TECHNISCHE
UNIVERSITÄT
DARMSTADT

Urban Efficiency and Sectoral Structure

Empirical Results for German Cities

Dissertation

Approved as fulfilling the requirements for the degree of
doctor rerum politicarum (Dr. rer. pol.)
by the Department of Law and Economics of the
TU Darmstadt, Germany

by
Diplom-Volkswirt Stephan Hitzschke
(born in Jena, Germany)

Supervisor:	Prof. Dr. Jens Krüger
Second advisor:	Prof. Dr. Volker Nitsch
Filing date	05.12.2013
Date of defense:	12.02.2014
Place of publication:	Darmstadt
Year of publication:	2014

D 17

Acknowledgments

At this point I want to thank my supervisor, Prof. Dr. Jens Krüger, for his patient and helpful advice during all the years of creating this dissertation. He did not only teach me in applied econometrics and programming but also encouraged me to finish the dissertation. Secondly, I want to thank my second advisor, Prof. Dr. Volker Nitsch, for providing a survey on my dissertation. Additionally, I want to thank Prof. Dr. Volker Caspari for chairing the examination committee as well as the committee members Prof. Dr. Michael Neugart and Prof. Dr. Dirk Schiereck for their questions and insights during the defense.

Additionally, I especially want to thank my colleague and friend Dr. Andreas Dietrich for discussing drafts of the earliest chapters as well as Vanessa G. Belew for reading drafts and providing language assistance. Furthermore, I want to thank my colleagues at the chair of Empirical Economics at TU Darmstadt Dr. Nikolas Gatzke, Dr. Benjamin Hampf, Julian Hoss, Clemens Otte, Mathias Rhiel, Maike Schmitt and my colleague at the TU Darmstadt Johannes Rode for the inspiring and supporting talks during the last years.

An additional appreciation should be given to the participants of the second and third ifo Workshop on regional economics, the eighth Regional Innovation Policies Seminar, the ninth and tenth Irish Society of New Economists Conference, the twelfth International Conference of the Economic Society of Thessaloniki on Regional Economic Development, the fourteenth International Joseph A. Schumpeter Society Conference, the ninth Augustin Cournot Doctoral Days, the fifth RGS Doctoral Conference in Economics, the Workshop of the Center for Statistics at the TU Darmstadt, and the URBANgrad-Workshop on knowledge and development of cities, who discussed earlier versions of parts of the chapters with me. It was a great pleasure to present my drafts there which also was only possible by the support of my supervisor.

Last but not least, I want to thank my family and in particular my parents for not only providing financial but also and - even more important - personal support during all the years of education.

Contents

List of Tables

List of Figures

1 Introduction

1.1 Urban Development and Characterization

We are living in a new era. Since 2007, more than 50% of the world's population is living in urban areas (Nijkamp and Kourtit, 2013, p. 296). The process began around 3000 B.C., when people built a society around a market place and civilization started. Although the first cities were founded earlier, such as Jericho, often cited as the oldest city, civilization is closely connected to the urbanization process. The word "civilization" has its origins in the Latin word "civis", which describes a citizen or more precisely a townsman or townswoman (see Sullivan (2009, p. 71)). In the epoch of ancient Rom, the function of cities, which evolved from military camps, was to protect trade routes. In more modern times, towns provided security through massive walls and fortresses. In Germany, the settlements were granted their town charter by sovereigns. The town charter included the right to hold a local market, to have a city court and to build a fortification. In the Middle Ages, cities became even more attractive due to their increased power, which entailed the freedom of their citizens and the abolition of villeinage as well as the right to mint and issue coins. By providing these additional rights, cities became even more attractive and the urban population boomed. Hence, cities always provided market places, where not only goods but also the latest information were exchanged.

In the current time, however, cities do not provide additional rights to their citizens but cities still attract people which results in the ongoing urbanization process. Furthermore, cities did not emerge within a few years in developed countries (as we observe, for example in centrally planned countries such as the People's Republic of China, in the case of Ordos and its new district Kangbashi). German cities have a long history and are steadily developing which results in a path dependency process. Martin (2014) summarizes the most common sources of path dependency and its importance in regional development. Figures 1.1 and 1.2 show the development of the population living in urban area since 1950 and how it is expected to grow until 2050.[1] The figures show that the development of the share of population living in urban areas is increasing although not increasing steadily in the case of Germany as shown in figure 1.2. The share of population living in urban areas worldwide will catch up.

The catch-up, in terms of share of population living in urban area, is not only driven by higher growth rates of urban population in developing countries but also by the smaller increase of urban population share in developed countries like Germany. In the case of Germany, there even was a decline in the share of population living in urban areas as figure 1.2 shows. But the forecasts of the UN also predict a steady but small increase of the share of urban population on average.

[1]The data source is United Nations, Department of Economic and Social Affairs, Population Division (2012).

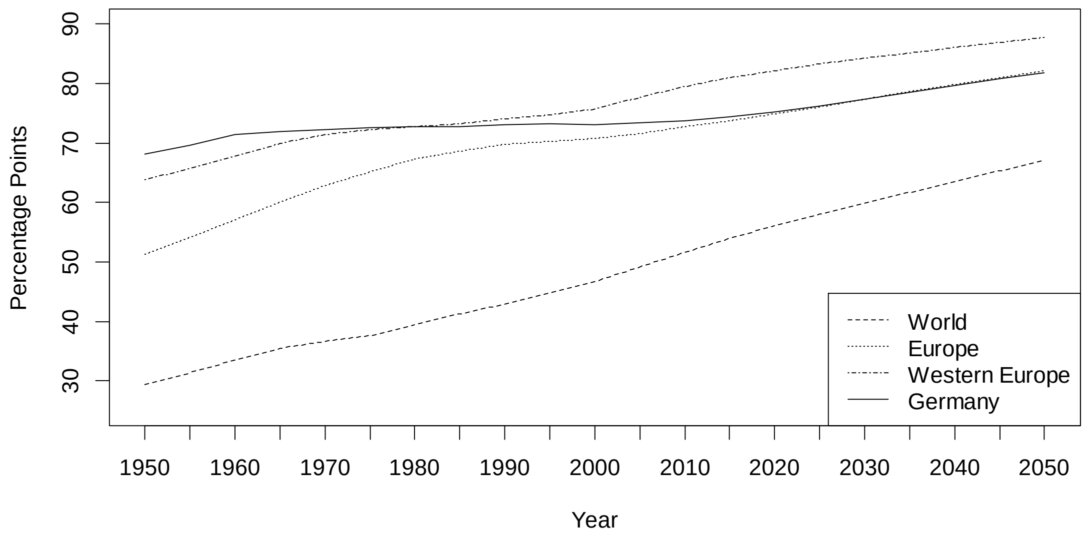

Figure 1.1: Percentage of Population Living in Urban Area

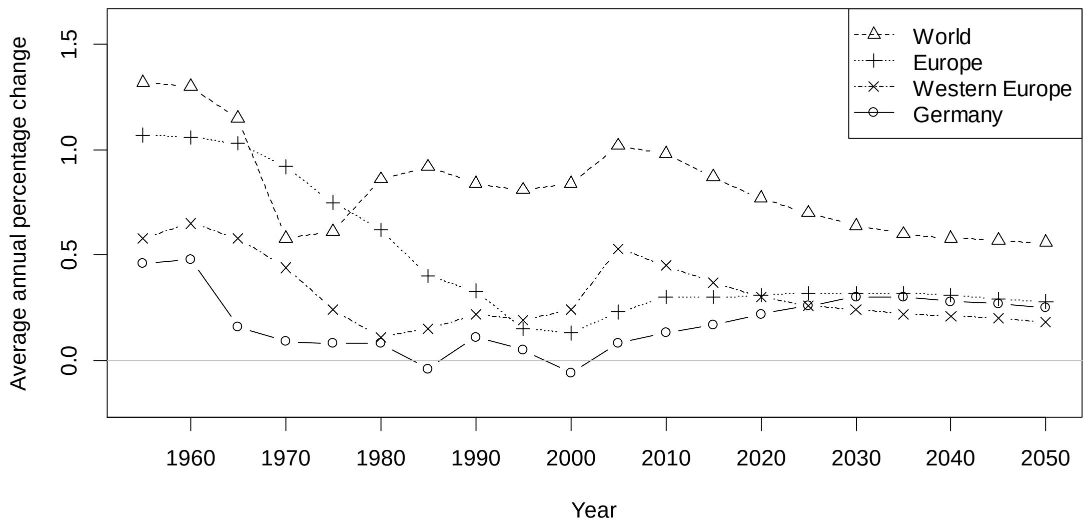

Figure 1.2: Average Annual Change of the Percentage of Urban Population

But why are urban areas or cities attracting a higher share of population? The additional utility that cities provide has changed from freedom for its citizens to economic forces. Recently, Shefer and Antonio (2013) investigated Israeli cities and found that income within the city is higher than in the surrounding area. This relationship is found for many countries and refers to urban externalities (Nijkamp and Kourtit, 2013). Additional reasons are manifold as given in e.g., Morrison (2014). They include well-being as a measurement of quality of life within a region and the supply of public goods such as infrastructure, cultural and higher educational facilities. Firms within a city are able to increase productivity by benefiting from urban externalities which additionally attracts other firms by specialization or diversification of economic sectors.

The presence of work lures people into the cities. Cities become more important, due to the increase in the number of people living in cities, which results in a larger market for new products and services.

Bettencourt et al. (2007) find that GDP is growing faster than city size. This rule is sometimes referred to as the "15% rule", as the GDP and other urban characteristics are additionally increasing by 15% as city size doubles.[2] In addition, cities need 15% less road surface and electricity supply as city size doubles. The 15% rule is therefore turned into the statement that cities become 15% more efficient as city size doubles.

In their recently published article, Nijkamp and Kourtit (2013) claim the urbanized Europe and the urbanized world and when the urbanization process moves on as expected, the role of cities will further increase, especially of mega-cities, as globally trading actors. They predict that cities will directly compete against each other and should therefore have to fulfill specific functions, such as providing innovations, connections, cultural diversity and livable area. That will have consequences on the sectoral composition within those specialized cities. The sectoral composition is therefore important to understand the function of cities.

Structural change, as the transition from the manufacturing sector to the service sector, already increases the demand of higher skills as stated by Mellander and Florida (2014). The new services sector operates on the basis of knowledge, innovation and skills. It is not only the presence of job opportunities that attracts people to move into the cities, but also education opportunities. In Germany, large universities play a dominant role in higher education by the creation of human capital, knowledge and innovation which are local in the cities. Thus, cities are the major source of high skilled employees and innovation, which are key drivers of economic growth and the structural change with technological progress. Knowledge and innovation are also identified as the reason for the 15% rule in Bettencourt et al. (2007).

Although urban population is about 70% of the total population, this dissertation mainly focuses on large urban areas like bigger cities and metropolitan areas which are statistically defined in NUTS-3 classification as free or independent German cities. On the one hand, an average of 26 million people lived in the free cities in Germany within the period between 1995 and 2011, which accounts for 32% of the total population. On the other hand, these free cities account for only 4.5% of the total area in Germany. Both figures result in different population densities. While there are on average about 229 people living per square kilometer in Germany as a whole, there are about 1622 people living per square kilometer in the cities. The closeness is highly correlated with innovation. As Carlino et al. (2007) show, the employment size and employment density positively affect patent intensity in US metropolitan areas. They estimate that doubling the employment density leads to a 20% higher number of inventions per person within the city, all other things being equal. The physical proximity within cities leads to more possibilities for interacting with other people. Furthermore, the result is in accordance with the 15% rule of Bettencourt et al. (2007), as innovation per capita is about 15% higher in cities with twice inhabitants. The question therefore is: Does everything follow the simple 15% rule? Carlino et al. (2007) additionally estimate the effects with quadratic terms and show that the quadratic terms were also significant, leading to inverse U-shaped development for patent intensity, employment size and employment densities. Therefore, the development is not simply linear as the 15% rule implies. There are non-linear relationships that have to be considered in urban investigations.

The geography of innovations is complex and interactions of innovations have to be considered (Malec-

[2]E.g., in The Economist article, "The law of the Cities", in the printed version of June 23rd 2012 [online: http://www.economist.com/node/21557313].

ki, 2014). Innovations and the subsequent technological progress are therefore not only an isolated explanatory variable for urban development, but also by their interaction with other urban characteristics. Figure 1.3 shows the development of the share of the gross value added (GVA) produced by each economic sector within the cities as well as the population share, compared to total Germany.

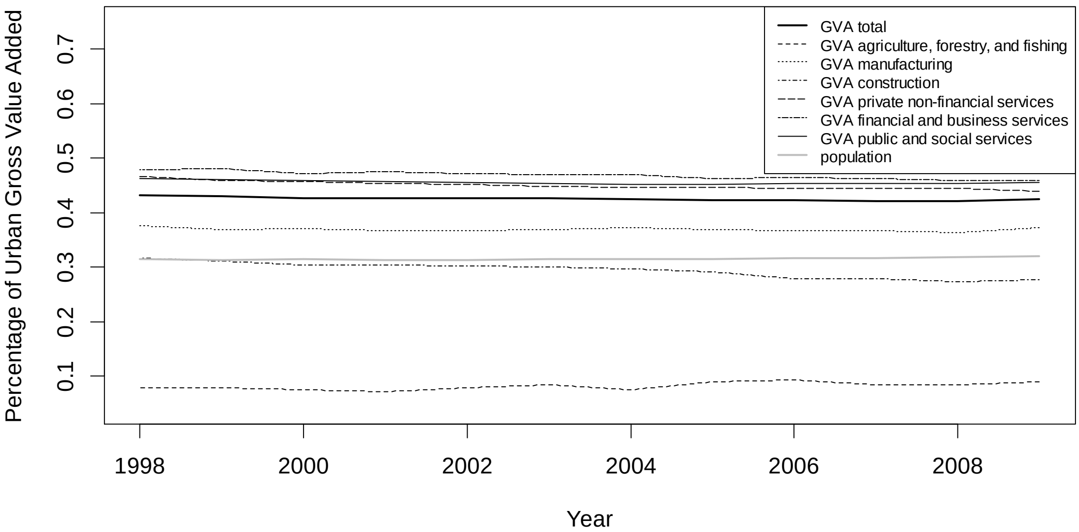

Figure 1.3: Urban Shares of Gross Value Added and Population

The figure demonstrates that the free cities are not the location of firms which produce a large share of the agricultural, fishing and forestry sector. Cities only account for 10% of the production in that sector. Furthermore, the share of gross value added produced within the construction sector has slightly declined over the past 11 years, from around 30% in 2000, which is about the same size as the population share of the cities. In addition, all other sectors have higher shares of gross value added compared to the population share, which indicates that these sectors are predominantly located within cities. Therefore, it is important to understand the development of the sectors as part of the development of cities. The main conclusion of urban economic literature is that cities provide additional value. This additional value results in higher productivity and efficiency in cities reasoning higher wages. And as Mameli et al. (2014) point out, the sectoral mix of the urban economy causes regionally different demands in employment. Furthermore, they emphasize the important role of regional productivity changes for explaining regional employment changes.

Furthermore, differences between cities are not only explained by size and economic characteristics but also by geographical location. Nijkamp and Kourtit (2013) emphasize the proximity externalities as another source of externalities in cities. Spatial inequality between cities such as in the case of Israeli cities (Shefer and Antonio, 2013) is investigated. Additionally, they show that spatial proximity measured by geographical distance is, by itself, an important factor to explain regional differences. Differences between cities have to be considered and modeled. Spatial differences within one country cause prices to be on average higher in urban areas than in rural areas. The higher income in cities, as found in e.g., Shefer and Antonio (2013), is therefore offset by higher prices within cities which results in similar real incomes across the country, as shown in BBSR (2009) for Germany. Figure 1.4 shows the densities of price indexes in rural and urban areas in Germany with the bandwidth calculated by Silverman's rule of thumb (Silverman, 1986, p. 48).

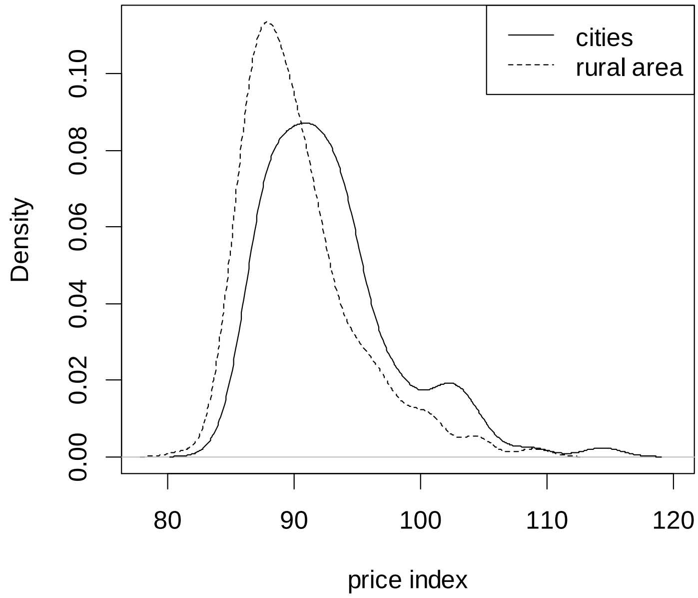

Figure 1.4: Densities of Price Indexes in Urban and Rural Area in Germany

Figure 1.4 shows, that price levels are higher in urban areas than in rural areas. This does not only hold true for absolute price levels, since the most expensive areas are urban areas, but also for average price levels. Comparing price indexes, the average price index in urban areas is 92.84 whereas it is 90.38 in rural areas (see chapter five for the discussion of the data and the analysis for cities). For rational actors, the decision to move into a city is a question of maximizing utility. Do the profits of positive urban externalities outweigh the losses of negative externalities? The question is similar for people deciding where they should live. Is it better to live within the city where prices are high or in the surrounding rural area with lower prices but higher costs of commuting? The ongoing growth of the share of population living in urban areas indicates that these areas compensate the higher prices by generated additional utility. Although people do not always act rational, see Kahneman (2012) for examples.

1.2 The Role of Productivity and Efficiency

Productivity is a key variable for understanding the booming development of urbanization. As already mentioned, it is fueled by many different externalities caused by, e.g., agglomeration, specialization, diversification and proximity. The schools of thoughts explicitly accounting for productivity in economic geographical theory are summarized in McCann (2014) with reference to the evolutionary and institutional economics. Productivity change plays an important role in economic development not only within neoclassical endogenous growth models but also within Schumpeterian evolutionary economic models.

But what is productivity? Productivity is a performance measure, which compares the generated output with the engaged inputs. Productivity can therefore be measured by the ratio of produced output divided by the used inputs. If all outputs and inputs of the production process are considered, the productivity measurement is called total productivity (Coelli et al., 2005, p. 3). The productivity measurement is called partial productivity if the output variable is only related to one input factor. Partial measures of productivity disregard the effect of the other input variables which might have been substituted and gives an insufficient performance measure. For example, a commonly employed productivity measure is the labor productivity of firms in a certain area within a specified time period. The gross value added is thereby used as the output measurement divided by engaged employment which was needed to produce this output within the period. To measure productivity within regions, Abreu (2014) points out that gross value added for regions is the counterpart to the gross domestic product of nations.

On the other hand, efficiency is the comparison of the observed output to an optimal amount of outputs for a given set of inputs. Alternatively it can be measured by the comparison of the observed quantity of inputs compared with the optimal quantity of inputs needed to produce a certain amount of outputs. Of course, the measurement of efficiency not only accounts for the quantities of outputs and inputs but also for the actual production technology. The input and output variables are the same as for the productivity measurement. Since efficiency measurements are technically the comparison of different productivities, the production frontier has to be estimated, which is set by the technology (Coelli et al., 2005, p. 3). A unit is described as technically efficient if that unit performs according to the production function that represents the actual best practice transformation of inputs to outputs. Meanwhile a unit's production is technically inefficient if it produces less output for a given set of inputs or needs more inputs to produce the same amount of outputs as more efficient units. To estimate the production frontier and to gain the resulting efficiency scores of the units, many different approaches are available: the parametric, the non-parametric and, as a mixture of both, the semi-parametric approaches.

Parametric approaches imply a specific assumed functional form for the production function in which the parameters have to be specified. These production functions are for instance the Cobb-Douglas or translog model (see Greene, 2008b, pp. 97ff.). These models consist of specific parameters, which are the factor elasticities and the total productivity in case of the Cobb-Douglas production function. These parameters have to be estimated. The estimation could be performed by ordinary least squares (OLS) which would result in a production function which does not envelop all observations since the error term of the OLS estimation could also be positive. Positive residuals would imply that the units are producing more than possible with the production function. One way to overcome that problem of not enveloping all units is to adjust the constant term, which is the total productivity, until all observations are enveloped, which is then called a corrected OLS constant. As Greene (2008b) points out, there are additional possibilities for estimating the enveloping production function such as the modified OLS estimation or by computing linear and quadratic programming of the residuals subject to the assumed production function. It results in non-negative residuals in a equation with negatively added residuals as proposed by Aigner and Chu (1968) (see Greene, 2008b, p. 108). However, the problem here is that the residuals, which are the measure of inefficiency, include a stochastic variation and random variation of the frontier across the units which are not separated (see Greene, 2008b, p. 114). This problem can be solved by estimating stochastic frontier models by Bayesian methods or method of moments estimations. As a result the parametric production function is estimated and the residuals indicate the inefficiency of the units.

The strict assumption of a specific production function is relaxed by the semi-parametric approach as used in e.g., Koop et al. (1994), in which the production function is approximated by polynomials in the logarithm of the output. As in the case of parametric approaches the error term has to be divided into a stochastic term and an inefficiency term which can be solved by stochastic frontier models. The estimation of these models has specific requirements of the available data structure such as a panel structure, which includes different units, observed over many time periods.

The non-parametric approaches go back to the seminal paper by Farrell (1957). The efficiency measure is derived by building a convex hull of the input-output vectors. This convex hull gives the production possibility and does not need any predefined parametric or semi-parametric model. Therefore, no parameter is needed to construct the production frontier. Farrell (1957) analyzed a production process with a single input and a single output. This measurement was generalized by Charnes et al. (1978) in the case of multiple inputs and multiple outputs, in which case linear programming is needed. The estimation procedure was then named as data envelopment analysis (DEA) by Charnes et al. (1981). An introduction to DEA is given in Coelli et al. (2005) and Cantner et al. (2007). Furthermore, productivity change can be estimated by the non-parametric approach. The coefficient for productivity change is called the Malmquist index and defined in Caves et al. (1982). Productivity change is derived by constructing the non-parametric production frontier for two different periods and calculating the resulting estimated efficiency scores. The productivity change is fueled by technological change, which moves the production frontier, and efficiency change of the observed units, which leads to a relative movement of the input-output combination below the production frontier. Färe et al. (1994) show how the Malmquist index can be decomposed to technological change and efficiency change. Productivity change is driven by the creation of knowledge and the implementation of innovations within the production process of firms. Thus, knowledge and innovation are crucial to explain efficiency and productivity change.

The diffusion of innovation and knowledge externalities results in efficiency externalities and innovation externalities, as explained in Johansson (2014). Thus, economic development is caused by an attractive environment, which causes path dependencies (see Karlsson and Grasjö (2014) for a recent literature review). Regional path dependency is also emphasized in Bathelt and Li (2014), who connect urban path dependency to the cluster life cycle. Crescenzi (2014) demonstrates the modern approach of innovation geography in which technological change does not result from a linear model but a more interacting model which also accounts for local innovation policies and non-spatial proximity. This non-spatial proximity corresponds to the hierarchical clusters in my investigations. Finally, van Oort and Lambooy (2014) provide an overview of knowledge production and diffusion and innovation in cities leading to urban growth. They emphasize that knowledge is not only context-specific but also person-specific, which causes knowledge spillovers to be localized. Therefore, analyses must focus on the local knowledge base and the innovation within the cities and allow for interactions between urban characteristics.

1.3 Research Questions and Outline

This dissertation investigates different time-variable criteria of German cities within the period between 1998 and 2007. By mainly focusing on the effects of economic performance of the manufacturing and service industries the empirical analyses are based on a data set with 112 larger cities which are classified as free cities. Efficiency and its improvement is a major economic goal since it either maximizes output for a given set of inputs or minimizes the use of inputs for a given output quantity. Thus, urban analysis should consider efficiency and its development. In addition, cities are not separate and closed economies, but instead are interacting parts within a country with specific functions as, e.g., financial centers or regional administrative centers. As such, cities are part of an urban hierarchy which has to be considered too. A special case of regional disparities arises in Germany caused by reunification in recent history and the administrative organizational structure within federal states.

Therefore, the main questions for this dissertation are listed in the order in which they are answered:

- What is the effect of city size on industrial efficiency?
- Is there an optimal city size of the free German cities with respect to urban efficiency?

- What are the consequences of the urban hierarch for the free German cities?
- What is the effect of industrial efficiency on urban price levels?
- Is there a difference between East and West Germany with respect to urban price levels, which is not explained by economic and social factors?
- What are the effects of technical progress on the development of value added and employment in German cities?
- Does technical progress lead to an increase of economic output or to a decrease of employment?

In addition, the consequences of the above issues for urban policy are considered. Analyses of complex regional development require complex modeling as claimed by Mellander and Florida (2014). The assumptions of the OLS model are often violated by the data, causing biased estimates and erroneous test results. Therefore, the main data set in which the subsequent analyses are rooted is explained first. Then the estimation approach of the main variable of interest, which is the economic and sectoral efficiency of the cities, is introduced. This approach involves non-parametric estimation, as already mentioned and justified in section 1.2. Furthermore, in the third subsection of the data set explaining section, the concept of urban hierarchy is introduced. A cluster analysis is performed for grouping the cities into hierarchical clusters. To overcome the drawbacks of the OLS model, advanced estimation methods can be used. Chapter two presents in detail the estimation methodologies, which are applied in the course of the dissertation. The estimation methods solve the problems of heterogeneity observed within the cities and are explained in the order in which they appear in subsequent chapters. First, robust regressions are introduced, which account for outliers without excluding the information of these observations. Second, an introduction to spatial data analysis is given with a sequential testing procedure in order to derive an estimation model which can explain most of the variation of the considered endogenous variable based on a low number of additional spatial parameters. Third, multilevel analysis is explained, which is the most complex analysis in this dissertation and has stringent requirements on the observed data structure.

Chapters three and four explain the size of cities and find an optimal city size determined by the efficient use of the production factors, capital and labor, to gain value added in the different sectors located in each city. The main difference between these two chapters is the measurement of efficiency, which results in different optimal city sizes. Chapter three, which is titled "The Optimal Size of German Cities: An Efficiency Analysis Perspective" investigates the interrelation between productive efficiency and population size for German cities.[3] The productive efficiency in this context is the scale efficiency, which is a result of positive and negative agglomeration externalities. The investigation is performed in a process which simultaneously estimates efficiency in terms of scale efficiency in a non-parametric setup and its relation to population size. It transpires that the optimal city size in Germany is about 220,000 inhabitants which is close to the mean city size. In addition, it is found that the optimal city size differs in East and West Germany and that it is close to the size of the mean city size within the region of investigation. The finding that optimal city size is the mean city size of its group holds true for geographically separated estimations. Furthermore, urban hierarchy plays a role for city size, which results in different city sizes for different city types.

Chapter four has the title "Technical Efficiency and (Optimal) City Size" and investigates the relationship between technical efficiency and city size for German cities. Technical efficiency is measured by a robust non-parametric approach for input-orientation, output-orientation and a combination of both. It

[3] This chapter as well as parts of chaper two are based on the working paper Hitzschke (2011) which is a previous version of this analysis.

turns out that smaller cities are more efficient in input direction and less in output direction, whereas the pattern for larger cities is reverse. Medium-sized cities are least efficient when efficiency is measured by a combination. It points towards a specific set of problems for medium-sized cities. As city size increases, congestion increases input wastage but also improves the utilization of scale economies in the output direction.

Chapter five takes the city size, efficiency, and urban hierarchy to explain price level differences between the cities. The title of chapter five is "Price Level Differences between German Cities: A Spatial Autoregressive Investigation". Within a single country the monetary aggregate and currency is the same for all inhabitants, but the law of one price does not hold. This chapter estimates price level differences observed between large cities in Germany. Prices should converge as a result of higher productivity growth in low price/low productivity areas. Since price level data at the required level of detail are only available for the year 2006, the static analysis prevents one from investigating convergence in price levels across Germany. The model includes price determination rules as well as productivity differences, which account for wage differences and thereby induce price level differences. The estimation model accounts for the presence of spatial dependencies and geographical heterogeneity in the number of neighboring cities within a radius of 100 km (about 62 miles). The chapter shows that prices are higher in cities with higher productivity and a higher share of the service industry as compared to the manufacturing industry. This is characteristic for cities of higher hierarchical order. Although the period of investigation starts 16 years after German Reunification, some analyses find significant differences between the price levels in East and West Germany. However, the analysis in chapter five does not find any geographically discriminating effects. Price level differences are already explained by economic characteristics such as income and population in the free cities.

From the static analyses in the previous chapters, chapter six takes a dynamic look on the development of the industries within the cities. Chapter six has the title "Industrial Growth and Productivity Change in German Cities: A Multilevel Investigation" and analyzes the role of productivity change and city-specific characteristics on economic growth within the cities. Productivity change is measured by the Malmquist index and its components, which are estimated by non-parametric data envelopment analysis. The dynamic analysis also incorporates the change of efficiency and technology which is the result of technological progress. Within an evolutionary economic geographical theory, the co-evolution of employment and value added growth of the industries is investigated. The nested structure as well as the interaction between industries within cities and over time is accounted for by estimating multilevel models. It is shown that there are differences for industrial growth for different cities and years. Therefore, the use of multilevel models is required. Schumpeter's creative destruction is found to hold for efficiency change on industrial growth. Efficiency change measures the catching-up to the best practice production function, reducing both value added growth and employment growth. Technological progress shifts the best practice production function and leads only to a rise in value added growth and not in employment growth. The estimations indicate a converging growth of urban industrial value added while employment growth diverges. As in the previous analyses, the regressions in chapter six account for urban hierarchy, city size, and east-west differences.

Chapter seven concludes, summarizes the main findings and connects the chapters. Furthermore, prospects for future research are given.

2 Estimation and Measurement Approaches

In the empirical analyses of the following chapters the baseline model is the ordinary least squares (OLS) regression model. The advantage of OLS estimation is that it is easily applied, well known throughout in economic science and gives first advices on the correlation between various variables. In many cases OLS estimations are fine because there is no violation of the assumed properties caused by the utilized data. In other cases, however, the OLS estimates are biased or wrong estimations of the variance of the coefficients cause wrong tests of significance (Greene, 2008a, pp. 148ff.). Since the violations of the OLS assumptions is caused by the data, the general data set as well as the measurement and classification approaches will be presented in the first part of this chapter. The three sections in the second part of the chapter contain estimation approaches that can be applied in these cases.

The first approach will be the robust regression analysis which accounts for data outliers. Data outliers might cause changing variance over the observation, which is called "heteroscedasticity" and it results in wrong variance estimates and thus in wrong tests of significance within the OLS results. Outliers might also cause biased estimates if the outliers do not follow the same linear function as the other observations.

Furthermore, the OLS model assumes that there is no correlation between the error terms which is called absence of autocorrelation (Greene, 2008a, pp. 16f.). Autocorrelation is in general only observable in time series since time structures the observation to a specific order whereas there is no meaningful criterion to structure cross-section data sets. However, also for cross-section analyses autocorrelation might be present within spatial data sets due to the effects of neighboring observation since the geographical location orders the observations within space. Because autocorrelation causes wrong variance estimations and thus wrong tests of significance, the estimation should correct for autocorrelation (Anselin, 2003a). Therefore, the second approach is the spatial model analysis which will be presented in section 2.3.

Additionally, omitted variables cause biased OLS estimates and wrong variance (Greene, 2008a, pp. 133f.). The omission of variables is often due to the fact that not every variable is measurable and also by keeping the models as simple as possible. A model is by definition always a simplification of the complex real world. In econometrics, the error term should include the random variation caused by omitted variables. However, the omitted variables are sometimes correlated with the variables which induce the estimates to be biased and the variances to be wrong. The effects of omitted variables are reduced by approaches which need a specific structure of the data like a panel data set or a multilevel model. The multilevel model approach will be presented in section 2.4.

2.1 General Data

The empirical analyses in the following four chapters root on a core data set that covers all 112 urban districts by NUTS3 classification 2009, although each chapter uses the data set in a different way and with additional variables, which will be explained within each chapter.

The 112 NUTS3-districts are those which are classified as free city districts (so-called 'kreisfreie Städte' or 'Stadtkreise' in Germany).[4] These cities are characterized by an independent local government that determines local environmental variables within the highly restricted legislative framework, like local tax structure and expenditure on local public issues. The time period for which data are available is 1998 until 2007. The data are taken from the regional database of the Statistical Offices of Germany[5] ("Statistische Ämter des Bundes und der Länder") and the INKAR database of the Federal Agency of

[4]A list of the included cities is given in the Appendix A1.

[5]The database is available on the Internet at https://www.regionalstatistik.de/genesis/online/logon (last check on 30th May 2011).

Building and Urban Development[6] ("Bundesamt für Bauwesen und Raumordnung"). It is a balanced panel, so for all cities the number of employees and the value added is known for each sector in every year. The sectors are defined at a one-digit industry specification (WZ 2003 of the Federal Statistical Office of Germany (Statistisches Bundesamt (2003)) which is a level of aggregation equivalent to the European wide classification NACE Rev. 1.1):

AB	agriculture, forestry, and fishing
CDE	wide manufacturing (including mining/quarrying, energy and water supply)
D	core manufacturing
F	construction
GHI	private non-financial services
JK	financial and business services (finance, insurance, and real estate)
LMNOP	public and social services

Figure 1.3 in the introduction shows that the free cities account for only less than 10 percent of total gross value added produced in the agriculture, forestry and fishing sector (AB) in Germany. As a result of the minor importance of this sector in German cities, the sector AB has been omitted.

For estimating an urban production function the capital stock has to be added as input factor, since Moomaw (1981) made the criticism that the disregard of the capital stock in Sveikauskas (1975) leads to biased estimates of an urban production function. The capital stock for each city and the wide manufacturing sector is computed with the perpetual inventory method (Park, 1995) supposing capital stocks $cap_{j,t}$ develop as

$$cap_{j,t} = (1-d)cap_{j,t-1} + inv_{j,t}, \tag{1}$$

with d the constant depreciation rate and $inv_{j,t}$ the city-specific investments in the wide manufacturing sector for each city j at time t. Furthermore, if investments change with constant growth rates $g_{inv,j}$, the starting capital stock at time $t = 0$, can be calculated as

$$cap_{j,0} = inv_{j,0} \cdot \frac{1 - g_{inv,j}}{d + g_{inv,j}}. \tag{2}$$

Eq. (2) is the result of the capital accumulation with investments growing at a constant rate and therefore leading to an infinite geometrical series.

The data of investments in the wide manufacturing sector are also taken from the regional database of the Statistical Offices in Germany for the time period 1995 to 2007 in real units and are given without the energy and water supply industry. The starting capital stock is estimated for 1995. The average annual depreciation rate is set to 10 percent per annum ($d = 0$), which is quite high but results in positive capital estimation caused by massive changes in investments in the first period of observation. The average growth rates of investments are calculated by the development of investment figures. Unfortunately, for some cities (Cottbus, Potsdam, and Stralsund) the growth rates of investment were shrinking by more than 10 percent, caused by immense changes after German Reunification and the associated structural changes in industry. Therefore, the average growth rates for all cities in East Germany were applied,

[6]The database is available on CD-ROM upon request from the Federal Agency of Building and Urban Development at http://www.bbsr.bund.de (last accessed on 30th May 2011).

which were above minus 10 percent, meaning that the denominator in Eq. (2) is positive. This results in positive starting capital stocks for all cities. Because of the higher uncertainty in the estimates of capital figures for the first years of observation, the figures should be treated with caution, especially for the first years until the starting capital stock is depreciated and the capital stock is predominately driven by last investments. However, the starting capital stock depreciated to 40 percent in 2004 and thereby reduces the involved uncertainty in the input factor. The capital stock for the other industry sectors is calculated based on the capital intensity in the wide manufacturing sector for each city and the ratio of capital intensity of the wide manufacturing sector compared to the other industry sectors in whole Germany. The information is given by the OECD Database for Structural Analysis (STAN[7]). The ratio for the whole of Germany is multiplied by the calculated capital in each city.

2.1.1 Productivity and Efficiency Measurement

Efficiency is measured within a non-parametric framework because the production function, which transforms inputs into outputs, is not known. Thus, a parametric setup would be questionable because of the unknown structure of the process specific for industries. The non-parametric framework to measure efficiency of cities is the data envelopment analysis (DEA), which was developed by Charnes et al. (1978). Within the DEA, the observed combinations of inputs and outputs for all cities are taken into account. The aim of the DEA is to find those cities that envelop all others. These cities building the enveloping frontier represent actual best practice and thus are efficient. All other cities could improve their efficiency by either reducing inputs for the same production of outputs or by increasing the production output for their used inputs, depending on the orientation, e.g., input- or output-orientation, respectively. The approach enables the construction and analysis of efficiency of general decision making units, which are in this dissertation cities or economic sectors within the cities with multiple inputs and outputs without requiring any information about specific prices or the underlying production function. A good introductory overview about DEA and distance functions is given by Coelli et al. (2005).

The output distance functions implemented in the analyses are those described by Shephard (1970) for constant returns to scale (CRS) as well as variable returns to scale (VRS). These distance functions are the reciprocals of those described by Farrell (1957). It is convenient to use both approaches because the underlying production function in all cities does not have to be described by CRS. However, the applicability of the VRS function has to be tested. These efficiency measurements by Shephard have values between zero and one. The value of one marks the most productive cities. The measure of scale efficiency (SE) for each city is the ratio of the distance function at CRS divided by the distance function at VRS.

There are two possible representations for an output distance function of an industry i a city j. These are defined as:

$$\theta_{ij}\left(\boldsymbol{x}_{ij}, y_{ij}\right) = \max\left\{\theta \mid \left(y_{ij} \cdot \theta\right) \in P\left(\boldsymbol{x}_{ij}\right)\right\}, \tag{3}$$

and

$$\delta_{ij}\left(\boldsymbol{x}_{ij}, y_{ij}\right) = \min\left\{\delta \mid \left(y_{ij}/\delta\right) \in P\left(\boldsymbol{x}_{ij}\right)\right\}, \tag{4}$$

where $\boldsymbol{x}_{ij}$ is the (2×1) vectors for inputs and y_{ij} is the scalar for the output for city j. $P\left(\boldsymbol{x}_{ij}\right)$ is the output set, which describes the production functions. In Eq. (3) θ is the distance for which θ_{ij} is the

[7] The database is available on the Internet by http://stats.oecd.org.

maximum value, and in Eq. (4) δ is the distance function for which δ_{ij} is the minimum, respectively. Eq. (3) is the representation equivalent to Shephard (1970, p. 207) with output-orientation, which is used in the analyses of this dissertation. Contrarily, Eq. (4) is the equivalent to Eq. (3) based on Farrell (1957), which is the more common representation used for example in Coelli et al. (2005). The calculation of the distance functions needs non-negative inputs and outputs, thus it is necessary to put special emphasis on the inputs and outputs as well as on their proper measurement.

To get estimates for the output distance functions for each city the DEA approach is used. For DEA measurements a linear programming model has to be solved. The linear programming involves finding the maximum of weighted outputs, which are still part of the production possibility set. Due to the duality in linear programming it is equivalent to find the minimum of weighted inputs and is called envelopment form. A distance function for output-orientation is defined by Farrell (1957) and calculated in a linear program for each industry separately. The envelopment form for constant returns to scale is

$$\begin{aligned} \min_{\theta,\boldsymbol{\lambda}} \theta_{CRS,ij}, & & & \\ st \qquad & -y_{ij} + \boldsymbol{y}\boldsymbol{\lambda} & \geq & \ 0 \\ & \theta\boldsymbol{x}_{ij} - \boldsymbol{X}\boldsymbol{\lambda} & \geq & \ \mathbf{0} \\ & \boldsymbol{\lambda} & \geq & \ \mathbf{0}, \end{aligned} \tag{5}$$

where $\theta_{CRS,ij}$ is the efficiency score for industry i in city j, $\boldsymbol{y}$ is a (1×112) vector containing the one output of each of the 112 cities, $\boldsymbol{\lambda}$ is a (112×1) vector of weights, and $\boldsymbol{X}$ is a (2×112) matrix for the two inputs in the 112 cities. The outputs are the gross value added of each of the industries investigated within the city, and the input matrix contains the two inputs capital and labor used in each industry within each city.

The calculation of distance functions with variable returns to scale is almost the same as for constant returns to scale in Eq. (5), except for one further constraint:

$$\begin{aligned} \min_{\theta,\boldsymbol{\lambda}} \theta_{VRS,ij}, & & & \\ st \qquad & -y_{ij} + \boldsymbol{y}\boldsymbol{\lambda} & \geq & \ 0, \\ & \theta\boldsymbol{x}_{ij} - \boldsymbol{X}\boldsymbol{\lambda} & \geq & \ \mathbf{0}, \\ & \mathbf{1}'\boldsymbol{\lambda} & = & \ 1 \\ & \boldsymbol{\lambda} & \geq & \ \mathbf{0}. \end{aligned} \tag{6}$$

The additional condition expressed in Eq. (6) constrains the weights to sum to unity. It is also called the convexity condition in Coelli et al. (2005). In the literature, there is a controversy about using VRS distance functions (see, e.g., Ray and Desli (1997) and Färe et al. (1997)). Therefore, tests have to be carried out to see whether the underlying production function, which generates the data, can be described by VRS or only CRS.

2.1.2 Productivity and Efficiency Change Measurement

The productivity and efficiency changes are measured by an index, which is named after a similar index in Malmquist (1953). Here, the definition of Färe et al. (1992) for the index is used. The Malmquist index $malm_{ij}(t_1, t_2)$ (denoted simply by *malm* later on) for two different periods in time t_1 and t_2, with $t_1 < t_2$ is defined as

$$malm_{ij}(t_1, t_2) = \sqrt{\frac{\Delta_{ij,t_1}(\boldsymbol{x}_{j,t_2}, y_{j,t_2})}{\Delta_{ij,t_1}(\boldsymbol{x}_{j,t_1}, y_{j,t_1})} \times \frac{\Delta_{ij,t_2}(\boldsymbol{x}_{j,t_2}, y_{j,t_2})}{\Delta_{ij,t_2}(\boldsymbol{x}_{j,t_1}, y_{j,t_1})}}, \tag{7}$$

with $\Delta_{ij,t_k}(\boldsymbol{x}_{ij,t_l}, y_{ij,t_l})$, the distance function of industry i in city j in period t_k in comparison to the frontier in period t_l $\Delta_{ij,t_k}(\boldsymbol{x}_{ij,t_l}, y_{ij,t_l}) = (\max\{\theta : (\boldsymbol{x}_{ij,t_l}, \theta y_{ij,t_l}) \in T(t_k)\})^{-1}$. The first factor in Eq. (7) measures the change of industry in city j from period t_1 to period t_2, and both relative to the frontier in period t_1. Analogously, the second factor in Eq. (7) gives the change of industry i in city j from period t_1 to period t_2, but both relative to the frontier in period t_2. Thus, the Malmquist index is the geometrical average of the productivity changes measured on the basis of the new and old frontier in period t_2 and period t_1, respectively. Values of the Malmquist index which are smaller than unity indicate decreases in productivity between period t_1 and period t_2, while values larger than unity indicate improvements in productivity between both periods. There are many different decompositions of this index. Because I am interested in the most common factors, I use the decomposition of Simar and Wilson (1999). The first decomposition of the Malmquist index is as described in Färe et al. (1992)

$$malm_{ij}(t_1, t_2) = \frac{\Delta_{ij,t_2}(\boldsymbol{x}_{ij,t_2}, y_{ij,t_2})}{\Delta_{ij,t_1}(\boldsymbol{x}_{ij,t_1}, y_{ij,t_1})} \times \sqrt{\frac{\Delta_{ij,t_1}(\boldsymbol{x}_{ij,t_2}, y_{ij,t_2})}{\Delta_{ij,t_2}(\boldsymbol{x}_{ij,t_2}, y_{ij,t_2})} \times \frac{\Delta_{ij,t_1}(\boldsymbol{x}_{ij,t_1}, y_{ij,t_1})}{\Delta_{ij,t_2}(\boldsymbol{x}_{ij,t_1}, y_{ij,t_1})}}. \tag{8}$$

The productivity change is still the same but the effect can be observed separately. The first factor of the Malmquist index in Eq. (8) indicates changes in efficiency (denoted by *eff* later on). The second factor expresses the technological change (denoted by *tech* later on) from period t_1 and period t_2. The change in efficiency is related to the catching-up of the industry in a particular city, whereas technological change measures shifts in the technology captured by the best practice production frontier. It should be noticed, that I only use distance functions under CRS up unto this point. As used in Wheelock and Wilson (1999), the change in efficiency can be split further to

$$\begin{aligned} eff_{ij}(t_1, t_2) &= \frac{\Delta_{ij,t_2}(\boldsymbol{x}_{ij,t_2}, y_{ij,t_2})}{\Delta_{ij,t_1}(\boldsymbol{x}_{ij,t_1}, y_{ij,t_1})} \\ &= \frac{\widetilde{\Delta}_{ij,t_2}(\boldsymbol{x}_{ij,t_2}, y_{ij,t_2})}{\widetilde{\Delta}_{ij,t_1}(\boldsymbol{x}_{ij,t_1}, y_{ij,t_1})} \times \frac{\Delta_{ij,t_2}(\boldsymbol{x}_{ij,t_2}, \boldsymbol{y}_{ij,t_2}) / \widetilde{\Delta}_{ij,t_2}(\boldsymbol{x}_{ij,t_2}, y_{ij,t_2})}{\Delta_{ij,t_1}(\boldsymbol{x}_{ij,t_1}, \boldsymbol{y}_{ij,t_1}) / \widetilde{\Delta}_{ij,t_1}(\boldsymbol{x}_{ij,t_1}, y_{ij,t_1})}, \end{aligned} \tag{9}$$

with $\widetilde{\Delta}_{ij,t}(\boldsymbol{x}_{ij,t}, y_{ij,t})$ for $t = t_1, t_2$ the distance function under VRS. The first decomposed factor in Eq. (9) is the change of pure efficiency and the second factor is the change of scale efficiency. The change of pure efficiency (denoted as *pure.eff* later on) is calculated by the ratio of the distance functions only to the VRS best practice frontier. The change in scale efficiency (denoted as *scale* later on) is the ratio of the scale efficiencies in period t_2 by period t_1. It is ratio of the distance function under CRS and that under VRS at the same time as the reference observation for the frontier in that particular period. The scale efficiency change component captures the change to the most productive scale in which the VRS and CRS frontier are equal.

In a similar way, the change in technological efficiency can be decomposed as shown in Wheelock and

Wilson (1999):

$$
\begin{aligned}
tech_{ij}\left(t_1,t_2\right) &= \sqrt{\frac{\Delta_{ij,t_1}\left(\boldsymbol{x}_{ij,t_2},y_{ij,t_2}\right)}{\Delta_{ij,t_2}\left(\boldsymbol{x}_{ij,t_2},y_{ij,t_2}\right)}\times\frac{\Delta_{ij,t_1}\left(\boldsymbol{x}_{ij,t_1},y_{ij,t_1}\right)}{\Delta_{ij,t_2}\left(\boldsymbol{x}_{ij,t_1},y_{ij,t_1}\right)}} \qquad (10)\\
&= \sqrt{\frac{\widetilde{\Delta}_{ij,t_1}\left(\boldsymbol{x}_{ij,t_2},y_{ij,t_2}\right)}{\widetilde{\Delta}_{ij,t_2}\left(\boldsymbol{x}_{ij,t_2},y_{ij,t_2}\right)}\times\frac{\widetilde{\Delta}_{ij,t_1}\left(\boldsymbol{x}_{ij,t_1},y_{ij,t_1}\right)}{\widetilde{\Delta}_{ij,t_2}\left(\boldsymbol{x}_{ij,t_1},y_{ij,t_1}\right)}}\\
&\quad\times\sqrt{\frac{\Delta_{ij,t_1}\left(\boldsymbol{x}_{ij,t_2},y_{ij,t_2}\right)/\widetilde{\Delta}_{ij,t_1}\left(\boldsymbol{x}_{ij,t_2},y_{ij,t_2}\right)}{\Delta_{ij,t_2}\left(\boldsymbol{x}_{ij,t_2},y_{ij,t_2}\right)/\widetilde{\Delta}_{ij,t_2}\left(\boldsymbol{x}_{ij,t_2},y_{ij,t_2}\right)}}\\
&\quad\times\sqrt{\frac{\Delta_{ij,t_1}\left(\boldsymbol{x}_{ij,t_1},y_{ij,t_1}\right)/\widetilde{\Delta}_{ij,t_1}\left(\boldsymbol{x}_{ij,t_1},y_{ij,t_1}\right)}{\Delta_{ij,t_2}\left(\boldsymbol{x}_{ij,t_1},y_{ij,t_1}\right)/\widetilde{\Delta}_{ij,t_2}\left(\boldsymbol{x}_{ij,t_1},y_{ij,t_1}\right)}}. \qquad (11)
\end{aligned}
$$

The first factor in the second line measures the pure change in technology, and the second factor in the third and fourth line quantifies the change in scale of technology. The pure change in technology (denoted as *pure.tech* later on) is the geometric mean of the distance ratio to the VRS frontier for each time period. The change in scale of technology (denoted as *scale.tech* later on) measures the change of returns to scale for VRS technology for the two time periods. Both components include distance functions under VRS with time different observations and reference frontiers $\widetilde{\Delta}_{ij,t_k}\left(\boldsymbol{x}_{ij,t_l},y_{ij,t_l}\right)$ with $t_k \neq t_l$. These mixed distance functions do not have to be calculable for every observation (see, e.g., Ray and Desli (1997)). Computations are performed with R using the package `FEAR` which is described in Wilson (2008).

2.1.3 Hierarchy of Cities

The idea of urban hierarchy goes back to the central place theory of Christaller (1933) and Lösch (1940) in which specific goods and services are produced at a central place for the neighboring area due to transportation cost and the need of a large market size to gain economies of scale. The central place theory classifies cities as "Oberzentrum", "Mittelzentrum", and so on. These central places are classified by the states but unfortunately almost all cities are classified as an "Oberzentrum" in the sample. Therefore, it is not appropriate to distinguish cities by their classification. However, higher order cities have to supply additional goods and services for a larger serve of people in the city and satellite cities of lower order as shown in Fujita et al. (1999). For example, cities highest in urban hierarchy have to supply not only administrative services, large shopping and leisure facilities but also cultural facilities like operas and so on. Lower ordered cities within the urban hierarchy will be specialized in producing only some goods and services. Henderson (2010) implements the manufacturing share of employment to account for urban hierarchy which is also used for example in Glaeser et al. (1995). In addition and as a refinement, Au and Henderson (2006a,b) use the manufacturing to service ratio to account for urban hierarchy. Evidently, the agricultural sector is of minor importance in urban analysis.

To classify hierarchical groups with the service to manufacturing ratio (*SMr*), I use the Ward approach (Ward, 1963). The advantage of using the Ward approach is that each resulting cluster has approximately the same amount of cities. Furthermore, it provides an optimal number of clusters to restore the information of variety within the data. According to the Milligan/Cooper criterion (Milligan and Cooper, 1985) the optimal number of clusters for the 112 cities is four. Figure 2.1 gives an overview of the cities with the values of the *SMr* as well as the cluster arrangement. In the figure the order of the clusters is from top to bottom three, four, two and one. Figure 2.1 shows, the city with the lowest *SMr* is Wolfsburg arranged in cluster one and the city with the highest *SMr* is Potsdam which is arranged in cluster four.

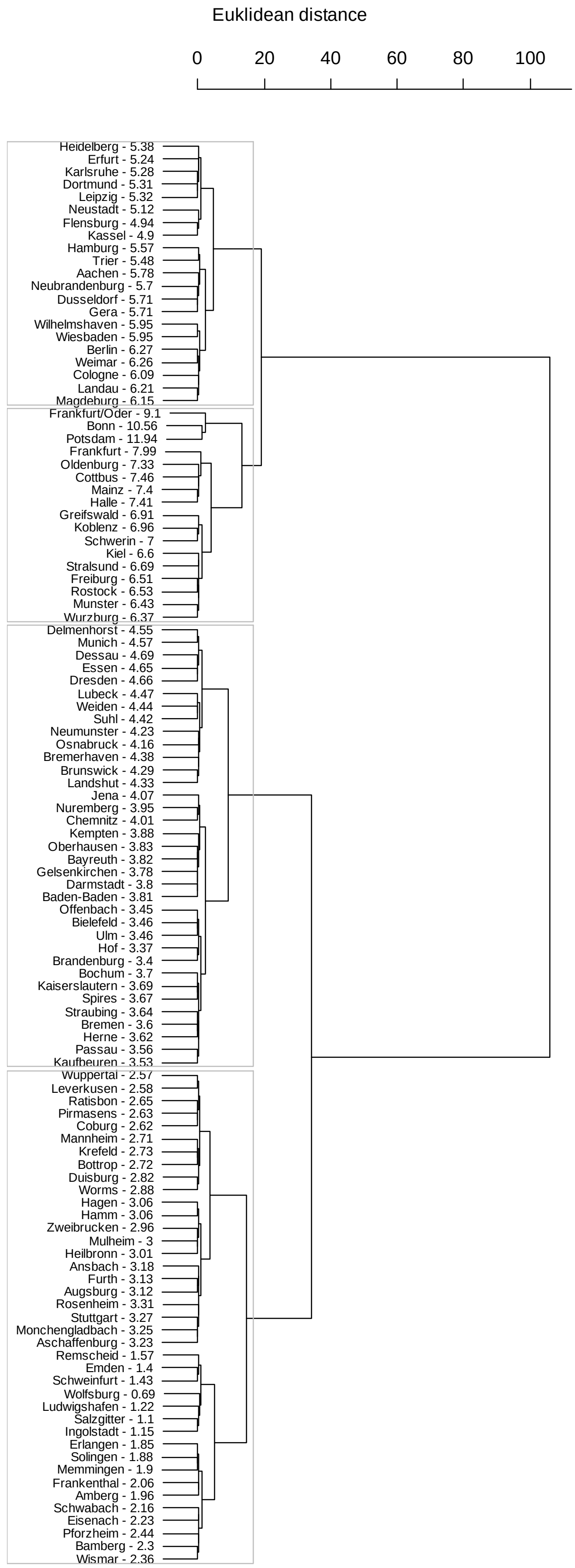

Figure 2.1: Hierarchical Clusters for Employment Service to Manufacturing Ratio

2.2 Robust Regression Analysis

The following estimations correct the variance estimation in the cause of heteroscedasticity and autocorrelation as well as the bias caused by outliers within the data.

The OLS estimation model in general is

$$\boldsymbol{y} = \boldsymbol{X}\beta + \boldsymbol{e}, \tag{12}$$

with $\boldsymbol{y}$ the $(n \times 1)$ vector of the observation of the dependent variable for all n cities, $\boldsymbol{X}$ the $(n \times m)$ matrix with the m independent variables including the intercept of all n cities, the error term which should be $\boldsymbol{e} \sim N(\boldsymbol{0}, \sigma^2 \boldsymbol{I})$, and the vector of the m regression parameters β. The OLS estimation minimizes the sum of squared residuals e_j for all j

$$\min_{\beta} \sum_{j=1}^{n} (e_j)^2 = \min_{\beta} \sum_{j=1}^{n} (y_j - \boldsymbol{x}_j \beta)^2 \tag{13}$$

or in matrix notation

$$\min_{\beta} \boldsymbol{e}'\boldsymbol{e} = \min_{\beta} (\boldsymbol{y} - \boldsymbol{X}\beta)' (\boldsymbol{y} - \boldsymbol{X}\beta)$$

with the solution $\widehat{\beta}_{OLS} = \left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1} \boldsymbol{X}'\boldsymbol{y}$ and $\boldsymbol{x}_j$ the j-th row, which contains the m variables (including the intercept) of city j, within the explanatory matrix $\boldsymbol{X}$ and y_j the observation for the dependent variable of city j. The OLS method is the best linear unbiased estimator, if some assumptions hold, such as the homoscedasticity of the residuals (see Greene (2008a, p. 11)). In that case, the covariance matrix of the coefficient is $\sigma^2 \left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1}$, with σ^2 the variance of the error term which is estimated by $\hat{\sigma}^2 = (n-m)^{-1} \sum_{j=1}^{n} \left(\hat{e}_j^2\right)$ and is equal for all coefficients.

By considering the distribution of city size with a large number of small cities and a fewer number of the largest cities, the variance declines. The decline results in a heteroscedasticity problem. Heteroscedasticity leads to inefficient OLS estimations and miscalculated standard errors and results therefore in erroneous significance test. Statements on the significance are permitted in these cases. Therefore, standard errors for the OLS estimates have to be calculated with a heteroscedasticity consistent covariance matrix, which is estimated by

$$\left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1} \boldsymbol{X}'\Omega\boldsymbol{X} \left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1} \tag{14}$$

with

$$\Omega = \operatorname{diag}\left[\frac{\hat{e}_j^2}{(1-h_{jj})^2}\right]. \tag{15}$$

Eq. (14) together with Eq. (15) represent the $HC3$ matrix as denoted in Long and Ervin (2000). As weighting the leverage of observation j $h_{jj} = \boldsymbol{x}_j \left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1} \boldsymbol{x}_j'$ is taken, which is the jj-element in the Hat-Matrix. The leverage of the observation h_{jj} is between $1/n$ and unity with high values for leverage points. MacKinnon and White (1985, p. 313) and also Long and Ervin (2000, p. 222) show by experiments that the $HC3$ matrix is preferable especially for small sample sizes smaller 250, which is the case in this dissertation. For computation the R package `sandwich` is used as explained in Zeileis (2004).

The remaining problem is the presence of outliers in explaining and explanatory variables. The OLS approach is not robust against outliers, which causes the estimates to be biased. Further methods are therefore applied that are better able to guard against outliers. One robust estimation approach is the least median of squares (LMS) method of Rousseeuw (1984), which minimizes the median of squared residuals instead of the sum of squared residuals as OLS does. The objective function is

$$\min_{\beta} \text{med}_j e_j^2. \tag{16}$$

There is no analytic solution for the LMS method and thus the residuals have to be compared with other solutions with respect to minimize the median of squared residuals. The advantage of the LMS estimation is that it is much more robust against residual outliers than the OLS estimation. The breakdown point, which states by how many percentage points of the observation is allowed to diverge without changing the estimates, is 0.5 for the LMS method indicating the robustness of this method and its results, respectively. These estimations are performed by use of the R package `MASS`. Although LMS results are highly robust, the results are inefficient. Additional methods are consequently needed which are not only robust but also efficient.

An additional robust method is the MM-estimation as described in Yohai (1987) in order to regard random regressors with possible outliers. The MM-estimator consists of three steps, where two different maximum likelihood type estimations have to be solved. First, an initial regression estimate $\hat{\beta}_0$ has to be found, which should be robust by means of a high breakdown point. The breakdown point determines the robustness of the MM-estimation since this breakdown point is not decreased by the following steps. The applied robust method for getting the starting estimation is the iterated re-weighted least squares (IRWLS) method as described in Yohai (1987) or Maronna et al. (2006). The IRWLS approach is computed in the three following steps:

1. Compute an initial starting point for the estimate of $\hat{\beta}_0$ by least absolute deviation estimation and scale $\widehat{s}$, with
   $\min \sum_{j=1}^{n} |u_j|$ with $u_j = y_j - x_j' \hat{\beta}_0$ and $\widehat{s} = \frac{1}{0.675} \text{med}_j \left(u_j \mid u_j \neq 0 \right)$.
2. Iterate the estimation of $\hat{\beta}_k$ for $k = 0, 1, 2, 3, \ldots$ with the constant scale $\widehat{s}$ by solving $\sum_{j=1}^{n} w_{j,k} x_j \left(y_j - x_j' \hat{\beta}_{k+1} \right)$ with $u_{j,k} = y_j - x_j' \hat{\beta}_k$ and $w_{j,k} = W\left(\frac{u_{j,k}}{\widehat{s}} \right)$ with $W()$ a non-increasing function for positive arguments.
3. Stop the iteration when $\max_j \left(|u_{j,k} - u_{j,k+1}| \right) / \widehat{s} < \varepsilon$.

The second step of the MM-estimation computes the M-scale $\hat{s}_n$, which is the scale for the residuals u_j resulting of the initial regression estimate $\hat{\beta}_0$. This is a maximum likelihood type estimation and therefore gives the first M in the name of the method. The objective function is

$$\min_{\widehat{\beta}} \sum_{j=1}^{n} \rho_0 \left(\frac{u_j \left(\hat{\beta}_0 \right)}{\hat{s}_n} \right) \tag{17}$$

with the first order condition

$$\sum_{j=1}^{n} \psi_0 \left(\frac{u_j \left(\hat{\beta}_0 \right)}{\hat{s}_n} \right) \boldsymbol{x}_j = \boldsymbol{0}, \tag{18}$$

where $\rho_0()$ is a real function in the residuals which are scale invariant by the M-scale $\hat{s}_n$ and $\psi_0()$ is the first derivative of $\rho_0()$. The properties of the function $\rho_0()$ are given in Huber (1981) or Yohai (1987), for instance symmetry, continuity, a supremum between zero and infinity and monotonic increasing for positive values where $\rho_0(0) = 0$. The sum in Eq. (17) divided by n and the supremum of $\rho_0()$ has to be 0.5, so that the breakdown point of the estimator is 0.5. In that step the initial regression estimate and the resulting residuals of the first step are taken as given and Eq. (17) is minimized by the M-scale. The third step minimizes a different maximum likelihood-type function $\rho_1() \leq \rho_0()$ with the same supremum and the M-scale $\hat{s}_n$ of the second step taken as given

$$\min_{\widehat{\beta}} \sum_{j=1}^{n} \rho_1 \left(\frac{u_j\left(\hat{\beta}_1\right)}{s_n} \right). \tag{19}$$

This estimator is another maximum likelihood-type estimation, which justifies the second M. Yohai (1987) shows that the estimates found by those steps are as robust as the LMS method with a breakdown-point of 0.5, but are also highly efficient. For further explanations see Maronna et al. (2006).

2.3 Spatial Analysis

The general model is already shown in subsection 2.2 in Eq. (12). If the errors are $\boldsymbol{e} \sim N(\boldsymbol{0}, \sigma^2 \boldsymbol{I})$ and the other Gauss-Markov conditions are true, the estimates for the regression parameters $\hat{\beta} = \left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1}\boldsymbol{X}'\boldsymbol{y}$ are the best linear unbiased estimators. But for spatial data it is possible that the errors are not independent of each other but are spatially autocorrelated. Therefore it is necessary to test whether there is spatial correlation between the observations or not. If there is spatial autocorrelation within the error terms, spatial models have to be estimated to ensure efficient estimates. Otherwise, the estimates are not efficient which results in wrong test statistics as Cordy and Griffith (1993) show. There are two commonly used tests for spatial autocorrelation namely *Moran's I* (Moran (1950)) and *Geary's c* (Geary (1954)) statistics, which are described in Cliff and Ord (1981, p. 13 and pp. 42f.). These two tests give first evidence for the presence of spatial autocorrelation although there are many more test statistics, see Getis (2007). Additional tests will be performed for spatial models after introducing these models in the following.

On the one hand, a commonly used test for spatial autocorrelation is (global) *Moran's I* calculated as

$$I = \frac{n}{A} \frac{\sum_{j=1}^{n} \sum_{k=1}^{n} \delta_{jk} \left(y_j - \bar{y}\right)\left(y_k - \bar{y}\right)}{\sum_{j=1}^{n} \left(y_j - \bar{y}\right)^2}$$

with $A = \sum_{j=1}^{n} \sum_{k=1}^{n} \delta_{jk}$ and $\delta_{jk} = 1$ if city j and k are neighbors, otherwise $\delta_{jk} = 0$. On the other hand, *Geary's c* is statistics given as (see Hubert et al. (1981, p. 225) for both coefficients)

$$c = \frac{(n-1) \sum_{j=1}^{n} \sum_{k=1}^{n} \delta_{jk} \left(y_j - y_k\right)^2}{2A \sum_{j=1}^{n} \left(y_j - \bar{y}\right)^2}.$$

Moran's I takes the values between -1 and 1 as correlation measures whereas *Geary's c* is between 0 and 2 with 1 indicating no spatial autocorrelation. Both test statistics have the same structure with a measure of the spatial covariance in the numerator and the variance in the denominator. *Moran's I* becomes even simpler for row-standardized weights $w_{jk} = \delta_{jk} / \sum_{k=1}^{n} \delta_{jk}$ which leads to $\sum_{k=1}^{n} w_{jk} = 1$ for all j and assuming that each observation j has at least one neighbor. The $\sum_{j=1}^{n} \sum_{k=1}^{n} w_{jk} = n$ and *Moran's I* is

$$I = \frac{\boldsymbol{e}'\boldsymbol{W}\boldsymbol{e}}{\boldsymbol{e}'\boldsymbol{e}},$$

which is standard normally distributed, with $\boldsymbol{e}$ the error components in Eq. (12). Instead of calculating the weight matrix based on the binary measure δ_{ik} many neighborhood indexes are approached such as geographical distance, nearest neighbor approaches or closeness with regard to trade flows, etc. The tests for spatial autocorrelation take the test statistics reduced by the expected mean under the null hypothesis of no spatial autocorrelation and divide them by the specific variance which is calculated in Cliff and Ord (1981) or Haining (2003, pp. 245ff.).

Following Cliff and Ord (1981) *Moran's I* is preferable to *Geary's c* statistic because it is less affected by the distribution of the upper boundary under the null hypothesis of no spatial autocorrelation (Cliff and Ord, 1981, p. 21). The problem with both statistics is that they take into account only direct neighbors and not neighbors of the second order or higher. Furthermore, additional test coefficients are available such as *Getis and Ord's G (Getis and Ord, 1992)* and many others as set out in Getis (2007).

If there is correlation among the error terms the OLS estimator is inefficient, the variance estimator is downwards biased, coefficients of determination are too large and the tests of significance are wrong. To cope the problem spatial models have to be estimated.

Pure Spatial Autoregressive Model

To introduce and distinguish spatial models I present the two basic purely spatial models, namely the spatial autoregressive model and the spatial error model. The basic spatial autocorrelation model as described in Cliff and Ord (1981) explains observation by spatially lagged observation and an error term

$$y_j = \rho \sum_{k=1}^{n} w_{jk} y_k + e_j \tag{20}$$

or in matrix notation

$$\boldsymbol{y} = \rho \boldsymbol{W} \boldsymbol{y} + \boldsymbol{e},$$

with independent and identically distributed errors with common variance σ^2 and a set of weights w_{jk} for area k directly spatially related to y_j with $w_{jk} \geq 0$. One example for the spatial weight is $w_{jk} = (c + d_{jk})^{-\alpha}$, where d_{jk} is the distance between city j and k, α is a friction distance parameter from gravity and interaction models, and the constant $c \geq 0$. Specifically, it is also possible to have only $w_{jk} = d_{jk}^{-b}$ as spatial weight, as used in Gartell (1979). The parameter ρ is the spatial autocorrelation parameter which is typically $|\rho| < 1$ but of course depends on the spatial weight matrix $\boldsymbol{W}$ which is typically row standardized, see Kelejian and Robinson (1995). Otherwise the spatial autocorrelation parameter can be any real number.

Pure Spatial Error Model

The other basic model is the spatial error model

$$y_j = e_j + \lambda \sum_{k \neq j} w_{jk} e_k, \tag{21}$$

where $E(y_j) = 0$, $E(e_j) = 0$, $E\left(e_j^2\right) = \sigma^2$, and $E(e_j e_k) = 0$ for $j \neq k$ and λ the spatial parameter (not to be confused with the vector of weights in efficiency measurement in subsection 2.1.1).

Spatial Autoregressive Model with Explanatory Variables

Since the basic models only include spatially lagged observations of the explained variable, they have to be extended by independent or explanatory variables. For example, the pure spatial autoregressive model (SAM) can be extended in the matrix notation of Anselin (2003b)

$$\boldsymbol{y} = \rho \boldsymbol{W} \boldsymbol{y} + \boldsymbol{X} \beta + \boldsymbol{e}, \tag{22}$$

with $\boldsymbol{X}$, the $(n \times m)$ matrix with the m independent variables including the intercept for all n cities, and β, the $(m \times 1)$ vector for the parameters. The errors should again be normally distributed with zero means and constant variance σ^2, $\boldsymbol{e} \sim N\left(\boldsymbol{0}, \sigma^2 \boldsymbol{I}\right)$. Eq. (22) is equivalent to

$$\boldsymbol{y} = (\boldsymbol{I} - \rho \boldsymbol{W})^{-1} (\boldsymbol{X}\beta + \boldsymbol{e}) = (\boldsymbol{I} - \rho \boldsymbol{W})^{-1} \boldsymbol{X}\beta + (\boldsymbol{I} - \rho \boldsymbol{W})^{-1} \boldsymbol{e}. \tag{23}$$

According to Anselin (2003b) Eq. (22) is called a structural form while Eq. (23) is called a reduced form because it does not contain the dependent variable on the right hand side of the equation.

The log-likelihood for estimation is comparable to the one of LeSage (2008) and to the one proposed by Ord (1975)

$$\log \mathcal{L} = \frac{n}{2} \log 2\pi - \frac{n}{2} \log \sigma^2 + \log |\boldsymbol{I} - \rho \boldsymbol{W}| - \frac{1}{2\sigma^2} (\boldsymbol{y} - \rho \boldsymbol{W} \boldsymbol{y} - \boldsymbol{X}\beta)' (\boldsymbol{y} - \rho \boldsymbol{W} \boldsymbol{y} - \boldsymbol{X}\beta)$$

or with the eigenvalues ω_j it follows $|\boldsymbol{I} - \rho \boldsymbol{W}| = \prod_{j=1}^{n} (1 - \rho\omega_j)$ as shown in Ord (1975) and the log-likelihood is

$$\log \mathcal{L} = \frac{n}{2} \log 2\pi - \frac{n}{2} \log \sigma^2 + \sum_{j=1}^{n} \log (1 - \rho\omega_j) - \frac{1}{2\sigma^2} (\boldsymbol{y} - \rho \boldsymbol{W} \boldsymbol{y} - \boldsymbol{X}\beta)' (\boldsymbol{y} - \rho \boldsymbol{W} \boldsymbol{y} - \boldsymbol{X}\beta).$$

The solution is Anselin (2003a, p. 320)

$$\hat{\beta} = \left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1} \boldsymbol{X}'\boldsymbol{y} - \rho \left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1} \boldsymbol{X}'\boldsymbol{W}\boldsymbol{y}$$

and

$$\hat{\sigma}^2 = \frac{1}{n} \left(\rho \left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1} \boldsymbol{X}'\boldsymbol{W}\boldsymbol{y} - \boldsymbol{X}\left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1} \boldsymbol{X}'\boldsymbol{y}\right)' \left(\rho \left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1} \boldsymbol{X}'\boldsymbol{W}\boldsymbol{y} - \boldsymbol{X}\left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1} \boldsymbol{X}'\boldsymbol{y}\right).$$

Spatial Error Model with Explanatory Variables

The next spatial model is the extension of the spatial error model (SEM) with spatial autoregressive error in terms of

$$e_j = \sum_{k=1}^{M} b_{jk} e_k + u_j, \tag{24}$$

for all j with $b_{jj} = 0$ and $b_{jk} = \lambda w_{jk}$ for $j \neq k$ or in matrix notation

$$\mathbf{e} = \lambda \boldsymbol{W} \mathbf{e} + \boldsymbol{u} = (\boldsymbol{I} - \lambda \boldsymbol{W})^{-1} \boldsymbol{u} = \lambda \boldsymbol{W} (\boldsymbol{y} - \boldsymbol{X}\beta) + \boldsymbol{u},$$

with $\boldsymbol{I}$, the identity matrix. The matrix $(\boldsymbol{I} - \lambda \boldsymbol{W})$ must be non-singular. Therefore, the regression model becomes

$$\boldsymbol{y} = \boldsymbol{X}\beta + \lambda\boldsymbol{W}(\boldsymbol{y} - \boldsymbol{X}\beta) + \boldsymbol{u} = \boldsymbol{X}\beta + (\boldsymbol{I} - \lambda\boldsymbol{W})^{-1}\boldsymbol{u} \tag{25}$$

and the residuals are

$$(\boldsymbol{I} - \lambda\boldsymbol{W})(\boldsymbol{y} - \boldsymbol{X}\beta) = \boldsymbol{u}. \tag{26}$$

The SEM is sometimes written differently by replacing λ by $-\theta$ in Eq. (25), so it becomes $\boldsymbol{y} = \boldsymbol{X}\beta + (\boldsymbol{I} + \theta\boldsymbol{W})^{-1}\boldsymbol{u}$ and it is then called spatial moving average process as used in e.g., LeSage and Pace (2014). Eq. (25) is the representation of the data generating process for a spatial error model. Eq (25) may also be written as

$$\boldsymbol{y} = \lambda\boldsymbol{W}\boldsymbol{y} + (\boldsymbol{I} - \lambda\boldsymbol{W})\boldsymbol{X}\beta + \boldsymbol{u} = \lambda\boldsymbol{W}\boldsymbol{y} + \boldsymbol{X}\beta - \lambda\boldsymbol{W}\boldsymbol{X}\beta + \boldsymbol{u}, \tag{27}$$

with $\boldsymbol{u} \sim N\left(\boldsymbol{0}, \sigma^2\boldsymbol{I}\right)$. That expression is similar to the spatial autoregressive model with the addition of the spatially lagged independent variable.

The optimal estimation parameter is the result of the minimization of the sum of squared residuals by the regression parameters β, which is given as follows

$$\begin{aligned} \min_{\beta} \boldsymbol{u}'\boldsymbol{u} &= \left[(\boldsymbol{I} - \lambda\boldsymbol{W})(\boldsymbol{y} - \boldsymbol{X}\beta)\right]'(\boldsymbol{I} - \lambda\boldsymbol{W})(\boldsymbol{y} - \boldsymbol{X}\beta) \\ &= \boldsymbol{y}'\left(\boldsymbol{I} - \lambda\boldsymbol{W}'\right)(\boldsymbol{I} - \lambda\boldsymbol{W})\boldsymbol{y} - 2\boldsymbol{y}'\left(\boldsymbol{I} - \lambda\boldsymbol{W}'\right)(\boldsymbol{I} - \lambda\boldsymbol{W})\boldsymbol{X}\beta \\ &\quad + \beta'\boldsymbol{X}'\left(\boldsymbol{I} - \lambda\boldsymbol{W}'\right)(\boldsymbol{I} - \lambda\boldsymbol{W})\boldsymbol{X}\beta. \end{aligned}$$

The first order condition and the resulting solution for the estimates are

$$\frac{\partial \boldsymbol{u}'\boldsymbol{u}}{\partial \beta} = \left[-2\boldsymbol{y}'\left(\boldsymbol{I} - \lambda\boldsymbol{W}'\right)(\boldsymbol{I} - \lambda\boldsymbol{W})\boldsymbol{X}\right]' + 2\boldsymbol{X}'\left(\boldsymbol{I} - \lambda\boldsymbol{W}'\right)(\boldsymbol{I} - \lambda\boldsymbol{W})\boldsymbol{X}\beta = \boldsymbol{0}$$

$$\Rightarrow \hat{\beta} = \left[\boldsymbol{X}'\left(\boldsymbol{I} - \lambda\boldsymbol{W}'\right)(\boldsymbol{I} - \lambda\boldsymbol{W})\boldsymbol{X}\right]^{-1}\boldsymbol{X}'\left(\boldsymbol{I} - \lambda\boldsymbol{W}'\right)(\boldsymbol{I} - \lambda\boldsymbol{W})\boldsymbol{y}.$$

It should be noticed, that since the $\lambda\boldsymbol{W}$ is included within the estimation, the autoregressive parameter of the error terms λ is needed before the estimation of the normal coefficients. Furthermore, as Gallo (2014) points out, the SAR model already includes heterogeneity by spatial weights.

Spatial Durbin Model

An extension of the spatial error model in Eq. (27) is the spatial Durbin model (SDM) which is commonly used in spatial growth regressions (see LeSage and Fischer (2008)). The SDM is a generalized version of the SEM with its structural form (Eq. (27)) of $\boldsymbol{y} = \lambda\boldsymbol{W}\boldsymbol{y} + \boldsymbol{X}\beta - \lambda\boldsymbol{W}\boldsymbol{X}\beta + \boldsymbol{u}$. This SEM model is also called the common factor model because it contains the factor $\boldsymbol{X}\beta$ twice. A more generalized version of the model is the spatial Durbin model with the coefficients η for β in the second factor

$$\boldsymbol{y} = \lambda\boldsymbol{W}\boldsymbol{y} + \boldsymbol{X}\beta + \boldsymbol{W}\boldsymbol{X}\eta + \mathbf{e}, \tag{28}$$

with $\mathbf{e} \sim N\left(\boldsymbol{0}, \sigma_e^2\boldsymbol{I}\right)$, which is equivalent to a reduced form

$$\boldsymbol{y} = (\boldsymbol{I} - \lambda \boldsymbol{W})^{-1} (\boldsymbol{X}\beta + \boldsymbol{W}\boldsymbol{X}\eta + \mathbf{e}). \tag{29}$$

The SEM is therefore a special case of the SDM if the constraint $\eta = -\lambda\beta$ is true, which is called the common factor constraint (Elhorst, 2010). The SDM is also a generalized version of the SAM, as can be seen by replacing ρ by λ and including the additional term $\boldsymbol{W}\boldsymbol{X}\eta$ on the right hand side of a structural form of a SAR as in Eq. (22) or if $\eta = \mathbf{0}$ the SDM is the SAM.

More generalized models are presented in Elhorst (2010), although not needed in the analysis in dissertation.

Testing Procedure

Once the spatial autocorrelation within the OLS residuals is detected, the SAM and SEM specifications have to be calculated to ascertain which model has to be used in the second step of the specific-to-general approach of Mur and Angulo (2009). The tests compare the SAM or SEM model with the OLS results by Lagrange multiplier statistics as explained in Anselin (1988a), Anselin (1988b), Florax and Rey (1995), and Anselin et al. (1996) and preferred by Carriazo and Coulson (2010). Florax and Rey (1995) show the Lagrange multiplier tests are scaled squared *Moran*'s I coefficients as

$$LMerr = \frac{(\boldsymbol{e}'\boldsymbol{W}\boldsymbol{e}/(\boldsymbol{e}'\boldsymbol{e}/n))^2}{\mathsf{tr}\left(\boldsymbol{W}'\boldsymbol{W} + \boldsymbol{W}^2\right)} = \frac{n^2}{\mathsf{tr}\left(\boldsymbol{W}'\boldsymbol{W} + \boldsymbol{W}^2\right)} \left(\frac{\boldsymbol{e}'\boldsymbol{W}\boldsymbol{e}}{\boldsymbol{e}'\boldsymbol{e}}\right)^2 \tag{30}$$

and

$$\begin{aligned} LMlag &= \frac{(\boldsymbol{e}'\boldsymbol{W}\boldsymbol{e}/(\boldsymbol{e}'\boldsymbol{e}/n))^2}{\mathsf{tr}\left(\boldsymbol{W}'\boldsymbol{W} + \boldsymbol{W}^2\right) + (\boldsymbol{W}\boldsymbol{X}\beta)'\left(\boldsymbol{I} - \boldsymbol{X}\left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1}\boldsymbol{X}'\right)(\boldsymbol{W}\boldsymbol{X}\beta)/(\boldsymbol{e}'\boldsymbol{e}/n)} \\ &= \frac{n}{\frac{(\boldsymbol{W}\boldsymbol{X}\beta)'\left(\boldsymbol{I} - \boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\right)(\boldsymbol{W}\boldsymbol{X}\beta)}{\boldsymbol{e}'\boldsymbol{e}} + \mathsf{tr}\left(\boldsymbol{W}'\boldsymbol{W} + \boldsymbol{W}^2\right)/n} \left(\frac{\boldsymbol{e}'\boldsymbol{W}\boldsymbol{e}}{\boldsymbol{e}'\boldsymbol{e}}\right)^2. \end{aligned} \tag{31}$$

Both test-statistics are χ^2 distributed with one degree of freedom (Anselin, 2003a, p. 324). Anselin and Florax (1995) show that the tests perform badly with small samples of around 50 while they perform well for sample sizes of about 100. Both tests outperform *Moran's I* tests as shown in Florax and Folmer (1992). If both test statistics are significant and the test statistic of *LMlag* is higher than *LMerr* the SAM model has to be used whereas SEM hast to be used if *LMerr* is higher than *LMlag*.

In the third step, the residuals of the spatial model have to be tested for remaining spatial autocorrelation. If the residuals of the SAM or SEM contain autocorrelation the SDM has to be used, as shown by Mur and Angulo (2009). The properties of the test for spatial dependence are explained by Anselin and Rey (1991) and Anselin (2002).

The test for the necessity of the SDM is a test for omitted variables as proposed by LeSage and Pace (2009, pp. 61-68) because the SDM is able to account for such omitted variables. Furthermore, a Hausman test involves the inefficient but consistent OLS estimators and the efficient SEM. An analogous test for SAM is not presented because the coefficients of OLS models are not consistent under the SAM. Due to the inconsistency of the OLS estimation, a Hausman test is not possible because both estimators have to be consistent and only differ in efficiency. Since the OLS estimator is

$$\hat{\beta}_{OLS} = \left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1}\boldsymbol{X}'\left((\boldsymbol{I} - \rho\boldsymbol{W})^{-1}\boldsymbol{X}\beta + (\boldsymbol{I} - \rho\boldsymbol{W})^{-1}\boldsymbol{e}\right)$$

$$E\left(\hat{\beta}_{OLS}\right) = \left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1}\boldsymbol{X}'\left(\boldsymbol{I}-\rho\boldsymbol{W}\right)^{-1}\boldsymbol{X}\beta, \tag{32}$$

whereas the SAM estimator is

$$\hat{\beta}_{SAM} \quad = \quad \left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1}\boldsymbol{X}'\left(\boldsymbol{I}-\rho\boldsymbol{W}\right)\left(\boldsymbol{X}\beta+\boldsymbol{e}\right)$$

with the maximum likelihood solution (see above)

$$E\left(\hat{\beta}_{SAM}\right) = \left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1}\boldsymbol{X}'\boldsymbol{y} - \rho\left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1}\boldsymbol{X}'\boldsymbol{W}\boldsymbol{y}$$

and is therefore not the same due to the inversion of the matrix $(\boldsymbol{I}-\rho\boldsymbol{W})$ in the OLS model.

But a similar test is possible with the total impacts of the variable. The total impacts are similar to the coefficients of the OLS estimations because they account for the partial effects of the independent variable on the same independent variable. The total impacts are calculated by the data generating process of the SAM as presented in Eq. (23), which will be repeated here

$$\boldsymbol{y} = \left(\boldsymbol{I}-\rho\boldsymbol{W}\right)^{-1}\boldsymbol{X}\boldsymbol{\beta} + \left(\boldsymbol{I}-\rho\boldsymbol{W}\right)^{-1}\boldsymbol{e}$$

or explicitly with each vector in which $\boldsymbol{x}_0 = \mathbf{1}$ a $(n \times 1)$ vector for the intercept

$$\boldsymbol{y} = \sum_{r=0}^{m}\left(\boldsymbol{I}-\rho\boldsymbol{W}\right)^{-1}\boldsymbol{I}\boldsymbol{\beta}_r\boldsymbol{x}_r + \left(\boldsymbol{I}-\rho\boldsymbol{W}\right)^{-1}\boldsymbol{e}.$$

For comparison, the matrix $\left(\boldsymbol{I}-\rho\boldsymbol{W}\right)^{-1}\boldsymbol{I}\boldsymbol{\beta}_r$ is a $(n \times n)$ matrix and called $\boldsymbol{S}_r(\boldsymbol{W})$ in LeSage and Pace (2009, pp. 34ff.). For each observation $j = 1, ..., n$ it is

$$y_j = \sum_{r=0}^{m}\left[\left(\left(\boldsymbol{I}-\rho\boldsymbol{W}\right)^{-1}\boldsymbol{I}\beta_r\right)_{jr}\boldsymbol{x}_{jr}\right] + \left(\left(\boldsymbol{I}-\rho\boldsymbol{W}\right)^{-1}\right)_j\mathbf{e}, \tag{33}$$

with $\left(\left(\boldsymbol{I}-\rho\boldsymbol{W}\right)^{-1}\boldsymbol{I}\beta_r\right)_{jr}$ the j-th row of the matrix $\left(\boldsymbol{I}-\rho\boldsymbol{W}\right)^{-1}\boldsymbol{I}\beta_r$, $\boldsymbol{x}_{jr}$ the observation vector of city j for all $q+1$ explanatory variables including the intercept, and $\left(\left(\boldsymbol{I}-\rho\boldsymbol{W}\right)^{-1}\right)_j$ the j-th row of the matrix $\left(\boldsymbol{I}-\rho\boldsymbol{W}\right)^{-1}$. This matrix $\left(\boldsymbol{I}-\rho\boldsymbol{W}\right)^{-1}$ represents an infinite geometric series of $\boldsymbol{I}+\rho\boldsymbol{W}+\rho^2\boldsymbol{W}^2+\rho^3\boldsymbol{W}^3+...$, with $\boldsymbol{W}^d$, the weight matrix of the d-th order, which accounts for the d-th neighbor. Therefore, every city has an impact on the observation on every other city although this is very small. Thus, it is easily shown that the impact of one explanatory variable depends on the observation it is calculated for and its specific neighborhood structure. The partial derivative of the dependent variable with respect to each explanatory variable and the intercept $(r = 0, ..., q)$ varies over all cities j and is

$$\frac{\partial y_j}{\partial x_{jr}} = \left(\left(\boldsymbol{I}-\rho\boldsymbol{W}\right)^{-1}\boldsymbol{I}\beta_r\right)_{jjr}$$

which is the j-th element of vector $\left(\left(\boldsymbol{I}-\rho\boldsymbol{W}\right)^{-1}\boldsymbol{I}\beta_r\right)_{jr}$ and which is not β_r as soon as the one off-diagonal element of the weight matrix is not equal to zero. The average direct impact is then the sum of the partial derivatives for every city divided by the number of city which is $n^{-1}\mathsf{tr}\left(\left(\boldsymbol{I}-\rho\boldsymbol{W}\right)^{-1}\boldsymbol{I}\beta_r\right)$ for each coefficient $r = 0, ..., m$. It only measures the effect of changes in a single city-specific observation.

Similar to the coefficients of the OLS model is the average total impact, which measures the effect occurring if the r-th variable is changed by the same amount across all observations. The average total impact is the sum of all impact of r-th variable divided by the sum of cities $n^{-1}\mathbf{1}'\left(\boldsymbol{I}-\rho\boldsymbol{W}\right)^{-1}\boldsymbol{I}\beta_r\mathbf{1}$ for each $r = 0, ..., m$. LeSage and Pace (2009, p. 39) recommend the presentation of both impact measures, which are comparable to the OLS estimates as shown in Eq. (32). Furthermore, LeSage and Pace (2009) call the difference between average total impact and average direct impact the average indirect effect.

The OLS estimates are biased and inconsistent due to simultaneity bias if the data is generated by the spatial autoregressive model as pointed out in Anselin and Bera (1998, p. 247). The simultaneity by $\boldsymbol{W}\boldsymbol{y}$ in Eq. (22) causes

$$E\left(\left(\boldsymbol{W}\boldsymbol{y}\right)_i e_i\right) = E\left(\left(\boldsymbol{W}\left(\boldsymbol{I}-\rho\boldsymbol{W}\right)^{-1}\boldsymbol{e}\right)_i e_i\right) \neq 0$$

and

$$Var\left(\boldsymbol{y}\right) = \sigma^2\left(\boldsymbol{I}-\rho\boldsymbol{W}\right)^{-1}\left(\boldsymbol{I}-\rho\boldsymbol{W}'\right)^{-1},$$

where the variance matrix has full entries due to the infinite geometric series property which implies that each observation y is correlated with all other observations. For consistency $\operatorname{plim} n^{-1}(\boldsymbol{W}\boldsymbol{y}'\boldsymbol{e}) = 0$ should be true but in SAM it is $\operatorname{plim} n^{-1}\boldsymbol{e}'\boldsymbol{W}(\boldsymbol{I}-\rho\boldsymbol{W})^{-1}\boldsymbol{e} \neq 0$ for the simple spatial autocorrelation model without any further independent explanatory variables in Eq. (20), see Anselin (1988c, p. 58) or Anselin (2003a, p. 316). This variance structure has to be implemented in maximum likelihood estimation or by a stepwise estimation procedure by first estimating the spatial correlation coefficient and then adopting the variance structure in GLS estimation for the regression coefficients.

As in all spatial models the results depend on the weight matrix. For constructing the weight matrix, the S-coding scheme by Tiefelsdorf et al. (1999) is used which accounts for the heterogeneity of the number of connections or neighbors given by geographical coordinates. In contrast, the C- and W-coding scheme is preferable if the number of neighbors is high or low, respectively. The C-coding scheme standardizes the neighborhood matrix. The original neighborhood matrix before standardization has values bigger than zero or zero if the objects are neighbors or not, by the sum of all neighborhood connections. In contrast the W-coding scheme which is commonly used in spatial analyses standardizes each element by its row-sum. The S-coding scheme first standardizes the neighborhood matrix by the square root of the row-quadratic-sum, which is the standard deviation of each row, and second globally standardizes by the sum over all variance standardized values of the first step.

2.4 Multilevel Analysis

Economic activities take place at different levels, such as the micro-, meso-, and macro-level (Dopfer et al. (2004)), and few recent econometric investigations account for this nested level structure by multilevel analysis or hierarchical model analysis. The notation of the levels in this dissertation is made according to Pinheiro and Bates (2000) and the multilevel and mixed-effects model literature, in contrast to the notation for hierarchical model analysis. The first level is the industry, as all industries are nested within the second level, which are the cities, and both levels are repeatedly measured over the third level, which is the time. This level orientation is the opposite to those sometimes found in the literature on hierarchical models (e.g., Bryk and Raudenbush (1988)). It might also be possible to use the time dimension as the most nested level, as proposed, e.g., in West et al. (2007) or Tabachnick and Fidell (2007), but the data structure with the least observations within the separate industries and more observations in the city and time level reason the proposed choice of levels. Thus, the data is structured first by industries, second by cities, and third by time to calculate the multilevel models as explained in Pinheiro and Bates (2000) for the package `nlme` in R (see Pinheiro et al. (2013)).

One big advantage of the analysis with multilevel models is that independence in the errors is not required. Independence is generally violated, because the objects in my case industries and cities within each level might influence each other. Furthermore, the interaction among the levels might be present, which can be taken into account within the multilevel analysis. Multilevel models enable us to include explanatory variables on each level.

Figure 2.2 illustrates the scheme of the multilevel model used in this dissertation.

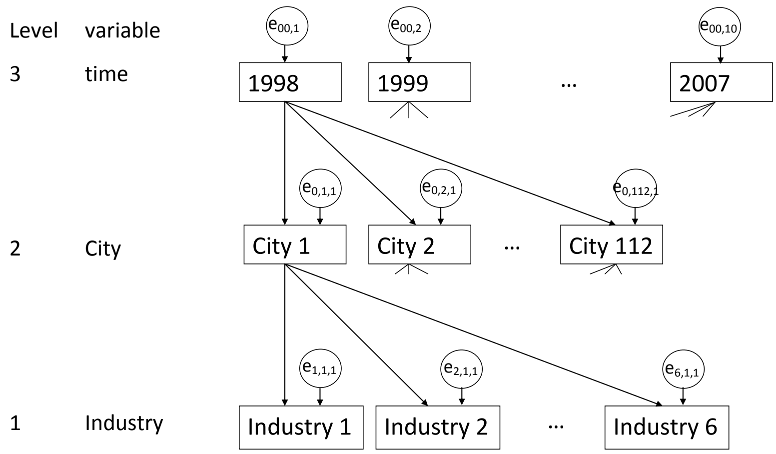

Figure 2.2: Multilevel Model Structure

In figure 2.2, units are indicated by boxes. All units of a lower level are observed in each unit of the higher level indicated by arrows from the units to the lower level units (for convenience only to the first two and last units are shown). Several errors or unobserved factors, indicated by circles in figure 2.2 affect each of these units at every level. Within multilevel models, it is possible to account for each of the unobserved factors at each level separately by specific random effects, and thus care for the nesting structure of the units. The multilevel models are developed from the most specific model, which is the basic multilevel model with the least number of random effects and no interaction terms (see, e.g. Goedhuys and Srholec (2010) or Zuur et al. (2009)). The basic multilevel model is constructed to investigate the necessity of the level structure by calculating the intraclass correlation coefficients. The basic multilevel model is then generalized by additional random effects, which allow the intercept coefficients to vary. The more generalized model is therefore called the intercept-as-outcome model. By further generalizing and allowing the slope coefficients in the model to vary by additional random effects, the most generalized model is developed, which is called the intercept-and-slope-as-outcome model. The estimations will be performed and presented in section 6.5 in the same structure beginning with the basic multilevel model, followed by the intercept-as-outcome model and finally the most generalized intercept-and-slope-as-outcome model.

2.4.1 The Basic Multilevel Model

The basic multilevel model is the starting point for the analysis. This is comprised of the industry growth trajectories $(Y_{ijt} - Y_{ij(t-1)})$ as the level-1 model in Eq. (34). Because the endogenous variables Y_{ijt}, which are either gross value added or employment, are in logarithm, the first differences measure the growth within one year. To capture a growth path and measure the effect of past productivity change within an industry on value added or employment growth, the one year lagged growth $(Y_{ij(t-1)} - Y_{ij(t-2)})$ and the productivity change $(PC_{ij(t-2)})$ measured by the Malmquist index and its components are included as explanatory variables in the level-1 model. Because the Malmquist index and its components are correlated to each other by construction, the level-1 model includes only the Malmquist index or one component for the estimation. Therefore, a separate estimation is calculated for the Malmquist index and each component. The variation in growth parameters among industries within a city is captured in the level-2 model in Eqs. (35) to (37) by the five city-specific variables $X_{ij(t-1)}$ with $i = 1, ..., 5$ and their quadratic terms $X^2_{ij(t-1)}$ for every city j and time (t-1). The variation among industries and cities over time is represented in the level-3 model in Eqs. (38) to (43), as described in Raudenbush and Bryke (2002) but with the notation from Pinheiro and Bates (2000)

$$Y_{ijt} - Y_{ij(t-1)} = \pi_{0jt} + \pi_{1jt}\left(Y_{ij(t-1)} - Y_{ij(t-2)}\right) + \pi_{2jt}PC_{ij(t-2)} + e_{ijt} \tag{34}$$

with city-level equations

$$\begin{aligned}
\pi_{0jt} &= \gamma_{00t} + \gamma_{01t}X_{1j(t-1)} + \gamma_{02t}X^2_{1j(t-1)} \\
&\quad + \gamma_{03t}X_{2j(t-1)} + \gamma_{04t}X^2_{2j(t-1)} + \ldots + \gamma_{010t}X^2_{5j(t-1)} + b_{0jt} && (35) \\
\pi_{1jt} &= \gamma_{10t} && (36) \\
\pi_{2jt} &= \gamma_{20t} && (37)
\end{aligned}$$

and time-level equations

$$\begin{aligned}
\gamma_{00t} &= \beta_{000} + b_{00t} && (38) \\
\gamma_{01t} &= \beta_{010} && (39) \\
\gamma_{02t} &= \beta_{020} && (40) \\
&\vdots \\
\gamma_{010t} &= \beta_{0100} && (41) \\
\gamma_{10t} &= \beta_{100} && (42) \\
\gamma_{20t} &= \beta_{200}, && (43)
\end{aligned}$$

with the fixed coefficients, β, and the random coefficients, b, in the Eqs. (35) and (38) to (43), with the b_{00t} the random effect for the intercept at time-level, and b_{0jt} the random effect for the intercept at city-level. Each random coefficient, b, is assumed to be normally distributed with a mean of zero and a specific standard error σ_2 and σ_3, $b_{00t} \sim N\left(0, \sigma_3^2\right)$ and $b_{0jt} \sim N\left(0, \sigma_2^2\right)$, which has to be calculated. The remaining residual e_{ijt} is also normally distributed with a mean of zero and a constant and unique standard error σ, $e_{ijt} \sim N\left(0, \sigma^2\right)$. Altogether, this leads to the following estimation equation:

$$
\begin{aligned}
Y_{ijt} - Y_{ij(t-1)} &= \beta_{000} + b_{00t} + \beta_{010} X_{1j(t-1)} + \beta_{020} X^2_{1j(t-1)} + \beta_{030} X_{2j(t-1)} + \dots \\
&\quad + \beta_{0100} X^2_{5j(t-1)} + \beta_{200} PC_{ij(t-2)} + b_{0jt} + \beta_{100} \left(Y_{ij(t-1)} - Y_{ij(t-2)} \right) + e_{ijt} \\
&= \beta_{000} + \beta_{100} \left(Y_{ij(t-1)} - Y_{ij(t-2)} \right) + \beta_{200} PC_{ij(t-2)} + \beta_{010} X_{1j(t-1)} + \dots \\
&\quad + \beta_{0110} X^2_{5j(t-1)} + b_{00t} + b_{0jt} + e_{ijt}.
\end{aligned} \tag{44}
$$

2.4.2 The Intercept-as-outcome Model

The more generalized version is the intercept-as-outcome model. The industry-level equation for the industrial growth is the same as for the basic multilevel model in Eq. (34). The equations at the city-level are the same as the equations in the basic multilevel model, except that the coefficient of the variable of interest, which is productivity change, is randomized. The Eqs. (35) and (36) are unchanged but Eq. (37) is modified:

$$
\begin{aligned}
\pi_{0jt} &= \gamma_{00t} + \gamma_{01t} X_{1j(t-1)} + \gamma_{02t} X^2_{1j(t-1)} \\
&\quad + \gamma_{03t} X_{2j(t-1)} + \gamma_{04t} X^2_{2j(t-1)} + \dots + \gamma_{010t} X^2_{5j(t-1)} + b_{0jt} \\
\pi_{1jt} &= \gamma_{10t} \\
\pi_{2jt} &= \gamma_{20t} + b_{2jt}.
\end{aligned} \tag{45}
$$

The time-level equations are also generalized by randomizing the coefficients in Eq. (35), which explains the intercept in the industry-level equation. Only the coefficients of the linear terms of the city-specific variables are randomized, to reduce the number of random effects; otherwise the analysis would not be computable in an adequate time. The equations for the coefficients at the time-level are:

$$\gamma_{00t} = \beta_{000} + b_{00t} \tag{46}$$
$$\gamma_{01t} = \beta_{010} + b_{01t} \tag{47}$$
$$\gamma_{02t} = \beta_{020} \tag{48}$$
$$\vdots$$
$$\gamma_{09t} = \beta_{090} + b_{09t} \tag{49}$$
$$\gamma_{010t} = \beta_{0100} \tag{50}$$
$$\gamma_{10t} = \beta_{100} \tag{51}$$
$$\gamma_{20t} = \beta_{200} + b_{20t}, \tag{52}$$

with the fixed coefficients, β, which might differ from those in the basic multilevel model because of the additional random coefficients, b, in the Eqs. (47), (49) and (52). Each random coefficient, b, is assumed to be normally distributed, with a mean of zero and a specific standard error that has to be calculated.

Altogether, this results in the simple intercept-as-outcome model

$$
\begin{aligned}
Y_{ijt} - Y_{ij(t-1)} &= \beta_{000} + b_{00t} + (\beta_{010} + b_{01t})\, X_{1j(t-1)} + \beta_{020} X^2_{1j(t-1)} \\
&\quad + (\beta_{030} + b_{03t})\, X_{2j(t-1)} + \ldots + \beta_{0100} X^2_{5j(t-1)} + b_{0jt} \\
&\quad + \beta_{100} \left(Y_{ij(t-1)} - Y_{ij(t-2)}\right) + (\beta_{200} + b_{20t} + b_{2jt})\, PC_{ij(t-2)} + e_{ijt} \\
&= \beta_{000} + \beta_{010} X_{1j(t-1)} + \beta_{020} X^2_{1j(t-1)} + \beta_{030} X_{2j(t-1)} + \ldots + \beta_{0100} X^2_{5j(t-1)} \\
&\quad + \beta_{100} \left(Y_{ij(t-1)} - Y_{ij(t-2)}\right) + \beta_{200} PC_{ij(t-2)} \\
&\quad + b_{00t} + b_{20t} PC_{ij(t-2)} + b_{01t} X_{1j(t-1)} + b_{03t} X_{2j(t-1)} + \ldots + b_{09t} X_{5j(t-1)} \\
&\quad + b_{0jt} + b_{2jt} PC_{ij(t-2)} + e_{ijt}. \qquad (53)
\end{aligned}
$$

2.4.3 The Intercept-and-slope-as-outcome Model

The intercept-and-slope-as-outcome model additionally explains the slope for the variable of interest, in my case the productivity change, which is π_{20t}. Eq. (45) becomes

$$
\pi_{20t} = \gamma_{20t} + \gamma_{21t} X_{1j(t-1)} + \gamma_{22t} X_{2j(t-1)} + \ldots + \gamma_{25t} X_{5j(t-1)} + b_{2jt} \qquad (54)
$$

with each $\gamma_{2jt}, j = 0, 1, ..., 5$ as a fixed coefficient $\gamma_{2jt} = \beta_{2j0}, j = 1, ..., 5$ except for γ_{20t} for which Eq. (52) holds. Please note, that only the linear and not the quadratic terms are added to explain variations of the effect (slope) of past productivity change on value added and employment growth, which results in the intercept-and slope-as-outcome model

$$
\begin{aligned}
Y_{ijt} - Y_{ij(t-1)} &= \beta_{000} + b_{00t} + (\beta_{010} + b_{01t})\, X_{1j(t-1)} + \beta_{020} X^2_{1j(t-1)} \\
&\quad + (\beta_{030} + b_{03t})\, X_{2j(t-1)} + \ldots + \beta_{011t} X^2_{5j(t-1)} + b_{0jt} + \beta_{100} \left(Y_{ij(t-1)} - Y_{ij(t-2)}\right) \\
&\quad + \left(\beta_{200} + \beta_{210} X_{1j(t-1)} + \beta_{220} X_{2j(t-1)} + \ldots + \beta_{250} X_{5j(t-1)} + b_{20t}\right) PC_{ij(t-2)} \\
&\quad + e_{ijt} \\
&= \beta_{000} + \beta_{010} X_{1j(t-1)} + \beta_{020} X^2_{1j(t-1)} + \beta_{030} X_{2j(t-1)} + \ldots + \beta_{011t} X^2_{5j(t-1)} \\
&\quad + \beta_{100} \left(Y_{ij(t-1)} - Y_{ij(t-2)}\right) + \beta_{200} PC_{ij(t-2)} + \beta_{210} X_{1j(t-1)} PC_{ij(t-2)} \\
&\quad + \beta_{220} X_{2j(t-1)} PC_{ij(t-2)} + \ldots + \beta_{250} X_{5j(t-1)} PC_{ij(t-2)} \\
&\quad + b_{00t} + b_{20t} PC_{ij(t-2)} + b_{01t} X_{1j(t-1)} + b_{03t} X_{2j(t-1)} + \ldots + b_{010t} X_{5j(t-1)} \\
&\quad + b_{0jt} + b_{1jt} PC_{ij(t-2)} + e_{ijt}, \qquad (55)
\end{aligned}
$$

with additional addends for the fixed effects resulting from explaining the slope. These fixed effects result from the interaction of the productivity change component and the city-specific variables. The computational details are explained in the Appendix D. The intercept-and-slope-as-outcome model tests therefore whether the variation in the slope of productivity change can be explained by the other variables.

2.4.4 Model Selection

The model selection approach is similar to that in Goedhuys and Srholec (2010) and standard in multilevel analysis. First, I estimate linear models by OLS estimation and heteroscedasticity-consistent standard errors are calculated to account for heteroscedasticity in general. The heteroscedasticity-consistent standard errors are HC3, as already mentioned in section 2.2, are introduced by MacKinnon and White (1985). By estimating OLS models with heteroscedasticity-consistent standard errors, it is possible to identify insignificant relationships and thus to reduce the number of coefficients which are estimated in the next steps. The OLS model is a reduced version of the intercept-and-slope-as-outcome multilevel model, which includes all city-specific explanatory variables as well as interaction terms of the city-specific variables with the Malmquist index and its components.

To compare the model fit of each model and estimation, the Pseudo R^2 of McFadden (1973) is calculated as

$$\text{McFadden} - \boldsymbol{R^2} = 1 - \ln L / \ln L_0, \tag{56}$$

with $\ln L$ as the log-likelihood of the actual model and $\ln L_0$ as the log-likelihood of the null model with only the intercept in the fixed effects and random effects part on each level (for multilevel models).

To verify the model and to check whether the additional variables add further explanatory power, many different measures can be used.

A general test for restrictions is the likelihood ratio test. For this, the likelihood of the more general model L_2 is divided by those of the more restricted model L_1. In general, the likelihood of the more unrestricted model is higher than the one of the restricted model. The test statistic is

$$LR = 2\ln\left(\frac{L_2}{L_1}\right) = 2\left(\ln L_2 - \ln L_1\right) \tag{57}$$

and is also always positive. Under the null hypothesis that the restricted model is sufficient, the likelihood ratio test statistic is χ^2 distributed with $k_2 - k_1$ degrees of freedom, where k_2 and k_1 are the number of parameters in the general model and the restricted model, respectively. Pinheiro and Bates (2000, chapter 2.4) show that the test can also be performed if both models are estimated by the restricted maximum likelihood (REML). The test enables us to test random effects but also the fixed effects similar to an F-statistic in OLS estimation, depending on the reference model.

Other possible instruments for evaluating the necessity of levels and random effects include information criteria as measures of the relative goodness of fit. I use the Akaikes information criterion (AIC) as well as the Bayesian information criterion (BIC), also known as the Schwarz information criterion. The criteria are generally formulated as

$$AIC = 2k - 2\ln L$$

and

$$BIC = k\ln n - 2\ln L,$$

with the value of the log-likelihood function and k, which is the number of estimates, as well as n, the number of observations, in the BIC. The value of the log-likelihood gives the goodness-of-fit and the number of estimates to reach that goodness-of-fit is added as a positive penalty term. The penalty term is needed, because more estimates increase the goodness of fit, yet induce uncertainty and cause over-fitting. Thus, the principle of parsimony is considered by minimizing the information criteria.

The intraclass correlation coefficient is one possible instrument which is commonly used in the multilevel literature and is based on the variances of the random effects in the basic multilevel model

$$ICC_3 = \frac{\sigma_3^2}{\sigma_2^2 + \sigma_3^2 + \sigma^2}. \tag{58}$$

Eq. (58) is the intraclass correlation at the third level (namely at the time-level), taken from West et al. (2007). For the second level the intraclass correlation is

$$ICC_2 = \frac{\sigma_2^2 + \sigma_3^2}{\sigma_2^2 + \sigma_3^2 + \sigma^2} \tag{59}$$

as described in Tabachnick and Fidell (2007) and discussed in Hox (2002). If the intraclass correlations are high, the correlation of the observation within that level is large. Unfortunately, there is no rule or distribution for any consideration in the test statistics.

Another way of testing the need for different levels is to look at the plots of the distribution of the observation at a specific level. The plot investigation also helps in the visualization of complicated multilevel models and is used, e.g., in Ieno et al. (2009). Generally, it is proposed for use in a protocol-based multilevel analysis as described in Zuur et al. (2009).

3 The Optimal Size of German Cities: An Efficiency Analysis Perspective

3.1 Motivation

There are many factors that encourage people and firms to settle in a city. Cities are centers for forces that improve quality of life, such as proximity to other people, jobs, recreational and shopping facilities and other institutions necessary for modern life. The proximity of these institutions is time-saving and increases leisure time; it also enhances convenience and utility for city-dwellers and thus helps to directly and indirectly increase productivity for firms in that city. These forces are mainly acknowledged as agglomeration externalities, which depend on the size of the specific urban agglomeration. Forces for agglomeration externalities include education facilities, health services, skilled jobs, infrastructure and social contacts supported by close relationships within a city. These affect the productivity of individuals and firms because closeness to related people and firms saves time and increases, or at least enhances, the flow of knowledge and ideas. Local education facilities improve the knowledge of the inhabitants and employees in firms. As is often pointed out, larger job market within a city which facilitates finding employment is another important aspect; it encourages participation in the production process as well as knowledge transfer from one firm to another. However, there are also signs of an urban overload as city size increases. There is evidence indicating that too many inhabitants in a certain area induce negative externalities in addition to the costs of urbanization. Recognized examples of these forces include pollution, intensive energy use, noise caused for example by traffic, high urban rents, as well as long and time-intensive journey to or from the working place. The factors affect productivity because overpopulation within a city produces negative externalities such as traffic jams and noise which decrease productivity by causing higher transportation costs and higher rents, or social friction in the labor market. These positive and negative aspects of urbanization are widely discussed and often referred to as average urban benefits, taking into account the agglomeration externalities on the one hand and average location costs on the other hand (Capello and Camagni, 2000, pp. 1484ff.). Of course these forces occur at specific levels of population or population densities. Although these effects may also arise in non-urban areas, they are mainly connected to urbanization effects.

Thus, a growing population in cities leads to increasing agglomeration effects, which are later neutralized by the negative effects of overpopulation. These considerations should obviously lead to an optimal city size. Contrariwise, it is observable that the largest cities in most industrialized countries in particular are continually growing. This fact presents a challenge to an optimal city size, if the optimal size is below the size of the largest city. Dascher (2002) shows that capital cities remain stable for a given population and set of institutional circumstances. The research questions that this chapter seeks to answer are therefore as follows: Is there an optimal city size in Germany? If there is, is the optimal city size within the range of the observed cities, or does it predict further growth of all cities? Does the optimal city size depend on the region in which the cities are located within Germany or is it constant for all cities?

A more detailed look into the industry data of the cities could lead to distinguishing different city types. Cities can be characterized by specialization in different industries and services as suggested by Henderson (1974). Different city types could explain a range of different optimal city sizes as a result of different requirements of specialized local industries. Cities benefit by specialization externalities which are often referred to as Marshall-Arrow-Romer-externalities (Marshall (1890), Arrow (1962) and Romer (1986)). An additional question is therefore: Does the optimal city size depend on the different types of cities as observed in urban hierarchy?

The purpose of the analysis in this chapter is to investigate the relationship between the efficiency and population size of German cities in a static setting. The efficiency in this context is the scale efficiency, which takes account of the specific size of the particular city. Therefore, the approach employed in this chapter is a two-stage process. At the first stage, the efficiency is measured in terms of scale efficiency, which involves estimating the efficiency of each city twice: once for constant returns to scale and once for variable returns to scale, and taking the ratio of both. The second stage investigates the relationship between the efficiency and the population size of the cities. The results show, that there is an optimal city size in Germany. The optimal city size is about 220,000 inhabitants, which is almost the mean of all German cities involved in this investigation. Furthermore, I use the bootstrap algorithms of Simar and Wilson (2007) to verify the two-stage approach. This confirms the results of the two-stage setup. In addition and although there are regional differences, optimal city size remains stable as the mean size of the inpected cities.

The analysis is organized as follows. The following section 3.2 presents the relevant literature. Section 3.3 gives a brief overview on the applied methods and section 3.4 describes the data used in the estimations. The empirical results are given and discussed in section 3.5. At the end of this chapter conclusions are drawn in section 3.6.

3.2 Literature Review

There are two separate branches of literature related to this analysis. One is concerned with optimal city size and the other investigates spatial efficiency or productivity. A brief review of the received literature in this regard is given in the following. The other aspect of the literature concerning optimal city size focuses on the size in terms of pure area size, see for example Henderson (1975). Since the present investigation deals with size in terms of population of the city, that aspect of the literature is not considered further.

The Henry George Theorem (George, 1880) is often applied for the analysis of optimal city size and has been tested to assess whether it leads to an analytic rule to test city sizes and thus to encourage cities to grow to an optimal size. The Henry George Theorem originally states that, in any Pareto optimal allocation, government spending on a pure local public good will be equal to land rents (Arnott, 2004, p. 1058).

Early insights into optimal city size with respect to the Henry George Theorem are given by Arnott and Stiglitz (1979). Their model of an optimal city size can be analyzed by the relation of aggregate land rents and expenditure on a pure local public good in a city. Furthermore, they outline conditions in which the model does not work, such as in cases of small economies and differing land rents. One result is that in large competitive economies with Pareto optimal distribution of economic activity and defined land rents the Henry George Theorem holds. Consequently Arnott and Stiglitz derived rules for ascertaining whether population size is optimal. For a single city, Arnott (1979) shows that for optimal city size the differential land rent has to be equal to the expenditure on the public good, which is the only incentive for urbanization. Arnott (2004) investigates whether the Henry George Theorem leads to a practical rule for optimal city size. Discussing the Henry George Theorem and presenting the disadvantages, Arnott works out a generalized version of the Theorem which allows for multiple outputs and multiple factors as well as for heterogeneous individuals. This version states that in an optimally-sized city the aggregate land rents have to be equal to the expenditure on the pure local public good. But for the generalized Henry George Theorem it is revealed that aggregate profits should be zero for each spatial unit of replication for any Pareto optimal allocation.

Based on available data it is difficult to estimate those models directly, as Arnott discusses in relation to the analysis of Kanemoto et al. (1996), who used the ideas of Henry George in their investigation of optimal city size. Kanemoto et al. use data for Tokyo with many restrictions on the Henry George Theorem. Kanemoto et al. (1996) investigate the size of Japanese cities with special emphasis on Tokyo. On the one hand, they are able to state that Tokyo is not too large. The conclusion is that there is no optimal size for cities within the observable range of city size in Japan. On the other hand, Arnott (2004) states that they do not use the main idea of Henry George, and thus it is still questionable whether an optimal city size in the terms of the Henry George Theorem can be established.

Combes et al. (2005) derive a U-form shaped cost-of-living curve, depending on population size, on the basis of different transport cost models. The minimum cost of living would then determine the optimal city size.

With respect to urban productivity, Sveikauskas (1975) and Segal (1976) show that Hicks-neutral productivity increases by city size. The critique by Moomaw (1981) makes clear that Sveikauskas´s estimates are biased in an upward direction as a result of omitting capital intensity or capital as explanatory variables. However, Sveikauskas (1975) estimates significantly positive linear correlation between city size and productivity for almost all manufacturing industries while also controlling for education and regional differences. Since he only investigates a linear relationship, Sveikauskas is only able to show that bigger cities have higher labor productivities. Thus, it would not result into an optimal city size below infinity or the whole citizenship. Segal (1976) on the other hand finds scale effects in cities by estimating production functions of the 58 largest US cities. Using OLS estimation he finds constant returns to scale for production output and labor productivity but also positive and significant effects on city size. Thus, metropolitan areas with two million and more inhabitants have significantly higher labor productivity compared to smaller metropolitan areas. Since Segal only investigates the largest cities the only viable result is that it is optimal for already large cities to grow further and therefore no optimal size exists within the range of observed city sizes.

Yezer and Goldfarb (1978) estimate that the optimal city size is in the range of 1.5 to 2.5 million inhabitants for US cities. They investigate the wage changes by region, occupation and population size and compare these with changes in firm efficiency related to changes in city size. The first effect reflects the household costs. The latter is based on Segal (1976) but with differing production functions for different industries and therefore different types of cities in line with Henderson (1974). Thus, the optimal city size is in the equilibrium of output value maximization of firms and household decisions based on average household costs.

A cross-section analysis for 58 Italian cities is given by Capello and Camagni (2000). They separately estimate average location benefits and costs as a function of city size as well as their squares and interactions with other variables, which include urban functional development and network integration level. Average location benefits and costs are calculated as unweighted sums of many different indicators (for instance the use of energy per capital as benefits and number of vehicles per square kilometers). With respect to average location benefits, they estimate an inverted U-shaped curve with a maximum of 361,000 inhabitants. Concerning urban overload by investigating average location costs, they estimate a U-shaped curve with a minimum of 55,500 inhabitants. Unfortunately they do not show in which range of city size average location benefits are above average location costs, which is obviously a result of the somewhat questionable measurement which renders them incomparable.

As Alonso (1971) points out, it is not minimum of the average urban costs which are of interest but the point where marginal costs equals marginal product for cities. At this point, the level at which average products minus average urban costs is of the highest positive amount for the whole economy, because at this level disposable income is maximized. Thus, it is not a question of optimal city size but of efficient city size! Comparing Germany, Japan, and the US he shows that the highest excess of average product over average public costs occurs in the population size class of 200,000 and was larger for Germany in 1964 without further disaggregation.

A brief literature overview on the topic of optimal city sizes developed by minimizing urban public costs is given by Richardson (1972). He concludes that there are many mostly philosophical approaches in the analysis that do not address a specific range of efficient city sizes or even a measurable dimension for city size. This range would encourage a critical minimum size as well as a theoretical maximum size.

On the other hand, there is plenty of literature regarding non-parametric estimation of efficiency and productivity. Because this analysis deals with cities as aggregate of many firms and households, a brief review of literature dealing with efficiency analysis for spatial decision-making units now follows. A good overview is given by Worthington and Dollery (2000), who also include efficiency analysis for firms and specific industries.

The work of Charnes et al. (1989) is closely related, who employ DEA techniques for analyzing the economic performance of Chinese cities. They also investigate returns to scale for depicting the most productive scale size, which is introduced by Banker (1984). The results show that Shanghai and smaller cities define the most productive scale size but these results are not linked with population figures as a measure of size.

Susiluoto and Loikkanen (2001) and Loikkanen and Susiluoto (2004, 2006) investigate Finnish regions and cities by DEA methods. In Susiluoto and Loikkanen (2001) it is obvious, even it is not the goal of that work, that bigger cities, including Helsinki, achieve the highest DEA efficiency scores while the lowest results are examined for smaller cities between 1988 and 1999. Although it is not empirically supported and there are indicators of geographical (north-south) patterns, the results support agglomeration effects. Loikkanen and Susiluoto (2006) estimate a significantly negative correlation between DEA results and population size for the whole period of investigation from 1994 to 2002, which stands in contrast to the observations of the former work. Furthermore, Loikkanen and Susiluoto (2004) use Tobit regressions, which result in positive estimates of population size in accounting for inefficiency. Therefore, smaller cities are more efficient in Finland according to this study.

Halkos and Tzeremes (2010) analyze Greek prefectures by DEA methods and also present population density and changes in it. This reveals that the most efficient areas are not the most densely populated, although it is not an analysis focusing on cities. Major changes in the industry structure and institutional setups as well as EU regulations in Greece cause some doubts regarding the results, because they do not control for those changes.

Altogether, there is a great deal of evidence for optimal city size but also for strong agglomeration effects, which could dominate increasing urbanization costs for the whole range of possible population sizes. This would lead to continually increasing efficiency by city size. The aim of this analysis is to merge both approaches, which are efficiency analysis and the investigation of optimal city size, and apply them to data for the free German cities to produce a measure of efficient city size.

3.3 Theory

The investigation in this chapter is mainly implemented by a two-stage analysis. In the first stage, the specific efficiency is measured for every unit of interest, i.e., the cities in Germany. In a second stage, these efficiency measurements are taken as given and their relationship with population size is examined in different regression setups. These setups incorporate quadratic, cubic in OLS, robust linear fit and LMS as well as non-parametric models.

The efficiency measurement is presented in subsection 2.1.1. In contrast to the general specification in subsection 2.1.1, efficiency is estimated for the sum of all industries of each city and not industry specific. The five major industry sectors with one output and two input factors are accumulated to an entire sector. Solving all models for all cities results in one estimate for the technical efficiency for CRS and one for VRS for every city. The scale efficiency SE_j in city j is calculated as the division of the technical efficiency for CRS divided by the technical efficiency for VRS

$$SE_j = \frac{\theta_{CRS,j}}{\theta_{VRS,j}}. \tag{60}$$

Given that the technical efficiency for CRS is smaller or equal to the technical efficiency for VRS calculated as Shephard distance functions, the measurement for scale efficiency is always in the range between zero and one, with one for a scale efficient and below one for a scale inefficient city. The detailed calculated results are listed in table A1 in the Appendix. Note, that the measurements for technical efficiency are in terms of Shephard (1970) with output-orientation and they are therefore smaller or equal to one, which represents the proportion of efficiency. The measurement for scale efficiency gives the percentage of efficiency of the city. Furthermore, a scale efficiency of one indicates the most productive scale size measured in output quantities caused by the output-orientation of the DEA (Banker and Thrall, 1992). The most productive scale size is characterized either by one city or a range of cities. Cities with a scale efficiency coefficient of less than one are not of efficient size and are either too small or too large. However, it should be taken into consideration that scale efficiency does not imply that the city or the sectors within the city are technically efficient by CRS or VRS. This can be seen in table A1 in the Appendix, which points out that Wolfsburg has the value of one for scale efficiency but the same technical inefficiency for CRS and VRS. All other scale efficient cities are technically efficient for CRS and VRS. These scale efficient cities could be exclusively used to determine the optimal or efficient city size. However, because this takes into account a range of city sizes and because some extant measurement errors are present, the interval is reduced to one sole measure for optimal city size by a linear regression. To be clear, the scale efficiency is the ratio of the average productivity at that point to the average productivity at the most productive scale size. Thus, it measures to what extent the average productivity could be improved by achieving optimal size. Therefore, scale efficiency is the appropriate measure for gaining statements about optimal size.

Since the productivity measurement are estimated in output-orientation, the optimal size with respect to average productivity is the output variable. The output variable is gross value added of the economic sectors within each city. The optimal city size should therefore be estimated in terms of gross value added. On the one hand, cities larger than the optimal size are able to improve average productivity by reducing gross value added because they produce in the area characterized by decreasing return to scale. On the other hand, cities smaller than the optimal size are able to produce with higher average productivity by increasing gross value added since they are producing with increasing return to scale. However, gross value added of cities is unhandy, since it is commonly not known. Bettencourt et al. (2007) estimate that gross value added is increasing more as city size increases for many countries. The rule is widely known as the 15% rule, because gross value added improves by additional 15% as cities' size doubles. The empirical analysis is an estimation of the logarithm of gross value added on the logarithm of city size. Applying the analysis for the free German cities results in an estimate of 1.026 with a standard error of 0.034. The 90% confidence interval for the coefficient is therefore from 0.97 to 1.081. This interval shows, that gross value added is not significantly improving additionally as city size increases, which contradicts the 15% rule for the cities within the sample. The following figure illustrates the relationship.

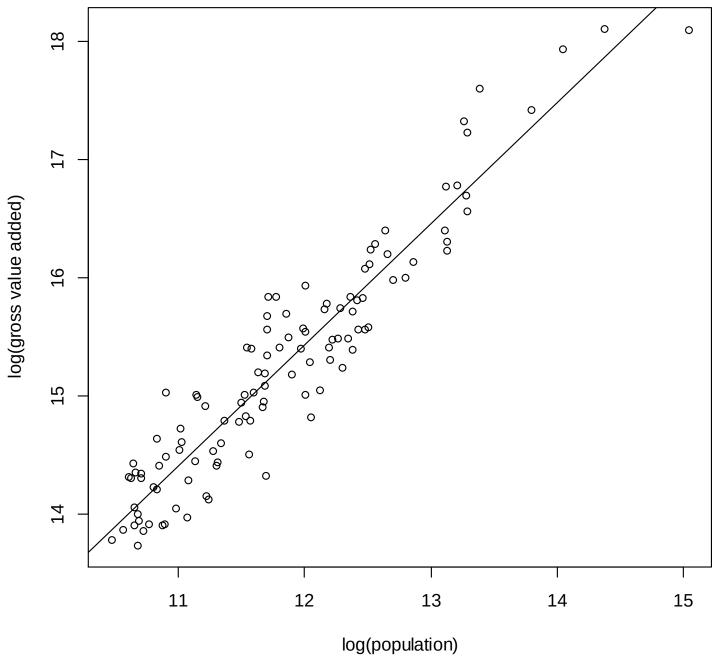

Figure 3.1: Gross Value Added against City Size

As figure 3.1 illustrates, the logarithm of gross value added is directly proportional to the logarithm of cities' population. An increase of city size by one percent is corresponding by an increase of gross value added by one percent. Therefore, optimal city size can be estimated in terms of population.

To test whether population size is related to scale efficiency, two different linear models are applied. These models need a quadratic term for population size to estimate optimal population size with respect to scale efficiency or average productivity, respectively. Thus, an optimal city size exists when the linear term has a positive coefficient and the quadratic term has a negative coefficient. In addition, cities population distribution in Germany follows an exponential rule, i.e., the number of cities decreases by a constant when the population increases by that constant, an effect commonly known as Zipf's law described by Zipf (1949). Zipf's law is also called the rank-size rule which is described for instance in Richardson (1972) and Nitsch (2005). Zipf's law states that the rank of a city is described by the number of inhabitants of the largest city divided by the population of that city. Thus, the distribution of cities can be described by an exponential function. Therefore, it is appropriate to use logarithms of population in the specifications in order to avoid the biggest cities leveraging the estimates caused by the exponential distribution. The model, which is tested, is in specification I:

$$SE_j = \beta_0 + \beta_1 \log(population_j) + \beta_2 \left(\log(population_j)\right)^2 + e_j, \tag{61}$$

with e_j the residuals in city j. The empirical results for Eq. (61) should result in an inverted U-shaped functional form with a maximum point. Thus β_1 should be significantly positive and β_2 significantly negative.

To check the correctness of the quadratic specification in specification I, a cubic function is also estimated. This cubic model is represented by specification II in form

$$SE_j = \beta_0 + \beta_1 \log(population_j) + \beta_2 \left(\log(population_j)\right)^2 + \beta_3 \left(\log(population_j)\right)^3 + e_j, \quad (62)$$

with β_3 the estimator for the cubic term, which should not be significant if the correct model is quadratic. Therefore, the test for the quadratic model is whether the cubic term is significant on scale efficiency or not.

That two-stage setup should be considered with caution for several reasons. The efficiency scores are serially correlated, because they are calculated with all observations and depend on the frontier, which is built on few other cities. Thus the errors within the efficiency scores are serially correlated. Since the inputs and outputs are also correlated with the explanatory variables (i.e., the logarithm of population and its squares, see Eq. (60)), these explanatory variables are correlated with the errors in the efficiency scores (see Simar and Wilson (2007)). To overcome the difficulties of the inherent defects of estimation within the two-stage setup, Simar and Wilson (2007) propose two algorithms. Both algorithms are formulated for (Farrell, 1957) efficiency measures in output-orientation. Those measures are equal and larger than unity, which results in left truncations at unity within both algorithms. However, my efficiency scores for each city are the estimates of Shephards output distance measures. The corresponding scale efficiency measures in Eq. (60) are between zero and unity, where zero is not possible for producing cities and unity is the score for cities with efficient size. In order to apply these algorithms I prepared two different setups. On the one hand, I calculated the inverse of the scale efficiency scores or the scale inefficiency scores, respectively, thereby essentially using algorithms as described in Simar and Wilson (2007, pp. 41-43). On the other hand, I took the scale efficiency scores as calculated in Eq. (60) and changed both algorithms to account for the two truncation points for the scale efficiency scores at zero and unity.

First, both algorithms are explained for the scale inefficiency measures, which are larger or equal unity and thus similar to the approaches described by Simar and Wilson (2007).

Algorithm 1:

1. Compute scale inefficiency $(1/SE_j)$ measures for all cities $j = 1, ..., n$ by the inverse of Eq. (60), on the basis of the efficiency scores for constant returns to scale and variable returns to scale for each city.

2. Estimate the coefficients $\widehat{\boldsymbol{\beta}}$ and $\widehat{\sigma}$ of a truncated regression similar to Eq. (61)

   $$(1/SE_j) = \beta_0 + \beta_1 \log\left(population_j\right) + \beta_2 \left(\log\left(population_j\right)\right)^2 + e_j$$

   but only for those m cities $(m < n)$ with scale inefficiency score larger than unity by the maximum likelihood method.

3. Loop over the next three steps for 10,000 times to obtain bootstrap estimates of $\widehat{\boldsymbol{\beta}}^*$ and $\widehat{\sigma}^*$ in a set $\mathcal{A} = \left\{ \left(\widehat{\boldsymbol{\beta}}^*, \widehat{\sigma}^* \right)_b \right\}_{b=1}^{10{,}000}$:

   (a) Draw ε_j for each $j = 1, ..., m$ from the $N\left(0, \widehat{\sigma}^2\right)$ distribution with left truncation at $\left(1 - \boldsymbol{z}_j \widehat{\boldsymbol{\beta}}\right)$, with $\boldsymbol{z}_j$ the explanatory variables in Eq. (61) containing the intercept, logarithm of population, and the squared logarithm of population.

   (b) Compute $(\widehat{1/SE_j}) = \boldsymbol{z}_j \widehat{\boldsymbol{\beta}} + \varepsilon_j$ for each $j = 1, ..., m$.

   (c) Estimate $\widehat{\boldsymbol{\beta}}^*$ and $\widehat{\sigma}^*$ of a truncated regression of Eq. (61) only for the cities with scale inefficiency of step 3b larger than unity by maximum likelihood method.

4. Construct the estimated confidence intervals of the estimates of $\widehat{\boldsymbol{\beta}}$ and $\widehat{\sigma}$ by the bootstrap values of $\mathcal{A}$.

Also algorithm 2 of Simar and Wilson (2007) is implemented for scale inefficiency measures, which are larger than unity. But the following algorithm differs from those described by Simar and Wilson (2007, p. 42f.) because their algorithm 2 computes pseudo observations for the output $\boldsymbol{y}^*$ in the first bootstrap estimation to ascertain bias-corrected efficiency measures. Those pseudo outputs $\boldsymbol{y}^*$ represent the reference observations for the estimation of the efficiency scores for all cities by solving the linear program as described in Eq. (5) for CRS and Eq. (6) for VRS. Since $\boldsymbol{y}^*$ is calculated as the ratio of the first and randomly drawn efficiency scores in the original work of Simar and Wilson (2007), these pseudo-observations are inherent the new pseudo-frontier. However, this calculation is not appropriate for scale efficiency measures which are simultaneously ratios of efficiency measures and do not project the observations onto the frontier. Once again, $\boldsymbol{y}^*$ could not be calculated within two separate bootstrap estimations, since scale efficiency measures are computed simultaneously by the efficiency measures for constant and variable returns to scale. The first bootstrap estimation is therefore replaced by the bias-correction of Simar and Wilson (1998). Altogether, the modified algorithm 2 for scale inefficiency measures is as follows:

1. Compute bias-corrected scale inefficiency $(\widehat{\widehat{1/SE_j}})$ measures for all cities $j = 1, ..., n$ by the inverse of Eq. (60), on the basis of the bias-corrected efficiency scores for CRS and VRS for each city. The bias-correction is implemented by the bootstrap method of Simar and Wilson (1998) with 10,000 replications.

2. Estimate the coefficients $\widehat{\boldsymbol{\beta}}$ and $\widehat{\sigma}$ of the truncated regression of the equation

$$(\widehat{\widehat{1/SE_j}}) = \beta_0 + \beta_1 \log\left(population_j\right) + \beta_2 \left(\log\left(population_j\right)\right)^2 + e_j$$

   but only for those m cities ($m < n$) with scale inefficiency score larger than unity by maximum likelihood method.

3. Loop over the next three steps 10,000 times to obtain a set of bootstrap estimates of $\widehat{\boldsymbol{\beta}}^*$ and $\widehat{\sigma}^*$ $\mathcal{B} = \left\{ \left(\widehat{\boldsymbol{\beta}}^*, \widehat{\sigma}^* \right)_b \right\}_{b=1}^{10,000}$:

   (a) Draw ε_j for each $j = 1, ..., n$ from the $N\left(0, \widehat{\sigma}^2\right)$ distribution with left truncation at $\left(1 - z_j \widehat{\boldsymbol{\beta}}\right)$, with $\boldsymbol{z}_i$ the explanatory variables in Eq. (61) including the intercept, logarithm of population, and the squared logarithm of population.

   (b) Compute $(\widehat{\widehat{1/SE_j}})^* = \boldsymbol{z}_j \widehat{\boldsymbol{\beta}} + \varepsilon_j$ for each $j = 1, ..., n$.

   (c) Estimate $\widehat{\boldsymbol{\beta}}^*$ and $\widehat{\sigma}^*$ of a truncated regression of Eq. (61) only for the cities with scale inefficiency of step 3b larger than unity by maximum likelihood method.

4. Construct the estimated confidence intervals of the estimates $\widehat{\boldsymbol{\beta}}$ and $\widehat{\sigma}$ by the bootstrap values of $\mathcal{B}$.

The confidence intervals in both algorithms are for each estimate $\widehat{\beta}_k$, $k = 0, 1, 2$, and $\widehat{\sigma}$ is found by the bootstrap sets of $\mathcal{A}$ and $\mathcal{B}$ with values for $a^*_{0.05}$ and $b^*_{0.05}$ to fulfill

$$\Pr\left[-b^*_{0.05} \leq \left(\widehat{\beta}^*_k - \widehat{\beta}_k \right) \leq -a^*_{0.05} \right] \approx 0.95. \tag{63}$$

The approximation is improving as the number of bootstrap replication increases. Thus the number of replications was set to 10,000, which is already ten times larger than the suggested number in Simar and Wilson (2007). That leads to the confidence interval $\left[\widehat{\beta}_k + a^*_{0.05}, \widehat{\beta}_k + b^*_{0.05}\right]$ for each estimate $k = 0, 1$ and 2.

Second, both algorithms are explained for the scale efficiency scores as given and used in the previous introduction. One should therefore keep in mind that these scores are between zero and unity. These boundaries have to be considered in both algorithms. Furthermore, scale efficiency scores are the ratio of the efficiency scores by CRS and VRS and thus the modified algorithm 2 has to be considered.

The modified algorithm 1 is:

1. Compute scale efficiency measures SE_j for all cities $j = 1, ..., n$ by Eq. (60), on the basis of the efficiency scores for CRS and VRS for each city.

2. Estimate the coefficients $\widehat{\boldsymbol{\beta}}$ and $\widehat{\sigma}$ of a truncated regression of Eq. (61)

$$SE_j = \beta_0 + \beta_1 \log (population_j) + \beta_2 \left(\log (population_j)\right)^2 + e_j$$

   but only for those m cities ($m < n$) with a scale efficiency score smaller than unity by maximum likelihood method.

3. Loop over the next three steps for 10,000 times to obtain bootstrap estimates of $\widehat{\boldsymbol{\beta}}^*$ and $\widehat{\sigma}^*$ in a set $\mathcal{C} = \left\{\left(\widehat{\boldsymbol{\beta}}^*, \widehat{\sigma}^*\right)_b\right\}_{b=1}^{10,000}$:

   (a) Draw ε_j for each $j = 1, ..., m$ from the $N\left(0, \widehat{\sigma}^2\right)$ distribution with left truncation at $-\boldsymbol{z}_j\widehat{\boldsymbol{\beta}}$ and right truncation at $\left(1 - \boldsymbol{z}_j\widehat{\boldsymbol{\beta}}\right)$, with $\boldsymbol{z}_j$ the explanatory variables in Eq. (61) including the intercept.

   (b) Compute $\widehat{SE_j} = \boldsymbol{z}_j\widehat{\boldsymbol{\beta}} + \varepsilon_j$ for each $j = 1, ..., m$. Both truncation points in 3a ensure that the resulting random scale efficiency score is in the range between zero and unity.

   (c) Estimate $\widehat{\boldsymbol{\beta}}^*$ and $\widehat{\sigma}^*$ of a truncated regression of Eq. (61) only for these cities with scale efficiency value of step 3b really smaller than unity (unequal one) by maximum likelihood method.

4. Construct the estimated confidence intervals of the estimates of $\widehat{\boldsymbol{\beta}}$ and $\widehat{\sigma}$ by the bootstrap values of $\mathcal{C}$.

Algorithm 2 is:

1. Compute bias-corrected scale efficiency measures $\widehat{\widehat{SE_j}}$ for all cities $j = 1, ..., n$ by Eq. (60), on the basis of the bias-corrected efficiency scores for constant returns to scale and variable returns to scale for each city. The bias-correction is implemented by the bootstrap method of Simar and Wilson (1998) with 10,000 replications.

2. Estimate the coefficients $\widehat{\boldsymbol{\beta}}$ and $\widehat{\sigma}$ of the truncated regression of the equation

$$\widehat{\widehat{SE_j}} = \beta_0 + \beta_1 \log (population_j) + \beta_2 \left(\log (population_j)\right)^2 + e_j$$

   but only for those m cities ($m < n$) with scale efficiency score smaller than unity by maximum likelihood method.

3. Loop over the next three steps 10,000 times to obtain a set of bootstrap estimates of $\widehat{\boldsymbol{\beta}}^*$ and $\widehat{\sigma}^*$ $\mathcal{D} = \left\{ \left(\widehat{\boldsymbol{\beta}}^*, \widehat{\sigma}^* \right)_b \right\}_{b=1}^{10,000}$:

   (a) Draw ε_j for each $j = 1, ..., n$ from the $N\left(0, \widehat{\sigma}^2\right)$ distribution with left truncation at $-z_j \widehat{\boldsymbol{\beta}}$ and a right truncation at $\left(1 - z_j \widehat{\boldsymbol{\beta}}\right)$, with $\boldsymbol{z}_j$ the explanatory variables in Eq. (61) including the intercept.

   (b) Compute $\left(1/\widehat{SE_j}\right)^* = \boldsymbol{z}_j \widehat{\boldsymbol{\beta}} + \varepsilon_j$ for each $j = 1, ..., n$. Both truncation points in 3a ensure that the resulting random scale efficiency score is in the range between zero and unity.

   (c) Estimate $\widehat{\boldsymbol{\beta}}^*$ and $\widehat{\sigma}^*$ of a truncated regression of Eq. (61) only for these cities with scale efficiency score of step 3b really smaller than unity (unequal one) by maximum likelihood method.

4. Construct the estimated confidence intervals of the estimates $\widehat{\boldsymbol{\beta}}$ and $\widehat{\sigma}$ by the bootstrap values of $\mathcal{D}$.

The main difference in both algorithms is that algorithm 1 estimates the coefficients only for those observations unequal to one (thus either larger or smaller than unity) and algorithm 2 uses bias-corrected efficiency scores and uses bootstrap estimates for all observations.

3.4 Data

In this analysis the data set for 112 NUTS3-districts is used, which are classified and explained in subsection 2.1.[8] As a consequence of the uncertainty in the capital estimation, only the average of the last five years is used in further estimations, which implies that the estimates for capital in the first nine years before the year 2004 are out of consideration.

Population figures are also taken from the regional data base of the Statistical Offices in Germany. A person is only counted in a city's population if the person has the principal residence within that city. So the figure does not account for people with secondary residence in order to avoid double counting, although many people have a secondary residence in a city and are part of those productive employees. Nevertheless, the use of the population figures for the number of inhabitants within a city is acceptable, because people who spend more than half of their time in the city are required to have their principal residence in that particular city.

All variables in the analysis of section 3.5 are used as the arithmetic average from 2004 until 2008. Descriptive statistics are given in table 3.1 with value added and capital stock given in thousand Euros and labor force and population given in thousands.

Table 3.1: Descriptive statistics

variable	Min.	1st Quartile	Median	Mean	3rd Quartile	Max.	sd
value added	919	1,875	3,975	7,740	7,372	73,390	12,470.1
capital	100.4	4,535	10,140	26,410	25,610	282,000	46,803.
labor force	18.42	43.33	76.12	139.70	137.30	1,551.00	209.38
population	35.28	64.67	120.60	231.70	239.80	3,396.00	387.15

Table 3.1 shows that there are many small cities with low average value added in the total industry in the years 2004 till 2008 as well as low capital stock, labor force, and population in the time span. This

[8] A list of the included cities and their scale efficiency scores is given in the Appendix A.

distribution results in a median of each of these variables which is much lower than the respective mean. The median is almost half the size of the respective mean for each variable and the mean is in the fourth quartile except for population. This indicates that the largest cities are of such a size that they have a strong influential power on the estimation of the mean and consequently on the standard deviation (sd). The descriptive statistics indicate the skewness of the data, which results in a heteroscedastic distribution of the data, with a decreasing variance in city size caused by a plurality of small cities. The skewness produced by these cities does not affect the efficiency analysis, which relies on relative measurements. For further analysis the data has to be transformed to become narrower. That is done by taking the logarithm of population (compare figure 3.1). Furthermore, all input variables as well as value added as output are non-negative as required in DEA.

3.5 Empirical Results

In this section the estimation results for population size on scale efficiency are presented and discussed.

First of all, tests have to be carried out to determine whether the underlying production function is characterized by CRS or VRS. If the production function can be described by CRS it is not appropriate to estimate the DEA with CRS and thus no scale efficiency rate can be measured. Simar and Wilson (2002) propose two different non-parametric tests for returns to scale by bootstrapping algorithms. The two approaches involve either examining all cities together by measuring the mean of scale efficiency over all cities or examining each city separately. With both approaches they propose to first test CRS as the most restricted production function and afterwards non-increasing returns to scale. These two possibilities were estimated for both null hypotheses. Thus, the first null-hypothesis is that the production function follows CRS and the second null-hypothesis is that the production function follows non increasing returns to scale. By testing each city separately, it was not possible to reject the null hypothesis of CRS for only three cities. Moreover, in the case of only eight cities could the null hypothesis of the second test of non-increasing returns to scale not be rejected. That strongly supports the appropriate use of the DEA with VRS and the possibility to calculate scale efficiency measures. These findings are also supported by the second approach with the mean of scale efficiency of the bootstrap algorithm for all cities simultaneously.

Furthermore, an illustration (figure A1) of scale efficiency is given in the Appendix, where the largest cities with a population size of over half a million inhabitants are indicated by boxes. The figure demonstrates the local distribution of the cities with their efficiency scores as well as the regional distribution of the largest cities. In addition, it can be seen in the figure that the largest cities are not necessarily the cities with highest efficiency scores. There is furthermore no specific region in Germany that only locates cities with low efficiency score and thus it is obvious that the efficiency scores are not asymmetrically distributed over the German regions.

3.5.1 Empirical Methods

The following estimations are performed by several methods. Since there are just four observations with the value of one for scale efficiency (compare table A1 in the Appendix) the estimations can be performed by normal regressions and do not have to be performed by Tobit or Logit regressions for truncated observations. These methods involve OLS, LMS, and robust fit MM-estimations as presented in section 2.2. Within the estimation model the number of coefficients is three in the estimation of specification I and four for estimation of specification II for all 112 cities.

3.5.2 Results for Entire Germany

Table 3.2 shows the results for both specifications in Eq. (61) and Eq. (62) and for robust fit (MM), OLS, and least median of squares (LMS) estimation. All computations are performed with R using the package `FEAR` for DEA as well as `robustbase`, which is covered by the book Maronna et al. (2006), for the non-parametric methods and least median of squares estimations. The R-package `FEAR` is described in Wilson (2008).

Table 3.2: Regression results for specification I and II

	specification I			specification II		
	MM	OLS	LMS	MM	OLS	LMS
Intercept	-0.844 *	-2.625 *	0.088	-3.948	0.659	6.329
	(0.476)	(1.544)	(0.357)	(4.544)	(27.058)	(4.141)
ln(*population*)	0.297 ***	0.598 **	0.153 **	1.063	-0.200	-1.360
	(0.076)	(0.261)	(0.059)	(1.093)	(6.772)	(1.018)
$(\ln(population))^2$	-0.012 ***	-0.025 **	-0.006 ***	-0.075	0.039	0.115
	(0.003)	(0.011)	(0.002)	(0.087)	(0.563)	(0.083)
$(\ln(population))^3$				0.002	-0.002	-0.003
				(0.002)	(0.016)	(0.002)
R^2	0.195	0.239	0.106	0.169	0.241	0.134

Significance codes: '***', '**', '*' significant up to 1%, 5%, and 10%, respectively.
Standard errors are below the estimates in parentheses.

Table 3.2 shows that the linear term and the quadratic term for logarithm population size is significant, with at least a five percent level of significance for every approach. Th sensitivity with respect to outliers explains the differences of the estimates of OLS approach. Furthermore, the estimates are of the type expected so that population has an inverse U-shaped distribution on scale efficiency. The results for specification II are not significant at all. Therefore, collinearity diagnostics have been computed, such as the condition number of the X matrix for specification II as well as variance inflation factors for both specifications (Fox and Monette (1992)). Both diagnostics indicate that collinearity is a problem in the underlying data by a condition number larger than 100 and high variance inflation factors for the population size variables. This is not surprising since the logarithm of population is in the range of between 10 and slightly above 15 which results in almost proportional quadratic and cubic terms. Altogether, the regression results in table 3.2 indicate that specification I is preferable and the distribution is quadratic with an inverted U-shaped design. Additionally, a regression equation specification error test (RESET, Ramsey (1969)) has been performed, which could not reject the null hypothesis of no misspecification for specification I.

The coefficient of determination R^2 is calculated as described in Hayfield and Racine (2008) by

$$R^2 = \frac{\left(\Sigma_{j=1}^{n}\left(SE_j - \overline{SE}\right)\left(\widehat{SE}_j - \overline{\widehat{SE}}\right)\right)^2}{\Sigma_{j=1}^{n}\left(SE_j - \overline{SE}\right)^2 \Sigma_{j=1}^{n}\left(\widehat{SE}_j - \overline{\widehat{SE}}\right)^2}, \tag{64}$$

where the fitted values for the regressand have a different mean ($\overline{\widehat{SE}}$) to the observations ($\overline{SE}$), in which case the sum of the residuals is not equal to zero. The robust estimations leave some observations out of recognition or down weight these observations, respectively. As a result the estimated errors do not have a zero mean entailing the use of Eq. (64), with the mean of the fitted values instead of the mean of the observed values as stated in Hayfield and Racine (2008).

By the definition in Eq. (64) the coefficient of determination is the squared correlation of the observed regressand to the fitted values of the regressand. Because the correlation is in the range between minus one and plus one, the resulting coefficient of determination is between zero and one. It should be noticed that this coefficient of determination is used in the cases of the robust linear fit estimations (MM) as well as the LMS estimations. These estimations are more robust than the OLS estimations in preventing some violations of its underlying assumptions, i.e., the normal distribution of the error term, which implies no outliers. In some cases where outliers are present, the robust estimations better match most observations except the outliers, which results in lower coefficients of determination.

The maximum of scale efficiency with respect to the logarithm of population in the case of MM-estimation is at $\frac{0.296831}{2 \cdot 0.012059} = 12.30745$ (with exact figures) or about 221,338 for population in total. In case of the OLS estimation the optimal city size is $\frac{0.598295}{2 \cdot 0.024885} = 12.02120$ (with exact figures) or about 166,242 for population in total, and for LMS estimation the result is $\frac{0.152813}{2 \cdot 0.006499} = 11.75665$ (with exact figures) or about 127,600 for population in total, respectively. Thus, the maximum points are always in the range of observed population size. Figure 3.2 illustrates the fitted values for specification I as well as a kernel fit estimation.

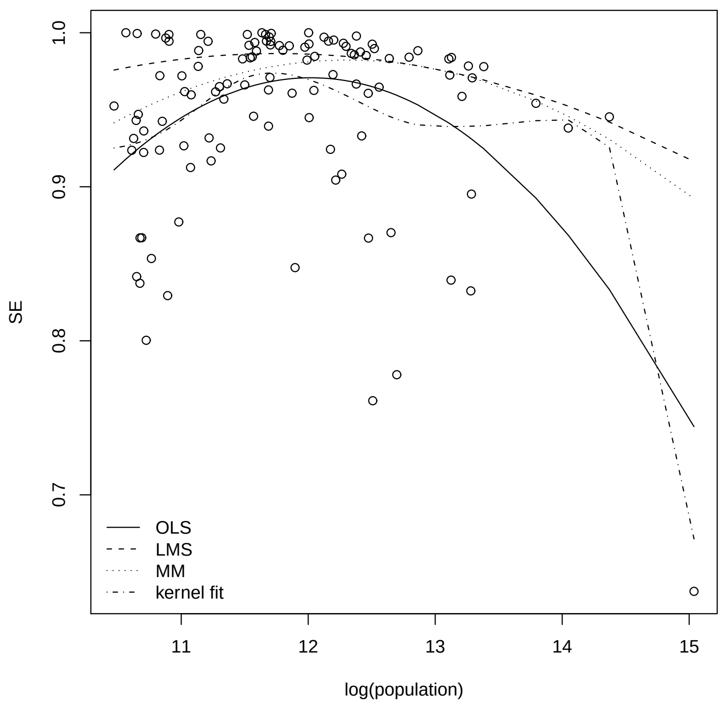

Figure 3.2: Fitted Estimates for Quadratic Models and Kernel Regression

For the non-parametric kernel fit regression a bandwidth has to be chosen. This bandwidth is fixed at 0.32 with respect to the underlying data in logarithm of population as the only explanatory variable by least squares cross-validation. A kernel function is a weighting function for the observation and the weights depend on the bandwidth. Since the underlying explanatory variable is continuous, a second order Gaussian kernel is implemented as described by Hayfield and Racine (2008). The computation is performed with R and the package `np`, which is explained by Hayfield and Racine (2008). For more details on kernel functions see Aitchison and Aitken (1976) or Li and Racine (2003).

As can been seen from the figure 3.2, the global maximum point for the kernel fit regression is below 12 for the logarithm of population and thus almost the same amount as the LMS estimation. There is a local maximum point at over 14 for the logarithm of population, which results in a spurious outcome as a result of the data sparsity at this size range and which can be explained by the 4th to 2nd largest cities.

Figure 3.2 points out the estimated graph and the maximum points of each approach. It also gives reasons for the specific results. For instance the OLS estimation is influenced by a small number of inefficient observations with low population size as well as the largest observation, which is the capital Berlin, with relatively low scale efficiency. Furthermore, the maximum points are around 12 for the logarithm of population, or 160,000 in total population size. Tukey (1979, p. 103) writes: "It is perfectly proper to use both classical and robust/resistant methods routinely, and only worry when they differ enough to matter. BUT when they differ, you should think HARD.". Thus, the question is whether the differences in the optimal city size are of a reasonable magnitude, and if this is the case, which result is more trustworthy. The figure demonstrates that there are many inefficient observations for small cities with populations of less than 100,000 and for cities with populations between 270,000 and 730,000 inhabitants. These observations influence the OLS estimation (the fitted or estimated scale efficiency is not as high as for the robust estimations) and cause heteroscedasticity[9] and contains outliers. Therefore, the OLS estimation does not seem to be appropriate for these observations. In addition, the kernel regression fit has two maximum points, with the global maximum point at below 12 and a local maximum point at over 14 for the logarithm of population. The local minimum point between 12 and 13 for the logarithm of population is caused by the inefficient observation in the range between 12.5 and 13.5. Thus, the kernel regression also seems to be inadequate to describe the observation. The observation may be different for other bandwidth but such a case is not further estimated in this analysis because its non-parametric character prevents further interpretations.

The robust estimations, especially the robust linear MM-estimation, fit the observations best. Therefore, an optimal city size of about 220,000 inhabitants, which is the result for the MM-estimation, is most proper for these observations. These findings are also supported by the empirical results in table 3.2, which states that the robust linear fit MM-estimator in specification I has a higher coefficient of determination (R^2) than the LMS model and is only slightly below the R^2 of the OLS estimation. Interestingly the maximum point at about 220,000 inhabitants is almost the mean of the observed cities' population in Germany (compare with table 3.1).

3.5.3 Validation of the Results by Bootstrap Algorithms for Entire Germany

As already explained in section 3.3 the results of subsection 3.5.2 should be validated for instance by the bootstrap algorithms explained in section 3.3 and Simar and Wilson (2007). Table 3.3 shows the results for scale efficiency as well as the scale inefficiency estimations for both algorithms.

In each case the results significantly support the findings of the previous section of an efficient city size of about 226.000 inhabitants or 12.33 in logarithm of population. This can be seen by the positive estimate of β_1 for the logarithm of population and the negative estimate of β_2 for the squared logarithm of population in the estimations for scale efficiency in the first two columns of table 3.3.

[9] A Breusch-Pagan-Test has been performed but not reported here, and which rejects the null-hypothesis of no heteroscedasticity at 5% level of significance.

Table 3.3: Results for both algorithms and both orientations

	scale efficiency		scale inefficiency	
	Algorithm 1	Algorithm 2	Algorithm 1	Algorithm 2
Intercept	-4.856	-2.831	6.312	4.350
	[-7.537, -2.493]	[-5.137, -0.991]	[3.617, 9.440]	[2.395, 6.610]
log(*population*)	0.982	0.617	-0.880	-0.546
	[0.592, 1.413]	[0.311, 0.991]	[-1.326, -0,433]	[-0.908, -0.222]
$(\log(population))^2$	-0.041	-0.025	0.038	0.023
	[-0.059, -0.026]	[-0.040, -0.013]	[0.019, 0.057]	[0.009, 0.037]
optimum	11.842	12.359	11.969	12.126

Note: 95% confidence boundaries are reported in squared parentheses below the estimates.

The reported confidence intervals demonstrate the significance, since the boundaries are also larger or smaller than zero for β_1 or β_2, respectively. Similar results are estimated for scale inefficiency measures as the inverse of the scale efficiency in the last two columns. The estimates demonstrate significance according to the U-shape design by the negative estimates of β_1 and their confidence boundaries as well as the positive estimates of β_2 and their confidence boundaries of both algorithms. Each optimum point is about the same as the estimated optimum point of the robust linear MM-estimation in the previous subsection.

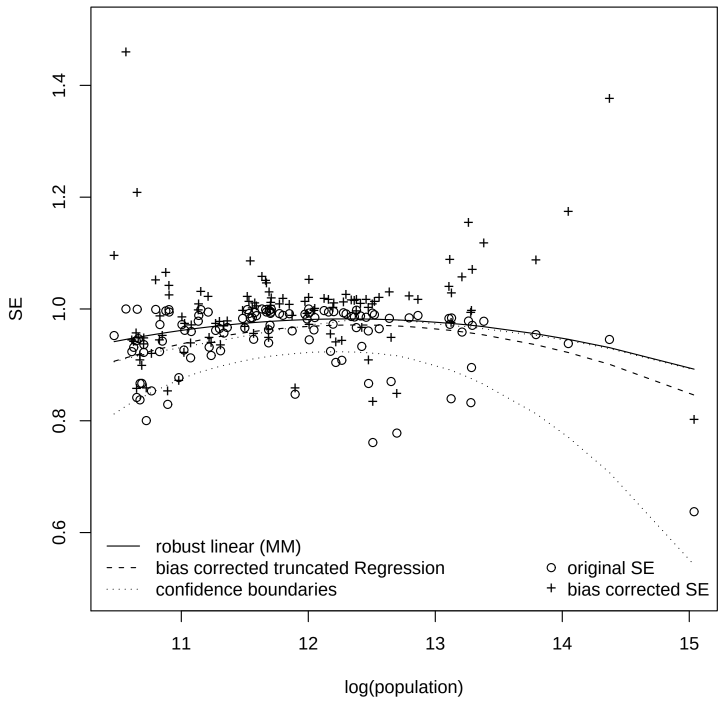

Figure 3.3: Results of Algorithm 2 Compared to MM-Estimation
Note: The confidence boundaries are for the level of significance of 1% and are drawn for the estimated scale efficiency scores in each of the 10.000 bootstrap replications.

Figure 3.3 shows the original observations for the scale efficiency scores by cycles. It also shows the bias-corrected efficiency scores which are also larger than unity; in these cases the bias-correction leads to disputable results for Shephards distance functions in regards to CRS larger than the accordant results for VRS. Nevertheless, the figure demonstrates that the maximum point is the same as the one estimated by the linear robust MM-estimation in the previous subsection. It also shows that the fitted graph of the MM-estimation is also within the range of the confidence interval of the results on the basis of those questionable bias-corrected observations. Therefore, the estimation of the robust linear regression is still preferred and will be used for the following analysis.

3.5.4 Geographically Separated Results within Germany

Furthermore, the robust linear fit model (MM-estimation) is employed for investigating geographical differences in Germany, to answer the question of whether or not there are differences in optimal city size. Comparisons are drawn for East and West Germany as well as for North and South Germany. Possible geographic differences may be explained by the historical transitions of some areas in Germany. Such a transition was precipitated by German Reunification 1990 which led to a change of the economic system associated with massive subsidies from the West German economy. There are still observable differences in the economic performance in East and West Germany as shown in Kirbach and Schmiedeberg (2008) and Sinn (2002). Another transition is caused by the structural change in North Germany by the decline of the shipbuilding industry and the economic transformation in North Rhine-Westphalia from a coal and steel region to one dominated by high-tech industries such as micro-systems (Jakoby, 2006). The comparisons are performed by separated estimations for each area. The estimation approach is the robust linear MM-estimator, which best considers the heterogeneity of the observed cities and outliers. The border is the former inner-German border before German reunification in 1990. Thus, all 22 cities of the former German Democratic Republic are counted for East Germany except Berlin which is viewed as a West German city. The results are presented in table 3.4.

Table 3.4: Geographically separated results for optimal city size in Germany

	East	West		North		South	
Intercept	-10.316	-0.836	**	-1.761	*	-0.966	
	(7.154)	(0.408)		(1.044)		(0.629)	
$\log(population)$	1.895	0.297	***	0.437	***	0.321	***
	(1.234)	(0.065)		(0.166)		(0.100)	
$(\log(population))^2$	-0.080	-0.012	***	-0.017	***	-0.013	***
	(0.053)	(0.003)		(0.007)		(0.004)	
n	22	90		56		56	
R^2	0.274	0.218		0.192		0.182	

Significance codes: '***', '**', '*' significant up to 1%, 5%, and 10%, respectively. Standard errors are below the estimates in parentheses.

As table 3.4 shows, there are differences in the estimates for East and West Germany and the estimates for East Germany are not significant at a level of 10% due to the small number of observations (n). The optimal city size for West German cities is $\frac{0.29694}{2 \cdot 0.012105} = 12.26518$ for the logarithm of population or 212,178 for the total population. This result is almost the same as for whole Germany since most cities are treated as West German cities. The optimal city size for East German cities is $\frac{1.8954}{2 \cdot 0.0795} = 11.92075$ for the logarithm of population or 150,355 for the total population, respectively. Thus, the optimal city size for cities in East Germany is much smaller than its counterpart for West German cities. In addition, the estimate for the quadratic term in East Germany is much higher than the one for West Germany, which indicates that exceeding or falling short of city size by the same amount results in a much higher loss of scale efficiency for East German cities. This conclusion is also visualized by figure A2 in the Appendix. Even when Berlin is treated as an East German city (which is not reported here, but is available upon request) the results remain stable indicating the robustness of the previous results. In addition, the coefficients of determination are quite high although the robust linear MM-estimator is applied.

The second comparison, namely between North and South Germany looks for north-south-differences within Germany. To have an equal number of cities in both regions the line of discrimination between both regions is drawn at the latitude of the median city which has the latitude of 50.94°N. Thus, almost all East German cities except the three most southerly-located are treated as North German cities.

The results shown in table 3.4 are compareable. An optimal city size exists for both regions caused by the inverted U-shape form of the relation between scale efficiency and population size. Although the estimates in specification I for North and South German cities are similar, the resulting optimal city size for each area is different. The estimated optimal city size of North German cities is $\frac{0.437112}{2 \cdot 0.017401} = 12.55997$ in logarithm of population or 284,921 for total population. Likewise, the optimal city size for South German cities is $\frac{0.320978}{2 \cdot 0.013206} = 12.15273$ in logarithm of population or 189,612 for total population, respectively. The coefficients of determination for the north and the south are almost the same at 18 or 19%. That is not very high but is due to the heterogeneity of the cities and the use of the robust linear MM-estimator.

3.5.5 Comparison of the Results for Geographical Distinction

It is noteworthy, that all estimations result in an inverted U-shaped representation of the relation of scale efficiency to population size with a positive estimate for the linear population size regressor and a negative estimate for the quadratic population size regressor. Furthermore, the optimal city size resulting from these estimates is always within the observed range of German cities and at the level of the mean sized city. The mean of German cities' population is 231,700 (see table 3.1) and the mean of population in South Germany is 155,600 and is almost half the size of the mean of the population in North Germany with 307,500. It is therefore unsurprising that the optimal city size in South Germany is smaller than in the northern part of Germany. Moreover, the optimal city size in West Germany including Berlin is slightly less than its mean with 251,900 inhabitants. Meanwhile, the optimal city size for the eastern part of Germany is slightly larger than its mean of 148,200 inhabitants. The regional distributions of population are summarized in table A2 in the Appendix as well as in table 3.5 which presents the mean, median, and comparison between the calculated optimal city size and the mean city size.

Table 3.5: Comparison between calculated optimal, mean and median city size

area	median size	optimal size	mean size	optimal - mean size
whole	120,600	221,301	231,700	-10,399
East	99,560	150,355	148,300	2,055
West	125,700	212,178	252,100	-39,922
North	177,800	284,921	307,600	-22,679
South	101,400	189,612	155,800	33,812

Note: Optimal city size is calculated by MM-estimation results.

As shown in table 3.5 the optimal city size is in all cases close to the mean of the underlying cities and thus differs in respect to the median of those cities. This result is stable even for different regional areas with remarkable differences in their mean and median city sizes.

3.5.6 Results for Urban Hierarchical Distinction

As already motivated, urban hierarchy explains many different developments of cities. Cities within an urban hierarchy are specialized in producing only few goods and services. The types of cities are separated into the four hierarchical clusters as analyzed and grouped in chapter 2.

Table 3.6: Results for hierarchically ordered cities

	Cluster 1	Cluster 2	Cluster 3	Cluster 4
Intercept	-2.932	-0.512	-0.473 ***	-1.102
	(1.905)	(0.156)	(1.237)	(3.515)
log(*population*)	0.642 *	0.252	0.921 ***	-0.014
	(0.327)	(0.26)	(0.193)	(0.587)
$(\log(population))^2$	-0.026 *	-0.011	-0.037 ***	0.000
	(0.014)	(0.011)	(0.008)	(0.002)
n	39	35	21	17
R^2	0.355	0.314	0.178	0.151
optimal city size	12.194	11.689	12.377	102.275
mean size	11.566	11.790	12.307	11.951

Significance codes: '***', '**', '*' significant up to 1%, 5%, and 10%, respectively. Standard errors are below the estimates in parentheses.

Table 3.6 shows the results of the estimation of scale efficiency on the natural logarithm of population and its squared value, distinguishing the hierarchical cluster. Although the number of observations is small for each cluster, the results are similar to the findings for geographical distinction, except for cluster four, which contains only 17 cities. Cluster four does not result in an optimal city size, since the linear coefficient is negative and the coefficient of the quadratic term is positive, resulting in a minimum point of scale efficiency with respect to population size. However, the estimates for cluster two and four are not significantly different from zero. Only the coefficients for cluster one and three are significant and result in a significant optimal city size and similarly the mean population size of the included cities. For both clusters one and three the optimal city size is larger than the mean city size, which indicates that the cities within these clusters should on average grow, or that most cities which are smaller than the optimal city size should grow, whereas only a few cities are too large. Figure 3.4 illustrates the results with the fitted quadratic curve for each of the hierarchical clusters.

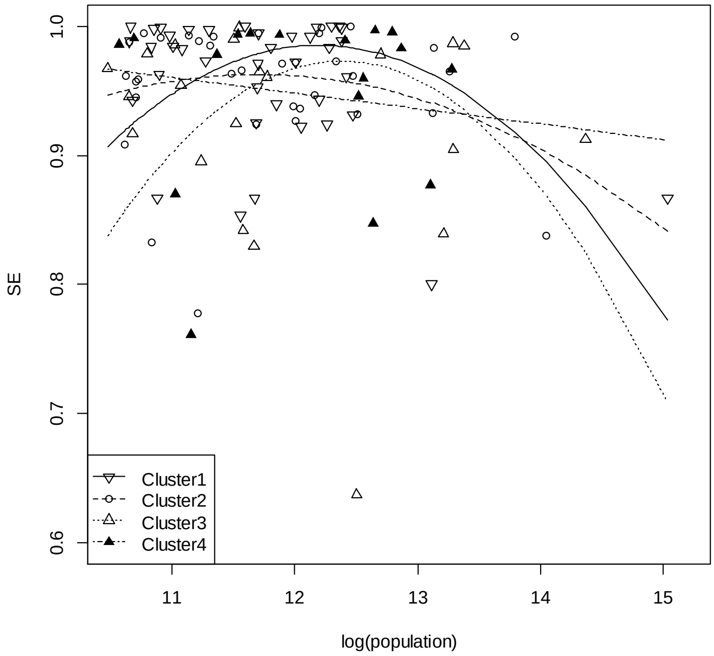

Figure 3.4: Fitted Estimates for Quadratic Models with Hierarchical Distinction

The figure shows that the cities within each cluster are not characterized by a similar population size. Instead the clusters are mixed across population size. The results are similar to those found for geographically-distinguished cities, since the optimal city size within each group is around the average city size within that group. That holds true for cities in hierarchical clusters one and three, which are cities with the lowest service to manufacturing ratio and an above average service to manufacturing ratio, respectively. The distribution of the optimal city size is not steadily developing over the urban hierarchy. As a result, cities in cluster four, which have the highest service to manufacturing ratio, would not gain scale efficiency by growing. In contrast, the optimal city size is the smallest for cluster two. However, the hierarchical discrimination of cities with respect to scale efficiency is not as beneficial as the geographical distinction.

3.6 Summary

This analysis investigates the relation between efficiency and population size of German cities. The relevant efficiency in this context is the scale efficiency, which takes into account the specific size of the particular city. Consequently, the approach employed in this chapter is a two stage process. First, the efficiency in terms of scale efficiency is measured, which involves estimating the efficiency of each city once for CRS and once for VRS, and then taking the ratio of both. The second stage investigates the relationship between the efficiency and the population size of the cities. The central result is, that there is an optimal city size in Germany. The optimal city size is about 220,000 inhabitants, which is almost the mean of all German cities involved in this investigation. Although there are regional differences in optimal city size, it remains stable that the optimal city size is the mean of the underlying cities. The findings are similar to the estimate of Alonso (1971) and consistent to those estimates of Capello and Camagni (2000). Furthermore, it turns out the largest cities have too many inhabitants which contrasts with the findings of Kanemoto et al. (1996) for Japan but occurs with the results of Loikkanen and Susiluoto (2004, 2006) for Finland. The two stage process was proven to be adequate as the findings were supported by modified bootstrap algorithms according to Simar and Wilson (2007).

Achievements in efficiency could be improved by a more unified population distribution over all cities in relation to the size of the mean. Cities with a suboptimal city size should attract people from cities with a surplus population. This is not only maintained by the economic performance of the industries within a city but also by the higher attractiveness of small cities compared to overpopulated cities, which lose attractiveness by the negative externalities caused by overpopulation such as traffic jams, noise, pollution, and so on, which results in negative economic performance.

On the agenda for further research is the analysis of industry-specific optimal city size, depending on the degree of specialization in the cities. This could be analyzed in a dynamic approach to optimal city size as part of a panel data analysis to account for unobserved effects. These effects could be accounted for by adding further variables such as the costs of living indexes, area sizes of cities, geographic distances which affect network possibilities between neighboring cities, especially in relation to the groups of cities in figure A1.

Another promising direction is the adoption of multilevel analysis since industries are part of the next level namely the city, and these cities are part of a further level, the level of federal states or time. On each level there are specific rules and laws, which influence firms and people in their decision to settle in a specific city. An analysis could account for these different levels. A further approach might take the form of an investigation of the time series, either by adopting dynamic models or by taking time as another level in a multilevel analysis. The analysis in chapter six picks up this idea.

4 Technical Efficiency and (Optimal) City Size

4.1 Motivation

Concerning the efficiency of local agglomerations, simply viewed as a production process in which inputs are transformed into outputs, it is an interesting question whether there is an optimal size. Taking agglomerations as cities we are here concerned with the particular size at which the efficiency of cites is maximized. Therefore, the central question is whether smaller or larger cities are most efficient? Or is there a city size between particularly small and particularly large which is associated with maximum efficiency?

One can easily imagine certain advantages of small cities which may lead small cites to be particularly efficient. These are greater homogeneity, shorter physical distances, more intense communication, stronger identification with the place and a slim and effective administration. However, there are also distinguished benefits of large cities, i.e. economies of scale and scope created by the multitude of economic opportunities which also permit diversification as well as lead to large cities being attractive locations for firm headquarters. Cities of a medium size may be able to combine the advantages of both small and large cities. On the contrary, cities of median size might already be too large to have the advantages of small cities and may not yet be large enough to benefit from the scale and scope economies of large cities.

As Glaeser (1998) argues, agglomeration forces as well as congestion forces play crucial roles for the relationship between city size and productivity. Small cities are disadvantaged by the lesser extent of agglomeration forces. Those forces are lower costs of moving goods and workers as input factors. This results from the local presence of suppliers and customers, labor market pooling, as well as higher levels of human capital as a result of learning within a city and its institutions. By contrast, as city size increases, congestion forces become more critical and may account for the disadvantages of more densely populated cities. These congestion forces raise costs of living and commuting, pollution and associated health issues, crime, and urban anonymity.

In this chapter we investigate the relation of efficiency and city size for German cities using a combination of non-parametric methods for efficiency measurement and non-parametric regression for assessing the functional form of the relationship. We show that the nature of the relation critically depends on the particular orientation chosen for the efficiency measurement. The plan of the chapter is first to provide a brief literature review in section 4.2. This is followed by an outline of the methodological approach chosen and the data base in section 4.3. Section 4.4 presents the results of the efficiency-size relation whereas section 4.5 contains a refinement of the analysis. Section 4.6 concludes.

4.2 Literature Review

Most urban analyses investigate only linear relationships between productivity (or efficiency) and city size. Sveikauskas (1975) finds mainly positive effects of city size on labor productivity for different manufacturing industries in the USA. Cingano and Schivardi (2004) also find empirical support for positive linear scale effects on total factor productivity at the firm level in Italian cities. Contrarily, Glaeser (1998) concludes that small homogeneous cities are more productive by estimating the size dependence of agglomeration and congestion forces. More recently, Baldwin et al. (2010) estimate negative effects of city size on labor productivity at the plant level in Canadian industries. This might be, however, caused by the fact that there are many smaller cities in Canada. The city size is ranging from 10 thousand to 4.6 million inhabitants but with a rather small median. In Germany the size ranges from 35 thousand to 3.4 million inhabitants for free cities with a larger median city size of 120 thousand inhabitants.

There are also some contributions in the literature investigating non-linear (i.e. non-monotonic) relationships. Most similar analyses investigate the optimal size of either labor productivity or total factor productivity (TFP) with respect to population size or total employment. Rigby and Essletzbichler (2002), for instance, investigate labor productivity within manufacturing industries in US cities and consider the possibility of a U-shaped relation of labor productivity and city size. Au and Henderson (2006a) investigate labor productivity in Chinese cities and find an inverted U-form of labor productivity on city employment depending on industry structure with a maximum range from 635,000 to 1.27 million. Relatedly, Au and Henderson (2006b) explicitly estimate an inverted U-shaped pattern for labor productivity and employment in Chinese cities with significant estimates for the period 1996-1997. Baldwin et al. (2007) find inverted U-form relations between labor productivity and population size with separate estimates for different industries in Canadian metropolitan areas. With respect to TFP, Fu and Hong (2011) estimate externalities on the level of firms in China. They estimate an inverted U-form for TFP in population size with an optimal point for firms in cities with a size between 1 and 2 million inhabitants.

Another argument goes back to various investigations of Henderson who mainly analyzes linear relationships and uncovers industry specific patterns in various countries such as the USA and Korea. Henderson (1986) shows that small and medium sized cities are highly specialized and that the input factors are not used more efficiently within larger cities. This may lead to a declining productivity with city size in input direction. In addition, Henderson (1997b) investigates medium-size cities with size between 50,000 and 500,000 inhabitants. Henderson finds that medium-size cities are highly specialized centers for services and manufacturing and that the industry structure changes with increasing size towards more skill-intensive industries. Furthermore, Henderson (2003) investigates localization and urbanization economies in different industries in medium-size and large cities. He finds only localization economies relevant for productivity of local manufacturing industries. Thus, together with his previous analyses he finds that scale economies should be present especially in relatively small and specialized cities. Henderson (2003, p. 24) concludes about urbanization-scale economies: "oddly, they then do not appear for non-affiliated machinery plants, where they ought to be more important". This follows because scale economies cannot be observed over the whole range of city sizes.

In sum, as a result of these findings in the received literature, we expect productivity levels to vary systematically with city size. Although there might be ranges of city sizes with decreasing as well as increasing productivity with the number of inhabitants, it may occur that positive and negative size effects nearly cancel out over the whole range of city sizes explaining non-significant population size effects on productivity.

Based on that, in the present analysis we want to tackle the question of the most efficient city size for a sample of German cities with a size of at least 35,000 inhabitants. Agglomerations smaller than this are not viewed as cities. Instead cities are defined as free urban municipalities by regional law.[10] Thus, our data set contains all 112 free cities classified as urban municipalities in Germany. Since efficiency can be measured either in terms of possible input reduction (the so-called input-orientation) or in terms of possible output enhancement (the so-called output-orientation) or by a mix of both orientations we put special emphasis on exploring the sensitivity of our results to the particular orientation chosen. This enables us to shed light on scale economies in small cities as distinguished from congestion effects in large cities.

[10]Labeled in German language as "kreisfreie Städte" and "Stadtkreise" in Baden-Wuerttemberg (status of 2008).

4.3 Method and Data

As surveyed in chapter 4.2 most urban economic analysis investigate U-shaped or inverted U-shaped patterns of labor productivity depending on population or employment size. With reference to the linear analysis of Sveikauskas (1975), Moomaw (1981) criticized that the use of labor only without capital leads to biased results. There are only some investigations explaining TFP such as Fu and Hong (2011) who, as already mentioned in section 4.2, find an inverted U-shaped pattern for TFP. By contrast, Combes et al. (2012) investigate TFP with a firm level data set for French cities. Although without direct investigation of TFP and city size, it becomes clear that larger cities are more productive which stands in contrast to the finding of Fu and Hong (2011).

Instead of most of the literature which relies on labor productivity and thus neglects the influence of capital as a factor of production, we use a non-parametric stochastic approach to compute a measure of total factor productivity accounting for both labor and capital inputs. This non-parametric approach does not require imposing a functional form of the production function and does not need price information to aggregate the production factors. It is thus independent of assumption regarding market behavior and equilibrium.

We measure efficiency in this chapter by the so-called order-α approach, introduced by Aragon et al. (2005) and fully developed to the multiple input-output case by Daouia and Simar (2007). This is a stochastic approach for computing a non-parametric efficiency measure which is more robust with respect to outliers compared to other non-parametric approaches such as classic DEA (Charnes et al., 1978). The approach is based on a probabilistic definition of the technology set. Stated for the case covered in this chapter with two inputs (capital and labor) in the vector $\boldsymbol{x}_j$ and a single output y_j (gross value added) for city j, the central concept is the probability of being dominated, meaning the probability of producing more output with less input usage. Formally, this probability is $H(y_j, \boldsymbol{x}_j) = \Pr(Y \geq y_j, \boldsymbol{X} \leq \boldsymbol{x}_j)$ with upper-case letters denoting random variables.

In the input-oriented perspective the efficiency measure of order α, $\alpha \in (0, 1]$, is defined as the Shephard distance (Shephard, 1970)

$$\theta_\alpha(y_j, \boldsymbol{x}_j) = \sup \left\{ \theta > 0 : H(y_j, \theta^{-1}\boldsymbol{x}_j) > 1 - \alpha \right\}. \tag{65}$$

Here, all inputs are reduced proportionally by the same factor θ^{-1} until the probability of being dominated is marginally larger than $1 - \alpha$, where domination is only by cities producing no less output. This assures that we do not need to enforce a complete envelopment of all DMUs as in conventional DEA but that a limited fraction $1 - \alpha$ of DMUs may be above the frontier function. Larger values of this measure reflect a greater level of *in*efficiency, meaning that the inputs have to be reduced by more for city j to become efficient. Therefore, we take the inverse $1/\theta_\alpha(y_j, \boldsymbol{x}_j)$ as the input-oriented efficiency measure.

The choice of α is usually within the interval $[0.90, 0.99]$. A choice of $\alpha = 0.95$ compares city j with the five percent of cities which may be more efficient. This means falsely classifying a DMU as efficient in 5% of the cases and is therefore analogous to committing a type-I error in statistical hypothesis testing. This value of α is used below.

Taking the output-oriented perspective the efficiency measure of order α is defined as the Shephard distance

$$\lambda_\alpha(y_j, \boldsymbol{x}_j) = \inf \left\{ \lambda > 0 : H(\lambda^{-1}y_j, \boldsymbol{x}_j) > 1 - \alpha \right\} \tag{66}$$

reflecting the level to which the output of city j has to be increased in order to become efficient. Larger values reflect a lower level of required output augmentation and thus in this case directly indicate a greater level of efficiency.

A combination of both orientations is the hyperbolic-graph efficiency measure (Färe et al., 1985, pp. 107ff.) which is defined for the order-α approach as

$$\gamma_\alpha(y_j, \boldsymbol{x}_j) = \sup\left\{\gamma > 0 : H(\gamma y_j, \gamma^{-1}\boldsymbol{x}_j) > 1 - \alpha\right\}. \tag{67}$$

By this measure the level of *in*efficiency of city j is expressed by a combination of possible output enhancement and input reduction along a hyperbolic path towards the frontier function. The inverse $1/\gamma_\alpha(y_j, \boldsymbol{x}_j)$ is taken as the efficiency measure.

Daouia and Simar (2007) provide an exact computational algorithm for the input- and output-oriented measures. Wheelock and Wilson (2008) provide a fast approximate algorithm for the computation of the hyperbolic-graph efficiency measure. These algorithms are implemented in the package FEAR for R as explained in Wilson (2008) which is used for the subsequent computations.

The three efficiency measures obtained in this way are related to city size by means of regressions in a next step. Applied are both parametric regressions in which efficiency is explained by a cubic polynomial in log size (with size measured by the number of inhabitants) and non-parametric local polynomial regressions. The latter are explained in detail in Loader (1999) who also outlines the functions of the R-package locfit used for the computations. We apply adaptive methods for bandwidth choice as described in Loader (1999, pp. 203f.). Reported confidence intervals are based on a local variance estimate.

The data set is that described in chapter 2 with gross value added, employment and the calculated capital stocks for the period 1998-2007. City size is measured by the number of inhabitants also taken from the regional database of the Statistical Offices in Germany. All estimates reported below are based on a cross-section of German cities with a minimum size of 35,000 inhabitants, where each efficiency measure is computed with averages of the last 5 years of available data (i.e. 2003-2007) for the inputs and the output. This has the benefit that possible measurement errors may be averaged away and that unavoidable errors in measuring the initial capital stocks are reduced.

Figure 4.1 shows the map of Germany with the location of the cities in our sample. The circles indicate the relative size of the cities and are drawn in a way that the area enclosed is proportional to the number inhabitants of the respective city. Berlin as the biggest German city can be easily identified as well as Munich in the south and Hamburg in the north. In the west the cluster of cities constituting the Rhine-Ruhr metropolitan region is also clearly visible. In addition, many quite dispersed smaller cities can be recognized which are somewhat more prevalent in the south than in the north.

4.4 Efficiency and Size

The regression results are presented in figure 4.2 in the form of scatter-plots of logarithm of efficiency and logarithm of city size[11] together with the regression fits. Shown are the local polynomial regression fit as black solid lines and the corresponding 95% confidence intervals as black dashed lines. A robust local polynomial regression fit is depicted by the dotted line. In addition, the solid gray line represents the parametric regression fit of a cubic polynomial in log city size and the dashed gray line is a simple linear regression fit. These regressions are perceived as merely descriptive devices summarizing the mean relation of size and efficiency.

[11] Size is measured by the number of inhabitants, taking for each city the average of the years 2003-2007.

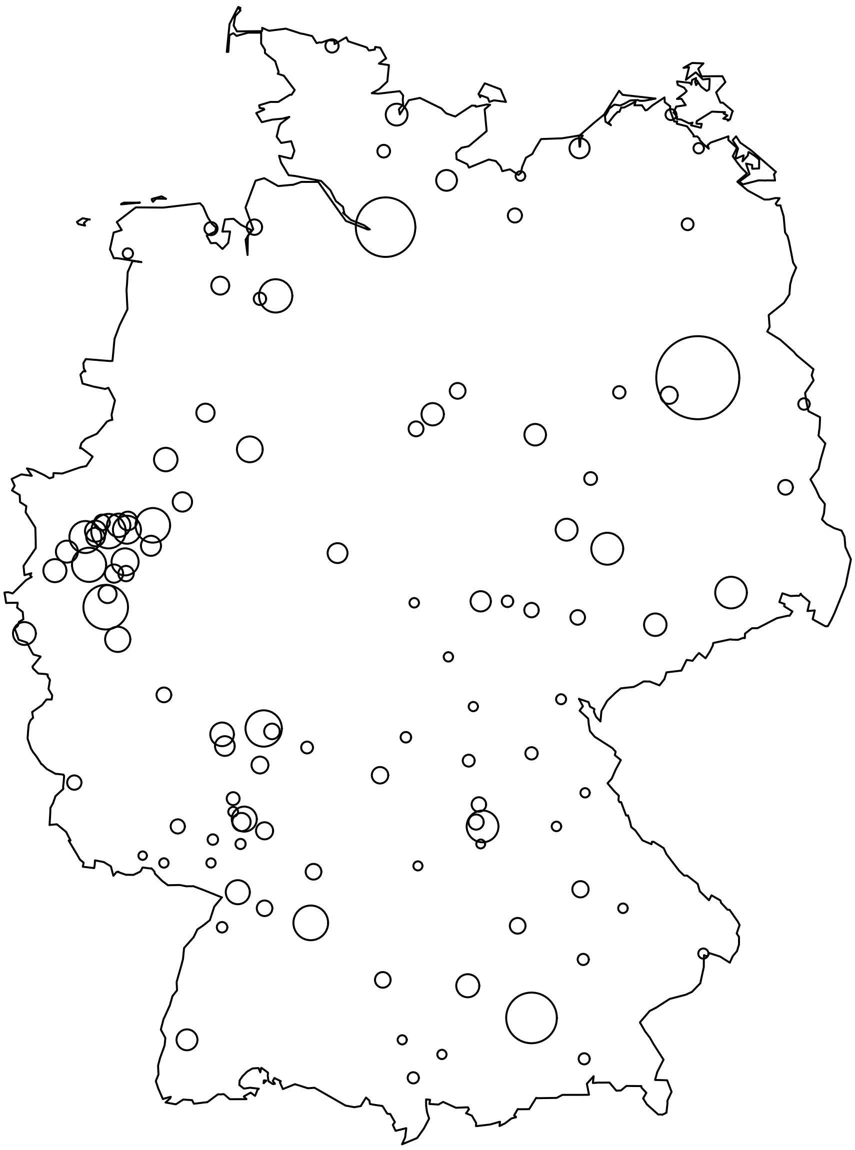

Figure 4.1: Map of Germany with Sample Cities

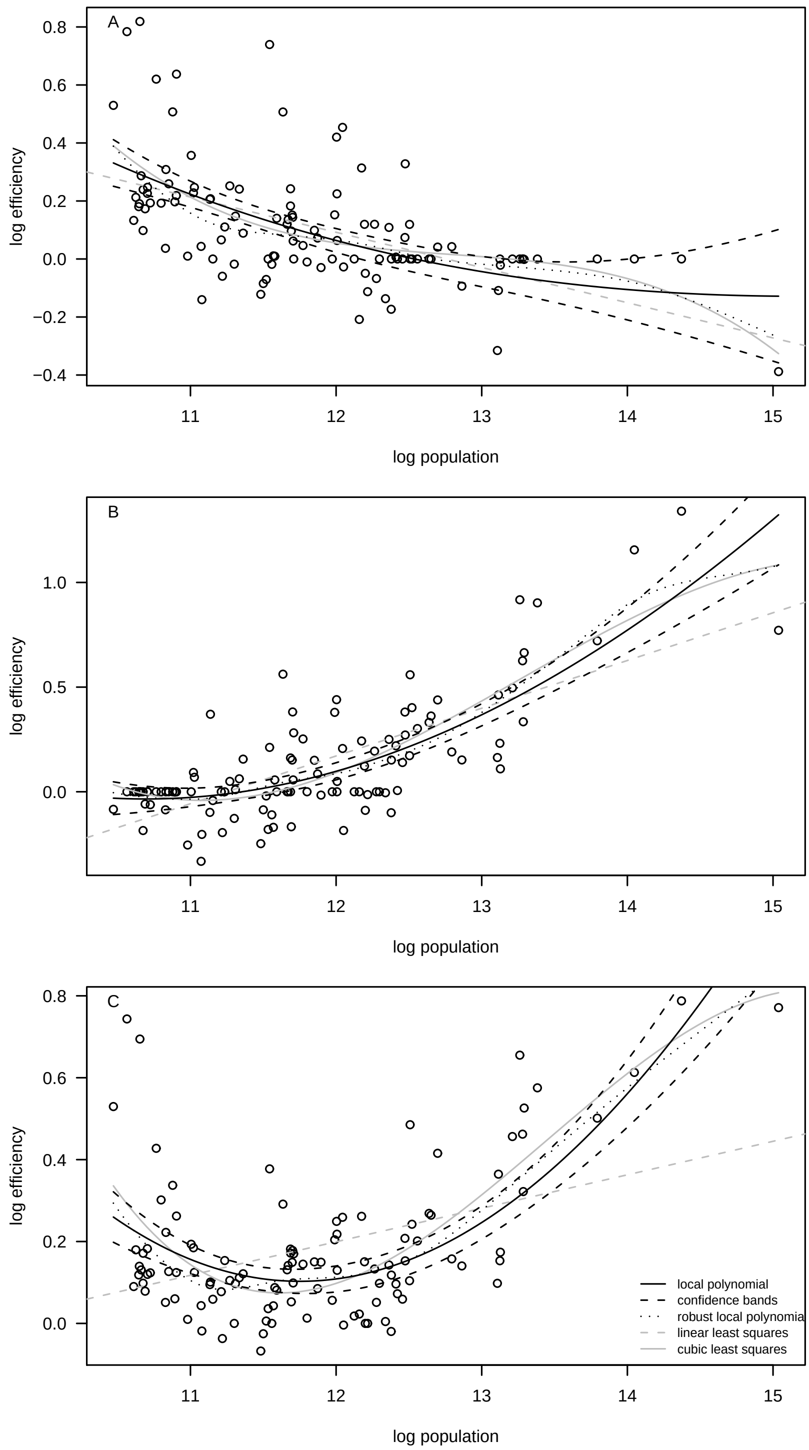

Figure 4.2: Efficiency-Size Relation with Regression Fits

Panel A of figure 4.3 is based on the input-oriented order-α efficiency measure. We observe that small cities appear to be systematically more efficient than large cities when efficiency is measured in terms of possible reduction of inputs. Efficiency is monotonically declining with size and all regressions agree on that. Panel B is analogous using the output-oriented order-α efficiency measure. When efficiency is measured in terms of possible output enhancement, the relation is converse with larger cities being systematically more efficient than smaller cities. The relation is also rather monotonic in this case and also robust across methods. This finding is similar to the results for labor productivity in Rigby and Essletzbichler (2002) for US cites.

Finally, panel C is based on the hyperbolic-graph order-α efficiency measure. Now we are faced with an approximate U-form relationship where smaller and larger cities appear to be systematically more efficient than medium-size cities. This occurs jointly with particularly large cities tending to be considerably more efficient than small cities as shown by the regression curves. For classifying city sizes, we use similar limits as Henderson (1997b) where medium-size cities are within the inter-quartile range, which is the interval between 60,000 and 250,000 inhabitants in Germany. This U-form relation seems to be a blend of the findings for the input-oriented and the output-oriented variants. It is, however, not overly surprising since the hyperbolic-graph efficiency measure expresses inefficiency in terms of both possible input reduction and output enhancement along a hyperbolic path towards the frontier function. The minimum can be located at a size of about 140,000 inhabitants ($\approx e^{11.85}$).[12] This size is larger than the minimum of 50,000 inhabitants in Sveikauskas (1975) for manufacturing industries in the US but similar to the results of Henderson (1997) for the US high-tech instruments industry. Labor productivity increases with city size from about 160,000 to 2 million inhabitants and decreases with population for cities with more than 2 million inhabitants. Moreover, in Korean manufacturing labor productivity is highest in medium-size cities with about one million inhabitants. In addition, there is a cluster of low productive cities with a mean size of about 100,000 inhabitants. In contrast, labor productivity in the textile industry is highest in US cities with a size below one million inhabitants, see Henderson (1997b).

From these results we see that the relation of efficiency and city size critically depends on the orientation chosen. On the one hand, when efficiency is measured purely in terms of input reduction small cities appear to be most efficient and thus closest to the frontier function. On the other hand, small cities are least efficient when efficiency is measured purely in terms of possible output enhancement and this implies that small cities are far from the frontier function in output direction. Small cities thus tend to use their inputs efficiently but are not able to realize their potential to produce output, possibly as a result of dismissed scale economies. Conversely, large cities are close to the frontier function in output direction and far from the frontier function in input direction. This is in line with an explanation based on exhausted scale economies not permitting much more enhancement of output and congestion effects indicated by the possibility to produce the same output with fewer inputs.

Thus, the two major problems of small (unexploited scale economies) and large (congestion effects) cities are shaping the results for input- and output-orientation, respectively. Cities of medium size are not particularly close to the frontier function in either direction and are thus least efficient once the hyperbolic-graph efficiency measure is used. Since the hyperbolic-graph efficiency measure is a combination of the pure input- and output-orientations, the U-form relationship can be viewed as the manifestation of both above mentioned problems affecting medium-size cities. In addition, the advantages of small cities are no longer present while the advantages of larger cities are not in reach.

[12]The qualitative findings are robust against other measures of city size, i.e. the total value added of the cities.

4.5 Influence of Hierarchy

We now turn to the consideration of the influence of the hierarchy on efficiency and on the relation of efficiency with city size, as argued in subsection 2.1.3. The cities are specialized according to their hierarchical order which leads to a formation of different types of cities as pointed out by the non-spatial model of Henderson (1974). Henderson (1974) shows that each city-type in the hierarchy has a specific optimal city size. Based on the urban hierarchy and specialization, Black and Henderson (2003) use the employment share of each industry to distinguish between the types of cities. With an increase of city size, the order in urban hierarchy increases and also the share of service sectors while the share of the manufacturing industries decreases as shown in Tabuchi and Thisse (2006). To assess the influence of hierarchy on the efficiency of the cities and the efficiency-size relation we apply the so-called frontier separation approach as suggested by Charnes et al. (1981). This approach provides a decomposition of an overall efficiency measure into a part which can be attributed to the decision making units *within* a particular group (managerial efficiency) and another part which is due to efficiency differences *across* the groups.[13] The decomposition can be formally stated as

$$\text{overall efficiency } (OE) = \text{managerial efficiency } (ME) \times \text{program efficiency } (PE)$$

where overall efficiency is the efficiency measure with all cities pooled together irrespectively of their hierarchy position. This is identical to the efficiencies analyzed in the previous section as shown in figure 4.2. Managerial efficiency is calculated for sub-samples of cities pertaining to a particular hierarchy group. Here, we consider one grouping into below-/above-median hierarchy levels and a second grouping according to terziles.[14] Thus, the managerial efficiency is calculated with respect to separate frontier functions specific to the hierarchy groupings. It reflects efficiency within a particular hierarchy grouping. Program efficiency is then computed as that part of overall efficiency which can *not* be accounted for by managerial efficiency, i.e. $PE = OE/ME$. It reflects the efficiency across hierarchy groupings.

The mean efficiencies are shown in table 4.1. Note that we compute geometric means ($(x_1 \cdots x_n)^{1/n}$) to summarize efficiencies instead of arithmetic means because of the multiplicative nature of the efficiency measures and the decomposition.

Table 4.1: Aggregate results of the frontier separation approach

	<median	>median	1st terzile	2nd terzile	3rd terzile
Input-orientation:					
overall efficiency (OE)	1.221	1.025	1.235	1.120	1.010
managerial efficiency (ME)	1.074	1.087	1.047	1.052	1.091
program efficiency (PE)	1.137	0.943	1.179	1.064	0.926
Output-orientation:					
overall efficiency (OE)	1.117	1.168	1.139	1.112	1.177
managerial efficiency (ME)	1.080	1.120	1.060	1.097	1.068
program efficiency (PE)	1.034	1.043	1.075	1.013	1.102
Hyperbolic-graph:					
overall efficiency (OE)	1.218	1.190	1.220	1.196	1.196
managerial efficiency (ME)	1.132	1.172	1.075	1.144	1.157
program efficiency (PE)	1.076	1.015	1.135	1.045	1.034

Note: Shown are the geometric means of the efficiency components across the indicated group.

[13] See e.g. Portela and Thanassoulis (2001) and Hampf and Krüger (2010) for more recent applications.
[14] The hierachical grouping in this chapter is different to the cluster analysis grouping of chapter 2.

For input-orientation we find a larger mean of overall efficiency for the below-median (<median) hierarchy group compared to the above-median (>median) hierarchy group. Overall efficiency is more due to program efficiency in the below-median group, whereas it is more due to managerial efficiency in the above-median group. Across terziles we find overall efficiency monotonically falling with a slightly increasing contribution of managerial efficiency and a noticeably decreasing contribution of program efficiency. Concerning output-orientation we find slightly smaller and larger overall efficiencies in the groups below and above the median hierarchy, respectively. In both groups this is more due to the managerial rather than the program component. Across terziles we now find a U-shaped pattern of overall efficiency, decreasing from the first to the second terzile and increasing from the second to the third terzile. This pattern is accompanied by a U-shaped pattern of program efficiency and an inverted U-shaped pattern of managerial efficiency. Regarding hyperbolic-graph efficiency we find overall efficiency larger in the below-median than in the above-median group and also only larger in the first terzile. Overall efficiency is here dominated by program efficiency in the first terzile and by managerial efficiency in the second and even more in the third terzile. Moving from the first to the third terzile we observe a rising contribution of managerial efficiency combined with a declining contribution of program efficiency.

We investigate the relation to size using the same type of figure as above for the overall efficiency in figure 4.2. Figures 4.3 and 4.4 show the scatter plots and the regressions for managerial and program efficiency, respectively, with grouping according to the median. Analogously, figures B.1 and B.2 in the Appendix B show the grouping according to terziles.

The focus is here in particular on managerial efficiency since this component corrects the efficiency measure for the influence of the hierarchy position. In figure 4.3 we find a decreasing shape of the regressions with size for the case of input-orientation (panel A) and an increasing shape for output-orientation (panel B). This is analogous to the findings for overall efficiency in figure 4.2. In the case of hyperbolic graph efficiency (panel C) we again find a U-shaped pattern, although with a bit alleviated curvature, however.

Corresponding results for program efficiency are depicted in figure 4.4. We see essentially flat regression lines in the case of input-orientation (panel A), implying that the decrease of efficiency with size is mainly caused by managerial efficiency. Program efficiency under output-orientation (panel B) is also increasing with size thus reinforcing the effect of managerial efficiency on overall efficiency for this case. For hyperbolic-graph efficiency (panel C) we find a clear U-shaped pattern so that the relation of overall efficiency to city size in this case seems to be mainly driven by the program component. Altogether, we can conclude that the efficiency-size relation here is mainly due to efficiency differences between the groups with comparably low (below the median) and high (above the median) hierarchy positions rather than differences within the groups.[15]

This set of results implies several conclusions concerning urban hierarchy and possible different production conditions across the hierarchical order. For efficiency measured in input direction the inefficient use of inputs is not caused by different levels of urban hierarchy. Therefore, highly ordered cities use the inputs inefficiently even if they are compared with similar ordered cities only (managerial efficiency). The relatively inefficient use of inputs of higher ordered cities might be due to the additional goods and services they have to supply to lower ordered cities. In the case of output-orientation the increasing efficiency by city size results from both cities within the same hierarchical order (managerial efficiency) and differences across the hierarchical order (program efficiency). Smaller cities with lower order are not only relatively more inefficient compared to cities with the same hierarchical order but also compared

[15]Figures B.1 and B.2 in the Appendix B show that these findings are robust to the alternative grouping according to terziles.

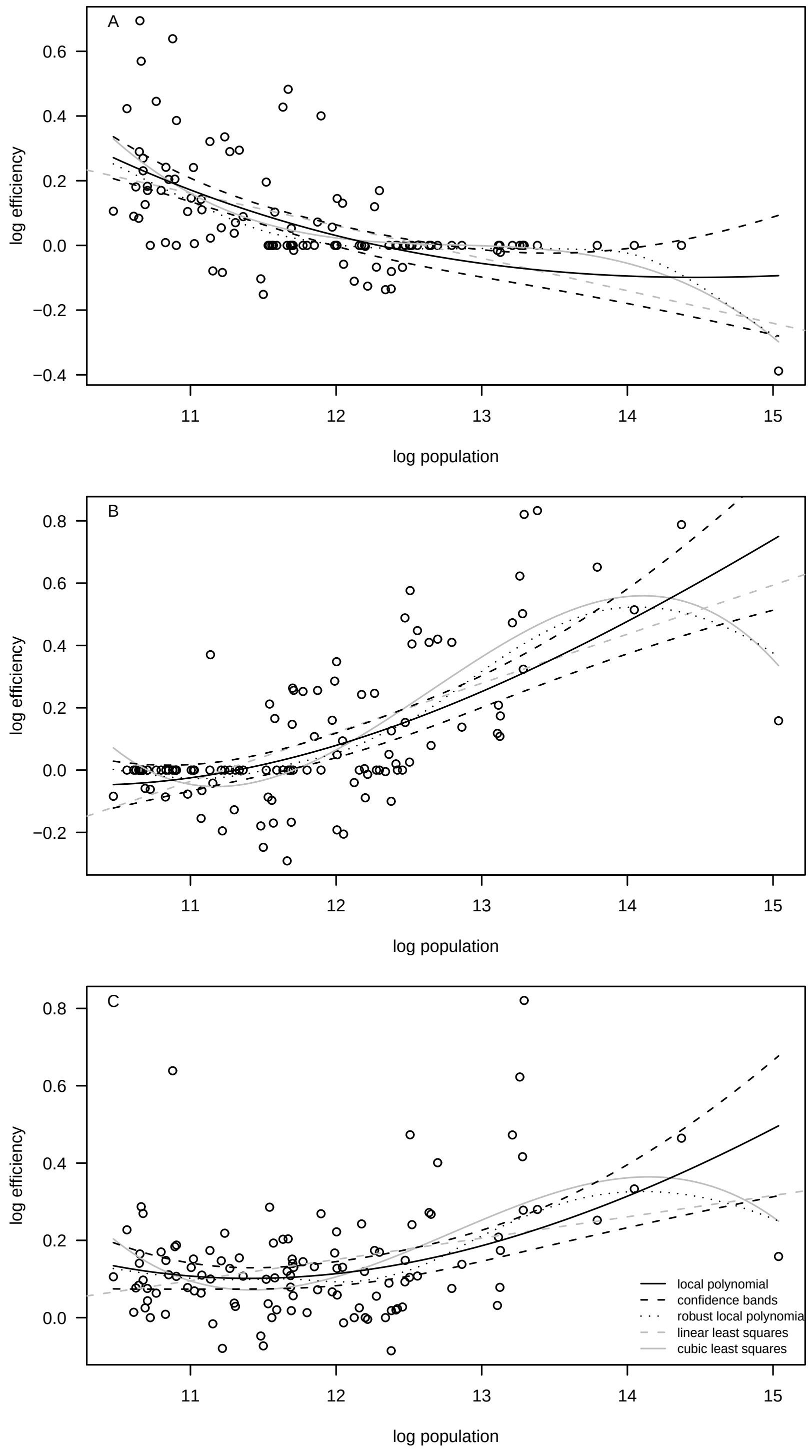

Figure 4.3: Scatter-Plots for Managerial Efficiency (Median Grouping)

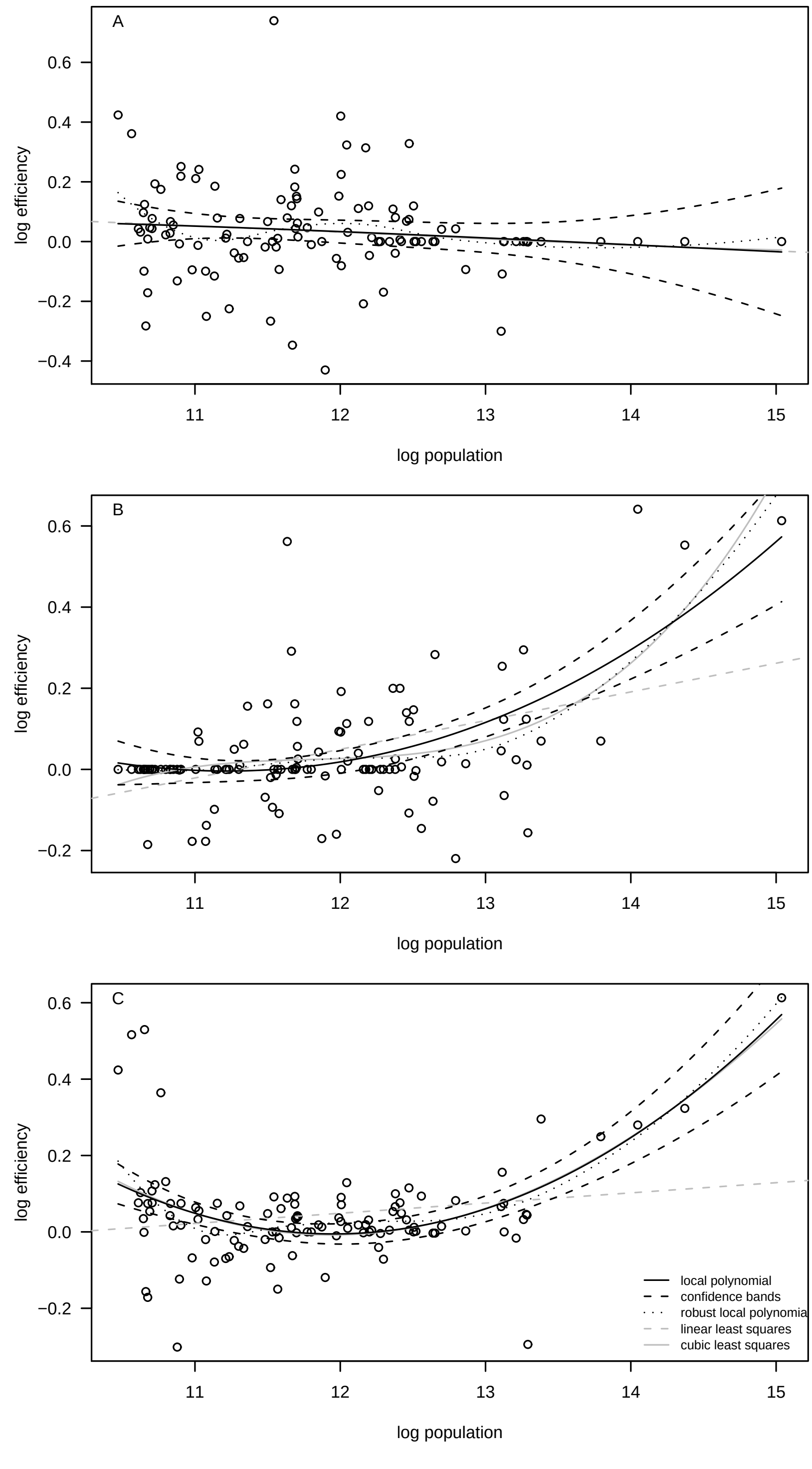

Figure 4.4: Scatter-Plots for Program Efficiency (Median Grouping)

to cities of higher order. That might be caused by the inability to realize economies of scale due to the lack of market size. Low ordered cities only provide local services for themselves and even smaller cities in the vicinity which results in a small market size. An optimal city size as found in several empirical investigations could be seen in the plot of managerial efficiency with hyperbolic-graph measurement. No optimal size seems to be present for lower ordered cities within the range of observed city sizes. With respect to program efficiency the findings indicate a systematic variation of efficiency with size for output-orientation and the hyperbolic-graph variant. Larger cities seem to be closer to the overall frontier and thus get a higher program efficiency measure. Panel C in figure 4.4 shows the U-form pattern with larger cities more efficient and medium sized cities least efficient.

The result is comparable to those in Fu and Hong (2011), although Chinese cities are larger and the corresponding medium size is larger. However, the last findings stand in contrast to the results in Au and Henderson (2006a,b) who find an *inverted* U-shaped pattern between labor productivity and population size and also account for urban hierarchy. On the one hand, in an urban system specialized cities gain localization economies but in an urban hierarchy only small, lower order cities can be specialized. On the other hand, a hierarchical urban system needs diversified cities with advantages in producing new ideas and new goods and services. These diversified cities are larger cities of highest order which cannot be specialized caused by the higher order and the obligation to produce all goods and services. Furthermore, the large cities gain from economies of scale by serving lower order satellite cities. This results in a more efficient production and different production functions for small and large cities as observed.

It can, however, also be the case that these differences may be due to the conceptual differences of partial and total factor productivity measurement. Partial productivity only accounts for the effects of one input factor, commonly labor productivity is used which is measured as output by labor. The partial productivity is however only comparable to total factor productivity in the case where the production function is characterized by constant returns to scales.

4.6 Summary

Our findings strongly support the existence of a relationship between productivity and city size in Germany. Productivity is not just labor productivity, but is estimated more generally within a non-parametric setup. It turns out, that smaller cities improve input-oriented efficiency compared to cities of medium size which are less efficient. We find the rise of productivity by city size only for output-orientation, which can be explained by benefits or urbanization economies.

Furthermore, our findings support the idea that economies appear to benefit from a mix of small specialized cities and large diversified cities as suggested by Duranton and Puga (2000). Using a frontier separation approach to assess the relation of efficiency and size in greater detail along a hierarchy of cities we find that the particular shape of the efficiency-size relation cannot be unequivocally attributed to the managerial or the program efficiency component. It seems that both the differences of cities within specific ranges of the hierarchy as well as the differences across hierarchy positions seem to shape the form of the efficiency-size relation jointly.

The monotonic relationship between productivity and city size assumed in many analyses is only an appropriate specification if larger (smaller) cities are considered and an output-oriented (input-oriented) productivity measure is used. Our results also shed light on inconclusive findings in the literature of sometimes positive, negative, or insignificant relationships between industry productivity and city size for several sub-samples. In any case the peculiarities of efficiency measurement play a substantial role in shaping the findings.

5 Price Level Differences between German Cities: A Spatial Autoregressive Investigation

5.1 Motivation

Even without trade barriers and the unrestricted free movement of factors, prices for goods and services differ. That observation is not surprising for inter-country comparisons due to incomplete integration caused by barriers of cultures or languages, for example. However, as is evident from the widespread experience of taking advantage of arbitrage, prices also differ within one country and even within one region. For example, if one drives through Germany by car one will notice that gasoline prices change in every region and even within a region or city, even if they belong to the same gas station brand. Of course, there are different rents within a city caused by different desirability levels of different areas. Unfortunately, there is a scarcity of regional price level data, although there are some available; for example the Office of Economic and Statistical Research of the Queensland Government regularly publishes regional price indexes. But it is costly to survey prices of the whole market basket in every region. Therefore, the data are only rarely available or only for one point in time, as is the case for Germany. It is interesting to ask what drives price level differences. Is the price level difference due to the high demand for goods, especially for non-tradable or immobile goods and services, or is it because of the low regional supply of some products and services? Which regionally-specific factors cause these patterns? A possible reason is that local monopolists causing price discrimination by a pricing to market strategy. Since it is hardly possible to collect all regional prices, an average price for the average market basket is observed. Thus, the question is: which region-specific factors account for the differences in the average price levels of the average household in German cities? These insights are not only useful for regional policy makers, who could influence the regional cost of living (henceforth: CoL) but also for everyone who rationally decides on where to live while accounting for the CoL differences. Although regional income differences may compensate CoL differences, there are more regional and systematic stimuli in the regional CoL, which cause the differences (Engel and Rogers, 1998).

Proximity, and its importance for the evolution of cities, is already mentioned in Duranton (1999). Thus, there might be spatial patterns within the CoL which should be accounted for and may have caused inaccurate estimates in previous studies. These spatial patterns are spatial spillovers which result not only from commuters earning income in one city while increasing demand in another city, but also from inter-regional trade. Furthermore, the urban hierarchy causes trade between cities of different hierarchical order because in an urban hierarchy each city has to produce specific goods and services. Therefore, since city are specialized in production, there has to be trade between cities. Commuting costs and urban hierarchy is already explained by Tabuchi and Thisse (2006) but is not linked to urban CoL. Moreover, although local governments are independent by law, they compete against each other for new firms and citizens and thus their decisions depend upon the past decisions of all of the cities. That structure should be accounted for in the analysis.

Reasons for price differences in international trade theory include trade barriers, the immobility of production factors, different monetary aggregates and non-traded goods. In the case of regions within one country, trade barriers and immobility of labor between regions are generally absent. On the other hand, production factors, especially labor, are immobile or hardly mobile between different sectors because there are not only teaching professions in some industries but also further labor legislation set by German labor unions which reduces the mobility of employees. This labor legislation includes regional wage agreements, dismissals protection, occupational skills, as well as entrance requirements for vocational training. Therefore, the effect of an demand increase to raise prices in other sectors might be low or depends on the interdependency of the considered sectors, although the CoL includes all consumer goods and services. Urban housing prices differ a great deal between cities for manifold reasons including the restriction of urban land for building, high population density, the mismatch between demand and supply of flats and houses in a price and quality range and the speculation of investors with real estate assets. Urban prices should therefore be reduced by the urban rents.

This investigation aims to shed light on the reasons for price level differences between German cities by investigating different urban characteristics. The questions addressed are: Does previously gained productivity change in different urban industry sectors, especially in the non-tradable service industry, accompany current price level differences? Does the price level increase with a higher business taxes in the city? Does the efficiency in the trade industry sector account for price level differences? Is there a systematic price level difference between East and West German cities as suggested by Kawka (2010)? Are spatial spillovers present with respect to urban price levels?

The chapter is organized as follows. Section 5.2 gives a short overview of the literature and theory on price level differences and develops the regression equation. The variables used in the regression are explained in section 5.4 together with the expected effects on price levels. The method which is used is explained in section 5.3. Section 5.5 presents the results. Finally, the results will be summarized and discussed in the last section 5.6.

5.2 Literature and Theory

Intimately connected with the determinants of regional price levels are the determinants of wage differentials in the literature on economic geography. Yankow (2006) estimates urban wage premium with wage levels and wage growth and finds significant differences between small and large US cities. Furthermore, Yankow corrects for CoL within these cities, which obviously has effects on wages with a stronger effect on large cities causing biased estimates. Wage differences thereby result from job changes, since he finds no differences in wages for the same jobs in cities and non-urban areas. Job change might also be due to the fact that cities have additional duties since they have to provide many facilities, which offer different jobs in urban areas compared to non-urban areas. That fact is referred to the so-called urban hierarchy which should be considered in urban analysis. Cities of the highest hierarchical order add value by high quality of life, as Albouy (2008) has shown. He concludes that population size, taxes, household income, and expenditure, which might improve quality of life, should therefore be included in regional analysis.

Bartik (1991) gives a review of literature for the United States where population growth significantly increases price levels although the effect decreases with no long-term effects due to the mobility of labor. Tabuchi (2001) shows that urban land constraints, interregional transportation costs, and commuting costs lead to an inability of spatial arbitrage and, together with urban agglomeration economies and diseconomies, result in price differentials. These externalities depend on the urban population size, which is correlated with urban efficiency, as shown for German cities by Hitzschke (2011).

The literature directly concerning urban and regional prices within one country is mainly influenced by the paper by Engel and Rogers (1996), who emphasize the role of the geographical localization of the observed cities. In urban literature, distance is used as proxy for trade costs by Crucini et al. (2009), who analyze time series for Japanese cities.

Metropolitan CoL is described by Haworth and Rasmussen (1973) who include population and population change along with other variables such as climatic and regional characteristics. Cebula (1989) estimates the determinants of geographical CoL differences in the US, controlling for heteroscedasticity and multicollinearity. Due to multicollinearity he cannot include all explanatory variables within one regression model. The explanatory variables for CoL are: per capita income, total population, geographical area, population density, tax rate, a coastal dummy, a dummy variable for working conditions and the proportion of population with a high school degree. He finds that each variable is significant in at least one specification and he calculates adjusted R^2 between 60% and 76%. Similar results are found by Cebula and Todd (2004) for counties in the US state of Florida in the year 2003. They find population size, per capita income, coastal location and population density to have significantly positive effects in relation to the average CoL. Cebula and Toma (2008) estimate the interstate CoL differentials for 2005 with a two-stage least squares estimation with per capita income, unemployment rate, geographical area, a dummy variable for working conditions, heating degree day, toxic chemical emission, and coastal location as independent variables. All these variables were found to be significant except the unemployment rate and the geographical area. By these geographical conditions, Cebula and Toma could account for 76% of the variation of the average CoL.

Glaeser and Maré (2001) show that for US cities wage increases are stronger than CoL increases by population size, although population size does not account a great deal for the variation of real income in the cities.

The costs of living and their spatial differences are observed and analyzed by many researchers such as McMahon (1991) who investigates the CoL for the states of the US individually. The statistical bureau of Japan periodically publishes price levels for 49 cities with Tokyo as the most expensive city, a situation especially driven by housing rents, with all other cities relatively equal in price levels. For Germany a similar pattern would imply that Berlin is the most expensive city since it is the largest city, the headquarters of many firms, as well as the capital city with many administrative duties and cultural sights attracting a large number of tourists. The effect of regional characteristics such as per capita personal income, the value of housing, and population change on the CoL index is estimated by McMahon and Chang (1991). They find that the value of housing and population change have significant positive effects, while per capita personal income is not a significant factor for metropolitan areas (for which they have only 24 observations made in 1981). For statewide regions per capita income is significant while change in population is not. They calculate coefficients of determination of 51% and 87% for metropolitan areas and statewide estimations, respectively.

With respect to regional price levels in Germany, Roos (2006) estimates regression models with population size, GDP per capita, average annual wage, rental rates of retail outlets, population density, a tourism dummy and an East Germany dummy for 51 German cities. One problem which arises is that he extrapolates the prices for ten years and so assumes the underlying market basket to be constant over ten years.

Blien et al. (2009) show the effect of price levels on wage differences in West Germany. Additionally, Kawka (2010) visually describes the regional price level differences in Germany for all NUTS3 regions and concludes that the average price level differences between East and West German regions is estimated to be above nine percent.

This present analysis is mainly influenced by the analysis of Kosfeld et al. (2008) who investigate the consumer price index without housing. They explain the price levels by disposable income, GDP, population, population density, wages and hotel overnight stays within a two-stage least squares regression with instruments for human capital and East Germany. They calculate impressive coefficients of determination of almost 94% for 50 German cities but the investigation is only applied to constructed data set presented by Roos (2006). Then they use these estimations for cities to estimate price levels in all 439 NUTS3 regions, including rural areas.

The regression model widely used for CoL is a reduced form equation model, see for instance Cebula (1989); McMahon (1991); McMahon and Chang (1991); Cebula and Todd (2004); Cebula and Toma (2008). I also adopt a reduced form equation, which I will now briefly develop.

The demand function for goods and services is

$$q = \alpha_1 p + \alpha_2 Y + \alpha_3 X + \alpha_4 D + e_1 \tag{68}$$

and the supply equation determines the price is

$$p = p_0 + \alpha_5 q + \alpha_6 Y + \alpha_7 X + \alpha_8 S + e_2. \tag{69}$$

The quantity of goods and services q in Eq. (68) is irrespective of the city since it is set by the market basket of an average family in Germany. That market basket is determined by: p, the prices for goods and services; Y, the income of the family; D, demand-specific factors within the city; X, urban factors that influence both demand and supply; and e_1 errors in the demand function. The price level in a city is described in Eq. (69) and determined by the initial price level p_0, the quantity q, supply specific factors S, X the urban factors influencing both demand and supply, and e_2 errors in the demand function. The initial price level p_0 is positive to cover fixed costs of production.

CoL is the price multiplied the quantity, which is assumed to be equal in the whole country and represents an average family in Germany with the standard market basket as set by the Federal Statistical Office according to Eurostat.

It can be easily shown (see Appendix C) that the resulting equation for the CoL is

$$CoL \quad = \quad p\bar{q} = \beta_0 + \beta_1 Y + \beta_2 X + \beta_3 D + \beta_4 S + u, \tag{70}$$

with $\beta_0 = \frac{\alpha_5^{-1} p_0 \bar{q}}{\alpha_5^{-1} - \alpha_1}$ $\beta_1 = \frac{(\alpha_2 + \alpha_6 \alpha_5^{-1})\bar{q}}{\alpha_5^{-1} - \alpha_1}$, $\beta_2 = \frac{(\alpha_3 + \alpha_5^{-1}\alpha_7)\bar{q}}{\alpha_5^{-1} - \alpha_1}$, $\beta_3 = \frac{\alpha_4 \bar{q}}{\alpha_5^{-1} - \alpha_1}$, $\beta_4 = \frac{\alpha_6 \alpha_5^{-1} \bar{q}}{\alpha_5^{-1} - \alpha_1}$, and $u = \frac{e_1 + \alpha_5^{-1} e_2}{\alpha_5^{-1} - \alpha_1}$. All linear coefficients share the same denominator $\alpha_5^{-1} - \alpha_1$ which is positive, because α_5 is positive and α_1 is for a normal good negative which means that a price increase normally leads to a decrease in demand except for in the case of a Giffen good. In addition, all coefficients incorporate the factor $\bar{q}$ in the nominator which is also positive because the average quantity $\bar{q}$ is positive. β_0 has to be positive, because the minimal supply price p_0 which includes fixed costs of production is positive. β_1, β_2, β_3 and β_4 can be positive or negative because the effects on demand and supply depend on the variable under consideration. The coefficients β_1 and β_2 capture effects of both the supply and demand side and it will be interesting to see which effect dominates. With respect to β_3, population increase the urban demand which is expressed by a positive coefficient α_4 and results in a positive over all coefficient β_3 in the linear CoL equation. Regarding β_4, an increase in local business tax should increase prices by a positive coefficient α_6 which leads to a positive coefficient β_4. The estimation results of Eq. (70) will be given in the section 5.5 based on the data set described in the next section 5.4 and section 2.1.

5.3 Estimation Method

The estimation method is explained in subsection 2.3, in which spatial model analysis is presented. The aim of the investigation is to test the error term of the OLS model for spatial correlation and if spatial correlation is found to find the recommended spatial model. The specific-to-general modeling approach in spatial investigations involves a testing procedure in accordance to Mur and Angulo (2009, pp. 202ff.)

and Haining (1997, p. 128).

The starting point of the analysis is the OLS estimation for urban prices, median rents and urban prices without rents as dependent variable $\boldsymbol{y}$, which is a vector containing the observation of the explained variable of all 112 German cities. That specific-to-general approach is discussed in comparison to the general-to-specific modeling which is favored by Florax and Folmer (1992) and Florax et al. (2003) but criticized and discussed by Hendry (2006) and Florax et al. (2006), respectively. However, the term "general" refers to the fact that the spatial parameters are not specifically fixed, whereas they are zero in the more specific models. The approach is therefore: First estimate the OLS model. Second, test for spatial correlation and if spatial correlation is present, estimate Lagrange multiplier (LM) test for SAM and SEM (see section 2.3). Third, if both LM tests are significant, take the spatial model with higher LM test and test for remaining spatial autocorrelation within the residuals. Fourth, if the residuals involve autocorrelation estimate the SDM. Moreover and in contrast to the spatial specific-to-general approach to find the correct spatial model, the evaluation of the necessity of the different independent variables is analyzed by the general-to-specific approach in each step to identify significant variables according to Lütkepohl (2007).

For comparison of different models I use likelihood-ratio statistics, information criteria, and estimate the likelihood based coefficient of determination ($\boldsymbol{R^2}$) described in Nagelkerke (1991). This $\boldsymbol{R^2}$ measure adjusts for the maximum value of the unadjusted $\boldsymbol{R^2}$ and is calculated as

$$R^2_{adj} = R^2 / \max\left(R^2\right) \tag{71}$$

with

$$R^2 = 1 - \left(L\left(0\right) / L\left(\hat{\beta}\right)\right)^{2/n}$$

and

$$\max\left(R^2\right) = 1 - L\left(0\right)^{2/n},$$

with $L\left(0\right)$ and $L\left(\hat{\beta}\right)$ the likelihood values of the model under the null hypothesis and the model with regression estimates respectively. The Nagelkerke $\boldsymbol{R^2}$ takes values between 0 and 1 and is therefore comparable to those standard in the OLS estimation. However, Nagelkerke $\boldsymbol{R^2}$ increases with the number of explanatory variables and therefore the information criterion, as the AIC, has to be taken into account to compare different models, as shown in Bivand (1984).

5.4 Data

The data set is described in chapter two and taken from the Federal Statistical Office of Germany together with additional data of the Statistical Offices of the Federal States, Empirica, and the Federal Institute for Research on Building, Urban Affairs and Spatial Development (in German: Bundesinstitut für Bau, Stadt- und Raumforschung or BBSR) of the Federal Office for Building and Regional Planning (in German: Bundesamt für Bauwesen und Raumordnung or BBR). Data on regional prices are taken from BBSR (2009). The BBSR estimated the price index for most German NUTS 3 regions conforming to their own observations and calculations on the basis of the market basket of the Federal Statistical Office of 2005. The BBSR reports the prices as an index for 2008, where the price level of Bonn is set to 100 as it is the place of the BBSR. The regional price indexes are calculated with observed or estimated local prices on many different goods and services in a period of 2006-2008, where the weights of these goods and services are the same for all cities. The goods and services are equivalent to those of the consumer price index of the Federal Statistical Office of Germany and the weights are similar. For example, the weight for rents is 20.33 percent. Furthermore, I have obtained rent prices for apartments with two and three rooms for 2005 from the German institute Empirica. To reduce the bias of extremely cheap and extremely expensive apartments the median of the rents for apartments with two and three rooms is taken to estimate a median for the rents for each city. These median rent prices were taken to calculate the regional absolute prices for each city. The absolute price without rent for Bonn can thus be calculated on hand of the median rent price multiplied by 0.7967, which is 1 minus the weight of rents within the price index. That scaling of the price level for Bonn is necessary and applied the underlying weight which reflects the share in the consumption market basket. Figure 5.1 illustrates the density of the price levels. The bandwidth is calculated by Silverman's rule of thumb (Silverman, 1986, p. 48) according to the data.

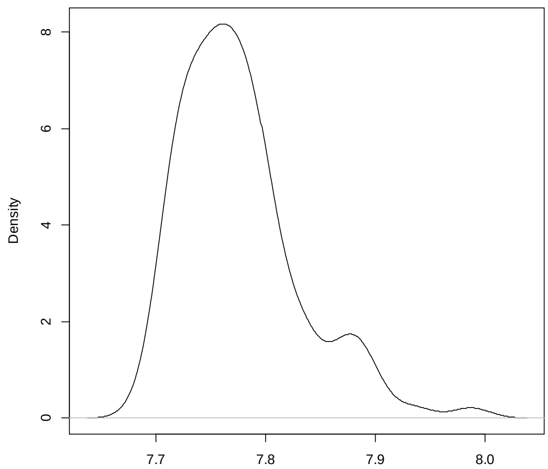

Figure 5.1: Density for the Logarithm of Urban Price Levels

The density of the urban price levels in figure 5.1 shows the skewness of the distribution which is positive with two local extreme values which are larger than the global maximum at the median value of around 7.77. As already mentioned above, the urban price level is a weighted average of housing rents and all other products and services consumed by an average German household. A different density distribution can be found for the natural logarithm of the median housing rents in German cities as illustrated in figure 5.2.

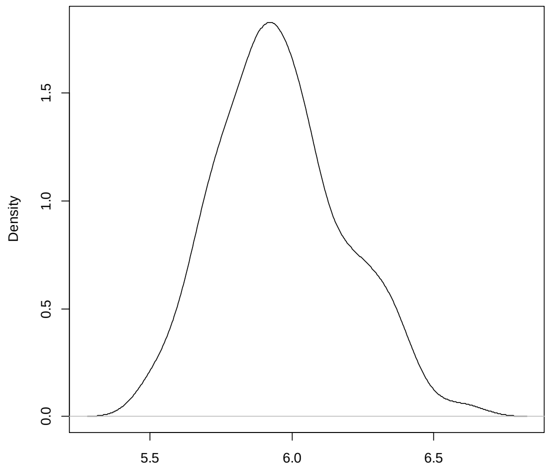

Figure 5.2: Density for the Logarithm of Median Urban Rents

Although the density of urban rents has a positive skew, it is much less skewed than urban price levels but it has a higher variation which causes the magnitude of the density graph to decrease.[16] Consequently, the resulting plot of the density of the logarithm of the price level without rents is presented in figure 5.3.

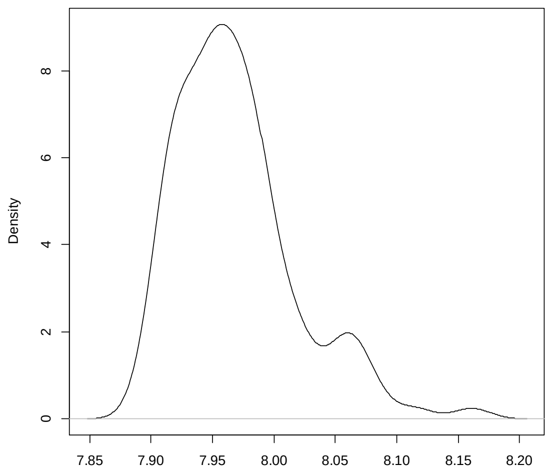

Figure 5.3: Density for the Logarithm of Urban Price Levels Without Rents

The prices of the other cities are calculated from their respective price indexes multiplied by the total price index with rents in Bonn minus their city-specific median rent prices. Price indexes can be viewed as an index for the CoL within a city because it incorporates the prices and demand quantities for a

[16]The magnitude of the density plot is the consequence of the variation in the variable since the area below the density is normalized to unity, which is measured by the integral over the variable. Therefore, an inverse relation between the variation of the variable and the magnitude of its density plot exists.

household. Unfortunately, the regional classification does not assign price indexes to all cities explicitly. In these cases the BBSR (2009) provides only data for the conglomerate of the city and its surrounding rural area. This is the case for small cities for which the published price indexes are probably smaller than their actual amounts.

In contrast to Engel and Rogers (1996), Crucini et al. (2009) and others, I have only a cross-sectional data set. So I only have observations for German cities at one point in time. A construction of growth rates of price levels is not possible since past observations do not provide information on all 112 cities, for example Ströhl (1994) contains only 50 German cities.

The first explanatory variable is the disposable income (*disp income*) in million Euros. It is supposed that a higher disposable income in a region is correlated with a higher price level, but the effect does not lead to a compensation of nominal income differences by equal real income across all regions, see BBSR (2009, p.21). It must, however, be mentioned that there are regional differences, as, for example, the case of the Bavarian capital and third largest city in Germany, Munich, which has high nominal income but unevenly higher price levels, which results in a lower real income as compared to other Bavarian cities. Disposable income includes the income of self-employed persons and capital income of citizen within the urban area which should increase urban demand and consequently raise price levels. Therefore, the coefficient should be positive.

The second explanatory variable is the squared amount of population (pop^2) as the measure for the size of the city. The importance of population size for German cities also with respect to industry efficiency as well as the heterogeneity of population size has already been shown in chapter three. In addition, the change in population (*dpop*) within the city between the years 1999 and 2006 was tested. The change in population induces employment growth within the city or at least an increase in the labor force. It is assumed that land prices will increase and as a result prices of goods and services within the city will also increase, as explained in Bartik (1991) and estimated in Haworth and Rasmussen (1973). As a result, population and change in population should positively affect price levels, while the effect of the squared size of population could be positive or negative. If the coefficient of squared population is negative (positive) than there is a maximum (minimum) price level with respect to population size. Additionally, the number of unemployed persons within the city (*unemp*) can be tested instead of the number of population since both variables are highly correlated.

The third variable is the number of students (*stu*) studying at a University or a University of Applied Science from the INKAR data base, which captures the knowledge base within the city. On the one hand, students increase demand as already explained for population and therefore raise prices. On the other hand, a larger knowledge base within the city should increase productivity and decrease prices. That effect might also occur since students are relatively cheap employees. The last two effects are controlled for by a direct measure of productivity and disposable income. The overall impact on prices depends whether the demand or supply side effect dominates.

A fourth variable accounts for regional policy as a result of the last federal state elections before 2006. Although, election results on county council would be better but not available for all regions, however, they are not supposed to be much different to the regional election behavior in state elections. In Germany, there are two major political parties the Christian Democratic Union of Germany (CDU) and the Social Democratic Party of Germany (SPD), who perform local and regional governance. A variable measuring the relative votes for the CDU within the city for the state parliament (*CDU*) is therefore included. Additionally, I tested alternative variables such as a dummy variable which takes the value of one if the CDU gained more votes than the SPD in the state election as well as dummy variables accounting for a high percentage of voters for the Free Democratic Party (FDP) and for the Party Alliance

'90/The Greens, but none of these was significant. If the state governments take the regional election results into account for their policies and are mainly expected to act liberally by reducing price cartels or restrictions influencing prices, then the coefficient should be negative.

Another explanatory variable is the dummy variable for West Germany ($west$), since for example Kosfeld and Eckey (2008) find that wages are significantly higher in West German labor market regions. Therefore, the coefficient for West Germany should be significantly positive.

A specific feature of the German federal-law system is that the business tax ($BusTax$) varies across regions in Germany and is set independently by each district and thus also by each city. A higher business tax at the place of production or at least the place of supply raises the consumer price for goods and services if the competition in the market is low. In such a case, the effect of business taxes on regional price levels is positive.

In urban analysis, the hierarchy of cities has to be taken into account. Cities within an urban hierarchy are specialized in producing only a few goods and services. Cities which are high in the urban hierarchy have to supply not only administrative services, large shopping and leisure facilities but also cultural facilities like opera houses, art galleries, and so on. These highly positioned cities serve the surrounding lower position cities and regions with these facilities, which do not have to be present in every city, to gain economies of scale. Cities which are lower in the urban hierarchy are by contrast specialized in producing goods and services especially in the manufacturing sector. As a consequence of this division of labor, the hierarchical order can be measured by the service to manufacturing industry ratio (SMr) as modeled for example in Tabuchi and Thisse (2006). Since highly ordered cities need goods and services of the surrounding lower ordered cities and because they have to be transported to cities higher in hierarchy, price levels should be higher in cities nearer the top of the urban hierarchy. The effect of urban hierarchy is not only estimated by the SMr in a linear estimation but also by clustering cities according to their urban hierarchy as applied in subsection 2.1.3. The allocation to the clusters two to four is expressed by three dummy variables. The coefficients indicate differences to cluster one, which consists of the cities with the lowest service to manufacturing ratio. Steadily increasing coefficients would indicate that price levels are increasing by urban hierarchy.

A further variable of interest is the concentration or diversification of economic activity within the city. The Gini coefficient of the gross value added of ten different industries ($GVA\ Gini$) serves as a measure of concentration. Current prices influence the value added in an industry and consequently the Gini coefficient of the cities before 2006 is used. The ten different industries are defined at one-digit industry specification of the Federal Statistical Office of Germany, which was WZ 2003 (Statistische Ämter des Bundes und der Länder (2009); Statistisches Bundesamt (2003) and is supplied by the statistical annals of the German Association of Cities).

To account for different levels of productivity a non-parametric efficiency score is estimated by DEA under constant returns to scale (Charnes et al., 1978). The efficiency scores are bias-corrected by homogeneous bootstrap methods as explained in Simar and Wilson (1998). The capital and labor serve as input variables of the industry and gross value added is the output. Higher levels of productivity indicate a more efficient use of inputs to produce the gross value added. These efficiency scores are estimates for the urban economy in total as well as for the five sectoral groups. If an industry is more productive it is able to produce output more cheaply and can therefore supply at a cheaper price, which should result in a negative coefficient. This negative coefficient should be especially significant for the private industry sectors CDE, F, GHI and JK (dea_CDE to dea_JK).

Furthermore, the change in productivity is measured by non-parametric DEA over time. Since the price index is available for 2006 the change in productivity is calculated until 2006. The change of productivity is thereby calculated using the Malmquist index as the overall productivity change and its components. These components are the technological change which caused changes of the productivity frontier on the one hand and on the other hand the efficiency change which causes changes to the position of the city in relation to the productivity frontier. Change in productivity is calculated for the total urban economy as well as individually for the five economic sectors (*change_ CDE, change_ F, change_ GHI, change_ JK* and *change_ LMNOP*). I tested which time difference is optimal not only to explain price level differences but also to reduce correlation with any other variable. It is evident, that the time difference for the change in productivity, with the highest explanatory impact on price levels, is five years, which accounts for the average length of business cycles during the observed period (Schirwitz, 2007). The difference of five years was tested against one, two, three, four and six years of differences which were proven to have less explanatory power. Only the decomposition of the Malmquist index into change of technology and change of efficiency is used because more decomposed measures as shown in section 2.1.2 are correlated with other explanatory variables (e.g., squared logarithm of population). Thus only productivity change measures are used which are estimated on CRS efficiency scores. The use of these two components is most common since it applies no restrictive returns to scale assumption. For some economic sectors, especially the service industries, non-increasing returns to scale hypothesis cannot be rejected for some cities.

Regional productivity and change of productivity is expected to influence regional price levels as investigated for European NUTS-2 level by Egert (2007). The expected influence roots on the effect introduced by Balassa (1964) and Samuelson (1964). The effect explains why services are cheaper in less developed or low productive countries. By viewing cities like open economies, the Lerner-Samuelson theorem, or theorem of factor price equalization, would predict equal wages and interest (factor prices for labor and capital) and therefore the law of one price should hold. Although the law of one price can be violated by non-tradable goods, as described by the Balassa-Samuelson effect, the overall purchasing power parity should hold in a free trade area. The difference between purchasing power parity and the law of one price is that the latter only account for one good or service while the former involves a market basket of many goods and services. An increase in productivity in one sector increases output and factor costs in that sector. That rise in factor costs will result in an increase in factor costs in the other sectors and induce surging prices in the other sectors and thereby lead to an increased price level. Thus, wages tend to rise when productivity increases, whereas firms are producing cheaper if wages do not rise while productivity increases. Price levels are therefore subjected to the wage agreements, the productivity and productivity change.

Table 5.1 contains the descriptive statistics for the urban variables that were considered.

Table 5.1: Descriptive statistics

	min	1st Quartile	mean	median	3rd Quartile	max	s.d.
price index	84.8	88.8	92.76	91.75	94.7	114.4	5.2
price without rents	2676.19	2786.17	2893.62	2864.54	2948	3503.88	146.45
ln(*price without rents*)	7.89	7.93	7.97	7.96	7.99	8.16	0.05
ln(*price level*)	7.69	7.73	7.78	7.77	7.8	7.99	0.05
ln(*rents*)	5.49	5.8	5.96	5.93	6.07	6.61	0.22
disp income	13.29	15.54	17.49	17.45	19.09	25.77	2.41
stu	0	17.52	61.59	43.65	82.62	244.5	60.47
pop	34.79	63.97	232.71	120.75	238.9	3407.62	390.15
CDU	0.14	0.33	0.4	0.38	0.47	0.66	0.11
GVA Gini	0.19	0.26	0.3	0.29	0.33	0.59	0.07
BusTax	300	390	413.36	410	445	490	36.92
dea_JK	0.39	0.54	0.62	0.62	0.68	0.94	0.12
SMr	0.69	2.99	4.29	3.82	5.61	11.94	1.97

Except in the case of the dummy variables, I used the natural logarithm from the explanatory variables as well as the explained variables. In order to take the logarithm and since some variables contain values of zero, I added the value of 0.001, which is for example one student in the case of the variable *stu*. Taking the logarithm also changes the variation of the variables, which decrease, and it also makes the variable in the data set more homogeneous. Furthermore, the estimates represent the price elasticities. Additionally, table C1 in the Appendix C2 shows the correlation matrix for the considered variables. Based on table C1, it becomes obvious that all variables cannot be included in one analysis due to high correlation and due to the construction of the Malmquist index as product of change in technology and efficiency. I therefore separately estimate and report the results for change in productivity for the Malmquist index and its two components. In order to reduce multicollinearity I tested population size, squared population size and over-night-stays. To account for non-linearity, I only include squared population size in the analyses in the following analysis. The non-linearity was already proven in chapter 3 and supported by specification test.

As table C3 in the Appendix C shows, the logarithm of price level without rents does not significantly contribute to future disposable income, number of students, population size and business tax given the past observations. These results indicate that an observed correlation is only from these variables to price level without rents. The test is similar to a Granger causality test although price level without rents is not observed for several years which prevents applying the original Granger causality test as presented in Greene (2008a, pp. 699f.).

In the analysis, distance between cities is calculated based on the geographical longitude and latitude instead of travel time by population as used in Rice et al. (2006), whereas Kosfeld and Eckey (2008) use travel time between labor market regions due to missing data for German commuters.

5.5 Results

In order to find the best model based on the available city-specific data set for Germany, I use the general-to-specific approach. The general-to-specific approach is a commonly applied approach that uses model results such as significances and information criteria to reduce a general model to a more parsimonious model, see Campos et al. (2005) for a general description and examples and Lütkepohl (2007) for a discussion of the approach with respect to time series analysis.

In table 5.2 columns one, two, and three show the OLS results of the regression for the price index in total, rents, and prices without housing prices, respectively. In table 5.2 change in productivity (*change*) is measured by the Malmquist index.

Table 5.2: OLS results

	price levels		rents		prices without rents	
Intercept	5.768	***	-3.985	***	6.244	***
	(0.299)		(1.458)		(0.268)	
disp income	0.203	***	1.048	***	0.173	***
	(0.03)		(0.148)		(0.027)	
stu	0.001	*	0.005		0.001	*
	(0.001)		(0.004)		(0.001)	
pop^2	0.001	***	0.001		0.001	***
	(0)		(0.001)		(0)	
CDU	-0.024	*	-0.045		-0.023	*
	(0.014)		(0.07)		(0.013)	
GVA Gini	0.069	***	0.29	***	0.061	***
	(0.016)		(0.079)		(0.014)	
BusTax	0.003		0.011		0.002	
	(0.004)		(0.017)		(0.003)	
dea_ JK	0.089	***	0.296	**	0.081	***
	(0.026)		(0.128)		(0.024)	
change_ CDE	0.038	**	0.192	**	0.033	**
	(0.017)		(0.084)		(0.015)	
change_ GHI	-0.046		-0.283	*	-0.039	
	(0.034)		(0.167)		(0.031)	
change_ JK	-0.063	*	-0.211		-0.058	**
	(0.032)		(0.156)		(0.029)	
*hcl*2	0.017	**	0.038		0.016	**
	(0.008)		(0.041)		(0.007)	
*hcl*3	0.006		0.038		0.005	
	(0.01)		(0.047)		(0.009)	
*hcl*4	0.028	**	0.136	**	0.024	**
	(0.011)		(0.052)		(0.01)	
west	-0.006		-0.113		-0.003	
	(0.013)		(0.064)		(0.012)	
AIC	-440.17		-85.02		-464.18	
R^2	0.7037		0.5835		0.7086	
ln$L$	236.084		58.511		248.088	
LR-test	136.238		98.099		138.097	

Significance codes: '***', '**', '*' significant up to 1%, 5%, and 10%, respectively. Standard errors are below the estimates in parentheses.

The results reported in table 5.2 show similar pattern for price levels, rents, and prices without rents, although the explained variation of rents is not as high as for the price levels. Significantly positive coefficients are found for disposable income, number of students, the Gini coefficient for the industrial gross value added, the efficiency score of the financial and business services sector (JK) and the change in efficiency within the wide manufacturing sector (CDE). The estimated coefficients are interesting, since the coefficients for disposable income are about 0.2 on price levels and price levels without rents whereas it is 1.05 on rents. On the one hand, these coefficients indicate that an increase of disposable income is corresponding to a 0.2 percent higher price level with and without rents. On the other hand, rents are even higher although the coefficient is not significantly larger than one. A one percent increase in disposable income corresponds to one percent higher rents within the city. The other coefficients are although significant rather small with less than 0.1 percent on price levels with and without rents. The dummy variables for the hierarchical clusters two and four (*hcl*2 and *hcl*4) are only significantly positive on price levels and prices without rents. Only hierarchical cluster four is significantly positive on rent per square meter. Cluster four only contains cities with the largest service to manufacturing ratio, these cities are thus classified as cities of highest hierarchical order. These cities provide all services as well as administrative and cultural facilities for their surrounding areas. The coefficients for the hierarchical clusters are relative to the hierarchical cluster one, which includes all cities with the lowest service to manufacturing ratio and are thus of the lowest hierarchical order. These cities provide goods for the higher order cities with little service industry such as Wolfsburg, which has the lowest service to manufacturing ratio due to its large automotive industry centering on Volkswagen and its suppliers. The coefficients are furthermore largest for hierarchical cluster four ($hcl4$) without a steady increase. It is therefore appropriate not to use a linear or quadratic term of the service to manufacturing ratio.

The change in productivity in the private non-financial services sector is only negative and weakly significant for rent per square meter. By contrast, the change in productivity in financial and business services is significantly negative only on the price levels and price levels without rents. That negative relation might be due to the fact that this sector includes the real estate business industry. If real estate services and financial industry have a higher change in productivity the prices are lower within the city. On the other hand, a higher increase in the change in productivity in the wide manufacturing sector leads to higher price levels without rents, higher rents and consequently higher price levels including rents.

Because rents are different than the other prices, I focus on prices without rents. The change in productivity is measured by the Malmquist index and can be split up into its components - namely change in efficiency and change in technology. The results are reported in table 5.3.

Table 5.3: OLS results for prices without rents

	Malmquist		efficiency		technology	
Intercept	6.244	***	6.174	***	6.137	***
	(0.268)		(0.27)		(0.281)	
disp income	0.173	***	0.181	***	0.184	***
	(0.027)		(0.028)		(0.029)	
stu	0.001	*	0.001		0.001	
	(0.001)		(0.001)		(0.001)	
pop^2	0.001	***	0.001	***	0.001	***
	(0)		(0)		(0)	
CDU	-0.023	*	-0.027	**	-0.02	
	(0.013)		(0.013)		(0.013)	
GVA Gini	0.061	***	0.063	***	0.043	***
	(0.014)		(0.014)		(0.015)	
BusTax	0.002		0.003		0.004	
	(0.003)		(0.003)		(0.003)	
dea_ JK	0.081	***	0.081	***	0.051	**
	(0.024)		(0.024)		(0.021)	
change_ CDE	0.033	**	0.032	**	0.132	
	(0.015)		(0.015)		(0.122)	
change_ GHI	-0.039		-0.063	*	0.064	
	(0.031)		(0.034)		(0.103)	
change_ JK	-0.058	**	-0.05	*	-0.617	*
	(0.029)		(0.029)		(0.328)	
*hcl*2	0.016	**	0.016	**	0.009	
	(0.007)		(0.007)		(0.007)	
*hcl*3	0.005		0.005		0.005	
	(0.009)		(0.009)		(0.009)	
*hcl*4	0.024	**	0.025	**	0.021	**
	(0.01)		(0.01)		(0.01)	
west	-0.003		-0.003		-0.002	
	(0.012)		(0.012)		(0.012)	
AIC	-464.18		-464.97		-458.34	
R^2	0.7086		0.7106		0.693	
ln$L$	248.088		248.485		245.172	
LR-test	138.097		138.89		132.265	

Significance codes: '***', '**', '*' significant up to 1%, 5%, and 10%, respectively. Standard errors are below the estimates in parentheses.

The first column in table 5.3 is the same as the third column in table 5.2 because it contains the estimates for which change in productivity is measured by the Malmquist index. Columns two and three of table 5.3 contain the estimates in which the components of change in productivity is measured with change in efficiency in column two and change in technology in column three. Since the Malmquist index is the product of change in efficiency and change in technology the effect on prices without rents can be different. The results for the regression with the specification including change in efficiency are almost the same as those with the specification including the Malmquist index which implies that the productivity change as the Malmquist index is mainly driven by change in efficiency. Change in efficiency only accounts the relative change of the industry in each city in relation to the best practice frontier. Column two in table 5.3 shows that change in efficiency in the private non-financial services sector (GHI) is weakly significant and negative although the Malmquist index for that sector is not significantly different from zero since the effect of change in efficiency is offset by the effect of the change in technology in the third column.

The results in table 5.3 are the basic results of the analysis. Tests for heterogeneity are not found to reject the null hypothesis of no heteroscedasticity. The standard deviations are therefore used for the tests of significance for each coefficient. Furthermore, the RESET does not reject the null hypothesis of no misspecification. In addition, tests are carried out for the quadratic effects of only a few variables because business tax and the Gini coefficient of gross value added are without significant results.

It is possible to have spatial dependence due to interactions between neighboring cities. For this reason the models are tested for spatial patterns. The test is described in Mur and Angulo (2009) within the specific-to-general approach to conclude whether the SAM or SEM has to be used in the case where spatial patterns are present. The results of those tests are presented in table 5.4.

Table 5.4: Spatial pattern test for prices without rents

	statistic	df	p-value
LMerr	12.877	1	0
LMlag	20.322	1	0
RLMerr	0.638	1	0.424
RLMlag	8.084	1	0.004

The tests are described in Anselin et al. (1996). Since cities are loosely distributed in the area and are not always closely located next to other cities, they are classified as neighbors if they are in a cycle with a radius of 100 km around the city. The maximum distance of 100 km was chosen, since it is within a maximum travel time for commuters and is within the range indicated by Kosfeld and Eckey (2008) and Niebuhr (2006). Furthermore, since only German cities are counted as neighbors within 100 km, every city has at least one neighbor, even those in remote areas at the German border.

As table 5.4 shows, the SAM and the SEM are significant since the p-values of LMlag and LMerr are below 5%. The higher value of the Lagrange Multiplier test for the spatial autoregressive model (LMlag) indicates the preference of the SAM. The SAM is also preferred by the restrictive tests as the restricted Lagrange Multiplier test of the lags is also significant in contrast to the error model.

For these reasons the spatial autoregressive model was estimated and the results for prices without rents are reported in table 5.5. In table 5.5 ρ captures the spatial autocorrelation of the prices without rents of a city and its neighbor weighted pendants.

Table 5.5: Spatial autoregression results for prices without rents

	Malmquist		efficiency		technology	
Intercept	3.054	***	3.068	***	2.889	***
	(0.654)		(0.658)		(0.66)	
disp income	0.155	***	0.158	***	0.161	***
	(0.023)		(0.023)		(0.024)	
stu	0.001	**	0.001	*	0.001	
	(0.001)		(0.001)		(0.001)	
pop^2	0.001	***	0.001	***	0.001	***
	(0)		(0)		(0)	
CDU	-0.016		-0.018		-0.016	
	(0.011)		(0.011)		(0.011)	
GVA Gini	0.06	***	0.061	***	0.049	***
	(0.012)		(0.012)		(0.012)	
BusTax	0.002		0.002		0.003	
	(0.003)		(0.003)		(0.002)	
dea_ JK	0.072	***	0.073	***	0.05	***
	(0.02)		(0.02)		(0.017)	
change_ CDE	0.026	**	0.026	**	0.071	
	(0.013)		(0.013)		(0.102)	
change_ GHI	-0.011		-0.022		0.02	
	(0.026)		(0.028)		(0.086)	
change_ JK	-0.048	**	-0.045	*	-0.442	
	(0.024)		(0.024)		(0.272)	
hcl2	0.019	***	0.019	***	0.015	**
	(0.006)		(0.006)		(0.006)	
hcl3	0.008		0.008		0.009	
	(0.007)		(0.007)		(0.007)	
hcl4	0.023	***	0.024	***	0.02	**
	(0.008)		(0.008)		(0.008)	
west	-0.016		-0.016		-0.015	
	(0.01)		(0.01)		(0.01)	
ρ	0.424	***	0.419	***	0.438	***
	(0.081)		(0.082)		(0.082)	
$RS_{\lambda\backslash\rho}$	0.5026		0.5546		0.003	
	[0.4784]		[0.4564]		[0.9869]	
AIC	-482.44		-482.04		-477.91	
R^2	0.7442		0.7442		0.7442	
$\ln L$	258.221		258.019		255.956	
LR-test	20.266		19.068		21.568	
Wald-test	27.445		26.225		28.449	

Significance codes: '***', '**', '*' significant up to 1%, 5%, and 10%, respectively. Asymptotic standard errors are below the estimates in parentheses and p-value for Lagrange multiplier test for residual autocorrelation in square brackets below the test statistic.

The results in table 5.5 show that changes in productivity measured by the Malmquist index and its component, the change in efficiency, are only significant for industry of wide manufacturing (CDE) as well as the financial and business services sector (JK). Change in technology is not significant for any sector, which indicates that the effects of productivity change are driven mainly by the effect of the change in efficiency. The effect of efficiency change in the private non-financial service sector, which was significant in the OLS estimation (table 5.3), is induced by spatial influences that are eliminated within the SAM. The static relative productivity measure is also only significant for the financial and business services sector with a positive coefficient, which is in accordance to the Balassa-Samuelson effect causing wages to be higher but not over all wages in the city (thus not captured by the disposable income variable) and thus results in higher prices. State capital cities and additionally Frankfurt/Main are those cities with the highest efficiency in the financial and real estate services sector. Frankfurt/Main, as the main financial center in Germany, is the location of many company headquarters and the German as well as European Central Bank with wages considerably above average wages seeming to drive local price levels significantly. Furthermore, the geographical localization of a city in East or West Germany as implemented by the west-dummy variable has no significant effect on the prices without rents in German cities. Contrarily, urban hierarchy measured by the service to manufacturing ratio and allocated to four different clusters is significant with a positive coefficient for the second and fourth cluster. This is in accordance with the considerations in the theoretical section, although the third cluster is not significant.

Due to spatial dependence, the coefficients should not be interpreted as those in the OLS results.

The spatial autoregressive parameter ρ is significantly positive as already expected by the Lagrange multiplier test of the OLS models. Because of the low differences of the change in efficiency components, all three columns in table C2 reveal almost the same information, which results in similar or equal performance information at the bottom of table C2. Below the second line is the Lagrange multiplier test for spatial autocorrelation in the error term ($RS_{\lambda\backslash\rho}$) with the p-values in square brackets below the estimates. This test has to be carried out in the specific-to-general modeling approach as Mur and Angulo (2009) show. The test statistic is asymptotically $\chi^2(1)$ distributed. The resulting p-values of the test statistics show that the null-hypothesis of no spatial autocorrelation in error terms cannot be rejected. Therefore, the SAM model is appropriate and preferred to the SDM model. Due to the fact that the SDM is not needed, the model seems not to be afflicted by omitted variables, as pointed out in section 2.3. The adjusted R^2 calculated by the log-likelihood ($\ln L$) is 74% and thus fairly large for cross-sectional analysis with a small number of parameters. A similar analysis is performed by accounting for the differences between the German federal states and presented in the Appendix C.

To test for spatial dependence, I estimated the SAM with weight matrix that includes weights for cities within a radius of 200 km around each city. It transpires that the spatial autocorrelation parameter ρ is not significantly different from zero for that neighboring classification. This indicates that not only the radius of 100 km around a city is sufficient but also that the spatial interaction decreases by distance. This is similar to the findings of Alecke et al. (2011), who investigate regional German labor productivity and show that, for labor market effects measured by spatial autocorrelation with *Getis and Ord's G*, the effects are at a maximum for about 100 km although still significantly present until a distance of 150 km.

As already mentioned in section 2.3, the coefficients in the SAM do not immediately represent the effect of the independent variable on the dependent variable due to the feedback loops with neighboring cities. The following tables 5.6 to 5.8 present the comparisons of the direct impacts in the first three columns, the OLS estimations in the fourth to sixth columns, and the total impacts in the last three columns. For each of the estimates the 90% confidence range is calculated with the lower five percentage quantile in the first row, the estimate which is the median quantile in the second row and the upper five percentage quantile in the third row for every coefficient and each impact. This procedure allows for a comparison of the impacts and the OLS estimates and incorporates the uncertainty in the estimations and is similar to the traditional Hausman test, which is only implemented for the SEM but not the SAM, although the Hausman test measures the difference of all coefficients simultaneously.

Table 5.6: Comparisons between impacts and OLS estimates for the Malmquist index

	direct impact			OLS estimates			total impact		
	5% Qu.	Median	95% Qu.	5% Qu.	Median	95% Qu.	5% Qu.	Median	95% Qu.
Intercept	1.926	3.173	4.453	5.796	6.244	6.692	4.292	5.304	6.026
disp income	0.119	0.161	0.206	0.128	0.173	0.219	0.194	0.269	0.376
stu	0	0.001	0.003	0	0.001	0.002	0.001	0.002	0.004
pop^2	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.002
CDU	-0.039	-0.016	0.008	-0.045	-0.023	-0.002	-0.062	-0.027	0.008
GVA Gini	0.039	0.062	0.088	0.036	0.061	0.085	0.068	0.104	0.148
BusTax	-0.004	0.002	0.007	-0.003	0.002	0.008	-0.005	0.003	0.01
dea_ JK	0.035	0.075	0.115	0.042	0.081	0.12	0.062	0.126	0.206
change_ CDE	0.001	0.027	0.052	0.008	0.033	0.059	0.009	0.046	0.081
change_ GHI	-0.061	-0.012	0.041	-0.09	-0.039	0.012	-0.088	-0.02	0.056
change_ JK	-0.095	-0.05	-0.003	-0.106	-0.058	-0.01	-0.165	-0.084	-0.016
hcl2	0.007	0.02	0.032	0.003	0.016	0.028	0.015	0.033	0.056
hcl3	-0.007	0.009	0.024	-0.009	0.005	0.019	-0.008	0.015	0.038
hcl4	0.009	0.024	0.042	0.008	0.024	0.04	0.016	0.041	0.07
west	-0.038	-0.017	0.005	-0.022	-0.003	0.017	-0.066	-0.028	0.004

As can be seen in table 5.6, the impacts are not significantly different from the OLS estimates, except in only one case. This case is the intercept as direct impact. The intercept as direct impact is much too low compared to the OLS estimate such that the 90% confidence intervals are not overlapping. Furthermore, the confidence intervals show for which variable the impact is significantly different from zero in the case zero is not included within the confidence interval. For example if disposable income increases within one city by one percent, price level is going to be higher between 0.119 and 0.206% (direct impact), whereas if disposable income increases by one percent in all cities, price levels without rents are between 0.194 and 0.376% higher (total impact) or on average 0.269%. This means that a doubling of disposable income in all cities, as it might be the goal of a labor unit in the collective negotiations, prices without rents increase on average by 26.9%. A doubling of the Gini coefficient of the gross value added within one city implies an on average six percent higher price level. A doubling of efficiency in the finance and business service sector denotes an eight percent higher price level without rents on average, whereas the change of productivity within this sector is denoted by a significant lower price level without rents.

Table 5.7: Comparisons between impacts and OLS estimates for change in efficiency

	direct impact			OLS estimates			total impact		
	5% Qu.	Median	95% Qu.	5% Qu.	Median	95% Qu.	5% Qu.	Median	95% Qu.
Intercept	2.147	3.184	4.142	5.723	6.174	6.626	4.338	5.278	5.959
disp income	0.128	0.164	0.2	0.135	0.181	0.227	0.201	0.272	0.371
stu	0	0.001	0.002	0	0.001	0.002	0	0.002	0.004
pop^2	0.001	0.001	0.001	0	0.001	0.001	0.001	0.001	0.002
CDU	-0.035	-0.018	-0.002	-0.048	-0.027	-0.005	-0.058	-0.03	-0.005
GVA Gini	0.043	0.064	0.083	0.039	0.063	0.087	0.068	0.105	0.154
BusTax	-0.003	0.002	0.006	-0.002	0.003	0.008	-0.005	0.003	0.011
dea_ JK	0.041	0.075	0.113	0.041	0.081	0.12	0.065	0.125	0.206
change_ CDE	0.004	0.027	0.047	0.007	0.032	0.058	0.007	0.044	0.085
change_ GHI	-0.072	-0.023	0.027	-0.119	-0.063	-0.007	-0.119	-0.039	0.047
change_ JK	-0.083	-0.046	-0.002	-0.098	-0.05	-0.002	-0.149	-0.077	-0.003
hcl2	0.009	0.02	0.031	0.004	0.016	0.029	0.015	0.033	0.055
hcl3	-0.003	0.009	0.023	-0.009	0.005	0.019	-0.005	0.014	0.039
hcl4	0.011	0.024	0.039	0.009	0.025	0.041	0.017	0.041	0.071
west	-0.036	-0.017	0.001	-0.023	-0.003	0.016	-0.068	-0.028	0.001

Similar to table 5.6, table 5.7 shows the same pattern, which demonstrates again that the estimates for productivity change in total is mainly driven by the change of efficiency of the inspected sectors. The interpretation of the coefficients and the impacts is therefore the same as for table 5.6.

Table 5.8: Comparisons between impacts and OLS estimates for change in technology

	direct impact			OLS estimates			total impact		
	5% Qu.	Median	95% Qu.	5% Qu.	Median	95% Qu.	5% Qu.	Median	95% Qu.
Intercept	1.852	3.011	4.036	5.667	6.137	6.607	4.077	5.137	5.854
disp income	0.132	0.168	0.209	0.137	0.184	0.232	0.215	0.286	0.407
stu	0	0.001	0.002	-0.001	0.001	0.002	0	0.001	0.003
pop^2	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.002
CDU	-0.035	-0.017	0.001	-0.041	-0.02	0.001	-0.061	-0.029	0.001
GVA Gini	0.032	0.051	0.072	0.018	0.043	0.068	0.051	0.088	0.137
BusTax	0	0.003	0.007	0	0.004	0.009	-0.001	0.006	0.013
dea_ JK	0.025	0.052	0.081	0.016	0.051	0.086	0.042	0.089	0.159
change_ CDE	-0.102	0.074	0.252	-0.073	0.132	0.336	-0.179	0.126	0.459
change_ GHI	-0.13	0.021	0.169	-0.108	0.064	0.237	-0.233	0.036	0.281
change_ JK	-0.961	-0.461	0.019	-1.165	-0.617	-0.069	-1.699	-0.787	0.029
hcl2	0.005	0.016	0.028	-0.003	0.009	0.022	0.009	0.027	0.051
hcl3	-0.003	0.009	0.022	-0.01	0.005	0.02	-0.005	0.015	0.038
hcl4	0.005	0.021	0.035	0.005	0.021	0.038	0.008	0.036	0.063
west	-0.035	-0.016	0	-0.022	-0.002	0.017	-0.067	-0.027	0

The last table 5.8, which includes the quantiles for the regression with change in technology, shows once again the same pattern as shown in the tables 5.6 and 5.7 for the regression model, which include the Malmquist index and change in efficiency, respectively. The only difference is that the impacts of the sectoral change in technologies are not significantly different from zero. This means that neither the change in technology within one city nor the change in technology for all cities has any impact on prices without rents. Technological change, which corresponds to changes in technology within the production process, does not denote different prices without rents. Altogether the comparison shows that the estimates are not significantly different in total, which indicates a correct specification and the significant higher likelihood of the SAM demonstrates the advantage over the OLS model, although the 90% confidence intervals are narrow. The price elasticities are higher than those in Roos (2006), who applied OLS estimations on estimated price data for all German regions.

The impacts are significantly different compared to the estimates in the SAM. The effect is driven by the significant spatial autocorrelation parameter ρ. The impact estimations are a combination of both the spatial autocorrelation parameter and the spatial weighting matrix. It shows that spatial interaction is a significant occurrence

Tests for outliers, which are explained in Mur and Lauridsen (2007), reveal no spatial outliers within the data. In addition and because the LMerr test was significant the SEM result are presented as robustness check.

Table 5.9 presents the results for the spatial error model with the Hausman test which has 15 degrees of freedom in the case of three cluster dummies accounting for urban hierarchy and 13 degrees of freedom if the service to manufacturing ratio (*SMr*) is used.

Table 5.9: Spatial error model results for price level without rents

	Malmquist		efficiency		technology	
Intercept	6.536 ***	6.464 ***	6.51 ***	6.436 ***	6.471 ***	6.396 ***
	(0.233)	(0.242)	(0.239)	(0.248)	(0.249)	(0.253)
disp income	0.14 ***	0.146 ***	0.143 ***	0.149 ***	0.147 ***	0.155 ***
	(0.023)	(0.024)	(0.024)	(0.025)	(0.025)	(0.026)
stu	0.001	0.001	0.001	0.001	0	0
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
pop^2	0.001 ***	0.001 ***	0.001 ***	0.001 ***	0.001 ***	0.001 ***
	(0)	(0)	(0)	(0)	(0)	(0)
CDU	-0.019	-0.025 *	-0.02	-0.026 *	-0.021	-0.025 *
	(0.014)	(0.014)	(0.014)	(0.014)	(0.014)	(0.014)
GVA Gini	0.052 ***	0.055 ***	0.054 ***	0.057 ***	0.049 ***	0.051 ***
	(0.012)	(0.012)	(0.012)	(0.012)	(0.013)	(0.012)
BusTax	0.003	0.004	0.003	0.004	0.004 *	0.005 *
	(0.003)	(0.003)	(0.003)	(0.003)	(0.002)	(0.003)
dea_JK	0.056 ***	0.046 **	0.057 ***	0.046 **	0.039 **	0.031 *
	(0.019)	(0.02)	(0.02)	(0.02)	(0.017)	(0.018)
change_CDE	0.031 **	0.025 *	0.03 **	0.025 *	0.022	0.044
	(0.013)	(0.013)	(0.013)	(0.013)	(0.105)	(0.108)
change_GHI	-0.01	-0.008	-0.018	-0.017	0.008	0.014
	(0.027)	(0.028)	(0.029)	(0.03)	(0.084)	(0.087)
change_JK	-0.043 *	-0.039	-0.042 *	-0.036	-0.254	-0.376
	(0.023)	(0.024)	(0.024)	(0.025)	(0.262)	(0.268)
hcl2	0.015 ***		0.016 ***		0.013 **	
	(0.006)		(0.006)		(0.006)	
hcl3	0.003		0.003		0.002	
	(0.007)		(0.007)		(0.007)	
hcl4	0.024 ***		0.024 ***		0.019 **	
	(0.008)		(0.008)		(0.009)	
SMr		0.012 **		0.012 **		0.011 *
		(0.006)		(0.006)		(0.006)
west	-0.006	0.001	-0.006	0.001	-0.003	0.004
	(0.013)	(0.013)	(0.013)	(0.013)	(0.013)	(0.013)
λ	0.57 ***	0.529 ***	0.569 ***	0.527 ***	0.543 ***	0.482 ***
	(0.125)	(0.129)	(0.128)	(0.126)	(0.126)	(0.141)
AIC	-477.42	-473.04	-477.47	-472.85	-469.95	-469.35
R^2	0.7457	0.7259	0.7458	0.7254	0.7281	0.7167
log*Lik*	255.708	251.52	255.734	251.426	251.974	249.674
LR-test	15.239	12.459	14.499	11.849	13.604	9.954
Hausman-test	29.020 **	22.517 **	31.899 ***	23.485 **	29.742 **	28.323 ***

Significance codes: '***', '**', '*' significant up to 1%, 5%, and 10%, respectively.
Standard errors are below the estimates in parentheses.

The R^2 reported in 5.9 are adjusted R^2 values from Nagelkerke (1991). The λ values for the spatial lag model are tested for significance by likelihood ratio statistics with the non-spatial model under the null hypothesis. It is important to notice, that the null hypothesis of the Hausman test is rejected in all SEM estimations. The null hypothesis of the Hausman test is that there is no bias in the OLS estimation and both methods, OLS and SEM, should return similar estimates. A rejection of the null hypothesis of the Hausman test indicates that the SEM is not the correct model and indicates a misspecification. This is due to the fact that the data generating process is a SAM and thus the OLS estimates are different.

5.6 Summary

The analysis in this chapter explains the price level differences without rents between German cities. Price levels without rents are spatially autocorrelated which brings into doubt previous analyses ignoring spatial patterns. For example, price levels are found significantly larger in West Germany in other investigations with OLS estimation including the economic explanatory variables for all German regions (BBSR (2009)). The price level without rents is not systematically larger in West Germany if one accounts for economic and political city-specific factors. Furthermore, regional prices without rents are higher in cities which experienced higher productivity increases over the past five years in the wide manufacturing sector. In contrast, prices without rents are lower in cities with higher productivity change in the financial and business services sector. This is also interesting, since it is the only sector for which the static relative productivity measure has a significant effect on prices without rents. The effects of the change in productivity are only driven by changes in the production process relative to the best practice production function (change in efficiency). Technological changes, which improve the best practice production function, do not correspond with different price levels. Regarding urban hierarchy, cities occupying the highest position in the hierarchical order have the highest prices although price levels are not rendered increasingly linear or quadratic by the service to manufacturing ratio.

Compared to the results in BBSR (2009) the estimates in this investigation differ. This can be explained by three reasons. First, in this chapter spatial interdependence is considered by estimating spatial autoregressive models. Second, this analysis mainly focuses on the price level without rents and not the price indexes. Third, this investigation includes only cities and not all regions including every rural area which is done in BBSR (2009).

6 Industrial Growth and Productivity Change in German Cities - A Multilevel Investigation

6.1 Motivation

"Why do some cities perform better than others?" remains a puzzling question in urban economics for rational actors. For example, entrepreneurs looking where to establish a new firm, consider different regional factors and might ask: Which city incorporates valuable opportunities and increases the productivity of the labor force to increase profit? Local governments are interested in how to set local variables to attract new firms within their areas and to increase the profit of the existing firms to gain more tax revenue and further increase the attractiveness of their own area. These questions are almost the same as for multicountry analyses understanding why some countries are poorer than others and do not converge as expected. Moreover, entrepreneurs have to think about settlement in different countries as well as national governments try to increase the attractiveness of their own countries for foreign and domestic firms. This is important, especially to open economies with long-run, future-oriented governments which compete with each other.

However, these questions are in fact somewhat different in national urban decisions, in which the major economic circumstances are the same and other barriers are absent, for instance laws, political uncertainty and language difficulties. These forces are known to reduce the attitude of movements among countries and even within economic unions. Thus, altogether the decision for settlement is set by many local characteristics and is determined by various circumstances as well as being predetermined by local governmental parameter settings. By winning new companies and increasing the productivity of established industries, industries within a city grow faster. Local urban politicians try to foster industrial growth by setting the parameters of the local business environment. But what forces really affect local industrial growth? What parameter settings are optimal for local industrial growth? And is there any difference between those parameters with reference to either value added or employment growth? The goal of this chapter is to find answers to these questions. The analysis is rooted in the literature on urban endogenous growth that is fueled by technological improvements and their effects on the change of sectoral composition. Technological improvements are not just technical changes by the generation of new ideas, but also efficiency changes and catching-up by imitating technologies. Boschma and Lambooy (1999) present the framework of technical change in a regional context by the evolution of regional economics, called 'evolutionary economic geography' by Fratesi (2010) and Boschma and Frenken (2011).

In this chapter, value added growth and employment growth are analyzed, whether productivity change contributes to a rise or causes creative destruction. Since Schumpeter (1934, 1939) innovations as the implementation of innovative ideas are known as the major drivers of economic development. Furthermore, he emphasizes the different levels of activity, namely the micro-, meso- and macro-spheres jointly within the economic development process. In this chapter, the results of innovations are utilized, namely the increase in productivity and efficiency. Productivity and efficiency changes are implemented by a bias-corrected Malmquist index and its components, which are generated by the results of non-parametric data envelopment analysis, as described in Wheelock and Wilson (1999).

Additionally, the growth path estimation model is extended by local variables such as public expenditure and business taxation. These variables might have an effect on the productive performance of entrepreneurs within the city and the settlement decisions for new firms, which therefore change growth patterns. The data set contains 112 German cities with independent local political authorities over the period 1998 until 2007. Furthermore, and in tradition of urban economic analysis, the investigation contains variables for concentration, diversification and city size, and indicators related to technology and the knowledge base. The seminal paper is Glaeser et al. (1992) and the subsequent work of Henderson et al. (1995) and Henderson (1997a), who estimate sector-specific regressions for different sectors, explaining employment change by local industrial and city-specific variables.

The contribution of this investigation consists of several extensions and refinements of this type of analysis. Primarily, three major extensions are incorporated. First, the role of productivity changes using bias-corrected Malmquist components estimated by data envelopment analysis is explored. Second, structural change is investigated, because aggregate economic growth is inevitably associated with structural change. Third, non-linearities are considered (i.e., interaction effects and quadratic effects) to expand the linear model to a more generalized version and to reveal the optimal conditions for future industry growth. The main refinement consists of the adoption of multilevel models to account for the nested structure of the data, the unobserved city-specific effects and to estimate unbiased estimates. Thus, these regressions are estimated by using multilevel analysis methods to account for the importance of the meso- and macro-spheres. This method allows varying coefficients on each level, which are industries as the first level, cities as the second level and time as the third level. Multilevel models include fixed as well as random effects for considering the dependency structures on each level, as explained in Raudenbush and Bryke (2002). It is possible to include exogenous variables in the estimation, which are observed at different levels, like city-specific variables as well as industry-specific variables which are nested within cities.

The purpose of this chapter lies in the identification of the effect of productivity and efficiency changes on value added growth and employment growth within different industries in German cities. Tests are carried out to investigate whether those effects of productivity changes vary between cities and over time. The results provide mixed evidence on the nature of the value added growth and employment growth. On average, and over all the industries included, historical changes affect value added growth and employment growth, supporting creative destruction. In addition, non-linearities seem to be characteristic features of several explanatory variables. As a consequence, some local political parameters seem to have a minimum point which leads to a decrease of value added growth and employment growth.
The chapter is organized as follows. The next section presents the related literature. Section 6.3 clarifies the data used in the estimations, and section 6.4 gives a brief overview of the applied methods. The empirical results are presented and analyzed in section 6.5. At the end of this chapter, a short conclusion is drawn in section 6.6.

6.2 Literature Review

Evolutionary economic geography is classified by Fratesi (2010), who shows the connection of regional innovations and dynamics via competitiveness. He furthermore emphasizes its roots in meso-economic applications, although it is possible to extend the analyses to regional micro- as well as macro-economics. Regional innovations and their effect on competitiveness are crucial elements in evolutionary economic geography, but they also provide feedback on income growth with innovations in a dynamic process. That feedback is often used, e.g. in Gaffard (2008), who incorporates the Schumpeterian ideas of creative destruction within a Hicksian framework in which competition increases efficiency, and therefore increases the effect of innovations and growth. The creative destruction and the contributions of innovation to regional growth are theoretically implemented in dynamic analyses by Batabyal and Nijkamp (2012, 2013), in which innovations and technological progress are key divers in the regional growth path.

Frenken and Boschma (2007) build an analytical framework for the effects of innovations on firm and city growth with innovations generated exogenously. Innovations affect urban growth by increasing urban diversification. They notice that the correlation between size and innovation might be caused by the correlation between size and diversification. The positive feedback relationship is non-linear because of the routines in evolutionary economic developments. Furthermore, they incorporate negative feedback effects on urban growth, where cities and industries decline without innovations. The theoretical analysis in Martin and Sunley (2006) connects regional path dependency and lock-in effects within a region, which could be positive by stimulating innovations and increasing economic performance. It could also be negative by creating negative externalities through inflexibility and reducing economic performance, institutional hysteresis, local external economies of industrial specialization, economics of agglomeration, and region-specific institutions.

Noseleit (2013) estimates the relationship between structural change and concentration measure namely, the Gini coefficient, on employment growth for West German regions and agglomerations. He finds a negative effect of the Gini coefficient on employment growth in agglomerations and the structural change, measured by the similarity between entries and exits of firms, which has negative effect on employment growth.

Illy et al. (2011) investigate employment growth for German cities with respect to the local economic structure. They find U-shaped functional forms of specialization and size on employment growth of the free German cities for the period 2003-2007.

The necessity for different levels of activity, namely the micro-, meso-, and macro-levels, is demonstrated in Rozenblat (2012) for the agglomeration economies of firms in international cities. The importance of intercity networks is emphasized with respect to agglomeration economies, because interaction takes place between people and institutions at the micro-level. These micro-level interactions affect urban externalities by city size and growth at the meso-level. However, Rozenblat does not identify location economies emerging from cities' specialization.

One elegant way to implement different levels of activity is by using multilevel or mixed-effects models. The multilevel analysis is already a wide-spread feature, applied in different academic fields like biology, as in the various analyses assembled in Zuur et al. (2009); or sociology as variously shown in Hox (2002). So far, it is rarely implemented in economics; especially in urban economics, even the observed data are predestined for that kind of investigation. For example individuals at the same level, for instance within a city, are likely to interact and are faced with the same environmental factors, which might be observed or unobserved. Therefore, those individuals are endogenously dependent, which leads to biased estimates. This problem can be solved by mixed-effects models. Mixed-effects models are explained in Pinheiro and Bates (2000) and in section 2.4 and account for each of the levels and nesting structures within the observations.

A regional multilevel model is estimated by Srholec (2010). He investigates the likelihood of innovations in the Czech Republic within a two-level Logit approach. As explanatory variables he includes a bunch of local variables, like population density, urbanization, average wage, long-term unemployment, number of murders, as well as a few other variables, and builds three factors for these. With those factors, he calculates a basic multilevel model with fixed and random effects for all factors and the intercept. Next, he extends the basic multilevel model to the so-called intercept-as-outcome model by additionally explaining the fixed effect intercept of the model. Furthermore, he generalized that model to the so-called 'slope-as-outcome' model by adding the explanatory factors for the slope estimates of each factor. In doing so, he includes all possible interaction terms to consider non-linearities. Unfortunately, he does not include any test for the model or at least any measure for the explanatory power of the model like the likelihood of the model or a resulting information criterion or likelihood ratio test. He only includes an index of dispersion that is not helpful to see whether the extensions add substantially explanatory power to the model or only increase the uncertainty of the estimates. For reasons of parsimony, likelihood ratio tests of the different models would be fruitful.

Other multilevel analyses within economics include Giovannetti et al. (2009), Goedhuys and Srholec (2010), and Srholec (2011). Giovannetti et al. (2009) analyze firm performances in Italy. They test the necessary use of the provincial level by a likelihood ratio test and conclude that the multilevel model is appropriate. They show that the effects of provincial variables like social capital have a larger effect on smaller firms than on larger firms, which still remains significant. Thus, location variables have to be considered like local governmental expenditures, as well as local circumstances like airports and other transportation facilities.

Goedhuys and Srholec (2010) perform another application of multilevel analysis within economics. They use a two-level model to analyze productivity at the firm-level within different countries. The investigation includes the stepwise approach similar to that in Srholec (2010). However, the productivity is estimated for a Cobb-Douglas production function, therefore, all the parameters of the production function have to be estimated. To derive accurate productivity measurements from the production function, all estimated parameters should be unbiased. Therefore, they apply a multilevel approach to reduce the bias resulting from the nested structure. Nevertheless, the functional form is also questionable, as indicated by analyses using translog functions as generalized versions of the Cobb-Douglas production function.

Srholec (2011) investigates the likelihood of innovations in 32 developing countries, which is similar to his analysis in 2010. He uses a two-level Logit model and finds necessary support for adoption of multilevel analysis, because country-specific variables contribute to the explanatory power for the likelihood of a successful innovation. Nonetheless, he finds no empirical evidence of an effect of population size on innovation. He records a highly significant negative estimate for the local income tax rate. Furthermore, he shows that the explanatory power soars with the random effects.

6.3 Data

The general data set is explained in subsection 2.1. In addition, population figures are taken from the regional database of the Statistical Offices in Germany. Under German registration law, a person is only added for a city if they have their principal residence within that city. So, the figure does not account for people with secondary residences in order to avoid double counting, even though many people have a secondary residence in a city and are part of its productive employees. Nonetheless, the use of the population figures for the number of inhabitants within a city is reasonable, since people who spend more than half of their time in the city are required to have their principal residence in that particular city.

Comparable studies estimate the effects of various additional variables on productivity growth by least squared methods. These analyses involve different city-specific variables, which might not have only linear effects on value added growth and employment growth. To account for the non-linear relationships proposed by Frenken and Boschma (2007), these factors are additionally included with quadratic as well as interaction terms within the linear regression, to test the significance of these terms. The factors are observed variables as, e.g., population changes ($dPop$) and the number of students within each city. For the analysis, the data has been transformed to become narrower. This is done for the number of students by taking the logarithm ($\ln Stu$). However, there are several cities in the sample with no University or University of Applied Science at all, so the amount of one is added to each city, which results in positive figures for the logarithm of all students.

According to Frenken and Boschma (2007), a growing city is assumed to have negative feedback slopes on employment growth and value added growth in the industries if there is no innovative activity within the city. The number of students represents the knowledge base within the city and serves as a measure of the ability to implement and generate innovation. Therefore, a larger number of students should be correlated with larger value added and employment growth, and might interact with a technological progress measure. Additionally, I propose a variable indicating the composition of the industry within each city. Urban analyses find support for the view that homogeneous distribution for industries support the generation and flow of new ideas by the localization externalities. The Gini-coefficient ($Gini$) is calculated on the basis of employment of a more disaggregated level by ten industries, which is observed and supplied by the federal labor agency of Germany.

Furthermore, a factor for the change of the structural composition of the industries within each city is calculated by the modified Lilien-index (SC), which indicates to what extent the change within one year has taken place, and is measured by

$$SC_{jt} = \sqrt{\sum_{i=1}^{10} x_{ijt} \cdot x_{ij(t-1)} \left(\ln \frac{x_{ijt}}{x_{ij(t-1)}} \right)^2}, \tag{72}$$

with x_{ijt}, the share of industry i in city j at time t, and the sum of all industries equals one for each city and every year. This measurement is used and discussed in the literature examining structural change, e.g., Stamer (1999) and Dietrich (2009). It is a dispersion index, in which smaller sectors and sectors with lower growth are considered with a smaller weight. That structural change measure is also calculated on the basis of the employment figures of the 10 disaggregated industries.

Additionally, spatial variables implemented include the whole area as well as the share of recreational area to the total area within each city. A larger recreational area within a city enables workers to recreate faster and thereby increases labor productivity or contributes to growth.

As an additional feature, the German tax system enables every city to set its own local business tax ($BusTax$) (by setting its own so-called 'Hebesatz', a collection rate in Germany) as well as its own tax on land and buildings ($LTax$), which is by German tax law a tax on land and buildings for non-agriculture land-use (so-called 'Grundsteuer B' in Germany). On the one hand, cities with higher taxes increase the costs of living and production within that city and thereby attract firms with higher productivity. On the other hand, cities with higher income are also able to spend more on infrastructure, education, administration, and so on. Although these variables are significant in some studies, expenses on transportation facilities, tax on land and building and the recreational area share have not been proven to be significant in this investigation for German cities and have, therefore, been excluded.

All the variables in the analysis in section 6.5 are observed over the years 1997 until 2008. Descriptive statistics are given in table 6.1 with the number of students measured in thousand and industrial growth rates in percentage changes.

Table 6.1: Descriptive statistics

Variable	Min.	1st Quartile	Median	Mean	3rd Quartile	Max.	s.d.
gross value added growth	-0.829	-0.019	0.017	0.015	0.054	0.901	0.083
employment growth	-0.330	-0.026	0	-0.004	0.018	0.571	0.047
students	0	12.2	43.2	58.91	82.45	250.4	58.003
population change	-20970	-659.8	-56.5	149.7	441.2	47880	2939.2
industrial Gini coefficient	0.415	0.489	0.535	0.544	0.580	0.748	0.070
structural change	0.003	0.009	0.012	0.016	0.016	0.465	0.024

Table 6.1 shows that there are many cities with a low number of students as a measure for the local knowledge base, which results in a median which is considerably lower than the mean. In addition, the standard deviation (s.d.) is very large for the number of students. Furthermore, all input variables as well as value added as output are non-negative, as required in DEA. The growth rates of gross value added have a mean and a median which are positive, meaning that on average value added is growing. Employment growth is zero on average, which indicates that there is on average no change in employment for all industries and cities. Table 6.1 also shows that the Gini coefficient as a concentration measure is in a narrow band. There is no extreme observation and, therefore, no absolute concentration with a city with only one industry and also no absolute equally distributed industry share. The structural change index has the value of almost zero for most cities, indicating almost no change in industry shares for subsequent years, but at least there is always some change. The descriptive statistics also show a high fluctuation among cities by massive changes in population.

6.4 Theory

6.4.1 Productivity Change

Productivity change is estimated by non-parametric DEA as the Malmquist index and its components as described in subsection 2.1.2. To test whether the variable returns to scale measure is advisable, Simar and Wilson (2002) propose different non-parametric tests for returns to scale based on the bootstrap algorithms of Simar and Wilson (1998). I run two different tests, one with the mean of efficiency for all cities (used, e.g. in Cullmann and von Hirschhausen (2008)), and one with efficiency for each city separately (used, e.g., in Badunenko (2010)). Each test is carried out for testing first the null hypothesis of constant returns to scale against decreasing returns to scale, and second for non-increasing returns to scale against increasing returns to scale. For using the variable returns to scale measurements, both null hypotheses must be rejected. It turns out that for the means of scale efficiency over all cities, both null hypotheses can be rejected. The null hypothesis of constant returns to scale can be rejected for all industries in every year. In addition, the null hypothesis of non-increasing returns to scale is also rejected for every industry in each year. Moreover, the second test for all cities separately generally rejects both null hypotheses. Tables 6.2 and 6.3 show the results for both null hypotheses, with the percentage share of cities for which the null hypotheses cannot be rejected, depending on the sector and year.

Table 6.2: Results for Simar and Wilson (2002) test for constant returns to scale

Year	CDE	D	F	GHI	JK	LMNOP
1999	3	0	2	1	2	1
2000	0	0	0	0	1	1
2001	0	1	1	1	1	1
2002	2	1	1	0	1	2
2003	2	2	0	0	0	0
2004	3	0	2	0	0	1
2005	3	2	0	0	1	1
2006	1	2	3	0	1	1
2007	0	1	0	0	1	1

Table 6.3: Results for Simar and Wilson (2002) test for non-increasing returns to scale

Year	CDE	D	F	GHI	JK	LMNOP
1999	1	2	13	3	58	39
2000	9	9	26	4	67	33
2001	15	16	14	7	63	49
2002	23	22	34	5	59	38
2003	5	19	41	5	51	40
2004	17	21	21	7	60	45
2005	7	29	15	6	60	34
2006	4	10	20	4	57	44
2007	2	5	16	8	36	39

Both tables 6.2 and 6.3 show that the null hypotheses are not rejected in just a few cases. However, there are many cases for the second test of non-increasing returns to scale in some industries especially for financial and business services (JK) and in public and social services (LMNOP). These findings support the test, which rejects the null hypothesis of non-increasing returns to scale for all cities together. Therefore, the results overall indicate that the underlying production function is characterized by variable returns to scale, and that the detailed decomposition of the Malmquist index proposed by Wheelock and Wilson (1999) is possible.

Productivity change results from technological change and change in efficiency and is the observable achievement of innovative activity. In evolutionary economics, innovations are key drivers of economic growth although there is a creative destruction component of innovation, as already mentioned by Schumpeter (1934). Productivity change and its components should therefore have a positive effect on value added growth but also a negative effect on employment growth caused by of the creative destruction. Of course, the effects of innovation do not lead to a linear increasing development for the number of firms or the demand, both decline after some periods, as shown for example by Saviotti and Pyka (2004). Also concordant to Schumpeter (1939), business cycle and product life cycle developments induce a decline in economic development after an increase caused by innovations. So, the effect of productivity change on value added and employment growth depends on the considered time frame.

6.4.2 Multilevel Models

The estimation models are the multilevel models as explained in subsection 2.4. Multilevel models account for the variation caused by unobserved variables within each level by random variables. Adding further random variables increases the explanatory power of the model but also uncertainty on the estimates. To achieve parsimony in the number of coefficients likelihood ratio tests are implemented. The procedure to gain the final model is a specific-to general approach which starts by the basic OLS model continues with the basic multilevel model to the intercept-as-outcome model and finally ends in the intercept-and-slope-as-outcome model. The intraclass correlation coefficients as well as the information criteria help to decide whether the more general model is recommended. This organization will be used in the presentation of the result in the next section.

Because the variables are measured at different levels of scale and to reduce the effects on other variables and gain estimates of similar scale, all the variables are centered and scaled (also called studentized), for a discussion see Hill and Adkins (2003, pp. 263f.). By centering, the intercept indicates the mean outcome of the mean observation, and thus the mean industry in an average city on average over time.

6.5 Empirical Results

The empirical investigation starts by analyzing the explanatory power and significance of the explanatory variables in a pooled setup. This pooled setup is estimated by standard OLS. Because the variables are studentized the constant term, specified as the intercept, has to be insignificant in every specification.

Table 6.4 shows the results of the OLS estimation for the linear models for gross value added growth on the preceding of years of value added growth, each of the components of yearly productivity change, the other explanatory variables, as well as the quadratic terms of these variables and the interaction terms of these variables, and components of productivity change. The model is similar to the intercept-and-slope-as-outcome model, but without any random effects. It helps to reduce the number of parameter estimates, all insignificant variables are already deleted. The Gini coefficient is not significant for the gross value added and employment growth, but it is significant for gross value added and employment in absolute numbers. Therefore, the change in the Gini coefficient is tested for significance in first differences.

Table 6.4: OLS results for gross value added growth

	A	B	C	D	E	F	G
Intercept	-0.0408	-0.0448	-0.0416	-0.0397	-0.0335	-0.0338	-0.0448
	(0.029)	(0.0285)	(0.0289)	(0.0292)	(0.0284)	(0.0284)	(0.0286)
$dGVAL1$	-0.1052 ***	-0.1065 ***	-0.1099 ***	-0.1111 ***	-0.1063 ***	-0.1094 ***	-0.1062 ***
	(0.031)	(0.0321)	(0.0314)	(0.031)	(0.0331)	(0.0333)	(0.0323)
$malm$	-0.0276						
	(0.0251)						
$tech$		0.0103					
		(0.0198)					
eff			-0.0479 **				
			(0.0239)				
$pure.eff$				-0.0697 ***			
				(0.0252)			
$pure.tech$					0.0524 **		
					(0.023)		
$scale.tech$						-0.0344	
						(0.0217)	
$scale$							0.0426 *
							(0.0256)
$PC{\cdot}\ln Stu$	0.0193	-0.0465 **	0.0423 *	0.0058	0.0021	-0.0427 *	0.0348
	(0.0257)	(0.0211)	(0.0246)	(0.0239)	(0.0241)	(0.0251)	(0.0333)
$\ln Stu$	0.045 *	0.0453 **	0.0447 *	0.0464 **	0.0243	0.025	0.0465 **
	(0.0231)	(0.0231)	(0.0231)	(0.0231)	(0.0224)	(0.0223)	(0.023)
$\ln Stu^2$	0.0352	0.0394 *	0.0364	0.0369	0.0217	0.0249	0.0407 *
	(0.0225)	(0.0223)	(0.0225)	(0.0225)	(0.0218)	(0.0217)	(0.0223)
$PC{\cdot}dPop$	0.0107	-0.0166	0.0202	0.0489	-0.052 **	0.0728 ***	-0.05 **
	(0.0391)	(0.0221)	(0.0378)	(0.038)	(0.0228)	(0.0229)	(0.0241)
$dPop$	0.0204	0.0221	0.0185	0.0153	0.0299	0.0305	0.022
	(0.0223)	(0.0225)	(0.0224)	(0.0219)	(0.0228)	(0.0229)	(0.0221)
$dPop^2$	0.02	0.0209	0.0178	-0.0042	-0.0043	-0.0146	-0.0017
	(0.0619)	(0.0588)	(0.0652)	(0.0681)	(0.0625)	(0.059)	(0.0567)
$PC{\cdot}Gini$	0.0591 *	0.0419 **	0.0323	0.039	0.0487	-0.0006	-0.0057
	(0.0302)	(0.0207)	(0.0371)	(0.0478)	(0.0298)	(0.0217)	(0.025)
$Gini$	0.0323 **	0.0399 **	0.0343 **	0.0341 **	0.0373 **	0.0387 **	0.0386 **
	(0.0158)	(0.017)	(0.0165)	(0.0168)	(0.0172)	(0.0177)	(0.0176)
$Gini^2$	0.0155	0.0065	0.0104	0.0099	0.0081	0.0036	0.0042
	(0.0157)	(0.0164)	(0.0175)	(0.0183)	(0.0161)	(0.0191)	(0.0191)
$PC{\cdot}SC$	0.028	0.0156	0.0083	0.012	0.0148	0.0014	-0.0141
	(0.0239)	(0.0223)	(0.0233)	(0.0235)	(0.0229)	(0.0276)	(0.0333)
SC	-0.1144 ***	-0.1177 ***	-0.1204 ***	-0.1198 ***	-0.1221 ***	-0.1208 ***	-0.1195 ***
	(0.0316)	(0.032)	(0.0314)	(0.0312)	(0.032)	(0.0318)	(0.0318)
SC^2	0.1102 ***	0.1075 ***	0.1159 ***	0.1153 ***	0.1126 ***	0.1146 ***	0.112 ***
	(0.0324)	(0.0331)	(0.0372)	(0.0362)	(0.0332)	(0.0357)	(0.0355)
$PC{\cdot}BusTax$	0.02	-0.0062	0.0253	0.0027	0.0537 **	-0.0272	0.013
	(0.024)	(0.0196)	(0.0212)	(0.0232)	(0.0252)	(0.0182)	(0.0208)
$BusTax$	-0.0299 *	-0.0297 *	-0.031 *	-0.029 *	-0.0188	-0.0234	-0.0309 *
	(0.016)	(0.016)	(0.016)	(0.0159)	(0.0157)	(0.0157)	(0.0159)
$BusTax^2$	0.0337 **	0.034 **	0.0354 **	0.0335 **	0.0314 *	0.0298 *	0.0323 *
	(0.0167)	(0.0166)	(0.0165)	(0.0166)	(0.0166)	(0.0166)	(0.0166)
$\ln L$	-5518.886	-5517.6	-5517.596	-5514.262	-5364.374	-5371.791	-5520.373
R^2	0.2765	0.2766	0.2766	0.2771	0.2967	0.2958	0.2763

Significance codes: '***', '**', '*' significant up to 1%, 5%, and 10%, respectively. Heteroscedasticity consistent standard errors in parentheses. R^2 is McFadden-R^2 for comparison reason.

The results in table 6.4 show many interesting features. Each column contains the estimates for one OLS estimation, with the heteroscedasticity consistent standard error below the estimates in parentheses. Each OLS estimation contains a different measurement of productivity change (PC). The first column A shows the results for the Malmquist index ($malm$). The intercept is not significant, with a small negative estimate of -0.0408, indicating no gross value added growth on average for an average city because the variables are standardized. The intercept is the average of the endogenous variable if every exogenous variable is zero, which stands for the average city. The next line in column A shows the results for gross value added growth lagged by one period ($dGVAL1$), which has a significant negative estimate of -0.1052. Therefore, past gross value added growth, which is also standardized, leads to a catching-up of growth rates. A growth rate below the average, measured by a negative standardized growth rate, will result in a growth rate above average or a positive standardized growth rate in the next period.

The estimate for the Malmquist index is not significantly different from zero by -0.0276, indicating that the Malmquist index in total does not affect the gross value added growth. The next three rows contain the estimates for the logarithm of the number of students. Whereas the first of the three rows includes the interaction term of productivity change, which is the Malmquist index in the first column, with the logarithm of the number of students. The following row contains the linear term of the logarithm of the number of students. The last of the three rows contains the quadratic term of the logarithm of the number of students within the city. Only the linear term is significantly positive, with an estimate of 0.045, indicating that an increase in the number of students is correlated with higher gross value added growth.

The next three rows show the estimates for population change where, again, the first of these three rows gives the estimate for the interaction term of the component with population growth, the second row gives the estimate for the change in population, and the third row contains the estimate for the quadratic term of population change. All the estimates including population change are not significantly different from zero in column A.

The next three rows comprise the estimates with the change in the Gini coefficient and, as for all other variables, with the interaction term, the linear term and the quadratic term in the first, second and third row, respectively. In the case of the Malmquist index as productivity change measure in column A, the interaction term and the linear term of the change of the Gini coefficient are significantly positive, with estimates of 0.0591 and 0.0323 for the interaction term and the linear term, respectively. Thus, for an average city with all standardized variables equal to zero, gross value added growth increases by a further increase in the Gini coefficient. Therefore, gross value added growth is correlated with a stronger industrial specialization. This effect is further increased if the city has a Malmquist index which is above average.

The next three rows show the estimates for the corresponding terms of structural change variable. The interaction term of the Malmquist index and the structural change is not significantly different from zero, the linear term is significantly negative with an estimate of -0.1144, and the quadratic term is significantly positive with an estimate of 0.1102. Therefore, structural change affects gross value added growth with a U-form and a minimum point of about 0.52; for an average city the effect of the standardized structural change on gross value added growth is only positive for negative values and values above 1.04 (or values of structural change below average or with 1.04 times the standard deviation greater than the average structural change, while for structural change slightly above average the effect in gross value added growth is negative).

The last three rows in the first column contain the estimates for business tax. Similar to structural change, the linear term of business tax is significantly negative and the quadratic term is significantly positive, with values of -0.0299 and 0.0337, respectively. The U-form effect of business tax on gross value added growth is minimal at about 0.89 and is positive for standardized values of business tax below zero and above 1.78. This means that values of business tax below average and larger than 1.78 times the standard deviation above average are correlated with positive gross value added growth.

The value of the McFadden-R^2 is remarkably large for an industry pooled cross-city growth analysis, with a value of about 27 percent. The next columns contain the estimates for the components of the Malmquist index, namely the technological change in column B, efficiency change in column C, pure efficiency change in column D, pure technological change in column E, scale technological change in column F, and scale efficiency change in column G. Because of the different components and the interaction terms of these with the other city-specific variables, the estimates are likely to change except for the intercept, because all variables are standardized. So, the intercept always estimates the gross value added growth of an average city with all standardized variables being zero.

Without going into too much detail, the results are explained in general without the exact estimates which can be found in table 6.4. First of all, the intercept is insignificant, as expected. Secondly, past gross value added growth is negatively significant. Value added growth above average is associated with value added growth below average and vice versa, which supports the catching-up hypothesis. Thirdly, neither the Malmquist index nor the change in technology has a significant effect on value added growth, though efficiency change and its component pure change in efficiency have a negative effect on value added growth. These components measure the catching-up to production frontier by process innovations. However, the catching-up results in lower value added growth. Furthermore, pure technological change as well as change in scale efficiency have a positive effect on value added growth. Thus, a shift in the production frontier, as measured by pure technological change, results in higher value added growth. In addition, some interaction terms are significant, depending on the component. Fourthly, the structural change index and business tax have a maximum effect on value added growth because the linear term is significantly positive and the quadratic term is significantly negative. Furthermore, the number of students and the change in concentration have significant positive effects on gross value added growth.

The corresponding results of table 6.4 are reported in table 6.5 for employment growth.

Table 6.5: OLS results for employment growth

	H	I	J	K	L	M	N
Intercept	-0.0113	-0.0125	-0.0102	-0.0105	-0.0102	-0.0107	-0.0128
	(0.0251)	(0.0251)	(0.0251)	(0.0252)	(0.025)	(0.025)	(0.0251)
$dEmpL1$	0.4498 ***	0.4463 ***	0.447 ***	0.4487 ***	0.46 ***	0.4613 ***	0.4511 ***
	(0.0289)	(0.0288)	(0.0288)	(0.029)	(0.0296)	(0.0298)	(0.0291)
$malm$	-0.0299 *						
	(0.0181)						
$tech$		0.0032					
		(0.0144)					
eff			-0.0391 **				
			(0.0181)				
$pure.eff$				-0.065 ***			
				(0.0162)			
$pure.tech$					0.0228		
					(0.0164)		
$scale.tech$						-0.0108	
						(0.0217)	
$scale$							0.0306
							(0.0244)
$PC{\cdot}\ln Stu$	0	0.0132	-0.002	-0.0247 *	0.0302 **	-0.0234	0.0143
	(0.0188)	(0.0157)	(0.0175)	(0.0148)	(0.0141)	(0.0294)	(0.0266)
$\ln Stu$	-0.0092	-0.0083	-0.0102	-0.0093	-0.0135	-0.0147	-0.0088
	(0.022)	(0.0219)	(0.0219)	(0.0219)	(0.0217)	(0.0218)	(0.022)
$\ln Stu^2$	0.0106	0.0111	0.0104	0.0114	0.0081	0.0091	0.0124
	(0.0192)	(0.0192)	(0.0193)	(0.0192)	(0.0191)	(0.0193)	(0.0193)
$PC{\cdot}dPop$	-0.0094	-0.0186	0.0044	0.0175	-0.0146	0.0093	-0.019
	(0.0241)	(0.0174)	(0.0215)	(0.0175)	(0.0169)	(0.0184)	(0.0289)
$dPop$	0.0483 **	0.0479 **	0.048 **	0.0458 **	0.0516 ***	0.0505 ***	0.0475 **
	(0.0201)	(0.0202)	(0.0198)	(0.0204)	(0.0189)	(0.0189)	(0.0207)
$dPop^2$	0.0699 **	0.0599 *	0.064 **	0.0606 *	0.0553 *	0.0584 **	0.0527 *
	(0.0329)	(0.0319)	(0.0321)	(0.0321)	(0.0301)	(0.0274)	(0.0294)
$PC{\cdot}Gini$	-0.0107	0.023	-0.0234	-0.0035	0.0009	0.0192	-0.0119
	(0.0225)	(0.0154)	(0.0212)	(0.0197)	(0.0166)	(0.0192)	(0.0309)
$Gini$	-0.0004	0	0.0009	-0.0016	0.0025	0.0047	-0.0015
	(0.0182)	(0.0182)	(0.0182)	(0.0181)	(0.0178)	(0.0179)	(0.0181)
$Gini^2$	-0.004	-0.0013	-0.0015	-0.0045	-0.0043	-0.0043	-0.0046
	(0.0322)	(0.0326)	(0.0315)	(0.0321)	(0.0335)	(0.0335)	(0.0323)
$PC{\cdot}SC$	0.0065	0.0028	-0.0029	-0.0019	0.0113	-0.0031	-0.0097
	(0.0094)	(0.0119)	(0.0123)	(0.0101)	(0.0123)	(0.0203)	(0.0167)
SC	-0.0937 ***	-0.0966 ***	-0.0989 ***	-0.0986 ***	-0.0913 ***	-0.088 ***	-0.0956 ***
	(0.0275)	(0.0276)	(0.0274)	(0.0275)	(0.0264)	(0.0268)	(0.0276)
SC^2	0.077 ***	0.0796 ***	0.0801 ***	0.0806 ***	0.073 ***	0.0719 ***	0.0771 ***
	(0.0263)	(0.027)	(0.026)	(0.0258)	(0.0252)	(0.026)	(0.0267)
$PC{\cdot}BusTax$	-0.0363 **	-0.0298 **	-0.0016	-0.0121	0.0041	-0.0269	-0.0072
	(0.0154)	(0.0122)	(0.0138)	(0.0142)	(0.0137)	(0.0171)	(0.0146)
$BusTax$	-0.0301 **	-0.0318 **	-0.0314 **	-0.0312 **	-0.0275 **	-0.0295 **	-0.0308 **
	(0.0133)	(0.0133)	(0.0133)	(0.0133)	(0.0133)	(0.0135)	(0.0133)
$BusTax^2$	0.0127	0.014	0.0127	0.0119	0.0123	0.0146	0.013
	(0.0133)	(0.0131)	(0.0132)	(0.0132)	(0.0132)	(0.0131)	(0.0132)
$\ln L$	-5043.595	-5043.066	-5040.982	-5035.681	-4854.772	-4853.936	-5046.078
R^2	0.3388	0.3388	0.3391	0.3398	0.3635	0.3636	0.3385

Significance codes: '***', '**', '*' significant up to 1%, 5%, and 10%, respectively. Heteroscedasticity consistent standard errors in parentheses. R^2 is McFadden-R^2 for comparison reason.

The results for employment growth are somewhat different from those for gross value added growth, not only by higher coefficients of determination but also by different significant explanatory variables. Columns H to M in table 6.5 have exactly the same structure as columns A to G in table 6.4 and are, therefore, interpreted the same way simply for employment growth instead of gross value added growth. First of all, past employment growth which is above average results in positive employment growth in the next period, which indicates a divergence of employment growth for industries in German cities. Secondly, productivity change measured by the Malmquist index is significantly negative for employment growth one year later. By decomposing productivity change into its components, as illustrated in subsection 2.1.2, it becomes obvious that this effect is driven only by changes in efficiency, as indicated by columns J and K. These effects are the same as for value added growth. Thus, the catching-up process seems to have a negative overall effect.

Interaction terms are only significant for business tax and the number of students. In addition, structural change has an inverted U-shaped effect on employment growth with a minimum point because the linear term is significantly positive and the quadratic term is significantly negative. The point at which structural change minimally affects employment growth is at 0.61. Therefore, standardized structural change has a positive effect on employment growth for values below zero (below average for non-standardized structural change) and for values above 1.22 standardized structural change (or 1.22 times the standard deviation of structural change above the average). Business tax is significantly negative. Thus, business tax rates below average foster employment growth, whereas business tax rates above average reduce employment growth in contrast to the effect on value added growth. Change in population has a significantly positive linear and quadratic terms, so larger growth of inhabitants within a city has even further positive effects on employment growth.

The results indicate the importance of productivity change as well as the other observed explanatory variables and the significance of some non-linear effects. Because industries are nested within cities and these are observed for many consecutive years, the results of the OLS estimation should be treated with care because the observations are correlated from the common factor within one level and some variables are only observed at lower levels. Thus, the variance was not considered correctly, which I tried to account for with the heteroscedasticity-corrected standard errors; furthermore, the estimates can be biased in the case of different slopes for each object within the levels. Therefore, multilevel analyses have to be used to gain unbiased estimates.

To test whether the levels should be considered, the residual plots of the OLS estimation can be visually analyzed at each level, as suggested by Zuur et al. (2009). The exemplary residual box plot for the city-level of the OLS estimation for gross value added growth is shown in figure 6.1 and the corresponding box plot for the time-level is shown in figure 6.2. (The equivalent residual box plots for employment growth are included in Appendix D.)

As seen in the residual box plots, they change over both levels, namely the city and time. Even the variation over the years is not large but it is nonetheless present, and the variance declines over the time, indicated by narrower boxes that illustrating the interquartile range for later years. The variation should result in a relatively large intraclass correlation coefficient for the city-level and a relatively low intraclass correlation coefficient for the time-level because of smaller changes in the residual variation in the time-level.

To compare the following results of the basic multilevel model and because the interaction terms of the productivity change measurements are often not significant and for the sake of completeness, the results of the OLS model without interaction terms are reported within the next two tables.

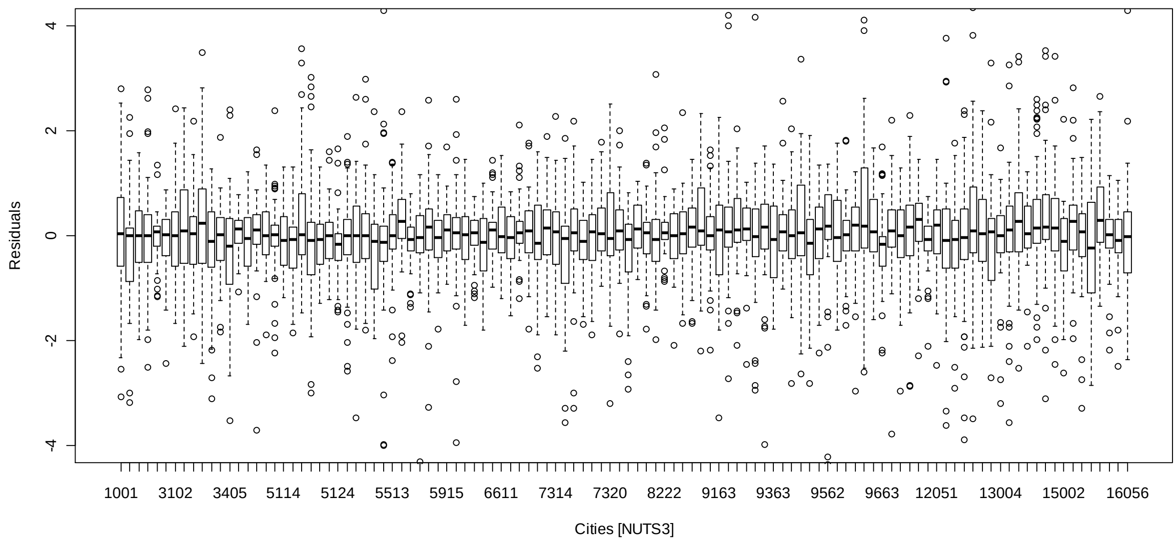

Figure 6.1: Residual Plot for Gross Value Added Growth at City Level

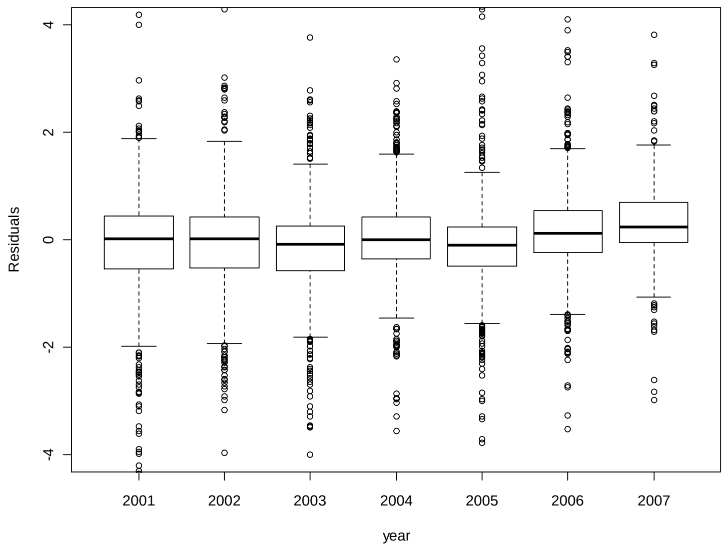

Figure 6.2: Residual Plot for Gross Value Added Growth at Time Level

Table 6.6: OLS results for gross value added growth without interaction terms

	A	B	C	D	E	F	G
Intercept	-0.044	-0.043	-0.043	-0.045	-0.034	-0.035	-0.045
	(0.029)	(0.029)	(0.028)	(0.029)	(0.028)	(0.028)	(0.029)
$dGVAL1$	-0.111 ***	-0.106 ***	-0.113 ***	-0.118 ***	-0.107 ***	-0.107 ***	-0.105 ***
	(0.031)	(0.032)	(0.031)	(0.031)	(0.033)	(0.033)	(0.032)
$malm$	-0.024						
	(0.027)						
$tech$		0.023					
		(0.018)					
eff			-0.04 *				
			(0.024)				
$pure.eff$				-0.067 **			
				(0.027)			
$pure.tech$					0.054 **		
					(0.021)		
$scale.tech$						-0.039 **	
						(0.018)	
$scale$							0.042 **
							(0.021)
$\ln Stu$	0.047 **	0.046 **	0.047 **	0.047 **	0.024	0.025	0.047 **
	(0.023)	(0.023)	(0.023)	(0.023)	(0.022)	(0.022)	(0.023)
$\ln Stu^2$	0.038 *	0.038 *	0.038 *	0.038 *	0.022	0.022	0.038 *
	(0.022)	(0.022)	(0.022)	(0.022)	(0.022)	(0.022)	(0.022)
$dPop$	0.021	0.022	0.021	0.021	0.032	0.03	0.022
	(0.022)	(0.022)	(0.022)	(0.022)	(0.023)	(0.023)	(0.022)
$dPop^2$	0.026	0.025	0.028	0.033	0.019	0.019	0.027
	(0.056)	(0.057)	(0.056)	(0.056)	(0.057)	(0.057)	(0.057)
$Gini$	0.038 **	0.038 **	0.038 **	0.037 **	0.038 **	0.038 **	0.038 **
	(0.017)	(0.017)	(0.017)	(0.017)	(0.018)	(0.018)	(0.017)
$Gini^2$	0.006	0.005	0.006	0.007	0.005	0.005	0.006
	(0.018)	(0.019)	(0.018)	(0.018)	(0.02)	(0.019)	(0.019)
SC	-0.12 ***	-0.12 ***	-0.121 ***	-0.122 ***	-0.124 ***	-0.119 ***	-0.117 ***
	(0.031)	(0.031)	(0.031)	(0.031)	(0.032)	(0.032)	(0.031)
SC^2	0.114 ***	0.115 ***	0.115 ***	0.116 ***	0.117 ***	0.113 ***	0.112 ***
	(0.034)	(0.034)	(0.034)	(0.034)	(0.034)	(0.034)	(0.034)
$BusTax$	-0.03 *	-0.03 *	-0.03 *	-0.03 *	-0.02	-0.02	-0.029 *
	(0.016)	(0.016)	(0.016)	(0.016)	(0.016)	(0.016)	(0.016)
$BusTax^2$	0.034 **	0.033 **	0.034 **	0.034 **	0.029 *	0.029 *	0.033 ***
	(0.017)	(0.017)	(0.017)	(0.017)	(0.017)	(0.017)	(0.017)
$\ln L$	-5530.445	-5530.476	-5527.917	-5521.765	-5379.731	-5383.068	-5530.476
R^2	0.275	0.275	0.275	0.276	0.295	0.294	0.275

Significance codes: '***', '**', '*' significant up to 1%, 5%, and 10%, respectively. Heteroscedasticity consistent standard errors in parentheses. R^2 is McFadden-R^2 for comparison reason.

Table 6.7: OLS results for employment growth without interaction terms

	H	I	J	K	L	M	N
Intercept	-0.012	-0.012	-0.011	-0.012	-0.01	-0.01	-0.012
	(0.025)	(0.025)	(0.025)	(0.025)	(0.025)	(0.025)	(0.025)
*dEmpL*1	0.45 ***	0.449 ***	0.447 ***	0.448 ***	0.462 ***	0.463 ***	0.452 ***
	(0.029)	(0.029)	(0.029)	(0.029)	(0.03)	(0.03)	(0.029)
malm	-0.027 *						
	(0.015)						
tech		0.013					
		(0.013)					
eff			-0.049 ***				
			(0.015)				
pure.eff				-0.064 ***			
				(0.015)			
pure.tech					0.021		
					(0.013)		
scale.tech						-0.011	
						(0.016)	
scale							0.024
							(0.018)
ln*Stu*	-0.009	-0.01	-0.01	-0.01	-0.014	-0.014	-0.009
	(0.022)	(0.022)	(0.022)	(0.022)	(0.022)	(0.022)	(0.022)
lnStu^2	0.011	0.011	0.011	0.011	0.007	0.007	0.011
	(0.019)	(0.019)	(0.019)	(0.019)	(0.019)	(0.019)	(0.019)
dPop	0.048 **	0.049 **	0.048 **	0.048 **	0.052 ***	0.051 ***	0.048 **
	(0.02)	(0.02)	(0.02)	(0.02)	(0.019)	(0.019)	(0.02)
$dPop^2$	0.064 **	0.062 **	0.065 **	0.07 **	0.06 **	0.06 **	0.064 **
	(0.029)	(0.029)	(0.029)	(0.029)	(0.027)	(0.027)	(0.029)
Gini	0	0	0.001	-0.002	0.003	0.003	-0.003
	(0.018)	(0.018)	(0.018)	(0.018)	(0.018)	(0.018)	(0.018)
$Gini^2$	-0.004	-0.004	-0.004	-0.005	-0.004	-0.004	-0.005
	(0.032)	(0.032)	(0.032)	(0.032)	(0.033)	(0.034)	(0.032)
SC	-0.098 ***	-0.097 ***	-0.099 ***	-0.099 ***	-0.093 ***	-0.091 ***	-0.095 ***
	(0.027)	(0.027)	(0.027)	(0.027)	(0.026)	(0.026)	(0.027)
SC^2	0.08 ***	0.08 ***	0.081 ***	0.081 ***	0.076 ***	0.075 ***	0.079 ***
	(0.026)	(0.026)	(0.026)	(0.026)	(0.025)	(0.025)	(0.026)
BusTax	-0.031 **	-0.031 **	-0.032 **	-0.031 **	-0.027 **	-0.027 **	-0.031 **
	(0.013)	(0.013)	(0.013)	(0.013)	(0.013)	(0.013)	(0.013)
$BusTax^2$	0.013	0.013	0.014	0.014	0.013	0.013	0.013
	(0.013)	(0.013)	(0.013)	(0.013)	(0.013)	(0.013)	(0.013)
ln*L*	-5047.817	-5049.447	-5042.517	-5038.199	-4857.7	-4858.672	-5048.101
R^2	0.338	0.338	0.339	0.339	0.363	0.363	0.338

Significance codes: '***', '**', '*' significant up to 1%, 5%, and 10%, respectively. Heteroscedasticity consistent standard errors in parentheses. R^2 is McFadden-R^2 for comparison reason.

The results presented in the tables 6.6 and 6.7 are comparable to the results of the basic and intercept-as-outcome multilevel models in the next subsections, because the interaction terms with productivity change measures are absent.

6.5.1 Results for the Basic Multilevel Model

To detect different variations within the level and to justify the need for the incorporation of each level, the intraclass correlation coefficients have to be calculated. This is done by estimating the basic multilevel models without considering all the intercepts and slopes as being heterogeneous and all possible random effects in the different levels similar to the stepwise procedure in Goedhuys and Srholec (2010). The basic model only includes the intercepts as random. The results for the basic multilevel estimation are presented in the tables 6.8 and 6.9.

Table 6.8: Multilevel results for gross value added growth in the basic multilevel model

	A	B	C	D	E	F	G
Fixed effects							
Intercept	-0.0359	-0.0357	-0.0358	-0.0372	-0.0259	-0.0265	-0.0377
	(0.0843)	(0.0836)	(0.083)	(0.0828)	(0.0828)	(0.0852)	(0.085)
dGVAL1	-0.1134 ***	-0.108 ***	-0.1126 ***	-0.118 ***	-0.1086 ***	-0.1081 ***	-0.1069 ***
	(0.0152)	(0.0149)	(0.0151)	(0.0151)	(0.015)	(0.015)	(0.0149)
malm	-0.0275 *						
	(0.0151)						
tech		0.0054					
		(0.0139)					
eff			-0.0262 *				
			(0.0145)				
pure.eff				-0.0587 ***			
				(0.015)			
pure.tech					0.0407 ***		
					(0.0143)		
scale.tech						-0.0462 ***	
						(0.0141)	
scale							0.0498 ***
							(0.014)
$\ln Stu$	0.0386	0.0383	0.0387	0.039	0.0154	0.0152	0.038
	(0.0281)	(0.0282)	(0.0282)	(0.0281)	(0.0284)	(0.0284)	(0.0281)
$\ln Stu^2$	0.0305	0.0308	0.031	0.0311	0.0148	0.0139	0.0303
	(0.0271)	(0.0272)	(0.0272)	(0.0271)	(0.0272)	(0.0272)	(0.0271)
dPop	0.0255	0.0262	0.0257	0.0253	0.038	0.0378	0.027
	(0.0231)	(0.0231)	(0.0231)	(0.0231)	(0.0235)	(0.0235)	(0.0231)
$dPop^2$	0.0296	0.0276	0.029	0.0337	0.0172	0.019	0.0307
	(0.0425)	(0.0426)	(0.0425)	(0.0424)	(0.0425)	(0.0425)	(0.0425)
Gini	0.0454 **	0.0457 **	0.0454 **	0.0442 **	0.0448 **	0.0451 **	0.045 **
	(0.0185)	(0.0185)	(0.0185)	(0.0185)	(0.0185)	(0.0185)	(0.0185)
$Gini^2$	0.018	0.0168	0.0175	0.0186	0.0172	0.0171	0.0174
	(0.0174)	(0.0174)	(0.0174)	(0.0174)	(0.0174)	(0.0174)	(0.0174)
SC	-0.0929 **	-0.0922 **	-0.0926 **	-0.0932 **	-0.0946 **	-0.0951 **	-0.0917 **
	(0.0427)	(0.0428)	(0.0428)	(0.0427)	(0.0426)	(0.0426)	(0.0427)
SC^2	0.0912 **	0.0908 **	0.091 **	0.0914 **	0.0925 **	0.093 **	0.0904 **
	(0.0421)	(0.0422)	(0.0421)	(0.042)	(0.042)	(0.0419)	(0.0421)
BusTax	-0.033 *	-0.0331 *	-0.0332 *	-0.0333 *	-0.023	-0.0221	-0.0324 *
	(0.0183)	(0.0184)	(0.0184)	(0.0183)	(0.0184)	(0.0184)	(0.0183)
$BusTax^2$	0.0301	0.0297	0.0301	0.0306	0.0247	0.0246	0.0299
	(0.019)	(0.0191)	(0.019)	(0.019)	(0.019)	(0.019)	(0.019)
Random effects							
time- level							
Intercept	0.1898	0.1877	0.1863	0.1857	0.1856	0.1922	0.1916
city- level							
Intercept	0.2794	0.2809	0.28	0.279	0.2789	0.2786	0.2799
residuals	0.8978	0.8979	0.8977	0.8965	0.8892	0.8889	0.8966
AIC	10992	10995	10992	10980	10698	10696	10983
BIC	11093	11096	11093	11081	10799	10796	11083
$\ln L$	-5479.9	-5481.5	-5480	-5473.9	-5333.2	-5331.9	-5475.3
ICCtime	0.0392	0.0383	0.0378	0.0377	0.0382	0.0408	0.0399
ICCcity	0.124	0.124	0.1231	0.1226	0.1243	0.1266	0.1252
R^2	0.2719	0.2716	0.2719	0.2727	0.2914	0.2915	0.2725

Significance codes: '***', '**', '*' significant up to 1%, 5%, and 10%, respectively. Standard errors for fixed effects are in parentheses below the estimates, for random effects standard errors are reported. R^2 is McFadden-R^2 for comparison reason.

Table 6.9: Multilevel results for employment growth in the basic multilevel model

	H	I	J	K	L	M	N
Fixed effects							
Intercept	-0.0058	-0.006	-0.0057	-0.0066	-0.0042	-0.0043	-0.0068
	(0.1146)	(0.1135)	(0.1128)	(0.1129)	(0.1105)	(0.1107)	(0.1138)
*dEmpL*1	0.4433 ***	0.443 ***	0.4417 ***	0.4427 ***	0.4555 ***	0.4558 ***	0.4449 ***
	(0.0141)	(0.0142)	(0.0142)	(0.0141)	(0.0142)	(0.0142)	(0.0142)
malm	-0.041 ***						
	(0.0127)						
tech		0.0004					
		(0.0125)					
eff			-0.0422 ***				
			(0.0125)				
pure.eff				-0.0592 ***			
				(0.0128)			
pure.tech					0.0098		
					(0.0125)		
scale.tech						-0.0107	
						(0.0123)	
scale							0.0294 **
							(0.0123)
$\ln Stu$	-0.029	-0.0293	-0.0287	-0.0284	-0.0328	-0.0329	-0.029
	(0.0206)	(0.0206)	(0.0206)	(0.0205)	(0.0204)	(0.0204)	(0.0205)
$\ln Stu^2$	-0.0012	-0.0009	-0.0006	-0.0005	-0.0038	-0.004	-0.001
	(0.0197)	(0.0197)	(0.0197)	(0.0196)	(0.0194)	(0.0194)	(0.0196)
dPop	0.0597 ***	0.0609 ***	0.06 ***	0.0593 ***	0.0639 ***	0.0638 ***	0.0608 ***
	(0.017)	(0.017)	(0.017)	(0.017)	(0.017)	(0.017)	(0.0169)
$dPop^2$	0.0523 *	0.0493	0.0514 *	0.0561 *	0.0464	0.0469	0.0518 *
	(0.0311)	(0.031)	(0.0311)	(0.031)	(0.0305)	(0.0305)	(0.0309)
Gini	0.0527 ***	0.052 ***	0.0522 ***	0.0496 ***	0.0526 ***	0.0526 ***	0.0497 ***
	(0.0169)	(0.0169)	(0.0169)	(0.0169)	(0.0168)	(0.0168)	(0.0169)
$Gini^2$	0.0155	0.0154	0.0153	0.014	0.0141	0.0141	0.0142
	(0.0127)	(0.0127)	(0.0127)	(0.0127)	(0.0125)	(0.0125)	(0.0126)
SC	-0.0013	-0.0006	-0.0009	-0.001	0.0036	0.0035	-0.0001
	(0.0313)	(0.0313)	(0.0313)	(0.0312)	(0.0306)	(0.0306)	(0.0312)
SC^2	-0.0088	-0.0091	-0.009	-0.009	-0.0131	-0.0129	-0.0095
	(0.0308)	(0.0308)	(0.0308)	(0.0307)	(0.0301)	(0.0301)	(0.0307)
BusTax	-0.0352 ***	-0.0353 ***	-0.0355 ***	-0.0354 ***	-0.0313 **	-0.0311 **	-0.0349 ***
	(0.0133)	(0.0133)	(0.0133)	(0.0132)	(0.013)	(0.013)	(0.0132)
$BusTax^2$	0.0065	0.0062	0.0066	0.0067	0.0054	0.0054	0.0063
	(0.0137)	(0.0137)	(0.0137)	(0.0136)	(0.0134)	(0.0134)	(0.0136)
Random effects							
time- level							
Intercept	0.2744	0.2718	0.27	0.2702	0.2644	0.2649	0.2725
city- level							
Intercept	0.0443	0.0388	0.0444	0.0363	0.0001	0.0001	0.0227
residuals	0.811	0.8123	0.8109	0.8103	0.7954	0.7953	0.8124
AIC	9890	9900.4	9889	9878.9	9525.9	9525.8	9894.7
BIC	9990.7	10001.1	9989.7	9979.7	9626.4	9626.3	9995.5
$\ln L$	-4929	-4934.2	-4928.5	-4923.5	-4747	-4746.9	-4931.3
ICCtime	0.1024	0.1004	0.0995	0.0999	0.0995	0.0999	0.1011
ICCcity	0.1051	0.1025	0.1022	0.1017	0.0995	0.0999	0.1018
R^2	0.3378	0.3371	0.3378	0.3385	0.3622	0.3622	0.3374

Significance codes: '***', '**', '*' significant up to 1%, 5%, and 10%, respectively. Standard errors for fixed effects are in parentheses below the estimates, for random effects standard errors are reported. R^2 is McFadden-R^2 for comparison reason.

The basic multilevel model estimations indicate different results for the necessity of the levels. On the one hand, both tables for the basic multilevel model show large intraclass correlation coefficients, except for the time-level for gross value added growth. Therefore, the time-level may not be kept for the estimations. The low intraclass correlation coefficient was expected to be relatively low by the residual box plots, although there is little variation and a decrease in the residual variance. However, the intraclass correlations are calculated only on the basis of the random effects of the intercepts in the basic multilevel model, even though there might be some slope variations not considered in the basic multilevel model. These results of the basic multilevel model are comparable to those results of the OLS model in the tables 6.6 and 6.7, except for the presence of random effects in the multilevel model. Therefore, these coefficients differ from those of the OLS model, not only in altitude, but also, consequently, in significance. The structure of both tables is the same as in the OLS estimation tables.

The results for value added growth in table 6.8 show similar significant results to those in table 6.7, with some minor differences resulting from the absence of the random effects in the OLS model. The coefficient of the Malmquist index in column A as well as its components of efficiency change in column C fueled by change of pure efficiency change in column D and scale technological change in column F are significantly negative, with values of -0.0275, -0.0262, -0.0587 and -0.0462, respectively. The significant positive coefficient of change in pure technology (*pure.tech*) in column E is important to notice. It indicates that technological progress has a positive effect on value added growth although it is offset by the negative scale technological change.

The coefficients of the city-specific variables are the same in every estimation, because there is no changing interaction involved in any estimation. The change in the Gini coefficient is significantly positive, with an estimate of around 0.045. Structural change has a U-form effect on value added growth with a minimum of about 0.5, which indicates that a moderate change above average has the lowest effect on value added growth. The quadratic term of business tax is not as significant as in the OLS estimation. Only the linear term of business tax is significantly negative, which shows that an increase in business tax in the city has a negative effect on value added growth in that city in the next year. Furthermore and in contrast to the OLS results in table 6.6 students are not significant in the multilevel model.

A similar pattern occurs for the estimations of employment growth in table 6.9. In contrast to the OLS results, in table 6.5 the change in scale efficiency is significantly positive in column N. Productivity change measured by the Malmquist index affects employment growth significantly negative. That effect is caused by the negative effect of the change in efficiency which is mainly fueled by the change of pure efficiency with estimates of -0.0422 and -0.0592 in the columns J and K, respectively. The change in population has a U-form effect on employment growth with a negative minimum value at about -0.57, which indicates that a moderate change in population below the average has the smallest effect on employment growth. Additionally, the change in the Gini coefficient has a significantly positive effect on employment growth while business tax affects employment growth significantly negative, because only the linear terms are significantly different from zero for both variables. Therefore, an increase in the change of the Gini coefficient increases employment growth, whereas an increase in business tax within the city leads to a decrease in employment growth for the next year. Compared with the OLS results, structural change is not significant in the basic multilevel model.

As in the following multilevel model estimations, I am not interpreting the random effects because of sparse observations which do not affect the accuracy of fixed parameter estimates (Hox (1998, p. 150) and Moerbeek et al. (2000)). However, the power of the tests in multilevel models depends on the number of levels, which is only two in Hox (1998), the design of the model, number of groups within each level and the intraclass correlation as shown in Maas and Hox (2005). According to Roy et al. (2007) the sufficient sample size for longitudinal multilevel model without an attrition rate for 7 years and intraclass correlation of 10 percent is 5, which is met by my data set. In the basic multilevel model the random effects are used to estimate the intraclass correlation coefficients which are about 10 percent indicating correlation within each level and, therefore, the necessity of accounting for different levels.

6.5.2 Results for the Intercept-as-outcome Model

The intercept-as-outcome model, additionally, has the intercepts of the basic model as random coefficients. The results for the intercept-as-outcome for gross value added and employment growth are presented in the tables 6.10 and 6.11, respectively.

Table 6.10: Multilevel results for gross value added growth in the intercept-as outcome-model

	A	B	C	D	E	F	G
Fixed effects							
Intercept	-0.0285	-0.0316	-0.0463	-0.0504	-0.0338	-0.0309	-0.0249
	(0.0729)	(0.086)	(0.0759)	(0.0738)	(0.0775)	(0.0854)	(0.0855)
$dGVAL1$	-0.0711 ***	-0.0903 ***	-0.0613 ***	-0.0718 ***	-0.0862 ***	-0.1103 ***	-0.1013 ***
	(0.0151)	(0.0148)	(0.015)	(0.0152)	(0.0147)	(0.015)	(0.0148)
$malm$	-0.0368						
	(0.0314)						
$tech$		-0.0099					
		(0.0407)					
eff			-0.0261				
			(0.0288)				
$pure.eff$				-0.0766 ***			
				(0.0253)			
$pure.tech$					0.1025 **		
					(0.0468)		
$scale.tech$						-0.1121 **	
						(0.045)	
$scale$							0.1117 **
							(0.0481)
$\ln Stu$	0.0182	0.0269	0.0278	0.0399	0.0109	0.0124	0.0192
	(0.0253)	(0.0268)	(0.0262)	(0.0264)	(0.0269)	(0.0288)	(0.0286)
$\ln Stu^2$	0.0166	0.0155	0.0231	0.0301	0.0059	0.0124	0.02
	(0.0235)	(0.0255)	(0.0246)	(0.0247)	(0.0249)	(0.0263)	(0.0257)
$dPop$	0.0163	0.0358	0.0318	0.0351	0.0532 *	0.0404	0.0337
	(0.0291)	(0.0274)	(0.0289)	(0.0307)	(0.0274)	(0.0273)	(0.0278)
$dPop^2$	0.0669	-0.0029	0.0398	0.0392	-0.0047	0.0058	-0.0223
	(0.0417)	(0.0401)	(0.041)	(0.0455)	(0.0421)	(0.0436)	(0.0464)
$Gini$	0.0222	0.0304	0.0174	0.0229	0.0229	0.0276	0.0274
	(0.0207)	(0.029)	(0.0237)	(0.0244)	(0.031)	(0.0286)	(0.0271)
$Gini^2$	0.03 *	0.0257	0.0362 **	0.0234	0.0346 **	0.0299 *	0.0282
	(0.0157)	(0.0166)	(0.0166)	(0.0164)	(0.0169)	(0.0179)	(0.0173)
SC	-0.0925 **	-0.0832 **	-0.0905 **	-0.0892 *	-0.0966 **	-0.0979 **	-0.1014 **
	(0.041)	(0.0412)	(0.0457)	(0.0456)	(0.0444)	(0.0444)	(0.0426)
SC^2	0.0949 **	0.0836 **	0.1166 ***	0.1253 ***	0.1318 ***	0.1104 **	0.1039 **
	(0.0413)	(0.0406)	(0.0451)	(0.045)	(0.0427)	(0.0439)	(0.0418)
$BusTax$	-0.0341 *	-0.0334 *	-0.0327 *	-0.0391 *	-0.0283	-0.0293	-0.0363 **
	(0.0177)	(0.019)	(0.0185)	(0.0199)	(0.0211)	(0.0182)	(0.0178)
$BusTax^2$	0.0243	0.0197	0.0292 *	0.0302 *	0.0274	0.0329 *	0.0278
	(0.0164)	(0.0178)	(0.0171)	(0.0173)	(0.0176)	(0.0185)	(0.018)
Random effects							
time- level							
Intercept	0.1638	0.1961	0.1705	0.1645	0.1744	0.1936	0.1946
PC	0.0324	0.0745	0.034	0.0034	0.0876	0.0919	0.0977
$\ln Stu$	0.0159	0.0115	0.015	0.0169	0.0158	0.0211	0.0256
$dPop$	0.051	0.041	0.0481	0.0537	0.0402	0.0362	0.0407
$Gini$	0.031	0.0566	0.0401	0.0427	0.0627	0.054	0.05
SC	0.0448	0.0214	0.0599	0.0592	0.0508	0.04	0.0303
$BusTax$	0.0194	0.02	0.02	0.0265	0.0311	0.0098	0.0095
city- level							
Intercept	0.2116	0.2674	0.2427	0.2353	0.244	0.2626	0.2497
PC	0.543	0.4158	0.4674	0.4435	0.5183	0.3163	0.3739
$residuals$	0.7674	0.8165	0.776	0.795	0.7967	0.8422	0.8418
AIC	10378	10778	10490	10567	10429	10630	10870
BIC	10661	11061	10774	10851	10712	10912	11153
$\ln L$	-5144	-5343.9	-5200.1	-5238.6	-5169.6	-5269.9	-5389.9
R^2	0.3165	0.2899	0.309	0.3039	0.3131	0.2998	0.2838

Significance codes: '***', '**', '*' significant up to 1%, 5%, and 10%, respectively. Standard errors for fixed effects are in parentheses below the estimates, for random effects standard errors are reported. R^2 is McFadden-R^2 for comparison reason.

Table 6.11: Multilevel results for employment growth in the intercept-as outcome-model

	H	I	J	K	L	M	N
Fixed effects							
Intercept	0.0126	0.0222	0.0073	0.0021	0.0094	0.0003	-0.0025
	(0.1112)	(0.1267)	(0.1148)	(0.1115)	(0.1112)	(0.1109)	(0.1111)
dEmpL1	0.4347 ***	0.4036 ***	0.4332 ***	0.4421 ***	0.4503 ***	0.4392 ***	0.4368 ***
	(0.0141)	(0.0143)	(0.0141)	(0.0141)	(0.0142)	(0.014)	(0.0139)
malm	-0.0334						
	(0.0218)						
tech		-0.0538					
		(0.1008)					
eff			-0.0278				
			(0.0367)				
pure.eff				-0.0515 ***			
				(0.0193)			
pure.tech					-0.0022		
					(0.0382)		
scale.tech						-0.0135	
						(0.0688)	
scale							-0.0052
							(0.0597)
$\ln Stu$	-0.036	-0.0279	-0.0299	-0.0273	-0.0339	-0.0283	-0.0299
	(0.022)	(0.024)	(0.0229)	(0.023)	(0.0223)	(0.0225)	(0.024)
$\ln Stu^2$	-0.0073	-0.0035	-0.0005	-0.0024	-0.0041	0.0022	0.0018
	(0.0195)	(0.0191)	(0.0193)	(0.0195)	(0.0193)	(0.0188)	(0.0189)
dPop	0.0677 ***	0.0626 *	0.0611 *	0.0586	0.0621 *	0.0581 *	0.0645 *
	(0.017)	(0.036)	(0.0367)	(0.0377)	(0.0317)	(0.0328)	(0.0369)
$dPop^2$	0.0366	-0.0015	-0.0236	0.0059	0.017	-0.018	-0.0531
	(0.0315)	(0.0388)	(0.035)	(0.0396)	(0.0381)	(0.0389)	(0.0389)
Gini	0.0566 ***	0.0442 *	0.0595 ***	0.0542 **	0.0499 **	0.0424 *	0.0448 **
	(0.017)	(0.0249)	(0.0201)	(0.0214)	(0.0217)	(0.023)	(0.0226)
$Gini^2$	0.0172	0.0295 **	0.0242 *	0.0245 *	0.0237 *	0.022	0.0223
	(0.0123)	(0.0132)	(0.0142)	(0.0146)	(0.0143)	(0.0136)	(0.0146)
SC	0.0014	-0.0053	0.0006	-0.0057	0.0006	0.0032	0.0082
	(0.0308)	(0.031)	(0.0309)	(0.0314)	(0.0309)	(0.0301)	(0.0305)
SC^2	-0.0086	0.0024	-0.01	-0.0038	0.0019	-0.0027	-0.0095
	(0.0305)	(0.0301)	(0.0302)	(0.0306)	(0.0301)	(0.0292)	(0.0297)
BusTax	-0.0369	-0.0295	-0.0333	-0.0328	-0.0306	-0.0327	-0.0284
	(0.0262)	(0.0237)	(0.0243)	(0.0231)	(0.0242)	(0.024)	(0.0251)
$BusTax^2$	0.0067	0.0062	0.0021	0.0041	0.0047	0.0126	0.0112
	(0.0135)	(0.0133)	(0.0134)	(0.0136)	(0.0135)	(0.0133)	(0.0132)
Random effects							
time- level							
Intercept	0.2659	0.3046	0.2752	0.2667	0.2661	0.2655	0.266
PC	0.0262	0.2407	0.0797	0.0266	0.0813	0.1597	0.1351
$\ln Stu$	0.0201	0.0327	0.0268	0.0264	0.0223	0.026	0.033
dPop	NA	0.078	0.0797	0.0821	0.0651	0.0687	0.0805
Gini	NA	0.0448	0.026	0.0308	0.0324	0.0384	0.0368
SC	0.0025	0.0146	0.0133	0.0134	0.0118	0.0111	0.0096
BusTax	0.0556	0.0486	0.0502	0.0464	0.0501	0.0498	0.0528
city- level							
Intercept	0.0046	0.0165	0.0223	0.0267	0.0334	0.0127	0.003
PC	0.2623	0.1192	0.214	0.1637	0.1163	0.277	0.3098
residuals	0.7765	0.7875	0.777	0.7913	0.7826	0.7515	0.7622
AIC	9835	9811	9834	9886	9550	9457	9780
BIC	10037	10094	10117	10170	9833	9740	10064
$\ln L$	-4885.6	-4860.4	-4872	-4898.2	-4730.1	-4683.6	-4845.2
R^2	0.3436	0.347	0.3454	0.3419	0.3645	0.3707	0.349

Significance codes: '***', '**', '*' significant up to 1%, 5%, and 10%, respectively. Standard errors for fixed effects are in parentheses below the estimates, for random effects standard errors are reported. R^2 is McFadden-R^2 for comparison reason.

Both tables of results for the intercept-as-outcome model show similar patterns compared with those of the basic multilevel model. The results presented in the tables 6.10 and 6.11 are similar to these results with interaction terms in the tables 6.8 and 6.9. Productivity change components are significant with respect to gross value added growth in table 6.10, although both standard errors and estimates changed compared with the basic multilevel model. The Malmquist index and its component of efficiency change are not significant, but the more detailed components are significantly different from zero. Pure efficiency change (catching-up) and change in scale technology are significantly negative, while, again, pure technological change (technical progress) and change in scale efficiency are significantly positive on value added growth. Moreover, the significance of the change in the Gini coefficient is changed when compared with the basic multilevel model; in table 6.10, only the quadratic term is significantly positive and not the linear term. Therefore, the minimum point for the effect on value added growth is zero, which is the average change in the Gini coefficient. Structural change has a U-form effect on value added growth with a minimum point at 0.5, which is the same as in the estimations of the basic multilevel model.

For employment growth, the Malmquist index, change in efficiency and change in scale efficiency are not significant. Only the change in pure efficiency remains significantly negative in table 6.11 compared to table 6.9. Furthermore, table 6.11 shows that the business tax structure does not affect employment growth in the intercept-as-outcome model. Structural change also has a U-form effect on employment growth, with a minimum point at 0.5. Therefore, a structural change within the city below or equal to the average positively affects employment growth as well as a large structural change, which is more than 0.5 times the standard deviation above the average of all cities.

The McFadden-R^2 increases in the intercept-as-outcome models of both dependent variables. The information criteria show different results compared with the basic multilevel model. On the one hand and with respect to value added growth, both information criteria decline, except for the BIC for the estimation with change of scale technology and scale efficiency change as components (in the last two columns). On the other hand, for employment growth as an explanatory variable, the BIC always declines, compared with the basic multilevel model, but the AIC declines except for pure technological change and pure efficiency change (columns J and K in table 6.11).

The intercept-as-outcome configuration clearly demonstrates that the random effect standard errors, especially for the productivity change components, are of considerable size. Thus, it would be wrong to ignore the nesting structure with both levels. Furthermore, the random effects at the time-level achieve considerably high standard errors. Unfortunately, in this estimation some random effects cannot be estimated because of the correlation at that level with the error terms and are, therefore, unavailable (NA). The likelihood ratio test statistic in Eq. (57) for the intercept-as-outcome model and the basic multilevel for value added growth reach values between 124 and 671.8, which are more than the 99% quantile of the χ^2 distribution with eight degrees of freedom, which is about 20.1. Therefore, the null hypothesis, of no effect of the additional random effect, can be rejected on 1% level of significance. This null hypothesis can also be rejected for the employment growth estimations, because the test statistic is still larger than the 99% quantile with values varying between 33.8 and 172.2. The significances of the random effects indicate that the level must not be eliminated. Thus, variation in the slopes of the dependent variables is present within the levels.

6.5.3 Results for the Intercept-and-slope-as-outcome Model

Furthermore, I calculate the most general multilevel model, namely the intercept-and-slope-as-outcome model, which adds the higher-level variables as explanatory variables for the slope parameter of productivity change. Therefore, interaction terms of the explanatory variables with productivity change are included, as shown in Eq. (55). The results for value added growth with interaction terms are shown in table 6.12 and for employment growth in table 6.13.

Table 6.12: Multilevel results for gross value added growth in the intercept-and-slope-as-outcome model

	A	B	C	D	E	F	G
Fixed effects							
Intercept	-0.0288	-0.0319	-0.0462	-0.049	-0.031	-0.0298	-0.0241
	(0.0727)	(0.0858)	(0.0759)	(0.0739)	(0.078)	(0.086)	(0.0863)
$dGVAL1$	-0.0701 ***	-0.0902 ***	-0.0608 ***	-0.071 ***	-0.0873 ***	-0.1125 ***	-0.1031 ***
	(0.0152)	(0.0148)	(0.015)	(0.0153)	(0.0147)	(0.015)	(0.0148)
$malm$	-0.0437						
	(0.032)						
$tech$		-0.0142					
		(0.039)					
eff			-0.0281				
			(0.0314)				
$pure.eff$				-0.0776 ***			
				(0.026)			
$pure.tech$					0.1151 **		
					(0.0459)		
$scale.tech$						-0.1052 **	
						(0.0456)	
$scale$							0.1062 **
							(0.0492)
$PC{\cdot}\ln Stu$	0.0244	-0.0288	0.0328	-0.0039	0.0016	-0.0339	0.0452 *
	(0.0288)	(0.0245)	(0.0251)	(0.0261)	(0.0316)	(0.0233)	(0.0272)
$\ln Stu$	0.0197	0.0229	0.0263	0.0397	0.0098	0.0113	0.0192
	(0.0254)	(0.027)	(0.0264)	(0.0265)	(0.0271)	(0.0292)	(0.0288)
$\ln Stu^2$	0.0163	0.0158	0.0229	0.0301	0.0059	0.0132	0.0213
	(0.0235)	(0.0255)	(0.0246)	(0.0247)	(0.025)	(0.0262)	(0.0256)
$PC{\cdot}dPop$	-0.0004	-0.0221	0.0237	0.0517 *	-0.048	0.058 **	-0.0882 ***
	(0.0301)	(0.0271)	(0.0277)	(0.0302)	(0.0338)	(0.0279)	(0.0289)
$dPop$	0.0167	0.0334	0.0315	0.0327	0.0487 *	0.0414	0.0332
	(0.0287)	(0.0286)	(0.0291)	(0.0314)	(0.0288)	(0.027)	(0.0265)
$dPop^2$	0.0707 *	-0.0006	0.0377	0.0299	-0.0112	-0.0143	-0.0535
	(0.0413)	(0.0402)	(0.0411)	(0.0459)	(0.0425)	(0.0451)	(0.0477)
$PC{\cdot}Gini$	0.0367	0.0182	0.0052	-0.0014	0.0016	-0.0133	-0.0018
	(0.0379)	(0.0264)	(0.0276)	(0.027)	(0.0355)	(0.0224)	(0.0247)
$Gini$	0.0244	0.0328	0.0176	0.0231	0.0215	0.0266	0.0283
	(0.0207)	(0.03)	(0.0233)	(0.0245)	(0.0313)	(0.0285)	(0.028)
$Gini^2$	0.03 *	0.0254	0.0367 **	0.0239	0.0346 **	0.0301 *	0.0276
	(0.0157)	(0.0166)	(0.0165)	(0.0164)	(0.0169)	(0.0178)	(0.0174)
$PC{\cdot}SC$	0.042 *	0.0231	0.0321	0.0127	0.0173	-0.0066	0.0075
	(0.023)	(0.0192)	(0.0203)	(0.02)	(0.0201)	(0.0188)	(0.019)
SC	-0.0901 **	-0.0768 *	-0.0896 **	-0.0887 **	-0.097 **	-0.0978 **	-0.103 **
	(0.0424)	(0.0418)	(0.0444)	(0.0449)	(0.0464)	(0.0441)	(0.0422)
SC^2	0.1057 **	0.0763 *	0.1149 ***	0.1224 ***	0.1391 ***	0.1086 **	0.105 **
	(0.0427)	(0.0412)	(0.0441)	(0.0446)	(0.0439)	(0.0435)	(0.0414)
$PC{\cdot}BusTax$	0.0254	-0.0039	0.033	-0.0146	0.1044 ***	-0.0608 **	0.0508 **
	(0.0276)	(0.0243)	(0.0248)	(0.0249)	(0.03)	(0.0239)	(0.0257)
$BusTax$	-0.0326 *	-0.034 *	-0.0341 *	-0.0386 *	-0.0194	-0.0322 *	-0.0381 **
	(0.0181)	(0.0191)	(0.0181)	(0.0201)	(0.0234)	(0.018)	(0.0176)
$BusTax^2$	0.024	0.0205	0.0292 *	0.0293 *	0.0257	0.0343 *	0.027
	(0.0164)	(0.0178)	(0.0171)	(0.0173)	(0.0176)	(0.0185)	(0.0179)
Random effects							
time- level							
Intercept	0.1635	0.1954	0.1704	0.1649	0.1757	0.1954	0.1968
PC	0.0327	0.0682	0.0438	0.0046	0.0831	0.0927	0.1
$\ln Stu$	0.0164	0.0102	0.0166	0.017	0.0154	0.0241	0.0272
$dPop$	0.0493	0.0438	0.0484	0.0561	0.0442	0.0349	0.0355
$Gini$	0.0302	0.0589	0.0386	0.0426	0.0632	0.0535	0.053
SC	0.0519	0.0248	0.0539	0.0559	0.0591	0.0376	0.0278
$BusTax$	0.0206	0.0187	0.0179	0.0272	0.0393	0.0067	0.0068
city- level							
Intercept	0.211	0.2675	0.2426	0.2352	0.2456	0.262	0.2487
PC	0.5437	0.4156	0.4669	0.4451	0.5108	0.3097	0.3688
residuals	0.7673	0.8165	0.7758	0.7949	0.7963	0.8425	0.8417
AIC	10409	10811	10521	10601	10451	10656	10894
BIC	10724	11126	10835	10916	10765	10970	11209
$\ln L$	-5154.3	-5355.5	-5210.3	-5250.7	-5175.7	-5278.2	-5397.1
R^2	0.3151	0.2884	0.3077	0.3023	0.3123	0.2987	0.2829

Significance codes: '***', '**', '*' significant up to 1%, 5%, and 10%, respectively. Standard errors for fixed effects are in parentheses below the estimates, for random effects standard errors are reported. R^2 is McFadden-R^2 for comparison reason.

Table 6.13: Multilevel results for employment growth in the intercept-and-slope-as-outcome model

	H	I	J	K	L	M	N
Fixed effects							
Intercept	0.0125	0.0219	0.0073	0.0039	0.0106	-0.0002	-0.0038
	(0.111)	(0.128)	(0.1144)	(0.1117)	(0.1117)	(0.1112)	(0.1112)
*dEmpL*1	0.4342 ***	0.4014 ***	0.4331 ***	0.4425 ***	0.4488 ***	0.4377 ***	0.4358 ***
	(0.0142)	(0.0143)	(0.0141)	(0.0141)	(0.0142)	(0.014)	(0.0139)
malm	-0.0374 *						
	(0.0227)						
tech		-0.0615					
		(0.1022)					
eff			-0.0224				
			(0.0374)				
pure.eff				-0.056 ***			
				(0.0207)			
pure.tech					-0.0015		
					(0.0415)		
scale.tech						-0.0122	
						(0.0699)	
scale							-0.0011
							(0.0615)
$PC{\cdot}\ln Stu$	0.0054	0.0128	0.0001	-0.0261	0.0377 **	-0.0413 **	0.0474 **
	(0.0193)	(0.014)	(0.017)	(0.0165)	(0.0159)	(0.0207)	(0.0237)
$\ln Stu$	-0.0358	-0.027	-0.0297	-0.0263	-0.031	-0.0292	-0.0294
	(0.0219)	(0.0235)	(0.0228)	(0.0229)	(0.0216)	(0.023)	(0.0243)
$\ln Stu^2$	-0.0078	-0.0041	-0.0005	-0.0017	-0.0036	0.0027	0.0019
	(0.0195)	(0.0192)	(0.0193)	(0.0195)	(0.0193)	(0.0188)	(0.0189)
$PC{\cdot}dPop$	-0.0136	-0.0275 *	0.0108	0.0171	-0.0196	-0.0009	-0.0005
	(0.0198)	(0.0158)	(0.0182)	(0.0183)	(0.0178)	(0.0249)	(0.0254)
$dPop$	0.0679 ***	0.0622 *	0.0595	0.0568	0.0601 *	0.0579 *	0.0646 *
	(0.017)	(0.0356)	(0.0363)	(0.0375)	(0.0321)	(0.0326)	(0.0374)
$dPop^2$	0.0418	0.0023	-0.0251	-0.0029	0.0134	-0.0157	-0.0526
	(0.032)	(0.0385)	(0.0354)	(0.0404)	(0.0388)	(0.0396)	(0.0404)
$PC{\cdot}Gini$	0.0045	0.0168	-0.0141	0.0087	0.0042	0.0067	-0.0255
	(0.0227)	(0.0154)	(0.0198)	(0.0184)	(0.0184)	(0.0221)	(0.0276)
$Gini$	0.0567 ***	0.0445 *	0.06 ***	0.0545 **	0.0505 **	0.0432 *	0.045 **
	(0.0171)	(0.0249)	(0.0205)	(0.0216)	(0.0214)	(0.023)	(0.0219)
$Gini^2$	0.017	0.0274 **	0.0262 *	0.0247 *	0.0221	0.0215	0.0206
	(0.0123)	(0.0133)	(0.0145)	(0.0148)	(0.0141)	(0.0136)	(0.0143)
$PC{\cdot}SC$	0.002	0.0258 *	-0.0162	-0.0123	0.0219	0.0096	-0.006
	(0.0149)	(0.0138)	(0.0147)	(0.0135)	(0.0137)	(0.0162)	(0.0163)
SC	0.0022	-0.0038	0.0027	-0.0054	0.0055	0.0021	0.008
	(0.0309)	(0.0315)	(0.0309)	(0.0315)	(0.0309)	(0.0303)	(0.0306)
SC^2	-0.0093	0.0039	-0.0177	-0.0063	-0.0067	-0.0014	-0.009
	(0.0305)	(0.0304)	(0.0303)	(0.0307)	(0.0304)	(0.0292)	(0.0298)
$PC{\cdot}BusTax$	-0.0368 **	-0.0247 *	-0.0022	-0.0131	-0.0027	-0.0183	0.0117
	(0.0185)	(0.0136)	(0.0166)	(0.0161)	(0.0161)	(0.0208)	(0.0221)
$BusTax$	-0.0371	-0.0312	-0.033	-0.0324	-0.0322	-0.0336	-0.0291
	(0.0255)	(0.0231)	(0.0242)	(0.023)	(0.0241)	(0.0234)	(0.0246)
$BusTax^2$	0.0065	0.0067	0.0022	0.0029	0.0039	0.013	0.0108
	(0.0135)	(0.0133)	(0.0134)	(0.0136)	(0.0135)	(0.0133)	(0.0132)
Random effects							
time- level							
Intercept	0.2654	0.308	0.2741	0.2671	0.2672	0.2664	0.2664
PC	0.0248	0.2439	0.0797	0.0272	0.0883	0.1609	0.1378
$\ln Stu$	0.0192	0.0298	0.0262	0.0255	0.0173	0.0284	0.0343
$dPop$	NA	0.0767	0.0786	0.0814	0.066	0.0683	0.0817
$Gini$	NA	0.0446	0.0279	0.0314	0.0311	0.0383	0.0345
SC	0.0028	0.0193	0.0117	0.0146	0.0106	0.0137	0.0112
$BusTax$	0.0535	0.0469	0.0499	0.0461	0.0497	0.0479	0.0513
city- level							
Intercept	0.0043	0.0224	0.0229	0.0288	0.031	0.0137	0.0028
PC	0.2634	0.0962	0.2169	0.1672	0.113	0.2746	0.311
residuals	0.7763	0.7892	0.7769	0.7909	0.7826	0.7516	0.7619
AIC	9871	9842	9873	9924	9585	9490	9813
BIC	10104	10157	10188	10239	9898	9804	10128
$\ln L$	-4898.4	-4870.8	-4886.7	-4912.2	-4742.3	-4695.1	-4856.7
R^2	0.3419	0.3456	0.3434	0.34	0.3628	0.3692	0.3475

Significance codes: '***', '**', '*' significant up to 1%, 5%, and 10%, respectively. Standard errors for fixed effects are in parentheses below the estimates, for random effects standard errors are reported. R^2 is McFadden-R^2 for comparison reason.

The intercept-and-slope-as-outcome model changes the results further because of the additional fixed effects. The results in the tables 6.12 and 6.13 are comparable with the results of the OLS estimations in the tables 6.4 and 6.5. The interaction terms are only significant in a few cases, e.g., with business tax. Nonetheless, productivity change and the components as well as structural change, change in the Gini coefficient and business tax are significant as in the intercept-as-outcome model. Pure technological change affects significantly positive value added growth with an estimate of 0.1151 in column E. Therefore, technological progress has a positive effect on value added growth in German cities as expected in evolutionary economic geography. However, this effect is offset by the negative coefficient of change in scale technology in column F. Furthermore, change in pure efficiency has a negative effect on value added growth with a coefficient of -0.0776 in column D, although there are interaction terms. The effect was already observable in the basic multilevel and intercept-as-outcome model. A possible explanation is that the increase of pure efficiency and thus the catching-up to the production frontier increases the degree of competition in the market and by doing so results in a decrease of growth (Aghion and Howitt (2009, p. 92)). Additional explanations are the negative feedback of firm growth in Frenken and Boschma (2007, p. 643) because the effect on gross value added is two years later (see Eq. (55)).

With respect to the results for employment growth presented in table 6.13, the Malmquist index and the component pure efficiency change significantly affect employment growth negatively. In case of employment growth as explained variable, technological progress measured by pure technological change has no effect while change of pure efficiency in column K is significantly negative with a value of -0.056 although there are interaction terms. However, the interaction terms are even less important than in the case of value added growth. The interactions of productivity change are only significant for the number of students and business tax. Because of the insignificance of most interaction terms with productivity change, the log-likelihood does not improve and even decreases in the REML estimation, which is not best for evaluating the significance of fixed effects. This shows, that the effect of technological change and efficiency change on industrial growth does not depend on the local variables investigated.

The intercept-as-outcome model without interaction terms already generates the best results. This finding is supported by the increasing AIC and BIC, which increase for every estimation.

An evaluation of the coefficients is possible with the OLS model only because all the fixed effects are the same except that the intercept-and-slope-as-outcome model contains random effects. The likelihood-ratio test statistic in Eq. (57) with the intercept-and-slope-as-outcome model and the OLS estimates for value added growth varies between 187.2 and 729.2. Therefore, the null hypothesis of no significant level effects (no random effects) can be rejected at the 1% level of significance and nine degrees of freedom (nine random effects are estimated). The same applies for employment growth, whose results of the test statistic vary between 225 and 378.8, which is still larger than the 99% quantile of the χ^2 distribution of 21.7. This shows the importance of the multilevel random effects.

Furthermore, the results are proven to be robust to the elimination of insignificant variables as well as the three large cities, namely Berlin, Hamburg, and Bremen (together with Bremerhaven), which are not only free cities but also sovereign states in Germany, with further competences and a large number of inhabitants.

6.6 Summary

Because we live in an urban world with more than half of the world's population living in cities and given that most economic activity takes place in cities, it is important to know what forces foster industrial growth and to learn about the role of city-specific circumstances. In Germany, cities classified as urban municipalities have the power to influence many variables, like business tax structure and expenditure for transportation facilities. These cities compete against each other by their individual characteristics in order to attract new entrepreneurs and support established industries, increase income and tax revenue. In addition, the industrial structure is different between cities, and several analyses have observed externalities which arise by closeness and innovation in the same or a different but related industry. However, if individuals interact with each other within a city, the nesting structure should be an important feature and without its consideration results in biased estimates. Multilevel analyses offer the tools to solve this problem and to estimate unbiased results with corrected standard errors by accounting for the nesting structure.

It turns out that the multilevel structure is appropriate for analyzing the industrial performance of cities observed over subsequent years. The development of industries is different between different cities, offering a specific environment. Yearly productivity change as estimated with the non-parametric DEA, such as the Malmquist index and its components, affect value added growth and employment growth. In particular, efficiency change, which captures catching-up to the best practice frontier of industries, is negatively associated with both value added growth and employment growth. This can be interpreted as Schumpeter's creative destruction of innovations. Pure technological progress fosters industrial value added growth.

Furthermore, the growth path leads to an adoption of value added growth and a divergence in employment growth in German cities. Several additional forces are found to be significantly related to value added growth and employment growth. The effects are not only linear but also quadratic. For example, the structural change of the industrial composition in the cities shows a U-shaped form, indicating that both large changes in the industrial structure increase value added growth and employment growth, but also that no or lower than average structural change is fruitful. However, interactions are not found to have a significant effect on industrial growth. This implies that the effect of productivity change is independent of the other city specifics. A negative effect of industrial concentration on employment growth, as found by Noseleit (2013) for West German agglomerations over the period between 1983 and 2002, has not been found in the data set of all cities in the most recent years. Instead, only the increase of the change in industrial concentration has been found to have positive effects on value added growth and employment growth with significant quadratic terms.

7 Conclusion

In this dissertation the role of urban efficiency and the sectoral structure in relation to a range of topics has been empirically investigated. This chapter summarizes the main results of the previous chapters, connects the chapters and suggests further ideas for future research on the topic of urban efficiency.

Chapter 2 summarizes the empirical methods for the investigations of the following chapters. Since all empirical analyses investigate efficiency and productivity change, the concept of the non-parametric DEA is presented. DEA and the Malmquist index can be individually estimated for each economic sector and separately for each year. To account for different types of cities as a consequence of urban hierarchy as implemented e.g., in Au and Henderson (2006a), a cluster analysis for the cities is performed. The cluster analysis with the Ward approach results in almost equally-sized clusters. An advantage of performing a hierarchical cluster analysis is that an optimal number of clusters can be estimated, which is found to be four in the case of hierarchically ordered cities. With these approaches and the estimated and collected urban data set, the empirical analysis is performed.

Within the empirical analysis one problem is the heterogeneity within regional data. Heterogeneity results from the diversity of regional units and omitted variables in the estimations. Cities are specialized in some industries within the urban hierarchy and also diversified to benefit from diversification externalities (Jacobs, 1970). Therefore, diversity is expected to be present for the free cities. In addition, outliers are observable within the data especially with respect to cities' size because there are only a few very large cities but many small cities. As a result, not only are outliers present but also heteroscedasticity as variation decreases by city size. These problems cause the OLS estimates to be biased and inefficient; in such a case, the tests of significance are not trustworthy. Robust estimates, which are not affected by data outliers, and heteroscedasticity consistent estimations are introduced in chapter 2.

Cities are geographical units with competing firms and city councils or local/municipality governments. Trade and for example travel connections between cities also cause spatial dependence between them. This spatial dependence, causing correlation between the error terms in OLS estimations but also heteroscedasticity within the data, can be accounted for by spatial models which are also introduced in chapter 2. The introduction of spatial models enables different approaches to account for spatial dependence. A testing procedure is explained to analyze the necessity for the application of spatial models which account for omitted variables as presented in Mur and Angulo (2009).

Furthermore, an introduction to multilevel models is given in chapter2. Multilevel models account for the nesting structure of the data, since industries are nested within cities and these are repeatedly observed. This nesting structure allows for the estimation of fixed and random effects and thereby allows for unobserved factors on each level (Pinheiro and Bates, 2000). As in the case of spatial regression models, a testing procedure is introduced with the goal of identifying the multilevel model which fits the data best, while accounting for the number of coefficients to be estimated.

In chapter 3, efficiency is related to city size. Efficiency is the scale efficiency which is measured as the ratio of efficiency under constant returns to scale and under variable returns to scale. It is measured in an output-orientated way, in which case the comparison is in output quantities for equal input quantities. The size of the urban economy is scale efficient when it is operating at the technical optimal productive scale. Since economic output and population size are highly correlated, a comparison with the handy population size is possible. On the one hand, cities operating below the optimal size are producing in the range of increasing returns to scale and are therefore able to increase productivity by growing. On the other hand, cities larger than the optimal size produce at decreasing returns to scale, with decreasing productivity by larger population and consequently larger output size. By estimating a quadratic function, an optimal city size is found similar to Capello and Camagni (2000) for Italian cities. The optimal size, at which scale efficiency is highest, is 220,000 inhabitants.

The optimal city size refers to the mean city size. This result is robust e.g., by accounting for the endogeneity in the city size and the estimated efficiency estimation, because city size includes the urban employment force as one of the input factors. For resolving endogeneity, modified bootstrap algorithms are constructed and estimated. In addition, the analysis is performed by separating geographical and hierarchical groups. If the coefficients are significant, the optimal size as indicated by the estimates is the mean city size for each of the separate analyses. Consequently, there are cities which are too large and too small with respect to scale efficiency in each group. However, for a combined separation of hierarchy and geography, the number of observations within each group is too small for reasonable investigations. Not only is the combined investigation of non-parametric efficiency measures and optimal city size for Germany novel, but also the application of bootstrap algorithms of Simar and Wilson (2007), which were adapted to the scale efficiency measures. In chapter 3, the scale efficiency score is estimated in output-orientation and the effects of outliers are compensated by bias-corrected efficiency scores and robust regressions.

In the 4th chapter, efficiency is estimated by stochastic non-parametric approaches which account for measurement errors and take only a fraction of more productive units to construct the reference points on the production function. Furthermore, the restrictive output-orientation within chapter 3 is abandoned in chapter 4 in favor of an input- and output-orientated measurement and a combination of both with the hyperbolic-graph approach. The results show that input efficiency is decreasing by city size as the smallest cities have the highest efficiency scores in input-orientation. Conversely, output-orientated efficiency scores are increasing with city size. The hyperbolic-graph efficiency measurements, which represent a combination of both orientations, confirm the findings. Urban hierarchy is considered by comparing cities with hierarchically-similar cities only within managerial efficiency. The U-shaped form of population to the hyperbolic graph efficiency score is mainly caused by differences between cities of different hierarchical orders. Hence, urban hierarchy is one characteristic for explaining differences between cities. The population influences the efficiency scores not only linearly but also quadratically.

Policy makers should consider that a change in urban size also includes a change in scale efficiency. Cities above the mean size should not consider increasing their size while cities with population below the mean size should be supported in increasing city size. By doing so, the hierarchical order should also be considered. In addition, it would be very interesting to evaluate the causality between city growth and change in efficiency. This analysis could be similar to the investigation of efficiency in Spanish regions by industry sector and time conducted by Maudos et al. (2000), who test for convergence and increasing efficiency in a developing process. Moreover, other reasons for the observed inefficiency should be determined in order to propose solutions for increasing efficiency without changing the population size in cities which are too large or too small. Furthermore, future research could possibly analyze additional variables in order to explain the urban efficiency or inefficiency without focusing on optimal city size.

The variables of quadratic population as well as urban hierarchy together with the non-parametric efficiency scores are then related to urban price levels in chapter 5. Following a specific-to-general approach to identify the appropriate model with which to explain urban price levels, the tests confirm the presence of spatial correlation. The price level is reduced by the median housing prices for each city. Spatial correlation appears to be an existing feature of the data. As a result the spatial autoregressive model is proved to be appropriate and not the spatial Durbin model, which indicates that omitted variables are not a concern for the investigation.

The results show that price levels have a U-shaped form with respect to population confirming the quadratic feature of city size. Efficiency only has an effect on price levels in the case of the financial and business services sector. While the effect on price level without housing costs is positive for the efficiency score, the change of efficiency observed over five years has a negative impact. Conversely, change of efficiency positively affects price level in the case of the manufacturing sector. However, change in the efficiency within the wholesale and retail sector is not found to significantly affect price levels although price differences are observed for retail prices. Controlling for economic and business tax differences, no significant difference between East and West German cities is found. In addition, as the price level investigations indicate, population size affects price levels. A doubling in the number of population for an average sized city with 232,710 inhabitants increases the urban price level by only 1.1 percent. This effect is very small and therefore is of minor importance for local government when considering whether to generate higher tax revenues by higher prices. Conversely, disposable income has the highest effect on future regional price levels. Thus, a responsible consideration of the price effects is required in the collective wage and salary negotiations.

The results strongly suggest fundamental differences between the different sectors. This suggestion is further proven by the fact that price levels differ for the hierarchically ordered clusters. Although the price level is highest in the most highly ordered cities, price levels do not steadily increase by hierarchical order. For the spatial autoregressive model tests in the sense of a Hausman test are individually applied to the estimates. Neighboring cities positively influence price levels within a radius of 100 km.

Opportunities for future research would be offered by additional years of observed regional price levels. Additional price levels could then be used to investigate the development of regional price levels and the effect of regional productivity change. In addition, price level data over subsequent years could be used to regionally deflate monetary urban characteristics or regional inflation dynamics similar to the analysis in Beck et al. (2006) for larger European regions. It is therefore not intuitive that the official statistical institute is not interested in local prices. The newly implemented market transparency unit for wholesale trade in electricity and gas is one new way to gain price information. Although in this case only used for gas prices which is only a small fraction of the average market basket with 3.6 percent (BBSR, 2009, p. 88). However, the new information of the market transparency unit offer new possibilities to investigate specific local prices, their development and spatial dependency.

The empirical analysis in the chapter 6 accounts for the sectoral differences emphasized in chapter 5. Chapter 6 employs a dynamic analysis of value added growth and employment growth of the economic sectors within the cities. The data set in chapter 6 is generally a panel in which cities are observed for subsequent years, but in addition within each city the economic development of the economic sectors is observed. The nesting structure is implemented within the empirical analysis by adopting a multilevel analysis with economic sectors, cities and time as the three levels. The novel contribution is to empirically analyze three distinct levels within one model and not only two levels. The multilevel model helps to account for the data structure and the dependence of the observed units within each level. Furthermore, variation caused by unobserved variables is incorporated at each level by adding random effects at each level. Tests are carried out proving the presence of random variation and the need for applying multilevel models.

Path dependency, as emphasized by evolutionary economic geography models, is found to be a significant feature for value added and employment development of the economic sectors in the free German cities. While value added develops on a converging growth path, urban employment growth is diverging. Since the variables are centered, the estimates show that cities with above average employment growth in the past are expected to have positive recent employment growth. Productivity change fueled by changes in technology was only found to explain value added growth but not employment growth. Conversely, change in efficiency negatively affects value added growth and employment growth as explained in Frenken and Boschma (2007). In contrast to the previous analyses, population size is not found to significantly add explanatory information. However, change of population as well as sectoral change, sectoral specialization and business tax are implemented in a quadratic form within the analysis to account for non-linearities. U-shaped effects are found for the sectoral change, business tax and the sectoral specialization but not for population change, which only linearly affects value added growth. Although business tax does not have an effect on urban price levels and employment growth, value added growth is negatively affected, which indicates that an increase in local business tax lowers future value added growth.

A more detailed look into industry-specific results is only possible with further monitoring and information about cities over an extended period of time. Moreover, the lag structure of the variables might be refined, because decisions do not have to be based on the observations of the last year. However, to integrate more time lag structures, a larger data set is needed to be able to pass additional yearly observations. It is also interesting to see how the recently published regional child care rate affects regional employment with sectoral distinction.

Furthermore, future investigations should extend the analysis to other countries. A multicountry model would be possible within a multilevel analysis with a further added level as country level. However, researchers should be aware of the fact that it is problematic to extend the analysis with other countries because not only are the economic sectoral structure likely to differ but also the tax systems likely to be different and incomparable. Investigations of multilevel models remain computer- and time-intensive. Adding further random effects and levels increases the time required to calculate the models. The analyzed models are considerably large and already provide interesting insights into the urban level of sectoral growth. Nonetheless, further refinements of the estimation approaches would be possible. Promising future work includes the implementation of spatial correlation within multilevel analysis by combining spatial weight matrices at the city level and temporal weight matrices to account for autocorrelation of the first and higher order at the time level. Corrado and Fingleton (2011) present and discuss some ways of modeling spatial effects within multilevel analysis. They include separate possibilities for modeling spatial effects as additional random effects within the error term or as part of the fixed effects included with interaction effects. Including spatial weight matrices within multilevel models might be fruitful.

Future analyses of employment and value added growth model will have to close the evolutionary cycle between income growth and the generation of innovations, as shown in Fratesi (2010). Unfortunately, the Malmquist index and its components are difficult to implement as endogenous variables in such an analysis, as the result of their construction, which includes endogeneity problems analogous to those discussed in Simar and Wilson (2007) and Thanassoulis et al. (2008, p. 343).

References

Abreu, M. (2014). Neoclassical Regional Growth Models. In M. M. Fischer and P. Nijkamp (Eds.), *Handbook of Regional Science*, Volume 1, Chapter 10, pp. 169–191. Berlin: Springer.

Aghion, P. and P. W. Howitt (2009). *The Economics of Growth*. Cambridge: MIT Press.

Aigner, D. and S. Chu (1968). On Estimating the Industry Production Function. *American Economic Review 58*, 826–839.

Aitchison, J. and C. G. G. Aitken (1976). Multivariate Binary Discrimination by Kernel Method. *Biometrika 63*, 413–420.

Albouy, D. (2008). Are Big Cities Bad Places to Live? Estimating Quality of Life Across Metropolitan Areas. NBER Working Paper Series 14472, November.

Alecke, B., T. Mitze, and G. Untiedt (2011). Regionale Wachstumseffekte der GRW-Förderung? Eine Räumlich-Ökonometrische Analyse auf Basis Deutscher Arbeitsmarktregionen. In C. Dreger, R. Kosfeld, and M. Türck (Eds.), *Empirische Regionalforschung heute: Festschrift für Professor Dr. Hans-Friedrich Eckey*, Chapter 3, pp. 51–109. Wiesbaden: Gabler.

Alonso, W. (1971). The Economics of Urban Size. *Papers in Regional Science 26*, 67–83.

Anselin, L. (1988a). A Test for Spatial Autocorrelation in Seemingly Unrelated Regressions. *Economic Letters 28*, 335–341.

Anselin, L. (1988b). Lagrange Multiplier Test Diagnostics for Spatial Dependence and Spatial Heterogeneity. *Geographical Analysis 20*, 1–17.

Anselin, L. (1988c). *Spatial Econometrics: Methods and Models*. Dordrecht: Kluwer Academic Publishers.

Anselin, L. (2002). Under the Hood: Issues in the Specification and Interpretation of Spatial Regression Models. *Agricultural Economics 27*, 247–267.

Anselin, L. (2003a). Spatial Econometrics. In B. H. Baltagi (Ed.), *A Companion to Theoretical Econometrics*, Chapter 14, pp. 310–330. Malden: Blackwell Publishing.

Anselin, L. (2003b). Spatial Externalities, Spatial Multipliers, and Spatial Econometrics. *International Regional Science Review 26*, 153–166.

Anselin, L. and A. K. Bera (1998). Spatial Dependence in Linear Regression Models with an Introduction to Spatial Econometrics. In A. Ullah and D. E. A. Giles (Eds.), *Handbook of Applied Economic Statistics*, Volume 155 of *Statistics: Textbooks and Monographs*, Chapter 7, pp. 237–289. New York: Marcel Dekker, Inc.

Anselin, L., A. K. Bera, R. Florax, and M. J. Yoon (1996). Simple Diagnostic Test for Spatial Dependence. *Regional Science and Urban Economics 26*, 77–104.

Anselin, L. and R. J. Florax (1995). Small Sample Properties of Tests for Spatial Dependence in Regression Models: Some Further Results. In L. Anselin and R. J. Florax (Eds.), *New Directions in Spatial Econometrics: Advances in Spatial Science*, Chapter 2, pp. 21–74. Berlin: Springer.

Anselin, L. and S. Rey (1991). Properties of Tests for Spatial Dependence in Linear Regression Models. *Geographical Analysis 23*, 112–131.

Aragon, Y., A. Daouia, and C. Thomas-Agnan (2005). Nonparametric Frontier Estimation: A Conditional Quantile-Based Approach. *Econometric Theory 21*, 358–389.

Arnott, R. J. (1979). Optimal City Size in a Spatial Economy. *Journal of Urban Economics 6*, 65–89.

Arnott, R. J. (2004). Does the Henry George Theorem Provide a Practical Guide to Optimal City Size? *The American Journal of Economics and Sociology 63*, 1057–1090.

Arnott, R. J. and J. E. Stiglitz (1979). Aggregate Land Rents, Expenditure on Public Goods, and Optimal City Size. *The Quarterly Journal of Economics 18*, 471–500.

Arrow, K. J. (1962). The Economic Implications of Learning by Doing. *Review of Economic Studies 29*, 155–173.

Au, C.-C. and J. V. Henderson (2006a). Are Chinese Cities Too Small. *Review of Economic Studies 73*, 549–576.

Au, C.-C. and J. V. Henderson (2006b). How Migration Restrictions Limit Agglomeration and Productivity in China. *Journal of Development Economics 80*, 350–388.

Badunenko, O. (2010). Downsizing in the German Chemical Manufacturing Industry During the 1990s. Why is Small Beautiful? *Small Business Economics 34*, 413–431.

Balassa, B. (1964). The Purchasing-Power-Parity Doctrine: A Reappraisal. *Journal of Political Economy 72*, 584–596.

Baldwin, J. R., D. Beckstead, W. M. Brown, and D. L. Rigby (2007). Urban Economies and Productivity. Economic Analysis Research Paper Series 2007045e.

Baldwin, J. R., W. M. Brown, and D. L. Rigby (2010). Agglomeration Economies: Microdata Panel Estimation from Canadian Manufacturing. *Journal of Regional Science 50*, 915–934.

Banker, R. D. (1984). Estimating Most Productive Scale Size in Data Envelopment Analysis. *European Journal of Operation Research 17*, 35–44.

Banker, R. D. and R. M. Thrall (1992). Estimation of Returns to Scale Using Data Envelopment Analysis. *European Journal of Operation Research 62*, 74–84.

Bartik, T. J. (1991). *Who Benefits From State and Local Economic Development Policies?* Kalamazoo (Mich.): W. E. Upjohn Institute for Employment Research.

Batabyal, A. A. and P. Nijkamp (2012). Retraction of "A Schumpeterian Model of Entrepreneurship, Innovation, and Regional Economic Growth". *International Regional Science Review 35*, 464–486.

Batabyal, A. A. and P. Nijkamp (2013). A Multi-region Model of Economic Growth with Human Capital and Negative Externalities in Innovation. *Journal of Evolutionary Economics 23*, 909–924.

Bathelt, H. and P.-F. Li (2014). Evolutionary Economic Geography and Relational Geography. In M. M. Fischer and P. Nijkamp (Eds.), *Handbook of Regional Science*, Volume 2, Chapter 31, pp. 591–607. Berlin: Springer.

Beck, G. W., K. Hubrich, and M. Marcellino (2006). Regional Inflation Dynamics within and Across Euro Area Countries and a Comparison with the US. European Central Bank Working Paper Series No 681, October 2006.

Bergstrand, J. H. (1991). Structural Determinants of Real Exchange Rates and National Price Levels: Some Empirical Evidence. *American Economic Review 81*, 325–334.

Bettencourt, L. M. A., J. Lobo, D. Helbing, C. Kühnert, and G. B. West (2007). Growth, Innovation, Scaling, and the Pace of Life in Cities. *Proceedings of the National Academy of Sciences 104*, 7301–7306.

Bivand, R. S. (1984). Regression Modeling with Spatial Dependence: An Application of Some Class Selection and Estimation Methods. *Geographical Analysis 16*, 25–37.

Black, D. and J. V. Henderson (2003). Urban Evolution in the USA. *Journal of Economic Geography 3*, 343–372.

Blien, U., H. Gartner, H. Stüber, and K. Wolf (2009). Regional Price Levels and Agglomeration Wage Differential in Western Germany. *Annals of Regional Science 43*, 71–88.

Boschma, R. A. and K. Frenken (2011). The Emerging Empirics of Evolutionary Economic Geography. *Journal of Economic Geography 11*, 295–307.

Boschma, R. A. and J. G. Lambooy (1999). Evolutionary Economics and Economic Geography. *Journal of Evolutionary Economics 9*, 411–429.

Bryk, A. S. and S. W. Raudenbush (1988). Toward a More Appropriate Conceptualization of Research on School Effects: A Three-Level Hierarchical Linear Model. *American Journal of Education 97*, 65–108.

Bundesinstitut für Bau-, Stadt- und Raumforschung (2009). *Regionaler Preisindex.* Berichte Band 30. Bonn: Bundesamt für Bauwesen und Raumordnung. (BBSR) or Federal Institute for Research on Building, Urban Affairs and Spatial Development.

Campos, J., N. R. Ericsson, and D. F. Hendry (2005, August). General-to-specific Modeling: An Overview and selected Bibliography. International Finance Discussion Papers No. 838, Board of the Governors of the Federal Reserve System.

Cantner, U., J. Krüger, and H. Hanusch (2007). *Produktivitäts- und Effizienzanalyse: Der nicht-parametrische Ansatz.* Berlin: Springer.

Capello, R. and R. Camagni (2000). Beyond Optimal City Size: An Evaluation of Alternative Urban Growth Patterns. *Urban Studies 37*, 1479–1496.

Carlino, G. A., S. Chatterjee, and R. M. Hunt (2007). Urban Density and the Rate of Invention. *Journal of Urban Economics 61*, 389–419.

Carriazo, F. and E. Coulson (2010). A Note on Testing for Spatial Error Components. *Regional Science and Urban Economics 40*, 331–335.

Caves, D. W., L. R. Christensen, and W. E. Diewert (1982). The Economic Theory of Index Numbers and Measurement of Input, Output, and Productivity. *Econometrica 50*, 1393–1414.

Cebula, R. J. (1989). The Analysis of Geographic Living-Cost Differentials: A Brief Empirical Note. *Land Economics 65*, 64–67.

Cebula, R. J. and S. Todd (2004). An Empirical Note on Determinants of Geographical Living-Cost Differentials for Counties in the State of Florida, 2003. *Review of Regional Studies 34*, 112–119.

Cebula, R. J. and M. Toma (2008). An Empirical Analysis of Determinants of Interstate Living-Cost Differentials, 2005. *Journal of Regional Analysis and Policy 38*, 222–228.

Charnes, A., W. W. Cooper, and S. Li (1989). Using Data Envelopment Analysis to Evaluate Efficiency in the Economic Performance of Chinese Cities. *Socio-Economic Planning Sciences 23*, 325–344.

Charnes, A., W. W. Cooper, and E. Rhodes (1978). Measuring the Efficiency of Decision Making Units. *European Journal of Operation Research 2*, 429–444.

Charnes, A., W. W. Cooper, and E. Rhodes (1981). Evaluating Program and Managerial Efficiency: An Application of Data Envelopment Analysis to Program Follow Through. *Managment Science 27*, 668–697.

Christaller, W. (1933). *Die zentralen Orte in Süddeutschland: Eine ökonomisch-geographische Untersuchung über die Gesetzmäßigkeit der Verbreitung und Entwicklung der Siedlungen mit städtischen Funktionen*. Jena: Gustav Fischer.

Cingano, F. and F. Schivardi (2004). Identifying the Sources of Local Productivity Growth. *Journal of the European Economic Association 2*, 720–742.

Cliff, A. D. and J. K. Ord (1981). *Spatial Processes: Models and Applications*. London: Pion.

Coelli, T. J., D. P. Rao, C. J. O'Donnell, and G. E. Battese (2005). *An Introduction to Efficiency and Productivity Analysis, 2nd Ed.* (2nd ed.). New York: Springer.

Combes, P.-P., G. Duranton, L. Gobillon, D. Puga, and S. Roux (2012). The Productivity Advantages of Large Cities: Distinguishing Agglomeration from Firm Selection. *Econometrica 80*, 2451–2594.

Combes, P. P., G. Duranton, and H. G. Overman (2005). Agglomeration and the Adjustment of the Spatial Economy. *Papers in Regional Science 84*, 311–349.

Cordy, C. B. and D. A. Griffith (1993). Efficiency of least squared estimators in the presence of spatial autocorrelation. *Communications in Statistics - Simulation and Computation 22*, 1161–1179.

Corrado, L. and B. Fingleton (2011). Multilevel Modelling with Spatial Effects. Strathclyde Discussion Papers in Economics No. 11-05, February 2011.

Crescenzi, R. (2014). Changes in Economic Geography Theory and the Dynamics of Technological Change. In M. M. Fischer and P. Nijkamp (Eds.), *Handbook of Regional Science*, Volume 2, Chapter 34, pp. 649–666. Berlin: Springer.

Crucini, M. J., M. Shintani, and T. Tsuruga (2009, April). The Law of One Price Without the Border: The Role of Distance Versus Sticky Prices. NBER Working Paper 14835.

Cullmann, A. and C. von Hirschhausen (2008). Efficiency Analysis of East European Electricity Distribution in Transition: Legacy of the Past? *Journal of Productivity Analysis 29*, 155–167.

Daouia, A. and L. Simar (2007). Nonparametric Efficiency Analsis: A Multivariate Conditional Quantile Approach. *Journal of Econometrics 140*, 375–400.

Dascher, K. (2002). Capital Cities: When Do They Stop Growing? *Papers in Regional Science 81*, 49–62.

Dietrich, A. (2009). Does Growth Cause Structural Change, or Is it the Other Way Round? A Dynamic Panel Data Analyses for Seven OECD Countries. Jena Economic Research Papers 2009-034.

Dopfer, K., J. Foster, and J. Potts (2004). Micro-Meso-Macro. *Journal of Evolutionary Economics 14*, 263–279.

Duranton, G. (1999). Distance, Land, and Proximity: Economic Analysis and the Evolution of Cities. *Environment and Planning A 31*, 2169–2188.

Duranton, G. and D. Puga (2000). Diversity and Specialisation in Cities: Why, Where and When Does it Matter? *Urban Studies 37*, 533–555.

Egert, B. (2007). Real Convergence, Price Level Convergence and Inflation Differentials in Europe. CESifo Working Paper No. 2127, October 2007.

Elhorst, J. P. (2010). Applied Spatial Econometrics: Rising the Bar. *Spatial Econometric Analysis 5*, 9–28.

Engel, C. and J. H. Rogers (1996). How Wide Is the Border? *American Economic Review 86*, 1112–1125.

Engel, C. and J. H. Rogers (1998). Regional Patterns in the Law of One Price: The Role of Geography versus Currencies. In J. A. Frenkel (Ed.), *The Regionalization of the World Economy*, Chapter 6, pp. 153–188. University of Chicago Press. January 1998.

Farrell, M. J. (1957). The Measurement of Productive Efficiency. *Journal of the Royal Statistical Society Series A 120*, 253–281.

Florax, R. J. and S. Rey (1995). The Impacts of Misspecified Spatial Interactions in Linear Regression Models. In L. Anselin and R. J. Florax (Eds.), *New Directions in Spatial Econometrics: Advances in Spatial Science*, Chapter 5, pp. 111–135. Berlin: Springer.

Florax, R. J. G. M. and H. Folmer (1992). Specification and Estimation of Spatial Linear Regression Models, Monta Carlo Evaluation of Pre-test Estimators. *Regional Science and Urban Economics 22*, 405–432.

Florax, R. J. G. M., H. Folmer, and S. J. Rey (2003). Specification Searches in Spatial Econometrics: The Relevance of Hendry's Methodology. *Regional Science and Urban Economics 33*, 557–579.

Florax, R. J. G. M., H. Folmer, and S. J. Rey (2006). A Comment on Specification Searches in Spatial Econometrics: The Relevance of Hendry's Methodology: A Reply. *Regional Science and Urban Economics 36*, 300–308.

Fox, J. and G. Monette (1992). Generalized Collinearity Diagnostics. *Journal of the American Statistical Association 87*, 178–183.

Fratesi, U. (2010). Regional Innovation and Competitiveness in a Dynamic Representation. *Journal of Evolutionary Economics 20*, 515–552.

Färe, R., S. Grosskopf, B. Lindgren, and P. Roos (1992). Productivity Changes in Swedish Pharmacies 1980-1989: A Non-Parametric Malmquist Approach. *Journal of Productivity Analysis 3*, 85–101.

Färe, R., S. Grosskopf, and C. A. K. Lovell (1985). *The Measurement of Efficiency of Production*. Boston: Kluwer Academic Publishers.

Färe, R., S. Grosskopf, and M. Norris (1997). Productivity Growth, Technical Progress, and Efficiency Change in Industrialized Countries: Reply. *American Economic Review 87*, 1040–1043.

Färe, R., S. Grosskopf, M. Norris, and Z. Zhang (1994). Productivity Growth, Technical Progress, and Efficiency Change in Industrialized Countries. *American Economic Review 84*, 66–83.

Frenken, K. and R. A. Boschma (2007). A Theoretical Framework for Evolutionary Economic Geography: Industrial Dynamics and Urban Growth as a Branching Process. *Journal of Economic Geography 7*, 635–649.

Fu, S. and J. Hong (2011). Testing Urbanization Economies in Manufacturing Industries: Urban Diversity or Urban Size? *Journal of Regional Science 51*, 585–603.

Fujita, M., P. Krugman, and T. Mori (1999). On the Evolution of Hierarchical Urban Systems. *European Economic Review 43*, 209–251.

Gaffard, J.-L. (2008). Innovation, Competition, and Growth: Schumpeterian Ideas within a Hicksian Framework. *Journal of Evolutionary Economics 18*, 295–311.

Gallo, J. L. (2014). Cross-Section Spatial Regression Models. In M. M. Fischer and P. Nijkamp (Eds.), *Handbook of Regional Science*, Volume 3, Chapter 34, pp. 1511–1533. Berlin: Springer.

Gartell, A. C. (1979). Autocorrelation in Spaces. *Environment and Planning A 11*, 507–516.

Geary, R. C. (1954). The Contiguity Ratio and Statistical Mapping. *The Incorporated Statistican 5*, 115–145.

George, H. (1880). *Progress and Poverty*. New York: Robert Schalkenbach Foundation.

Getis, A. (2007). Reflections on Spatial Autocorrelation. *Regional Science and Urban Economics 37*, 491–496.

Getis, A. and J. K. Ord (1992). The Analsis of Spatial Association by Use of Distance Statistics. *Geographical Analysis 24*, 189–206.

Giovannetti, G., G. Ricchiuti, and M. Velucchi (2009, September). Location, Internationalization and Performance of Firms in Italy: A Multilevel Approach. Universita' degli Studi di Firenze, Dipartimento di Scienze Economiche, Working Papers Series N. 09/2009.

Glaeser, E. L. (1998). Are Cities Dying? *Journal of Economic Perspectives 12*, 139–160.

Glaeser, E. L., H. D. Kallal, J. A. Scheinkman, and A. Shleifer (1992). Growth in Cities. *Journal of Political Economy 100*, 1126–1152.

Glaeser, E. L. and D. C. Maré (2001). Cities and Skills. *Journal of Labor Economics 19*, 316–342.

Glaeser, E. L., J. A. Scheinkman, and A. Shleifer (1995). Economic Growth in a Cross-Section of Cities. *Journal of Monetary Economics 36*, 117–143.

Goedhuys, M. and M. Srholec (2010). Understanding Multilevel Interactions in Economic Development. TIK Working Papers on Innovation Studies No. 20100208.

Greene, W. H. (2008a). *Econometric Anylsis.* Upper Saddle River (NJ): Pearson Education.

Greene, W. H. (2008b). The Econometric Approach of Efficiency Analysis. In H. O. Fried, C. A. K. Lovell, and S. S. Schmidt (Eds.), *The Measurement of Productivity Efficiency and Productivity Growth*, Chapter 2, pp. 92–250. New York: Oxford.

Haining, R. P. (1997). *Spatial Data Analysis in the Social and Environmental Science.* Cambridge: Cambridge University Press.

Haining, R. P. (2003). *Spatial Data Analysis: Theory and Practice.* Cambridge: Cambridge University Press.

Halkos, G. E. and N. Tzeremes (2010). Measuring Regional Economic Efficiency: The Case of Greek Prefectures. *Annals of Regional Science 45*, 603–632.

Hampf, B. and J. J. Krüger (2010). Technical Efficiency of Automobiles - A Nonparametric Approach Incorporating Carbon Dioxide Emissions. Darmstadt Discussion Papers in Economics 198.

Harville, D. A. (1977). Maximum Likelihood Approaches to Variance Component Estimation and to Related Problems. *Journal of the American Statistical Association 72*, 320–38.

Haworth, C. T. and D. W. Rasmussen (1973). Determinants of the Metropolitan Cost of Living Variation. *Southern Economic Journal 40*, 183–192.

Hayfield, T. and J. S. Racine (2008). Nonparametric Econometrics: The np Package. *Journal of Statistical Software 27*, 1–32.

Henderson, J. V. (1974). The Sizes and Types of Cities. *American Economic Review 64*, 640–656.

Henderson, J. V. (1975). Congestion and Optimal City Size. *Journal of Urban Economics 2*, 48–62.

Henderson, J. V. (1986). Efficiency of Resource Usage and City Size. *Journal of Urban Economics 19*, 47–70.

Henderson, J. V. (1997a). Externalities and Industrial Development. *Journal of Urban Economics 42*, 449–470.

Henderson, J. V. (1997b). Medium Size Cities. *Regional Science and Urban Economics 27*, 583–612.

Henderson, J. V. (2003). Marshall's Scale Economies. *Journal of Urban Economics 53*, 1–28.

Henderson, J. V. (2010). Cities and Development. *Journal of Regional Science 50*, 512–540.

Henderson, J. V., A. Kuncoro, and M. Turner (1995). Industrial Development in Cities. *Journal of Political Economy 103*, 1067–1090.

Hendry, D. F. (2006). A Comment on Specification Searches in Spatial Econometrics: The Relevance of Hendry's Methodology. *Regional Science and Urban Economics 36*, 309–312.

Hill, R. C. and L. C. Adkins (2003). *Collinearity*, Chapter 12, pp. 256–278. Malden: Blackwell Publishing.

Hitzschke, S. (2011). The Optimal Size of German Cities: An Efficiency Analysis Perspective. Darmstadt Discussion Papers in Economics 202.

Hox, J. J. (1998). Multilevel Modeling: When and Why. In I. Balderjahn, R. Mathar, and M. Schader (Eds.), *Classification, Data Analysis, and Data Highways*, pp. 147–154. New York: Springer.

Hox, J. J. (2002). *Multilevel Analysis: Techniques and Applications.* Mahwahn (NJ): Erlbaum.

Huber, P. J. (1981). *Robust Statistics.* New York: John Wiley and Sons.

Hubert, L. J., R. G. Golledge, and C. M. Costanzo (1981). Gerneralized Procedure for Evaluating Spatial Autocorrelation. *Geographical Analysis 13*, 224–233.

Ieno, E. N., P. L. Luque, G. J. Pierce, A. F. Zuur, M. B. Santos, N. J. Walker, A. A. Saveliev, and G. Smith (2009). Three-Way Nested Data for Age Determination Techniques Applied to Cetaceans. In A. F. Zuur, E. N. Ieno, N. J. Walker, A. A. Saveliev, and G. M. Smith (Eds.), *Mixed Effects Models and Extensions in Ecology with R*, Chapter 20, pp. 459–492. New York: Springer.

Illy, A., M. Schwartz, C. Hornych, and M. T. W. Rosenfeld (2011). Local Economic Structure and Sectoral Employment Growth in German Cities. *Tijdschrift voor Economische en Sociale Geografie 102*, 582–593.

Jacobs, J. (1970). *The Economy of Cities.* New York: Vintage Books.

Jakoby, H. (2006). Evaluation as Part of the Regional Policy Life Cycle: The Example of North Rhine-Westphalia, Germany. *Regional Studies 40*, 281–284.

Johansson, B. (2014). Generation and Diffusion of Innovation. In M. M. Fischer and P. Nijkamp (Eds.), *Handbook of Regional Science*, Volume 1, Chapter 21, pp. 391–412. Berlin: Springer.

Kahneman, D. (2012). *Thinking, Fast and Slow.* London: Penguin Books.

Kanemoto, Y., T. Ohkawara, and T. Suzuki (1996). Agglomeration Economies and a Test for Optimal City Sizes in Japan. *Journal of the Japanese and International Economies 10*, 379–398.

Karlsson, C. and U. Grasjö (2014). Knowledge Flows, Knowledge Externalities, and Regional Economic Development. In M. M. Fischer and P. Nijkamp (Eds.), *Handbook of Regional Science*, Volume 1, Chapter 22, pp. 413–437. Berlin: Springer.

Kawka, R. (2010). Regionale Preisunterschiede in den Alten und Neuen Ländern. *ifo Dresden berichtet 2*, 5–16.

Kelejian, H. H. and D. P. Robinson (1995). Spatial Correlation: A Suggested Alternative to the Autoregressive Model. In L. Anselin and R. J. Florax (Eds.), *New Directions in Spatial Econometrics: Advances in Spatial Science*, Chapter 3, pp. 75–95. Berlin: Springer.

Kirbach, M. and C. Schmiedeberg (2008). Innovation and Export Performance: Adjustment and Remaining Differences in East and West German Manufacturing. *Economics of Innovation and New Technology 17*, 435–457.

Koop, G., J. Osiewalski, and M. F. J. Steel (1994). Bayesian Efficiency Analysis with a Flexible Form: The AIM Cost Function. *Journal of Business and Economic Statistics 12*, 339–346.

Kosfeld, R. and H.-F. Eckey (2008). Market Access, Regional Price Level and Wage Disparities: The German Case. MAGKS Joint Discussion Paper Series in Economics No. 14-2008.

Kosfeld, R., H.-F. Eckey, and J. Lauridsen (2008). Disparities in Prices and Income across German NUTS 3 Regions. *Applied Economics Quarterly 54*, 123–141.

Laird, N., N. Lange, and D. Stram (1987). Maximum Likelihood Computations with Repeated Measures: Application of the EM Algorithm. *Journal of the American Statistical Association 82*, 97–105.

Laird, N. M. and J. H. Ware (1982). Random-Effects Models for Longitudinal Data. *Biometrics 38*, 963–974.

LeSage, J. P. (2008). An Introduction to Spatial Econometrics. *Revue d'Economie Industrielle 123*, 19–44.

LeSage, J. P. and M. M. Fischer (2008). Spatial Growth Regression: Model Specification, Estimation and Interpretation. *Spatial Economic Analysis 3*, 275–304.

LeSage, J. P. and R. K. Pace (2009). *Introduction to Spatial Econometrics.* Boca Raton (FL): Chapman & Hall/CRC Taylor & Francis Group.

LeSage, J. P. and R. K. Pace (2014). Interpreting Spatial Econometric Models. In M. M. Fischer and P. Nijkamp (Eds.), *Handbook of Regional Science*, Volume 3, Chapter 77, pp. 1535–1552. Berlin: Springer.

Li, Q. and J. Racine (2003). Nonparametric Estimation of Distributions with Categorical and Continuous Data. *Journal of Multivariate Analysis 86*, 266–292.

Lindstrom, M. J. and D. M. Bates (1988). Newton-Raphson and EM Algorithms for Linear Mixed-Effects Models for Repeated-Measures Data. *Journal of the American Statistical Association 83*, 1014–1022.

Lindstrom, M. J. and D. M. Bates (1990). Nonlinear Mixed Effects Models for Repeated Measures Data. *Biometrics 46*, 673–687.

Loader, C. (1999). *Local Regression and Likelihood.* Berlin: Springer.

Loikkanen, H. A. and I. Susiluoto (2004). Cost Efficiency of Finnish Municipalities 1994-2002. An Application of DEA and Tobit Methods. Paper, European Regional Science Association. August 2004.

Loikkanen, H. A. and I. Susiluoto (2006). Cost Efficiency of Finnish Municipalities in Basic Service Provision 1994-2002. Discussion Paper 96, Helsinki Center of Economic Research. February 2006.

Long, J. S. and L. H. Ervin (2000). Using Heteroscedasticity Consistent Standard Errors in the Linear Regression Model. *American Statistician 54*, 217–224.

Lösch, A. (1940). *Die räumliche Ordnung der Wirtschaft: Eine Untersuchung über Standort, Wirtschaftsgebiete und internationalen Handel.* Jena: Gustav Fischer.

Lütkepohl, H. (2007). General-to-Specific or Specific-to-General Modelling? An Opinion on Current Econometric Terminology. *Journal of Econometrics 136*, 319–324.

Maas, C. J. M. and J. J. Hox (2005). Sufficient Sample Sizes for Multilevel Modeling. *Methodology 1*, 86–92.

MacKinnon, J. G. and H. White (1985). Some Heteroskedasticity-Consistent Covariance Matrix Estimators with Improved Finite Sample Properties. *Journal of Econometrics 29*, 305–325.

Malecki, E. J. (2014). The Geography of Innovation. In M. M. Fischer and P. Nijkamp (Eds.), *Handbook of Regional Science*, Volume 1, Chapter 20, pp. 375–389. Berlin: Springer.

Malmquist, S. (1953). Index Numbers and Indifference Surfaces. *Trabajos de Estadística y de Investigación Operativa 4*, 209–242.

Mameli, F., V. Tselios, and A. Rodriguez-Pose (2014). Regional Employment and Unemployment. In M. M. Fischer and P. Nijkamp (Eds.), *Handbook of Regional Science*, Volume 1, Chapter 7, pp. 109–124. Berlin: Springer.

Maronna, R. A., R. D. Martin, and V. J. Yohai (2006). *Robust Statistics: Theory and Methods*. New York: John Wiley and Sons.

Marshall, A. (1890). *Principles of Economics*. London: Macmillan.

Martin, R. (2014). Path Dependence and the Spatial Economy: A Key Concept in Retrospect and Prospect. In M. M. Fischer and P. Nijkamp (Eds.), *Handbook of Regional Science*, Volume 2, Chapter 32, pp. 609–629. Berlin: Springer.

Martin, R. and P. Sunley (2006). Path Dependence and Regional Economic Evolution. *Journal of Economic Geography 6*, 395–437.

Maudos, J., J. M. Pastor, and L. Serrano (2000). Efficiency and Productive Specialization: An Application to the Spanish Regions. *Regional Studies 34*, 829–842.

McCann, P. (2014). Schools of Thought on Economic Geography, Institutions, and Development. In M. M. Fischer and P. Nijkamp (Eds.), *Handbook of Regional Science*, Volume 2, Chapter 28, pp. 527–538. Berlin: Springer.

McFadden, D. (1973). Conditional Logit Analysis of Qualitative Choice Behavior. In P. Zaremka (Ed.), *Frontiers in Econometrics:*, pp. 105–142. Academic Press, New York.

McMahon, W. W. (1991). Geographical Cost of Living Differences: An Update. *Real Estate Economics 19*, 426–450.

McMahon, W. W. and S.-C. Chang (1991). Geographical Cost of Living Differences: Interstate and Intrastate, Update 1991. MacArthur/Spencer Special Series on Illinois School Finance Number 20.

Mellander, C. and R. Florida (2014). The Rise of Skills: Human Capital, the Creative Class, and Regional Development. In M. M. Fischer and P. Nijkamp (Eds.), *Handbook of Regional Science*, Volume 1, Chapter 17, pp. 317–330. Berlin: Springer.

Milligan, G. W. and M. C. Cooper (1985). An Examination of Procedure for Determining the Number of Clusters in a Data Set. *Psychometrika 50*, 159–179.

Moerbeek, M., G. J. P. Breukelen, and M. P. F. Berger (2000). Design Issues for Experiments in Multilevel Populations. *Journal of Educational and Behavioral Statistics 25*, 271–284.

Moomaw, R. L. (1981). Productivity and City Size: A Critique of Evidence. *Quarterly Journal of Economics 96*, 675–688.

Moran, P. A. P. (1950). Notes On Continuous Stochastic Phenomena. *Biometrika 37*, 17–23.

Morrison, P. S. (2014). The Measurement of Regional Growth and Wellbeing. In M. M. Fischer and P. Nijkamp (Eds.), *Handbook of Regional Science*, Volume 1, Chapter 15, pp. 277–289. Berlin: Springer.

Mur, J. and A. Angulo (2009). Model Selection Strategies in a Spatial Setting: Some Additional Results. *Regional Science and Urban Economics 39*, 200–213.

Mur, J. and J. Lauridsen (2007). Outliers and Spatial Dependence in Cross-Sectoral Regression. *Environment and Planning A 39*, 1752–1769.

Nagelkerke, N. J. D. (1991). A Note on a General Definition of the Coefficient of Determinance. *Biometrika 73*, 691–692.

Niebuhr, A. (2006). Market Access and Regional Disparities: New Economic Geography in Europe. *Annals of Regional Science 40*, 313–334.

Nijkamp, P. and K. Kourtit (2013). The "New Urban Europe": Global Challenges and Local Responses in the Urban Century. *European Planning Studies 21*, 291–315.

Nitsch, V. (2005). Zipf Zipped. *Journal of Urban Economics 57*, 86–100.

Noseleit, F. (2013). Entrepreneurship, Structural Change, and Economic Growth. *Journal of Evolutionary Economics 23*, 735–766.

Ord, K. (1975). Estimation Methods for Models of Spatial Interaction. *Journal of the American Statistical Association 70*, 120–126.

Park, W. G. (1995). International R&D Spillovers and OECD Economic Growth. *Economic Inquiry 33*, 571–591.

Pinheiro, J. C. and D. Bates (2000). *Mixed-effects Models in S and S-Plus*. New York: Springer.

Pinheiro, J. C., D. Bates, S. DebRoy, D. Sarkar, and R Development Core Team (2013). *nlme: Linear and Nonlinear Mixed Effects Models*. R package version 3.1-111.

Portela, M. C. A. S. and E. Thanassoulis (2001). Decomposing School and School Type Efficiency. *European Journal of Operational Research 132*, 114–130.

Ramsey, J. B. (1969). Test for Specification Errors in Classical Linear Least Squares Regression Analysis. *Journal of the Royal Statistical Society B 31*, 350–371.

Raudenbush, S. W. and A. S. Bryke (2002). *Hierarchical Linear Models, Application and Data Analysis Methods, Second Edition*. Thousand Oaks: Sage, Advanced Quantitative Techniques in the Social Science Series.

Ray, S. C. and E. Desli (1997). Productivity Growth, Technical Progress, and Efficiency Change in Industrialized Countries: Comment. *American Economic Review 87*, 1033–1039.

Rice, P., A. J. Venables, and E. Patacchini (2006). Spatial Determinants of Productivity: Analysis for the Regions of Great Britain. *Regional Science and Urban Economics 36*, 727–752.

Richardson, H. W. (1972). Optimality in City Size, Systems of Cities and Urban Policy: A Sceptic's View. *Urban Studies 9*, 29–48.

Rigby, D. L. and J. Essletzbichler (2002). Agglomeration Economies and Productivity Differences in US Cities. *Journal of Economic Geography 2*, 407–432.

Romer, P. M. (1986). Increasing Returns and Long Run Growth. *Journal of Political Economy 94*, 1002–1037.

Roos, M. W. M. (2006). Regional Price Levels in Germany. *Applied Economics 38*, 1553–1566.

Rousseeuw, P. J. (1984). Least Median of Squares Regression. *Journal of the American Statistical Association 79*, 871–881.

Roy, A., D. K. Bhaumik, S. Aryal, and R. D. Gibbons (2007). Sample Size Determination for Hierarchical Longitudinal Designs with Different Attrition Rates. *Biometrics 63*, 699–707.

Rozenblat, C. (2012). Opening the Black Box of Agglomeration Economies for Measuring Cities' Competitiveness through International Firm Networks. *Urban Studies 47*, 2841–2865.

Samuelson, P. (1964). Theoretical Notes on Trade Problems. *Review of Economics and Statistics 46*, 142–154.

Saviotti, P. P. and A. Pyka (2004). Economic Development by the Creation of New Sectors. *Journal of Evolutionary Economics 14*, 1–35.

Schirwitz, B. (2007). Eine Chronik klassischer Konjunkturzyklen für Deutschland. *ifo Dresden berichtet 5*(5), 3–11.

Schumpeter, J. A. (1934). *The Theory of Economic Development.* Cambridge (MA): Harvard University Press.

Schumpeter, J. A. (1939). *Business Cycles.* New York: McGraw Hill.

Segal, D. (1976). Are There Returns to Scale in City Size? *Review of Economics and Statistics 58*, 339–350.

Shefer, D. and M. Antonio (2013). Spatial Inequality Between and Within Urban Areas: The Case of Israeli Cities. *European Planning Studies 21*, 373–387.

Shephard, R. W. (1970). *Theory of Cost and Production Function.* Princeton: Princeton University Press.

Silverman, B. W. (1986). *Density Estimation.* London: Chapman and Hall.

Simar, L. and P. W. Wilson (1998). Sensitivity Analysis of Efficiency Scores: How to Bootstrap in Nonparametric Frontier Models. *Management Science 44*, 49–61.

Simar, L. and P. W. Wilson (1999). Estimating and Bootstrapping Malmquist Indices. *European Journal of Operation Research 115*, 459–471.

Simar, L. and P. W. Wilson (2002). Non-Parametric Test of Returns to Scale. *European Journal of Operational Research 139*, 115–132.

Simar, L. and P. W. Wilson (2007). Estimations and Inference in Two-stage, Semi-parametric Models of Production Processes. *Journal of Econometrics 136*, 31–64.

Sinn, H.-W. (2002). Germany's Economic Unification: An Assessment after Ten Years. *Review of International Economics 10*, 113–128.

Srholec, M. (2010). A Multilevel Approach to Geography of Innovation. *Regional Studies 44*, 1208–1220.

Srholec, M. (2011). A Multilevel Analysis of Innovation in Developing Countries. *Industrial and Corporate Change 22*, 1539–1569.

Stamer, M. (1999). *Strukturwandel und wirtschaftliche Entwicklung in Deutschland, den USA und Japan.* Aachen: Shaker.

Statistische Ämter des Bundes und der Länder (2009). Bruttowertschöpfung zu Herstellungspreisen in jeweiligen Preisen; regionale Tiefe: Kreise und krfr. Städte. online. August.

Statistisches Bundesamt (2003). *German Classification of Economic Activities, Edition 2003 (WZ 2003).* Wiesbaden: Statistisches Bundesamt.

Ströhl, G. (1994). Zwischenörtlicher Vergleich des Verbraucherpreisniveaus in 50 Städten. *Wirtschaft und Statistik 6*, 415–434.

Sullivan, L. (2009). *The Sage Glossary of the Social and Behavior Sciences.* Los Angeles: Sage Publications.

Susiluoto, I. and H. A. Loikkanen (2001). Efficiency of Economic Regions in Finland 1988-1999: Application of the DEA Method. In *ERSA conference papers, Congress of the European Regional Science Association.*

Sveikauskas, L. A. (1975). The Productivity of Cities. *Quarterly Journal of Economics 89*, 392–413.

Tabachnick, B. G. and L. S. Fidell (2007). *Using Multivariate Statistics, 5th Edition.* Boston: Pearson International Edition.

Tabuchi, T. (2001). On Interregional Price Differentials. *Japanese Economic Review 52*, 104–115.

Tabuchi, T. and J.-F. Thisse (2006). Regional Specialization, Urban Hierarchy, and Commuting Costs. *International Economic Review 47*, 1295–1317.

Thanassoulis, E., M. C. S. Portela, and O. Despic (2008). Data Envelopment Analysis: The Mathematical Programming Approach to Efficiency Analysis. In H. O. Fried., C. A. K. Lovell, and S. S. Schmidt (Eds.), *The Measurement of Productivity Efficiency and Productivity Growth*, Chapter 3, pp. 251–420. New York: Oxford.

Tiefelsdorf, M., D. A. Griffith, and B. Boots (1999). A Variance-stabilizing Coding Scheme for Spatial Link Matrices. *Environment and Planning A 31*, 165–180.

Tukey, J. W. (1979). Robust Techniques for the User. In R. L. Launer and G. N. Wilkinson (Eds.), *Robustness in statistics: Proceedings, held at Army Research Office, Research Triangle Park, North Carolina, 11.-12.04.1979*, pp. 103–106.

United Nations, Department of Economic and Social Affairs, Population Division (2012). World Urbanization Prospects: The 2011 Revision. online.

van Oort, F. G. and J. G. Lambooy (2014). Cities, Knowledge, and Innovation. In M. M. Fischer and P. Nijkamp (Eds.), *Handbook of Regional Science*, Volume 1, Chapter 22, pp. 475–488. Berlin: Springer.

Ward, J. H. J. (1963). Hierarchical Grouping to Optimize an Objective Function. *Journal of the American Statistical Association 58*, 236–244.

Ware, J. H. (1985). Linear Models for the Analysis of Longitudinal Studies. *The American Statistician 39*, 95–101.

West, B. T., K. B. Welch, and A. T. Galecki (2007). *Linear Mixed Models: A Practical Guide Using Statistical Software*. Boca Raton (FL): Chapman & Hall/CRC Taylor & Francis Group.

Wheelock, D. C. and P. W. Wilson (1999). Technical Progress, Inefficiency, and Productivity Change in U.S. Banking, 1984-1993. *Journal of Money, Credit, and Banking 31*, 212–234.

Wheelock, D. C. and P. W. Wilson (2008). Non-parametric, Unconditional Quantile Estimation for Efficiency Analysis with an Application to Federal Reserve Check Processing Operations. *Journal of Econometrics 145*, 209–225.

Wilson, P. W. (2008). FEAR 1.0: A Software Package for Frontier Efficiency Analysis with R. *Socio-Economic Planning Sciences 42*, 247–254.

Worthington, A. and B. Dollery (2000). Efficiency Measurement in the Local Public Sector: Econometric and Mathematical Programming Frontier Techniques. School of Economics and Finance Discussion Papers and Working Papers Series 078, School of Economics and Finance, Queensland University of Technology.

Yankow, J. J. (2006). Why Do Cities Pay More? An Empirical Examination of Some Competing Theories of Urban Wage Premium. *Journal of Urban Economics 40*, 139–161.

Yezer, A. M. J. and R. S. Goldfarb (1978). An Indirect Test of Efficient City Size. *Journal of Urban Economics 5*, 46–65.

Yohai, V. J. (1987). High Breakdown-Point and High Efficiency Robust Estimates for Regression. *Annals of Statistics 15*, 642–656.

Zeileis, A. (2004). Econometric Computing with HC and HAC Covariance Matrix Estimators. *Journal of Statistical Software 11*, 1–17.

Zipf, G. K. (1949). *Human Behavior and the Principle of Least Effort*. Cambridge: Addison-Wesley Press.

Zuur, A. F., L. B. Gende, E. N. Ieno, N. J. Fernández, M. J. Eguaras, R. Fritz, N. J. Walker, A. A. Saveliev, and G. M. Smith (2009). Mixed Effects Modelling Applied on American Foulbrood Affecting Honey Bees Larvae. In *Mixed Effects Models and Extensions in Ecology with R*, Chapter 19, pp. 447–458. Alain F. Zuur and Elena N. Ieno and Neil J. Walker and Anatoly A. Saveliev and Graham M. Smith.

Zuur, A. F., E. N. Ieno, N. J. Walker, A. A. Saveliev, and G. M. Smith (2009). *Mixed Effects Models and Extensions in Ecology with R*. New York: Springer.

Appendix A

Appendix A.1

Table A1 comprises all cities included in the analysis in chapter 3 together with its average population size and the estimated scale efficiency (SE).

Table A1: Cities included

City	Population	θ_{CRS}	θ_{VRS}	SE	City	Population	θ_{CRS}	θ_{VRS}	SE
Aachen	256779	0.646	0.655	0.985	Kempten	61508	0.695	0.722	0.962
Amberg	44542	0.744	0.807	0.922	Kiel	233994	0.694	0.704	0.986
Ansbach	40559	0.706	0.764	0.924	Koblenz	106833	0.705	0.709	0.994
Aschaffenburg	68680	0.829	0.839	0.988	Krefeld	237905	0.739	0.740	0.998
Augsburg	261109	0.762	0.793	0.961	Landau	42150	0.621	0.737	0.842
Baden-Baden	54398	0.886	0.891	0.994	Landshut	61125	0.724	0.782	0.927
Bamberg	69824	0.706	0.706	0.999	Leipzig	501137	0.561	0.668	0.839
Bayreuth	73992	0.733	0.737	0.994	Leverkusen	161190	0.940	0.957	0.982
Berlin	3395673	0.627	0.984	0.637	Lubeck	211928	0.670	0.737	0.908
Bielefeld	326873	0.690	0.887	0.778	Ludwigshafen	163290	1.000	1.000	1.000
Bochum	386003	0.737	0.746	0.988	Magdeburg	228424	0.565	0.573	0.987
Bonn	312397	0.710	0.816	0.870	Mainz	190854	0.655	0.659	0.995
Bottrop	119671	0.585	0.587	0.997	Mannheim	308159	0.851	0.866	0.983
Brandenburg	74552	0.546	0.586	0.932	Memmingen	41156	0.706	0.758	0.931
Bremen	546149	0.824	0.859	0.959	Monchengladbach	261638	0.671	0.774	0.867
Bremerhaven	117114	0.734	0.738	0.995	Mulheim	170160	0.810	0.841	0.963
Brunswick	245506	0.673	0.681	0.988	Munich	1261289	0.885	0.944	0.938
Chemnitz	247963	0.554	0.594	0.933	Munster	270535	0.761	1.000	0.761
Coburg	41981	0.696	0.738	0.943	Neubrandenburg	68431	0.577	0.590	0.978
Cologne	977860	0.762	0.799	0.954	Neumunster	78459	0.669	0.696	0.962
Cottbus	105822	0.567	0.599	0.946	Neustadt	53800	0.571	0.689	0.829
Darmstadt	140213	0.748	0.754	0.992	Nuremberg	496129	0.729	0.749	0.973
Delmenhorst	75762	0.626	0.683	0.917	Oberhausen	219141	0.650	0.656	0.991
Dessau	80881	0.569	0.589	0.965	Offenbach	119022	0.819	0.851	0.963
Dortmund	588420	0.751	0.839	0.895	Oldenburg	158512	0.676	0.682	0.991
Dresden	492613	0.596	0.606	0.983	Osnabruck	163969	0.694	0.735	0.945
Duisburg	502805	0.753	0.765	0.984	Passau	50593	0.721	0.742	0.972
Dusseldorf	574621	0.978	1.000	0.978	Pforzheim	119061	0.742	0.790	0.939
Eisenach	43867	0.561	0.647	0.867	Pirmasens	43268	0.639	0.737	0.867
Emden	51576	0.721	0.765	0.942	Potsdam	146746	0.575	0.679	0.848
Erfurt	202103	0.563	0.623	0.904	Regensburg	129679	0.763	0.769	0.992
Erlangen	103101	0.984	1.000	0.984	Remscheid	116299	0.705	0.706	0.999
Essen	585589	0.832	1.000	0.832	Rosenheim	60191	0.674	0.693	0.972
Flensburg	85980	0.683	0.706	0.967	Rostock	198910	0.585	0.588	0.995
Frankenthal	47340	0.723	0.847	0.853	Salzgitter	108257	0.781	0.791	0.988
Frankfurt/M	648383	0.978	1.000	0.978	Schwabach	38716	1.000	1.000	1.000
Frankfurt/O	64754	0.543	0.565	0.960	Schweinfurt	54338	0.778	0.778	0.999
Freiburg	214682	0.654	0.659	0.993	Schwerin	97044	0.544	0.554	0.983
Furth	113006	1.000	1.000	1.000	Solingen	163743	0.788	0.794	0.993
Gelsenkirchen	269518	0.734	0.740	0.993	Spires	50440	0.642	0.695	0.924
Gera	104692	0.515	0.523	0.984	Stralsund	58725	0.511	0.582	0.877
Greifswald	52997	0.694	0.697	0.997	Straubing	44590	0.703	0.751	0.936
Hagen	197802	0.719	0.740	0.973	Stuttgart	592028	0.872	0.898	0.971
Halle	237600	0.545	0.564	0.967	Suhl	43245	0.535	0.639	0.837
Hamburg	1743712	0.945	1.000	0.945	Trier	100740	0.622	0.623	0.999
Hamm	184330	0.619	0.621	0.997	Ulm	120398	0.769	0.792	0.971
Heidelberg	143350	0.710	0.739	0.961	Weiden	42695	0.675	0.713	0.947
Heilbronn	121198	0.703	0.707	0.994	Weimar	64421	0.516	0.566	0.912
Herne	171334	0.622	0.632	0.985	Wiesbaden	274002	0.904	0.914	0.990
Hof	48950	0.687	0.688	0.999	Wilhelmshaven	83739	0.693	0.724	0.957
Ingolstadt	120777	0.888	0.895	0.992	Wismar	45448	0.608	0.759	0.800
Jena	102075	0.655	0.660	0.992	Wolfsburg	121647	0.898	0.898	1.000
Kaiserslautern	98735	0.616	0.638	0.966	Worms	81551	0.669	0.723	0.925
Karlsruhe	284487	0.765	0.793	0.965	Wuppertal	360140	0.731	0.742	0.984
Kassel	193914	0.766	0.828	0.924	Wurzburg	133318	0.642	0.649	0.989
Kaufbeuren	42321	0.991	0.992	0.999	Zweibrucken	35275	0.745	0.782	0.952

Appendix A.2

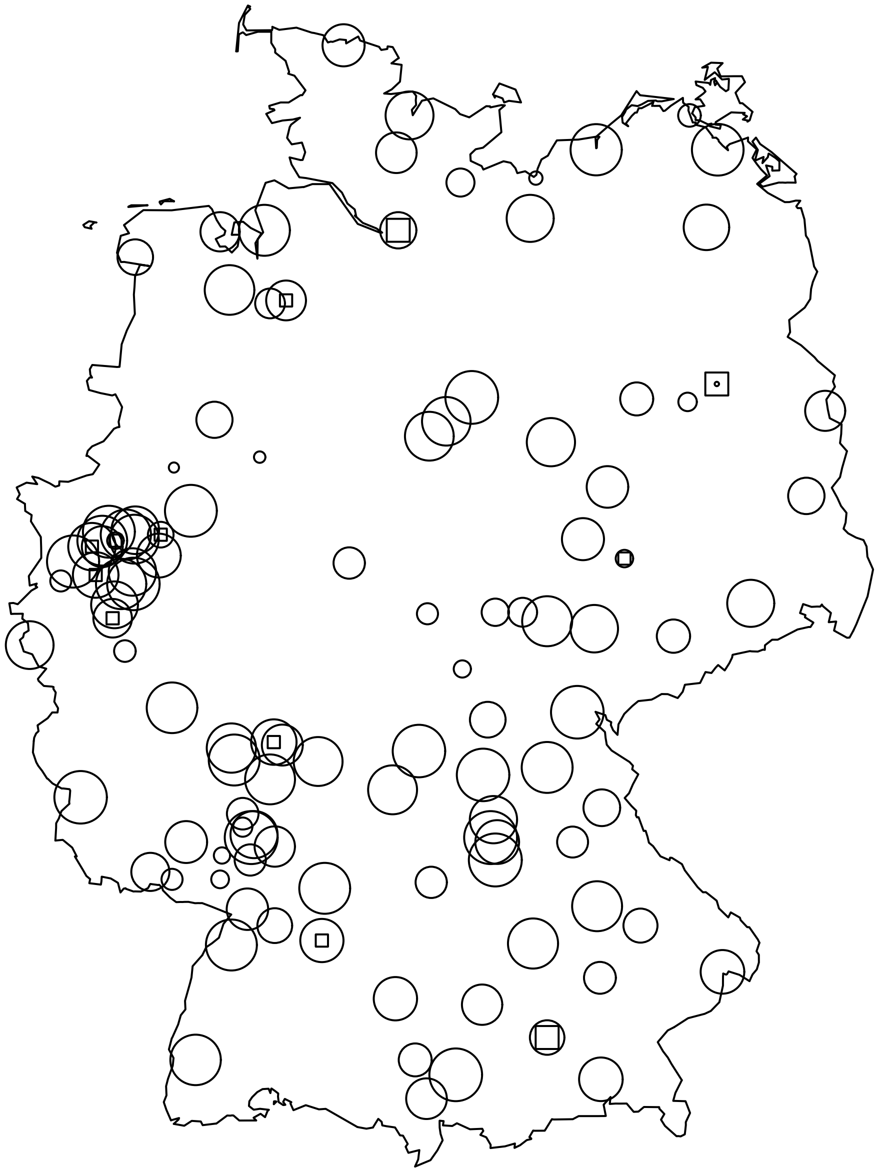

Figure A1: Map of Germany with Scale Efficiency and Major Cities
Note: Circles indicate the scale efficiency and boxes highlight cities with population size of over half a million. Larger circles are used for more efficient cities and larger boxes for larger cities. The relative size of the circles is not equal to the relative efficiency. It visualizes that the largest cities are not the most efficient. The figure was drawn with R using the package `mapdata`.

Appendix A.3

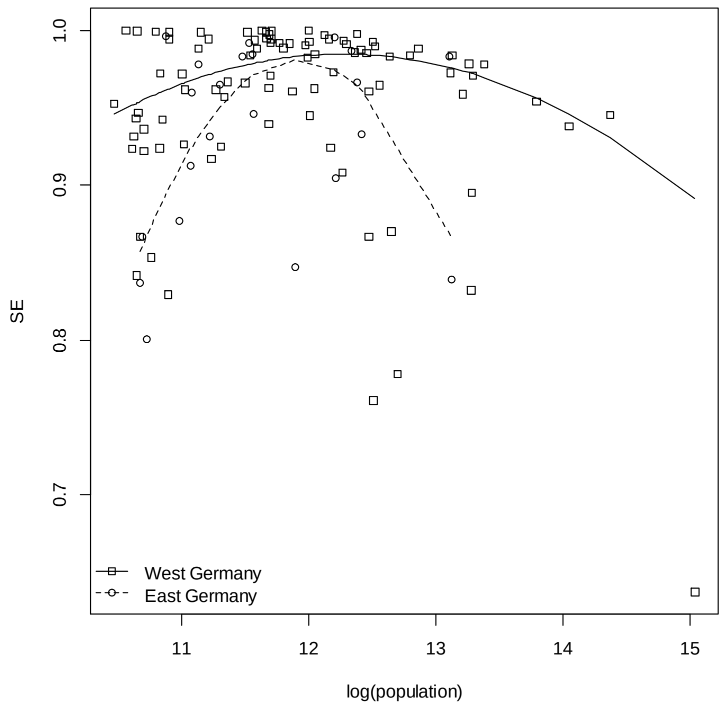

Figure A2: Optimal City Size for East and West Germany
Note: Circles indicate East German cities and boxes indicate West German cities. Fits are robust linear regressions separated for East and West German cities.

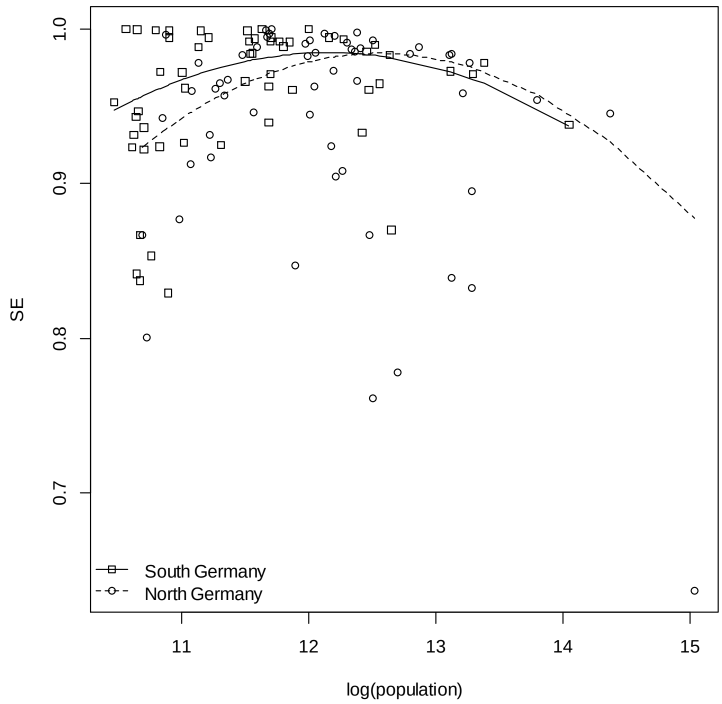

Figure A3: Optimal City Size for North and South Germany
Note: Circles indicate North German cities and boxes indicate South German cities. Fits are robust linear regressions separated for North and South German cities.

Appendix A.4

Table A2: Descriptive statistics for population in thousands in different Germany areas

	Min.	1st Quartile	Median	Mean	3rd Quartile	Max.	s.d.
East	43.2	64.5	99.6	148.3	201.3	501.1	131.2
West	35.3	69.0	125.7	252.1	260.0	3396.0	425.1
North	43.9	94.3	177.8	307.6	269.8	3396.0	500.8
South	35.3	50.1	101.4	155.8	148.3	1261.0	199.9

Appendix B

The following figures show the scatter-plot for managerial efficiency and program efficiency of chapter 4.

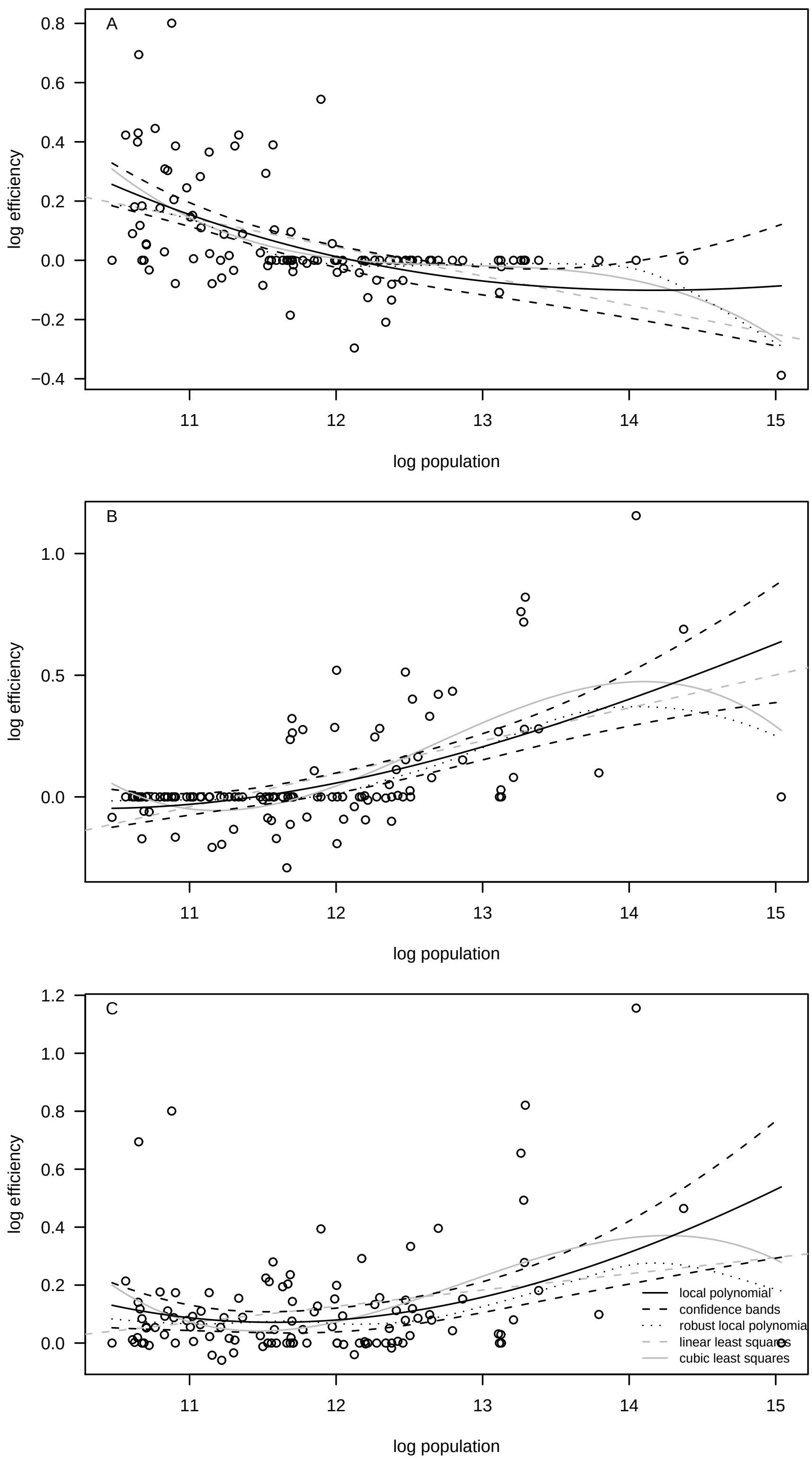

Figure B1: Scatter-Plots for Managerial Efficiency (Terzile Grouping)

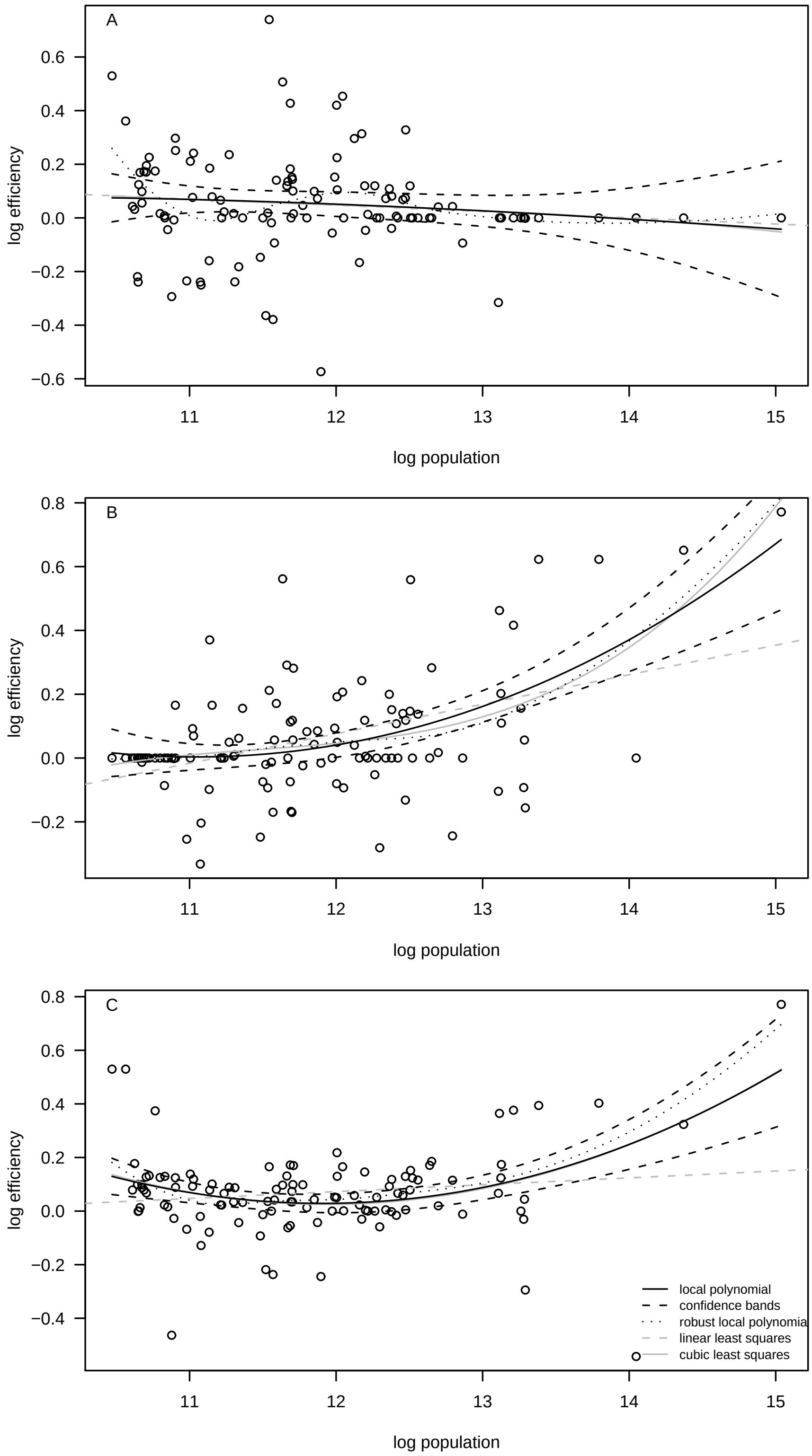

Figure B2: Scatter-Plots for Program Efficiency (Terzile Grouping)

Appendix C

Appendix C.1

Deriving the reduced form is simply found by solving the equations for prices and multiplying by the average quantity. The average quantity represents the standard market basket, which is consumed by the representative household. The reduced form of the simultaneous equations model is derived as follows. At the beginning, price and quantity are brought on the right hand side of the equation to combine them in one vector model. Eq. (68) becomes

$$q - \alpha_1 p = \alpha_2 Y + \alpha_3 X + \alpha_4 D + e_1$$

and Eq. (69) becomes similarly

$$q - \alpha_5^{-1} p = -\alpha_5^{-1} p_0 - \alpha_6 \alpha_5^{-1} Y - \alpha_7 \alpha_5^{-1} X - \alpha_8 \alpha_5^{-1} S - \alpha_5^{-1} e_2,$$

thus the resulting vector model is

$$\begin{aligned} \begin{bmatrix} 1 & -\alpha_1 \\ 1 & -\alpha_5^{-1} \end{bmatrix} \begin{pmatrix} q \\ p \end{pmatrix} &= \begin{bmatrix} \alpha_2 & \alpha_3 \\ -\alpha_6 \alpha_5^{-1} & -\alpha_7 \alpha_5^{-1} \end{bmatrix} \begin{pmatrix} Y \\ X \end{pmatrix} \\ &+ \begin{bmatrix} \alpha_4 & 0 \\ 0 & -\alpha_5^{-1} \end{bmatrix} \begin{pmatrix} D \\ p_0 + \alpha_6 S \end{pmatrix} + \begin{pmatrix} e_1 \\ -\alpha_5^{-1} e_2 \end{pmatrix}. \end{aligned}$$

The vector model has to be transformed to gain the price function

$$\begin{aligned} \begin{pmatrix} q \\ p \end{pmatrix} &= \begin{bmatrix} 1 & -\alpha_1 \\ 1 & -\alpha_5^{-1} \end{bmatrix}^{-1} \begin{bmatrix} \alpha_2 & \alpha_3 \\ -\alpha_6 \alpha_5^{-1} & -\alpha_7 \alpha_5^{-1} \end{bmatrix} \begin{pmatrix} Y \\ X \end{pmatrix} \\ &+ \begin{bmatrix} 1 & -\alpha_1 \\ 1 & -\alpha_5^{-1} \end{bmatrix}^{-1} \begin{bmatrix} \alpha_4 & 0 \\ 0 & -\alpha_5^{-1} \end{bmatrix} \begin{pmatrix} D \\ p_0 + \alpha_6 S \end{pmatrix} \\ &+ \begin{bmatrix} 1 & -\alpha_1 \\ 1 & -\alpha_5^{-1} \end{bmatrix}^{-1} \begin{pmatrix} e_1 \\ -\alpha_5^{-1} e_2 \end{pmatrix}. \end{aligned}$$

The inverted matrix is $\begin{bmatrix} 1 & -\alpha_1 \\ 1 & -\alpha_5^{-1} \end{bmatrix}^{-1} = \frac{1}{-\alpha_5^{-1} + \alpha_1} \begin{bmatrix} -\alpha_5^{-1} & \alpha_1 \\ -1 & 1 \end{bmatrix} = \frac{1}{\alpha_5^{-1} - \alpha_1} \begin{bmatrix} \alpha_5^{-1} & -\alpha_1 \\ 1 & -1 \end{bmatrix}$. Therefore, the vector model becomes

$$\begin{aligned} \begin{pmatrix} q \\ p \end{pmatrix} &= \frac{1}{\alpha_5^{-1} - \alpha_1} \begin{bmatrix} \alpha_5^{-1} & -\alpha_1 \\ 1 & -1 \end{bmatrix} \begin{bmatrix} \alpha_2 & \alpha_3 \\ -\alpha_6 \alpha_5^{-1} & -\alpha_7 \alpha_5^{-1} \end{bmatrix} \begin{pmatrix} Y \\ X \end{pmatrix} \\ &+ \frac{1}{\alpha_5^{-1} - \alpha_1} \begin{bmatrix} \alpha_5^{-1} & -\alpha_1 \\ 1 & -1 \end{bmatrix} \begin{bmatrix} \alpha_4 & 0 \\ 0 & -\alpha_5^{-1} \end{bmatrix} \begin{pmatrix} D \\ p_0 + \alpha_6 S \end{pmatrix} \\ &+ \frac{1}{\alpha_5^{-1} - \alpha_1} \begin{bmatrix} \alpha_5^{-1} & -\alpha_1 \\ 1 & -1 \end{bmatrix} \begin{pmatrix} e_1 \\ -\alpha_5^{-1} e_2 \end{pmatrix}. \end{aligned}$$

I get the equation for the price

$$
\begin{aligned}
p &= \frac{1}{\alpha_5^{-1} - \alpha_1} (\left[\alpha_2 + \alpha_6 \alpha_5^{-1} \right] Y + \left(\alpha_3 + \alpha_7 \alpha_5^{-1} \right) X \\
&\quad + \alpha_4 D + \alpha_5^{-1} p_0 + \alpha_6 \alpha_5^{-1} S + e_1 + \alpha_5^{-1} e_2).
\end{aligned}
$$

The CoL can be derived from the price equation by multiplying with the average quantity $\bar{q}$, which results in the Eq. (70)

$$
CoL = p\bar{q} = \beta_0 + \beta_1 Y + \beta_2 X + \beta_3 D + \beta_4 S + u.
$$

Appendix C.2

The following table presents the correlation matrix for used data.

Table C1: Correlation matrix

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)	(16)
(1) *disp income*	1	-0.05	0.12	0.09	-0.02	0.42	-0.12	0.02	0.48	-0.14	-0.07	-0.09	0.04	-0.09	-0.19	0.54
(2) *stu*	-0.05	1	0.02	-0.03	-0.04	-0.13	0.04	0.04	0	0.03	-0.12	0.14	-0.08	0.08	0.3	-0.06
(3) *o-n-stay*	0.12	0.02	1	0.95	0.85	-0.08	0.1	0.27	0.16	0.06	0.02	-0.01	-0.03	0.28	0.01	0.06
(4) *pop*	0.09	-0.03	0.95	1	0.88	-0.12	0.04	0.35	0.15	0.09	0.05	-0.05	-0.04	0.3	-0.05	0.11
(5) pop^2	-0.02	-0.04	0.85	0.88	1	-0.11	0.03	0.05	0.05	0.01	-0.03	-0.03	-0.06	0.24	-0.06	0.06
(6) *CDU*	0.42	-0.13	-0.08	-0.12	-0.11	1	-0.08	-0.17	0.41	-0.27	-0.27	0.08	0.12	-0.17	-0.32	0.45
(7) *GVA Gini*	-0.12	0.04	0.1	0.04	0.03	-0.08	1	-0.12	-0.08	0.06	0.02	0.1	-0.39	0.02	0.27	-0.15
(8) *BusTax*	0.02	0.04	0.27	0.35	0.05	-0.17	-0.12	1	0.11	0.23	0.1	-0.14	0.02	0.07	0.06	0.08
(9) *dea_JK*	0.48	0	0.16	0.15	0.05	0.41	-0.08	0.11	1	-0.06	-0.02	0.34	-0.03	0.05	-0.21	0.68
(10) *change_CDE*	-0.14	0.03	0.06	0.09	0.01	-0.27	0.06	0.23	-0.06	1	0.23	0.14	-0.02	0.07	0.04	-0.24
(11) *change_GHI*	-0.07	-0.12	0.02	0.05	-0.03	-0.27	0.02	0.1	-0.02	0.23	1	0.26	0.15	0.03	0	-0.03
(12) *change_JK*	-0.09	0.14	-0.01	-0.05	-0.03	0.08	0.1	-0.14	0.34	0.14	0.26	1	0.07	-0.02	0	-0.01
(13) *hcl2*	0.04	-0.08	-0.03	-0.04	-0.06	0.12	-0.39	0.02	-0.03	-0.02	0.15	0.07	1	-0.32	-0.29	0.04
(14) *hcl3*	-0.09	0.08	0.28	0.3	0.24	-0.17	0.02	0.07	0.05	0.07	0.03	-0.02	-0.32	1	-0.2	-0.11
(15) *hcl4*	-0.19	0.3	0.01	-0.05	-0.06	-0.32	0.27	0.06	-0.21	0.04	0	0	-0.29	-0.2	1	-0.29
(16) *west*	0.54	-0.06	0.06	0.11	0.06	0.45	-0.15	0.08	0.68	-0.24	-0.03	-0.01	0.04	-0.11	-0.29	1

Appendix C.3

The following table presents the results of table 5.5 with additional distinction for the federal states.

Table C2: Spatial autoregression results for prices without rents and states

	Malmquist		efficiency		technology	
Intercept	4.224 ***	4.313 ***	4.186 ***	4.273 ***	4.116 ***	4.251 ***
	(0.63)	(0.653)	(0.632)	(0.655)	(0.636)	(0.649)
disp income	0.118 ***	0.12 ***	0.119 ***	0.121 ***	0.122 ***	0.124 ***
	(0.021)	(0.022)	(0.022)	(0.023)	(0.022)	(0.023)
stu	0	0	0	0	0	0
	(0)	(0.001)	(0)	(0.001)	(0)	(0)
pop^2	0.001 ***	0.001 ***	0.001 ***	0.001 ***	0.001 ***	0.001 ***
	(0)	(0)	(0)	(0)	(0)	(0)
CDU	-0.018	-0.019	-0.017	-0.019	-0.013	-0.016
	(0.015)	(0.016)	(0.015)	(0.016)	(0.014)	(0.015)
GVA Gini	0.051 ***	0.047 ***	0.052 ***	0.047 ***	0.046 ***	0.041 ***
	(0.01)	(0.01)	(0.01)	(0.01)	(0.011)	(0.01)
BusTax	0.001	0.002	0.001	0.002	0.003	0.003
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
dea_JK	0.059 ***	0.05 ***	0.06 ***	0.051 ***	0.039 ***	0.033 **
	(0.016)	(0.017)	(0.016)	(0.017)	(0.014)	(0.015)
change_CDE	0.02 *	0.016	0.016	0.012	0.093	0.109
	(0.011)	(0.011)	(0.011)	(0.011)	(0.083)	(0.085)
change_GHI	-0.018	-0.014	-0.012	-0.008	-0.026	-0.019
	(0.021)	(0.022)	(0.024)	(0.025)	(0.068)	(0.07)
change_JK	-0.041 **	-0.035 *	-0.042 **	-0.035 *	-0.261	-0.332
	(0.02)	(0.021)	(0.02)	(0.021)	(0.215)	(0.219)
hcl2	0.015 ***		0.014 ***		0.012 **	
	(0.005)		(0.005)		(0.005)	
hcl3	0.002		0.002		0.002	
	(0.006)		(0.006)		(0.006)	
hcl4	0.014 **		0.014 **		0.01	
	(0.007)		(0.007)		(0.007)	
SMr		0.006		0.006		0.006
		(0.005)		(0.005)		(0.005)
SH	0.06 **	0.07 ***	0.063 **	0.072 ***	0.07 ***	0.074 ***
	(0.024)	(0.025)	(0.024)	(0.025)	(0.024)	(0.025)
NI	0.047 **	0.055 **	0.047 **	0.055 **	0.044 **	0.05 **
	(0.022)	(0.023)	(0.022)	(0.023)	(0.022)	(0.023)
HB	0.019	0.03	0.02	0.031	0.018	0.027
	(0.025)	(0.025)	(0.025)	(0.025)	(0.025)	(0.025)
NW	0.033	0.038 *	0.035 *	0.04 *	0.038 *	0.041 *
	(0.021)	(0.021)	(0.021)	(0.021)	(0.021)	(0.021)
HE	0.084 ***	0.094 ***	0.085 ***	0.095 ***	0.089 ***	0.095 ***
	(0.022)	(0.023)	(0.022)	(0.023)	(0.022)	(0.023)
RP	0.062 ***	0.068 ***	0.064 ***	0.069 ***	0.064 ***	0.065 ***
	(0.023)	(0.024)	(0.023)	(0.024)	(0.024)	(0.025)
BW	0.081 ***	0.086 ***	0.083 ***	0.087 ***	0.084 ***	0.086 ***
	(0.021)	(0.022)	(0.022)	(0.022)	(0.022)	(0.022)
BY	0.061 ***	0.07 ***	0.063 **	0.071 ***	0.06 ***	0.067 ***
	(0.022)	(0.022)	(0.022)	(0.022)	(0.022)	(0.022)
BE	-0.06 **	-0.062 **	-0.057 ***	-0.059 *	-0.052 *	-0.057 *
	(0.029)	(0.03)	(0.029)	(0.03)	(0.029)	(0.03)
BB	0.07 ***	0.075 ***	0.073	0.079 ***	0.071 ***	0.071 ***
	(0.027)	(0.028)	(0.027)	(0.028)	(0.027)	(0.028)
SN	0.021	0.026	0.024	0.029	0.018	0.02
	(0.024)	(0.025)	(0.024)	(0.025)	(0.024)	(0.025)
SA	0.04	0.044	0.042	0.046 *	0.044 *	0.044 *
	(0.026)	(0.027)	(0.026)	(0.027)	(0.026)	(0.026)
TH	0.077 ***	0.079 ***	0.079 ***	0.081 **	0.08 ***	0.079 ***
	(0.024)	(0.026)	(0.025)	(0.026)	(0.025)	(0.026)
lambda	0.301 ***	0.284 ***	0.305 ***	0.289 ***	0.309 ***	0.288 ***
	(0.077)	(0.08)	(0.078)	(0.081)	(0.078)	(0.08)
LM test for residuals	3.0499	2.4279	2.7372	2.2684	5.6662	5.0932
	[0.0807]	[0.1192]	[0.098]	[0.132]	[0.0173]	[0.024]
AIC	-523.94	-518.11	-522.41	-516.81	-520.14	-518.19
R^2	0.7442	0.7442	0.7442	0.7442	0.7442	0.7442
LR-test	290.969	286.054	290.207	285.403	289.068	286.095

Significance codes: '***', '**', '*' significant up to 1%, 5%, and 10%, respectively. Standard errores are below the estimates in parentheses and p-value for the LM test of spatial correlation in the residuals is in square brackets below the test.

Appendix C.4

The following table C3 presents the results for the logarithm of price level without rents on former variables (of the year 2009) which are included as explanatory variable in section 5. All variables are in logarithms. The lags shown in table C3 are the lags of the explained variable presented in the top row of the table. Due to unit root processes of the explained variable the first difference is also taken as explanatory variable although the differences are not applied within the analysis in chapter five. In chapter five the variables are only included for one point in time which is the year before price level data was collected. Prices without rents do not affect the variables in 2009 given their past observations. Prices without rents have only a significantly negative effect on disposable income growth.

Table C3: Results of regression of various variables on price levels without rents

	disp income	Δ*disp income*	*bev*	Δ*bev*	*stu*	Δ*stu*	*BusTax*	Δ*BusTax*
Intercept	1.329 ***	1.488 ***	-0.045	-0.06	-10.986	-35.926	0.27	60.142
	(0.15)	(0.227)	(0.227)	(0.073)	(16.683)	(16.683)	(0.257)	(73.363)
prices without rents	-0.038	-0.185 ***	0.005	0.007	1.418	4.767	-0.029	-7.516
	(0.023)	(0.028)	(0.028)	(0.009)	(2.094)	(2.094)	(0.031)	(9.202)
lag1	1.111 ***	-0.352 **	1.411 ***	0.425 ***	0.353	0.534 ***	0.986 ***	-0.008
	(0.111)	(0.156)	(0.156)	(0.102)	(1.013)	(1.013)	(0.041)	(0.049)
lag2	-0.116	0.078	-0.131	0.236 *	0.684	0.332 **	0.029	0.002
	(0.134)	(0.127)	(0.127)	(0.133)	(1.081)	(1.081)	(0.087)	(0.043)
lag3	-0.131	-0.071	-0.262 *	0.009	0.023	-0.031	-0.026	-0.004
	(0.166)	(0.154)	(0.154)	(0.062)	(1.133)	(1.133)	(0.087)	(0.046)
lag4	0.139	-0.278 **	-0.018	0.007	-0.12	0.139	0.004	0.01
	(0.094)	(0.131)	(0.131)	(0.009)	(1.171)	(1.171)	(0.038)	(0.035)
R^2	0.996	0.425	0.999	0.61	0.958	0.293	0.996	0.008

Significance codes: '***', '**', '*' significant up to 1%, 5%, and 10%, respectively. Standard errors are below the estimates in parentheses.

Appendix C.5

The regression results in Bergstrand (1991) for absolute and relative price level on gross value added per capita explain large fraction of variation with an adjusted R^2 of 87 percent of absolute price level and 85 percent of relative price levels. In my case of cities in one nation the explained variation by gross value added per capita only is much smaller. I use the same regression model for absolute price levels and in logarithm without standardization. The results are presented in table C4.

Table C4: Results of regression of gross value added per capita on price levels

	OLS	SAM(100)	SAM(200)
Intercept	7.514 ***	7.492 ***	7.502 ***
	(0.056)	(0.050)	(0.051)
$\log(GVA/population)$	0.064 ***	0.067 ***	0.064 ***
	(0.016)	(0.014)	(0.014)
R^2	0.1654	0.3408	0.2722
$\ln L$	189.685	202.389	196.844

Significance codes: '***', '**', '*' significant up to 1%, 5%, and 10%, respectively. Adjusted standard errors are below the estimates in parentheses.

The results in table C4 show that gross value added per capita is significantly positive correlated with price level. This is also reported in Bergstrand (1991). The elasticities are much smaller with 0.064 compared to those in Bergstrand (1991) with 0.5. Furthermore, the adjusted R^2 in the OLS model is much smaller with 16.5% compared with 85%. This indicates that gross value added per capita is not a good explanatory variable for price level on nation level as it is for different countries. The estimation of the SAM model with neighbors within a radius of 100 km (SAM(100) in the second column) or 200 km (SAM(200) in the third column) (62.14 miles or 124.28 miles) does not significantly change the estimates and significance as well as the log likelihood. The coefficients of variation are the Nagelkerke R^2, which also takes values within the range of 0 and 1, and are as similar low as in the case of the OLS model.

Appendix D

Appendix D.1

In general and in the formulation of Pinheiro and Bates (2000) a three level model with two levels of random effects is written as

$$y_{ijt} = \boldsymbol{X}_{ijt}\beta_{ijt} + \boldsymbol{Z}_{ij,t}\boldsymbol{b}_{ij} + \boldsymbol{Z}_{ijt}\boldsymbol{b}_{ijt} + e_{ijk}, \tag{73}$$

with $i = 1, ..., N$, $j = 1, ..., n$, and $t = 2, ..., T$, and $\boldsymbol{b}_{ij} \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{\Sigma}_1)$, $\boldsymbol{b}_{ijt} \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{\Sigma}_2)$, $\boldsymbol{e}_{ijk} \sim \mathcal{N}(0, \sigma^2 \boldsymbol{I})$. For simplification the number observations is the same for every level and group so that no observation is missing and it does not vary by lower level groups. In the mixed or random effects literature Eq. (73) is written in vector notation for all i as

$$\boldsymbol{y}_{jt} = \boldsymbol{X}_{jt}\boldsymbol{\beta}_{jt} + \boldsymbol{Z}_{j,t}\boldsymbol{b}_j + \boldsymbol{Z}_{jt}\boldsymbol{b}_{jt} + \boldsymbol{e}_{jt}. \tag{74}$$

Eq. (73) and accordingly Eq. (73) incorporate $\boldsymbol{X}_{jt}$ the regressor matrix for the vector of the p fixed effects β_{jt}, $\boldsymbol{Z}_{j,t}$ the regressor matrix for the random effects $\boldsymbol{b}_j$ of the second level, and $\boldsymbol{Z}_{jt}$ the regressor matrix for the random effect $\boldsymbol{b}_{jt}$ of the third level. The variance-covariance matrices $\boldsymbol{\Sigma}_l$ for $l = 1, 2$ and in each of the two levels of random effects have to be symmetric and positive definite and can be expressed as $\sigma^2\boldsymbol{D}_l$ with σ^2 the variance of the error term and $\boldsymbol{D}_l$ a scaled variance-covariance matrix for the random effects of level l.

The estimation procedure is developed from the simple model with one level of random effects to two levels of random effects and can be extended by further levels of random effects.

For one level of random effects with $l = 1$ the calculation is performed as follows. The general model equation without the third level denoted with t or the second level of random effects is in vector notation

$$y_{ij} = \boldsymbol{X}_{ij}\beta_{ij} + \boldsymbol{Z}_{ij}\boldsymbol{b}_{ij} + e_{ij}, \tag{75}$$

for $i = 1, ..., N$, $j = 1, ..., n$, and $\boldsymbol{X}_{ij}$ the $(N \cdot n \times p)$ regressor matrix for the $(p \times 1)$ vector of fixed effects β_{ij}, $\boldsymbol{Z}_{ij}$ is the $(N \cdot n \times q)$ regressor matrix for the q random effects $\boldsymbol{b}_{ij}$. In notation for all i as vector it follows

$$\boldsymbol{y}_j = \boldsymbol{X}_j\boldsymbol{\beta}_j + \boldsymbol{Z}_j\boldsymbol{b}_j + \boldsymbol{e}_j, \tag{76}$$

for $j = 1, ..., n$. As Lindstrom and Bates (1988) show in general without restriction on the error term structure $\boldsymbol{e}_j \sim N\left(\boldsymbol{0}, \sigma^2\boldsymbol{\Lambda}\right)$ where $\boldsymbol{\Lambda}$ is of size $N \times N$ and does not have to be the identity matrix $\boldsymbol{I}$

$$y_j|\boldsymbol{b}_j \sim N\left(\boldsymbol{X}_j\boldsymbol{\beta}_j + \boldsymbol{Z}_j\boldsymbol{b}_j, \sigma^2\boldsymbol{\Lambda}_j\right), \, j = 1, \, ..., \, n.$$

For all j, it becomes in vector notation

$$\boldsymbol{y}|\boldsymbol{b} \sim N\left(\boldsymbol{X\beta} + \boldsymbol{Zb}, \sigma^2\boldsymbol{\Lambda}\right)$$

with$\boldsymbol{Z} = \mathrm{diag}\left(\boldsymbol{Z}_1, \boldsymbol{Z}_2, ..., \boldsymbol{Z}_n\right)$, $\boldsymbol{\Lambda} = \mathrm{diag}\left(\boldsymbol{\Lambda}_1, \boldsymbol{\Lambda}_2, ..., \boldsymbol{\Lambda}_n\right)$ and $\boldsymbol{b} \sim N\left(0, \sigma^2\boldsymbol{\Sigma}\right)$

$$\boldsymbol{y} \sim N\left(\boldsymbol{X}\beta, \boldsymbol{D}\right), \boldsymbol{D} = \sigma^2\left(\boldsymbol{Z\Sigma Z}' + \boldsymbol{\Lambda}\right) \tag{77}$$

The likelihood function is

$$L\left(\boldsymbol{\beta},\boldsymbol{\theta},\sigma^2|\boldsymbol{y}\right)=\prod_{j=1}^{n}p\left(y_j|\boldsymbol{\beta},\boldsymbol{\theta},\sigma^2\right). \tag{78}$$

In Eq. (78) $\boldsymbol{\theta}$ contains the unique elements of $\boldsymbol{\Sigma}$ and the parameters in $\boldsymbol{\Lambda}$ which are the variance components without exact specification (Harville, 1977; Lindstrom and Bates, 1990). Because $\boldsymbol{b}_j$ and $\boldsymbol{e}_j$ are independent, as Eq. (77) indicates, Eq.(78) results in

$$\begin{aligned}
L\left(\boldsymbol{\beta},\boldsymbol{\theta},\sigma^2|\boldsymbol{y}\right) &= \prod_{j=1}^{n}\int p\left(y_j|\boldsymbol{b}_j,\boldsymbol{\beta},\sigma^2\right)p\left(\boldsymbol{b}_j|\boldsymbol{\theta},\sigma^2\right)d\boldsymbol{b}_j \\
&= \prod_{j=1}^{n}\int \frac{\exp(-\left\|\boldsymbol{y}_j-\boldsymbol{X}_j\boldsymbol{\beta}_j+\mathbf{Z}_j\boldsymbol{b}_j\right\|^2/2\sigma^2)}{\left(2\pi\sigma^2\right)^{N/2}}\frac{\exp\left(-\boldsymbol{b}_j'\boldsymbol{D}^{-1}\boldsymbol{b}_j/2\sigma^2\right)}{\left(2\pi\sigma^2\right)^{q/2}\sqrt{|\boldsymbol{D}|}}d\boldsymbol{b}_j \\
&= \prod_{j=1}^{n}\frac{1}{\sqrt{\left(2\pi\sigma^2\right)^{N/2}}}\int\frac{\exp\left[\frac{-1}{2\sigma^2}\left(\left\|y_j-\boldsymbol{X}_j\boldsymbol{\beta}-\boldsymbol{Z}_j\boldsymbol{b}_j\right\|^2+\boldsymbol{b}_j'\boldsymbol{D}^{-1}\boldsymbol{b}_j\right)\right]}{\left(2\pi\sigma^2\right)^{q/2}\sqrt{|\boldsymbol{D}|}}d\boldsymbol{b}_j \\
&= \prod_{j=1}^{n}\frac{1}{\sqrt{\left(2\pi\sigma^2\right)^{N/2}}}\int\frac{\exp\left[\frac{-1}{2\sigma^2}\left(\left\|y_j-\boldsymbol{X}_j\boldsymbol{\beta}-\boldsymbol{Z}_j\boldsymbol{b}_j\right\|^2-\left\|\Delta\boldsymbol{b}_j\right\|^2\right)\right]}{\left(2\pi\sigma^2\right)^{q/2}\mathsf{abs}\,|\Delta|^{-1}}d\boldsymbol{b}_j \\
&= \prod_{j=1}^{n}\frac{\mathsf{abs}\,|\Delta|}{\sqrt{\left(2\pi\sigma^2\right)^{N/2}}}\int\frac{\exp\left[\frac{-1}{2\sigma^2}\left(\left\|\tilde{\boldsymbol{y}}_j-\tilde{\boldsymbol{X}}_j\boldsymbol{\beta}-\tilde{\boldsymbol{Z}}_j\boldsymbol{b}_j\right\|^2\right)\right]}{\left(2\pi\sigma^2\right)^{q/2}}d\boldsymbol{b}_j,
\end{aligned} \tag{79}$$

with $\tilde{\boldsymbol{y}}_j=\begin{bmatrix}\boldsymbol{y}_j\\ \mathbf{0}\end{bmatrix}$, $\tilde{\boldsymbol{X}}_j=\begin{bmatrix}\boldsymbol{X}_j\\ \mathbf{0}\end{bmatrix}$, $\tilde{\boldsymbol{Z}}_j=\begin{bmatrix}\boldsymbol{Z}_j\\ \boldsymbol{\Delta}\end{bmatrix}$ as pseudo data, where $\boldsymbol{\Delta}$ a relative precision factor as the Cholesky factor of $\boldsymbol{D}^{-1}$, since $\boldsymbol{b}_j'\boldsymbol{D}^{-1}\boldsymbol{b}_j=\left\|\triangle\boldsymbol{b}_j\right\|^2=\left\|\mathbf{0}-\mathbf{0}\boldsymbol{\beta}-\triangle\boldsymbol{b}_j\right\|^2$ and therefore $\boldsymbol{D}^{-1}=\Delta'\Delta$ (Lindstrom and Bates, 1990).

So the exponent is the sum of squared residuals ($\|a\|=\sqrt{a'a}$ as the norm of a matrix). Eq. (79) clearly points out that the maximization of the log-likelihood requires the minimization of the quadratic norm within the exponential function within the integral. This quadratic norm includes the quadratic error terms and is therefore similar to other least squares problems except that the mean of the random effects have to be zero. To solve that least squares problem numerically the orthogonal-triangular decomposition of rectangular matrices is preferred since it provides stable and efficient results by reducing the condition, i.e. the complexity of $\boldsymbol{X}_j$ and $\boldsymbol{Z}_j$. The orthogonal-triangular decomposition uses is the QR-decomposition, with $\tilde{\boldsymbol{Z}}_j=\boldsymbol{Q}_{(j)}\begin{bmatrix}\boldsymbol{R}_{11(j)}\\ \mathbf{0}\end{bmatrix}$, where $\boldsymbol{Q}_{(j)}$ is a $(N+q)\times(N+q)$ orthogonal matrix $\left(Q_{(j)}'=Q_{(j)}^{-1}\right)$ and $\boldsymbol{R}_{11(j)}$ is an upper-triangular $(q\times q)$ matrix. This decomposition can be performed for every real matrix but in the case for positive elements in $\boldsymbol{R}_{11(j)}$ have to be invertible, so $\tilde{\boldsymbol{Z}}_j$ has to have full rank as for OLS regression there must not be any linear dependency structure within the random variables. Also $\tilde{\boldsymbol{X}}_j=\boldsymbol{Q}_{(j)}\begin{bmatrix}\boldsymbol{R}_{10(j)}\\ \boldsymbol{R}_{00(j)}\end{bmatrix}$ and $\tilde{\boldsymbol{y}}_j=\boldsymbol{Q}_{(j)}\begin{bmatrix}\boldsymbol{c}_{1(j)}\\ \boldsymbol{c}_{0(j)}\end{bmatrix}$. Therefore, it is also possible to orthogonal triangular decomposition (QR) of an augmented matrix

$$\begin{bmatrix}\boldsymbol{Z}_j & \boldsymbol{X}_j & \boldsymbol{y}_j\\ \boldsymbol{\Delta} & \mathbf{0} & \mathbf{0}\end{bmatrix}=\begin{pmatrix}\tilde{\boldsymbol{Z}}_j & \tilde{\boldsymbol{X}}_j & \tilde{\boldsymbol{y}}_j\end{pmatrix}=\boldsymbol{Q}_{(j)}\begin{bmatrix}\boldsymbol{R}_{11(j)} & \boldsymbol{R}_{10(j)} & \boldsymbol{c}_{1(j)}\\ \mathbf{0} & \boldsymbol{R}_{00(j)} & \boldsymbol{c}_{0(j)}\end{bmatrix}$$

or

$$\boldsymbol{Q}_{(j)}^{-1}\left(\begin{array}{ccc}\tilde{\boldsymbol{Z}}_j & \tilde{\boldsymbol{X}}_j & \tilde{\boldsymbol{y}}_j\end{array}\right)=\left[\begin{array}{ccc}\boldsymbol{R}_{11(j)} & \boldsymbol{R}_{10(j)} & \boldsymbol{c}_{1(j)}\\ \mathbf{0} & \boldsymbol{R}_{00(j)} & \boldsymbol{c}_{0(j)}\end{array}\right].$$

The exponent in Eq. (79) becomes

$$\begin{aligned}\left\|\tilde{\boldsymbol{y}}_j-\tilde{\boldsymbol{X}}_j\boldsymbol{\beta}-\tilde{\boldsymbol{Z}}_j\boldsymbol{b}_j\right\|^2 &= \left\|\boldsymbol{Q}'_{(j)}\left(\tilde{\boldsymbol{y}}_j-\tilde{\boldsymbol{X}}_j\boldsymbol{\beta}-\tilde{\boldsymbol{Z}}_j\boldsymbol{b}_j\right)\right\|^2\\ &= \left\|\boldsymbol{c}_{1(j)}-\boldsymbol{R}_{10(j)}\boldsymbol{\beta}-\boldsymbol{R}_{11(j)}\boldsymbol{b}_j\right\|^2+\left\|\boldsymbol{c}_{0(j)}-\boldsymbol{R}_{00(j)}\boldsymbol{\beta}\right\|.\end{aligned}$$

Thus the integral in Eq. (79) can be expressed as

$$\exp\left[\frac{\left\|\boldsymbol{c}_{0(j)}-\boldsymbol{R}_{00(j)}\boldsymbol{\beta}\right\|^2}{-2\sigma^2}\right]\int\frac{\exp\left[\frac{-1}{2\pi\sigma^2}\left(\left\|\boldsymbol{c}_{1(j)}-\boldsymbol{R}_{10(j)}\boldsymbol{\beta}-\boldsymbol{R}_{11(j)}\boldsymbol{b}_j\right\|^2\right)\right]}{(2\pi\sigma^2)^{q/2}}d\boldsymbol{b}_j. \tag{80}$$

Note because $\boldsymbol{R}_{11(j)}$ is a non-singular, Bates and Pinheiro construct the following variable $\boldsymbol{\phi}_j=\left(\boldsymbol{c}_{1(j)}-\boldsymbol{R}_{10(j)}\boldsymbol{\beta}-\boldsymbol{R}_{11(j)}\boldsymbol{b}_1\right)/\sigma$ with $d\boldsymbol{\phi}_j=\sigma^{-q}\mathrm{abs}|\boldsymbol{R}_{11(j)}|d\boldsymbol{b}_j$ to easily eliminate the integral. The integral expressed in Eq. (80) is

$$\begin{aligned}\int\frac{\exp\left[\frac{-1}{2\pi\sigma^2}\left(\left\|\boldsymbol{c}_{1(j)}-\boldsymbol{R}_{10(j)}\boldsymbol{\beta}-\boldsymbol{R}_{11(j)}\boldsymbol{b}_j\right\|^2\right)\right]}{(2\pi\sigma^2)^{q/2}}d\boldsymbol{b}_j &= \frac{1}{\mathrm{abs}|\boldsymbol{R}_{11(j)}|}\int\frac{\exp\left(-\left\|\boldsymbol{\phi}_j\right\|^2/2\right)}{(2\pi)^{q/2}}d\boldsymbol{\phi}_j\\ &= \mathrm{abs}|\boldsymbol{R}_{11(j)}|^{-1}\end{aligned}$$

because the integral is over a standard normal distribution, which is unity over the whole range.

And because the determinant of $\boldsymbol{R}_{11(j)}$ is the sum of its diagonal elements since it is an upper-triangular matrix by construction of QR decomposition. So altogether the likelihood function becomes

$$L\left(\boldsymbol{\beta},\boldsymbol{\theta},\sigma^2|\boldsymbol{y}\right)=\prod_{j=1}^{n}\frac{\exp\left[\frac{\left\|\boldsymbol{c}_{0j}-\boldsymbol{R}_{00(j)}\boldsymbol{\beta}\right\|^2}{-2\sigma^2}\right]}{\sqrt{(2\pi\sigma^2)^N|\boldsymbol{D}|}}\mathrm{abs}|\boldsymbol{R}_{11(j)}|^{-1}.$$

A further QR decomposition can be performed by

$$\left[\begin{array}{cc}\boldsymbol{R}_{00(1)} & \boldsymbol{c}_{0(1)}\\ \vdots & \vdots\\ \boldsymbol{R}_{00(M)} & \boldsymbol{c}_{0(M)}\end{array}\right]=\boldsymbol{Q}_0\left[\begin{array}{cc}\boldsymbol{R}_{00} & \boldsymbol{c}_0\\ \mathbf{0} & \boldsymbol{c}_{-1}\end{array}\right]$$

to

$$L\left(\boldsymbol{\beta},\boldsymbol{\theta},\sigma^2|\boldsymbol{y}\right)=\left(2\pi\sigma^2\right)^{-N_M/2}\exp\left(\frac{\|\boldsymbol{c}_{-1}\|^2+\|\boldsymbol{c}_0-\boldsymbol{R}_{00}\boldsymbol{\beta}\|^2}{-2\sigma^2}\right)\prod_{j=1}^{n}\mathrm{abs}\left(\frac{|\boldsymbol{\Delta}|}{|\boldsymbol{R}_{11(j)}|}\right)$$

with $N_n=\sum_{j=1}^{n}N=n\cdot N$ and $1/\sqrt{|\boldsymbol{D}|}=\mathrm{abs}|\boldsymbol{\Delta}|$. The estimate of fixed effects $\boldsymbol{\beta}$ follows from $\|\boldsymbol{c}_0-\boldsymbol{R}_{00}\boldsymbol{\beta}\|^2$ and is

$$\hat{\boldsymbol{\beta}}=\boldsymbol{R}_{00}^{-1}\boldsymbol{c}_0$$

and

$$\boldsymbol{\sigma}^2 = \|\boldsymbol{c}_{-1}\|^2 / N_n.$$

Maximum likelihood estimates are then performed by setting an estimate for $\boldsymbol{\theta}$. The random effects are evaluated by

$$\hat{\boldsymbol{b}}_j(\boldsymbol{\theta}) = \boldsymbol{R}_{11(j)}^{-1}\left(\boldsymbol{c}_{1j} - \boldsymbol{R}_{10(j)}\hat{\boldsymbol{\beta}}(\boldsymbol{\theta})\right).$$

This is the best linear unbiased predictor for the random effects, where $\boldsymbol{\theta} = \hat{\boldsymbol{\theta}}$ as the maximum likelihood estimate.

Lindstrom and Bates (1988; 1990) show the computation for full maximum likelihood and restricted maximum likelihood estimation. Since the maximum likelihood estimation does not account for the loss in degrees of freedom $(N_M - p)$ the estimators are generally downward biased for example if the estimator for the variance component is $\theta_i (N_n - p)/N$ its bias is $\theta_i p / N_n$ (Harville, 1977). The estimation is therefore performed with the restricted maximum likelihood estimation (REML) sometimes also called residual maximum likelihood which accounts for the degrees of freedom but results in incomparable results if the number of parameters differ. The restricted form as Laird and Ware (1982); Ware (1985)

$$L_R\left(\boldsymbol{\theta}, \sigma^2 | \boldsymbol{y}\right) = \int L\left(\boldsymbol{\beta}, \boldsymbol{\theta}, \sigma^2 | \boldsymbol{y}\right) d\boldsymbol{\beta} \tag{81}$$

logarithm

$$\begin{aligned} l_R\left(\boldsymbol{\theta}, \sigma^2 | \boldsymbol{y}\right) &= \log L_R\left(\boldsymbol{\theta}, \sigma^2 | \boldsymbol{y}\right) \\ &= -\frac{N_n - p}{2} \log\left(2\pi\sigma^2\right) - \frac{\|\boldsymbol{c}_{-1}\|^2}{2\sigma^2} - \log \operatorname{abs}|\boldsymbol{R}_{00}| + \sum_{j=1}^{n} \log \operatorname{abs}\left(\frac{|\boldsymbol{\Delta}|}{|\boldsymbol{R}_{11(j)}|}\right). \end{aligned}$$

As the result, the conditional estimate for $\boldsymbol{\beta}$ is

$$\hat{\boldsymbol{\beta}} = \boldsymbol{R}_{00}^{-1}\boldsymbol{c}_0$$

as the same as in the unconditional case but with $\boldsymbol{R}_{00}^{-1}$ different due to different $\boldsymbol{\Delta}$ and σ^2

$$\hat{\sigma}_R^2(\boldsymbol{\theta}) = \|c_{-1}\|^2 / (N_t - p).$$

So the restricted log-likelihood is

$$\begin{aligned} l_R(\boldsymbol{\theta}|\boldsymbol{y}) &= l_R\left(\boldsymbol{\theta}, \hat{\sigma}_{RE}^2(\boldsymbol{\theta})|\boldsymbol{y}\right) \\ &= \text{const} - (N_n - p)\log\|\boldsymbol{c}_{-1}\| - \log \operatorname{abs}|\boldsymbol{R}_{00}| + \sum_{j=1}^{n} \log \operatorname{abs}\left(\frac{|\boldsymbol{\Delta}|}{|\boldsymbol{R}_{11(j)}|}\right). \end{aligned}$$

In both cases the variance of the fixed effect coefficients is

$$\operatorname{Var}\left(\hat{\boldsymbol{\beta}}\right) = \hat{\sigma}^2 \boldsymbol{R}_{00}^{-1}\left(\boldsymbol{R}_{00}^{-1}\right)'.$$

The integral or respectively the sum becomes clear as soon as we rewrite the likelihood function for one level of random effects in Eq. (78) for two levels of random effects namely in my example the city level $j = 1, ..., n$ which is nested within the time level $t = 1, ..., T$, it becomes

$$L\left(\boldsymbol{\beta}, \boldsymbol{\theta}_1, \boldsymbol{\theta}_2, \sigma^2|\boldsymbol{y}\right) = \prod_{t=1}^{T} \int \prod_{j=1}^{n} \left[\int p\left(\boldsymbol{y}_{jt}|\boldsymbol{b}_{jt}, \boldsymbol{b}_{it}, \boldsymbol{\beta}, \sigma^2\right) p\left(\boldsymbol{b}_{jt}|\boldsymbol{\theta}_2, \sigma^2\right) d\boldsymbol{b}_{jt}\right] p\left(\boldsymbol{b}_t|\boldsymbol{\theta}_1, \sigma^2\right) d\boldsymbol{b}_t. \tag{82}$$

Decomposition is constructed similar to the case with one level of random effects

$$\begin{bmatrix} \boldsymbol{Z}_{jt} & \boldsymbol{Z}_{j,t} & \boldsymbol{X}_{jt} & \boldsymbol{y}_{jt} \\ \boldsymbol{\Delta}_2 & \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} \end{bmatrix} = \boldsymbol{Q}_{jt} \begin{bmatrix} \boldsymbol{R}_{22(jt)} & \boldsymbol{R}_{21(jt)} & \boldsymbol{R}_{20(jt)} & \boldsymbol{c}_{2(jt)} \\ \boldsymbol{0} & \boldsymbol{R}_{11(jt)} & \boldsymbol{R}_{10(jt)} & \boldsymbol{c}_{1(jt)} \end{bmatrix}, j = 1, ..., n, t = 1, ..., T$$

decomposition for that

$$\begin{bmatrix} \boldsymbol{R}_{11(1t)} & \boldsymbol{R}_{10(1t)} & \boldsymbol{c}_{1(1t)} \\ \vdots & \vdots & \vdots \\ \boldsymbol{R}_{11(Mt)} & \boldsymbol{R}_{1(Mt)} & \boldsymbol{c}_{1Mt)} \\ \boldsymbol{\Delta}_1 & \boldsymbol{0} & \boldsymbol{0} \end{bmatrix} = Q_{(i)} \begin{bmatrix} \boldsymbol{R}_{11(t)} & \boldsymbol{R}_{10(t)} & \boldsymbol{c}_{1(t)} \\ 0 & \boldsymbol{R}_{00(t)} & \boldsymbol{c}_{0(t)} \end{bmatrix}$$

the profiled log-likelihood becomes

$$\begin{aligned} l_R\left(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2|\boldsymbol{y}\right) &= \log L_R\left(\hat{\boldsymbol{\beta}}_R\left(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2\right), \boldsymbol{\theta}_1, \boldsymbol{\theta}_2, \hat{\sigma}_R^2\left(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2\right)|\boldsymbol{y}\right) \\ &= \text{const} - \left(N_T - p\right)\log\|\boldsymbol{c}_{-1}\| - \log\text{abs}|\boldsymbol{R}_{00}| + \sum_{t=1}^{T}\log\text{abs}\left(\frac{|\boldsymbol{\Delta}_1|}{|\boldsymbol{R}_{11(t)}|}\right) \\ &\quad + \sum_{t=1}^{T}\sum_{j=1}^{n}\log\text{abs}\left(\frac{|\boldsymbol{\Delta}_2|}{|\boldsymbol{R}_{22(jt)}|}\right), \end{aligned}$$

with $N_T = N \cdot n \cdot T$ the total number of observations. Compared to the two level model, the three level model just adds the last addend for the nested higher level.

The solution is straight forward according to one level estimation.

Multilevel models are solved by EM algorithm, which is an iteration of two steps, namely the expectation and maximization (Laird et al., 1987). The data are fitted to the model within the expectation step by estimating the fixed effects, random effects, and the pseudo data ($\tilde{\boldsymbol{y}}_j$, $\tilde{\boldsymbol{X}}_j$, and $\tilde{\boldsymbol{Z}}_j$) to the current values of variance components $\hat{\theta}$. The maximization step fits the parameter θ of the model to the data by maximizing the likelihood to achieve new variance component parameters $\hat{\theta}$ for the expectation step (Laird and Ware, 1982 and Lindstrom and Bates, 1988).

As described in Laird and Ware (1982) and Lindstrom and Bates (1990) it starts by setting an initial value for θ within the maximization-step. The error term depends on those variance components in $\hat{\theta}$ which is straightforward $\boldsymbol{e}_j = \boldsymbol{y}_j - \boldsymbol{X}_j\beta_j\left(\hat{\theta}\right) - \boldsymbol{Z}_j\boldsymbol{b}_j\left(\hat{\theta}\right)$. The expectation-step consists of estimation of the variance components namely for the error terms and the random effects, they basically are presented as in Laird and Ware (1982)

$$\mathsf{E}\left(\sum_{j=1}^{n}\boldsymbol{e}_j^T\boldsymbol{e}_j \mid \boldsymbol{y}_j, \hat{\theta}\right) = \sum_{j=1}^{n}\boldsymbol{e}_j^T\left(\hat{\theta}\right)\boldsymbol{e}_j\left(\hat{\theta}\right) + \mathsf{tr}\,\mathsf{var}\left(\boldsymbol{e}_j \mid \boldsymbol{y}_j, \hat{\theta}\right) \tag{83}$$

and

$$\mathsf{E}\left(\sum_{j=1}^{n} \boldsymbol{b}_j \boldsymbol{b}_j^T \mid \boldsymbol{y}_j, \hat{\theta}\right) = \sum_{j=1}^{n} \boldsymbol{b}_j \left(\hat{\theta}\right) \boldsymbol{b}_j^T \left(\hat{\theta}\right) + \mathsf{var}\left(\boldsymbol{b}_j \mid \boldsymbol{y}_j, \hat{\theta}\right). \tag{84}$$

The maximization steps then use the log-likelihood function depending on whether estimating by maximum likelihood or restricted maximum likelihood as presented above or in Lindstrom and Bates (1990) for both estimation in general and with computational improvements in Laird et al. (1987) as implemented in current software to achieve faster convergence.

Appendix D.2

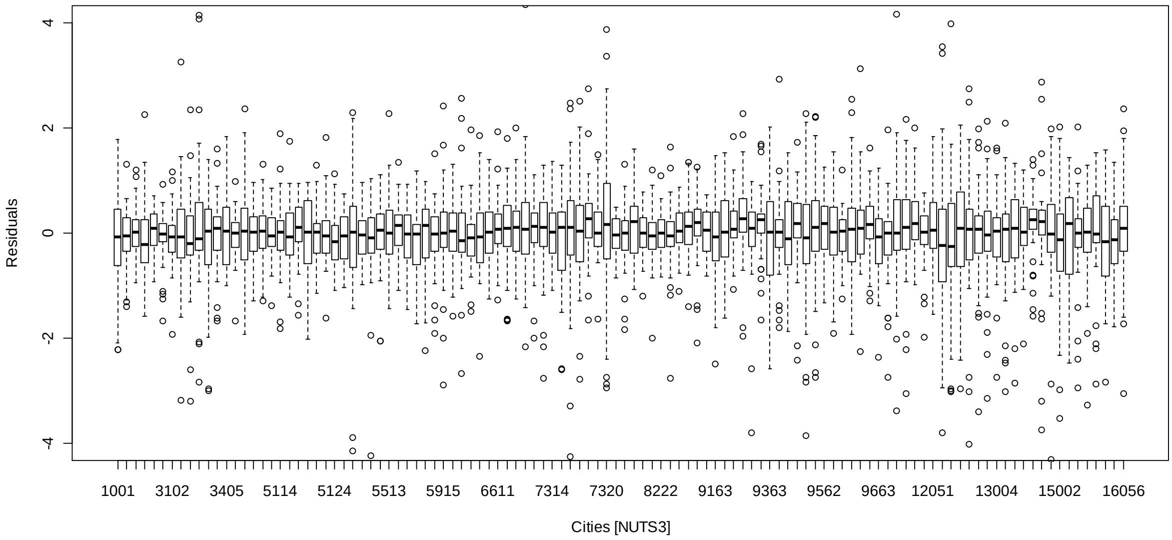

Figure D1: Residual Plot for Employment Growth at City Level

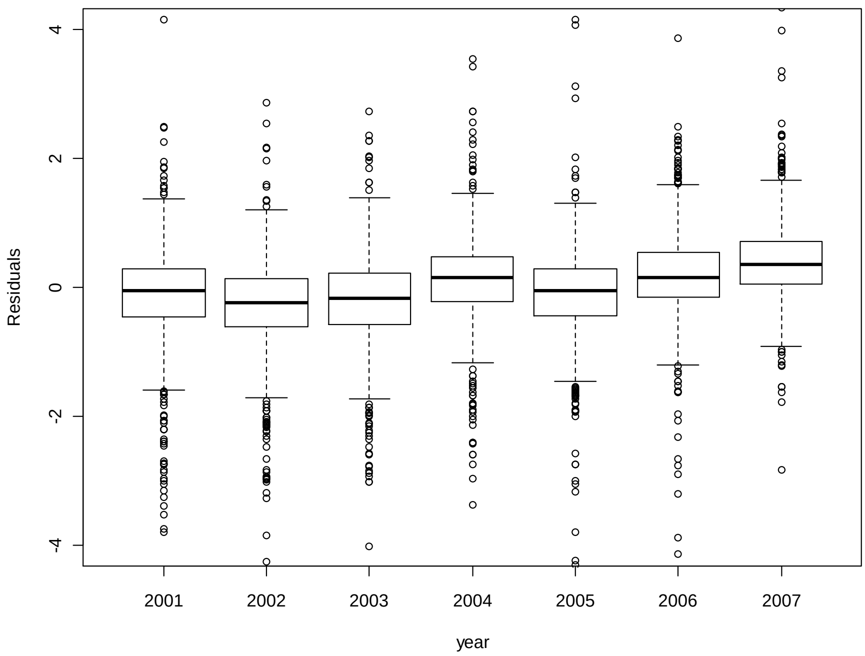

Figure D2: Residual Plot for Employment Growth at Time Level

Affirmation

I hereby declare that the dissertation entitled

"**Urban Efficiency and Sectoral Structure: Empirical Results for German Cities**"

is my own work. I have only used the sources indicated and have not made unauthorized use of services of a third party. Where the work of others has been quoted or reproduced, the source is always given. I have not presented this thesis or parts thereof to an university as part of an examination or degree.

....	....
Place, Date	Signature

Wissenschaftlicher Werdegang

01.04.2009 – 12.02.2014	Promotion an der Technischen Universität Darmstadt
18.02.2008 – 18.07.2008	Postgraduelles Studium Ökonomie an der University of Queensland in Brisbane, Australien Abschluss: "Graduate Certificate in Economics"
01.10.2002 – 31.10.2007	Studium der Volkswirtschaftslehre an der Friedrich-Schiller-Universität Jena mit Schwerpunkt in Wirtschaftsstatistik Abschluss: "Diplom-Volkswirt"
01.08.1994 – 30.09.2002	Goethe-Schule Ilmenau Staatliches Gymnasium in Ilmenau Abschluss: Abitur