by *ResFunding*, reduce significantly the occurrence of turbulences at the level of firms. Thus, regions with a strong (basic research oriented) science base are not necessarily those regions where firms are able to economically take off. On the contrary, *EmplUniv*, the qualification level of the region’s workforce, is strongly correlated with the fatness of the tails: the more employees with a university degree, the higher the likelihood of extreme growth events. This clearly confirms

the literature, which attributes the region's workforce skills an important role in its resilience and transformative capacity in general (e.g. CHAPPLE/LESTER 2010, GLAESER et al. 2011).

Finally, the industrial structure matters. Regions with a higher share of firms belonging to *Manufacturing* are more exposed to extreme events. This result is reasonable, as the manufacturing industry was stronger affected from the macroeconomic recession (e.g. GROOT et al. 2011, DAVIES 2011). Besides, it supports the expectation that increasing returns as a source of fat tails are more relevant among manufacturing industries. Regions with a higher share of the *Construction* industry do not show different stochastic properties in their firm dynamics. Besides the type of activity, also the way how these activities are distributed across a region's industrial portfolio matters. Particularly, a higher degree of similarity is associated with thinner tails. The same effect is observed for related variety, however only significant at the 10% level. Only in the case of variety of technologically unrelated activities, the sign of its effect on the tails changes – a less coherent and less interrelated technological base makes extreme growth events more likely. Thus, diversity and variety in the region's industrial structure seems to increase the likelihood that new technological trajectories unfold and old ones decline (CASTALDI et al. 2014). Put differently, specialized regions, where activities of the same or similar type concentrate, either constrain the necessary competition leading to such turbulent processes, probably due to a higher inertia of its actors and institutions, or simply provide less potential sources for path-breaking technological solutions, which "can be taken from one industry and used to create innovations that solve problems in other fields" (GILBERT 2012: 738).

Besides the explanation of the tails' fatness, which is conceptualized as an indicator for regional resilience from an evolutionary perspective, also the asymmetry of the distribution, measured by $(b_r - b_l)$ and representing a kind of vulnerability, can be analysed. However, the inclusion of further regional variables in model 3 and 4 do not increase the explanatory power, which remains at around 20% compared to the base model. With *PopDensity* only one variable shows a significant influence: the higher the population density, the more likely extreme negative growth events are to occur relatively to positive ones. This result can be regarded as a consequence of the recession, as urban regions turned out to be affected more adversely (HOLM/OSTERGAARD 2013).

6.6 Conclusions

This paper studies turbulences in the firm dynamics of regional economies. Turbulences are an indicator for processes like structural adaptation and technological re-orientation. Regional economies that provide a competitive environment, which facilitates that outmoded activities are substituted by new innovations and technologies, are assumed to be resilient in the long run from an evolutionary perspective. Such turbulences, otherwise hidden by aggregation, are revealed in the employment dynamics of firms. Here, the secret lies in the tails of the distribution of firm growth rates, i.e. in the extreme events, which tend to have a higher transformative impact on the regional economy. Therefore, this paper is a first attempt to assess the meaning of fat tails for the systems they correspond to, and the factors which make them particularly pronounced. Above all, this analysis shows that firm-level turbulences are more likely in regions with a higher aggregate growth

performance. Although the direction of causality is still unknown, this finding underlines the positive nuance of fat tails throughout the paper. In this vein, especially a diversified industrial structure that provides unrelated variety as well as the presence of a qualified workforce seems to make regional economies more resilient. In contrast, a strong science base surprisingly attenuates the tails.

However, it is yet an open issue which types of institutions can better accommodate these turbulences and translate it into fruitful changes and new long-term growth paths (BOSCHMA 2014). It should also be noted that this kind of resilience – with an indisputable neoliberal notion – might have negative consequences for some workers, who face difficulties in becoming reemployed by reason of their individual skills within the implied environment of more flexible working conditions (MARTIN 2012). Moreover, it is yet to be debated how much change is desirable from a societal point of view (REGGIANI et al. 2002). It is out of reach for this thesis to answer these questions, for which also the short-term and long-term benefits and social consequences have to be taken fully into account.

7 Chinese firm dynamics and the role of ownership type: A conditional estimation approach of the Asymmetric Exponential Power (AEP) density

Abstract: This paper investigates the impact of ownership type on the entire growth rate distributional mass of Chinese firms, using a conditional estimation approach of the Asymmetric Exponential Power (AEP) density that goes beyond simple location-shift analysis. We first find a Chinese growth puzzle, i.e. a deviation from the stylized fact of the variance-scaling relationship commonly observed in Western European economies. We then find that the ownership type mainly affects growth rates far-off the mean value. Our results also indicate that barriers to becoming a high-growth firm, such as financial constraints, are especially prevalent in the Chinese private sector.

Keywords: firm growth, growth rate distributions, AEP conditional estimation, variance-scaling relationship, China, ownership type

JEL Classification: C46, O53, L10

7.1 Introduction

Recently, China's growth momentum has been observed to slow down (EICHENGREEN et al. 2012). To find ways of sustaining the growth path, it is important to understand the dynamics at the level of firms. Since the adoption of an open-door policy in 1978 and subsequent liberalization efforts, the entrepreneurial environment in China has witnessed rapid and dramatic changes (MILANA/WANG 2013). After China's entry in the WTO in 2001, its firms have shown an exceptionally high productivity growth, driven by an increasing focus on global competition and innovation (BRANDT et al. 2012). China's private sector, in particular, is increasingly acknowledged as the new engine of growth, although it faces substantial financial constraints and other forms of institutional discrimination as compared with the more protected state sector (ALLEN et al. 2005, GUARIGLIA et al. 2011, CHAN et al. 2012). Besides the access to financial resources, the ownership type of Chinese firms implies differences in their innovation propensity (GUAN et al., 2009) and innovation efficiency (LIN et al. 2013), as well as in their ability to maintain and benefit from *guanxi*[42] (PARK/LUO 2001) and political connections (LI et al. 2008), and in their export-orientation (JEFFERSON et al. 2003). All these aspects are expected to translate into differential growth potential among state-, private- and foreign-owned firms (called SOE, POE, and FOE, respectively). Yet, the empirical literature on firm dynamics in China so far only finds weak and mixed effects of the ownership type on firm performance (PENG et al. 2004, CHOI et al. 2011, JU/ZHAO 2009).

In this paper, we argue that this lack of evidence stems partly from a mere focus on the average firm in the growth rate distribution. Yet, heterogeneity in the performance is omnipresent also within a group of ownership type (PENG et al. 2004). For example, some firms of the state sector have become quite competitive in the global market due to reform and transformation of their organizational culture, and are now even considered an important driver of aggregate economic growth, as "China's dynamo for the future" (RALSTON et al. 2006). This stands in contrast to many state-owned firms which are heavily in debt and perform rather poorly (SUN et al. 2002). Heterogeneity exists also within the private sector. For example, some private- or foreign-owned firms, although confronted with less favorable financial institutions, might overcome financial constraints by alternative channels, like retained earnings through cash flow (GUARIGLIA et al. 2011, PONCET et al. 2010). Strongly performing private firms are more likely to receive bank loans (CULL/XU 2005). In addition, other firms of institutional discrimination can be often compensated for by entrepreneurial and more aggressive strategies (PENG et al. 2004).

However, we expect that the mechanisms of how the ownership types affect firm performance can be observed mainly far off from the mean value of the growth rates. For example, a reliance on informal capital might be adequate for the average firm in the growth rate distribution, but become restrictively costly in the case of high growth events. Differences in innovation-orientation matter only for a rather small share of the best performing firms, whereas protective institutions affect the survival chances of the worst

[42] The concept of *guanxi* plays a prominent role in organizational studies in China. It refers to the "web of connections to secure favors in personal and organizational relations" (PARK/LUO 2001: 455) and it is sometimes regarded as a substitute for formal institutional support (XIN/PEARCE 1996).

performing ones. Limited exposure to domestic and international market fluctuations in the case of state-owned firms might primarily be a matter of general volatility and not of average growth performance. Recently, firms in the tails of the growth rate distribution have gained much attention from scientists and policy makers. These so-called high-growth firms, which would be neglected by a mere focus on the dynamics of the average growing firm, are acknowledged to be a main ingredient for aggregate growth dynamics (e.g., DELMAR/DAVIDSSON 200, DAVIDSSON et al. 2006, COAD et al. 2014a). In light of the recent slowdown of economic growth in China, it is of utmost interest to identify the barriers to becoming a high-growth firm.

Considering the heterogeneity among and within different groups of firms, we plead for an analysis of the entire growth rate distribution which explicitly takes into account the ownership type. It is a stylized fact that the distribution of growth rates of economic entities shows tails that are much fatter than the ones implied by a Gaussian normal distribution (e.g. STANLEY et al. 1996, BOTTAZZI et al. 2002). The literature suggests using the more flexible asymmetric exponential power (AEP) density to account for both extreme events and asymmetry in the shape (BOTTAZZI/SECCHI 2011). In this paper, we introduce a new approach of estimating this distribution model conditional on other variables, in our case the type of ownership. This approach combines advantages of (unconditional) distributional analysis with insights gained from traditional regression approaches.

The paper contributes to the literature in two ways. First, it tests whether the stylized facts on the distribution of growth rates, which have been established in the context of several Western economies, survive in an emerging East Asian economy. We identify especially a deviation in the variance-scaling relationship – a new Chinese growth puzzle that helps to shed light on particularities in the firm dynamics in China. Second, we investigate the role of ownership type on firm performance by considering its impact on the distributional mass that goes beyond simple location-shift effects. Thus, we provide a more complete picture on firm dynamics in China. In particular, we find that significant barriers to growth exist, making especially the realization of high growth events more difficult to achieve, which allows for space of policy intervention.

The paper is structured as follows. Section 7.2 provides a brief overview of the literature on firm dynamics in China, with a particular focus on the role of the ownership types. In addition, it positions this study in the literature on growth rate distributions. Section 7.3 discusses some data issues, while section 7.4 introduces the conditional estimation approach of the AEP distribution. The results are presented in section 7.5, and discussed in section 7.7. Section 7.7 concludes by shedding some light on the policy relevance of the findings as well as on the merits and limits of the proposed methodology.

7.2 Literature

In the literature on Chinese firm growth, the ownership type has assumed a pivotal role. In transition economies that experience a rapid institutional change, the organizational diversity, as reflected in ownership differences, is a key to understanding firm structure and behavior (e.g. PENG et al. 2004, PENG et al. 2008). In China in particular, different forms of

ownership types coexist and compete, and are exposed to different environmental constraints and competitive advantages (SHENKAR/VON GLINOW 1994). Depending upon their ownership type, firms face differences in access to external resources such as financial capital (e.g. ALLEN et al. 2005) or knowledge (e.g. CHOI et al. 2011). They also adopt and develop different competitive strategies (GEDAJLOVIC 1993), managerial heuristics and mentalities (PENG et al. 2004), or resource allocation and utilization decisions (JU/ZHAO 2009). All these aspects might translate into considerable variation in performance measures across ownership types (JEFFERSON et al. 2003).

Within each ownership group, however, firms are also heterogeneous. Hence, it might be insufficient and occasionally misleading to simply assess the impact on the average firm. We therefore re-investigate the role of ownership type on Chinese firm dynamics by looking at its impact on the entire shape of the growth rate distribution. In this section, we first review the literature on historical developments and characteristics of the various ownership types in China. Then, we position this study in the literature on firm growth rate distributions.

7.2.1 A brief historical account on the ownership types in China

During China's transition phase starting in the late 1980s, private-owned enterprises (POE) and foreign-owned enterprises (FOE) emerged as new ownership types to compete with state-owned enterprises (SOE), a communist-era legacy (PENG et al. 2004). These three main ownership types are complemented by collective-owned firms (COE), which are nominally owned by the local government but are often run like POE. Their organizational attributes fall somewhere in between those of SOE and POE (PENG et al. 2004). Although important in the 1980s, their relevance suffered from strong competition from POE (MILANA/WANG 2013). Because of their ambiguity and declining importance, we exclusively focus on SOE, POE and FOE, which can be characterized as follows.

State-owned enterprises (SOE)

In 1993, China started a gradual process of privatizing SOE, based on the reform policy of 'keeping the large and letting the small go'. Four years later, in 1997, the ideological discrimination against the private sector was formally removed and in 2004, private property rights were recognized by the constitution (MILANA/WANG 2013, RALSTON et al. 2006). This privatization process led to two important quantitative changes: the absolute number of SOE as well as the employment levels within the surviving SOE strongly declined (JEFFERSON et al. 2003). Yet, SOE remain crucial to the Chinese economy. MILANA and WANG (2013), for instance, report that the largest 118 SOE still contribute 43% of the GDP in 2012. Despite cutbacks of formal privileges, SOE are still favored by government policies, providing them with virtually unlimited access to loans from state banks (PONCET et al. 2010, GUARIGLIA et al. 2011) or with monopolistic positions in key strategic industries (RALSTON et al. 2006, JU/ZHAO 2009). These privileges, which are more informal in nature, are reinforced by their higher ability to maintain and benefit from *guanxi* (PARK/LUO 2001) or political connections (LI et al. 2008). Besides, the state as an owner

is also concerned about goals other than profit maximization, say, responding to social and political needs by absorbing surplus labor or ensuring stability in production in order to maintain social stability (BAI et al. 2006, CHOI et al. 2011). Hence, the managers' objectives are only partially related to market requirements (GUAN et al. 2009)

SOE are often described as the group with lowest productivity (GUARIGLIA et al. 2011). This can be explained, on the one hand, by their tendency to build up redundancies by investing in slack resources (QIN/SONG 2009, STAN et al. 2013), i.e. "potentially utilizable resources that can be diverted or redeployed for the achievement of organizational goals" (GEORGE 2005: 661). This phenomenon is especially relevant in emerging economies, where environmental uncertainties tend to be higher due to underdeveloped institutions and highly dynamic markets (STAN et al. 2013). These resources, however, are often diverted from the main business toward unproductive uses (JU/ZHAO 2009, MILANA/WANG 2013). On the other hand, they are less innovation- (GUAN et al. 2009) and export-oriented (JEFFERSON et al. 2003, BLONINGEN/MA 2007). In addition, a more risk-averse business culture has been developed under state protection (PONCET et al. 2010, MILANA/WANG 2013). Besides, they tend to be rigidly managed by older executives and confronted with bureaucratic and inefficient structures, lack of managerial knowledge, corruption, and agency problems emerging with malpractice or mismanagement (PENG et al. 2004, CHOI et al. 2011, MILANA/WANG 2013).

Although massive, inefficient and pre-reform SOE, to which the characterization above mainly refers, are still part of the business landscape, many have transformed to become viable and globally competitive. These are called by RALSTON et al. (2006) "China's economic dynamo for the future". Relying on the state's secure supply of capital and facilitated access to credit, they have become "protectively competitive" (MILANA/WANG 2013). Hence, it is indispensable to also consider heterogeneity among SOE when studying Chinese firm dynamics.

Private-owned enterprises (POE)

In the 1980s, POE emerged mainly from COE, which had then been leased out to private entrepreneurs, or from privatized small SOE (RALSTON et al. 2006). Although the private sector is now formally regarded as equally important as the state sector, POE still face institutional barriers and political discrimination, such as industrial entry barriers, lack of commercial conventions, insecure property and contract rights, taxation and asset seizure by local governments, and most notably, limited access to loans from state banks (RALSTON et al., 2006, JU/ZHAO 2009, MILANA/WANG 2013). Despite this unfavorable and rapidly changing environment, the mostly family-owned and, on average, smaller firms are increasingly regarded as the engine of aggregate growth (GUARIGLIA et al. 2011).

Many researchers have attempted to explain the "Chinese growth puzzle" (ALLEN et al. 2005, GUARIGLIA et al. 2011), i.e. China's high economic growth despite its malfunctioning financial system, by showing that POE are actually able to finance their investments through alternative channels, like internal savings (SONG et al. 2011) or informal capital (AYYAGARI et al. 2010). For instance, LIAO et al. (2009) reported that the financial sources for their interviewed POE consist of more than 50% from personal savings, 25% from

family and friends, and 10% from mortgages of own assets. Bank loans and other debt sources account for only 8% and 7%, respectively. Besides, it is found that especially top performing private firms are more likely to receive bank loans (CULL/XU 2005), implying heterogeneity as well as a complex endogenous relationship between firm performance and financial constraints.

Moreover, POE are more willing to take risks and to focus on innovation (PENG et al. 2004; JEFFERSON et al. 2003). Their simpler and relatively flexible organizational structure, often headed by younger managers, allows them to react more quickly to opportunities and to compete against the more established SOE, especially as they tend to follow more aggressive and growth-oriented business strategies (PENG et al. 2004).

Foreign-owned enterprises (FOE)

Since the adoption of the open door policy initiated by Deng Xiaoping in 1978, foreign investment has flourished in China. Foreign partners tend to contribute technological and managerial knowledge, equipment, capital and marketing experience, while their Chinese partners contribute land, buildings and *guanxi* (RALSTON et al. 2006).

FOE are the group with the highest productivity (HU et al. 2005, CHOI et al. 2011). On the one hand, this is due to advantages in accessing foreign knowledge. On the other hand, the parent multi-national enterprises (MNE), which in emerging economies operate in an environment that is less familiar to them and often also more volatile, tend to provide established technologies, which makes technology more predictable and allows FOE to better utilize their resources (JU/ZHAO 2009). Besides, they focus more strategically on the external environment in order to remain flexible within such a transitioning economic system (RALSTON et al. 2006).

Whereas POE and SOE primarily produce for the domestic market, FOE are the most export-oriented firms, accounting for over half of overall Chinese exports, according to BLONINGEN and MA (2007). These authors also argue that the spectacular success of Chinese exports within the global supply chain was only possible with the contribution of foreign direct investments. The other side of the coin is that MNE also tend to divest rapidly in uncertain times (MILANA/WANG 2013).

7.2.2 The distribution of firm growth rates

This paper is motivated by the observation that the average growth rate is insufficient to acknowledge the role of ownership type for firm performance. In line with the literature on industrial dynamics (e.g. DOSI et al. 2010), it calls for taking into account the entire distribution of growth rates. It is a stylized fact that growth rate distributions of economic entities show fat tails. Empirical evidence exists at the levels of firms (e.g. STANLEY et al. 1996), regional economies (DUSCHL/BRENNER 2013) and national economies (FAGIOLO et al. 2008), as well as for various size measures, like employees, turnover, productivity, profits and assets (ERLINGSSON et al. 2013). This stream of literature finds that growth

processes cannot be described by a normal distribution. Instead, it suggests using more flexible distributional models, say, the asymmetric exponential power (AEP) density, to account for two recurrent features, namely the high probability of extreme events and an asymmetric shape (BOTTAZZI/SECCHI 2011). Thus far, studies on firm dynamics using a distributional approach have been successfully applied in the context of several Western economies. Yet, only limited evidence on firm growth rate distributions exists for East Asian economies. Notable exceptions are found in MATHEW (2012) on Indian manufacturing firms and DOSI et al. (2013), who investigate labor productivity (growth rate) distributions of Chinese manufacturing firms. On the one hand, universality of the underlying growth mechanisms (AMARAL et al. 2001) suggests that the firm dynamics in China should show distributional properties similar to those observed for Western economies. On the other hand, particularities of emerging economies might give rise to substantial differences, demanding systematic comparisons.

Distributional models are usually fitted unconditionally to a population of economic entities without considering other aspects of heterogeneity apart from their growth performance. For instance, BOTTAZZI et al. (2014) observe that financially constrained firms differ in their growth patterns from unconstrained ones especially in the tails, and less so in the average. DUSCHL (2014) shows that the shape of firm growth rate distributions systematically varies across regional economies. In this study, we focus on the role of ownership type on firm growth. The impacts of the underlying mechanisms, however, are more complex than what can be captured by a simple location shift of the distribution. This is reflected in the empirical literature on ownership type and firm performance, which faces difficulties in establishing robust findings and reports positive, negative, curvilinear or often absent relationships (e.g. PENG et al. 2004, JU/ZHAO 2009, CHOI et al. 2011). But already DOSI et al. (2013), in the context of productivity, show that differences in firm performance across ownership types cannot be captured by the mean value alone. Hence, it might be a matter of the growth rates far off the mean.

7.3 Data

Data were obtained from the Chinese Industrial Enterprises Database conducted by the National Bureau of Statistics (NBS) of China for the years from 2001 to 2006. It contains all industrial firms that report directly to the statistics *xitong* (functional bureaucracy), that is, all SOE as well as all other types of firms with annual sales of more than 5 million yuan (HOLZ 2013). Empirical studies relying on data from emerging economies have to be particularly concerned with the issues of data quality and coverage.

In a recent survey on Chinese data, HOLZ (2013: 2) concludes that official data in China are plentiful as a product of the “legacy of a planned economy (with its need for a large volume of data) combined with a current government actively engaged in economic policy”. In the literature, Chinese official statistics are usually regarded as accurate and reliable (e.g. CHOW 2006). This holds especially true for firm level data, because they are not used as a direct benchmark for performance of local governments. Hence, they should not be subject to systematic manipulations. Data errors due to a lower quality of the data collection process might be higher, but can be regarded as random noise.

The data cover a share of around 87% in sales and 50% in employment of the entire Chinese economy in 2004 and are highly correlated with aggregate data at the industry level (HOLZ 2013). They cover many small and young firms (GUARIGLIA et al. 2011), excluding self-employment, which is not registered as an enterprise (HOLZ 2013). A break in the data occurred in 2004, following an economic census which implied new definitions and procedures in the data collection process (HOLZ 2013). To guarantee consistency, two symmetric balanced panels are constructed for the years before (2001 to 2003) and after the break (2004 to 2006).

Firm growth is a multi-dimensional phenomenon, with no universally accepted best size indicator (GILBERT et al. 2006, MCKELVIE/WIDKLUND 2010). We therefore study growth rates based on two alternative size measures: number of employees and sales. The latter was deflated relative to the base year of 2001 using an industry-specific producer price index provided by the NBS. For an extensive study on labor productivity dynamics in China, we refer to DOSI et al. (2013).

Firms are assigned to ownership types based on the fraction of paid-in-capital contributed by different types of investors. This procedure is suggested in the literature (e.g. JEFFERSON et al. 2003, GUARIGLIA et al. 2011), because official registration codes are not reliable, as they are only updated with delay, and state ownership can change meaning over time (HOLZ 2013). SOE consist of state investors, POE of legal entities and individuals, and FOE of foreign investors including Hong Kong, Macao and Taiwan. As multiple types of investors may own a company, the absolute majority rule is applied. GUARIGLIA et al. (2011) argue that this classification rule reduces measurement errors in the capital variable. The shares of investors are calculated over the entire growth period to reduce potential biases resulting from firm transitions, for example from SOE to POE (JEFFERSON et al. 2003; GUARIGLIA et al., 2011). The drop in number of firms due to the requirement of having at least 50% of one investor type is small (1.4% of total firms).

Finally, the following procedures for preparing the data are performed to increase the reliability of the results. First, all firms with negative or zero sales are excluded (resulting in a further drop of around 0.9%). Second, all firms with fewer than five employees are excluded (drop of 0.3%), because the growth process of micro-firms tends to be rather noisy. Third, outliers are excluded (drop of 0.2%) following the procedure proposed in BOTTAZZI et al. (2014). Some descriptive statistics are provided in Table 21.

Table 21: Descriptive statistics

	N_{firms} (relative share in growth period)		*Average size*	
	2001-03	2004-06	Employees	Sales
SOE	15,131 (17.5%)	8,452 (4.8%)	643.9	131,147.1
POE	54,767 (63.7%)	135,348 (76.4%)	214.8	507,26.2
FOE	16,100 (18.7%)	33,367 (18.8%)	353.5	119,401.5

Note: Sales are measured in 1000 Renminbi Yuan.
Source: Chinese Industrial Enterprises Database (2000-2006).

The sample share of SOE declined from the growth period of 2001-03 to 2004-06 almost fourfold, reflecting the structural transformations at that time, whereas the share of POE increased by 12.7% and the share of FOE remained constant. POE is the largest ownership group in numbers, but these firms are, on average, much smaller than SOE or FOE in terms of both employees and sales. SOE are, on average, the largest with these two size measures as well, especially with employees.

As a byproduct of their specific historic background and the existence of entry restrictions in certain industries, the ownership types are not equally distributed across industries. The Pearson's correlation coefficient is 0.707 between POE and SOE for the number of manufacturing firms at the 2-digit industry level, 0.617 between POE and FOE, and 0.198 between SOE and FOE. Given that industries exhibit different patterns of growth dynamics, innovation processes and competitive intensities (e.g. BRESCHI et al. 2000, MALERBA 2002), and thus may drive the observed firm performances instead, this industry composition effect is controlled for in our estimation, which will be discussed in the following section.

7.4 Methodology

The growth rates $g_{i,o,t}$ for the two growth periods t = {2001, 2004} are calculated for each firm i based on size $S_{i,o,t}$, represented by either employees or sales, using the logarithmic definition that follows from GIBRAT's law:

$$g_{i,o,t} = \log(S_{i,o,t+2}) - \log(S_{i,o,t}), \tag{40}$$

with o denoting the ownership type. To control for macroeconomic fluctuations, the mean within a growth period is subtracted from the growth rates. Next, the growth rates are also de-meaned by the ownership-specific average for each growth period, because growth within these ownership groups is affected by general structural transformations in China. An inverse relationship between S and variance of g is known to be a universal feature in the growth of complex economic organizations (AMARAL et al. 2001). To rescale the growth rates, we follow BOTTAZZI et al. (2014) in modelling this variance-scaling relationship directly by introducing a heteroskedasticity term into the stochastic growth process:

$$\mathrm{g_{i,o,t}} = \exp(\beta_o(\mathrm{s_{i,o,t}} - \bar{\mathrm{s}}_o))\ \mathrm{e_{i,o,t}}, \tag{41}$$

with $\mathrm{s_{i,o,t}} \equiv \log(\mathrm{S_{i,o,t}})$ and $\bar{\mathrm{s}}_i$ denotes the corresponding ownership-specific mean. This expression takes into account the observation that the functional form of heteroskedasticity might be non-linear (e.g. BOTTAZZI et al. 2011). Be rearranging equation (41), we get:

$$\mathrm{e_{i,o,t}} = \frac{\mathrm{g_{i,o,t}}}{\exp(\beta_o(\mathrm{s_{i,o,t}} - \bar{\mathrm{s}}_o))}, \tag{42}$$

which yields the rescaled growth rates $\tilde{\mathrm{g}}_{\mathrm{i,o,t}} := \mathrm{e_{i,o,t}}$. Anticipating non-normality, equation

(43), which solves for the rescaling parameter β, is estimated by minimizing absolute deviations (LAD):

$$\{\beta_o\} = argmin_\beta \sum_i \left| \frac{g_{i,o,t}}{exp(\beta_o(s_{i,o,t} - \bar{s}_o))} \right|. \tag{43}$$

The rescaling step, which cleans the data of heteroskedasticity, is performed for each ownership type separately to account for possible differences in the variance-scaling relationship.

The literature suggests that the asymmetric exponential power (AEP) density reasonably describes the (rescaled) growth rates (e.g. BOTTAZZI/SECCHI 2011, FAGIOLO ET AL. 2008, DUSCHL/BRENNER 2013). Its density function reads:

$$\begin{aligned} f_{AEP}(g;\ b_l, b_r, a_l, a_r, m) \\ = \frac{1}{C} exp\left(- \left[\frac{1}{b_l} \left| \frac{g-m}{a_l} \right|^{b_l} \theta(m-g) + \frac{1}{b_r} \left| \frac{g-m}{a_r} \right|^{b_r} \theta(g-m) \right] \right), \end{aligned} \tag{44}$$

where $\theta(.)$ is the Heaviside theta function, and $C = a_l b_l^{1/b_l - 1} \Gamma(1/b_l) + a_r b_r^{1/b_r - 1} \Gamma(1/b_r)$ a normalization constant, containing the gamma function $\Gamma(.)$. Five parameters define the distribution: location parameter m, which indicates the central tendency in the data, two scale parameters a_l and a_r, which determine the spread or dispersion of the distribution for the values below and above m, and two shape parameters b_l and b_r describing the tail behaviour on the left and right side of m. Both the normal ($b_l = b_r = 2$) and Laplace ($b_l = b_r = 1$) distributions are special cases of the AEP family of probability densities. It allows for a continuous variation from non-normality to normality, with a smaller shape parameter b_l or b_r representing fatter tails of the corresponding density. Finally, it is able to cope with asymmetries in the data by having independent left and right hand side parameters.

This distributional model is fitted to the sample data by using maximum likelihood. Mathematical details and related inference issues are discussed in BOTTAZZI and SECCHI (2011). Here, we focus instead on two aspects of heterogeneity in the firm population that might bias the results: firms belong to different industries and ownership types, and both aspects might affect not only the average value of the growth dynamics, but also the entire distribution in a complex way. As this paper focuses primarily on the latter aspect, two different strategies are proposed. First, to control for the industry composition effect, the estimation of the AEP likelihood function is weighted by the share of a firm's industry in the entire economy relative to its ownership-specific share. Put differently, the industries that are underrepresented in one ownership type will then be compensated by higher weights of their firms, calculated according to this share, in the estimation of the distribution. Second, to directly address and quantify the impact of ownership type, all distributional parameters are made conditional upon the ownership type. Thus, the distributional function of the conditional AEP reads

$$\begin{aligned} F_{AEP}(g;\ b_l &= b_{l,0} + \beta_{bl}\mathbf{X}, \\ b_r &= b_{r,0} + \beta_{br}\mathbf{X}, \end{aligned} \tag{45}$$

$$a_l = a_{l,0} + \beta_{al}\mathbf{X},$$
$$a_r = a_{r,0} + \beta_{ar}\mathbf{X},$$
$$m = m_0 + \beta_m\mathbf{X}),$$

where **X** is a matrix containing the conditioning variables and $\{\beta_{bl}, \beta_{br}, \beta_{al}, \beta_{ar}, \beta_m\}$ a set of coefficients containing the magnitude of the effects of **X** on the distributional parameters $\{b_{l,0}, b_{r,0}, a_{l,0}, a_{r,0}, m_0\}$. The ownership type is represented by a dummy variable. Here, we assume the benchmark to be POE, hence **X** is *{d_SOE, d_FOE}*.

Finally, significance of the effects of the conditioning variable on the parameters can be obtained within the maximum likelihood framework with the help of likelihood-ratio tests.

7.5 Results

The results for the variance-scaling parameter *β* are reported in Table 22. The findings suggest that there exists a clear negative relationship between the variance of growth rates and the size of firms. The scaling parameter *β*, however, is strikingly higher than observed in most of the Western economies: it ranges from -0.057 (employment growth in SOE) to -0.096 (sales growth in SOE). In contrast, studies on the US robustly report that *β* lies in a narrow range between -0.15 (e.g. STANLEY et al. 1996, AMARAL et al. 2001) and -0.20 (e.g. AMARAL et al. 1997). For Italy, BOTTAZZI et al. (2014) report a scaling parameter of -0.23. Only for France was it found to be higher than -0.1 in many industries (BOTTAZZI et al. 2011).

Table 22: The estimated variance-scaling parameter *β*

	Variance-scaling parameter β		
	Employees	Sales	Productivity
SOE	-0.057	-0.096	-0.120
POE	-0.085	-0.078	-0.061
FOE	-0.066	-0.064	-0.065

The results obtained from the conditional estimation of AEP distribution are reported in Table 22. The table's upper part reports the estimated coefficients for intercepts in equation (45). These intercepts represent the five distributional parameters after conditioning for ownership type, while the lower part contains the coefficients of the ownership dummies and the corresponding p-values. One can calculate the effect magnitudes (and their significance) of the ownership of SOE or FOE compared with the benchmark of POE for a distributional parameter, by adding up the corresponding estimated intercept and ownership-specific coefficient. For example, for the growth rate distribution of employees, the shape parameter of the left tail (b_l) for POE is 0.633, and that for SOE is 0.668 (by adding 0.035).

Table 23: Results from conditional AEP estimation

	Employees		*Sales*	
	coef.	p-values	coef.	p-values
$b_{l,0}$	0.633		0.904	
$b_{r,0}$	0.784		1.483	
$a_{l,0}$	0.262		0.399	
$a_{r,0}$	0.309		0.636	
m_0	-0.075		-0.205	
b_l *d_SOE*	0.035	0.0001 ***	-0.011	0.4978
b_l *d_FOE*	0.008	0.2380	0.007	0.5572
b_r *d_SOE*	-0.320	0.0000 ***	-0.333	0.0000 ***
b_r *d_FOE*	0.093	0.0000 ***	-0.096	0.0000 ***
a_l *d_SOE*	0.015	0.0000 ***	0.111	0.0000 ***
a_l *d_FOE*	-0.015	0.0000 ***	-0.019	0.0002 ***
a_r *d_SOE*	-0.129	0.0000 ***	-0.169	0.0000 ***
a_r *d_FOE*	0.015	0.0000 ***	-0.053	0.0000 ***
m d_SOE	0.062	0.0000 ***	0.017	0.1794
m d_FOE	0.013	0.0000 ***	-0.036	0.0014 **

*p-values: * < 0.05, ** < 0.01, *** < 0.001*

To provide a more intuitive account, the growth rate distributions in terms of employees and sales are also visually represented in Figure 28 and Figure 29. This is only possible because the conditioning variables are binary. Hence, the data can be split into three subsets and the parameters of the AEP can be estimated for each ownership type separately.

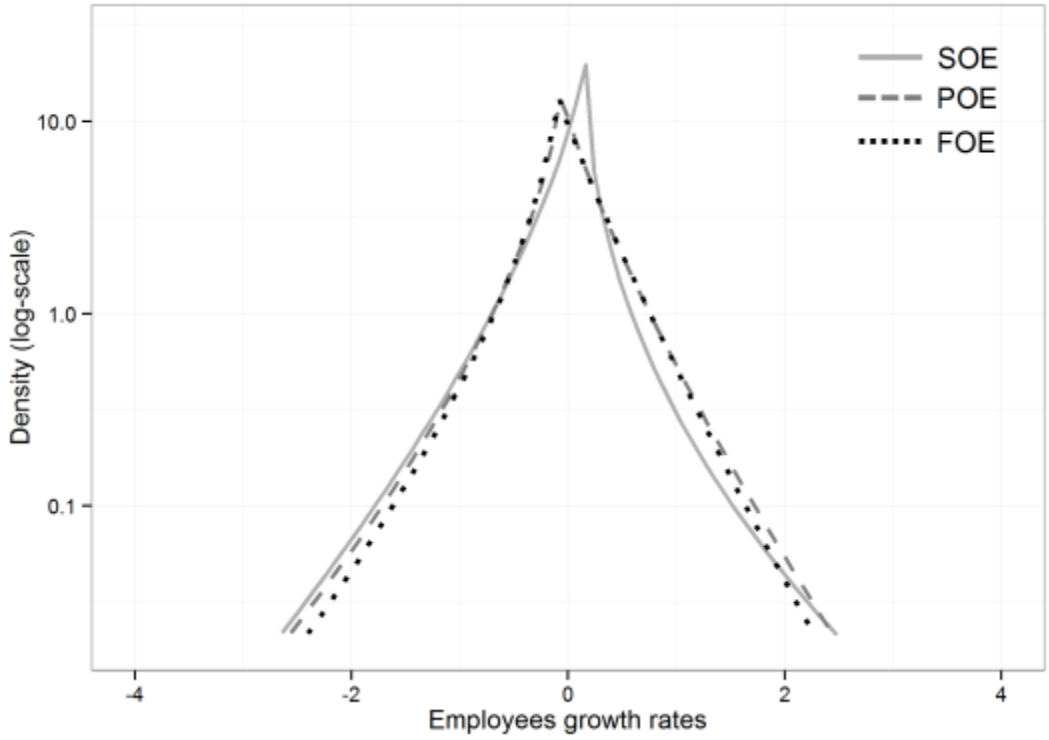


Figure 28: Growth rate distributions for employees

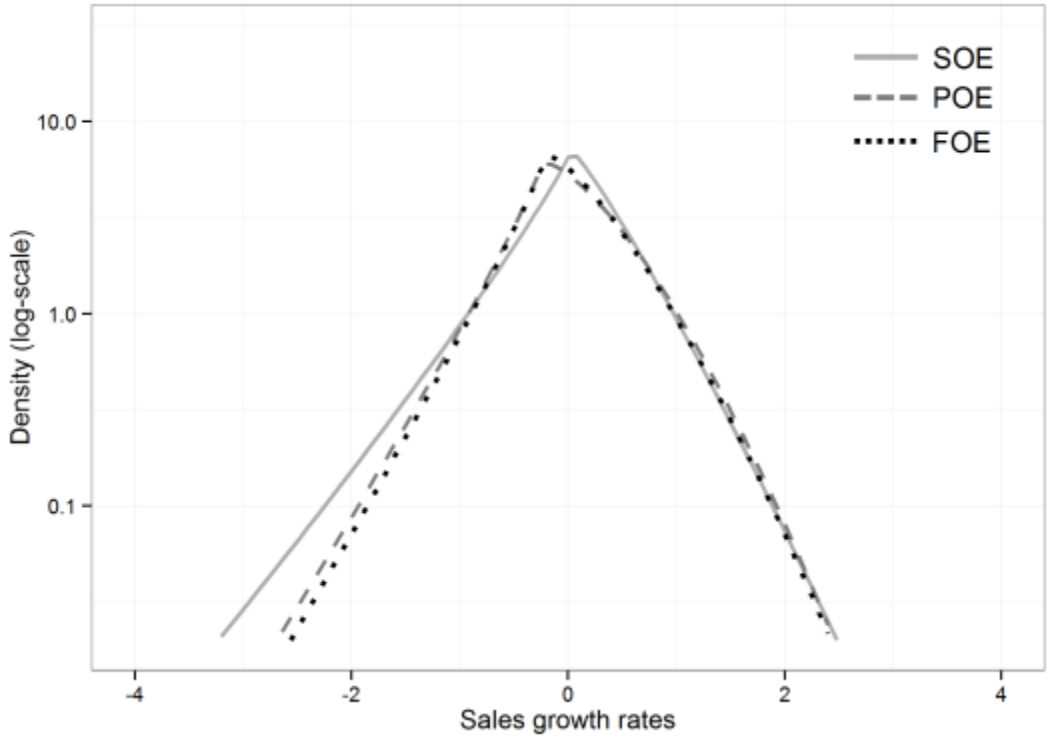


Figure 29: Growth rate distributions for sales

The robustness of the results is tested in two ways. First, the ownership type is inferred from the official registration codes instead of using the fraction of paid-in-capital. However, only marginal differences can be observed in the results. Second, to account for the general observation that the distributions tend to move closer to the normal one with an increased time lag (BOTTAZZI/SECCHI 2006a, DUSCHL/BRENNER 2013), also 1-year growth rates are studied. As expected, b_l and b_r become lower. For sales, b_r clearly remains above 1. The decrease of b_r was larger for sales than for employees, a finding which will be taken up again in the discussion.

7.6 Discussion

We separate the discussion of the results into two parts. In the first part, the variance-scaling parameter β from equation (43) is contrasted with the findings reported in the literature on Western economies. In the second part, the growth rate distributions for the three alternative size measures are investigated. Here, both general properties of the distributions and the impact of the ownership type on firm performance are discussed.

7.6.1 The variance-scaling relationship – another Chinese growth puzzle

In the literature, the variance-scaling relationship has not yet been investigated explicitly for Chinese firms. It turns out that the scaling parameter β tends to be much higher than in most Western economies. Is this deviation from a well-established stylized fact a statistical artefact (due to some data issues) or does it convey any economic meaning? Although the former cannot be completely ruled out, it might be fruitful to look into the models that explain the emergence of the variance-scaling relationship. Two main competing models exist. On the one hand, a tree-based hierarchical model (e.g. STANLEY et al. 1996, AMARAL et al. 1997) suggests that, by making assumptions on the internal structure of the firms, if the subunits of a firm fluctuate independently, β should be -0.5. The higher the positive correlations among the firm's units at different levels, the more β approaches 0. Our observed scaling relationship should therefore lie within these two limiting cases (see Figure 30).

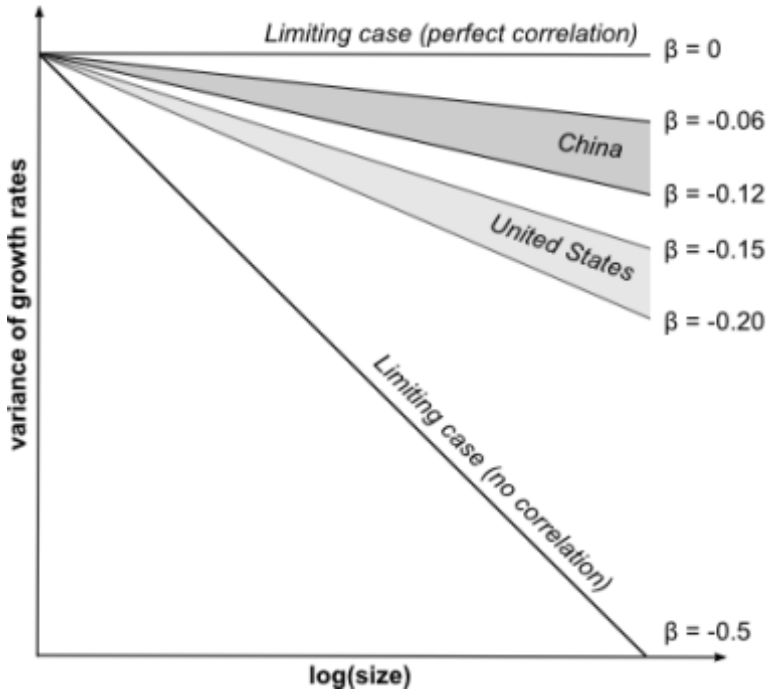


Figure 30: Variance-scaling relationship in China and US economies with limiting cases from tree-based hierarchical model

In the Chinese business culture, there is strong emphasis on hierarchical structures (RALSTON et al. 2006, PENG et al. 2004). The top-down control in Chinese firms tends to be more stringent, implying that the managerial decisions at upper levels should have a higher statistical impact on the decisions made at lower levels (STANLEY et al. 1996). As sub-units are less independent in China, this model would suggest a lower *β* compared to Western economies and is thus in line with our findings.

On the other hand, BOTTAZZI and SECCHI (2006b: 848) show that the variance-scaling can be explained by "a similar scale relation that exists between the size of the firm and its diversification structure". In other words, the variance-scaling relationship can be directly derived from the firms' diversification pattern. Using worldwide pharmaceutical data, they find that the number of active sub-markets is exponentially related to the size of the firm, as the firms' "ability to effectively enter in new sub-markets progressively increases with the number of sub-markets already penetrated" (BOTTAZZI/SECCHI 2006b: 871). Thus, they propose a stochastic branching model of firm diversification, in which all active sub-markets become possible sources of new diversification events. This competence-based diversification process is in line with the PENROSE's (1958) theory on the firm. In the context of an emerging economy like China, the evolutionary idea that firms learn how to diversify is less a matter of the incremental development of technological capabilities and product innovations (DOSI et al. 1995), but rather related to the daily struggle of survival in uncertain environments. The lower *β* found in China would imply that the "scope economy to diversification" argument due to learning (BOTTAZZI/SECCHI 2006b) becomes even more important. However, the variance-scaling relationship could have other possible origins,

like the way in which activities of a firm are distributed or concentrated among its active markets.

To sum up, the deviation from a well-established stylized fact found in the variance-scaling relationship is another "puzzle" in the Chinese growth literature. Therefore, future research should try to test the theoretical models more explicitly, for example by exploiting data at the product or sub-market level. The focus of this paper, however, is on the growth rate distribution and the role of the ownership type, which is the topic of the next section.

7.6.2 The growth rate distributions of Chinese firms – the type of ownership matters

Growth rate distributions provide a more comprehensive picture on firm dynamics than average values. Several mechanisms for how the ownership type might affect growth rates far off the mean are feasible. Therefore, the AEP distribution is estimated conditional on the ownership type. But first, general properties of the distributional shape in China are discussed.

General properties of the growth rate distributions of Chinese firms

The results of the distributional analysis of growth rates show that extreme growth events, especially negative ones, are a prominent feature of firm growth in China. This finding confirms one of the most robust stylized facts on firm growth also for the case of China – the presence of fat tailed and asymmetric growth rate distributions. Fat tails are observed in the literature independent of the measure for size (ERLINGSSON et al. 2013). However, differences among employees and sales are observed. For employees, both b_l and b_r are clearly lower than 1, hence the tails are much fatter than those of a Laplace distribution. For sales, the left tails are similar to the Laplacian ones, but the right tails come quite close to the shape of the normal distribution, with b_r around 1.5. Fitting the symmetric EP distribution to productivity growth rates, DOSI et al. (2013) report values for parameter b between 1.1 and 1.2. As labor productivity essentially resembles sales relative to employees (without taking into account the correlation structure between both variables), this intermediate value is in line with our findings.

The fatter tails for employees can be explained by the nature of the growth process. Firms often make lumpy adjustments to their employment level, which are due to non-convexities and irreversibility either large or nil (CABALLERO et al. 1997, COAD 2012). HALTIWANGER (1997), for instance, observed that plants spend a large fraction of time within +-30% of their desired employment level. But if they grow, extreme growth events become more likely (DUSCHL 2014). This is reflected in the peaked and fat tailed shape of the growth rate distribution. More intriguing is the finding of the right tails in case of sales. These are less fat than usually found in the literature on Western economies (e.g. REICHSTEIN/JENSEN 2005; BOTTAZZI et al. 2014). In other words, high positive growth events are less likely among Chinese firms. Two explanations can be brought forward. The first relies on the observation that China, instead of being an innovation-driven economy, is still an

efficiency-driven economy, with mostly capital-widening efforts in traditional primary and secondary industries (SCHWAB 2013). Innovation itself is a stochastic process showing a heavily skewed distribution with rare but highly influential events (SILVERBERG/VERSPAGEN 2007). If an innovation is successfully realized, however, it often translates into high growth performance. Hence, innovations will primarily affect the right tails of the growth rate distribution. As has been shown in DUSCHL and BRENNER (2013), an engagement in knowledge-based activities is correlated with the fatness of distributions' tails. To increase the pool of potential high-growth firms, the government could implement policies that offer incentives and provide support such that firms go beyond their focus on input factor accumulation and labor-intensive activities with low technology content, but undertake innovations and productivity improvements (MILANA/WANG 2013). In addition, CAO et al. (2013) discuss in the policy forum of the Science magazine the historical roots of the underperforming Chinese innovation system and possible pathways for political interventions, for example by connecting the research and business sector more strongly with help of institutional reforms.

The second explanation is based on the idea that there might be some barriers to growth, like the access to financial resources and knowledge, management and marketing problems, or regulatory obstacles (COAD/TAMVADA 2012, COAD et al. 2014b), making the realization of high growth events, in contrast to decline events, much more difficult to achieve. By comparing 1- and 2-year growth rates in the preceding robustness analysis, it can be shown that especially in the case of sales it is more difficult to sustain high growth over longer time periods, which is a feature of high-growth firms.

The role of the ownership type for firm dynamics in China

During the study period, the state sector has shown weaker growth than the private sector (ALLEN et al. 2005). As argued, this is an artefact stemming from current structural transformations and the demise of the state sector, implying lower growth potentials across the entire distribution of SOE. Once controlling for structural transformations by subtracting the mean value from the distribution, the central location of the distribution of SOE, as compared to POE, is shifted towards the right in the case of employees, while this shift becomes insignificant in the case of sales. The distribution of FOE is shifted slightly towards the right for employees, and towards the left for sales. Although these differences might be attributable to the institutional environment, it has been argued that the central location of the distribution is not of high interest, as it mainly concerns the firms that do not grow much in absolute terms. Rather, growth events far off the mean are relevant. These both crystallize the underlining mechanisms and significantly contribute to aggregate dynamics.

To begin with, SOE are much more likely to become high-growth firms in the case of both employees and sales. This shift of the distributional mass towards the right tail becomes most clearly visible for employees. This finding is in line with the literature (BOTTAZZI et al. 2014), which argues that financial constraints become increasingly binding the higher the growth rates are, as a jump in the volume of sales or in the number of employees requires large amounts of prior investments. This reasoning holds particularly true in China, for

which it has been argued that firms basically bear higher costs with alternative financial sources, say, internal funds through cash flow and returned earnings or informal finance. A non-negligible share of SOE, having virtually unlimited access to financial resource and operating under conditions of global competition, is able to show superb performance, resulting in the fatter tails of the SOEs' growth rate distribution. Put differently, the barriers to high growth discussed earlier are particularly strong for the private and foreign sectors due to financial constraints and other forms of institutional discrimination, a finding that offers important insights for macro-economic growth policies.

In addition, we find that FOE are significantly more likely to show high growth events in case of sales as compared to POE. The availability of technological and managerial knowledge is an important barrier for the exploitation of economic opportunities that can be more easily overcome by FOE (HU et al. 2005, RALSTON et al. 2006, COAD et al. 2014b). Foreign partners provide access to critical resources and imitation- and innovation-relevant knowledge (GUAN et al. 2009, CHOI et al. 2011). In the case of employees, it is the opposite: FOE have the thinnest right tails, reflecting the observation that POE tend to follow a more aggressive and growth-oriented business strategy, whereas FOE tend to concentrate on established technologies of their core business (PENG et al. 2004, CHOI et al. 2011).

Regarding high decline events, no difference is found in the case of sales. In the case of employees, SOE are less likely to show extreme negative growth events. The latter finding might be explained by the reluctance to abruptly lay off large amounts of employees in the state sector due to political considerations. Taking into account simultaneously the general negative fluctuations a_l, it becomes clear that there is a strong overall shift of distributional mass towards the left for SOE (but less strong for employees). This finding can be explained by a selection pressure, which is lowest for SOE (GUARGLIA et al. 2011). In key industries, they maintain monopolistic positions (JU/ZHAO 2009) and often pursue goals other than profit maximization (CHOI et al. 2011). In contrast hereto, increasing product market competition compels POE and FOE to be more efficient to survive (JU/ZHAO 2009). Hence, even the worst performing firms of the state sector tend to survive and continue to co-exist with better performing ones, for example by exploiting market niches (DOSI et al. 2012) or by relying on other strategic advantages, like facilitated access to financial resources (PONCET et al. 2010, SONG et al. 2011) or political connections (LI et al. 2008). Indirect evidence for the selection pressure argument can be drawn from the preceding robustness analysis, which compared 1- and 2-year growth rates: selection forces should become more effective in case of sustained poor performance. Indeed, both the tails and volatility of the left side decrease less strongly over time for SOE than for POE and FOE.

From the literature, it can be expected that the general volatility of the growth rate distribution should be lower for SOE, resulting in a more stable growth path. Slack resources buffer firms from internal changes by absorbing fluctuations in the external environment (PFEFFER/SALANCIK 2003). Also their higher self-sufficiency and lower export-orientation makes them less dependent on shorter term market fluctuations, especially in the context of international markets. The relatively small and highly significant coefficients confirm this expectation in the case of both employees and sales for a_r. For the left side, however, fluctuations are higher among SOE, conflicting with this expectation. We can explain this finding only by arguing that the selection pressure argument, as outlined above, might outweigh these mechanisms. To conclude, it is important to acknowledge the

complex interplay between the various parameters and to take into account measures of the assumed mechanisms more directly.

7.7 Conclusions

The conclusion is dedicated to the discussion of the policy relevance of the findings as well as the merits and limits of the proposed conditional estimation approach of the AEP distribution.

The analysis of Chinese firm dynamics has shown the importance of taking into account the entire distribution of growth rates. While many firms perform very poorly (but still survive), many other firms thrive and prosper. This heterogeneity in firm dynamics is particularly strong among SOE, confirming RALSTON et al. (2006), who distinguish the massive, inefficient and pre-reform SOE from the viable and globally competitive ones. Because of this within-group heterogeneity, the empirical literature faces difficulties in detecting robust differences among various groups because it tends to exclusively focus on the average growing firm. This paper contributes to the literature by showing that the ownership type mainly affects the growth rates far off the mean value.

Our findings are particularly relevant in light of the recent slowdown of China's growth momentum. In the literature, it is stressed that growth-oriented firms and innovative entrepreneurship are the engine of economic development and a key factor for maintaining sustainable growth paths (DELMAR/DAVIDSSON 2003, DAVIDSSON et al. 2006, COAD et al. 2014a). However, this paper finds that some significant barriers to growth exist, making especially the realization of high growth events more difficult to achieve. These barriers seem to be highest for the private sector due to substantial financial constraints and other forms of institutional discrimination. If external finances had been available less restrictively, POE and FOE, with many of them being already quite efficient, would have been able to grow at even higher rates, as they would not have had to rely on expensive alternative sources to finance their growth (GUARIGLIA et al. 2011, SONG et al. 2011). Even though high aggregate growth was possible in the past, some growth potential was missed and is in danger of being missed in the future. Hence, it is suggested that policy to be revised to increase efforts to improve the capital market and to focus on the remaining discrimination against the private sector (MILANA/WANG 2013, CHAN et al. 2012). In 2014, China approved trials for the first privately owned banks. It will be interesting to see how these reforms affect firm dynamics in the future. In addition, the discrimination is also visible in the relatively low selection pressure faced by SOE, which might be addressed by removing their monopolistic market positions and establishing a competitive economic environment.

The conditional estimation approach of the AEP distribution combines the advantages of both traditional regression models and unconditional distributional analyses. Regression models are designed to capture the "location-shift effect […] in the conditional distribution of the dependent variable" (BOTTAZZI/SECCHI 2013: 2). Following a distributional perspective it becomes clear that the variables, like ownership type, might affect not only the average or some specific arbitrarily chosen quantiles of the growth rate distribution, but

“may continue to impact on other distributional characteristics” (MAASOUMI et al. 2007: 449). More precisely, they might affect the central location, the variance, the tails or the asymmetry, i.e. the entire shape of the distribution, *simultaneously*. For instance, studying the shifts of distributional mass towards the tails is of high interest due to the importance of the fastest growing firms to aggregate dynamics. High-growth firms, which populate the right tail of the distribution, can be studied without the requirement of delimiting a sub-population *ex ante*, for example by choosing some arbitrary growth rate threshold (on the difficulties of measuring high-growth firms see also DELMAR et al. 2003; or more recently COAD et al. 2014a). But this approach also goes beyond the insights from an unconditional distributional analysis. The value of a conditional approach was already outlined by BOTTAZZI et al. (2014), who visually inspect the complex effects of financial constraints on the entire shape of the firm growth rates distribution. Unlike in DUSCHL (2014) or DOSI et al. (2013), a conditional estimation approach does not require a second-stage regression on the estimated parameters, which carries the potential risk of introducing additional biases (HAUSMAN 2001).

In this paper, we introduce this approach by investigating the simultaneous impact of binary variables, i.e. the ownership dummies, on the five parameters of the AEP. This allows us to inspect the results with a complementary visual assessment. But this approach is even more interesting for continuous variables, such as R&D, exports or financial measures, which provide a way to test the implicitly assumed mechanisms more directly. Visual comparisons are only possible by subsetting the data using some arbitrary thresholds. In contrast, the conditional estimation approach of the AEP allows us to assess the impact of a one-unit change of some variable on the probability of a certain growth rate occurring. However, two main disadvantages remain. First, a distributional model must be assumed in the first place. In the context of firm growth, however, there is widespread agreement that the AEP distribution provides a reasonable description of the reality (FAGIOLO et al. 2008). Second, the numerical optimization of the corresponding likelihood function is computationally expensive, with each conditioning variable introducing five new parameters to estimate. Considering the decreasing importance of computational costs as a restrictive factor, this approach will become even more promising for future research.

8 Modelling Firm and Market Dynamics – A Flexible Model Reproducing Existing Stylized Facts

Abstract: This paper presents a firm and market model that is able to reproduce the empirically observed patterns on firm growth and its statistical characteristics. It goes beyond the existing firm models by reproducing all stylized facts established in the literature. Furthermore, the model is flexible so that it can be adapted to certain industries and life-cycle stages. We analyse and discuss the options that are provided by the various parameters in this sense.

8.1 Introduction

Economic growth is an important issue in science as well as in society. Most research in this field is done on national economic growth and the growth of firms. In recent years regional growth processes have received increasing attention (see BREINLICH et al. 2014 for an overview). All kinds of studies can be found in this field, empirical studies as well as conceptual theoretical works and mathematical modelling approaches. On the level of firms the empirical literature is especially rich (see COAD et al. 2009 for an overview). As a consequence, the statistical knowledge about firm growth and firm size distribution is quite comprehensive. Although many firm theories exist (e.g. PENROSE 1959), so far no firm model exists that is able to reproduce all well-established stylized facts.

In this paper we develop such a model. Conceptually, we build on the model of monopolistic competition by DIXIT and STIGLITZ (1977). As a consequence, we argue that firm dynamics can only be adequately modelled in connection with modelling market dynamics. The proposed model represents market dynamics, including the emergence and disappearance of submarkets, as well as firm dynamics, including innovation processes and firm foundation. The approach by BRENNER and WERKER (2007) is used to validate the model.

We show that the model is able to reproduce all well-known stylized facts on firm growth and their relation to firm size and age. We neither intend to nor obtain one simulation model. Instead, we identify a number of parameter sets that all lead to realistic outcomes. The different parameter sets are able to represent different market situation, such as different industries or different stages in the industrial life-cycle. The meanings of the various parameters in this context are discussed. The paper proceeds as follows. In the next section the simulation approach is described. Section 8.3 provides an overview on the existing firm models as well as on the empirical knowledge about firm growth. Our simulation model is developed in section 8.4. Section 8.5 contains the calibration of the model and the discussion of the implied characteristics of firm growth. Section 8.6 concludes.

8.2 Simulation approach

8.2.1 Fundamental considerations

Our simulation approach follows the proposal for an abductive approach by BRENNER and WERKER (2007). The basic idea of this approach is a distinction between two spheres: (A) The sphere of the model and (B) the sphere of the implications. This distinction is an analytical tool, which is helpful because separate literature – theoretical and empirical – exists for each of the spheres. The spheres might refer to different spatial levels, but this is not a necessary condition. For example, the two spheres might be (A) firm growth and (B) regional growth, but they might also be (A) the mechanisms in firms underlying firm growth and (B) the statistical characteristics of firm growth in a firm population.

The simulation approach builds on the fact that (1) knowledge is available on both spheres and (2) there is a logical link (L) between the spheres that can be built into the simulation model. Most simulation approaches in the literature try to find one model (model sphere) that is able to reproduce the known facts on the implication sphere, given the logical link. The intention of the approach by BRENNER and WERKER is to find all models that are in line with the knowledge about the model sphere and the implication sphere. This is done in a two-step procedure.

The first step is to build various possible models which all together build the set of model specifications (see Figure 31). The aim is to develop a set of models that is quite general, restricting the models only as far as it can be justified by the available knowledge (this is done here in section 8.4). Generality of the model is reached by using many parameters that are not or only vaguely fixed. A set of possible model specifications results (see Figure 31). Simulating one model specification leads to a theoretical realisation (logical link). Due to random elements in the models, rerunning a model might lead to different results, so that for each model specification a set of theoretical realisations results (see Figure 31).

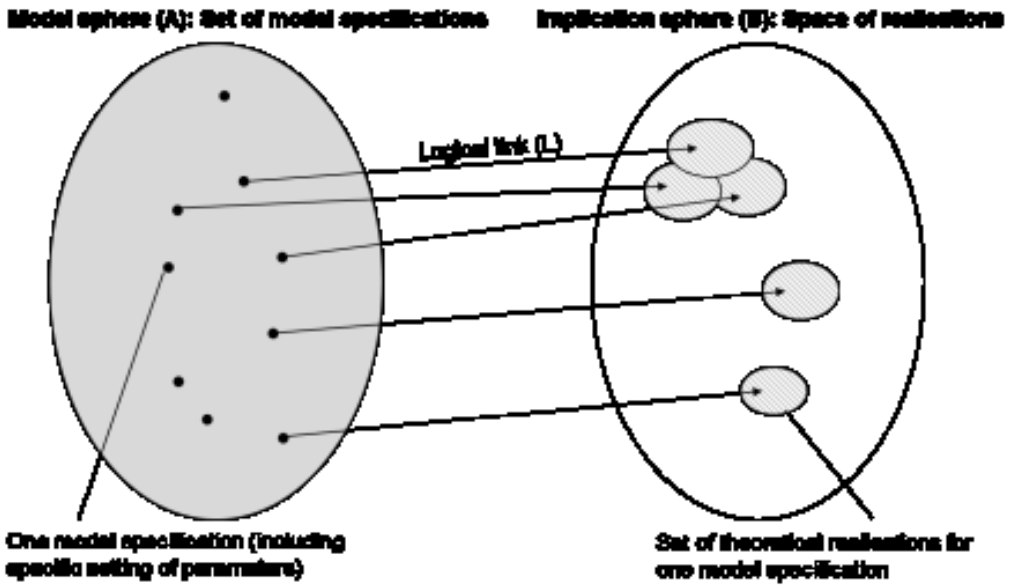


Figure 31: First step: Simulating the logical link from the model sphere to the implication sphere

In a second step, numerous model specifications (parameter settings) are simulated and the implications are compared with the knowledge on the implication level (this is done here in section 8.5). We might represent the empirical knowledge within the implication sphere as a subset representing the empirical realizations (see Figure 32). All model specifications that lead, at least partly, to realizations within this subset are in line with the knowledge on the implication level. Through this, we are able to decide for each simulated model specification whether it is a realistic setting and obtain an area of realistic parameter settings (see Figure 32). This validation step narrows down the set of model specifications to a smaller potentially realistic set.

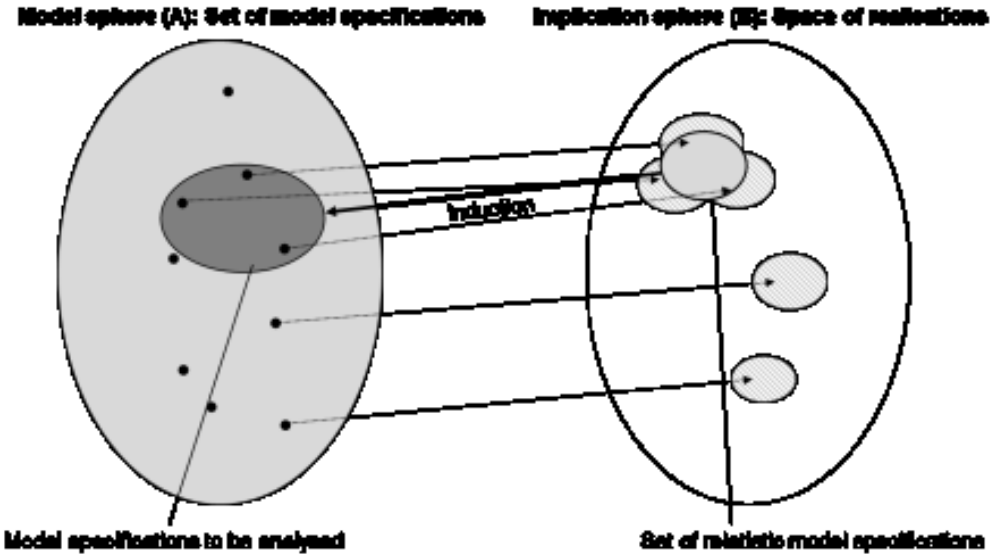


Figure 32: Second step: Inducing from knowledge on the implication level (empirical realisations) on the set of realistic model specifications in the model sphere

The final result of this simulation approach is still not one simulation model, but a set of model specifications, containing all specifications that are in line with the available knowledge on the model sphere as well as the implication sphere.

8.2.2 Application to firm growth

In the context of this paper the model sphere (A) is defined as the sphere containing the mechanisms and processes within firms that lead to firm growth. The implication sphere (B) is defined as the level on which the statistical characteristics of firm growth are analysed. Hence, the first step is to set up a firm model that is in line with the mainly theoretical knowledge about firm and market processes (section 8.4). The model is kept very general, containing many only vaguely restricted parameters. The second step is to simulate various parameter settings of this model (section 8.5, in total 34200 specifications are simulated). The results of these simulations are compared to the empirical knowledge about the statistical characteristics of firm growth (this knowledge is described in section 8.3). This allows to narrow down the set of realistic firm models tremendously.

8.3 Literature on firm growth

8.3.1 Modelling firms

Many perspectives are used in the literature to model firms and their size. (1) A traditional approach is to model firm output (as a measure of size) with a production function, which relates output to inputs such as labour and capital (e.g. GRILICHES/MAIRESSE 1995, BRYNJOLFSSON/HITT 1995). (2) Another traditional approach focuses on sales and deduces them from competition models. Various market structures and mechanisms are used in this strand of literature. For our approach only the idea of monopolistic competition (DIXIT/STIGLITZ 1977) is relevant. (3) A quite different perspective is taken in studies that aim to determine the optimal size of firms (e.g. LUCAS 1978). (4) Finally, within the literature on firm growth a number of models have appeared in recent years that are built in an attempt to reproduce the known statistical characteristics of firm growth. Interestingly, these models do not build on the other three approaches.

The main reason for this is that the first three approaches do not aim at explaining dynamics. Furthermore, the first approach rather deals with the question of how a firm can use inputs to produce a certain given output. Hence, the output is not the *explanandum* but the *explanans*. In contrast, the second and third approach have the size of firms as *explanandum*. Hence, although they do not aim at explaining the dynamics of firm sizes, they can be of help in modelling them. In particular, we believe that firm size and its dynamics cannot be explained without considering market success and, thus, competition on the market.

However, the starting point for our modelling are the already existing models. These models are foremost concerned to explain the emergence of empirically observed patterns in firm growth, like the puzzling deviation from normality in the growth rate distributions or the scaling of their variance with firms' size. To move beyond GIBRAT's (1931) model of proportional growth, which by most researchers is rejected on ground of these empirical findings and the assumption of an independent, random growth process, several models have been developed. AMARAL et al. (1998) introduced a hierarchical model with subunits that evolve according to a random multiplicative process, which yields a Laplace distribution of growth rates and a power-law in the variance-scaling relationship. FU et al. (2005) show that proportional growth is still able to reproduce these two statistical patterns if both the number of subunits and the size of the subunits are allowed to grow. In SCHWARZKOPF et al. (2010), a different mechanism within the hierarchical structure is assumed. Here, the subunits are replaced from a replication distribution, and originate either as new ones or are taken from another firm. In order to avoid any assumptions on the internal structure of firms, BOTTAZZI and SECCHI (2006a) model the growth rate distribution by drawing on PENROSE's (1959) idea of competencies and learning and by explicitly considering the market dimension. Basically, they assume that the firms' ability to take up new business opportunities increases with the number of opportunities already exploited. A self-reinforcing mechanism in the competition for limited market opportunities ultimately leads to fat-tailed growth rate distributions. However, these models focus only on specific stylized facts. Other empirical observations, like the temporal auto-correlation structure, are often neglected or stand even in contrast to the models' implications (COAD 2012). Besides, as these models aim to be very general, they are only loosely grounded

on economic mechanisms. For instance, in SCHWARZKOPF et al. (2010), the Laplace distributed growth rates directly follow from the assumption of power-law replication function.

Hence, more complex Agent-Based models have been developed, which try to include knowledge on economic processes and mechanisms in their assumptions. While DELLI GATTI et al. (2005) model explicitly the demand and supply on the credit market and the financial fragility of firms, METZIG and GORDON (2014) match employees to firms, which are required to produce goods and to generate profit. Here, firms compete both for aggregate demand and workforce. While these models are able to reproduce a larger number of stylized facts, they are quite specific, focusing on specific aspects. As discussed in section 2, our aim is to develop a rather general model based on the knowledge about processes and mechanisms within firms. According to the simulation approach by BRENNER and WERKER (2007), the modelling on the model sphere should be not influenced by any knowledge on the implication level. In the case of firms, this is not completely possible because firms develop individually so that little general knowledge about their development is available. The existing theoretical models are often developed with some statistical characteristics of firm growth in mind (see above). All that we can do is to orient our own modelling approach more on the firm models that are more general and do not explicitly aim to reproduce all statistical characteristics of firm growth.

Furthermore, we can use theoretical literature on firms that is not connected to any empirical examination of firm growth. Two theoretical ideas are of interest here. First, DIXIT and STIGLITZ (1977) developed the model of monopolistic competition, which is based on the idea that products differ so that submarkets exist in which certain firms might dominate and are able to decide about prices more freely. Second, IJIRI and SIMON (1977) developed the idea that each firm faces a set of growth opportunities. As a consequence, firms grow in dependence of how many of these opportunities they are able to realise. BOTTAZZI and SECCHI (2006a) developed this idea further by assuming that the total number of growth opportunities for all firms is limited.

By combining both ideas we might assume that a market consists of many sub-markets (determined by specific consumer preferences) with firms competing for these sub-markets and each firm dominating some of them. According to PENROSE's (1959) theory of the growth of the firm, firms compete on basis of their competencies, and from a dynamic perspective they also compete on how they develop and advance their capabilities (TEECE et al. 1997). However, markets are also not static. Here, innovation processes are repeatedly seen in the literature as an important aspect in the competition for new markets. NELSON and WINTER (1978) have framed the notion of „Schumpeterian competition“ in this context. ERICSON and PAKES (1995) modelled firm behavior as an exploration of evolving market places by investing in research. Hence, we might assume that firms compete for existing and emerging sub-markets on basis of their competencies and the research they perform. Such a perspective is used below.

8.3.2 Characteristics of firm growth

For the second step of our simulation approach we have to collect the available empirical knowledge on the implication sphere, meaning the knowledge about the statistical characteristics of firm growth. Many empirical studies have been conducted examining various aspects of firm growth.

First of all, the observed patterns of firm growth are more stable and invariant than the corresponding size and age distributions, which are strongly history-, industry- and country-specific (BOTTAZZI et al. 2007, DOSI et al. 2010). Some of the detected statistical characteristics, like the variance-scaling relationship or the fat-tailed growth rate distributions already mentioned above, are even regarded as universal laws which seem to hold in the growth of all complex organizations. That means they are independent of the particular details of systems (LEE et al. 1998). Hence, stylized facts on firm growth represent adequate selection criteria in the implication sphere for identifying reasonable specifications of the model.

Studies testing GIBRAT's law of proportionate growth find that the average growth rate is not independent of but decreases with the size of firms (e.g., EVANS 1987, DUNNE/HUGHES 1994, BOTTAZZI et al. 2011). Although some deviation from this negative relationship is observed for specific sub-populations of firms (see COAD 2009 for a more extensive discussion), this "statistical regularity" (SUTTON 1997) is often considered for the construction and validation of theoretical firm growth models. Furthermore, these studies observe that this negative dependence of average growth is also valid for firm age. In the literature, it is often argued that firm age, which is strongly correlated with firm size, is causally even closer connected with average growth (EVANS 1987, DUNNE et al. 1989, GEROSKI/GUGLER 2004).

More recently, the literature has departed from merely focusing on average growth rates. In the econophysics literature (e.g., STANLEY et al. 1996) it was pointed out that the variance of growth rates scales with the logarithm of firm size as a power law. A variance-scaling parameter coefficient between -0.15 (STANLEY et al., 1996) and -0.20 (AMARAL et al., 1997) is often reported, although some studies on firm dynamics in France (Bottazzi et al. 2011) and China (DUSCHL/PENG 2014) find a coefficient as low as -0.07 and -0.06, respectively. As it is generally true for power laws (GABAIX 2009), many possible generating mechanisms exist for the variance-scaling relationship of firm growth rates, with AMARAL et al. (2001) and BOTTAZZI and SECCHI (2006b) providing two alternative theoretical models.

Another stylized fact concerns the unconditional distribution of firms' growth rates. The earlier studies in the literature (e.g., STANLEY et al. 1996, BOTTAZZI et al. 2001) show that growth rates significantly deviate from normality due to the presence of fat tails and are much closer to the tent-like shaped Laplace distribution. This finding is confirmed for firms in various countries (e.g., DUSCHL et al. 2014 for Germany), in different industries (BOTTAZZI/SECCHI 2003) and for alternative measures of firm size (ERLINGSSON et al. 2013). More recently, tails that are even fatter than expected from the Laplace distribution and an asymmetric shape of the growth rates distribution is observed (BOTTAZZI et al. 2011, REICHSTEIN/JENSEN 2005). Hence, BOTTAZZI and SECCHI (2011) suggest to use the more

flexible Asymmetric Exponential Power (AEP) distribution, which has five parameters, to account both for fatness of the tails and asymmetry in the growth rate distributions. By comparing this theoretical distribution to alternative distributions, like the Student-t, Cauchy or Levy-stable distribution, FAGIOLO et al. (2008) demonstrate empirically that it is the preferred specification, as it is especially flexible in the tail behaviour.

Finally, the empirical literature on firm dynamics has also strongly focused on the temporal auto-correlation structure of growth rates. The results from decades of research are, however, mixed and often conflicting (CAVES 1998, COAD 2009). Depending on the country, industry and number of lags considered, some studies report positive autocorrelations over time (e.g. CHESHER 1979), others report negative ones (e.g., BOTTAZZI et al. 2011) or even fail to detect any temporal correlation structure (LOTTI et al. 2003). New light on the study of temporal auto-correlation has been shed by COAD (2007) who additionally takes into account the size of the firm as well as the growth rate itself. First, he demonstrated that the larger the firm, the larger also the auto-correlation coefficient. For very small firms, the coefficient tends to be negative, as they follow a more stochastic growth path, while for large firms it becomes positive and indicates a more stable growth path. Secondly, firms in the tails of the unconditional growth rate distribution are less likely to repeat their extreme growth experience in the subsequent year.

In contrast to the statistical characteristics of growth rates, a more heterogeneous picture emerges for the corresponding distributions of firm size and firm age. Whereas an exponential distribution seems to fit reasonable well the age distribution (COAD 2010), an ongoing debate exists on the appropriate distributional model for firm size. Theoretical and empirical evidence exists both for the log-normal (e.g. CABRAL/MATA 2003) as well as the Pareto distribution (e.g. AXTELL 2001), which seems to fit better beyond a certain size threshold. Even if common patterns in terms of the functional form might exist, the parameters are strongly influenced by the size of the economy (GAFFEO et al. 2003) or by historical contingencies, like the bumps in the age distributions due to both world wars. Furthermore, these patterns do not survive at the more disaggregated level of industries. Here, often bimodal size distributions are observed (BOTTAZZI et al. 2011). Therefore, we restrict the analysis of the simulation outcomes to one of the few robust stylized facts about the size and age distribution of firms, which has not been put into question since GIBRAT (1931): both distributions are found to be highly skewed to the right. Put differently, the coefficients of the location and scale of the size and age distribution can be only used as additional information in the assessment of the simulation outcomes, but not as criteria to reject the corresponding simulation model specification.

8.4 Simulation model

8.4.1 Basic considerations

A common feature of many previous firm models that are able to resemble the known statistical characteristics of firm growth is the definition of sub-units or business opportunities. Shortly after realising that firm growth does not show completely unstructured, random dynamics – as assumed by GIBRAT (1931) – the existence of sub-

units was used to explain the observed statistical characteristics (AMARAL et al. 1998). This leads to discrete steps in the development of the size of firms instead of continuous dynamics, which is more in line with empirical findings. An alternative model (e.g. BOTTAZZI/SECCHI 2006a), which tries to avoid assumptions on the internal structure of the firm, uses business opportunities as units, moving the concept of units to the market side. The original idea was introduced by IJIRI and SIMON (1977), who argue that growth results from the number of opportunities a firm is able to take up.

While BOTTAZZI and SECCHI (2006a) focus on new opportunities, their approach can be generalised by interpreting each business opportunity as an existing sub-market. BOTTAZZI and SECCHI (2006a) represent competition in their model by limiting the growth of the complete firm population to a given finite set of new opportunities. Interpreting this in the framework of monopolistic competition (DIXIT/STIGLITZ 1977), each business opportunity can be seen as a sub-market that is dominated by one firm.

Since we believe that competition is the main underlying mechanism of firm growth we go beyond the existing models and explicitly consider sub-markets and competition for these sub-markets. Sub-markets are not assigned to firms once at the time they appear, but there is permanent competition for sub-markets. However, following the idea of monopolistic competition, at each time each sub-market is dominated by one firm that is a quasi-monopolist in this sub-market. Furthermore, we explicitly consider firm characteristics that might change with time and assume that firms compete for sub-markets on the basis of these characteristics, an idea rooted in the dynamic capabilities approach (PENROSE 1959, TEECE et al. 1997).

As a consequence, the model contains two kinds of units of observation: markets and firms. In the following two subsections we discuss the modelling of the dynamics of markets and firms one after the other. All parameters of the model are denoted by μ and σ with the respective indices. $N(\mu,\sigma)$ stands for the normal distribution with the respective mean and variance.

8.4.2 Market dynamics

As outlined above, we assume monopolistic competition and model the market as a set of sub-markets that are each dominated by one firm. We call the sub-markets 'market packages' to signify that they have a certain *size* and are *possessed* by certain firms. Competition works on these market packages with always one firm winning the competition and possessing the whole market package.

As a consequence, the respective market is given at any point in time t by the set of market packages $M_{tot,t}$. Each market package $m \in M_{tot,t}$ is characterised at each point in time t by its size $d_{m,t}$ and its age $a_{m,t}$. Hence, the firm that possesses a certain market package m – meaning that it supplies this sub-market – faces a demand of $d_{m,t}$ from this sub-market. To keep the model simple, we do not care about further characteristics of market packages. As a consequence, we have to model three processes: (1) new market packages might appear (the process modelled by BOTTAZZI/SECCHI 2006a), (2) market packages might disappear, and (3) market packages might change their size.

Appearance of market packages

The simplest way of modelling the appearance of market packages is an independent random process. Hence, new market packages appear at each point in time (each day) with a certain probability $\mu_{m,new}$.

If a market package appears, its size $d_{m,t}$ is randomly drawn from a uniform distribution between *0* and $\mu_{d,max}$.

Disappearance of market packages

A similarly simple way is used for the disappearance: Each existing market package has at each point in time (each day) a certain probability ($\mu_{m,ex0}+\mu_{m,exa}*a_{m,t}$) *to disappear*. This probability increases with the age of the market package.

However, there is a further process that leads to the disappearance of market packages. The market packages randomly change in size (see below). If the size of a market package falls below *0.005* (million EUR), it also disappears. However, in most settings this happens very rarely.

Changes in the size of market packages

Market packages change their size randomly. This random change is assumed to follow a normal distribution and to be proportional to the actual size. Hence, it can be mathematically written by

$$d_{m,t+1} = d_{m,t} * (1 + N(\mu_{d,mean}, \sigma_d), \tag{46}$$

Each day, the market packages fluctuate in their size randomly, some packages becoming larger while others become smaller. Hence, GIBRAT's (1931) proportional growth model is introduced here at the level of market packages. In addition, there is also a trend in the development of the whole market given by $\mu_{d,mean}$. Besides the appearance probability $\mu_{m,new}$, this allows to model growing and shrinking markets.

In addition, we assume a yearly fluctuation in order to consider fashion trends or other medium-term fluctuations in demand. Hence, the real demand $D_{m,t}$ for each market package is given by

$$D_{m,t} = d_{m,t} * N_{m,y}(0, \sigma_y), \tag{47}$$

where $N_{m,y}$ is independently drawn for each market package *m* in each year *y*.

8.4.3 Firm dynamics

Firms are modelled on the basis of the market packages that they possess. Hence, competition plays a major role in our model for determining the size of firms. Firm exits are not explicitly modelled. Firms exit if they lose their last market package. Each firm *f* is characterised at each point in time *t* by the set of market packages $M_{f,t}$ that it possesses. Thus, the demand it faces is given by

$$D_{f,t} = \sum_{m \in M_{f,t}} D_{m,t}. \tag{48}$$

Therefore, we have to model mainly the competition process that determines which market package is possessed by which firm. Empirical studies show that only a small part of firm success can be explained by factors that are easily observed from the outside. Decisions and structures within the firm are decisive. Part of the story is whether firms approach the right markets at the right time, part of the story are internal processes and decisions that affect competitiveness. In an abstract approach, both parts have to be modelled as random processes. The former is modelled below. In order to reflect variations in the competitiveness of firms, a competitiveness value $C_{f,t}$ is assigned to each firm when it is founded. The initial competitiveness $\mu_{c,init}$ is the same for all firms. Then every year the competitiveness is updated according to

$$C_{f,t} = \mu_{c,r} * C_{f,t} + N(0, \sigma_c), \tag{49}$$

meaning that the development of competitiveness follows a random walk process. Two kinds of competitions have to be modelled: (1) the competition for new market packages and (2) the competition for existing market packages. While some firms focus on new markets – we might call them innovators –, other firms focus on existing markets – we might call them imitators. To model this difference, a random value s_f between 0 and 1 is drawn for each firm when it is founded. The higher s_f, the more innovative is a firm.

Competition for existing market packages

We assume that markets show a certain stability in the sense that, in general, sub-markets stay with the same firm. Specific events, such as technological changes or fashion changes, are necessary to lower the grip of firms on sub-markets and allow other firms to enter. For simplicity, we assume that such events occur randomly, so that at each point in time *t* there is a certain probability $p_{mov,m,t}$ that an existing sub-market becomes the object of competition. This probability might increase with the age of a market package as well as with the time that it is owned by the same firm. Therefore, we define the probability $p_{mov,m,t}$ as

$$p_{mov,m,t} = \mu_{mov,0} + \mu_{mov,a} * a_{m,t} + \mu_{mov,o} * o_{m,t}, \tag{50}$$

where $o_{m,t}$ denotes the time that market package *m* is owned by the same firm.

If a market package becomes the object of competition, the firm currently owning the market package has to compete with one randomly drawn other firm. The competing firm might be a new firm or an already existing firm. We define a probability $\mu_{com,start,exist}$ which stands for the basic probability that the competitor is a new firm. Hence, with this probability a new firm is created that tries to gather the market package. Otherwise, the competing firm is drawn from the existing firms with the likelihood of each firm being proportional to

$$1 + \mu_{p,turn} * T_{f,t}, \tag{51}$$

meaning that larger firms have a higher likelihood. In this draw, a new firm is also considered with a likelihood proportional to $\mu_{T,new}$ in order to reflect the fact that the probability of new firms entering is higher if the actual market activity (total turnover of existing firms) is lower. This can be seen as part of a life cycle development in which entries become less likely the more established the market and the existing firms are (e.g. KLEPPER 1996).

The competition strength $C_{exist,m,f,t}$ of a firm *f* with respect to an existing market package is calculated according to

$$C_{exist,f,t} = \left(1 + C_{f,t} + \mu_{exist,cap} * \left(T_{f,t} - D_{f,t}\right)^3\right) * (1 + \mu_{exist,s} * (1 - s_f), \tag{52}$$

where $\mu_{exist,cap}$ and $\mu_{exist,s}$ are parameters. $\mu_{exist,cap}$ determines the importance of capacity effects. Production and, hence, turnover gradually adapts to demand (see below). As a consequence, if demand changes quickly, a firm might have problems to adapt, which might have an impact on the chance and willingness of the firm to obtain or loose further market packages. We assume that such an effect becomes relevant only for large difference between turnover and demand and, therefore, use the cubic form of this difference above. $\mu_{exist,s}$ determines the importance of the strategy variable, implying that firms that focus on imitative behaviour have higher chances to obtain existing market packages.

As mentioned above, we assume that the ownership of market packages shows some stability. Therefore, another firm is only able to gather a market package if it has a competence value $C_{exist,m,f,t}$ that exceeds the one of the currently owning firm by a certain amount $\mu_{threshold}$.

Competition for new market packages

The appearance of new market packages is determined as described above (section 8.4.2). The competition for new market packages is modelled similar to the competition for existing market packages. Again there is a certain probability $\mu_{com,start,new}$ that a new firm obtains the new market package. For each existing firm f a competition strength $C_{new,m,f,t}$ is calculated according to

$$C_{new,f,t} = 1 + \mu_{new,turn} * T_{f,t} + \mu_{new,size} * T_{f,t} * D_{m,t} + \mu_{new,cap} * \left(T_{f,t} - D_{f,t}\right)^3 + \mu_{new,inno} * I_{f,t}, \tag{53}$$

where $\mu_{new,turn}$, $\mu_{new,size}$, $\mu_{new,cap}$ and $\mu_{new,inno}$ are parameters. $\mu_{new,turn}$ determines whether larger firm have a higher probability to win new market packages. $\mu_{new,size}$ stands for the fact that larger market packages might be more likely won by larger firms while smaller market packages are more likely won by smaller firms. $\mu_{new,cap}$ has the same meaning as described for $\mu_{exist,cap}$ above. $\mu_{new,inno}$ determines the dependence of the competition strength on the innovation potential $I_{f,t}$ of the firm (see below). Again, a potential start-up is assigned a competition strength of $\mu_{start,new}$, and the competition process is a random draw of a winning firm.

Development of firm characteristics

The potential demand $D_{f,t}$ that a firm f faces at time t is given by equation (48). However, the turnover of a firm must not automatically be the same as the potential demand. Especially if new sub-markets appear, it might take some time before a firm is able to satisfy the demand and/or before the customers are aware of the new possibility. Hence, we define a turnover variable $T_{f,t}$ that represents the real sales of a firm. We assume that this variable adapts towards the potential demand at each time step according to

$$T_{f,t+1} = T_{f,t} + \mu_{T,adapt} * (T_{f,t} - D_{f,t}), \tag{54}$$

where $\mu_{T,adapt}$ is a parameter that determines the speed of this adaptation.

Furthermore, we explicitly model the innovation process of firms. Following other approaches (such as NELSON/WINTER 1978 and ERICSON/PAKES 1995), we assume that innovation activities play an important role in firm growth and that firms develop routines and strategies that make them more or less innovative. Therefore, in our model firms are characterised by their innovativeness $I_{f,t}$ at each point in time t. A parameter $\mu_{inno,start}$ is defined which represents the innovativeness of a firm when it is founded. During the existence of a firm, its innovativeness develops according to

$$I_{f,t+1} = \mu_{red,inno} * I_{f,t} + D_I(f,t) * s_f, \tag{55}$$

where $\mu_{red,inno}$ is a parameter and $D_I(f,t)$ is a function that is 1 if firm f wins a new market package – which can be interpreted as an innovation – at time t and 0 otherwise. Hence, the innovativeness of firms increases with each successful innovation and decreases slowly (the speed is given by $\mu_{red,inno}$) thereafter. Innovation success is assumed to increase the innovative capability of firms. The increase of the innovative capability with each innovation success is modelled proportional to the strategy variable s_f, meaning that firms with a higher focus on innovation stay on average on a higher level of innovativeness.

8.5 Model calibration and implications

8.5.1 Simulation procedure

The above model contains in total 28 parameters. The following model calibration has two aims. First, we check whether the model is able to generate realistic firm dynamics. Second, we identify those parameter sets that lead to realistic firm dynamics.

One central finding regarding the first aim is that realistic firm dynamics are only obtained if we simulate the whole development of an industry or economy. Starting the simulation model with a random firm population has caused deviations from the known characteristics of firm growth. It would have been necessary to start with an empirical firm population, which would not be in line with testing whether the simulated firm population matches empirical knowledge. Therefore, we start all simulations with a situation with one market package and one firm. We run each simulation for a randomly drawn time period between 10 and 50 years. Realistic industry developments are represented by changing probabilities for the appearance of new market packages and the disappearance of existing market packages. New market packages become first more and then less likely. The disappearance of market packages becomes more likely with time. Through this, we resemble industry life-cycle dynamics. Then, 10 further years are simulated and the statistical characteristics of firms and firm growth are studied. All simulations are run with time steps of one day.

For each simulation run we randomly draw each parameter from its range (see Table X.3 in the appendix X.4). Then it is checked whether the parameter set leads to realistic firm characteristics (section 8.5.2). The aim is to identify not only one but many parameter sets. We argue that a firm model with some flexibility is needed so that, for example, different industries can be simulated.

8.5.2 Model calibration

Knowledge about the statistical characteristics of firm growth, as outlined in section 8.3.2, is used to identify the realistic parameter space of the simulation model (see the description of the calibration approach in section 8.2). As common in the literature, growth rate g of firm f is defined as the log-difference of its size, here measured by its turnover T_f, between time $t+1$ and t:

$$g_f = \log(T_{f,t+1}) - \log(T_{f,t}), \tag{56}$$

According to the literature discussed in section 8.3.2, the growth rates g_f should fulfil the following criteria (see also Table 24). First, if g_f is fitted unconditionally to the Asymmetric Exponential Power (AEP) density (for mathematical details and related issues regarding the inference, see BOTTAZZI/SECCHI 2011), the estimated shape parameters b_l and b_r should indicate the existence of fat tails on both sides of the distribution. The smaller these parameters, the fatter the tails of the density function at the respective side of the mode. In case of $b_l = b_r = 2$, it converges to the normal distribution, and in case of $b_l = b_r = 1$ to

the Laplace distribution. Hence, we set the cut-off value to 1.3, which is one of the highest values observed in literature for b_l or b_r, although most studies find values lower than 1.

Secondly, the variance of g_f should scale negatively as a power law with the log size of *f*. We follow BOTTAZZI et al. (2014) in modelling this variance-scaling relationship directly by introducing a heteroskedasticity term into the stochastic growth process. The resulting scaling parameter *beta* should lie within the highest and lowest values observed in the literature, that is in the interval [-0.25 < *beta* < -0.06].

Next, to assess the stylized facts concerning the average growth rate, we estimated a GIBRAT-like growth regression (AUDRETSCH/LEHMANN 2005), in which g_f is regressed on the log of size log($T_{f,t}$), age a_f, past growth rate $g_{f,t-1}$ and an interaction term between $g_{f,t-1}$ and log($T_{f,t}$):

$$g_f = b_0 + b_1 \log(T_{f,t}) + b_2 a_f + b_3 g_{f,t-1} + b_3 \log(T_{f,t}) * g_{f,t-1} + u \qquad (57)$$

with *u* denoting an independent, but possibly not normal distributed error term. Therefore, equation (57) is estimated using least absolute deviations (DASGUPTA/MISHRA 2004). To assess the temporal auto-correlation structure along the entire conditional growth rate distribution, the equation is also estimated at the quantiles $\theta_{0.1}$, $\theta_{0.25}$, $\theta_{0.75}$ and $\theta_{0.9}$ applying quantile regression techniques. To conform to the literature, b_1 and b_2 should be negative, indicating negative dependence of average growth on size and age, respectively. Although the literature is inconclusive about the absolute value of b_3, it shows that the temporal auto-correlation should be smaller at the tails of the conditional growth rate distribution, meaning that extreme growth events are less likely to repeat. Furthermore, temporal auto-correlation is known to be larger for larger firms, hence the coefficient of the interaction term b_4 should be positive.

Finally, less robust knowledge exists about the distribution of firm population characteristics, like size and age. Here, we only exclude simulation outcomes in which one basic fact is not fulfilled: the right-skewedness of the size and age distribution. This can be assessed by comparing the mean and median of the corresponding distributions.

Table 24: Selection criteria for empirical model validation

Statistical characteristic	*Parameters*	*Selection values*
Fat-tailed growth rate distribution	Shape parameter of AEP (b_l, b_r)	[b_l < 1.3] [b_r < 1.3]
Variance-scaling relationship	Scaling parameter *beta*	[-0.25 < *beta* < -0.06]
Negative dependence of average growth on size	Coefficient of size b_1 growth regression	$b_1<0$
Negative dependence of average growth on age	Coefficient of age b_2 growth regression	$b_2<0$
Smaller auto-correlation coefficient at tails of growth rate distribution	Auto-correlation coefficient b_3 in growth regression at different quantiles θ	$b_{3,\theta 0.1} < b_{3,\theta 0.25} < b_{3,\theta 0.5}$ $b_{3,\theta 0.9} < b_{3,\theta 0.75} < b_{3,\theta 0.5}$
Positive size-dependence of auto-correlation coefficient	Coefficient of interaction term b_4 of lagged growth and size in growth regression	$b_4>0$
Right-skewed size distribution	Mean and median of size T_f	mean(T_f) > median(T_f)
Right-skewed age distribution	Mean and median of age a_f	mean(a_f) > median(a_f)

In total we examine 34200 model specifications and find nine specifications that lead to firm growth processes that satisfy all our selection criteria. Hence, our first aim is reached. The developed model is, indeed, able to reproduce many characteristics on firm growth that are observed in reality. However, we also find that the ranges of the model parameters have been chosen much too large. Only approximately 0.03 % of our model specifications show realistic dynamics. The huge parameter ranges have been chosen to be sure that no realistic specification is missed. For future approaches the ranges can be narrowed and the discussion in the next subsection helps to define adequate restrictions.

8.5.3 Identified models

As argued above, we identify a number of parameter sets that lead to realistic firm characteristics and dynamics. In total, nine realistic parameter sets result from 34200 model specification that we tested. They are listed in Table 25. In addition, the average size and age is given in the bottom lines of Table 25.

Table 25: Realistic parameter sets

Parameter	*Set 1*	*Set 2*	*Set 3*	*Set 4*	*Set 5*	*Set 6*	*Set 7*	*Set 8*	*Set 9*
$\mu_{d,max}$	8.12	8.15	.21	.26	.21	1.88	1.45	2.66	2.91
$\mu_{d,mean}$	-.00010	-.00015	-.00097	-.00041	-.00024	.0000048	.00014	.00036	.00021
$\mu_{m,ex0}$	.00082	.000022	.00020	.00027	.00031	.000048	.00011	.00064	.00025
$\mu_{m,exa}$	.00017	.00019	.00019	.00012	.00059	.00048	.00082	.00060	.00029
$\mu_{c,init}$	.012	.015	.000040	.0044	.000058	.012	.0012	.0030	.045
$\mu_{c,r}$	.94	.85	.81	.85	.91	.82	.80	.99	.83
$\mu_{T,adapt}$	.0014	.00021	.00065	.00019	.00014	.071	.0050	.047	.024
$\mu_{inno,start}$	.37	.079	.0012	.0021	.037	.059	.21	.0071	.37
$\mu_{red,inno}$	.994	.998	.994	.997	.995	.996	.997	.994	.9998
$\mu_{move,0}$	.035	.036	.058	.069	.063	.032	.018	.050	.0052
$\mu_{move,a}$	.000048	.00087	.000038	.00067	.00092	.000025	.000014	.00075	.00067
$\mu_{move,o}$	.000005	.000013	.000012	.000026	.000010	.00078	.000029	.000018	.000064
$\mu_{threshold}$	.00011	.00023	.00013	.00023	.0019	.0036	.0016	.0086	.025
$\mu_{com,start,exist}$	.0020	.0021	.0181	.0082	.0097	.0095	.0081	.0087	.064
$\mu_{p,turn}$	.0094	.0014	.0111	.0082	.018	.0046	.0062	.0059	.0076
$\mu_{T,new}$	21.87	1.43	18.05	28.48	15.87	5.05	3.40	1.68	3.35
$\mu_{exist,cap}$	.00014	.000025	.000031	.051	.034	.00086	.00049	.00038	.0043
$\mu_{exist,s}$	.000011	.0040	.128	.17	.033	.000046	.012	.045	.00066
$\mu_{m,new}$	4.80	.918	6.68	5.04	3.32	.67	.16	.171	.20
$\mu_{start,new}$	308	353	2.02	4.18	9.41	333	550	233	4.67
$\mu_{com,start,new}$	.066	.013	.10	.26	.30	.025	.013	.071	.24
$\mu_{new,turn}$	-.0084	.0055	.0092	.0023	.0039	.0027	-.0031	.0093	.0018
$\mu_{new,size}$	4.0E-07	.000005	1.1E-08	4.8E-07	9.9E-09	1.7E-10	1.2E-09	1.6E-10	1.4E-09
$\mu_{new,cap}$	.89	.0016	.0022	.13	.0000042	.32	.000016	.83	.0025
$\mu_{new,inno}$	.015	.00012	.00067	.17	.036	.000021	.072	.00089	.021
σ_d	.0022	.0025	.0079	.0066	.0021	.00055	.0056	.0048	.0070
σ_y	.00065	.0052	.0056	.0070	.069	.053	.0057	.10	.065
σ_c	.00013	.0097	.00090	.00071	.00012	.0039	.00093	.00056	.0032
Mean $T_{f,t}$	10.0	45.8	.25	.32	.32	1.85	1.55	3.30	3.39
Mean $a_{f,t}$	5.00	6.21	10.5	8.75	8.07	16.4	10.8	15.1	5.52

The sets of realistic parameter settings are interesting for two reasons. First, since we tested wide ranges for the parameters, the realistic parameter values provide information about the strength and importance of mechanisms. Second, the various realistic parameter sets represent different situations, so that they enable the modelling of industries with different characteristics.

Quite some of the parameters seem to be not of crucial importance for whether the simulation shows realistic firm behaviour or not. However, some parameters are in all realistic simulation runs quite similar and the originally fixed parameter ranges are found

to be too large. Six parameters of this kind can be identified and will be discussed in the following. The increase ($\mu_{m,exa}$) of the disappearance probability with the age of market packages is rather on the upper end of the originally set range, meaning that this is an important, realistic mechanism. The stability ($\mu_{c,r}$) of firm's competition strength never takes values below 0.8 in realistic simulation runs, meaning that competition strength changes rather slowly in reality. This aspect is strengthened by the fact that also the variance (σ_c) of competition strength never takes higher values in realistic simulation runs. Values on the upper end of the original range are not found for the innovative ability of start-ups ($\mu_{inno,start}$), meaning that an extremely high innovativeness of new firms is not a realistic feature. The increase ($\mu_{move,o}$) of the probability to lose a market package with the time of owning it lays in all realistic simulation runs on the lower end of the parameter range. Put differently, this dependence is very weak if existing at all. Maybe we could even withdraw this effect from the model. A similar result is obtained for the dependence ($\mu_{new,size}$) of winning a new market package on the matching between the size of the firm and the size of the market package. This dependence is either weak or not existing.

The second intention is to find parameter variations that can be used to simulate different realities such as different industries or different stages in the industrial life-cycle. The parameter sets identified and listed in Table 25 show that this aim is reached. While three parameter sets (Sets 3 to 5) are quite similar, all other identified parameter sets show clear differences. Hence, they indeed represent different kinds of markets and can be used to simulated different industrial realities. This will be discussed in more detail in the following.

A general observation is that the maximal market package size ($\mu_{d,max}$) influences strongly the average size of firms. However, this often comes along with other characteristics. Some of the parameters are not independent of each other. To produce realistic model behaviour, some parameters have to take corresponding values. The maximal market package size ($\mu_{d,max}$) is clearly connected to two other parameters: High average market package sizes come together with high initial competition strengths ($\mu_{c,init}$), implying a high stability of competitiveness, and low probabilities of losing market packages to start-ups ($\mu_{com,start,exist}$). Hence, large market packages are connected to a high stability of market package ownership and lead to large average firm size.

In general, larger market package size come also together with lower competition strengths of start-ups ($\mu_{T,new}$) and a lower number of new (innovative) market packages ($\mu_{m,new}$). However, Set 1 builds an exception in this context and deserves more detailed discussion. While in general Set 1 is quite similar to Set 2 with relatively large market packages and firm sizes, it represents a market with a high number of new (innovative) market packages ($\mu_{m,new}$) and a high competitiveness of start-ups ($\mu_{T,new}$), two characteristics that are found rather in volatile markets with small average firm sizes (Sets 3 to 5). This higher volatility results also in a lower average firms size compared to Set 2. However, to some extent this volatility is counterbalanced by a high probability of firms with free production capacities to enter new (innovative) markets ($\mu_{new,cap}$). Thus, Set 1 represents a market that is dominated by rather large and stable firms, but shows also a high dynamic in terms of a frequent occurrence of new sub-markets and firms.

While Sets 2 to 5 can be seen as the typical cases for a market dominated by large incumbent firms (Set 2) and a volatile market with many small firms (Sets 3 to 5), the Sets

6 to 8 lie somehow between these two extreme cases. Besides this, Set 6 shows only three specificities: firms are able to adapt their production capacities quickly ($\mu_{T,adapt}$), it is impossible for firms to own market packages for a very long time ($\mu_{move,o}$) and firm strategy ($\mu_{exist,s}$) and prior innovation success ($\mu_{new,inno}$) are of low importance. Hence, Set 6 seems to represent a rather low-tech, competitive market with low fix costs in production and a very stable firm population (high average firm age).

In contrast, Set 7 is characterised by a high dependence on prior innovation success ($\mu_{new,inno}$), a high probability of smaller firms to gather new, innovative market packages ($\mu_{new,turn}$<0) and a high innovativeness of start-ups. This is somewhat counterbalanced by a low frequency of new market packages ($\mu_{m,new}$). Thus, Set 7 represents a market that is, in general, quite stable but contains a part with very innovative small firms and start-ups.

Set 8 is characterised by a very high stability of firm's competitive strength ($\mu_{c,r}$). From year to year 99% of the competitive strength remains constant. In addition, the year to year variation in competitive strength (σ_c) is also rather low. In contrast, the year to year variation in market package size (σ_y) is quite high. New market packages are rather rare ($\mu_{m,new}$). Furthermore, firms are able to adapt their production capacities quickly ($\mu_{T,adapt}$) to the demand for their products. Hence, Set 8 describes a situation in which markets rather change in size, innovations are rare and firms are flexible and quite stable (high average age of firms).

Set 9 has a number of characteristics that makes it quite specific, so that it is discussed here in more detail. First, it shows the highest value for the innovativeness of start-ups ($\mu_{inno,start}$) and an exceptionally high stability of the innovativeness of firms ($\mu_{red,inno}$). In combination with a high success-breads-success element in the innovation activity ($\mu_{new,inno}$), this leads to a strong accumulation of research capacities within firms. In addition, the ownership of market packages is very stable (highest value of $\mu_{threshold}$ and lowest value of $\mu_{move,0}$). If the ownership of a market package changes, start-ups have good chances ($\mu_{com,start,exist}$). New market packages are rather rare ($\mu_{m,new}$). Thus, Set 9 reflects a market in which existing firms build innovation capacities and dominate the rather rare events of sub-market emergence. Sub-market ownership changes rarely, but if it happens, start-ups are most likely to enter.

8.6 Conclusions

A model of firm and market dynamics is developed in this paper. The model builds on general assumptions about competition and innovation processes: firms compete for monopolistic positions in narrow sub-markets, which in turn develop independently. Depending on their competitive strength, existing as well as new firms gain or lose these so-called market packages. New market packages are conquered as an outcome of successful innovation.

Following the approach by BRENNER and WERKER (2007), all processes and mechanisms, originating from knowledge about firms and markets, are modelled as general as necessary by using many parameters that are not or only vaguely fixed. Established

stylized facts about the statistical characteristics of firm growth provide the knowledge to identify those parameter sets that lead to realistic behaviours of the model. Running 34200 parameter sets, nine parameter sets are identified as realistic.

A detailed view on these parameter sets has shown that they, indeed, can be seen as representing different markets or industries with specific characteristics. Hence, the aim to develop a flexible model that is in line with well-established knowledge about firm growth and that is able to represent different industrial situations is reached. Testing further parameter sets would provide further realistic parameter sets. However, the range of potential characteristics that the model is able to represent is already well outlined by the nine parameter sets that are identified and discussed here.

Various extensions of this simulation model are possible. First, employment dynamics can be modelled by assuming that changes in the number of employees are adjustments to the turnover a firm has realized (or is expected to realize). This requires the introduction of an adjustment function, like in SCHLUMP and BRENNER (2010). Besides, the firm model developed here might be used to study growth processes at higher levels of aggregations. This is straightforward because regional or national growth is the sum of the growth of all firms and production sites, plus firm founding, site opening and site/firm closure, within these spatial units. Hence, to study the implications of the above firm model for regional growth, the location of the firms has to be considered. Additional parameters might take into account different kinds of benefits from location, namely technological spillovers, agglomeration externalities for market competition and agglomeration externalities for innovation activities. In the literature on New Economic Geography and agglomeration studies, it is argued that these mechanisms are crucial for spatial concentration of economic activities to emerge and stabilize. An alternative explanation of industrial clusters can be found in KLEPPER (2006) showing that new firms are more often established in locations in which similar firms are already located. Our firm simulation model would allow to discriminate which mechanisms matter most under which circumstances, such as the life cycle phase or the innovation propensity of an industry.

9 Conclusions

Economic growth is a complex phenomenon, in which the precise conditions of the spatial economic landscape, the particularities of the industries and the firm level dynamics assume a crucial role. Departing from the omnipresent heterogeneity at the layers of firms, industries and space, the thesis investigated the patterns, processes and causes of how firms and regions grow. Hereby, it focused particularly on the following aspects: the distribution of growth rates, causality in growth dynamics, the role of extreme events at the firm level for regional development, the spatial dimension of agglomeration economies and the emergence of growth rate distributions.

The concluding chapter begins with a summary of the main findings. These findings answer the research questions raised in the introductory chapter 1. Second, important limitations of this thesis are discussed. Third, some conclusions concerning policy and research are drawn from the findings. Referring to the outlined limitations and conclusions, an outlook on future research challenges and opportunities completes this thesis.

9.1 Main findings

The patterns, processes and causes of economic growth are analysed in seven research papers at the level of firms and regions applying and combining quantitative methods such as distributional analyses, regression models and simulation approaches. Table 26 provides a brief overview on the main findings of each of these papers, indexed by their chapter number. In the remaining of this section, the results are discussed in a more synthetic way. Therefore, first the findings on the stochastic patterns of growth processes are brought together. Secondly, the factors influencing these patterns and causing growth processes are discussed. Finally, the spatial aspects of growth processes are summarized.

Table 26: Overview on the main findings from the papers

C.	Title of publication	Main findings
2	Characteristics of regional industry-specific employment growth rates' distributions	As observed for firms, also regional growth rate distributions are fat tailed and asymmetric; they differ systematically across industries and approach the Laplace for longer time-horizons.
3	Growth dynamics in regional systems of technological activities – A SVAR approach	Depending on the industry, a success-driven model of regional growth can be distinguished from a linear model. Some industries, like electronics, are more likely to show self-reinforcing growth dynamics.
4	Firm growth and the spatial impact of geolocated external factors	The spatial impact of external factors, like economic, educational or research activities, differs along firms' size and their growth performance. The larger the firms, the more diverse are the activities they benefit from.
5	Industry-specific firm growth and agglomeration	Firm growth is hampered by agglomeration of own-industry employment, but improved by proximate scientific activities. The growth effects and the spatial scale of industrial clusters depend on the kind of industry.
6	Firm dynamics and regional resilience: an empirical and evolutionary perspective	The tails of the growth rate distributions are fatter in regions with higher presence of qualified workers and with better overall economic performance. These regions are conceptualized as re-inventive and *resilient* from an evolutionary perspective.
7	Chinese firm dynamics and the role of ownership type - A conditional estimation approach of the Asymmetric Exponential Power (AEP) density	The distributional mass of the growth rates is affected by the ownership type in complex ways. Financial constraints seem to be an important barrier for high growth events particularly among private-owned firms. The coefficient of the variance-scaling relationship is lower than in Western economies.
8	Modelling Firm and Market Dynamics – A Flexible Model Reproducing Existing Stylized Facts	The stylized facts can be replicated by assuming general theoretical mechanisms on market and innovation dynamics. The model is flexible so that it can be adapted to certain industries and life-cycle stages.

9.1.1 The stochastic patterns of growth processes

One of the main findings of this thesis is that fat tails in growth rate distribution, a stylized fact in the literature of industrial organization and economic growth, survive at the intermediate level of regions. Chapter 2 studies industry-specific regional growth rates and chapter 6 firm growth rates within regions. For both cases, that is to say regions as the unit of analysis as well as the reference system for firm growth, the tails of the respective distributions are significantly fatter than Laplace, for which $b_l = b_r = 1$. Moreover, b_l tends to be smaller than b_r, meaning that extreme negative events are more likely to occur than positive ones. The presence of fat tails and the observed asymmetry justifies the AEP as the appropriate distributional function for describing the empirical growth rates of and within regional economies. Chapter 2 even shows that this flexible density function significantly outperforms the less general alternatives, like the symmetric EP or Laplace. Hence, this thesis strongly confirms the literature which argues to use more flexible distributional models that are able to cope with fat-tailed growth dynamics of economic entities.

Additionally hereto, this thesis contributes to the literature by identifying important differences in the parameters of the distributions. These differences are observed especially across industries and regions. Industries show different growth dynamics: both chapter 2 for industry-specific regional growth rates and chapter 6 for firm growth rates within regions confirm that manufacturing is more prone to extreme growth events in

comparison to services. Besides, particularly knowledge intensive industries show pronounced fat tails. It follows that the composition of the industry-mix of a regional aggregate is an important variable to account for in the study of regional growth dynamics. By investigating the properties of industries and relating them to the observed growth patterns, even some preliminary conclusions on the mechanisms responsible for fat tails can be drawn. For instance, industries for which knowledge assumes an important role are naturally also more susceptible to local knowledge spillovers, and thus making the latter a candidate mechanism leading to the emergence of fat tails. These mechanisms are studied in chapter 8 in more detail by devising a simulation model based on basic assumptions about markets and competition in the growth process of firms. In addition to the industry composition of a region, also the way of how and the degree to which these industries are related to each other matter. In chapter 6 it is found, for instance, that a less coherent and less interrelated industrial portfolio makes extreme growth events more likely, arguing with CASTALDI et al. (2014) that unrelated variety increases the likelihood of technological breakthroughs, which is expected to result ultimately in turbulent growth processes.

Chapter 6 identifies further regional variables that correlate with the fatness of the tails. The two most notable ones are the qualification level of the workforce as well as the overall performance of the regional economy. Extreme growth events of firms are more likely to occur in better performing regions that are equipped with a high level of human capital. Although the direction of causality is still unknown, this finding uncovers an intriguing positive association of fat tails and regional development. However, especially the long-term effects of fat tails on regional development deserve more attention in future research (see also section 9.4)

A comparison of the empirical densities of firm growth rates (pooled across regions) and industry-specific regional growth rates, as shown in Figure 33 for the yearly growth rates between 2008 and 2010, reveals that the tails of firm growth rates (black line) are less pronounced than the tails of regional growth rates (grey line). This discrepancy mainly can be attributed to an artefact residing in the properties of the data. In contrast to regional data, firm data does not include entry and exit events. These events, lumpy in nature, seem to contribute significantly to the fatness of the tails. Therefore, it remains an interesting research question of how much each of the three different firm level events – growth, exit and entry – contribute individually to the aggregate regional level distribution.

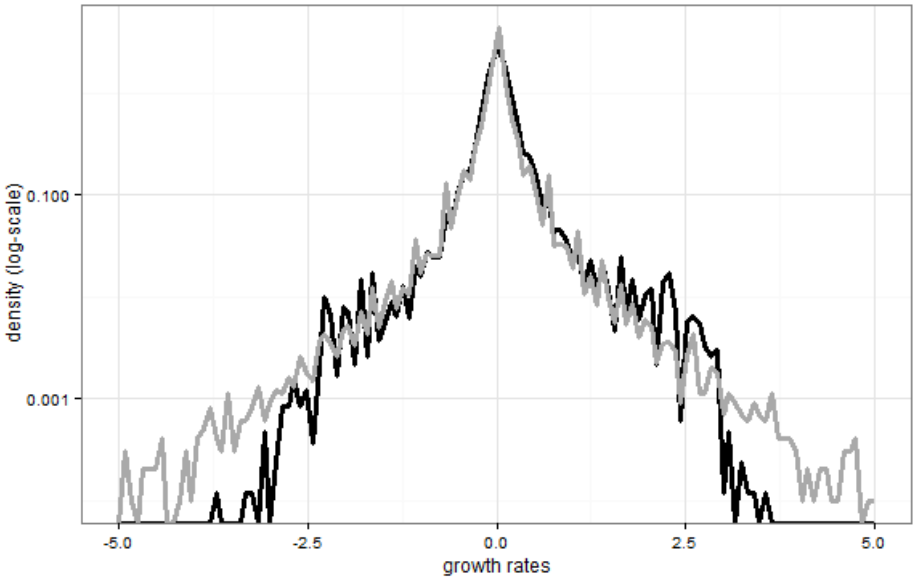


Figure 33: Empirical densities of firm growth rates (black line) and pooled industry-specific regional growth rates (grey line) for the yearly growth rates between 2008 and 2010.

In chapter 2 it is hypothesized that the distributions might be affected by the necessarily arbitrary choices of regional boundaries, the industrial aggregation level or the time scale. Notwithstanding, the results survive in various robustness checks that are performed for the named issues. Interestingly, increasing the time lag in the calculation of growth rates leads to a convergence towards the Laplace distribution instead of a progressive normalization, which could have been expected from assuming an increasing independence of the elementary growth shocks within GIBRAT's framework.

In the context of the analysis of stochastic characteristics of growth processes, the main focus of this thesis is on the growth rate distribution. However, further stylized facts are discussed in the literature, like the variance-scaling relationship, which can be confirmed here also for the level of regions. The variance-scaling parameter β, which indicates the relationship between the units' size and the observed variance of its growth rate, lies for regional economies closely around the expected value of -0.2. This finding underlines the universality of the scaling feature (AMARAL et al. 2001).

In chapter 7, the stylized facts on growth rates, which have been established in the context of several Western market economies, are investigated again to test if they survive in an East Asian economy, namely China. Although similar patterns in the growth rate distribution and variance-scaling relationship exist, supporting the universality hypothesis, some important deviations are observed. For instance, the scaling parameter is much lower and lies in the range between -0.06 and -0.12. By referring to theoretical models, this finding is explained by a rather hierarchical business culture and the more uncertain environment usually encountered in an emerging economy. Furthermore, this chapter provides evidence that high growth events are less likely among Chinese firms compared to Western countries. This fact is attributed to their lower innovation orientation and existing

barriers to growth, like financial constraints or other institutional discriminations, which are strongest among the privately owned firms. In the vein of SIMON (1968), these systematic deviations in patterns of growth rates can shed new light on particularities in firm dynamics of China.

9.1.2 Regional growth factors

The factors causing growth are analysed throughout chapters 2 to 4. In chapter 2, causal relationships among the dynamics of various technological activities of a regional economy are studied. In chapter 3 and 4, a firm level perspective is assumed to investigate the impact of various growth factors external to the firm. Hence, two different levels of aggregation are assumed to study growth factors. Working at the firm level attenuates problems stemming from endogeneity and spatial aggregation. However, it also means that a systematic comparison of the firm level and regional level results is required, because results not necessarily survive at higher levels of spatial aggregation.

Starting with universities and public research institutes, the findings of the thesis agree with the literature at the level of firms (e.g. CASSIA et al. 2009) as well as regions (e.g. SCHLUMP/BRENNER 2010) that both of them are important growth factors. By taking into account the heterogeneity of firms and industries as well as the various functional roles performed by universities, foremost education and research, a more complete picture can be drawn. First, not all firms are able to benefit from nearby public research and higher educational activities. As chapter 4 shows, these activities become beneficial only for firms above a certain size threshold: only larger firms have the required capacities to absorb external knowledge in the first place. Besides, especially large firms benefit from an adequate pool of available and qualified workers to expand strongly in absolute terms. Comparing universities' research activities with educational activities, the former show in general a higher explanatory power regarding firm growth. Hence, chapter 5 focuses exclusively on publication data and additionally introduces an industry-specific perspective. It is found, on the one hand, that nearby knowledge-generating activities tend to increase the firms' growth prospects in industries which are knowledge-intensive and in an earlier phase of their life cycle. Examples to be named are financial services, information technology, or medical devices. In these industries, firms strongly rely on input from science and also possess the necessary absorptive capacities to implement external knowledge. In more mature industries, yet science based, like chemistry, metal manufacturing or production technology, nearby knowledge-generating activities are less relevant. Firms from these industries probably tend to rely increasingly on internal research activities and are also more able to maintain collaborations with the best possible partner, irrespective of the geographical location (BRENNER/SCHLUMP 2011). On the other hand, it is found that innovation-driven growth, a feature of firms in science-based industries, results also in more negative events at the left tails of the conditional growth rate distribution. This finding will become relevant again in the comparison with the regional level results.

Regarding agglomerations with other firms, negative effects tend to prevail. In chapter 4 it is shown that only small firms are able to benefit from nearby economic activities to which

they are closely related in a technological sense. Such specialized agglomerations significantly hamper particularly the growth prospects of larger firms. One interpretation might be that an agglomeration of very similar activities cannot provide complementary knowledge, but rather tends to become a source of rivalry. This is confirmed in chapter 5, in which the effects of being agglomerated with other firms from the same cluster industries, which are defined by their technological relatedness and spatial co-location patterns, are studied. Also at an industry-specific level, negative agglomeration effects on the growth of incumbent firms are observed. By focusing on the entire conditional growth rate distribution using quantile regression techniques, it is additionally shown that an agglomeration of more diverse yet related activities reduces the firms' risk of adverse negative growth shocks (chapter 4), but the opposite holds true if the surrounding firms are from the same cluster industry (chapter 5). Put differently, specialized agglomerations seem to increase the firms' dependence on industry-specific problems.

This seeming contradiction between the observed firm level evidence and the general positive notion of industrial clusters in the literature deserves further explanation. It is proposed that firms in clusters benefit from their strongly agglomerated surroundings in terms of higher innovativeness and competitiveness, but at the same time they have to pay higher wages for their employees and higher prices for real estate due to the intensified competition (KETELS 2013). As a consequence, the broader region is often able to benefit from clusters (PORTER 2003a), but firms within agglomerations do not show higher profits or growth rates. This underlines the intuition that not all results found at the firm level might survive at the regional level. This is mainly due to the heterogeneity of firms as well as the Modifiable Areal Unit Problem (MAUP). First, this research has shown that the effects of agglomerations, universities and public research institutes strongly depend on firms' size, industry affiliation or position in the conditional growth rate distribution. Secondly, it has shown that the effects not always coincide with the spatial scales that are used in regional level studies. By ignoring the heterogeneity of the spatial landscape at the micro scale, these studies tend to either over- or underestimate the effects, depending on the chosen level of aggregation. In addition, regional level studies tend to be more affected by endogeneity issues (see also the discussion of the limitations in section 9.2.5). Hence, it is interesting to consult also the findings from chapter 3, which explicitly addresses the question of causality within an endogenous setup.

In chapter 3 the previously reported finding that research activities tend to outweigh educational activities in their contribution to regional economic growth is confirmed. Although human capital is a key variable in all recent growth models (e.g. ROMER 1990), the picture looks different at the regional level if it is operationalized by university graduates. These represent the most mobile group in our society and often leave their region of education (FAGGIAN/MCCANN 2009). Hence, they might contribute to economic growth of the national economy, but not necessarily to the growth of the region of education. Besides, it is important to note that in chapter 3 all research activities are considered, both within private firms as well as within universities and public research institutes. The firm level studies in chapter 4 and 5 focus only on public research activities. However, the general conclusion of the importance of knowledge generating activities for mainly science-based industries still remains valid. Furthermore, almost no direct impact is found for innovation on regional economic growth. The firm level results suggest that

innovation-driven growth is strongly a matter of turbulences, embracing many high-growth firms, but also a non-negligible share of strongly declining firms. Considering the micro-level results, the absence of any effects at the regional level, which rather looks at aggregates and averages, seems reasonable (the issue of aggregating out important firm level turbulences is explicitly addressed in chapter 6). Finally, there are important differences among industries in the causal structure of growth processes of technological activities. Whereas in machine tools, for instance, successful innovations seem to spur further research and economic activities, basic research takes a more fundamental position in the causal chain of growth dynamics in more science-based industries. Hence, an industry-specific perspective is mandatory both at the firm and regional level, and a derivation of the regional level effects from the firm level processes would require a complete consideration of the industrial composition of a regional economy.

Further insights on the role of agglomerations in the micro-macro nexus can be gained by a future extension of the firm growth simulation model in chapter 8 to the domain of regional growth. The simulation model can contribute to the question of the role of increasing returns due to competition advantages from agglomerations (e.g. KRUGMAN 1991), spin-off dynamics in industrial clusters (e.g. KLEPPER 1997) or local knowledge spillovers (e.g. BRESCHI/LISSONI 2001) and on how firm dynamics aggregate up to regional growth by implementing additional regional parameters on market and competition processes of firms. A simulation approach into this direction can be found also in BRENNER (2001).

9.1.3 Spatial aspects of growth and agglomeration

In chapter 4 and 5, the spatial dimension of regional growth is explicitly addressed. In anticipation of the finding that regional boundaries might bias the assessment of effects like knowledge spillovers or agglomeration economies on economic growth, the analysis is performed at the micro-level of firms. The spatial dimension of agglomeration economies is directly determined from the data by modelling the impact of external factors to firm growth using a flexible distance decay function approach, in which firms are located in a continuous economic landscape,.

The most important finding is that in some cases spatially modelled effects abruptly decay at a very narrow local range of a few minutes, whereas in other cases they go beyond traditionally defined regional boundaries. Sometimes, it is even found that different spatial scales may matter at the same time. Hence, the use of predefined regional boundaries in empirical studies would either underestimate or overestimate the impact of agglomeration on growth. The relevant distances depend much on the specific industry, the size and growth performance of the firms, or the source of agglomeration economies. For instance, larger distances seem to be required to trade-off the close cognitive distances implied by related activities of the same 4-digit industry. This finding confirms ERIKSSON (2011), who argues that the variety in heuristics and routines increases with geographic distance, thus offering a substitute to strong technological overlap. In contrast hereto, externalities that originate from public research and higher educational activities mainly occur at a geographical scale much smaller than usually assumed as “regional”. It is reasoned that small firms cooperate only with local universities or public research institutes due to travel

costs, and that larger firms either cooperate with local ones or with the best global alternatives. Finally, firms at the tails of the conditional growth rate distribution are less affected by their spatial surroundings, as already discussed in the previous section.

9.2 Limitations of the approaches used

The above discussed findings have been elicited by applying innovative statistical approaches to fine-grained empirical data. In the course of the research process several limitations have come to light. In the following, the most important limitations are discussed in more detail. These limitations concern the dependent (section 9.2.1) as well as the independent variables, or rather a lack of some important explanatory variables (section 9.2.2). Related hereto is the ignorance of other sources of heterogeneity (section 9.2.3) and a simplified conception of space (section 9.2.4). The critical reflections move on to more methodological issues, like the issue of endogeneity (section 9.2.5), the ignorance of the panel structure in the quantile regression approaches (section 9.2.6), and more specific problems encountered in the SVAR setup (section 9.2.7). Finally, the limitations of a (unconditional) distributional analysis are reflected (section 9.2.8) and opposed to the higher computational costs of a conditional estimation (section 9.2.9).

9.2.1 Focus on growth rates instead of growth paths

Growth processes are operationalized throughout all chapters, and within a large part of the literature, as growth rates, i.e. as temporal, usually yearly changes in a size measure. However, the (short-term) growth rates ignore valuable information that is contained within the (long-run) growth paths. These boil down to the question of “how” a firm or region grows instead of simply asking “how much” it grows (McKelvie/Wiklund 2010). Instead of a consistent, linear growth over time, mostly discontinuous, non-linear growth paths are observed empirically (Garnsey et al. 2006, McKelvie/Wiklund 2010). Various issues related to long-run growth paths – which forms they take (e.g. continuousness, turning points, reversals, cumulative growth; Garnsey et al. 2006), why they emerge (e.g. as a branching process, Neffke et al. 2011), or how they evolve (e.g. as path dependent process, Arthur 1994) – are still under-researched in the literature. In a recent publication, Brenner and Schimke (2014) find that the determinants of firm growth paths, operationalized as temporal sequences of expansion or decline steps, differ from the determinants of isolated growth events. Further research into this direction is currently undertaken for high-growth firms by analysing the persistence of these events (e.g. Hölzl 2014, Bianchini et al. 2014).

However, the critical reflections should not be exclusively limited to the dependent variable, the use of growth rates instead of growth paths, but also embrace the independent variables, the growth factors. Instead of repeating all the arguments that are found in the literature in favour or against specific independent variables used in this thesis, it is more fruitful to focus on the variables that are not considered to explain the growth of firms and regions.

9.2.2 The lack of institutions, networks and entrepreneurship

Three major important growth factors are ignored: institutions, networks and entrepreneurship. Much has been written on the crucial role of institutions for regional development by shaping economic (inter-)actions (e.g. RODRÍGUEZ-POSE 2013), of a-spatial networks for spillovers by directing mostly intentional knowledge flows (e.g. MAGGIONI et al. 2007), or of entrepreneurship for growth by transforming knowledge into economic opportunities (e.g. AUDRETSCH/KEILBACH 2008).

As concerns institutions and networks, these issues are often too complex and difficult to operationalize for large-scale empirical studies. At the regional level, adequate measures for both formal and informal institutions are rarely available (RODRÍGUEZ-POSE 2013). Although a huge body of literature on networks has recently surged in the context of innovation and economic growth (see, amongst others, the special issue edited by BRENNER et al. 2011), it is difficult to construct reliable network measures for a large population of firms or for a cross-section of regions. HUGGINS and THOMPSONS (2014), for example, sketch a theoretical regional endogenous growth model, in which inter-organizational networks underpinning the flow of knowledge within and across regions are explicitly modelled as a key capital input, without scrutinizing the model with empirical data. Especially completeness as an important requirement on network data is a huge barrier for empirical analyses. However, new and original data sources (e.g. BRÖKEL et al. *forthcoming*) might remedy this issue in the future.

Entrepreneurship is a further fundamental factor explaining regional growth and change (AUDRETSCH et al. 2006, GLAESER et al. 2014). Although data is available on regional entrepreneurial activities, it is difficult to be integrated into the empirical framework of this thesis. Regional growth is the sum of firm growth plus net effects resulting from firm entry and exit, which often show different dynamics and determinants (COMBES et al. 2004). Hence, the analysis of firm growth rates covers regional growth dynamics only partially, and the analysis of growth processes at the regional level mixes up both firm growth as well as exit and entry processes. At least, it might be assumed that exit and entry matter for aggregate dynamics, for example by inducing structural change (NEFFKE et al. 2014), only at time periods that are longer than analysed in this thesis (METCALFE/FOSTER 2010).

9.2.3 An incomplete picture of heterogeneity

An important aim of this thesis is to acknowledge the heterogeneity encountered in the real world in order to gain a deeper understanding of growth processes and the related driving or causing forces. Being far away from covering all relevant growth factors, as argued in the previous section, the thesis ignores also many other sources of heterogeneity. In the following, some examples are outlined in a non-exhaustive manner, referring to the already established layers of heterogeneity: firms, industries and space.

Firms are often composed of multiple plants, which differ across several important dimensions, like products, technologies, employees or location. Moving even another level

deeper, the role of individuals, especially the firms' managers and their attitudes as well as decisions, increasingly gain importance in the literature on firm growth (e.g. MARRIS 1964). A managerial perspective might also explain the often observed unequal willingness to grow, or put differently, why some firms start small and remain small and why they lack an interest in growth (MCKELVIE/WIKLUND 2010). PYKETT (2013) discusses in a recent contribution the role that neuroeconomics and behavioural economics might play for economic geographical analyses.

Interestingly, by analysing a balanced panel of firms LUNARDI et al. (2014) observed that individual firms might be also heterogeneous over time with respect to the mean value, standard deviation and even the distributional shape of their growth rates. Only at the disaggregated industry level, they could not reject a common functional form. This finding again underlines the importance of an industry-specific perspective in the analysis of growth rates distributions. But the use of industrial classification systems is only one way to analyse separately their heterogeneous growth performances. However, the logic of such taxonomies strongly relies on similarities in the production processes or the products. Hence, they tend to ignore other important aspects, like the different capabilities or even technologies found among firms which are categorized into the same industrial class.

Although firms are analysed in this thesis separately along different size classes and industries, the results still provide an average picture, for instance on the relevant spatial extent of growth-related externalities. One single firm can be expected to interact and cooperate with partners at several geographical scales at the same time. It is still an unresolved issue if all distances matter equally, such that the average can be taken, and from how many potential sources a firm can benefit due to internal constraints in maintaining and coordinating an extensive network (the social network theory suggest a power-law scaling relationship, e.g. BARABÁSI/ALBERT 1999). At least, this thesis determines a flexible distance-decay function separately for various kinds of knowledge sources.

Which of these possible sources of heterogeneity truly matter is still an open question and deserves more attention in future research. In this vein, the next section continues by arguing that also the conception of space itself as geographical distances might be incomplete.

9.2.4 The one-dimensional conception of space

Travel times represent a more realistic distance metric with respect to economic interactions than orthodromic distances (see section 1.6.3). However, they still provide pure physical distance-based weights and thus fail to capture non-geographical distances, like cognitive, organizational, social or institutional forms of proximity (BOSCHMA 2005). Although physical distances remain the frame for all other types (RODRIGUEZ-POSE 2011), geographical proximity is neither a necessary nor a sufficient condition for knowledge spillovers to occur (PONDS et al. 2010). In the empirical literature, recently much effort is dedicated to determine the relevance of the other dimensions of proximity, like institutional (PONDS et al. 2007), social (AGRAWAL et al. 2008), cognitive (BOSCHMA/TER WAL 2007) or

technological proximity (AUTANT-BERNARD 2001). Some even question the concept of distances altogether and highlight the importance of being located in a place with specific qualities: “geographic proximity is often less important as a force driving cluster firms' innovativeness than the characteristics of the place and local firms” (LORENZEN et al. 2012).

This rather conceptual issue regarding space is accompanied by another empirical issue in determining the role of agglomerations and industrial clusters for economic growth, namely the problem of endogeneity between firm growth and distances.

9.2.5 Endogeneity of firm growth and distances

In the literature, the exogeneity assumption of the performance of firms and the geographic distance towards external growth factors is sometimes questioned: better performing firms might be less restricted in their location choice and thus be able to move closer to universities or locations within clusters (LYCHAGIN et al. 2010). If location choices are not exogenous, estimates of the effects of external factors on firm growth would be biased (MAINE et al. 2010). Using travel time distances, the endogeneity issue might be even more severe, as roads are built between centres of increasing activities and prospering firms might be favoured by infrastructure projects, hence moving virtually closer (GRAHAM et al. 2010). Unfortunately, distances are difficult to instrument (PINSKE/SLADE 2010) and controlled experiments merely possible (MAINE et al. 2010). However, the endogeneity problem has to be relativized, as this thesis looks at a rather short time period of a few years for which the spatial configuration of economic activities and infrastructure can be regarded as rather stable. Besides, it focuses on growth rates instead of levels, which contain a much higher stochastic part, and it performs the analysis at the firm level, which tends to be less biased than the regional level (BALDWIN/OKUBO 2006).

Hence, regression approaches seem to be sufficient at the level of firms. At the regional level, the endogeneity issue has to be explicitly addressed, for instance by employing some sophisticated identification procedure. This thesis relies on both of these approaches, which, not surprisingly, entail their own limitations, as the two subsequent sections will point out.

9.2.6 Neglecting the panel structure in the quantile regression setup

Quantile regression techniques are used in chapter 4 and 5 owing to the stochastic characteristics of firms’ growth rates. This method is applied in a conventional cross-sectional setup neglecting the panel information in the data. For instance, applying fixed effects would not only allow to account for the many factors that are commonly unobserved, say, managerial capabilities, but also to account for the issue of endogeneity, even if only incompletely.

Although quantile regression approaches for panel data have recently become an important econometric research field (see BACHE et al. 2013, GALVAO 2011, CANAY 2011),

they are not employed in this thesis mainly for two reasons. On the one hand, an established estimation strategy is still missing (in the context of firm growth, panel quantile regression has not yet been applied). On the other hand, and more importantly, only an unbalanced panel of firm data is available. Exploiting time-series information would therefore imply a loss of cross-sectional information. This trade-off was decided in favour of keeping those firms for which only few data points in the panel are available. On contrary, panel data is available at the purely regional level and methods exploiting this information become suitable, like the SVAR approach.

9.2.7 Some limitations of the SVAR approach

The SVAR approach, once an appropriate identification strategy has been found, provides many valuable insights into the underlying data generation process. Up to now, only few words have been dedicated in this thesis to the constraints and shortcomings of this approach. These can be categorized as being specific to the chosen identification strategy or as being inherent to the general model setup.

The chosen data-driven identification strategy, named LiNGAM, relies on three conditions: non-normality and independence of the shocks as well as acyclicity of the causal structure. Whereas non-normality is safe to assume and can be verified empirically, independence and acyclicity are much stronger assumptions. Cyclical relationships might be a frequent phenomenon, especially in regional growth, which is often characterized by dynamic feedback loops. Therefore, the cyclical version of LiNGAM (LACERDA et al. 2008) should be considered as a relevant extension. In chapter 3 it is argued that independence is less trivial to test. However, the identification is also robust to some degree of deviation from this assumption. Besides, the number of independent shocks is assumed to equal the number of variables. This assumption might perhaps be even more serious. Alternative models, which allow for some degree of freedom on the number of relevant independent shocks, have to be explored first.

The remaining shortcomings concern the general setup of the model and are extensively discussed in the econometric literature, like the omitted variable bias, the assumption of constant parameters over time or linearity in the relationships (STOCK/WATSON 2001). First extensions of the basic LiNGAM algorithms are developed to take into account these issues (e.g. HOYER et al. 2009 for non-linear models). Besides, inter-regional dependencies, for which various spatial extensions of the vector autoregression models exist (e.g. BEENSTOCK/FELSENSTEIN 2007 or GIACINTO 2010), yet cannot be explicitly addressed using VAR-LiNGAM. However, BUERGER et al. (2012) show that the spatial auto-correlation of the technological activities is rather negligible for functionally defined labour market regions.

9.2.8 The unconditional analysis of distributions

“The fact that heavy-tailed distributions occur in complex systems is certainly important (because it implies that extreme events occur more frequently than would otherwise be the

case) […] as it is certainly useful to estimate how often extreme events occur in a given system" (STUMPF/PORTER 2012: 666). This thesis has shown that the analysis of growth rate distributions provides a more realistic entry point to the investigation of dynamic economic phenomena. Knowledge about these distributions can be even exploited to learn more about the causal structure in growth processes. The same authors of the incipient citation, however, question further insights that can be drawn from identifying distributional characteristics. As GABAIX (2009) notes, from a simple unconditional analysis no direct information is revealed on the underlying causal mechanisms, because the same distribution can arise from a plethora of different mechanisms (ALFARANO/MILAKOVIC 2008: 274) or simply emerges as an artefact of aggregation of lower level (fat tailed) distributions (SCHWARZKOPF et al. 2010). Hence, merely finding a robust distribution is no evidence of universality without a concrete supporting theory (STUMPF/PORTER 2012).

In chapter 6 and 7, this issue is addressed by systematically comparing the estimated parameters of the firm growth rate distributions of various regions as well as by estimating the distribution conditional on further variables. Especially the latter approach is promising, although it suffers from high computational costs, as outlined in the following section.

9.2.9 Computational issues

Heterogeneity not only involves higher conceptual complexity, but also – if empirically implemented – higher computational costs. This concerns both the data as well as the estimation. Calculating driving distances between the locations of all actors, even though the locations are approximated by the municipality, was only possible in a reasonable amount of time by using pre-processed data and specific speedup techniques in the route planning algorithms. Finding the global optimum of the likelihood function of the AEP distribution becomes a less trivial task if the distributional parameters are estimated conditional on further variables. The risk of getting stuck in local optima can only be counterbalanced by a computational time that increases exponentially with the inclusion of each new variable.

However, as terms like cloud computing, parallelization or big data increasingly appear in academic vocabulary (e.g. VARIAN 2014), micro-data and estimation methods that go beyond the average value will become more attractive for future research. This thesis provides first examples into this direction in the context of economic growth at the level of firms and regions.

9.3 Implications for policy and research

Keeping the limitations of this thesis in mind, implications can be drawn from the findings, as summarized in section 9.1, for both policy and research. Whereas the policy implications are strongly aligned to the specific research questions and corresponding empirical findings of each single chapter, rather general conclusions are drawn for research focusing on the issue of heterogeneity and the presence of fat tails. A more specific research agenda, advised by these rather broad and general implications, is finally developed in section 9.4.

9.3.1 Policy implications

The importance of geography for economic policy was prominently acknowledged in a recent World Development Report, which was exclusively dedicated to economic geography (WORLD BANK 2009). In the context of growth, GARRETSEN et al. (2013) reassess the increasing awareness of the role that regions play in aggregate national growth. They observe that economic growth occurs unevenly across various geographical scales, leading to divergence of income among regions, notwithstanding the (slight) convergence at the national level. Besides, also more and more policies are decided and implemented at the regional level. For instance, new capital investments from the EU which aim to promote growth are mainly distributed and administered at the level of regional government (RODRÍGUEZ-POSE/GARCILAZO 2013). Despite the increasing importance of spatial matters, economic geographers traditionally face difficulties in influencing and shaping public policy (MARTIN 2001, RODRÍGUEZ-POSE 2010). One important problem in the formulation of policy recommendations with respect to growth strategies can be attributed to the inappropriateness of traditional (regional) growth regressions. The reasons are mainly twofold. First, ordinary regressions do not tell anything about regional specificities (LEVINE/ZERVOS 1993) and secondly, they ignore the issue of causality. Only causal knowledge makes prediction on consequences of actions possible (PEARL 2000), the *conditio sine qua non* for designing and implementing effective policies (COAD et al. 2012). Besides, most empirical work so far has been performed at a qualitative, case study level (with a limited degree to which the results can be generalised) or at the aggregated level, both in terms of industries and regions. Different growth dynamics across industries or (more or less strongly) agglomerated firms within regions call for disaggregated studies, which allow for non-generic policy conclusions that are customized to the composition of firms and industries within specific regions (RASPE/VANOORT 2008, KNOBEN et al. 2011).

Drawing on the theoretical underpinnings of this thesis, the following general recommendations for regional growth policies are formulated. One of the main normative implications which arises from an evolutionary perspective is that the aims of economic policy do not have to be sought in the achievement of an optimal state of whatever shape (CANTNER/HANUSCH 2002). Instead, taking into account the ubiquitous heterogeneity, policy instruments should be adaptive and region-specific, aiming to create some regional advantages based upon existing strengths and resources (ASHEIM/BOSCHMA 2009). Put differently, “policy builds on region-specific assets that provides opportunities but also sets

limits to what can be achieved by policy. Public intervention should neither apply 'one-size-fits-all" approaches nor adopt 'picking-the-winner' strategies, but should aim to connect complementary sectors and exploit related variety as a source of regional diversification" (BOSCHMA 2009: 1). However, policy instruments should also deal with the fact that regional growth – at all levels of aggregation – is characterized by the frequent occurrence of extreme events. How they affect and disturb a regional economy will strongly depend on its resilience and adaptability. Policy is advised to transform their perturbing character into an evolutionary force that brings change and new growth opportunities in the long-run.

The issue of heterogeneity becomes especially relevant for regional policies aiming at agglomerations. If the aim of cluster policies is to internalize agglomeration effects in existing clusters (KETELS 2013), then the "rationale for cluster policy appears strongest, when it is focused on those industries that enjoy such cluster effects" (BEAUDRY/SWANN 2009: 422). Therefore, cluster policies should focus on specific industries, because not all industries show positive agglomeration economies, as has been shown in this thesis. Moreover and obvious to argue, firms within clusters try to optimize their performance. But this behaviour not necessarily results in higher aggregate regional growth, because agglomeration economies affect only indirectly, via aggregate firm performance, regional growth (VANOORT 2014). In some cases agglomeration economies, for example in the form of higher wages, are beneficial for the region, but detrimental to the firms (KETELS 2013). Besides to their industry, firms differ according to many other dimensions and show quite specific requirements on their spatial surroundings. Hence, policies should target certain groups of firms, like high-growth firms (OECD 2010), and respond to their specific local needs (VANOORT 2014).

Considering these general perspectives on regional policies, the thesis draws more concrete conclusions based on the empirical findings. Above all, the thesis shows that extreme events are more frequent and have a larger magnitude than policy makers who are informed by standard economic models, i.e. models that are based on the normal distribution, tend to expect (MCKELVEY/ANDRIANI 2005). Instead of exclusively focusing on the average performance, policy makers should also look at the extreme events, which are not only more consequential in the short run, but also bear opportunities for structural change and evolutionary developments in the long run (CAPASSO et al. 2013). The fat tails are related to two further important policy issues, namely high-growth firms and regional resilience.

The promotion of high-growth firms is increasingly acknowledged as an important policy object (OECD 2010, EUROPEAN COMMISSION 2010), putting it in the recent agenda even on top of the support for new start-ups and entrepreneurs (NIGHTINGALE/COAD 2014). However, the stochastic nature of firm growth makes it difficult to predict which firms are going to be high-growth firms and hence might be worth to be targeted by policy (COAD et al. 2014, HÖLZL 2009). Instead, COAD et al. (2014) suggest that "the policy focus should be on the identification of barriers to firm growth dynamics". The barriers might be both internal to the firm, like their resource composition or financial constraints (BOTTAZZI et al. 2014), or external, that is resulting from the general industry dynamics or the regional environment (BASTESEN/VATNE 2014). As this thesis shows, the likelihood of extreme events differs among regions and industries. Concerning the factors that are amenable to policy makers it is found that fat tails are particularly pronounced in the presence of highly

educated workers, knowledge intensive activities and a more diverse and variegated industrial structure (see chapter 3 and 6). The role of agglomerations for high-growth firms has to be discussed more carefully, as the results strongly depend on the degree of relatedness, the internal characteristics of the firms and the type of industry. For instance, chapter 4 argues that high growth events are more likely for smaller firms, but only if the nearby activities are technologically closer related. Complementarily, chapter 5 shows that agglomeration matters for high growth only in industries which tend to be more knowledge intensive. These findings might be a helpful ingredient for the design of promotion strategies for high-growth firms by removing barriers in particular regions.

Moreover, high growth events are of great interest for the long-term regional development. Here, also negative extreme events play a crucial role, as the "creative destruction implies the destruction of some activities as a necessary element of the growth of others" (METCALFE 2001: 556). BRAVO-BIOSCA et al. (2013), for instance, show that more competitive and dynamic environments are accompanied by higher shares of both fast growing and shrinking firms. From an evolutionary perspective, it is argued that regional resilience can be understood as regional systems that are able to accommodate these extreme events, and translate their transformative character into structural change and a higher long-term performance. Regions that are more turbulent at the micro-level of firms tend to be more adaptive. Hence, policy makers might aim to support the capacity of firms to cope with and to adjust to change (SCHUBERT 2013). Chapter 6 observes these firm level turbulences mainly in regions with a highly skilled workforce, a more diverse industrial structure and a competitive regional environment. From this it can be concluded, once again, that institutions matter. However, it is yet an open issue which types of institutions can better accommodate these turbulences and translate them into fruitful changes and new long-term growth paths (BOSCHMA 2014). It should also be noted that this kind of resilience – with an indisputable neoliberal notion – might have negative consequences for some workers, who face difficulties in becoming reemployed by reason of their individual skills within the implied environment of more flexible working conditions (MARTIN 2012). Moreover, it is yet to be debated how much change is desirable from a societal point of view (REGGIANI et al. 2002). For this thesis it is out of reach to answer these questions, for which also the short-term and long-term benefits and social consequences have to be taken fully into account.

The study on causality in chapter 3 argues that the nature of the technological activities and the causal links among them differ across industries, underlining the importance of industry-specific regional policies (TÖDTLING/TRIPPL 2005, ASHEIM et al. 2011). More precisely, it is found that university graduates foster innovation especially in science-based industries. Hence, policy should closer align the university education to the local innovation activities. Besides, it is important to distinguish between the industries' knowledge bases. Stimuli like monetary incentives for basic research seem to be more promising in industries based on analytical knowledge, whilst a focus on the transfer of innovation might be suggested for industries based on synthetic knowledge. Finally, chapter 3 concludes that certain industries, say, electronics, are more suited to pick to become large export-oriented champions. By promising some channels on which policy shocks propagate throughout the entire system of regional technological activities, these translate more easily into longer-term success.

The findings from the firm-level approach in chapter 4 and 5 agree with the emerging notion in the literature that specialisation or diversification strategies of fostering growth lose value (for a recent overview on the discussion see also VANOORT 2014). Because of the various kinds of heterogeneity, including the different degrees of relatedness between firms and other activities, policies drawing rather on the concept of smart specialization become interesting. This concept is more about the generation of new specialization (FRENKEN 2014) and it builds on a profound analysis of the specific conditions, technologies and competencies of a regional economy (MCCANN/ORTEGA-ARGILÉS 2013). For example, it is shown that in most industries, policy measures that foster further clustering of economic activities are confronted with prevailing diseconomies of agglomeration. In such cases, scarce public resources might be more effectively invested in activities of universities or public research institutes, which are acknowledged in this thesis as well as in the literature to be an important driver of regional economic development (e.g. RODRIGUEZ-POSÈ/CRESCENZI 2008, SCHLUMP/BRENNER 2010). However, to fully realize their potential economic benefit, it is argued here that research and educational activities should match the (industrial) composition of the regional economic activities.

This thesis has some important limitations when it comes to policy conclusions. The design and implementation of policy measures require a sound understanding of the mechanisms at work. This thesis is exclusively about patterns, processes and causes of economic growth. With the exception of the simulation model in chapter 8, it is not studied what the mechanisms behind the fat tails, agglomeration externalities and knowledge spillovers are, or behind the causal links within a regional system of technological activities. Hence, future research, guided by the observed patterns, processes and causes, might focus more on the identification of the underlying mechanisms. Further general implications for research are outlined in the subsequent section.

9.3.2 Research implications

The implications for research are mainly twofold. First, this thesis concludes that research on the patterns, processes and causes of economic growth could strongly benefit from taking into account the distribution of growth rates seriously. Secondly, it concludes that heterogeneity must be a guiding principle for all research on economic growth.

Meanwhile, there exist well-established and robust *stylized facts* on growth rate distributions at various levels of aggregation. As put by BOTTAZZI et al. (2007: 150), the "underlying drivers of (firm) growth ought to involve relatively frequent and big events, unaccountable by a Gaussian one". This deviation of growth rates from a normal distribution, but also their variance-scaling which can be described as a power law, are interesting findings per se and hence entail a huge body of literature. However, more can be learnt about the generating mechanisms by explaining these stylized facts as emergent properties using *modelling* techniques. Furthermore, by fully acknowledging these empirical regularities in making assumptions about the underlying economic data generation mechanism, the quality and scope of *estimation* techniques can be improved. Figure 34 classifies the literature on growth rates into these three domains.

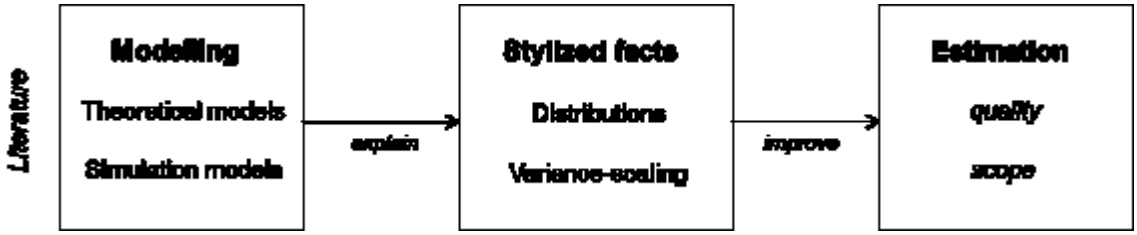


Figure 34: Domains of research on economic growth rates

Models are helpful to explain the emergence of properties in complex systems with interacting units. In the literature, there exist both theoretical models and simulation models. In the former mostly one stylized fact on firm dynamics is explained by assuming a specific mechanism. Well-known examples are the hierarchical sub-unit models based on AMARAL et al. (2001) or the increasing return model in catching limited market opportunities by BOTTAZZI and SECCHI (2006a), which yield a Laplace distribution of growth rates and a power-law in the variance-scaling relationship. Other models focus on different empirical observations, like the temporal auto-correlation structure (COAD 2012), and many of them are only loosely grounded on economic mechanisms. For instance, in SCHWARZKOPF et al. (2010) the Laplace distributed growth rates directly follow from the assumption of a power-law replication function. Consequently, more complex Agent-Based models have been developed which try to include knowledge on economic processes and mechanisms in their assumptions (e.g. METZIG/GORDON 2014, DELLI GATTI et al. 2005, AXTELL 2013). While these models are able to reproduce a larger number of stylized facts, they are quite specific, focusing on specific aspects. In chapter 8 of this thesis, a simulation model is devised on basis of the simulation approach by BRENNER and WERKER (2007). This model takes into account general knowledge on firm dynamics that draws on basic market, competition and innovation processes. For the empirical validation of this model, all robust stylized facts on firm growth can be used.

Important lessons from stylized facts can also be learnt for the empirical estimation. Because the assumption of a normal distribution is not met, MINEO (2003) suggests to use either more robust methods, like LAD instead of OLS, or to hypothesize a different, more general distributional model to adequately describe the error term. He cites FISHER (1922), who writes against the idea of excluding outliers: “as a statistical measure, however, the rejection of observations is too crude to be defended and unless there are other reasons for rejection than mere divergence from the majority, it would be more philosophical to accept these extreme values, not as gross errors, but as indications that the distribution of errors is not normal”. Furthermore, the focus of the estimation itself might shift away from the mean value towards other parts of the distribution. Compared to a mere improvement in the quality of estimation, this broadens also the scope of possible insights on many economic relationships. Eligible methods are the quantile regression or the conditional estimation of parameters of a more general (and thus more realistic) distributional model. Both options have been explored extensively in this thesis. Finally, the empirical estimation

can be enriched by exploiting the additional information which resides in the deviation from non-normality, for example to identify the underlying causal structure in the data.

Besides from their empirical relevance, extreme events bear also some theoretical meaning. The concept of hysteresis, for instance, assumes that extreme events have a long-term impact on a system, as they trigger threshold effects: "extreme experiences that propel systems sufficiently far from its current state are thought to result in structural change" (SETTERFIELD 2010: 8). This is in line with evolutionary thinking that attributes extreme growth events a crucial role for evolutionary processes (RICE 2004, METCALFE/FOSTER 2010, ANDERSEN/HOLM 2014). Some of these aspects are already explored in chapter 6 of this thesis. However, it is still too little known about the sources, characteristics and implications of extreme events, which are not mere noise, but represent "complex, specific development processes", which "should be taken seriously and even theorized" (STORPER 2011: 342). Performing case studies could be a first step towards a better understanding of the extremes that reside in the tails.

Secondly, this thesis concludes that research on the patterns, processes and causes of economic growth should take into account the issue of heterogeneity seriously. It shows, for instance, that in the context of spatial matters of growth and agglomeration, the Modifiable Areal Unit Problem (MAUP) becomes a relevant issue. The heterogeneity in the spatial economic landscape might have significant effects on the empirical results and can be only taken into account by the use of micro-data and a sound specification of the distance decay function. Besides, growth theories are advised to address the spatial dimension of growth externalities more explicitly. Although they cannot explain the exact distances, which are very sensitive to the industry or the national context, at least they should provide some statements on the relevant spatial scale – whether narrow urban areas, larger labour market regions or even the level of states are expected to matter. To conclude, cross-sectional studies should use a flexible distance decay function instead of pre-defining the regional boundaries to account for both the effects that go beyond political jurisdictions and the highly localized, within-region effects. Moreover, the thesis shows that concepts like regional resilience are ideally addressed by taking into account the full heterogeneity at the level of firms, which reveal dynamics that are otherwise hidden by averaging or summing up, that is, aggregation. This confirms the literature (e.g. MARTIN 2012) which is increasingly demanding a more rigorous statistical analysis of regional resilience, an important feature of regional economic development, at the level of firms.

9.4 Outlook on future research

Based on the conclusions for research (see section 9.3.2), especially the importance of taking into account the entire distribution of growth rates and the spatial heterogeneity of the economic landscape, and referring to the previously outlined limitations (see section 9.2), this final section elaborates a more detailed agenda for future research. First, it shows three possible directions for extending the emerging literature on the analysis of growth rates. Secondly, it sketches possible extensions by breaking down even further the already acknowledged heterogeneity – the heterogeneity at the layers of firms, industries and space.

9.4.1 Distributional analysis of growth rates: three possible extensions

In the previous section, the literature on the distributional analysis of growth rates was classified according to three basic domains: 1) the identification and confirmation of *stylized facts* regarding the distributional characteristics of the growing entities, 2) the explanation of these facts as emergent properties by theoretical or simulation *models*, and 3) the improvement of the quality and scope of *estimation* techniques by acknowledging these facts in the assumptions on the underlying economic data generation mechanism. According to the proposed taxonomy, this thesis contributes to all three domains in various ways (see Figure 35). It extends the evidence on the stylized facts to the intermediate level of regions and to an emerging East Asian economy. It explores and develops estimation strategies concerned with the non-normal distribution of the growth rates. And it devices a simulation model explaining the emergence of the empirically observed distributions. Yet, all of the three domains still provide many possible directions for future research. In the following, the most promising open research questions are worked out by combining the insights from this thesis with current discussions in the literature.

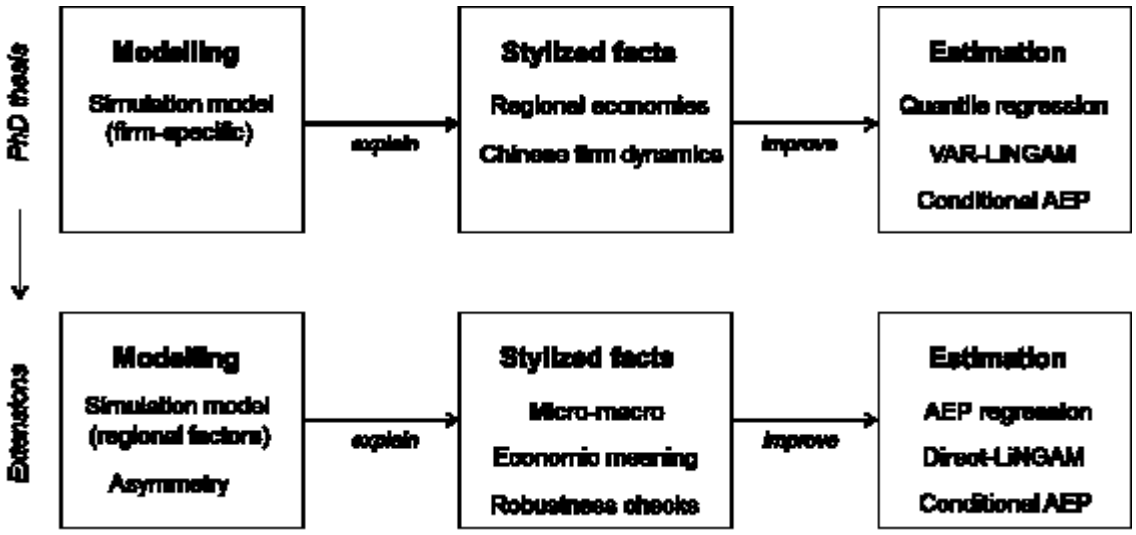


Figure 35: Possible extensions of the PhD thesis

9.4.1.1 Stylized facts

This thesis provides new evidence that the stylized facts on growth rate distributions survive also at the regional level, thus supporting the argument that distributional characteristics like non-normality or variance-scaling occur in all complex economic organisations independent of the level of aggregation (AMARAL et al. 2001). Furthermore, it shows that firm growth rate distributions remain asymmetric and fat tailed using regions instead of countries as a reference system, although the actual shape of the distribution markedly differs across regions. From the established stylized facts on economic growth processes, which mostly concern one level of aggregation at a time, an extension of the analysis to the micro-macro nexus seems to be the next logical step. Besides, the robustness of the findings at the regional level might be further scrutinized.

The micro-macro nexus and the economic meaning of fat tails

Naturally, the micro-macro nexus can be analysed into two directions: how do micro-level properties impact on the macro-level performance of a system, and how do aggregate macro-level shocks effect the micro-level growth rate distributions?

Investigating the former is tantamount of exploring the economic meaning of the distributional characteristics. In chapter 6, it is hypothesized that turbulences at the level of firms are an indicator of structural change and a competitive regional environment. Arguing with ARTHUR (1999) or BEINHOCKER (2007), extreme growth events have a potentially pervasive, path-breaking and long-lasting impact on the regional economy and industrial structure by crossing thresholds of robustness and triggering cascades of further changes in related industries (see also MCGLADE et al. 2006). Drawing on a metaphor from TALEB (2012), the damage to a car that crashes ten times against a wall with the speed of 10 km/h compared to a car that hits the wall only once but with 100 km/h will not be the same. In this vein, chapter 6 finds that fat tails of firm level growth rates are related to the presence of a highly qualified regional workforce, unrelated variety in the industrial structure and a positive aggregate economic performance. However, the direction of causality remains unclear and resilience-thinking is mostly concerned with a long-term adaptability (DAVIES 2011). To underpin the hypothesis, it is necessary to study the relationship of the distributional parameters also with the long-term performance and structural change of regional economies. Related work on the link from the micro to the macro level can be found also in GABAIX (2011), who decomposes aggregate fluctuations into their granular origins. By looking at productivity growth rates, DOSI et al. (2012) interpret the left tails as an indicator for degree of the sectoral tolerance of inefficient firms, that is, the (lack of) selection forces in markets. Finally, HOLLY et al. (2013: 477) stress that there is a real effect of dynamic distributional changes at the micro-level on aggregate growth: “asymmetric changes in the probability mass on either side of the mode may act as important amplifiers or attenuators of aggregate activity”.

Equally interesting is to investigate the impact of the macro-level performance on the micro-level growth rate distributions. Here, HIGSON et al. (2004) found that cyclical national shocks show a stronger impact on the central mass of the distribution and weaker effects on the tails. Put differently, firms at the extremes respond less to aggregate shocks. HOLLY

et al. (2013) extend this finding by showing that firms in the left side of the distribution are more responsive to aggregate business cycles than those in the right side. However, this relationship might look differently during non-cyclical economic crises and might also be mediated by some regional factors. For instance, ARCHIBUGI et al. (2013) found that the concentration of innovative activities within a small group of the best performing firms increased as a result of the economic crisis of 2008. Hence, to study the complex co-movements of the AEP parameters with the aggregate regional performance, cyclical fluctuations have to be differentiated from economic recessions. In future research, the micro-level firm growth rate distribution might be compared before, during and after historically recorded major recessions. Additional insights can be gained from analysing the factors that mediate these relationships at the regional level.

Robustness checks

In this thesis, the regional level is explicitly taken into account. While at the firm level a wide range of evidence exists for various national contexts (e.g. BOTTAZZI et al. 2002 for Italy; BOTTAZZI et al. 2011 for France; DUSCHL et al. 2014a for Germany; REICHSTEIN/JENSEN 2005 for Denmark; STANLEY et al. 1996 for the United States, MATHEW 2012 for India; DUSCHL/PENG 2014 for China) and growth measures (employees, turnover, productivity, profit, total assets, market value; see COAD 2014 or MUNDT et al. 2014), this thesis is the first to study regional growth processes. Hence, evidence exists only for regional employment dynamics in Germany.

It is well known that Germany has shown a specific response to the recent financial and economic crises (e.g. DAVIES 2011). Therefore, it would be interesting to systematically compare growth patterns both at the level of firms and regions across different countries within Europe or beyond. For instance, new insights can be gained by framing the analysis into the varieties of capitalism approach (HALL/SOSKICE 2001) or studying comparable historical events, like the Asian crisis 1997 or the dotcom bubble 2001. Employees as a measure of growth have been chosen based on theoretical considerations. However, growth is a multi-dimensional phenomenon and different forms of growth should be measured with different growth measures (COAD 2010, DELMAR et al. 2003). Systematic differences are expected also at the regional level: growth rates measured by employees tend to be more fat-tailed than alternative measures. Plants are indivisible due to technological constrains, resulting in lumpy employment changes reinforced through the entry of new plants and the exit of existing plants (BOTTAZZI et al. 2007). These differences in the growth measures themselves provide interesting insights into the underlying growth processes.

Finally, a deeper understanding of the distribution might be gained by carrying out some case studies. Studying specific regions at various quantiles of the growth rate distribution in more details would allow to evaluate the often context-specific reasons for their exceptionally well or poor performance.

9.4.1.2 Estimation

Regression models with the error term following the AEP distribution

In view of the now well-established stylized facts, FAGIOLO et al. (2008) make a general plea for non-Gaussian econometrics, which means the development and use of estimators and tests based on fat-tailed errors. It is widely acknowledged that OLS methods perform poorly in case of non-normal distributed errors (DASGUPTA/MISHRA 2004, MAASOUMI et al. 2007). As a more robust alternative, LAD is often preferred in the literature (e.g. MONETA et al. 2013). If the errors follow a Laplace distribution, LAD provides the maximum likelihood estimators. However, in the empirical data the tails are usually even fatter than Laplace and often an asymmetric shape is observed. Hence, regression models might be employed which take into account this knowledge on the distributional characteristics. Such a regression model based on EP-distributed errors is implemented by MINEO/RUGGIERI (2005) using methods of moments. In the context of regional growth, SCHLUMP and BRENNER (2010) introduce a ML approach for regression models with the error term following the AEP distribution. A ML estimation method for such AEP-error regression models is also discussed in LI (2011) and further analysed in NARANJO et al. (2014).

Conditional estimation of the AEP distribution parameters

Ordinary regression models, even if extended by more realistic assumptions on the error term, are designed to capture the effects on the average value of the dependent variable. From a distributional perspective it becomes clear that "the variables may continue to impact other distributional characteristics" (MAASOUMI et al. 2007: 449). By studying the effect of other variables on the entire shape of the AEP distribution, the conditional estimation, as developed in chapter 7, combines the advantages from both traditional regression models and unconditional distributional analyses. This approach might provide also new perspectives on many questions that are related to the topics of this thesis. For instance: how does the industrial composition of regional economies affect their growth rate distributions? Is the size of a firm or regional economy related only to the variance, as the literature on variance-scaling claims, or also to the shape of the distribution? How did the financial and economic crisis impact on the various distributional parameters? Or, how does the distribution of growth rates change with the distance of a region from the technological frontier – a question related to a new KALDOR fact as pointed out by JONES and ROMER (2010), which states that variation in the growth rates increases with the distance from the technological frontier.

Finally, by looking at the conditional effects on the tails of the distribution, high-growth firms or exceptionally well performing can be analysed without the requirement of delimiting a sub-population of these firms or regions, for instance, by choosing a growth rate threshold or by focusing on specific quantiles within the quantile regression framework (on the difficulties of measuring high-growth firms see DELMAR/DAVIDSSON 1998, DELMAR et al. 2003, or more recently COAD et al. 2014). An alternative approach, relying on a theoretical understanding of growth rate distributions, is also suggested by HALVARSSON (2012), who develops an explicit parametric test for the presence of high-growth firms.

Extension of the VAR-LiNGAM algorithm for identifying the causal structure of SVAR models

Only recently, with LiNGAM a data-driven approach has become available to identify the causal structure of vector autoregression models (HYVÄRINEN et al. 2010). In the field of machine learning, further extensions and improvements have been developed. On the one hand, LACERDA et al. (2008) propose a cyclical version that allows for feedback loops within the time period. Cyclicity as a condition might be meaningful to assume especially for annual data, as the causal direction of the regional growth effects probably show into both directions at this time scale. On the other hand, SHIMIZU et al. (2011) provide a new direct method to estimate the causal structure. Compared to the established method, which is based on iterative search algorithms, it guarantees convergence. This direct method, originally developed for structural equation models (SEM), might be also applied in the context of SVAR models.

Above that, as it is also prominently claimed by the chief economist of Google Inc. (VARIAN 2014), a stronger collaboration between computer scientists and econometricians might be fruitful. With LiNGAM, this thesis applies a method from machine learning to address the issue of causal inference, which usually lies in the domain of econometricians. This method might be further elaborated by a stronger engagement of the latter, for instance in the calculation of the standard errors (for example, by more advanced bootstrapping methods), which usually is not a major concern of the former.

9.4.1.3 Modelling

Already Herbert SIMON (1968) noted that there are many possible explanations for empirically observed regularities. This holds especially true for unconditional objects such as growth rate distributions (BROCK 1999). Therefore, models using assumptions based on theory are a way forward (COAD 2012). As MCKELVEY and ANDRIANI (2005) conclude, the modelling of fat-tailed phenomena has still received insufficient attention. Two research extensions of modelling growth rates are particularly promising.

Explaining the emergence of asymmetric growth rates distributions

Several models have been developed to explain the emergence of the distributional characteristic of growth rates. For instance, BOTTAZZI and SECCHI (2006a) take upon the classical island model by IJIRI and SIMON (1997) and assume a competition for limited resources. Combined with self-reinforcing dynamics in catching these resources, fat tailed distributions emerge – while few units seize the major share of growth opportunities, many units only seize a marginal share. COAD (2012) models fat tails by assuming that growth events are inherently lumpy due to the discreteness of the resources, say, managers, that firms need to employ to grow. His model is motivated by the empirical observation of a negative temporal autocorrelation of growth rates, and hence contradicting the increasing-

return models from a temporal perspective. Finally, AMARAL et al. (1997) or FU et al. (2005) introduce heterogeneity at the level of firms, which have different numbers of sub-units that interact and correlate in their dynamics. By assuming a hierarchy in the organizational structure of firms, the variance-scaling relationship and fat-tailed growth rate distributions can be closely reproduced.

Although the mentioned models and many others are all able to generate fat tailed distributions, they neglect the issue of asymmetry, another stylized fact that has recently been established in the empirical literature. In the existing models, decline is treated as growth or expansion with reversed sign. The observed asymmetry, however, implies that specific mechanisms might explain why processes of decline differ from processes of expansion. Models that explicitly take into account these differences are required. In this context, BREINLICH et al. (2014) conclude in a recent survey on economic growth that decline, which often implies large social and economic costs, as the case of Detroit well illustrates, is an under-researched topic. This argument holds true for the highly declining firms that populate the left tail of the distribution, which have received disproportionately less attention compared to the best performing firms in the right tail (COAD et al. 2014). As argued in chapter 6, the entire distribution is required to understand the development of an economic system – and it is often observed that higher aggregate performance is associated with more dynamic growth rate distributions, meaning a larger fraction of both highly growing and declining firms (COAD et al. 2014, HÖLZL 2011).

Explaining the emergence of regional growth rates distributions

Fat tails are common to all complex system dynamics (DOSI et al. 2010). But what are the specific mechanisms behind their emergence at the level of regional economies? In general, there are a "plethora of different mechanisms that lead to fat tailed distributions" (ALFARANO/MILAKOVIC 2008: 274), and the nature of the mechanisms might differ across levels of aggregation (CASTALDI/DOSI 2006). In the literature, there are two competing general explanations for the emergence of fat tails. On the one hand, they emerge from some underlying correlating mechanisms of a self-organized system (Dosi et al. 2010). This line of thinking is assumed by the models discussed in the preceding section. These models, like the increasing-return model of BOTTAZZI and SECCHI (2006a), can be adapted to the particular features of the regional level, just as CASTALDI and SAPIO (2008) adapted the model to the sectoral level. For instance, more successful regions might seize the growth opportunities of other regions. This phenomenon is discussed in the literature as an inter-regional competition for firm locations or as local spillover effects in mainly RD-based innovation activities leading to agglomeration tendencies. On the other hand, fat tails might result from the aggregation of growth rates of the entities at the lower level. Instead of compensating out in the aggregation process, firm level shocks may aggregate in a non-trivial way and produce relevant regional shocks (CASTALDI/SAPIO 2008). This is in line with GABAIX's (2011) model on the granular origins of aggregate fluctuations, in which idiosyncratic firm level shocks explain an important part of the aggregate fluctuations. Besides, the lumpy nature of the growth processes of the elementary units of a regional economy, manifested by the exit and entry of firms (HALTIWANGER 1997) and

the nonlinear adjustments due to technological constrains (COAD 2010), critically contributes to the propagation of the growth shocks to aggregate levels.

However, little is known on the specific contribution of these mechanisms in explaining the emergence of fat tailed growth rates. One way forward might be the design of empirically validated simulation models (WERKER/BRENNER 2007), which should aim to replicate a large number of stylized facts. Especially agent-based simulation models allow for complex interactions of locally interacting heterogeneous agents. In chapter 8, such an empirically validated simulation model is developed to study firm growth processes. This model can be extended in future research to account also for regional growth processes, as regional growth is simply the sum of the elementary firm growth processes. By introducing additional parameters on market, competition and innovation processes of the firms accounting for different kinds of agglomeration effects, the role of regional factors for the emergence of fat tailed distributions could be investigated.

9.4.2 Breaking down heterogeneity

The aim of this research is to explicitly take into account the heterogeneity of firms, industries and space. Not surprisingly, these three layers of heterogeneity can be broken down even further.

Here, firms are mostly distinguished according to their size as measured by the number of employees. Although size is an important proxy for many structural characteristics, other aspects, like their knowledge intensity (KOO 2005) or age (COAD et al. 2013) might impact on their growth performance and their embeddedness in regional innovation systems. In this vein, a study of agglomeration effects might also compare the strength and spatial extent of the externalities across other categories like firm size.

Technological relatedness among firms, which mediate knowledge spillovers, is often more heterogeneous and diverse as can be captured by industry affiliation. Other dimensions of relatedness might be considered to assess the effects of externalities on firm growth, by focusing for example on input-output relationships, labour mobility, or the study field and the first job of the graduates. It would be interesting to show how such a refinement of the relatedness measure shapes the spatial patterns of agglomeration effects.

Finally, spatial heterogeneity could be more fully acknowledged by introducing GIS-data into the spatial economics literature (OVERMAN 2010). The densities of important economic variables, like population or employment, vary considerably within the regions (GORDON 2013). To address this issue, ANDERSSON et al. (2011) study the relationship between agglomeration and productivity by dividing Sweden into areas of one square kilometre. As geo-coded data on individual employees is not available for most of the countries, employment might be down-scaled using satellite images. Such an approach is suggested in GALLEGO (2010), who calculate a population density grid for the EU based on land cover satellite data.

References

A |

ACEMOGLU, D., CARVALHO, V., OZDAGLAR, A. & TAHBAZ-SALEHI, A. (2010): The Network Origins of Aggregate Fluctuations. In: *Econometrica* 80: 1977-2016.

ACS, Z., AUDRETSCH, D. & FELDMAN, M. (1994): R&D spillovers and recipient firm size. In: *The Review of Economics and Statistics* 76: 336-340.

ALFARANO, S. & MILAKOVIC, M. (2008): Does classical competition explain the statistical features of firm growth? In: *Economics Letters* 101: 272–274.

AGHION, P., AKCIGIT, U. & HOWITT, P. (2013): What do we learn from Schumpeterian growth theory? NBER Working Paper 18824.

AGNEW, J. (2013): Arguing with Regions. In: *Regional Studies 47*: 6-17.

AGRAWAL, A., KAPUR, D. & MCHALE, J. (2008): How do spatial and social proximity influence knowledge flows? Evidence from patent data. In: *Journal of Urban Economics* 64*:* 258-269.

ALECKE, B., MITZE, T. & UNTIEDT, G. (2010): Internal migration, regional labour market dynamics and implications for German East-West disparities: results from a Panel VAR. In: *Jahrbuch für Regionalwissenschaften* 30: 159-189.

ALFARANO, S. & MILAKOVIC, M. (2008): Does classical competition explain the statistical features of firm growth? In: *Economics Letters* 101: 272-274.

ALLEN, F., QIAN, J. & QIAN, M. (2005): Law, finance and economic growth in China. In: *Journal of Financial Economics* 77: 57-116.

ALMEIDA, P., DOKKO, G. & ROSENKOPF, L. (2003): Startup size and the mechanisms of external learning: increasing opportunity and decreasing ability? In: *Research Policy* 32: 301-315.

AMARAL, L., BULDYREV, S., HAVLIN, S., MAASS, P., SALINGER, M., STANLEY, E. & STANLEY, M. (1997): Scaling behavior in economics: The problem of quantifying company growth. In: *Physica A* 244: 1–24.

AMARAL, L., GOPIKRISHNAN, P., PLEROU, V. & STANLEY, E. (2001): A model for the growth dynamics of economic organizations. In: *Physica A* 299: 127–136.

ANDERSEN, E. & HOLM, J. (2014): The signs of change in economic evolution. An Analysis of directional, stabilizing and diversifying selection based on Price's equation. In: *Journal of Evolutionary Economics* 24: 291-316.

ANDERSSON, M. & GRASJÖ, U. (2009): Spatial dependence and the representation of space in empirical models. In: *The Annals of Regional Science* 43: 159–180.

ANDERSSON, M. & LÖÖF, H. (2011): Agglomeration and productivity: evidence from firm-level data. In: *The Annals of Regional Sciences* 46: 601–620.

ANDERSSON, M. & KARLSSON, C. (2007): Knowledge in Regional Economic Growth - The Role of Knowledge Accessibility. In: *Industry and Innovation* 2: 129–149.

ANDERSSON, M., KLAESSON, J. & LARSSON, J. (2011): Spatial Density and Productivity - an analysis on one-by-one kilometer squares. CESIS Electronic Working Paper Series.

ANGRIST, J. & PISCHKE, J. (2008): *Mostly harmless econometrics: An empiricist's companion*. Princeton University Press: Princeton, New Jersey.

ANSCOMBE, F. & GLYNN, W. (1983): Distribution of kurtosis statistic for normal statistics. In: *Biometrika* 70: 227-234.

ANSELIN, L., VARGA, A. & ACS, Z. (1997): Local Geographic Spillovers between University Research and High Technology Innovations. In: *Journal of Urban Economics* 42: 422-448.

ANSELIN, L., VARGA, A. & ACS, Z. (2000): Geographic and sectoral characteristics

of academic knowledge externalities. In: *Papers in Regional Science* 79: 435–443.

ARCHIBUGI, D., FILIPPETTI, A. & FRENZ, M. (2013): Economic crisis and innovation: Is destruction prevailing over accumulation? In: *Research Policy* 42: 303-314.

ARTHUR, W. (1994): *Increrasing returns and path dependence in the economy*. Ann Arbor: University of Michigan Press.

ARTHUR, W. (1999): Complexity and the Economy. In: *Science* 284: 107–109.

ARTIS M., MIGUELEZ, E. & MORENO, R. (2012): Agglomeration economies and regional intangible assets: an empirical investigation. In: *Journal of Economic Geography* 12: 1167–1189.

ASHEIM, B. (2007): Differentiated knowledge bases and varieties of regional innovation systems. In: *Innovation: The European Journal of Social Science Research* 20: 223-241.

ASHEIM, B., COOKE, P. & MARTIN, R. (2006): The rise of the cluster concept in regional analysis and policy: a critical assessment. In: ASHEIM, B., COOKE, P. & MARTIN, R. (eds): *Clusters and regional development: critical reflections and explorations*. Routledge: London and New York, 1-29.

ASHEIM, B. & GERTLER, M. (2006): The Geography of Innovation: Regional Innovation Systems. In: FAGERBERG, J., MOWERY, D. & NELSON, R. (eds): *The Oxford Handbook of Innovation*. Oxford University Press: Oxford, 291-317.

ASHEIM, B., BOSCHMA, R. & COOKE, P. (2011): Constructing regional advantage: platform policies based on related variety and differentiated knowledge bases. In: *Regional Studies* 45: 893-904.

AUDRETSCH, D. & DOHSE, D. (2007): Location: A Neglected Determinant of Firm Growth. In: *Review of World Economics* 143: 79–107.

AUDRETSCH, D. & FELDMAN, M. (1996): Innovative clusters and the industry life cycle. In: *Review of Industrial Organization* 11: 253–273.

AUDRETSCH, D. & KEILBACH, M. (2008): Resolving the knowledge paradox: Knowledge-spillover entrepreneurship and economic growth. In: *Research Policy* 37: 1697–1705.

AUDRETSCH, D., KEILBACH, M. & LEHMANN, E. (2006): *Entrepreneurship and economic growth*. Oxford University Press: Oxford.

AUDRETSCH, D. & LEHMANN, E. (2005): Mansfield's Missing Link: The Impact of Knowledge Spillovers on Firm Growth. In: *Journal of Technology Transfer* 30: 207–210.

AUTANT-BERNARD, C. (2001): The geography of knowledge spillovers and technological proximity. In: *Economics of Innovation and New Technology* 10: 237-254.

AXTELL, R. (2001): Zipf distribution of US firm sizes. In: *Science* 293: 1818-1820.

AXTELL, R. (2013): Endogenous Firms and Their Dynamics. Working Paper.

AYYAGARI, M., DEMIRGÜÇ-KUNT, A. & MAKSIMOVIC, V. (2010): Formal versus informal finance: evidence from China. In: *Review of Financial Studies* 23: 3048–3097.

B |

BACHE, S., DAHL, C. & KRISTENSEN, J. (2013): Headlights on tobacco road to low birthweight outcomes. In: *Empirical Economics* 44: 1593-1633.

BADE, F. (1987): Regionale Beschäftigungs-entwicklung und produktionsorientierte Dienstleistungen. DIW Berlin: Sonderheft 143.

BAI, C., LU, J. & TAO, Z. (2006): The multitask theory of state enterprise reform: empirical evidence from China. In: *American Economic Review* 96: 353–357.

BALDWIN, J., BECKSTEAD, D., BROWN, W. & RIGBY, D. (2008): Agglomeration and the geography of localization economies in Canada. In: *Regional Studies* 42: 117–132.

BALDWIN, J., BROWN, M. & RIGBY, D. (2010): Agglomeration Economies: Microdata

Panel Estimates from Canadian Manufacturing. In: *Journal of Regional Science* 50: 915–934.

BARABÁSI, A. & ALBERT, R. (1999): Emergence of scaling in random networks. In: *Science* 286: 509-512.

BARBOSA, N. & EIRIZ, V. (2011): Regional Variation of Firm Size and Growth: The Portuguese Case. In: *Growth and Change* 42: 125–158.

BARRO, R. (1999): Notes on growth accounting. In: *Journal of Economic Growth* 4: 119-137.

BARRO, R. & SALA-I-MARTIN, X. (1992): Convergence. In: *Journal of Political Economy* 100: 223-251.

BASILE, R. & USAI, S. (2012): Analysis of regional endogenous growth. CRENoS Working Paper.

BASTESEN, J. & VATNE, E. (2014): Rapid-growth firms: exploring the role and location of entrepreneurial ventures. In: KARLSSON, C., JOHANSSON, B. & STOUGH, R. (eds): *Agglomeration, Clusters and Entrepreneurship: Studies in Regional Economic Development*. Edward Elgar: Cheltenham/UK and Northamption/USA, 159-198.

BATTY, M. (2008): The Size, Scale, and Shape of Cities. In: *Science* 319: 769–771.

BATTY, M. (2013): *The New Science of Cities*. MIT Press: Hong Kong.

BEAUDRY, C. & SWANN, P. (2009): Firm growth in industrial clusters in the United Kingdom. In: *Small Business Economics* 32: 409–424.

BEAUDRY, C. & SCHIFFAUEROVA, A. (2009): Who's right, Marshall or Jacobs? The localization versus urbanization debate. In: *Research Policy* 38: 318-337.

BEINHOCKER E. (2007): *The origin of wealth: Evolution, complexity, and the radical remaking of economics*. Harvard Business School Press: Boston, MA.

BERGEK, A., BERGGREN, C., MAGNUSSEN, T. & HOBDAY, M. (2013): Technological discontinuities and the challenge for incumbent firms: Destruction, disruption or creative destruction. In: *Research Policy* 42: 1210-1224.

BEUGELSDIJK, S. (2007): The Regional Environment and a Firm's Innovative Performance: A Plea for a Multilevel Interactionist Approach. In: *Economic Geography* 83: 181–199.

BEENSTOCK, M. & FELSENSTEIN, D. (2007): Spatial vector autoregressions. In: *Spatial Economic Analysis* 2: 167-196.

BIANCHINI, S., BOTTAZZI, G. & TAMAGNI, F. (2014): What does (or does not) determine persistent corporate high-growth. LEM Working Paper Series 2014/11.

BINDER, J. & SCHWENGLER, B. (2006): Neuer Gebietszuschnitt der Arbeitsmarktregionen im Raum Berlin und Brandenburg. Kritische Überprüfung der bisher gültigen Arbeitsmarktregionen und Vorschläge für einen Neuzuschnitt. IAB-Forschungsbericht 4/2006, Nürnberg.

BINDER, M. & COAD, A. (2014): Causal linkages between work and life satisfaction and their determinants in a structural VAR approach. In: *Economics Letters* 124: 263-268.

BLIEN, U., SUEDEKUM, J. & WOLF, K. (2006): Local employment growth in West Germany: A dynamic panel approach. In: *Labour Economics* 13: 445–458.

BLONINGEN, B. & MA, A. (2007): Please pass the catch-up: The relative performance of Chinese and foreign firms in Chinese exports. NBER working paper 13376.

BOGLIACINI, F. & Pianta, M. (2010): Innovation and Employment: a Reinvestigation using Revised Pavitt classes. In: *Research Policy* 39: 799-809.

BORTZ, J., LIENERT, G. & BOEHNKE, K. (2008): *Verteilungsfreie Methoden in der Biostatistik*. Springer: Heidelberg.

BOSCHMA, R. (2005): Proximity and Innovation: A Critical Assessment. In: *Regional Studies* 39: 61–74.

BOSCHMA, R. (2009): Evolutionary economic geography and its implications for regional innovation policy. Papers in Evolutionary Economic Geography *9*.12.

BOSCHMA, R. (2014): Towards an evolutionary perspective on regional

resilience. In: Regional Studies DOI: 10.1080/00343404.2014.959481.

BOSCHMA, R., ERIKSSON, R. & LINDGREN, U. (2009): How does labour mobility affect the performance of plants? The importance of relatedness and geographical proximity. In: *Journal of Economic Geography* 9: 169–190.

BOSCHMA, R. & FRENKEN, K. (2006): Why is economic geography not an evolutionary science? Towards an evolutionary economic geography. In: *Journal of Economic Geography* 6: 273-302.

BOSCHMA, R. & FRENKEN, K. (2011): The emerging empirics of evolutionary economic geography. In: *Journal of Economic Geography* 11: 295–307.

BOSCHMA, R. & IAMMARINO, S. (2009): Related Variety, Trade Linkages, and Regional Growth in Italy. In: *Economic Geography* 85: 289–311.

BOSCHMA, R. & LAMBOOY, J. (1999): Evolutionary economics and economic geography. In: *Journal of Evolutionary Economics* 9: 411-429.

BOSCHMA, R. & TER WAL, A. (2007): Knowledge networks and innovative performance in an industrial district: the case of a footwear district in the South of Italy. In: *Industry and Innovation* 14: 177-199.

BOTTAZZI, G. (2004): Subbotools user's manual. LEM Working Paper Series 2004/14

BOTTAZZI, G. (2008): On the relationship between firms' size and growth rate. In: *Economics Bulletin* 3: 1–7.

BOTTAZZI, G. (2012): Maximum likelihood with order statistics. Working paper.

BOTTAZZI G., CEFIS E. & DOSI G. (2002): Corporate growth and industrial structure: Some evidence from the Italian manufacturing industry. In: *Industrial and Corporate Change* 11: 705–723

BOTTAZZI G., CEFIS E., DOSI G. & SECCHI, A. (2007): Invariances and Diversities in the Patterns of Industrial Evolution: Some Evidence from Italian Manufacturing Industries. In: *Small Business Economics* 29: 137–159.

BOTTAZZI, G., COAD, A., JACOBY, N. & SECCHI, A. (2011): Corporate growth and industrial dynamics: evidence from French manufacturing. In: *Applied Economics* 43: 103–116.

BOTTAZZI, G., DOSI, G., LIPPI, M., PAMMOLLI, F. & RICCABONI, M. (2001): Innovation and corporate growth in the evolution of the drug industry. In: *International Journal of Industrial Organization 19*: 1161-1187.

BOTTAZZI, G., DOSI, G., FAGIOLO, G. & SECCHI, A. (2008): Sectoral and geographical specificities in the spatial structure of economic activities. In: *Structural Change and Economic Dynamics* 19: 189-202.

BOTTAZZI, G. & SECCHI, A. (2003): Common Properties and Sectoral Specificities in the Dynamics of U.S. Manufacturing Companies. In: *Review of Industrial Organization* 23: 217–232.

BOTTAZZI, G. & SECCHI, A. (2006a): Explaining the distribution of firm growth rates. In: *RAND Journal of Economics* 37: 235–256.

BOTTAZZI, G. & SECCHI, A. (2006b): Gibrat's Law and diversification. In: *Industrial and Corporate Change* 15: 847-875.

BOTTAZZI, G. & SECCHI, A. (2011): A new class of asymmetric exponential power densities with applications to economics and finance. In: *Industrial and Corporate Change* 20: 991-1030.

BOTTAZZI, G. & SECCHI, A. (2013): Financial Fragility and the Distribution of Firm Growth Rates. LEM Working Paper Series.

BOTTAZZI, G., SECCHI, A. & TAMAGNI, F. (2014): Financial Constraints and Firm Dynamics. In: *Small Business Economics* 42: 99-116.

BOTTAZZI, L. & PERI, G. (2003): Innovation and spillovers in regions: Evidence from European patent data. In: *European Economic Review* 47: 687-710.

BRANDT, L., BIESENBROECK, J. & ZHANG, Y. (2012): Creative Accounting or Creative Destruction? Firm-level Productivity Growth in Chinese Manufacturing. In: *Journal of Development Economics* 97: 339–351.

BRAUNERHJELM, P., ACS, Z., AUDRETSCH, D. & CARLSSON, B. (2010): The missing link: knowledge diffusion and entrepreneurship in endogenous growth. In: *Small Business Economics* 34: 105–125.

BRAVO-BIOSCA, A., CRISCUOLO, C. & MENON, C. (2013): What drives the dynamics of business growth? OECD Science, Technology and Industry Policy Papers No. 1.

BREINLICH, H., OTTAVIANO, G. & TEMPLE, J. (2014): Regional growth and regional decline. In: AGHION, P. & DURLAUF, S. (eds): *Handbook of Economic Growth* (Vol. 2). The Netherlands: North-Holland, 683-779.

BRENNER, T. (2001): Simulating the evolution of localised industrial clusters-an identification of the basic mechanisms. In: *Journal of Artificial Societies and Social Simulation* 4.

BRENNER, T. (2004): *Local Industrial Clusters - Existence, Emergence, and Evolution*. Routledge: London.

BRENNER, T. (2005): A Stochastic Theory of Geographic Concentration and the Empirical Evidence in Germany. MPI Jena: Papers on Economics and Evolution 05.23.

BRENNER, T. (2006): Identification of Local Industrial Clusters in Germany. In: *Regional Studies* 40: 991-1004.

BRENNER, T. & BROEKEL, T. (2011): Methodological issues in measuring innovation performance of spatial units. In: *Industry and Innovation* 18: 7-37.

BRENNER, T., CANTNER, U. & GRAF, H. (2011): Innovation Networks: Measurement, Performance and Regional Dimension. In: *Industry and Innovation* 18: 1-5.

BRENNER, T. & GILDNER, A. (2006): The Long-term Implications of Local Industrial Clusters. In: *European Planning Studies* 14: 1315-1328

BRENNER, T. & SCHIMKE, A. (2014): Growth Development Paths of Firms - A Study of Smaller Businesses. In: *Journal of Small Business Management* DOI: 10.1111/jsbm.12081.

BRENNER, T. & SCHLUMP, C. (2011): Policy Measures and their Effects in the Different Phases of the Cluster Life Cycle. In: *Regional Studies* 45: 1363–1386.

BRENNER, T. & SCHLUMP, C. (2013): Universities, Public Research and Regional Innovation Output: An Empirical Study of 19 Technologies in Germany. Marburg Working Paper on Innovation and Space 03.13.

BRENNER, T. & WERKER, C. (2007): A Taxonomy of Inference in Simulation Models. In: *Computational Economics* 30: 227-244.

BRESCHI, S. & LISSONI, F. (2001): Knowledge Spillovers and Local Innovation Systems: A Critical Survey. In: *Industrial and Corporate Change* 10: 975–1005.

BRESCHI, S. & LISSONI, F. (2009): Mobility of skilled workers and co-invention networks: an anatomy of localized knowledge flows. In: *Journal of Economic Geography* 9: 439-468.

BRESCHI, S., MALERBA, F. & ORSENIGA, L. (2000): Technological regimes and Schumpeterian patterns of innovation. In: *The Economic Journal* 110: 388-410.

BRIGUGLIO, L., CORDINA, G., FARRUGIA, N. & VELLA, S. (2009): Economic Vulnerability and Resilience: Concepts and Measurements. In: *Oxford Development Studies* 27:229-247

BROCK, W. (1999): Scaling in Economics: a Reader's Guide. In: *Industrial and Corporate Change* 8: 409–446.

BROEKEL, T. (2007): A Concordance between Industries and Technologies: Matching the technological fields of the Patentatlas to the German Industry Classification. Jena Economic Research Papers 2007-041.

BRÖKEL, T., BRACHERT, M., DUSCHL, M. & BRENNER, T. (*forthcoming*): Joint R&D subsidies, related variety and improvements in regions' innovation efficiency.

BRYNJOLFSSON, E. & HITT, L. (1995): Information technology as a factor of production: The role of differences among firms. In: *Economics of*

Innovation and New Technology 3: 183-200.

BUCHINSKY, M. (1998): Recent Advances in Quantile Regression Models. In: *The Journal of Human Resources* 33: 88-126.

BUENSTORF, G. & GUENTHER, C. (2011): No place like home? Relocation, capabilities, and firm survival in the German machine tool industry after World War II. In: *Industrial and Corporate Change* 20: 1-28.

BUENSTORF, G. & KLEPPER, S. (2009) Heritage and Agglomeration: The Akron Tyre Cluster Revisited. In: *Economic Journal* 119: 705–733.

BUERGER, M., BROEKEL, T. & COAD, A. (2012): Regional Dynamics of Innovation – Investigating the Co-Evolution of Patents, R&D, and Employment. In: *Regional Studies* 46: 565-582.

BUERGER, M., VANOORT, F. & VAN DER KNAAP, B. (2010): A Treatise on the Geographical Scale of Agglomeration Externalities and the MAUP. In: *Scienze Regionali* 9: 19-40.

BUSCH, O. (2007): When have all the graduates gone? Internal cross-state migration of graduates in Germany 1984-2004. In: SOEPpapers on Multidisciplinary Panel Data Research 26: 1-17.

C |

CABALLERO, R., ENGEL, E. & HALTIWANGER, J. (1997): Aggregate employment dynamics: building up from microeconomic evidence. In: *American Economic Review* 87: 115-137.

CABRAL, L. & MATA, J. (2003): On the evolution of the firm size distribution: Facts and theory. In: *American Economic Review* 93: 1075-1090.

CAINELLI, G. (2008): Spatial Agglomeration, Technological Innovations, and Firm Productivity: Evidence from Italian Industrial Districts. In: *Growth and Change* 39: 414–435.

CAINELLI, G. & LUPI, C. (2008): Does Spatial Proximity Matter? Micro-evidence from Italy. Economics & Statistics Disucssion Paper 42/08.

CAO, C., LI, N., LI, X. & L. LIU (2013): Reforming China's S&T system. In: *Science* 341: 460–462.

CANAY, I. (2011): A simple approach to quantile regression for panel data. In: *The Econometrics Journal* 14: 368-386.

CANNING, D., AMARAL, L., LEE, Y., MEYER, M. & STANLEY, E. (1998): Scaling the volatility of GDP growth rates. In: *Economics Letters* 60: 335–341.

CANTNER, U. & HANUSCH, H. (2002): Evolutionary economics, its basic concepts and methods. In: LIM, H., PARK, U. & HARCOURT, G. (eds) *Editing Economics: Essays in Honour of Mark Perlman*. Routledge: London, New York, 182-207.

CAPASSO, M., ALESSI, L., BARIGOZZI, M. & FAGIOLO, G. (2009): On approximating the distributions of goodness-of-fit test statistics based on the empirical distribution function: the case of unknown parameters. In: *Advances in Complex Systems* 12: 157-167.

CAPASSO, M., CEFIS, E. & SAPIO, S. (2011): Framing the empirical findings on firm growth. LEM Working Paper Series 2011/01.

CAPASSO, M., CEFIS, E. & SAPIO, S. (2013) Reconciling quantile autoregressions of firm size and variance-size scaling. In: *Small Business Economics* 41: 609-632.

CAPELLO, R. (2009): Space, growth and development. In: CAPELLO, R. & NIJKAMP, P. (eds): *Handbook of Regional Growth and Development Theories*. Edward Elgar: Cheltenham, UK, 33-52.

CAPELLO, R. & NIJKAMP, P. (2009): Introduction: regional growth and development theories in the twenty-first century – recent theoretical advances and future challenges. In: CAPELLO, R. & NIJKAMP, P. (eds): *Handbook of Regional Growth and Development Theories*. Edward Elgar: Cheltenham, UK, 1-19.

CASSIA, L., COLOMBELLI, A. & PALEARI, S. (2009): Firms' growth: Does the

innovation system matter? In: *Structural Change and Economic Dynamics* 20: 211–220.

CASTALDI, C. & DOSI, G. (2006): Income Levels and Income Growth. Some New Cross-Country Evidence and Some Interpretative Puzzles. LEM Working Paper Series.

CASTALDI, C. & DOSI, G. (2009): The patterns of output growth of firms and countries: Scale invariances and scale specificities. In: *Empirical Economics* 37: 475–495.

CASTALDI, C., FRENKEN, K. & LOS, B. (2014): Related Variety, Unrelated Variety and Technological Breakthroughs: an analysis of U.S. state-level patenting. Regional Studies DOI: 10.1080/00343404.2014.940305.

CASTALDI, C. & SAPIO, S. (2008): Growing like mushrooms? Sectoral evidence from four large European economies. In: *Journal of Evolutionary Economics* 18: 509–527.

CAVES, R. (1998): Industrial organization and new findings on the turnover and mobility of firms. In: *Journal of Economic Literature* 36: 1947-1982.

CHA, K., DANG, V. & YAN, I. (2012): Chinese firms' political connection, ownership, and financing constraints. In: *Economics Letters* 115: 164–167.

CHAPPLE, K. & LESTER, T. (2010): The resilient regional labour market? The US case. In: *Cambridge Journal of Regions, Economy and Society* 3: 85–104.

CHESHER, A. (1979): Testing the law of proportionate effect. In: *The Journal of Industrial Economics* 52: 403-411.

CHESHIRE, P. & MALECKI, E. (2004): Growth, development, and innovation: A look backward and forward. In: *Papers in Regional Science 83*(1): 249-267.

CHOI, S., LEE, S. & WILLIAMS, C. (2011): Ownership and firm innovation in a transition economy: Evidence from China. In: *Research Policy* 40: 441–452.

CHOW, G. (2006): Are Chinese Official Statistics Reliable? In: *CESifo Economic Studies* 52: 396–414.

CIRILLO, P. & HÜSKER, J. (2009): On the upper tail of Italian firms' size distributions. In: *Physika A* 388: 1546-1554.

CLARKE, J., HOLT, R. & BLUNDEL, R. (2014): Re-imagining the growth process: (co)-evolving metaphorical representations of entre-preneurial growth. In: *Entrepreneurship and Regional Development* 26: 234-256.

COAD, A. (2007): A Closer Look at Serial Growth Rate Correlation. In: *Review of Industrial Organization* 31: 69–82.

COAD, A. (2009): *The Growth of Firms: A Survey of Theories and Empirical Evidence*. Edward Elgar: Cheltenham.

COAD, A. (2010a): Exploring the processes of firm growth: evidence from a vector auto-regression. In: *Industrial and Corporate Change* 19: 1677–1703.

COAD, A. (2010b): The exponential age distribution and the Pareto firm size distribution. In: *Journal of Industry, Competition and Trade* 10: 389-395.

COAD, A (2012): Firms as Bundles of Discrete Resources - Towards an Explanation of the Exponential Distribution of Firm Growth Rates. In: *Eastern Economic Journal* 38: 189-209.

COAD, A. (forthcoming): The meso-level: structural persistency. In CANTNER, U. & GUERZONI M. (eds): *Handbook on industrial dynamics and evolutionary economics.* Edward Elgar: Cheltenham.

COAD, A., COWLING, M. & SIEPEL, J. (2012): Growth Processes of High-Growth Firms in the UK. Nesta Working Paper No. 12/10.

COAD A., DAUNFELDT S., HÖLZL W., JOHANSSON D. & NIGHTINGALE P. (2014): High-growth firms: introduction to the special section. In: *Industrial and Corporate Change* 23: 91-112.

COAD, A. & HÖLZL, W. (2009): On the Autocorrelation of Growth Rates. In: *Journal of Industry, Competition and Trade* 9: 136-166.

COAD, A. & RAO, R. (2008): Innovation and firm growth in high-tech sectors: A quantile regression approach. In: *Research Policy 37*: 633-648.

COAD, A., SAVONA, M. & PELLEGRINO, G. (2014): Don't Stop Me Now: Barriers to

innovation and firm productivity. SPRU-Science and Technology Policy Research No. 2014-04.

COAD, A., SEGARRA, A. & TERUEL, M. (2013). Like milk or wine: Does firm performance improve with age? In: *Structural Change and Economic Dynamics* 24: 173-189.

COAD, A. & TAMVADA, J. (2012): Firm growth and barriers to growth among small firms in India. In: *Small Business Economics* 39: 383–400.

COASE, R. (1937): The nature of the firm. In: *Economica* 4: 386-405.

COHEN, W. & LEVINTHAL, D. (1990): Absorptive capacity: a new perspective on learning and innovation. In: *Administrative Science Quarterly* 35: 128-152.

COHEN, S., DI MININ, A., MOTOYAMA, Y. & PALMBERG, C. (2009): The Persistence of Home Bias for Important R&D in Wireless Telecom and Automobiles. In: *Review of Policy Research* 26: 55-79.

COMBES, P., MAGNAC, T. & ROBIN, J. (2004): The dynamics of local employment in France. In: *Journal of Urban Economics* 56: 217–243.

COOKE, P., GOMEZ URANGA, M. & ETXEBARRIA, G. (1997): Regional innovation systems: Institutional and organisational dimensions. In: *Research policy* 26: 475-491.

CRESCENZI, R. (2005): Innovation and regional growth in the enlarged Europe: the role of local innovative capabilities, peripherality, and education. In: *Growth and Change*, 36: 471-507.

CROSS, R., MCNAMARA, H. & POKROVISKII, A. (2010): Memory of Recessions. Strathclyde Discussion Papers in Economics 10-09.

CULL, R. & XU, L. (2005): Institutions, ownership, and finance: the determinants of profit reinvestment among Chinese firms. In: *Journal of Financial Economics* 77: 117–146.

D |

DASGUPTA, M. & MISHRA, S. (2004): Least Absolute Deviation Estimation of Linear Econometric Models: A Literature Review. MPRA Paper 1781.

DAUNFELDT, S. & HALVARSSON, D. (2012): Are high-growth firms one hit wonders? Evidence from Sweden. HUI Stockholm Working Paper No 70.

DAUTEL, V. & WALTHER, O. (2013): The geography of innovation in a small metropolitan region: An intra-regional approach in Luxembourg. In: *Papers in Regional Science* doi:10/1111.pirs.12019

DAVIDSSON, P., DELMAR, F. & WIKLUND, J. (2006): Entrepreneurship as growth: Growth as entrepreneurship. In: DAVIDSSON, P., DELMAR, F. & WIKLUND, J. (eds): *Entrepreneurship and the Growth of Firms.* Edward Elgar Publishing: Cheltenham, UK.

DAVIES, S. (2011): Regional resilience in the 2008-2010 downturn: comparative evidence from European countries. In: *Cambridge Journal of Regions, Economy and Society* 4: 369-382.

DE GROOT, H., POOT, J. & SMIT, M. (2009): Agglomeration externalities, innovation and regional growth: theoretical reflections and meta-analysis. In: CAPELLO, R. & NIJKAMP, P. (eds): *Handbook of Regional Growth and Development Theories*. Edward Elgar: Cheltenham, 256-281.

DELGADO, M., PORTER, M. & STERN, S. (2010): Clusters and entrepreneurship. In: *Journal of Economic Geography* 10: 495–518.

DELGADO, M., PORTER, M. & STERN, S. (2012). Clusters, convergence, and economic performance. NBER Working Paper No. 18250.

DELLI GATTI, D., DI GUILMI, C., GAFFEO, E., GIULIONI, G., GALLEGATI, M. & PALESTRINI, A. (2005): A new approach to business fluctuations: heterogeneous interacting agents, scaling laws and financial fragility. In: *Journal of Economic Behavior & Organization* 56: 489–512.

DELMAR, F. & DAVIDSSON, P. (1998): A taxonomy of high-growth firms. In: Reynolds, P. (eds): *Frontiers of Entrepreneurship Research*. Babson College: Wellesley, MA, 399–413.

DELMAR, F., DAVIDSSON, P. & GARTNER, W. (2003): Arriving at the high-growth firm. In: *Journal of Business Venturing* 18: 189–216.

DEMIRALP, S. & HOOVER, K. (2003): Searching for the Causal Structure of a Vector Autoregression. In: *Oxford Bulletin of Economics and Statistics* 65: 745-767.

DI GIACINTO, V. (2010): On vector autoregressive modeling in space and time. In: *Journal of Geographical Systems 12*: 125-154.

DIODATO, D. & WETERINGS, A. (2012): The Resilience of Dutch Regions to Economics Shocks. PEEG Working Paper Series 12.15.

DIXIT, A. & STIGLITZ, J. (1977): Monopolistic competition and optimum product diversity. In: *The American Economic Review* 67: 297-308.

DÖRING, T. & SCHNELLENBACH, J. (2006): What Do We Know about Geographical Knowledge Spillovers and Regional Growth? A Survey of the Literature. In: *Regional Studies* 40: 375–395.

DOSI, G. (1988): Sources, Procedures, and Microeconomic Effects of Innovation. In: *Journal of Economic Literature* 26: 1120-1171.

DOSI, G. (2007): Statistical regularities in the evolution of industries: a guide through some evidence and challenges for the theory. In: MALERBA, F. & BRUSONI, S. (eds): *Perspectives on Innovation*. Cambridge University Press: Cambridge, UK, 153–186.

DOSI, G., GRAZZI, M., TOMASI, C. & ZELI, A. (2012): Turbulence underneath the big calm? The micro-evidence behind Italian productivity dynamics. In: *Small Business Economics 39:* 1043–1069.

DOSI, G., LECHEVALIER, S. & SECCHI, A. (2010): Introduction: Interfirm heterogeneity - nature, sources and consequences for industrial dynamics. In: *Industrial and Corporate Change* 19: 1867–1890.

DOSI, G., LEI, J. & YU X. (2013): Institutional Change and Productivity Growth in China's Manufacturing: The Microeconomics of Creative Restructuring. LEM Working Paper 2013/07.

DOSI, G., MARSILI, O., ORSENIGO, L. & SALVATORE, R. (1995): Learning, market selection and the evolution of industrial structures. In: *Small Business Economics* 7: 411–436.

DOSI, G. & NELSON, R. (1994): An introduction to evolutionary theories in economics. In: *Journal of Evolutionary Economics* 4:153-172.

DOSI, G. & TEECE, D. (1998): Organizational Competencies and the Boundaries of the Firm. In: ARENA R. & LONGHI C. (eds): *Markets and Organization*. Springer: Berlin, Heidelberg, 281-302.

DUMAIS, G., ELLISON, G. & GLAESER, E. (2002) Geographic concentration as a dynamic process. In: *Review of Economics and Statistics* 84: 193-204.

DUNNE, P. & HUGHES, A. (1994): Age, size, growth and survival: UK companies in the 1980s. In: *The Journal of Industrial Economics* 42: 115-140.

DUNNE, T., ROBERTS, M. & SAMUELSON, L. (1988): Patterns of firm entry and exit in US manufacturing industries. In: *The RAND Journal of Economics* 19: 495-515.

DUNNING, J. & LUNDAN, S. (2009): The Internationalization of Corporate R&D: A Review of the Evidence and Some Policy Implications for Home Countries. In: *Review of Policy Research* 26: 13-33.

DRUCKER, J. (2012): The spatial extent of agglomeration economies: evidence from three U.S. manufacturing industries. CES Working Paper 12-01.

DRUCKER J. & FESER E. (2012): Regional industrial structure and agglomeration economies: An analysis of productivity in three manufacturing industries. In: *Regional Science and Urban Economics* 42: 1–14.

DURANTON, G. (2008): Spatial Economics. In: DURLAUF S. & BLUME, L. (eds): *The New*

Palgrave Dictionary of Economics. Palgrave Macmillan, Second Edition.

DURANTON, G. & OVERMAN, H. (2005): Testing for Localization Using Micro-Geographic Data. In: *Review of Economic Studies* 72: 1077–1106.

DURLAUF, S. (2001): Manifesto for a growth econometrics. In: *Journal of Econometrics* 100: 65–69.

DURLAUF, S., JOHNSON, P. & TEMPLE, J. (2005): Growth econometrics. In: AGHION, P. & DURLAUF, S. (eds): *Handbook of economic growth* (Vol. 1). North Holland: The Netherlands, 555-677.

DUSCHL, M. (2014): Regional resilience and fat tails - A stochastic analysis of firm growth rate distributions of German regions. Working Papers on Innovation and Space Vol. 01.14.

DUSCHL, M. & BRENNER, T. (2013): Characteristics of regional industry-specific employment growth rates' distributions. In: *Papers in Regional Science* 92: 249-270.

DUSCHL, M. & PENG, S. (2014): Chinese firm dynamics and the role of ownership type: A conditional estimation approach of the Asymmetric Exponential Power (AEP) density. MPI Jena: Papers on Economics and Evolution 1401.

DUSCHL, M., SCHIMKE, A., BRENNER, T. & LUXEN, D. (2014a) Firm Growth and the Spatial Impact of Geolocated External Factors – Empirical Evidence for German Manufacturing Firms. In: *Jahrbuch für Nationalökonomie und Statistik* 234: 234-256.

DUSCHL, M., SCHOLL, T., BRENNER, T., LUXEN, D. & RASCHKE, F. (2014b): Industry-specific firm growth and agglomeration. In: *Regional Studies* DOI: 10.1080/00343404.2013.861059.

DUTZ, M., KESSIDES, I., O'CONNELL & S., WILLIG, R. (2011): Competition and Innovation-Driven Inclusive Growth. The World Bank: Policy Research Working Paper No. 5852.

E |

ECKEY, H., KOSFELD, R. & TÜRCK, M. (2006): Abgrenzung deutscher Arbeitsmarktregionen. In: *Raumforschung und Raumordnung 64*: 299-309.

ECKEY, H., SCHWENGLER, B. & TÜRCK, M. (2007): Vergleich von deutschen Arbeitsmarktregionen. IAB discussion paper 2007/3.

EICHENGREEN, B., PARK, D. & SHIN, K. (2012): When Fast-Growing Economies Slow Down: International Evidence and Implications for China. In: *Asian Economic Papers* 11: 42–87.

ELLISON, G. & GLAESER, E. (1997): Geographic concentration in U.S. manufacturing industries: a dartboard approach. In: *The Journal of Political Economy* 105: 889-927.

ERICSON, R. & PAKES, A. (1995): Markov-perfect industry dynamics: A framework for empirical work. In: *The Review of Economic Studies* 62: 53-82.

ERIKSSON, R. (2011): Localized Spillovers and Knowledge Flows: How Does Proximity Influence the Performance of Plants? In: *Economic Geography* 87: 127–152.

ERLINGSSON, E., ALFARANO, S., RABERTO, M. & STEFÁNSSON, H. (2013): On the distributional properties of size, profit and growth of Icelandic firms. In: *Journal of Economic Interaction and Coordination* 8: 57-74.

ETZKOWITZ, H. & LEYDESDORFF, L. (2000): The dynamics of innovation: from National Systems and "Mode 2" to a Triple Helix of university–industry–government relations. In: *Research Policy 29*: 109-123.

EUROPEAN COMMISSION (2003): Commission Recommendation of 6 May 2003 concerning the definition of micro, small and medium-sized enterprises. L 125/36.

EUROPEAN COMMISSION (2010): Europe 2020: A Strategy for Smart, Sustainable and Inclusive Growth: Communication From the Commission. European

Commission Publications Office: Brussels, Belgium.

EVANS, D. (1987): The relationship between firm growth, size, and age: estimates for 100 manufacturing industries. In: *The Journal of Industrial Economics* 35: 567-581.

F |

FAGIOLO, G., NAPOLETANO, M. & ROVENTINI, A. (2008): Are output growth-rate distributions fat-tailed? Some evidence from OECD countries. In: *Journal of Applied Econometrics* 23: 639–669.

FAGGIAN, A. & MCCANN, P. (2009): Universities, agglomerations and graduate human capital mobility. In: *Tijdschrift voor economische en sociale geografie* 100: 210-223.

FALCK, O., GUENTHER, C., HEBLICH, S. & KERR, W. (2013): From Russia with love: the impact of relocated firms on incumbent survival. In: *Journal of Economic Geography 13*: 419-449.

FELDMAN, M. (1994): *The geography of innovation* (Vol. 2). Kluwer Academic Publishers: Dordrecht, Netherlands.

FELDMAN, M. (1999): The New Economics of Innovation, Spillovers and Agglomeration: A Review of Empirical Studies. In: *Economics of Innovation and New Technology* 8: 5-25.

FELDMAN, M. (2000): Location and innovation: the new economic geography of innovation, spillovers, and agglomeration. In: CLARK, G., FELDMAN, M. & GERTLER, M. (eds): *Oxford Handbook of Economic Geography*. Oxford University Press: Oxford, 373 – 394.

FESER, E. (2002): Tracing the Sources of Local External Economies. In: *Urban Studies* 39: 2485–2506.

FINGLETON, B., GARRETSEN, H. & MARTIN, R. (2012): Recessionary Shocks and Regional Employment: Evidence on the Resilience of UK Regions. In: *Journal of Regional Science* 52: 109–133.

FLIGNER, M. & POLICELLO, G. (1981): Robust rank procedures for the Behrens-Fisher problem. In: *Journal of the American Statistical Association* 76: 141-206.

FLORIDA, R. (2005): *Cities and the Creative Class*. Routledge: New York.

FORNAHL, D. & OTTO, A. (2008): Between snails and gazelles? Analysing the effects of regional specialisation on firm growth: a quantile regression approach. Druid Conference Paper 2008.

FORNAHL, D. & BRENNER, T. (2009): Geographic concentration of innovative activities in Germany. In: *Structural Change and Economic Dynamics* 20: 163–182.

FORNAHL, D., BROEKEL, T. & BOSCHMA, R. (2011): What drives patent performance of German biotech firms? The impact of R&D subsidies, knowledge networks and their location. In: *Papers in Regional Science* 90: 395–418.

FRATESI, U. & SENN, L. (2009): Regional Growth, Connections and Economic Modelling: An Introduction. In: FRATESI, U. & SENN, L. (eds): *Growth and Innovation of Competitive Regions*, Advances in Spatial Science. Springer: Berlin, Heidelberg, 3-27.

FREEMAN, C. (1995): The 'National System of Innovation' in historical perspective. In: *Cambridge Journal of Economics 19*: 5-24.

FRENKEN, K. (2014): Smart Specialisation – an Empirical Assessment and Policy Implication. Keynote address at the 15th ISS conference in Jena, 28th of July.

FRENKEN, K. & BOSCHMA, R. (2007): A theoretical framework for evolutionary economic geography: industrial dynamics and urban growth as a branching process. In: *Journal of Economic Geography* 7: 635–649.

FRENKEN, K., CEFIS, E. & STAM, E. (2011): Industrial dynamics and economic geography: a survey. Eindhoven Centre for Innovation Studies (Working Paper 11.07).

FRENKEN, K., CEFIS, E. & STAM, E. (2014): Industrial Dynamics and Clusters: A Survey. Regional Studies DOI: 10.1080/00343404.2014.904505.

FRENKEN, K., VANOORT, F. & VERBURG, T. (2007): Related Variety, Unrelated

Variety and Regional Economic Growth. In: *Regional Studies* 41: 685-697.

FRIEDMAN, M. (1992): Do Old Fallacies Ever Die? In: *Journal of Economic Literature* 30: 2129-2132.

FRITSCH, M. & SLAVTCHEV, V. (2007): Universities and Innovation in Space. In: *Industry and Innovation* 14: 201-218.

FRITSCH, M. & SLAVTCHEV, V. (2011): Determinants of the Efficiency of Regional Innovation Systems. In: *Regional Studies* 45: 905-918.

FRITSCHE, U. & LOGEAY, C. (2002): Structural Unemployment and the Output Gap in Germany: Evidence from a SVAR Analysis within a Hysteresis Framework. DIW-Diskussionspapiere 312.

FU, D., PAMMOLLI, F., BULDYREV S., RICCABONI, M., MATIA, K., YAMASAKI, K., STANLEY, H. (2005): The growth of business firms: Theoretical framework and empirical evidence. In: *Proceedings of the National Academy of Sciences* 102: 18801-18806.

FUJITA, M. (2002): *Economics of agglomeration: Cities, industrial location, and regional growth*. Cambridge University Press: Cambridge.

G |

GABAIX X. (2009): Power Laws in Economics and Finance. In: *Annual Review of Economics* 1: 255-294.

GABAIX X. (2011): The granular origins of aggregate fluctuations. In: *Econometrica* 79: 733–772.

GAFFEO, E., GALLEGATI, M. & PALESTRINI, A. (2003): On the size distribution of firms: additional evidence from the G7 countries. In: *Physica A: Statistical Mechanics and its Applications* 324: 117-123.

GALLEGO, F. (2010): A population density grid of the European Union. In: *Population and Environment* 31: 460–473.

GALVAO, A. (2011): Quantile regression for dynamic panel data with fixed effects. In: *Journal of Econometrics 164*: 142-157.

GARNSEY, E., STAM, E. & HEFFERNAN, P. (2006): New Firm Growth: Exploring Processes and Paths. In: *Industry and Innovation* 13: 1–20.

GARRETSEN, H., MCCANN, P., MARTIN, R. & TYLER, P. (2013): The future of regional policy. In: *Cambridge Journal of Regions, Economy and Society* 6: 179-186.

GEDAJLOVIC, E. (1993): Ownership, strategy and performance: is the dichotomy sufficient? In: *Organization Studies* 14: 731–752.

GEHRKE, B., RAMMER, C., FRIETSCH, R. & NEUHÄUSLER, P. (2010): Listen wissens- und technologieintensiver Güter und Wirtschaftszweige. Studien zum deutschen Innovationssystem 19-2010.

GEISBERGER R., LUXEN, D., NEUBAUER, S., SANDERS, P. & VOLKER, L. (2010): Fast Detour Computation for Ride Sharing. In: ERLENBACH, T. & LÜBBECKE, M. (eds): 10thWorkshop on Algorithmic Approaches for Transportation Modelling, Optimization, and Systems (ATMOS'10), 88-99.

GEORGE, G. (2005): Slack resources and the performance of privately held firms. In: *Academy of Management Journal* 48: 661–676.

GEROSKI, P. (1989): Entry, Innovation and Productivity Growth. In: *The Review of Economics and Statistics* 71: 572–578.

GEROSKI, P. & GERGG, P. (1996): What makes firms vulnerable to recessionary pressure? In: *European Economic Review* 40: 551-557.

GEROSKI, P. & GUGLER, K. (2004): Corporate growth convergence in Europe. In: *Oxford Economic Papers* 56: 597-620.

GERTLER, M. (2003). Tacit knowledge and the economic geography of context, or the undefinable tacitness of being (there). In: *Journal of Economic Geography 3:* 75-99.

GIBRAT, R. (1931): *Les Inégalités Économiques*. Recueil Sirey: Paris.

GILBERT, B. (2012): Creative destruction: Identifying its geographic origins. In: *Research Policy* 41: 734–742.

GILBERT, B., MCDOUGALL, P. & AUDRETSCH, D. (2006): New venture growth: A review and extension. In: *Journal of Management 32:* 926-950.

GLAESER, E., KALLAL, H., SCHEINKMAN, J. & SHLEIFER, A. (1992): Growth in Cities. In: *Journal of Political Economy* 100: 1126-1152.

GLAESER, E., PONZETTO, G. & TOBIO K. (2014): Cities, Skills, and Regional Change. In: *Regional Studies* 48: 7-43

GORDON, P. (2013): Thinking about economic growth: cities, networks, creativity and supply chains for ideas. In: *Annals of Regional Science* 50: 667-684.

GOULD, W. (1992): Quantile regression with bootstrapped standard errors. In: *Stata Technical Bulletin* **9**: 19–21.

GOTTSCHALK, J. (2001): An Introduction into the SVAR Methodology: Identification, Interpretation and Limitations of SVAR models. Kiel Working Paper No. 1072.

GRAHAM, D. (2009): Identifying urbanisation and localisation externalities in manufacturing and service industries. In: *Papers in Regional Science* 88**:** 63–94.

GRAHAM, D., GIBBSON, S. & MARTIN, R. (2010): The spatial decay of agglomeration economies: estimates for use in transport appraisal. LSE Imperial College Working Paper.

GRETTON, A. & GYÖRFI, L. (2010): Consistent Nonparametric Tests of Independence. In: *Journal of Machine Learning Research*, 11: 1391-1423.

GRILICHES, Z. & MAIRESSE, J. (1995): Production Functions: The Search for Identification. NBER Working Paper Series No. 5067.

GROOT, S., MÖHLMANN, J., GARRETSEN, H. & GROOT, H. (2011): The crisis sensitivity of European countries and regions: stylized facts and spatial heterogeneity. In: *Cambridge Journal of Regions, Economy and Society* 4: 437–456.

GUAN, J., YAM, R., TANG, E. & LAU, A. (2009): Innovation strategy and performance during economic transition: Evidences from Beijing, China. In: *Research Policy* 38: 802–812.

GUARIGLIA, A., LIU, X. & SONG, L. (2011): Internal finance and growth: Microeconometric evidence on Chinese firms. In: *Journal of Development Economics* 96: 79–94.

GUIMÓN, J. (2009): Government strategies to attract R&D-intensive FDI. In: *Journal of Technological Transfer* 34: 364-379.

GUISO, L. & SCHIVARDI, F. (2007): Spillovers in industrial districts. In: *The Economic Journal* 117: 68–93.

H |

HALL, B. (1988): The relationship between firm size and firm growth in the US manufacturing sector. NBER Working Paper No. w1965.

HALL, P. & SOSKICE, D. (2001): *Varieties of capitalism: The institutional foundations of comparative advantage* (Vol. 8). Oxford University Press: Oxford.

HALTIWANGER, C. (1997): Measuring and Analyzing Aggregate Fluctuations: The Importance of Building from Microeconomic Evidence. Federal Reserve Bank of St. Louis.

HALVARSSON, D. (2012): Identifying High-Growth Firms. Ratio Working Paper No. 215.

HARRIS, R. (2011): Models of regional growth: past, present and future. In: *Journal of Economic Surveys* 25: 913-951.

HART, P. & PRAIS, S. (1956): The analysis of business concentration: a statistical approach. In: *Journal of the Royal Statistical Society A* 119: 150-191.

HASSNIK, R., KLAERDING, C. & MARQUES, P. (2014): Advancing Evolutionary Economic Geography by Engaged Pluralism. In: *Regional Studies* 48: 1295-1307.

HAUSMAN, J. (2001): Mismeasured variables in econometric analysis: problems from the right and problems from the left. In: *Journal of Economic Perspectives* 15: 57-67.

HECKMAN, J. (1979): Sample selection bias as a specification error. In: *Econometrica* 47: 153–61.

HENDERSON, J. (2003): Marshall's scale economies. In: *Journal of Urban Economics 53*: 1-28.

HENDERSON J., KUNCORO, A. & TURNER, M. (1995): Industrial Development in Cities. In: *Journal of Political Economy* 103: 1067–1090.

HIGSON, C., HOLLY, S., KATTUMAN, P. & PLATIS, S. (2004): The Business Cycle, Macroeconomic Shocks and the Cross-Section: The Growth of UK Quoted Companies. In: *Economica* 71: 299–318.

HILL, B. (1975): A simple general approach to inference about the tail of a distribution. In: *The Annals of Statistics* 3: 1163-1174.

HILL, E. & ATKINS, P. (2012): Economic Shocks and Regional Economic Resilience. In: WEIR, M., PINDUS, N, WIAL, H. & WOLMAN, H. (eds): *Building Resilient Regions: Urban and Regional Policy and Its Effects*. Brookings Institution Press: Washington, 193-274.

HODGSON, G. (2013): Understanding Organizational Evolution: Toward a Research Agenda Using Generalized Darwinism. In: *Organization Studies* 34: 973-992.

HOEKMAN, J., FRENKEN, K. & TIJSSEN, R. (2010): Research collaboration at a distance: Changing spatial patterns of scientific collaboration within Europe. In: *Research Policy* 39: 662-673.

HOLLY, S., PETRELLA, I., SANTORO, E. (2013): Aggregate Fluctuations and the Cross-sectional Dynamics of Firm Growth. In: *Journal of the Royal Statistical Society, Series A* 176: 459–479.

HOLM, J. & ØSTERGAARD, C. (2013): Regional Employment Growth, Shocks and Regional Industrial Resilience: A Quantitative Analysis of the Danish ICT Sector. In: *Regional Studies* DOI: 10.1080/00343404.2013.787159.

HOLZ, C. (2013): Chinese statistics: classification systems and data sources. In: MPRA Paper No. 43869.

HÖLZL, W. (2009): Is the R&D behaviour of fast-growing SMEs different? Evidence from CIS III data for 16 countries. In: *Small Business Economics* 33: 59–75.

HÖLZL, W. (2011): Unternehmenswachstum im internationalen Vergleich. In: WIFO Monatsberichte 84: 557–567.

HÖLZL, W. (2014): Persistence, survival, and growth: a closer look at 20 years of fast-growing firms in Austria. In: *Industrial and Corporate Change* 23: 199-231.

HOOGSTRA, G. & VAN DIJK, J. (2004): Explaining Firm Employment Growth: Does Location Matter? In: *Small Business Economics* 22: 179–192.

HOYER, P., JANZING, D., MOOJI, J., PETERS, J. & SCHÖLKOPF, B. (2009): Nonlinear causal discovery with additive noise models. In: *Advances in Neural Information Processing Systems* 21: 689-696.

HU, A., JEFFERSON, G. & JINCHANG, Q. (2005): R&D and technology transfer: firm-level evidence from Chinese industry. In: *Review of Economics and Statistics* 87: 780–786.

HUDSON, R. (2010): Resilient regions in an uncertain world: wishful thinking or a practical reality. In: *Cambridge Journal of Regions, Economy and Society* 3: 11-25.

HUGGINS, R. & THOMPSON, P. (2013): A Network-based view of regional growth. In: *Journal of Economic Geography* 14: 511-545.

HUNDT, C. & STERNBERG, R. (2014): Explaining new firm creating in Europe from a spatial and time perspective: A multilevel analysis based upon data of individuals, regions and countries. In: *Papers in Regional Science* DOI: 10.1111/pirs.12133.

HYMER, S. (1976): *The international operations of national firms: A study of direct foreign investment* (Vol. 14). MIT press: Cambridge, MA.

HYMER, S. & PASHIGIAN, P. (1962): Firm size and rate of growth. In: *The Journal of Political Economy* 70: 556-569.

HYVÄRINEN, A. (1999): Fast and robust fixed-point algorithms for independent component analysis. In: *IEEE Transactions on Neural networks* 10: 626-634.

HYVÄRINEN, A. (2013): Independent component analysis: recent advances.

In: *Philosophical Transactions of the Royal Society* A, 371: 1-19.

HYVÄRINEN, A. & Oja, E. (2000): Independent Component Analysis: Algorithms and Applications. In: *Neural Networks* 13: 411-430.

HYVÄRINEN, A., ZHANG, K., SHIMIZU, S. & HOYER, P. (2010): Estimation of a Structural Vector Autoregression Model Using Non-Gaussianity. In: *Journal of Machine Learning Research* 11: 1709-1731.

I |

IAMMARINO, S. & MCCANN, P. (2013): *Multinationals and economic geography: location, technology and innovation*. Edward Elgar Publishing: Cheltenham, UK.

IJIRI, Y. & SIMON, H. (1997): *Skew Distributions and the Sizes of Business Firms*. North Holland: Amsterdam.

J |

JAFFE, A. (1989): Real Effects of Academic Research. In: *The American Economic Review* 79: 957-970.

JARQUE, C. & BERA, A. (1980): Efficient test for normality, homoscedasticity and serial independence of residuals. In: *Economics Letters* 6: 255-259.

JEFFERSON, G., HU, A., GUAN, X. & YU, X. (2003): Ownership, performance, and innovation in China's large- and medium-sized industrial enterprise sector. In: *China Economic Review* 14: 89–113.

JOHANSSON, B., KLAESSON, J. & OLSSON, M. (2003): Commuters' non-linear response to time distances. In: *Journal of Geographical Systems* 5: 315–329.

JONES, C. (1997): In the Evolution of the World Income Distribution. In: *Journal of Economic Perspectives* 11: 19-36.

JONES, C. & ROMER, P. (2010): The New Kaldor Facts: Ideas, Institutions, Population, and Human Capital. In: *American Economic Journal, Macroeconomics* 2: 224-245.

JU, M. & ZHAO, H. (2009): Behind organizational slack and firm performance in China: The moderating roles of ownership and competitive industry. In: *Asia Pacific Journal of Management* 26: 701–717.

K |

KARLSSON, C. & MANDUCHI, A. (2001: Knowledge spillovers in a spatial context - a critical review and assessment. In: FISCHER, M. & FRÖHLICH, J. (eds): *Knowledge, Complexity and Innovation Systems*. Springer: Berlin, 101-123.

KETELS, C. (2013): Recent research on competitiveness and clusters: what are the implications for regional policy? In: *Cambridge Journal of Regions, Economy and Society* DOI: 10.1093/cjres/rst008.

KIRMAN, A. (2010): The economic crisis is a crisis for economic theory. In: *CESifo Economic Studies 56*: 498-535.

KLEPPER, S. (1996): Entry, Exit, Growth, and Innovation over the Product Life Cycle. In: *American Economic Review* 86: 562-583.

KLEPPER, S. (1997): Industry life cycles. In: *Industrial and corporate change* 6: 145-182.

KLEPPER, S. (2006): The evolution of geographic structure in new industries. In: *Revue de l'OFCE*: 135–158.

KLINE, P. & MORETTI, E. (2014): Local Economic Development, Agglomeration Economies, and the Big Push: 100 Years of Evidence from the Tennessee Valley Authority. In: *The Quarterly Journal of Economics* 129: 275-331.

KNOBEN, J., RASPE, O., ARIKAN, A. & OORT, F. (2011): Agglomerations and Firm Performance: one firm's medicine is another firm's poison. DRUID 2011 Conference Proceeding.

KNOPP, S., SANDERS, P., SCHULTES, D., SCHULZ, F. & WAGNER, D. (2007): Computing Many-to-Many Shortest Paths Using Highway Hierarchies. At:

Workshop on Algorithm Engineering and Experiments, January 2007.

KOENKER, R. & HALLOCK, K. (2001): Quantile Regression. In: *Journal of Economic Perspectives* 15: 143–156.

KOENKER, R. (2005): *Quantile Regression*. Cambridge University Press: New York.

KOGUT, B. & ZANDER, U. (1992): Knowledge of the firm, combinative capabilities, and the replication of technologies. In: *Organization Science* 3: 383–397

KOO, J. (2005): Agglomeration and spillovers in a simultaneous framework. In: *The Annals of Regional Science* 39: 35-47.

KRUGMAN, P. (1991): Increasing Returns and Economic Geography. In: *The Journal of Political Economy* 99: 483-399.

KRUSKAL, W. & WALLIS, W. (1952): Use of ranks in one-criterion variance analysis. In: *Journal of the American Statistical Association* 47: 583-621.

KUBIS, A., BRACHERT, M. & TITZE, M. (2009): Industrielle Cluster als Ursache regionaler Prosperität? Zur Konvergenz deutscher Arbeitsmarktregionen 1996-2005. Institut für Wirtschaftsforschung Halle (IWH), Halle.

KUERSTEINER, G. (2008): Granger-Sims causality. In: DURLAUF, S. & BLUME, L. (eds): *The New Palgrave Dictionary of Economics* (Vol. 2). Palgrave Macmillan. DOI:10.1057/9780230226203.0665.

L |

LACERDA, G., SPIRTES, P., RAMSEY J. & HOYER, P. (2008): Discovering cyclic causal models by independent components analysis. Proceedings of the 24th Conference on Uncertainty in Artificial Intelligence.

LAURSEN, K. & SALTER, A. (2004): Searching high and low: what types of firms use universities as a source of innovation? In: *Research Policy* 33: 1201-1205.

LAWSON, T. (2013): What is this 'school' called neoclassical economics? In: *Cambridge Journal of Economics* 37: 947-983.

LEE, Y., AMARAL, L., CANNING, D., MEYER, M. & STANLEY, E. (1998): Universal Features in the Growth Dynamics of Complex Organizations. In: *Physical Review Letters* 81: 3275–3278.

LEGLER, H. & FRIETSCH, R. (2006): Neuabgrenzung der Wissenswirtschaft –forschungsintensive Industrien und wissensintensive Dienstleistungen. In: BMBF (eds) *Studien zum deutschen Innovationssystem*, 22-2007.

LEVINE, R. (2005): Finance and growth: theory and evidence. In: AGHION, P. & DURLAUF, S. (eds): *Handbook of economic growth*, Vol. 1, North Holland: The Netherlands, 865–934.

LEVINE, R. & ZERVOS, S. (1993): What we have learned about policy and growth from cross-country regressions? In: *The American Economic Review* 83: 426-430.

LI, S. (2011): Regression models with the error term following the asymmetric exponential power distribution. Mimeo, Department of Economics, Rutgers University.

LI, H., MENG, L., WANG, Q. & ZHOU, L. (2008): Political connection, financing and firm performance: Evidence from Chinese private firms. In: *Journal of Development Economics* 87: 283–299.

LIAO, J., YE, Q., PISTRUI, D. & WELSCH, H. (2009): New venture financing: An empirical investigation of Chinese entrepreneurs. In: Li, J. & Hsu, S. (eds): *Informal Finance in China. American and Chinese Perspectives*. Oxford University Press: Oxford.

LIN, Z., Y. CHEN & X. QIN (2013): The effects of ownership types on enterprise innovation efficiency: Do industrial and regional heterogeneity matter? Management Science and Engineering International Conference IEEE: 833–844.

LORENZEN, M., MAURER, I. & STABER, U. (2012): Space and Inter-organizational Relations. In: *Industry and Innovation* 19: 181–186.

LOTTI, F., SANTARELLI, E. & VIVARELLI, M. (2009): Defending Gibrat's Law as a long run regularity. In: *Small Business Economics* 32:31-44.

LUCAS, R. (1978): On the size distribution of business firms. In: *The Bell Journal of Economics* 9: 508-523.

LUCAS, R. (1988): On the mechanics of economic development. In: *Journal of Monetary Economics* 1: 3-42.

LUNARDI, J., MICCICHE, S., LILLO, F., MANTEGNA, R. & GALLEGATI, M. (2014): Do firms share the same functional form of their growth rate distribution? A statistical test. In: *Journal of Economic Dynamics and Control* 39: 140-164.

LÜTKEPOHL, H. (2003): Vector Autoregression. In: BALTAGI, B. (edt): *A companion to theoretical econometrics*. Blackwell Publishing: Malden, MA, p. 678-699.

LYCHAGIN, S., PINKSE, J., SLADE, M. & VAN REENEN, J. (2010): Spillovers in Space: Does Geography Matter? NBER Working Paper 16188.

M |

MAASOUMI, E., RACINE, F. & STENGOS, T. (2007): Growth and convergence: A profile of distribution dynamics and mobility. In: *Journal of Econometrics* 136: 483–508.

MAGGIONI, M, NOSVELLI, M. & UBERTI, T. (2007): Space versus networks in the geography of innovation: A European analysis. In: *Papers in Regional Science* 86: 472–493.

MAINE, E., SHAPIRO, D. & VINING, A. (2010): The role of clustering in the growth of new technology-based firms. In: *Small Business Economics* 34: 127–146.

MALECKI, E. (1997): Technology and Economic Development: The Dynamics of Local, Regional, and National Change. Urbana-Champaign's Academy for Entrepreneurial Leadership, University of Illinois.

MALERBA, F. (2002): Sectoral systems of innovation and production. In: *Research Policy* 31: 247-264.

MALMBERG, A., MALMBERG, B. & LUNDQUIST, P. (2000): Agglomeration and firm performance: economies of scale, localisation, and urbanisation among Swedish export firms. In: *Environment and Planning* A 32: 305-321.

MANSFIELD, E. (1962): Entry, Gibrat's law, innovation, and the growth of firms. In: *The American Economic Review* 52: 1023-1051.

MANTEL, N. (1967): The detection of disease clustering and a generalized regression approach. In: *Cancer Research* 27: 209–220.

MARCON, E. & PUECH, F. (2012): A Typology of Distance-based Measures of Spatial Concentration. Working paper.

MARRIS, R. (1964): *The economic theory of managerial capitalism*. Macmillan: London.

MARSHALL, A. (1890): *Principles of Economics*. Macmillan: London.

MARTIN, R. (2001): Geography and public policy: The case of the missing agenda. In: *Progress in Human Geography* 25:189–210.

MARTIN, R. (2012): Regional economic resilience, hysteresis and recessionary shocks. In: *Journal of Economic Geography* 12: 1–32.

MARTIN, R. & SUNLEY, P. (2003): Deconstructing clusters: chaotic concept or policy panacea? In: *Journal of Economic Geography* 3: 5–35.

MARTIN, R. & SUNLEY, P. (2012): Forms of emergence and the evolution of economic landscapes. In: *Journal of Economic Behavior & Organization* 82: 338-351.

MARTIN, R. & SUNLEY, P. (2014): On the Notion of Regional Economic Resilience: Conceptualisation and Explanation. In: *Journal of Economic Geography*, Advanced Access.

MASKELL, P. (2001): The Firm in Economic Geography. In: *Economic Geography* 77: 329–344.

MATHEW, N. (2012): Drivers of Firm Growth: Micro-evidence from Indian Manufacturing. LEM Working Paper Series 2012/24.

MCCANN, P. & SHEFER, D. (2005): Agglomeration, economic geography and regional growth. In: *Papers in Regional Science* 84: 301-309.

McCann, P. & Folta, T. (2008): Location matters: where we have been and where we might go in agglomeration research. In: *Journal of Management* 34: 532–565.

McCann, P. & Folta, T. (2011): Performance differentials within geographic clusters. In: *Journal of Business Venturing* 26: 104-123.

McCann, P. & vanOort, F. (2009): Theories of agglomeration and regional economic growth: a historical review. In: Capello, R. & Nijkamp, P. (eds): *Handbook of Regional Growth and Development Theories*. Edward Elgar: Cheltenham, UK, p. 19-32.

McCann, P. & Ortega-Argilés, R. (2013): Smart specialization, regional growth and applications to European Union cohesion policy. In: *Regional Studies* DOI:10.1080/00343404.2013.799769

McGlade, J., Murray, R., Baldwin, J., Ridgway, K. & Winder, B. (2006): Industrial resilience and decline: A co-evolutionary approach. In: Garnsey, E. & McGlade, J. (eds): *Complexity and co-Evolution: Continuity and change in socio-economic systems*. Edward Elgar: Cheltenham, 147-176.

McKelvey, B. & Andriani, P. (2005): Why Gaussian statistics are mostly wrong for strategic organization. In: *Strategic Organization* 3: 219–228.

McKelvie, A. & Wiklund, J. (2010): Advancing Firm Growth Research: A Focus on Growth Mode Instead of Growth Rate. In: *Entrepreneurship Theory and Practice* 34: 261–288.

Melo, P., Graham, D. & Noland, R. (2008): A meta-analysis of estimates of agglomeration economies. In: *Regional Science and Urban Economics* 39: 332-342.

Metcalfe, J. (2001): Institutions and progress. In: *Industrial and Corporate Change* 10: 561–586.

Metcalfe, S. & Foster, J. (2010): Evolutionary growth theory. In: Setterfield, M. (edt): *Handbook of alternative theories of economic growth*. Edward Elgar: Cheltenham, UK, p. 64-94.

Metzig, C. & Gordon, M. (2014): A Model for Scaling in Firms' Size and Growth Rate Distribution. In: *Physica A* 398: 264-279.

Milana, C. & Wang, J. (2013): Fostering Entrepreneurship in China: A Survey of the Economic Literature. In: *Strategic Change* 22: 387–415.

Miller, H. (2010): The data avalanche is here. Shouldn't we be digging? In: *Journal of Regional Science* 50: 181-201.

Mineo, A. (2003): On the estimation of the structure parameter of a normal distribution of order p. In: *Statistica* anno LXIII: 109–122.

Mineo, A. & Ruggieri, M. (2005): A Software Tool for the Exponential Power Distribution: The normalp package. In: *Journal of Statistical Software* 12: 1–23.

Mohr, H. (2002): Räumliche Mobilität von Hochschulabsolventen. In: L. Bellmann & Velling, J. (eds): Arbeitsmärkte für Hochqualifizierte. IAB: Nürnberg, 249-281.

Moneta, A., Entner, D., Hoyer P. & Coad, A. (2013): Causal Inference by Independent Component Analysis: Theory and Application. In: *Oxford Bulletin of Economics and Statistics* 75: 705-730.

Mosteller, F. & Tukey, J. (1977): Data Analysis and Regression: A Second Course in Statistics. Reading.

Mullen, K., Ardia, D., Gil, D., Windover, D. & Cline, J. (2011): DEoptim: An R Package for Global Optimization by Differential Evolution. In: *Journal of Statistical Software* 40: 2–26.

Mundt, P., Milaković, M. & Alfarano, S. (2014): Gibrat's Law Redux: Think Profitability Instead of Growth. BERG Working Paper Series 92.

Myrdal, G. (1959): *Economic Theory and Under-developed Regions*. Gerald Duckworth & Co: London.

N |

Naranjo, L., Pérez, C. J. & Martín, J. (2014): Bayesian analysis of some models that use the asymmetric exponential power distribution. In: *Statistics and*

Computing DOI 10.1007/s11222-014-9449-1.

NEFFKE, F., HENNING, M. & BOSCHMA, R. (2011): How Do Regions Diversify over Time? Industry Relatedness and the Development of New Growth Paths in Regions. In: *Economic Geography* 87: 237–265.

NEFFKE, F., HENNING, M., BOSCHMA, R., LUNDQUIST, K. & OLANDER, L. (2011): The Dynamics of Agglomeration Externalities along the Life Cycle of Industries. In: *Regional Studies* 45: 49–65.

NEFFKE, F., Hartog, M., Boschma, R. & Henning, M. (2014): Agents of structural change. The role of firms and entrepreneurs in regional diversification. PEEG Working Paper Series 14.10.

NELSON, R. & WINTER, S. (1974): Neoclassical vs. evolutionary theories of economic growth: Critique and prospectus. In: *Economic Journal* 84: 886-905.

NELSON, R. & WINTER, S. (1978): Forces generating and limiting concentration under Schumpeterian competition. In: *The Bell Journal of Economics* 9: 524-548.

NELSON, R. & WINTER, S. (1982): *An Evolutionary Theory of Economic Change*. Harvard University Press: Cambridge, MA.

NELSON, R. (1998): The agenda for growth theory: a different point of view. In: *Cambridge Journal of Economics* 22: 497–520.

NICHOLS, A. (2007): Causal inference with observational data. In: *Stata Journal* 7: 507-541.

NIGHTINGALE, P. & COAD, A. (2014): Muppets and gazelles: political and methodological biases in entrepreneurship research. In: *Industrial and Corporate Change* 23: 113-143.

NOOTEBOOM, B. (2000): Learning by Interaction: Absorptive Capacity, Cognitive Distance and Governance. In: *Journal of Management and Governance* 4: 69-92.

O |

OECD (2001): *Measuring Productivity*. OECD: Paris.

OECD (2009): *How Regions Grow: Trends and Analysis*. OECD: Paris.

OECD (2010): *High-growth Enterprises: What Governments Can Do to Make a Difference*. OECD: Paris.

OECD (2011): *OECD Regional Outlook 2011: Building Resilient Regions for Stronger Recovery*. OECD: Paris.

OKUBO, T. & TOMIURA, E. (2014): Skew Productivity Distributions and Agglomeration: Evidence from Plant-Level Data. In: *Regional Studies* 48: 1514-1528.

OPENSHAW, S. (1984): *The modifiable areal unit problem*. Norwich.

OTTO, A. & FORNAHL, D. (2008): Long-term growth determinants of young businesses in Germany: Effects of regional concentration and specialisation, IAB-Discussion Paper 13/2008. Institut für Arbeitsmarkt- und Berufsforschung (IAB), Nürnberg.

OVERMAN, H. (2010): "GIS a job": What use geographical information systems in spatial economics. In: *Journal of Regional Science* 50: 165–180.

OTTAVIANO, G. (2011): 'New' new Economic Geography: firm heterogeneity and agglomeration economies. In: *Journal of Economic Geography* 11: 231-240.

P |

PACK, H. (1994): Endogenous Growth Theory: Intellectual Appeal and Empirical Shortcomings. In: *Journal of Economic Perspectives* 8: 55–72.

PARK, S. & LUO, Y. (2001): Guanxi and organizational dynamics: organizational networking in Chinese firms. In: *Strategic Management Journal* 22: 455–477.

PARTRIDGE, M. & RICKMAN, D. (2003): The waxing and waning of regional economies: the chicken-egg question of jobs versus people. In: *Journal of Urban Economics* 53: 76-97.

PAVITT, K. (1984): Sectoral patterns of technical change: Towards a taxonomy and a theory. In: *Research Policy* 13: 343–373.

PEARL, J. (2000): *Causality: models, Reasoning, and Inference*. Cambridge University Press: Cambridge.

PEARL, J. (2009): Causal inference in statistics: An overview. In: *Statistical Surveys* 3: 96-146.

PENG, M., TAN, J. & TONG, T. (2004): Ownership types and strategic groups in an emerging economy. In: *Journal of Management Studies* 41: 1105–1129.

PENG, M., D. WANG & Y. JIANG (2008): An institution-based view of international business strategy: A focus on emerging economies. In: *Journal of International Business Studies* 39: 920–936.

PENROSE, E. (1959): *The theory of the growth of the firm*. John Wiley and Sons: New York.

PERROUX, F. (1950): Economic space: theory and applications. In: *The Quarterly Journal of Economics 64*: 89-104.

PFEFFER, J. & SALANCIK, G. (2003): *The external control of organizations: A resource dependence perspective*. Stanford University Press: Stanford.

PINSKE, J. & SLADE, M. (2010): The Future of Spatial Econometrics. In: *Journal of Regional Science* 50: 103–117.

PIKE, A., DAWLEY, S. & TOMANEY, J. (2010): Resilience, adaptation and adaptability. In: *Cambridge Journal of Regions, Economy and Society* 3: 59–70.

PONCET, S., STEINGRESS, W. & VANDENBUSSCHE, H. (2010): Financial Constraints in China: Firm-level Evidence. In: *China Economic Review* 21: 411–422.

PONDS, R., VANOORT, F. & FRENKEN, K. (2007). The geographical and institutional proximity of research collaboration. In: *Papers in Regional Science 86*: 423-443.

PONDS, R., VANOORT, F. & FRENKEN, K. (2010): Innovation, spillovers and university-industry collaboration: an extended knowledge production function approach. In: *Journal of Economic Geography* 10: 231–255.

POPPER, K. (1973): *Objektive Erkenntnis*. Hoffmann und Campe: Hamburg.

POTTER, A. & WATTS, D. (2011): Evolutionary agglomeration theory: increasing returns, diminishing returns, and the industry life cycle. In: *Journal of Economic Geography* 11: 417–455.

PORTER, M. (1998): *On Competition*. Harvard Business School Press: Cambridge, MA.

PORTER, M. (2000): Location, Competition, and Economic Development: Local Clusters in a Global Economy. In: *Economic Development Quarterly* 14: 15-34.

PORTER, M. (2003a): The economic performance of regions. In: *Regional Studies 37:* 545-546.

PORTER, M. (2003b): Location, Clusters, and Company Strategy. In: CLARK, G., FELDMAN, G. & GERTLER, M. (eds): *The Oxford Handbook of Economic Geography*. Oxford University Press: Oxford, 253–274.

PYKETT, J. (2013): Neurocapitalism and the new neurons: using neuroeconoics, behavioural economics and picoeconomics for public policy. In: *Journal of Economic Geography* 13: 845-869.

Q |

QIN, D. & SONG, H. (2009): Sources of investment inefficiency: the case of fixed-asset investment in China. In: *Journal of Development Economics* 90: 94–105.

QUAH, D. (1996): Twin Peaks: Growth and Convergence in Models of Distribution Dynamics. In: *The Economic Journal* 106: 1045-1055.

QUAH, D. (1997): Empirics of Growth and Distribution: Stratification, Polarization, and Convergence Clubs. In: *Journal of Economic Growth* 2: 27-59.

R |

RALSTON, D., TERPSTRA-TONG, J., TERPSTRA, R., WANG, X. & EGRI, C. (2006): Today's State-Owned Enterprises of China: Are they dying dinosaurs or dynamic dynamos. In: *Strategic Management Journal* 27: 825–843.

RASPE, O. & VANOORT, F. (2008): Firm Growth and Localized Knowledge Externalities. In: *The Journal of Regional Analysis and Policy* 38: 110–116.

RASPE, O. & VANOORT, F. (2011): Growth of new firms and spatially bounded knowledge externalities. In: *The Annals of Regional Sciences* 46: 495–518.

REGGIANI, A., DE GRAAFF, T. & NIJKAMP, P. (2002): Resilience: An Evolutionary Approach to Spatial Economic Systems. In: *Networks and Spatial Economics* 2: 211–229.

REICHSTEIN, T., DAHL, M., EBERSBERGER, B. & JENSEN, M. (2010): The devil dwells in the tails. In: *Journal of Evolutionary Economics* 20: 219-231.

REICHSTEIN, T. & JENSEN, M. (2005): Firm size and firm growth rate distributions - The case of Denmark. In: *Industrial and Corporate Change* 14: 1145–1166.

REIERSOL, O. (1950): Identifiability of a Linear Relation between Variables Which Are Subject to Error. In: Econometrica 18: 375-389.

RENSKI, H. (2011): External economies of localization, urbanization and industrial diversity and new firm survival. In: *Papers in Regional Science* 90: 473-502.

RICE, P., VENABLES, A. & PATACCHINI, E. (2006): Spatial determinants of productivity: Analysis for the regions of Great Britain. In: *Regional Science and Urban Economics* 36: 727-752.

RICE, S. (2004): Evolutionary theory: mathematical and conceptual foundations. Sinauer: Sunderland, Massachusetts.

RIGBY, D. & BROWN, M. (2013): Who Benefits from Agglomeration? In: *Regional Studies* DOI: 10.1080/00343404.2012.753141.

RIGBY, D. & ESSLETZBICHLER, J. (2002): Agglomeration economies and productivity differences in US cities. In: *Journal of Economic Geography* 2: 407-432.

ROBERTS, M. & SETTERFIELD, M. (2010): Endogenous regional growth: a critical survey. In: SETTERFIELD, M. (edt): *Handbook of alternative theories of economic growth*. Edward Elgar: Cheltenham, 431-450.

RODRÍGUEZ-POSE, A. (1999). Innovation prone and innovation averse societies: Economic performance in Europe. In: *Growth and Change 30*: 75-105.

RODRÍGUEZ-POSE, A. (2010): Economic geographers and the limelight: institutions and policy in the World Development Report 2009. In: *Economic Geography 8*3: 361-370.

RODRÍGUEZ-POSE, A. (2011): Economists as geographers and geographers as something else: on the changing conception of distance in geography and economics. In: *Journal of Economic Geography* 11: 347–356.

RODRÍGUEZ-POSE, A. (2013): Do Institutions Matter for Regional Development. In: Regional Studies 47: 1034-1047.

RODRÍGUEZ-POSE, A. & CRESCENZI, R. (2008): Research and Development, Spillovers, Innovation Systems, and the Genesis of Regional Growth in Europe. In: *Regional Studies* 42: 51–67.

RODRIGUEZ-POSE, A. & GARCILAZO, E. (2013): *Quality of Government and the Returns of Investment: Examining the Impact of Cohesion Expenditure in European Regions*. OECD Publishing, Paris.

ROMER, P. (1986): Increasing Returns and Long-Run Growth. In: *The Journal of Political Economy* 94: 1002–1037.

ROMER, P. (1990): Endogenous Technological Change. In: *The Journal of Political Economy* 98: 71-102

ROMER, P. (2008): Economic Growth. In: *The Concise Encyclopedia of Economics.* Library of Economics and Liberty. http://www.econlib.org/library/Enc/Econ

omicGrowth.html (accessed: 3 March 2014).

ROSENTHAL, S. & STRANGE, W. (2001): The determinants of agglomeration. In: *Journal of Urban Economics 50*: 191-229.

ROSENTHAL, S. & STRANGE, W. (2003): Geography, industrial organization, and agglomeration. In: *The Review of Economics and Statistics* 85: 377-393.

S |

SANTARELLI, E., KLOMP, L. & THURIK, A. (2006): Gibrat's law: an overview of the empirical literature. In: *Entrepreneurship, growth, and innovation*. Springer: US, 41-73.

SAPIO, S. & THOMA, G. (2006): The growth of industrial sectors : theoritical insights and empirical evidence from US manufacturing. LEM working paper series 2006/09.

SCHLUMP, C. & BRENNER, T. (2010): University Education, Public Research and Employment Growth in Regions-An Empirical Study of Germany. In: *Ber. z. dt. Landeskunde* 84: 115-136.

SCHMOCH, U., LAVILLE, F., PATEL, P. & FRIETSCH, R. (2003): Linking Technology Areas to Industrial Sectors. DG Research: Final Report to the European Commission.

SCHOLL, T. & BRENNER, T. (2012): Detecting Spatial Clustering Using a Firm-Level Cluster Index. Marburg Working Papers on Innovation and Space No. 02.12.

SCHUBERT, C. (2013): How to evaluate creative destruction: reconstructing Schumpeter's approach. In: *Cambridge Journal of Economics* 37: 227–250.

SCHWAB, K. (2013): *The global competitiveness report, 2013-14 (full data edition)*. World Economic Forum.

SCHWARZKOPF, Y., AXTELL, R. & FARMER, J. (2010): The cause of universality in growth fluctuations. arXiv:*1004.5397*.

SETTERFIELD, M. (2010): Hysteresis. Trinity College Deparment of Economics Working Paper 10-04.

SHENKAR, O. & VON GLINOW, M. (1994): Paradoxes of organizational theory and research: Using the case of China to illustrate national contingency. In: *Management Science* 40: 56–71.

SHERMAN, H., LAMB, W. & ASPREGEN, K. (2009): The care and feeding of high-growth businesses in rural areas: the role of universities. In: VARGA, A. (edt): *Universities, Knowledge Transfer and Regional Development*. Edward Elgar: Cheltenham, 350-366

SHIMIZU, S. (2014): Estimation of data generating processes using SEM. – web reference: https://sites.google.com/site/sshimizu06/home/lingampapers (last visited: 28/08/2014).

SHIMIZU, S., HOYER, P., HYVÄRINEN, A. & KERMINEN, A. (2006): A Linear Non-Gaussian Acyclical Model for Causal Discovery. In: *Journal of Machine Learning Research* 7: 2003-2030.

SHIMIZU, S., INAZUMI, T., SOGAWA, Y., HYVÄRINEN, A., KAWAHARA, Y., WASHIO, T., HOYER, P. & BOLLEN, K. (2011): A direct method for estimating a causal ordering in a linear non-Gaussian acyclic model. In: *Journal of Machine Learning Research* 12: 1225–1248.

SIMMIE, J. & MARTIN, R. (2010): The economic resilience of regions: towards an evolutionary approach. In: *Cambridge Journal of Regions, Economy and Society* 3: 27–43.

SIMON, H. (1968): On judging the plausibility of theories. In: VAN ROOTSELAAR, B. & STAAHL, J. (eds): *Logic, Methodology and Philosophy of Sciences* 3. North-Holland: Amsterdam, 439–459.

SIMON, H. & BONINI, C. (1958): The size distribution of business firms. In: *The American Economic Review* 46: 607-617.

SIMS, C. (1980): Macroeconomics and reality. In: *Econometrica* 48: 1-48.

SINGH, J. (2005): Collaborative networks as determinants of knowledge diffusion patterns. In: *Management Science* 51: 756-770.

SMITH, K. (2002): What is the 'Knowledge Economy'? Knowledge Intensity and

Distributed Knowledge Bases. UNU/INTECH Discussion Papers 2002-6.

SMITH, K. (2005): Measuring Innovation. In: FAGERBERG, J., MOWERY, D. & NELSON, R. (eds): *The Oxford Handbook of Innovation*. Oxford University Press: Oxford, p. 148-177.

SONG, Z., STORESLETTEN, K. & ZILIBOTTI, F. (2011): Growing like China. In: *American Economic Review* 101: 196–233.

SONN, J. & STORPER, M. (2008): The increasing importance of geographical proximity in technological innovation: an analysis of US patent citations, 1975-1997. In: *Environment and Planning A* 40: 1020–1039.

SORENSEN, O. & AUDIA, P. (2000): The Social Structure of Entrepreneurial Activity: Geographic Concentration of Footwear Production in the United States, 1940-1989. In: *The American Journal of Sociology* 106: 424–262.

SORENSON, O., RIVKIN, J. & FLEMING, L. (2006): Complexity, networks and knowledge flow. In*: Research Policy* 25: 994-1017.

SPENCER, G., VINODRAI, T., GERTLER M. & WOLFE, D. (2010): Do Clusters Make a Difference? Defining and Assessing their Economic Performance. In: *Regional Studies* 44: 697–715.

SPIEZIA, V. & WEILER, S. (2007): Understanding Regional Growth. In: *The Review of Regional Studies* 37: 344-366.

STAN, C., PENG, M. & BRUTON, G. (2013): Slack and the performance of state-owned enterprises. In: *Asia Pacific Journal of Management* 31: 473-495.

STANLEY, M., AMARAL, L., BULDYREV, S., HAVLIN, S., LESCHHORN, H., MAASS, P., SALINGER, M., STANLEY, E. (1996): Scaling behaviour in the growth of companies. In: *Nature* 379: 804–806.

STEPHENS, M. (1974): EDF Statistics for Goodness of Fit and Some Comparisons. In: *Journal of the American Statistical Association* 69: 730-737.

STIGLITZ, J. (2011). Rethinking macroeconomics: what failed, and how to repair it. In: *Journal of the European Economic Association 9:* 591-645.

STOCK, J. & WATSON, M. (2001): Vector Autoregression. In: *The Journal of Economic Perspectives* 15: 101–115.

STOREY, D. (1994): *Understanding the Small Business Sector*. Routledge: London, UK.

STORPER, M. (2011): Why do regions develop and change? The challenge for geography and economics. In: *Journal of Economic Geography* 11: 333-346.

STORPER, M. & SCOTT, A. (2009): Rethinking human capital, creativity and urban growth. In: *Journal of Economic Geography* 9: 147-167.

STUMPF, M. & PORTER, M. (2012): Critical Truths About Power Laws. In: *Science* 335: 665–666.

SUN, S., TONG, W. & TONG, J. (2002): How does government ownership affect firm performance? Evidence from China's Privatization Experience. In: *Journal of Business Finance & Accounting* 29: 1–27.

SUTTON, J. (1997): Gibrat's Legacy. In: *Journal of Economic Literature* 35: 40–59.

T |

TALEB, N. (2012): *Antifragile: Things That Gain from Disorder*. Random House Digital: New York.

TEECE, D. J., PISANO, G. & SHUEN, A. (1997): Dynamic capabilities and strategic management. In: *Strategic Management Journal 18*: 509-533.

THORSEN I., UBOE, J. & NAVDAL, G. (1999): A Network Approach to Commuting. In: *Journal of Regional Science* 39: 73-101.

TOBLER, W. (1970): A computer movie simulating urban growth in the Detroit region. In: *Economic Geography 46:* 234-240.

TÖDTLING, F. & TRIPPL, M. (2005): One size fits all?: Towards a differentiated regional innovation policy approach. In: *Research Policy 34*: 1203-1219.

TORRE, A. (2008): On the role played by temporary geographical proximity in knowledge transmission. In: *Regional Studies* 42: 869-889.

U |

UNCTAD (2005): *World investment report 2005*. Geneva: United Nations Commission for Trade and Development.

V |

VANOORT, F. (2014): Unity in variety? Agglomeration economics beyond the specialization-diversity controversy. In: KARLSSON, C. & ANDERSSON, M. (eds): *Handbook of research methods and applications in economic geography*. Edward Elgar: Cheltenham, UK (forthcoming).

VANOORT, F., BURGER, M., KNOBEN, J. & RASPE, O. (2012): Multilevel approaches and the firm-agglomeration ambiguity in economic growth studies. In: *Journal of Economic Surveys* 26: 468–491.

VAN SOEST, D., GERKING, S. & VANOORT F. (2006): Spatial impacts of agglomeration externalities. In: *Journal of Regional Science* 46: 881–899.

VARGA, A. & PARAG, A. (2009): Academic knowledge transfers and the structure of internal research networks. Working paper IAREG 1.3d.

VARIAN, H. (2014): Big Data: New Tricks for Econometrics. In: *Journal of Economic Perspectives* 28: 3-28.

VEBLEN, T. (1900): The Preconceptions of Economic Science III. In: *Quarterly Journal of Economics* 14: 240-269.

VEGA, S. & ELHORST, J. (2013): Modelling regional labour market dynamics in space and time. In: *Papers in Regional Science* DOI: 10.1111/pirs.12018.

DE VRIES, J., NIJKAMP, P. & RIETVELD, P. (2009): Exponential or power distance-decay for commuting? An alternative specification. In: *Environment and Planning A* 41: 461–480.

W |

WALKER, B. & SALT, D. (2006): *Resilience thinking: sustaining ecosystems and people in a changing world*. IslandPress: Washington.

WEIBULL, J. (1980): On the numerical measurement of accessibility. In: *Environment and Planning A* 12: 53-67.

WENNERG, K. & LINDQVIST, G. (2010): The effect of clusters on the survival and performance of new firms. In: *Small Business Economics* 34: 221-241.

WERKER, C. & BRENNER, T. (2004): Empirical calibration of simulation models. MPI Jena: Papers on Economics and Evolution 0410.

WINTER, S. (1984): Schumpeterian competition in alternative technological regimes. In: *Journal of Economic Behavior & Organization* 5: 287-320.

WITT, U. (1987). *Individualistische Grundlagen der evolutorischen Ökonomik* (Vol. 47). Mohr Siebeck: Tübingen.

WITT, U. (2000): Changing cognitive frames – changing organizational forms: an entrepreneurial theory of organizational development. In: *Industrial and Corporate Change* 9: 733-755.

WITT, U. (2008): What is specific about evolutionary economics. In: *Journal of Evolutionary Economics* 18: 547–575.

WITT, U. (2013): The Future of Evolutionary Economics: Why Modalities Matter. MPI Jena: Papers on Economics and Evolution 1309.

WITT, U., BROEKEL, T. & BRENNER, T. (2012): Knowledge and its economic characteristics: a conceptual clarification. In: ARENA, R., FESTRÉ, A. & LAZARIC, N. (eds): *Handbook of Economics and Knowledge*. Edward Elgar: Cheltenham, 369-382.

WOODWARD, D. & GUIMARAES, P. (2009): Porter's cluster strategy and industrial targeting. In: GOETZ, S., DELLER, S. & HARRIS, T. (eds): *Targeting Regional Economic Development*. Routledge: New York, 68–83.

WORLD BANK (2009): *World Development Report 2009: Reshaping economic*

geography. World Bank: Washington, D.C.

X |

Xin, K. & Pearce, J. (1996): Guanxi: Connections As Substitutes for Formal Institutional Support. In: *Academy of Management Journal* 39: 1641-1658.

Z |

ZHU, D. & ZINDE-WALSH, V. (2009): Properties and estimation of asymmetric exponential power distribution. In: *Journal of Econometrics 148*: 86-99.

Appendix

Appendix X.1

Table X.1: Normality tests for $\tilde{g}_{r,i}$ – relative frequencies of significant rejections of the null hypothesis at 0.05 significance level

	Kolmogorov-Smirnov	Anderson-Darling	Jarque-Bera	Anscombe-Glynn
LMR-IAB 2-digit	100 %	100 %	100 %	100 %
LMR-IAB 3-digit	99.5 %	100 %	100 %	100 %
LMR-IAB 4-digit	99.5 %	99.5 %	100 %	100 %
LMR-Eckey 2-digit	98.3 %	98.3 %	100 %	100 %
LMR-Eckey 3-digit	99.0 %	99.5 %	100 %	100 %
LMR-Eckey 4-digit	98.3 %	99.1 %	100 %	100 %
Districts 2-digit	100 %	100 %	100 %	100 %
Districts 3-digit	100 %	100 %	100 %	100 %
Districts 4-digit	99.8 %	99.7 %	100 %	100 %

Appendix X.2

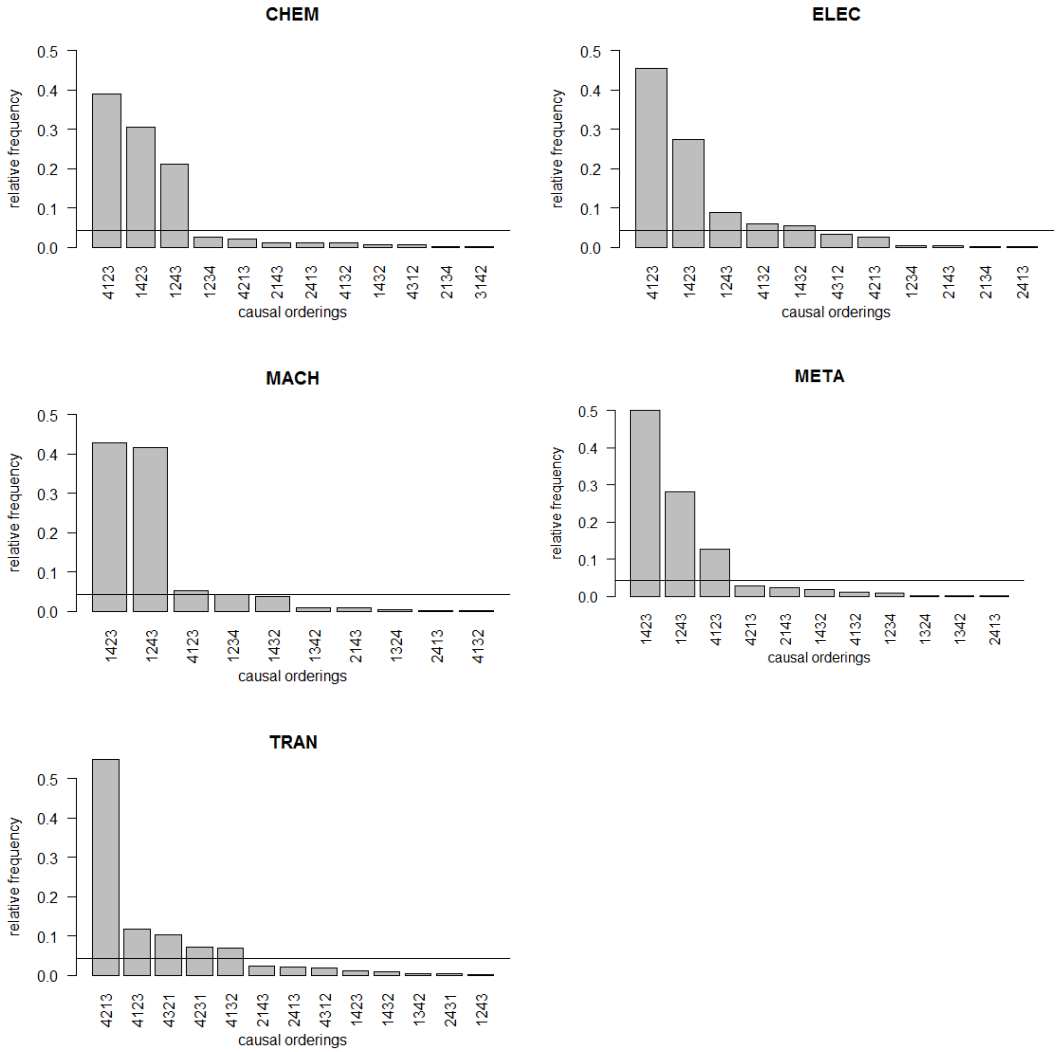


Figure X.2: Frequency distribution of alternative causal orderings resulting from the bootstrapping analysis (n=500). The number indicate the ordering of the variables, with 1=*Empl*, 2=*RD*, 3=*Pat* and 4=*Grad*.

Appendix X.3

	Descriptive statistics					Correlations (Pearson)											
	mean	sd	median	min	max	$(b_r + b_l)$	$(b_r - b_l)$	N_{firms}	*PopDensity*	*UnemplRate*	*RegGrowth*	*ResFunding*	*EmplUniv*	*Manufacturing*	*Construction*	*Similarity*	*RelVariety*
$(b_r + b_l)$	1.39	0.33	1.37	0.76	2.33												
$(b_r - b_l)$	0.19	0.46	0.28	-1.06	1.37	0.475											
N_{firms}	374.03	378.33	230.50	100.00	1763.00	-0.222	0.042										
PopDensity	293.90	257.03	208.42	51.22	1648.71	-0.084	0.136	0.477									
UnemplRate	0.13	0.06	0.11	0.04	0.34	-0.200	-0.021	-0.007	0.038								
RegGrowth	0.00	0.03	0.00	-0.08	0.09	-0.189	0.021	-0.064	-0.180	0.175							
ResFunding	34239.14	54319.46	12006	0.00	312414	-0.107	0.121	0.582	0.319	0.069	0.063						
EmplUniv	15719.25	24441.73	7251	1013.00	138038	-0.174	0.067	0.849	0.413	0.029	0.000	0.451					
Manufacturing	0.58	0.12	0.57	0.32	0.93	0.092	-0.108	-0.343	-0.138	-0.287	-0.331	-0.364	-0.440				
Construction	0.08	0.03	0.07	0.01	0.22	-0.097	-0.074	-0.224	-0.305	0.439	0.272	-0.149	-0.163	-0.405			
Similarity	0.14	0.00	0.13	0.13	0.16	0.244	-0.009	-0.563	-0.303	-0.234	0.126	-0.464	-0.482	0.346	0.069		
RelVariety	1.90	0.12	1.90	1.59	2.22	-0.110	0.017	0.470	0.304	0.139	-0.186	0.294	0.314	-0.039	-0.180	-0.695	
UnrelVariety	3.39	0.28	3.45	2.37	3.84	-0.252	0.064	0.327	0.171	0.627	0.281	0.412	0.402	-0.773	0.335	-0.598	0.441

Appendix X.4

Table X.3: Parameter ranges

Parameter	*Minimum*	*Maximum*	*Parameter*	*Minimum*	*Maximum*
$\mu_{d,max}$	.1	10	$\mu_{p,turn}$	0	.01
$\mu_{d,mean}$	-.002	.001	$\mu_{T,new}$	1	100
$\mu_{m,ex0}$	.00001	.002	$\mu_{exist,cap}$	.00001	.1
$\mu_{m,exa}$	.000001	.001	$\mu_{exist,s}$	.00001	1
$\mu_{c,init}$	.00001	1	$\mu_{m,new}$	.01	100
$\mu_{c,r}$	.2	1	$\mu_{start,new}$	1	1000
$\mu_{T,adapt}$	.0001	.1	$\mu_{com,start,new}$	.01	.3
$\mu_{inno,start}$	.001	10	$\mu_{new,turn}$	-.01	.01
$\mu_{red,inno}$	.99	1	$\mu_{new,size}$	.0000000001	.0001
$\mu_{move,0}$	.0001	.1	$\mu_{new,cap}$	.000001	1
$\mu_{move,a}$	.00001	.001	$\mu_{new,inno}$	.00001	1
$\mu_{move,o}$	.000005	.001	σ_d	0	.01
$\mu_{threshold}$	.0001	.1	σ_y	0	.1
$\mu_{com,start,exist}$	.001	.1	σ_c	.0001	.1

Eidesstattliche Erklärung

Matthias DUSCHL
Ketzerbach 45
35037 Marburg

An den
Vorsitzenden des
Promotionsprüfungsausschuss
Fachbereich Geographie
Philipps-Universität Marburg
Deutschhausstr. 10
35032 Marburg

Erklärung

(gemäß § 10, Abs. 1c der Promotionsordnung vom 15.07.2009)

Ich versichere wahrheitsgemäß, dass ich die vorliegende Dissertation selbst und ohne fremde Hilfe verfasst, keine anderen als die in ihr angegebenen Quellen oder Hilfsmittel benutzt sowie vollständig oder sinngemäß übernommenen Zitate als solche gekennzeichnet habe. Die Dissertation wurde in der vorliegenden oder einer ähnlichen Form noch bei keiner anderen in- oder ausländischen Hochschule eingereicht und hat noch keinen sonstigen Prüfungszwecken gedient.

Marburg, den 30.10.2014

Matthias Duschl