Darmstadt
Discussion Papers
in ECONOMICS

TECHNISCHE UNIVERSITÄT DARMSTADT

The Optimal Size of German Cities An Efficiency Analysis Perspective

Stephan Hitzschke

Nr. 202

Arbeitspapiere der
Volkswirtschaftlichen Fachgebiete der TU Darmstadt

ISSN: 1438-2733

The Optimal Size of German Cities
An Efficiency Analysis Perspective

June 2011

by

Stephan Hitzschke

Darmstadt University of Technology
Department of Law and Economics
Residenzschloss, Marktplatz 15, D-64283 Darmstadt, Germany
tel.: +49 6151 16 70371, fax: +49 6151 16 3897
e-mail: hitzschke@vwl.tu-darmstadt.de

Abstract

This work investigates the interrelation between production efficiency and population size of German cities. 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 two-stage process. First, the efficiency in terms of scale efficiency is measured using nonparametric methods. The second stage investigates the relation of scale efficiency and populations size. It turns out that the optimal city size in Germany is about 220,000 inhabitants, which is almost the mean size of all German cities involved. Although there are regional differences, optimal city size remains stable as the mean size.

JEL classification: D24, O47, R12, R15

Keywords: Efficiency analysis, optimal city size, population size, German cities

1 Introduction

There are many reasons that encourage people and firms settling in a city. On the one hand, cities embody forces making life more comfortable like closeness to other people, jobs, recreational and shopping facilities or institutions necessary for life in modern economies. That closeness helps in saving time every day and thereby increases leisure time as well as complement utility for a person within a city and helps direct and indirect increasing the productivity for firms in that city. These forces are mainly acknowledged as agglomeration externalities, which depend on size of the specific urban agglomeration. Those agglomeration externalities are often referred to as Marshall-Arrow-Romer-externalities (Marshall (1890), Arrow (1962), and Romer (1986)). On the other hand, there are signs for an urban overload as city size increases. These pros and cons for urbanization are widely discussed and often referred to as average urban benefits regarding the agglomeration externalities on the one hand and average location costs on the other hand (Capello and Camagni, 2000, pp. 1484ff .). Forces for agglomeration externalities are for example education facilities, health services, skilled jobs, infrastructure, and social contacts which are supported by close relationships within a city. The productivity of individuals and firm are effected because closeness to related people and firms helps saving time and increases the flow of knowledge and ideas or at least the possibility for it. Also located education facilities improve the knowledge of the inhabitants and employees in firms. In addition an often mentioned aspect is the larger job market within a city which helps finding a job and thus participating in the production process as well as knowledge transfer from one firm to another. On the opposite, there are forces indicating that too many inhabitants on a certain area is producing negative externalities as well as costs of urbanization. These forces are recognized for instance by pollution, intensive use of energy, noise caused for example by traffic, high urban rents, as well as long and time intensive duty strokes. They effect the productivity because too many inhabitants within a city produce negative externalities like traffic jam and noise which decrease the productivity by higher transportation costs and higher rents or social friction in the labor market. Of course these forces occur at specific levels of population or population densities, respectively. Although these effects may also rise in non-urban areas they are mainly connected to urbanization effects.

Thus a growing population in cities is leading to increasing agglomeration effects, which are later neutralized by negative effects of overpopulation. These considerations obviously should lead to an optimal city size. Contrariwise, it is observable that especially the largest cities in most industrialized countries are continuously growing, which stands in contrast to an optimal city size, if this is below the size of the largest city. Therefore, the research questions that are trying to answer in this work are the following. Is there an optimal city size in Germany? If that is the case, is the optimal city size in the range of the observed cities or does it predict further growing cities. Does the optimal city size depend on the region where the cities are located within Germany or is it the same for all cities?

The purpose of this work is to investigate the relationship between efficiency and population size of German cities in a static setting. The efficiency in this context is the scale efficiency, which takes care of the specific size of the particular decision making unit. Therefore, the approach employed in this work is a two stage process. Firstly, the efficiency is measured in terms of scale efficiency, which involves estimating the efficiency of each city 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 of the efficiency and the populations size of the cities. It results, 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, optimal city size remains stable as the mean of the underlying cities.

The work is organized as follows. The following Section 2 presents the related literature. Section 3 gives a brief overview on the applied methods and Section 4 describes the data used in the estimations. The empirical results are given and discussed in Section 5. At the end of this work conclusions are drawn in Section 6.

2 Literature Review

There are two separated branches in the literature related to this work. One is concerned with optimal city size and the other investigates spatial efficiency or productivity. A brief review of the received literatures is given in the following. Another part of the literature concerning the optimal city size focuses on the size in terms of pure area size, see for example Henderson (1975). Because this work deals with size in terms of population of the city, that part of the literature is not considered further.

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

Early insights of 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 point out conditions where the model does not work for, like 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. Therefore, they derived rules for checking whether population size is optimal. For a single city, Arnott (1979) shows that for the 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, he works out a generalized version of the theorem, which allows for multiple outputs and multiple factors as well as heterogeneous individuals. This version states that in an optimal 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 turns out, 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, thus Arnott discusses the analysis of Kanemoto et al. (1996), who firstly used the idea of Henry George in their investigation of optimal city size. They 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 Tokyo is not too big thus 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 thus it is still questionable whether an optimal city size in the terms of Henry George Theorem could be established.

Sveikauskas (1975) and Segal (1976) show that Hicks-neutral productivity increases by city size. The critique by Moomaw (1981) makes clear that Sveikauskas estimates are biased upward by omitting capital intensity or capital as explanatory variables, respectively. Nonetheless, Sveikauskas (1975) estimates significantly positive linear correlation between city size and productivity for almost all manufacturing industries while he also controlled for education and regional differences. Since he just investigates a linear relationship he 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 whole citizenship, respectively. Segal (1976) on the other hand finds scale effects in cities by estimating production functions of the 58 largest US cities. Using ordinary least squares estimation he finds constant returns to scale for production output and labor productivity but also positive and significantly effects of city size. Thus metropolitan areas with 2 million and more inhabitants have significantly higher labor productivity compared with smaller metropolitan areas. Since Segal only investigates the largest cities the only viable result is that, if a city is already large it is optimal to grow further and therefore no optimal size exists within the range of observable city sizes.

For the US cities Yezer and Goldfarb (1978) estimate that the optimal city size is in the range of 1.5 to 2.5 million. They investigate the wage changes by region, occupation, and population size and compare these with changes in firm efficiency related with changes in city size. The first effect reflects the household costs. While 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 are the type of urban function 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 squared kilometers). With respect to average location benefits they estimate an inverted U-shaped curve with maximum at 361,000 inhabitants. Concerning urban overload by investigating average location costs they get an U-shaped curve with minimum at 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 caused by the somewhat questionable measurement that makes them incomparable.

As Alonso (1971) points out, the minimum of the average urban costs are not of interest but the point where marginal costs equals marginal product for cities. At this point the level where average products minus average urban costs are 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 it turns out that the highest excess of average product over average public cost is for the population size class of 200,000 and bigger for Germany in 1964, which is not further disaggregated. A brief literature overview about optimal city sizes developed by minimizing urban public costs is given by Richardson (1972). He concludes that there are many mostly philosophic approaches in the analysis without concerning 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 a lot of literature about non-parametric estimation of efficiency and productivity. Because this work deals with cities as aggregate of many firms and households, a brief review of literature dealing with efficiency analysis for spatial decision making units is given.

A good overview is given by Worthington and Dollery (2000), who also include efficiency analyses for firms and specific industries. Closely related is the work of Charnes et al. (1989), who employ data envelopment analysis (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), Loikkanen and Susiluoto (2004), and Loikkanen and Susiluoto (2006) investigate Finish 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 for 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 explaining 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 of it. It turns out that the most efficient areas are not the most densely populated, although it is not an analysis of cities. Major changes in the industry structure and institutional setups as well as EU regularities in Greece cause some doubts regarding the results, because he does not control for that changes.

Altogether, there is a lot 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 and thus leads to continuous increasing efficiency by city size. The aim of this work is to merge both approaches, which are efficiency analysis and the investigation of optimal city size, and apply them to data for German cities for a measure of efficient city size.

3 Theory

The investigation in this work is 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. Those setups incorporate quadratic, cubic, and non-parametric models in ordinary least squares, robust linear fit, and least median of squares estimations.

The efficiency of cities is measured by data envelopment analysis (DEA), which was developed by Charnes et al. (1978). This approach enables the construction and analysis of efficiency of general decision making units, which are cities in this work 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 this work 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 constant returns to scale. 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 a city i. These are defined as:

$$\theta_i\left(\boldsymbol{x}_i, \boldsymbol{y}_i\right) = \max\left\{\theta \mid \left(\boldsymbol{y}_i \cdot \theta\right) \in P\left(\boldsymbol{x}_i\right)\right\}, \tag{1}$$

and

$$\delta_i\left(\boldsymbol{x}_i, \boldsymbol{y}_i\right) = \min\left\{\delta \mid \left(\boldsymbol{y}_i / \delta\right) \in P\left(\boldsymbol{x}_i\right)\right\}, \tag{2}$$

where $\boldsymbol{x_i}$ and $\boldsymbol{y_i}$ are the (12×1) and (6×1) vectors for inputs and outputs for city i, since there are 6 sectors in tha data set with one output and two input factors in each sector (see Section 4). $P\left(\boldsymbol{x_i}\right)$ is the output set, which describes the production functions. In Eq. (1) the θ is the distance for which θ_i is the maximum value, and in Eq. (2) δ is one possible distance functions for which δ_i is the minimum, respectively. Eq. (1) is the representation equivalent to Shephard (1970, p. 207) with output orientation, which is used in the analysis of this work. Contrarily, Eq. (2) is the equivalent to Eq. (1) 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. The envelopment form for constant returns to scale is

$$\begin{aligned} &\min_{\theta,\boldsymbol{\lambda}} \theta_{CRS,i}, && \\ &st \qquad -\boldsymbol{y}_i + \boldsymbol{Y\lambda} &\geq& \ \mathbf{0} \\ &\qquad\qquad \theta\boldsymbol{x}_i - \boldsymbol{X\lambda} &\geq& \ \mathbf{0} \\ &\qquad\qquad \boldsymbol{\lambda} &\geq& \ \mathbf{0}, \end{aligned} \tag{3}$$

where $\theta_{CRS,i}$ is the efficiency score for a particular city i, $\boldsymbol{Y}$ is a $6 \times N$ matrix containing all 6 outputs in the N cities, $\boldsymbol{\lambda}$ is a $N \times 1$ vector of weights, and $\boldsymbol{X}$ is a $12 \times N$ matrix for the 12 inputs in the N cities.

For variable returns to scale one constraint is added to Eq. (3). That additional constraint is $\mathbf{1}'\boldsymbol{\lambda} = 1$ and leads to the linear program in Eq. (4)

$$\begin{aligned} &\min_{\theta,\boldsymbol{\lambda}} \theta_{VRS,i}, && \\ &st \qquad -\boldsymbol{y}_i + \boldsymbol{Y\lambda} &\geq& \ \mathbf{0} \\ &\qquad\qquad \theta\boldsymbol{x}_i - \boldsymbol{X\lambda} &\geq& \ \mathbf{0} \\ &\qquad\qquad \mathbf{1}'\boldsymbol{\lambda} &=& \ 1 \\ &\qquad\qquad \lambda &\geq& \ \mathbf{0}. \end{aligned} \tag{4}$$

Solving all models for all cities results in one estimate for the technical efficiency for CRS and one for VRS for every city. Based on these results for technical efficiencies, scale efficiency SE_i in city i is calculated as the division of the technical efficiency for CRS divided by the technical efficiency for VRS

$$SE_i = \frac{\theta_{CRS,i}}{\theta_{VRS,i}}. \tag{5}$$

Because the technical efficiency for CRS is smaller or equal to the technical efficiency for VRS, the measurement for scale efficiency is always in the range between zero and one, with one for scale efficient and smaller than one for scale inefficient city. The detailed calculated results are listed in table 5 in the appendix. Notice, the measurements for technical efficiency are in terms of Shephard (1970) with output orientation and so they are smaller or equal to one, which represents the proportion of efficiency. The measurement for scale efficiency gives the percentage of inefficiency 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 scale efficiency coefficient less than one do not have the efficient size and are either too small or too large . Although, it should be considered that scale efficiency does not imply that the city or the sectors within the city are technical efficient by constant or variable returns to scale. This can be seen in table 5 in the appendix, which points out that Wolfsburg has the value of one for scale efficiency but the same technical inefficiency for constant and variable returns to scale. All other scale efficient cities are also technical efficient for constant and variable returns to scale. These scale efficient cities could be exclusively used to determine the optimal or efficient city size. But since it is a range of city sizes and to some extant measurement errors are present, these intervall is reduced to one solely measure for optimal city size by a linear regression.

To test whether population size has a relation with 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, which is commonly known as Zipf's-Law which described in Zipf (1949). Zipf's -Law is also called the rank-size rule which is described for instance in Richardson (1972) or Nitsch (2005) and 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 proper to use logarithms of population in the specifications to avoid that the biggest cities leverage the estimates caused by the exponential distribution. The model, which is tested, is in specification I

$$SE_i = \beta_0 + \beta_1 \log(population_i) + \beta_2 \left(\log(population_i)\right)^2 + u_i, \tag{6}$$

with u_i the residuals in city i. The empirical results for Eq. (6) should result in an inverted U-shaped functional form with one maximum point. Thus β_1 should be significantly positive and β_2 significantly negative. To check the correctness of the specification of the quadratic model of specification I a cubic function is also estimated. This cubic model is represented by specification II in form

$$SE_i = \beta_0 + \beta_1 \log(population_i) + \beta_2 \left(\log(population_i)\right)^2 + \beta_3 \left(\log(population_i)\right)^3 + u_i, \tag{7}$$

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 has a significant effect on scale efficiency.

4 Data

In this analysis a data set for 112 NUTS3-districts that are classified as cities (so called "kreisfreie Städte" or "Stadtkreis" in Germany)[1] is used. 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[2] ("Statistische Ämter des Bundes und der Länder") and the INKAR database of the Federal Agency of Building and Urban

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

[2] The dabase is available on internet by https://www.regionalstatistik.de/genesis/online/logon (last check on 30th May 2011).

Development[3] ("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 an one-digit industry specification (WZ 2003 of the Federal Statistical Office of Germany (Federal Statistical Office (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

Because of the minor importance of the agriculture, forestry and fishing sector in German cities the sector AB has been omitted. As Moomaw (1981) critized that the disregard of the capital stock in Sveikauskas (1975) leads to biased estimates, the capital stock has to be added. The capital stock for each city and the wide manufacturing sector is computed with the perpetual-inventory-method (Park, 1995) supposing capital stocks $cap_{i,t}$ develop as

$$cap_{i,t} = (1-d)cap_{i,t-1} + inv_{i,t}, \tag{8}$$

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

$$cap_{i,0} = inv_{i,0} \cdot \frac{1 - g_{inv,i}}{d + g_{inv,i}}. \tag{9}$$

Eq. (9) 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 is also taken from the regional database of the Statistical Offices in Germany for the time period 1995 until 2007 in real units and is 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 quiet 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 is 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 the German Reunification and the associated structural changes in the industry. Therefore, the average growth rates for all cities in East-Germany was applied which is above minus 10 percent and thus the denominator in Eq. (9) is positive which 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 furthermore depreciated and the capital stock is largely driven by last investments. Therefore, only the average of the last 5 years is used in further estimations, which implies that the first 9 years before the year 2004 for capital are out of consideration. Thus the starting capital stock depreciated away to 40 percent in 2004 and thereby reduces the involved uncertainty in that input factor. The captial 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 captial intensity of the wide manucaturing

[3] The database is available on CR-ROM upon request to the Federal Agency of Building and Urban Development at http://www.bbsr.bund.de/cln_032/nn_187652/BBSR/EN/Home/homepage__node.html?__nnn=true (last check on 30th May 2011).

sector compared to the other industry sectors in whole Germany, which is given by the OECD Database for Structural Analysis (STAN[4]) .

Population figures are also taken from the regional data base of the Statistical Offices in Germany. A person is only counted for a city if it has the principal residence within this city. So the figure does not account for people with secondary residence to avoid double countings, although many people have a secondary residence in a city and are part of those productive employees. Nonetheless, the use of the population figures for the number of inhabitant within a city is permitted, because people that 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 5 are used as the arithmetic average over the years 2004 until 2008. Descriptive statistics are given in table 1 with value added and capital stock in Euro [€] and labor force and population in thousand [T].

Table 1: Descriptive statistics

variable	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.	SD
value added [€]	919,000	1,875,000	3,975,000	7,740,000	7,372,000	73,390,000	12,470,104
capital [€]	100,400	4,535,000	10,140,000	26,410,000	25,610,000	282,000,000	46,803,062
labor force [T]	18.42	43.33	76.12	139.70	137.30	1,551.00	209.38
population [T]	35.28	64.67	120.60	231.70	239.80	3,396.00	387.15

Table 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. That distribution results in a median of each of these variables that is much lower than the respective mean. The median is almost of 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 an heteroscedastic distribution of the data, with a decreasing variance in city size caused by many different small cities. The skewness of the cities does not effect the efficiency analysis, which relies on relative measurements. For further analysis the data has to be transformed to become more narrow. That is done by taking the logarithm of population. Furthermore, all input variables as well as value added as output are non-negative as required in DEA.

5 Empirical Results

In this section the estimation results for population size on scale efficiency are presented and discussed. Furthermore, an illustration (figure 2) for scale efficiency is given in the appendix, where the largest cities with over half a million population size are pointed out by boxes. That 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. Also there is no specific region in Germany that only locates cities with low efficiency score and thus it is obvious that the efficiency scores are not asymmetric distributed over the German regions.

The estimations are performed by several methods. Since there are just four observations with the value of one for scale efficiency (compare table 5 in the appendix) the estimations can be performed by normal regressions and do not have to be performed by Tobit or Logit regressions for trancated observations. These methods involve ordinary least squares (OLS), least median of squares (LMS), and robust fit MM-estimations. The OLS estimation minimizes the sum of squared residuls

$$\min_{\beta} \sum_{i=1}^{N} (u_i)^2 = \min_{\beta} \sum_{i=1}^{N} (y_i - \boldsymbol{x}_i \beta)^2 \tag{10}$$

[4]The database is available on the internet by http://stats.oecd.org/Index.aspx?DatasetCode=STAN08BIS&lang=en (last check on 30th May 2011).

with the solution $\widehat{\beta}_{OLS} = \left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1}\boldsymbol{X}'\boldsymbol{y}$ and $\boldsymbol{x}_i$the i-th row within the explanatory matrix $\boldsymbol{X}$ in the city i. The OLS-method is the best linear unbiased estimator, if some assumptions hold, like homoscedasticity of the residuals.

By considering the distribution of city size with many small cities and only few largest cities it can be seen that the variance is declining and thus results in a heteroscedasticity problem. That problem leads to inefficient OLS estimations and wrong standard errors. Therefore, standard errors for the OLS estimates are calculated with a heteroscedasticity consistent covariance matrix, which is estimated by

$$\left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1}\boldsymbol{X}'\Omega\boldsymbol{X}\left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1} \tag{11}$$

with

$$\Omega = \operatorname{diag}\left[\frac{u_i^2}{(1-h_{ii})^2}\right]. \tag{12}$$

Eq. (11) together with Eq. (12) represent the $HC3$ matrix as denoted in and Long and Ervin (2000). $\boldsymbol{X}$ is the $p \times N$-matrix containing the explanatory variables and constant, with p the number of coefficients that are estimated, i.e. constant plus 2 explanatory variables in the estimation of specification I and 3 for estimation of specification II, and N the number of cities, which is 112. As weighting the leverage of observation i is taken $h_{ii} = \boldsymbol{x}_i\left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1}\boldsymbol{x}_i'$, which is the ii-element in the Hat-Matrix. The leverage of the observation h_{ii} is between $1/N$ and one 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 prefereable especially for small sample sizes smaller 250, which is the case in this analysis. The remaining problem is the presence of outliers in explaining and explanatory variables. Since the OLS approach is not robust against those outliers, the estimates are biased. Therefore, further methodes are applied that are more robust against outliers.

One robust estimation approach is the LMS method of Rousseeuw (1984), which minimizes the median of squared residuals instead of the sum as OLS does. The objective function is

$$\min_{\boldsymbol{\beta}} \operatorname{med}_i u_i^2 \tag{13}$$

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 the 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 how many percent of the observation which are allowed to diverge without changing the estimates is 0.5 for the LMS method indicating the robustness of this method and its results, respectively. This insensitivity with respect to outliers explains the differences of the estimates of LMS and OLS. These estimations are performed by use of the R package `MASS`. Although LMS-results are highly robust, the results are inefficiency. Thus additional methods are needed which are not only robust but also efficient.

An additional robust method is the MM-estimation as described in Yahai (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 $\widehat{\beta}_0$ has to be found, which should be robust by means of a high breakdown-point. That breakdown-point determines the robustness of the MM-estimation since this breakdown-point is not being 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 Yahai (1987) or Maronna et al. (2006). The IRWLS approach is computed in the three following steps:

1. Compute initial starting point for the estimate of $\widehat{\beta_0}$ by least absolute deviation estimation and scale $\widehat{s}$, with
   $\min \sum_{i=1}^{N} |u_i|$ with $u_i = y_i - x_i'\widehat{\beta_0}$ and $\widehat{s} = \frac{1}{0.675}\operatorname{med}_i\left(u_i \mid u_i \neq 0\right)$.
2. Iterate the estimatation of $\widehat{\beta_k}$ for $k = 0, 1, 2, 3, \ldots$ with the constant scale $\widehat{s}$ by solving $\sum_{i=1}^{N} w_{i,k} x_i \left(y_i - x_i'\widehat{\beta_{k+1}}\right)$ with $u_{i,k} = y_i - x_i'\widehat{\beta_k}$ and $w_{i,k} = W\left(\frac{u_{i,k}}{\widehat{s}}\right)$ with $W\left(\right)$ a nonincreasing function for positive arguments.

3. Stop the iteration when $\max_i \left(|u_{i,k} - u_{i,k+1}|\right) / \widehat{s} < \varepsilon$.

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

$$\min_{\widehat{\beta}} \sum_{i=1}^{N} \rho_0 \left(\frac{u_i\left(\widehat{\beta}_0\right)}{\widehat{s_N}} \right) \tag{14}$$

with the first order condition

$$\sum_{i=1}^{N} \psi_0 \left(\frac{u_i\left(\widehat{\beta}_0\right)}{\widehat{s_N}} \right) \boldsymbol{x}_i = \mathbf{0}, \tag{15}$$

where $\rho_0\left(\right)$ is a real function in the residuals which are scale invariant by the M-scale $\widehat{s_N}$ and $\psi_0\left(\right)$ is the first derivative of $\rho_0\left(\right)$. The properties of the function $\rho_0\left(\right)$ are given in Huber (1981) or Yahai (1987), for instance symmetry, continuity, a supremum between zero and infinity, and monotonic increasing for positive values where $\rho_0\left(0\right) = 0$. The sum in Eq. (14) devided by N and the supremum of $\rho_0\left(\right)$ has to be 0.5, such that the breakdown-point of the estimater is 0.5. In that step the initial regression estimate and the resulting residuals of the first step are taken as given and Eq. (14) is minimized by the M-scale. The third step minimizes a different maximum likelihood type function $\rho_1\left(\right) \leq \rho_0\left(\right)$ with the same supremum and the M-scale $\widehat{s_N}$ of the second step taken as given

$$\min_{\widehat{\beta}} \sum_{i=1}^{N} \rho_1 \left(\frac{u_i\left(\widehat{\beta}_1\right)}{s_N} \right). \tag{16}$$

This estimator is another maximum likelihood like estimation, which justifies the second M. Yahai (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 highly efficient. For further explanations see Maronna et al. (2006).

Table 2 shows the results for both specifications in Eq. (6) and Eq. (7) and for robust fit, ordinary least squares (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 covers the book Maronna et al. (2006), for the nonparametric methods and least median of squares estimations. The functions of the package `FEAR` are described in Wilson (2008).

Table 2: Regression results for specification II 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)
log(*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)
(log(*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)
(log(*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

Note: Significance is denoted by *** on 1%, ** on 5%, and * on 10% level (standard errors are reported in parentheses).

Table 2 shows that the linear term and the quadratic term for logarithm population size is significant at least at 5 percent level of significance for every approach. Furthermore, the estimates are of expected sign 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 like 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 surprisingly since the logarithm of population size is in the range between 10 and slightly above 15 which results in almost proportional quadratic and cubic terms. Altogether, the regression results in table 2 indicate that specification I is preferable and the distribution is quadratic with an inverted U-shaped design.[5]

The coefficient of determination R^2, which is maximized by OLS estimation, is calculated as described in Hayfield and Racine (2008)

$$\mathrm{R}^2 = \frac{\left(\Sigma_{i=1}^{N}\left(SE_i - \overline{SE}\right)\left(\widehat{SE}_i - \overline{\widehat{SE}}\right)\right)^2}{\Sigma_{i=1}^{N}\left(SE_i - \overline{SE}\right)^2 \Sigma_{i=1}^{N}\left(\widehat{SE}_i - \overline{\widehat{SE}}\right)^2}, \tag{17}$$

the fitted values for the regressand have a different mean ($\overline{\widehat{SE}}$) than 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 reasoning the use of Eq. (17), 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. (17) 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 least median of squares estimations. These estimations are more robust than the ordinary least square estimations against some violations of its underlying assumptions, i.e. the normal distribution of the error term, which implies no outliers. In cases, where outliers are present, the robust estimations better fit to 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 1 illustrates the fitted values for specification I as well as a kernel fit estimation.

[5] Additionally, a regression equation specification error test (RESET, Ramsey (1969)) has been performed, which was not able to reject the null hypothesis of no misspecification for specification I.

Figure 1: Fitted estimates for quadratic models and kernel regression

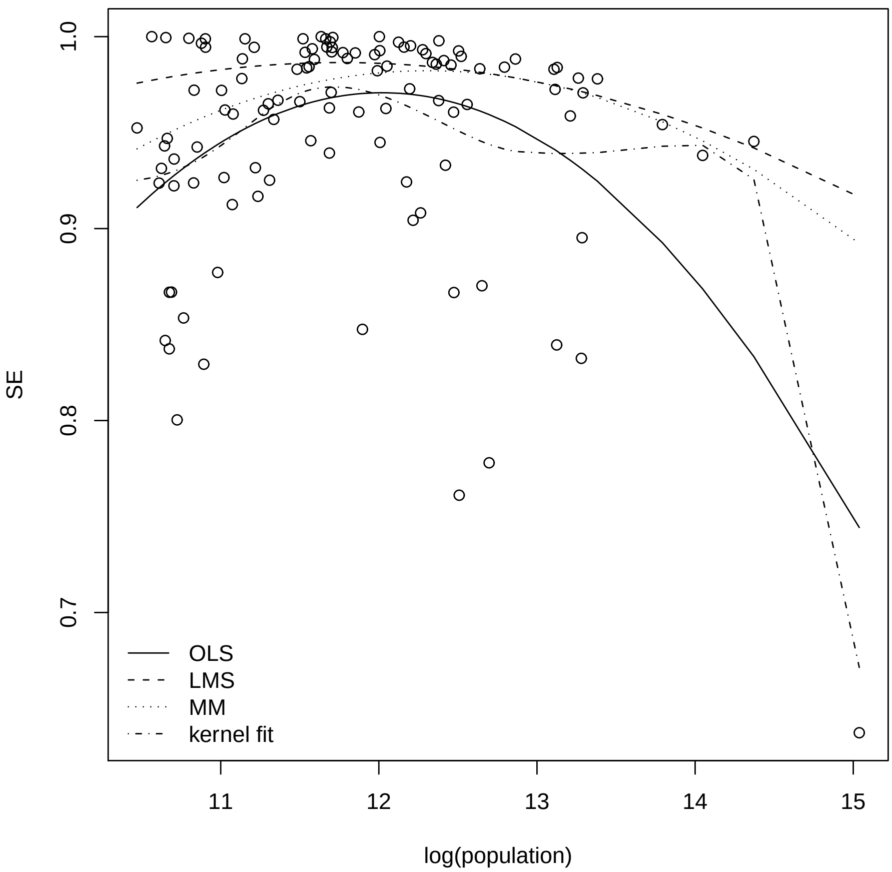


Figure 1 shows the fitted values for all approaches in specification I as well as the estimated function for a nonparametric kernel fit estimation. For the nonparametric kernel fit regression a bandwidth has to be chosen. This bandwidth is fixed at 0.32 with respect to the underlying data in logarithm 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 continuously, 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 it can been seen from the figure the global maximum point for the kernel fit regression is below 12 for the logarithm of population and thus of almost the same amount as the LMS estimation. There is a local maximum point at over 14 for the logarithm of population, which results as a spurious outcome of the data sparsity in this region/at this size and can be explained by the 4th to 2nd largest cities.

Figure 1 points out the estimated graph and the maximum points of each approach as well as reasons for the specific results. For instance the OLS estimation is influenced by few inefficient observation with low population size as well as the largest observation, which is the capital Berlin with a relative low scale efficiency. Furthermore, the maximum points are around 12 for the logarithm of population, that is 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 it is the question whether the differences in the optimal city size are of a reasonable magnitude and when this is the case which result is more trust-worthily. The figure demonstrates that there are many inefficient observations for small cities with population less than 100,000 and cities with population 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[6] and may contain outliers. Therefore, the OLS estimation does not seem to be proper 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, which is caused by the inefficient observation in the range between 12.5 and 13.5 that result in the local minimum point between 12 and 13 for the logarithm of population. Thus the kernel regression seems also to be inadequately to describe the observation, which may be different for other bandwidth but is not further estimated in this work caused by its nonparametric character which prevents further interpretations. The robust estimations, especially the robust linear MM-estimation, with quadratic term for the logarithm of population fits the observations best and therefore an optimal city size of about 220,000 inhabitants, which is the result for the robust linear fit model, is most proper for these observations. These findings are also supported by the empirical results in table 2, which states that the robust linear fit MM-estimator in specification I has an higher coefficient of determination 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 1).

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 for the optimal city size. The comparison is investigated 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 is cause by the German Reunification 1990 which leads to a change of the economic system associated with massive subsidies of west German economy. There were still differences in the economic performance in both parts observable as shown in Kirbach and Schmiedeberg (2008) or Sinn (2002). An other 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 hightech industries like microsystems (Jakoby (2006)). The comparisons ire performed by separated estimations for each area. The estimation approach is the robust linear MM-estimater, which best considers the heterogeneity of the observed cities and outliers. The border is the former inner-German border before the German Reunification 1990. Thus all 22 cities of the former German Democratic Republic are accounted for east Germany except Berlin which is viewed as a west German city. The results are presented in table 3.

Table 3: Results for optimal city size in West- and East-Germany

	east Germany	west Germany	north Germany	south Germany
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

Note: Significance is denoted by *** on 1%, ** on 5%, and * on 10% level (standard errors are reported in parentheses).

As table 3 shows there are differences in the estimates for east and west Germany and the estimates for east Germany are not significant at 10% level 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, respectively. 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 the counterpart for west Germany cities. In addition, the coefficient estimate for the quadratic term in east Germany is much higher as the one for west Germany, which indicates that the same amount of exceeding or shortfalling of city size results in a much higher loss

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

of scale efficiency for east German cities. This conclusion is also visualized by figure 3 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 for the three most southern located are treated as north German cities.

The results shown in table 3 is again that there is an optimal city size for both regions caused by the inverted U-shaped shape 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, respectively. 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 are for the north and the south almost the same at 18-19 percent. That is not very high but is due to the heterogeneity of the cities and the performed robust linear MM-estimator.

It is worthwhile to notice, that all estimations result in an inverted U-shaped shape of the relation of scale efficiency with respect to population size with an 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 Germans cities population is 231,700 (see table 1) and the mean of population in southern Germany is 155,600 and almost half the size of the mean of the population in northern Germany with 307,500. Therefore, it does not surprise that the optimal city size in southern Germany is smaller than in the northern part of Germany. Also the optimal city size in West-Germany including Berlin is slightly less than its mean with 251,900 inhabitants and 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 6 in the Appendix as well as in table 4 which presents the mean, median, and comparison between the calculated optimal city cize and mean city size.

Table 4: Comparison between calculated optimal city size, mean city size, and median city size

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

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

As shown in table 4 the optimal city size is in all cases close to the mean of the underlying cities and thus respective different 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.

6 Conclusion

This work investigates the relation between efficiency and population size of German cities. The relevant efficiency in this context is the scale efficiency, which considers the specific size of the particular city. Therefore, the approach employed in this work 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 constant returns to scale and once for variable returns to scale and taking the ratio of both. The second stage investigates the relation of 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 stand to contrast to findings of Kanemoto et al. (1996) for Japan but is similar to the results of Loikkanen and Susiluoto (2004) and Loikkanen and Susiluoto (2006) for Finland.

Improved achievements on efficiency could be obtained by a more unified population distribution over all cities with the size of the mean. Cities with a suboptimal low city size should attract people in 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 e.g. traffic jams, noise, pollution, and so on, resulting 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 of 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 as the costs of living indices, area sizes of cities, geographic distances which affect network possibilities between neighboring cities, especially facing the clusters of cities in figure 2.

Another promising extension is the adoption of multi-level analysis since industries are one level of interest and these are part of the next level namely the city, and these are part of states (Bundesländer). On each level are specific rules and laws, which influence firms and people to settle in a specific city. So the analysis should account for these different levels. A further approach might by the investigating of the time series either by adopting dynamic models or by taking time as an other level in a multi-level analysis. In addition, it would be very interesting to evaluate the causality between citiy growth and change in efficiency. That analysis could be similar to the investigation of efficiency in Spanish regions by industry sector and time of 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 to establish solutions for for getting more efficient without changing the population size in cities that are too large or too small.

One more detailed look into the industry data of the cities could lead to separating different city types characterized by specialization in different industries and services as introduced by Henderson (1974). A more industry disaggregated database is needed to answer this question. But different city types may be one explanation for the range of optimal city sizes, caused by the requirement of the specialized local industry.

Appendix

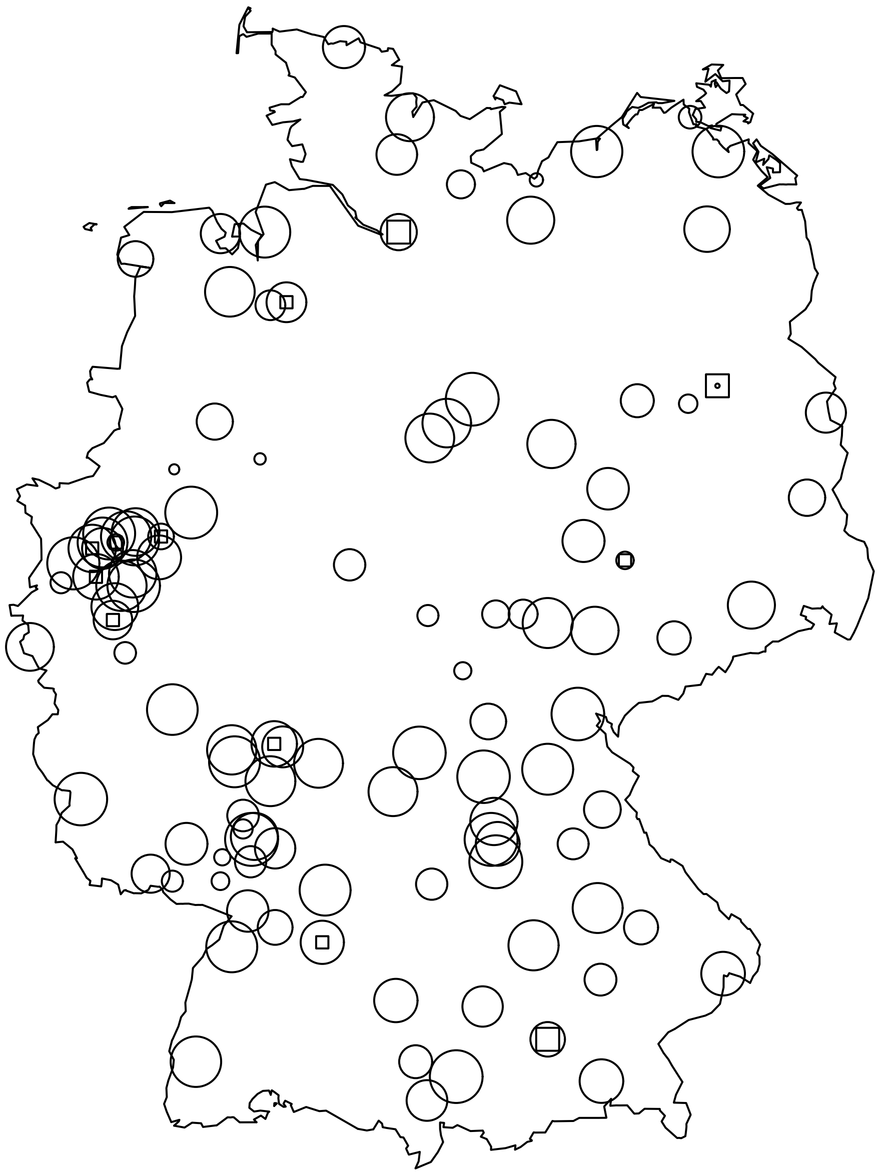
Appendix 1

Table 5: Cities

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	Ratisbon	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 2

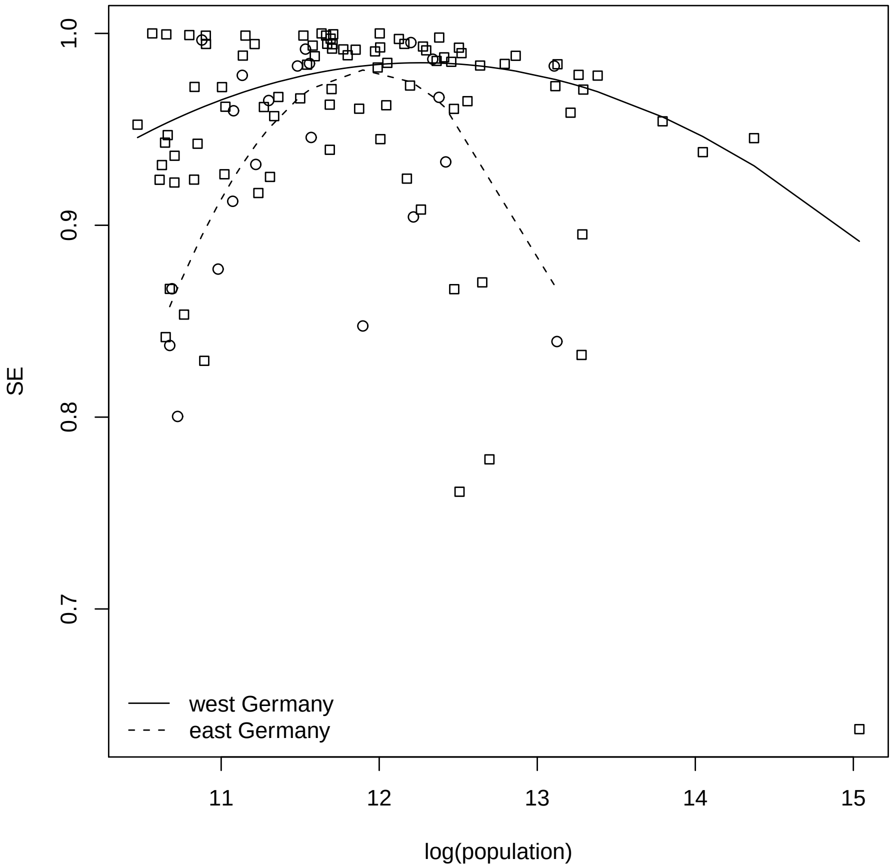
Figure 2: Germany map with scale efficiency and major cities



Note: Circles indicating the scale efficiency and boxes highlight cities with over half a million population size. Larger circles are for more efficient cities and larger boxes for larger cities but 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 drawn withn R using the package `mapdata`.

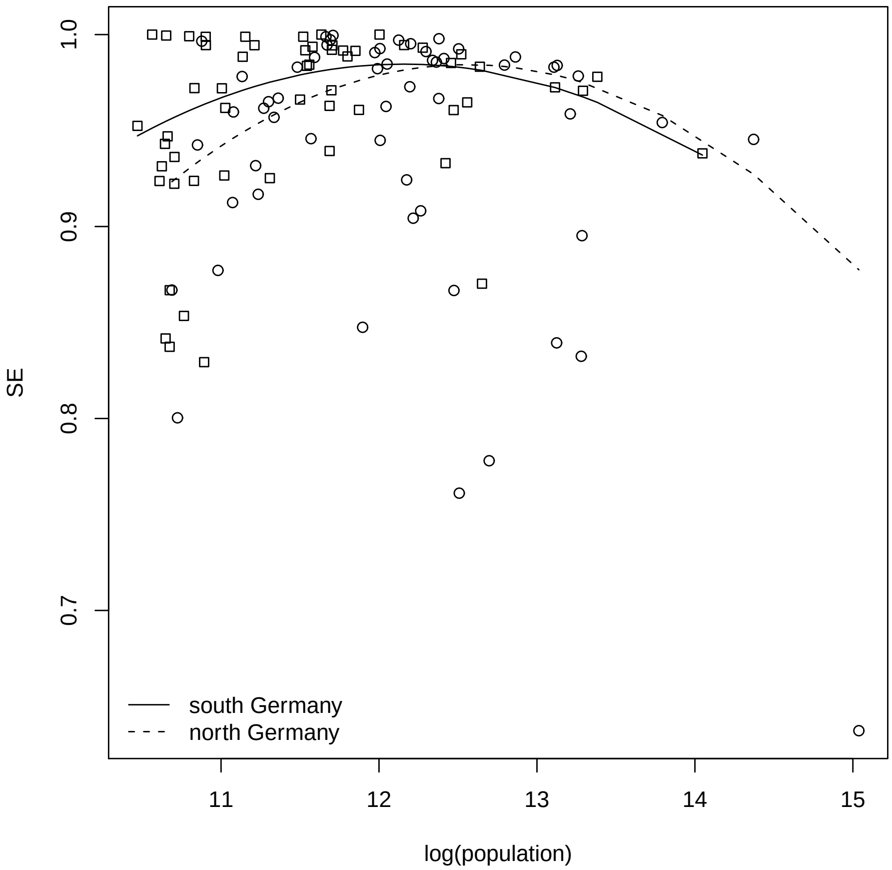
Appendix 3

Figure 3: Optimal city size for east and west Germany



Note: Circles indicate east German cities and boxes indicate west German cities. Firs are robust linear regressions separated for east and west German cities.

Figure 4: 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 4

Table 6: Descriptive statistics for population in thousands in different areas of Germany

	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.	SD
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

References

Aitchison, J. and C. G. G. Aitken (1976). Multivariate binary discrimination by kernel method. *Biometrika 63*, 413–420.

Alonso, W. (1971). The economics of urban size. *Papers in Regional Science 26*, 67–83.

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.

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.

Capello, R. and R. Camagni (2000). Beyond optimal city size: An evaluation of alternative urban growth patterns. *Urban Studies 37*, 1479–1496.

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.

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.). Springer.

Farrell, M. J. (1957). The measurement of productive efficiency. *Journal of the Royal Statistical Society Series A 120*, 253–281.

Federal Statistical Office (2003). *German Classification of Economic Activities, Edition 2003 (WZ 2003).* Wiesbaden: Federal Statistical Office.

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

Halkos, G. E. and N. Tzeremes (2010). Measuring regional economic efficiency: The case of greek prefectures. *The Annals of Regional Science 45*, 603–632.

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. *The Armerican Economic Review 64*, 640–656.

Henderson, J. V. (1975). Congestion and optimal city size. *Journal of Urban Economics 2*, 48–62.

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

Jakoby, H. (2006). Evaluation as part of the regional policy life cycle: The example of north rhine-westphalia, germany. *Regional Studies 40*, 281–284.

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.

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.

Li, Q. and J. Racine (2003). Nonparametric estimation of distributions with categorical and continuous data. *Journal of Multivariate Analysis 86*, 266–292.

Loikkanen, H. A. and I. Susiluoto (2004, August). Cost efficiency of finnish municipalities 1994-2002. an application of dea and tobit methods. Paper, European Regional Science Association.

Loikkanen, H. A. and I. Susiluoto (2006, February). Cost efficiency of finnish municipalities in basic service provision 1994-2002. Discussion Paper 96, Helsinki Center of Economic Research.

Long, J. S. and L. H. Ervin (2000). Using heteroscedasticity consistent standard errors in the linear regression model. *The American Statistician 54*, 217–224.

MacKinnon, J. G. and H. White (1985). Some heteroskedasticity-consistent covariance matrix estimators with improved finite sample properties. *Journal of Econometrics 29*, 305–325.

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

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

Maudos, J., J. M. Pastor, and L. Serrano (2000). Efficiency and productive specialization: An application to the spanish regions. *Regional Studies 34*, 829–842.

Moomaw, R. L. (1981). Productivity and city size: A critique of evidence. *The Quarterly Journal of Economics 96*, 675–688.

Nitsch, V. (2005). Zipf zipped. *Jounral of Urban Economics 57*, 86–100.

Park, W. G. (1995). International R&D spillovers and OECD economic growth. *Economic Inquiry 33*, 571–591.

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.

Richardson, H. W. (1972). Optimality in city size, systems of cities and urban policy: A sceptic's view. *Urban Studies 9*, 29–48.

Romer, P. M. (1986). Increase returns and long run growth. *Journal of Political Economy 94*, 1002–1037.

Rousseeuw, P. J. (1984). Least median of squares regression. *Journal of the American Statistical Association 79*, 871–881.

Segal, D. (1976). Are there returns to scale in city size? *The Review of Economics and Statistics 58*, 339–350.

Shephard, R. W. (1970). *Theory of Cost and Production Function.* Princeton: Princeton University Press.

Sinn, H.-W. (2002). Germany's economic unification: An assessment after ten years. *Review of International Economics 10*, 113–128.

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. *The Quarterly Journal of Economics 89*, 392–413.

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.

Wilson (2008). Fear: 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.

Yahai, V. J. (1987). High breakdown-point and high efficiency robust estimates for regression. *The Annals of Statistics 15*, 642–656.

Yezer, A. M. J. and R. S. Goldfarb (1978). An indirect test of efficient city size. *Journal of Urban Economics 5*, 46–65.

Zipf, G. K. (1949). *Human Behavior and the Principle of Least Effort.* Addison-Wesley Press, Cambridge.