Quality of Life in the Regions: An Exploratory Spatial Data Analysis for West German Labor Markets

Rusche, Karsten

Centrum für Angewandte Wirtschaftsforschung Münster

1 October 2008

Quality of Life in the Regions: An Exploratory Spatial Data Analysis for West German Labor Markets

CAWM Discussion Paper No. 10

October 2008

by

Karsten Rusche[1]

Abstract
Which of Germanys regions is the most attractive? Where is it best to live and work - on objective grounds? These questions are summed up in the concept “quality of life”. This paper uses recent research projects that determine this parameter to examine the spatial distribution of quality of life in Germany. For this purpose, an Exploratory Spatial Data Analysis is conducted which focuses on identifying statistically significant (dis-)similarities in space. An initial result of this research is that it is important to choose the aggregation level of administrative units carefully when considering a spatial analysis. The level plays a crucial role in the strength and impact of spatial effects. In concentrating on various labor market areas, this paper identifies a significant spatial autocorrelation in the quality of life, which seems to be characterized by a North-Mid-South divide. In addition, the ESDA results are used to augment the regression specifications, which helps to avoid the occurrence of spatial dependencies in the residuals.

Keywords: Quality of Life - Exploratory Spatial Data Analysis - Functional Economic Areas - Spatial Econometrics - LISA Dummies

JEL-Classifications: D63 - R10 - R12 - R50

1. Introduction

Which region in Germany offers the highest quality of life and which one offers the most attractive set of amenities? How are these areas distributed in space? Is there a tendency towards clustering? A wide range of currently published rankings aims at answering these questions. For example: INSM Regionalranking, Innovationsranking, Focus Landkreistest, Zukunftsstudie Berlin-Institut für Weltbevölkerung und globale Entwicklung, Städte-Ranking Capital, Bundesländerranking INSM, all deal with this issue. These studies analyze regions by means of both soft and hard location factors to assist policy makers in determining the strengths and weaknesses of their particular region (Buettner/Ebertz 2007a). Different methods of scoring and evaluating local data are used, so that the results vary widely. Given these disparities, the rankings are mostly also accompanied by intense discussion, when the results are published. Such reactions deal with appropriate way of interpreting the ranking positions or criticism concerning the validity of the scoring methods

[1] Correspondence: K Rusche, Center of Applied Economic Research Muenster, Institute for Spatial and Housing Economics, Am Stadtgraben 9, 48143 Muenster, Germany. Email: rusche@insiwo.de, phone: +49 251 83 22972.

(Uhde 2008). The studies employ different weighting schemes, non-congruent sets of variables and non-matching spatial aggregation levels (Counties vs. Planning Regions vs. Länder). Therefore, the results cannot be compared scientifically and they reveal a wide spread in the relative position of different regions to one another. In a recent research paper from Buettner/Ebertz (2007b), these criticisms are considered. Based on theoretical grounds, implicit prices are derived in a regression approach. They are summed to form an objective indicator of the quality of life in the German counties.
The present article considers and evaluates the analysis of the quality of life in the regions. In order to do this, the results of the "quality of life index" (QOL) are integrated into an Exploratory Spatial Data Analysis (ESDA). This approach has been established in the field of spatial econometrics and has recently been applied to various different sets of data and countries (Le Gallo/Ertur 2003, Dall'erba 2005, Patacchini/Rice 2007).
The statistical examination of the spatial distribution of the QOL adds important investigation issues to the results published so far. Specifically, these elements are:

a. What reference level of spatial aggregation is optimal for this kind of evaluation?
b. Is there a spatial concentration of the quality of life in Germany?
c. Is the distribution of quality of life characterized by significant local regimes?

The analysis is arranged as follows. In Section 2, the approach by Buettner/Ebertz - as the basis for the ESDA - is defined briefly. Section 3 begins with a definition of the objectives of an Exploratory Spatial Data Analysis. Then, the choice of appropriate spatial aggregation levels is discussed. Subsequently, the ESDA is employed, its specifications are determined and the results are visualized. A regression approach is then empowered with the results from the ESDA and compared to a non-spatially referred regression. Section 5 concludes.

2. Quality of Life in the Regions

Based on primary work by Rosen (1979) and Roback (1982), Buettner and Ebertz (2007b) apply a method for determining the quality of life in German counties ("Kreise und kreisfreie Städte"). Their methodology focuses on the assumption that the (dis-)amenities in regions capitalize into land prices and wages. If this holds true, the driving forces for these data can be specified by identifying and quantifying the significant influencing variables (Buettner/Ebertz 2007b).
For this purpose, a regression approach is implemented to analyze the prices of land and household incomes. In addition to the relevant control variables for the spatial character of the regions, a variety of data aimed at mirroring the (dis-)amenities are used. These ecological indicators are amended with survey data which characterizes regional attractiveness as a place to work and live. Table 1 lists the exogenous variables which were used.

Table 1: Variables influencing the quality of life

Indicator
Metropolitan area
Peripherality
Level of Tourism
Industrial emissions
Sunshine
Share of water

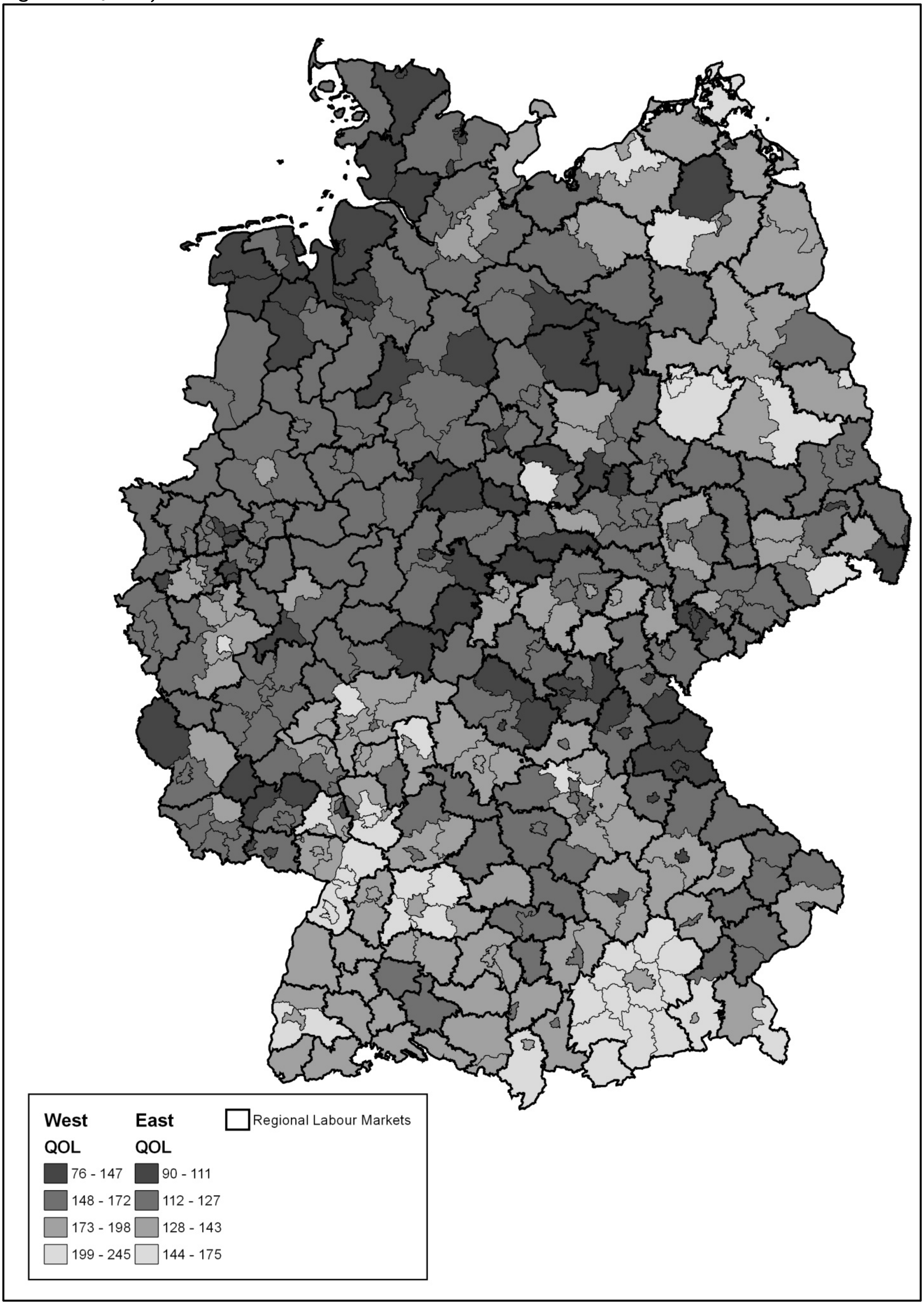
Share of forest
Level of Poverty
Leisure facilities
Accessibility
Education
Crime
Local labor market
Alternative job opportunities
East
Rural
Rural and East
Population density
Population growth

The regressions yield a very high explanatory power for the land prices. Because almost all the coefficients for household income regression indicate no statistical significance, this second approach was eliminated from further analysis (Buettner/Ebertz 2007b).

The remaining, significant coefficients of land price regression are then used to determine the willingness to pay in Euros for each variable. For this purpose, the provisional results from the regression are recalculated, focusing on the monthly household budget. The quality of life index for each region is then the sum of the implicit willingnesses to pay, weighted with the values of the local sets of (dis-)amenities (Buettner/Ebertz 2007a).

Figure 1 depicts the results for the German counties. West and East German regions are colored separately.

Figure 1: Quality of Life in German Counties



Buettner and Ebertz's method of determining the quality of life is open to criticism. First, when concentrating on the aggregation level of counties it is necessary to check for spatial autocorrelation and to implement the appropriate spatial econometric techniques. Furthermore, there is not necessarily a direct interlink between quality of life and the price of land. Because approval of land for development is heavily influenced by restrictive or active local policies, the price of land is distorted. The relevant scarcity of available land for development can only be described in a rudimentary manner by the use of the corrective variable "population density". Nevertheless, the results by Buettner/Ebertz offer a unique dataset which is worthy of further investigation.

3. Exploratory Spatial Data Analysis

3.1 Approach

Particularly for the analysis of relatively small regions, the existence of two spatial effects must be taken into account (Anselin 1996). The field of spatial econometrics labels these effects as <u>spatial autocorrelation</u> and <u>spatial heterogeneity</u>. The first effect is derived from the observation that, within a geographical reference area, "everything is related to everything else, but near things are more related than distant things" (Tobler 1979). Referring to this "First Law of Geography", nearby regions will yield a correlation with one another. In economic agglomerations, it is, for example, a common observation that there is not only a high concentration of economic activities at the centre. The surroundings of the core regions are also affected by network and transmission effects, so that they are also characterized by above-average economic prosperity. Economic activity tends to cluster in space (Le Gallo/Ertur 2003). The second spatial effect, heterogeneity, results from the existence of spatial autocorrelations. Interrelating effects of economic variables in a region under study may not have the same level of impact everywhere in space. Following the concept of time series analysis, this effect is also referred to as spatial non-stationarity (Kosfeld/Eckey/Türck 2007).

The concept of an Exploratory Spatial Data Analysis picks up these spatial phenomena and incorporates systems for different regional neighborhoods, so as to achieve four main objectives (Anselin 1996):

1. describe spatial distributions,
2. discover patterns of spatial association (spatial clustering),
3. suggest different spatial regimes (non-stationarity),
4. identify atypical observations (outliers).

Since the ESDA focuses especially on a combination of different (geo-)statistical tests, the issue of statistical significance is predominant. Before the ESDA starts, it is necessary to determine which spatial scale of aggregation should be used in the research context (Dall'erba 2005).

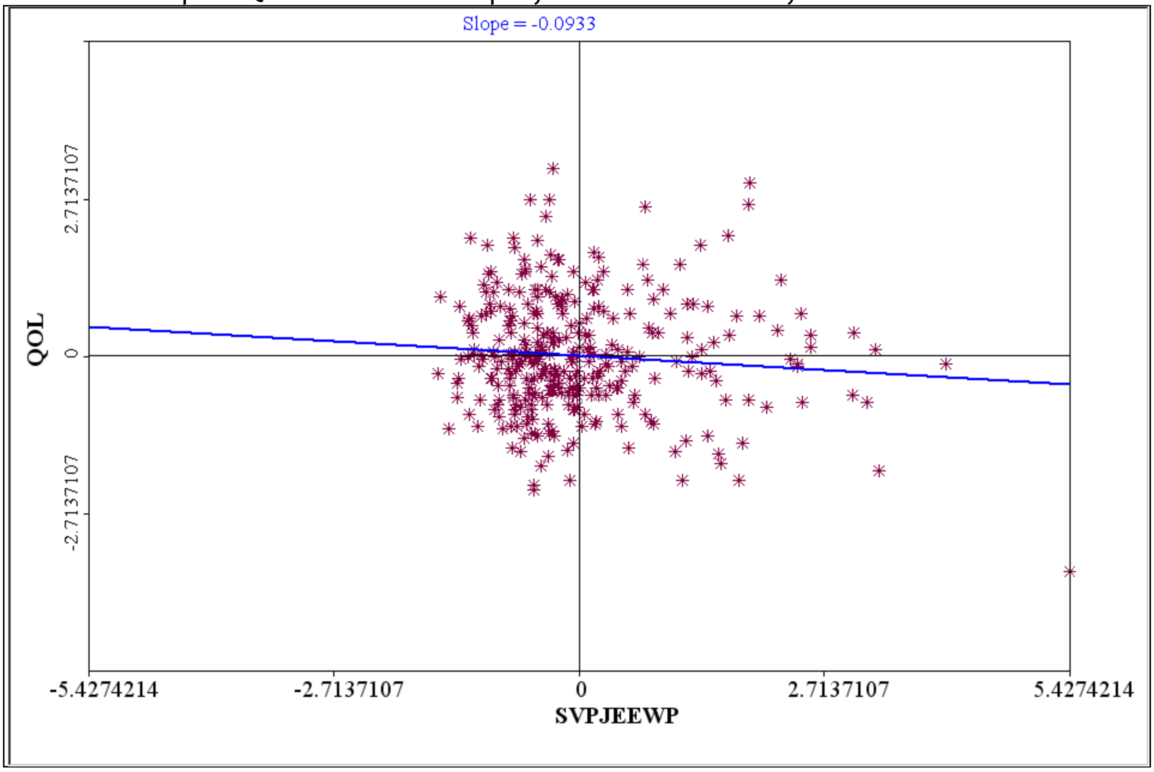
3.2 Scale of spatial aggregation

The strength of spatial relations and the interpretability of economic indicators are influenced by the choice of spatial aggregation level for administrative units. Depending on this spatial scale, relationships between regions may interfere with each other, which would bias the analytical results (Magrini 2004). This effect is named the "Modfiable Areal Unit Problem" (MAUP) and refers to the difference between administrative units and functional (economic) areas (Openshaw 1984). Interregional relationships spread continuously over space. Politically delineated regions often dissect these interrelations artificially (Openshaw/Taylor 1979). Therefore, it is necessary to aim for a high degree of congruency between the level of aggregation and the underlying spatial diffusion. An inappropriately selected spatial scale will lead to wrong conclusions (Eckey/Kosfeld/Türck 2006a). In order to avoid this, it is necessary to concentrate on functional areas. They aggregate administrative regions by means of their interconnectivity in commuter or migration flows (Eckey/Kosfeld/Türck 2006a and Michels/Rusche 2008).

With respect to the data on quality of life, the aggregation level of German counties turns out to be inappropriately low in scale for two reasons.
On the one hand, the usage of county level data neglects the linkage between the cities and their hinterland. In Figure 1, it is clearly evident that, at the core, cities often have a lower quality of life than in their direct surroundings. However the quality of life in the nearby counties cannot be evaluated in this way, because the main driver for a higher value of QOL is the proximity to a city with a greater degree of centrality. Accordingly, this connectivity between cities and their hinterlands must be taken into account when analyzing the spatial distribution of quality of life.
On the other hand, the way in which data for some exogenous variables was collected leads to a bias in favor of functional regions rather than county boundaries. The variables "local labor market" and "alternative job opportunity" are based on survey results. Because these indicators reveal a strong impact on local quality of life, a closer look at the interview questions is appropriate. Concerning the "local labor market", people in Germany were asked whether they consider the local labor market to be satisfactory or not. In the event of "alternative job opportunities", the interviewee should assess whether he or she would - if losing their job - find an equivalent job in their region within a reasonable period of time (Buettner/Ebertz 2007b). Thus, for both questions, the interview focused on local and regional aspects, rather than limiting the scale to the county borders. It can therefore be concluded that the answers to these questions focus mainly on the aggregation level of local labor market, rather than on administrative boundaries.
Altogether, both considerations indicate that the appropriate spatial scale for the spatial analysis of quality of life should focus on functional economic areas. The influence of disturbances would then be minimized. This article uses the "regional labor markets" (RLM) of Eckey/Kosfeld/Türck (2006a), which are the most recently delineated functional economic areas for Germany. Figure 1 highlights the functional areas with the aid of thick lines. It can be visually assessed that these labor markets indeed capture the relevant core-periphery pattern between German counties.

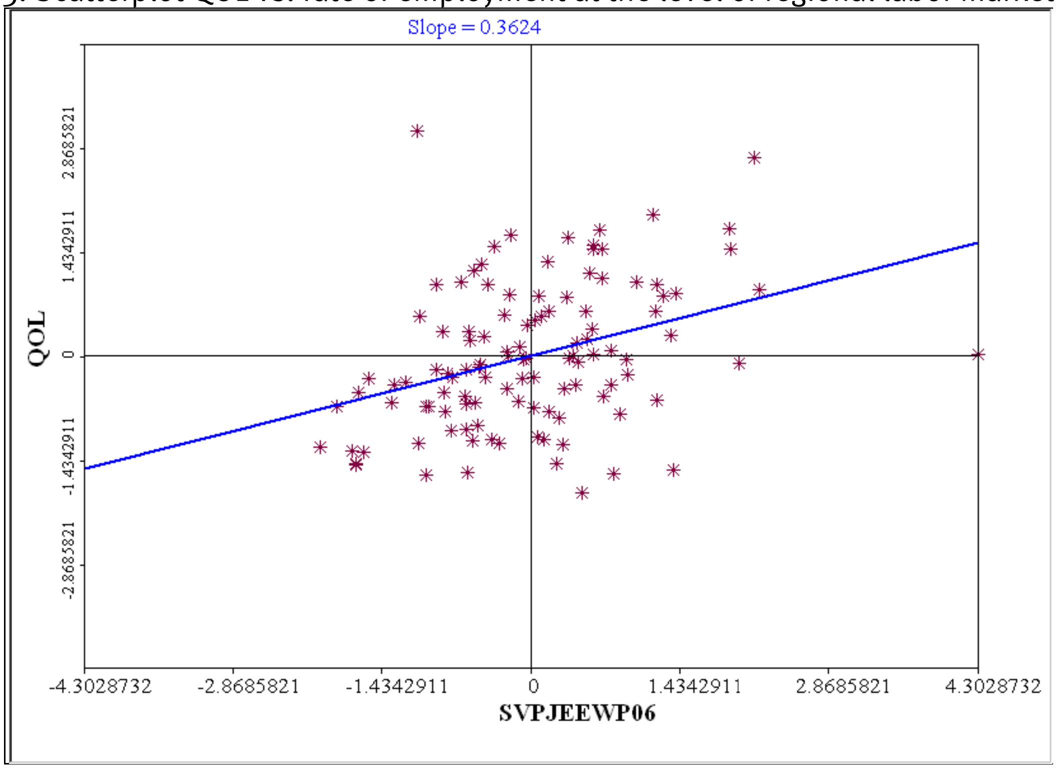
A good example of the misleading effect of an inappropriately chosen spatial scale is the relationship between quality of life and the rate of employment (i.e. employees liable for social insurance as a proportion of total economically active persons, age 15-65). The biasing effect of cities and their related surroundings results in a counterintuitive observation. That is, when plotting both variables against each other at the county level, it is evident that there is a slightly negative correlation between quality of life and attractiveness of the region for labor (Figure 2).

Figure 2: Scatterplot QOL vs. rate of employment at the county level



However, this conclusion of a negative relationship between labor and quality of life should not be drawn from the data. After all, for labor-market-related issues, the spatial scale of counties is not appropriate, because the radius of labor market effects extends beyond this relatively small level of aggregation (Kosfeld/Eckey/Dreger 2006). Here, it is obvious that an analysis at the county level fails to take into account the relationship between places to live and places to work. This clearly biases the results. A spatial econometric analysis would not resolve this problem completely, so that the results from the ESDA would remain imprecise (Eckey/Kosfeld/Türck 2007). Although this research paper aims at analyzing spatial dependencies, the focus lies on the relationship between regions. At the county level, the correlation within regions that are linked with one another would interfere with the results of the ESDA (Niebuhr 2001).
In Figure 3, the correlation between the QOL and the rate of employment is plotted at the level of regional labor markets.

Figure 3: Scatterplot QOL vs. rate of employment at the level of regional labor markets



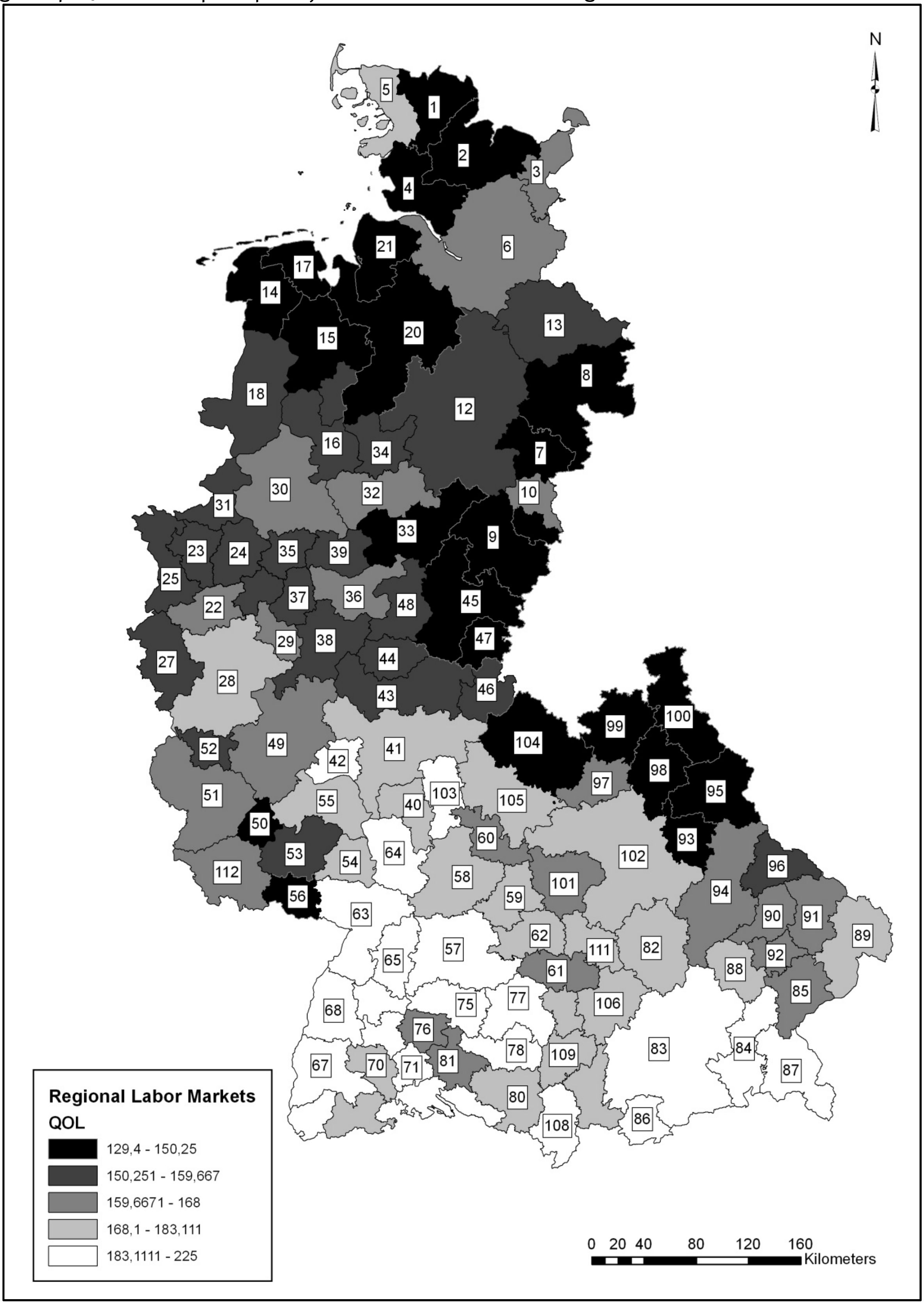
Here, the visual examination indicates a positive correlation. This emphasizes the importance of the appropriate spatial scale. Regional labor markets are apparently effective in demonstrating dependencies between areas within a specific region.

In addition, a further constraint has to be made. The ESDA analyses only the West German regional labor markets. This is important for a consistent assessment of the spatial dependencies. Because the values for quality of life are a result of a regression on land prices, there is still a need to differentiate between West and East Germany (Bode 2008). This due to the fact that the East German housing market has undergone strong politically driven distortions and is, therefore, far from equilibrium (Möller 2008). Accordingly, this article uses 112 regional labor markets that aggregate 330 counties (four East German counties are allocated by Eckey/Kosfeld/Türck to West German labor markets). The particular values for quality of life are calculated as mean averages of the county-based raw data.

3.3 Visualization

The first step in the ESDA is to find indications of spatial clustering of the data. On this basis, hypotheses for the spatial regimes are set up that are assessed statistically (Patacchini/Rice 2007). For this purpose, Figure 4 starts by depicting the distribution of quality of life in the regional labor markets.

Figure 4: Quintile Map of quality of life in West German regional labor markets



An initial hypothesis that can be derived from Figure 4 is that there is a strong south-north divide in quality of life. This distinction can also be identified when analyzing location factors (Fischer et.al. 2007). The regional quintile distribution reveals that all regional labor markets in the highest quintile are localized in Bavaria or Baden-Wuerttemberg. Beginning with the fourth quintile, only two regions are located in more northern latitudes: Cologne/Bonn (28) and North Friesland (5). The regional labor markets of Lower Saxony, North Rhine-Westphalia, North Hessen and Northern Bavaria can be identified as those with relatively lower qualities of life. The ESDA is primary a tool for measuring these tendencies towards local clustering of areas with high or low levels of quality of life. The analysis below will shed more light on the hypothesis of a north-south divide - measured on a statistical basis. Is there a significant spatial clustering of the quality of life or is it randomly distributed?

3.4 Spatial Weights Matrix

At the start of each spatial econometric analysis or Exploratory Spatial Data Analysis, it is necessary to define the way in which the spatial connectivity is encoded (Anselin 1988). Therefore, one or more spatial weights matrices $\underline{W}$ are formulated. The most straightforward way to incorporate the spatial neighborhood into a matrix is to use a 0-1 contiguity scheme. All regions with a common border are marked with a "1" in the corresponding cell of the weights matrix. Other regions are indicated with a Null. Given that no region is a neighbor to itself, the matrix diagonal is also filled with Nulls. The decision of whether or not regions should be allocated a common border is basically derived from the game of chess. The status of a direct neighbor is indicated by referring to the movements of the rook, bishop or queen. For a more precise statistical usage and an easier assessment of spatial lags, the spatial weights matrix is usually row-standardized.
Besides this fairly straightforward coding scheme, researchers have developed a large variety of different ways for dealing with neighborhood. The most commonly used examples are inverse distances, commuter flows, surface shares, k-nearest neighbors or local statistics as a basis for the coding scheme (Getis/Aldstadt 2004).

The choice of spatial weights matrices plays a crucial role in determining the spatial effects. The results of each spatially referenced analysis depend on the matrix used. Therefore, the appropriate weights matrix should be chosen with care. Although there are some rules of thumb for determining the weights matrix (Griffith 1996), empirical research tends to compare the results for a set of likely fitting coding schemes. If the spatial reference is chosen appropriately, the results from different spatial weights should be robust (Le Gallo/Ertur 2003, Dall'erba 2005).

Since this article focuses on the spatial scale of functional economic areas, using weights matrices that are based on distances is not recommended. As the delineation of regional labor markets is aimed at defining the spatial scope in which all inhabitants of the labor market live <u>and</u> work, the relevance of distances is already taken into account (Baumont/Ertur/Le Gallo 2004). Because the regional labor markets vary substantially in size, the use of different k-nearest neighbor matrices and contiguity based weightings is advisable. The concept of k-nearest neighbors defines neighborhood by means of the number of k surrounding regions that are geographically nearest to the assessed core (referring to the center of a regional labor market). The general form can be described as follows (Guillain/Le Gallo 2006):

$$\begin{cases} w_{ij}^{*}(k) = 0 \; if \; i = j \\ w_{ij}^{*}(k) = 1 \; if \; d_{ij} \leq d_i(k) \; and \;\; w_{ij}\,(k) = w_{ij}^{*}(k) / \sum_j w_{ij}^{*}(k) \\ w_{ij}^{*}(k) = 0 \; if \; d_{ij} > d_i(k) \end{cases} \qquad (1)$$

with $w^{*}_{ij}(k)$ as an element of the (non-standardized) weights matrix, $w_{ij}(k)$ as an element of the row-standardized matrix. $d_i(k)$ depicts the individual distance from which on each region has exactly k neighbors.
For the further analysis, three weights matrices are used: queen contiguity weights together with 5- and 10-nearest neighbors.

3.5 Global spatial autocorrelation

Moran's I is a widely used indicator for measuring the existence and impact of spatial autocorrelation. It has evolved as a standard tool for spatial analysis (Getis 2008) and is calculated as follows:

$$I = \frac{n}{S_0} * \frac{\sum_{i=1}^{n} \sum_{j=1}^{n} w_{ij}(x_i - \bar{x})(x_j - \bar{x})}{\sum_{i=1}^{n} (x_i - \bar{x})^2} \qquad (2)$$

Here, n is the number of observations, S_o stands for the sum of all elements in the weights matrix and x represents the values of the observations. In essence, Moran's I can be seen as a regression coefficient resulting from the regression of Wx* on x* (x* in deviations from the mean) (Anselin 1996). The interpretation of this measure is analogous to common correlation coefficients. A positive value of Moran's I indicates that there are more often regions with high (low) values surrounded by neighboring regions which also have high (low) values. Contrary to a purely random distribution in space, a spatial autocorrelation would then be detected.
Statistical inference from the significance of Moran's I can be based on a normalization approach or by means of conditional permutations (Patacchini/Rice 2007). The data on quality of life yields high statistical significance for both alternatives. Table 2 gives an overview of the results.

Table 2: Moran's I for quality of life in the regions

Weights matrix	*Moran's I*	*standard deviation*	*z-value*
contiguity	0.570	0.06073	9.535
5-nn	0.596	0.05501	11.004
10-nn	0.530	0.03831	14.058

All statistics are significant at a p-value = 0.0001
(normalization and 9999 permutations)

Even with this relatively high level of spatial aggregation, the values of Moran's I indicate a high degree of spatial autocorrelation for quality of life. The results for the 5-neareast-neighbor (5nn) weights matrix lead to the highest significant level of spatial correlation. Accordingly, this spatial weighting is chosen as the standard neighborhood scheme for further analysis. Two more rationales support this choice. First, this relatively small number of nearest neighbors is preferred to the higher one, even if the z-value is lower (Patacchini/Pace 2007). This analysis is already based on an appropriate, but higher-level spatial scale, so that the 5nn should cover the relevant spatial dependencies among regional labor markets. Second, the use of knn weights matrices offers certain methodological advantages compared to contiguity based weights. In fixing the number of neighbors for each region instead of varying them, these problems are avoided (Le Gallo/Ertur 2003). Nevertheless, all calculations in the subsequent analysis are computed for all three weights and the robustness of the results is assessed. If any significant distortions occur, they are documented.

Concerning the quality of life, it is evident that the distribution in space is far from random (Dall'erba 2005). In the sense of global autocorrelation, Moran's I clearly indicates that there are significant clustering tendencies for quality of life in Germany. Regions with a high (low) value of quality of life are often surrounded by regions that are also very (un-)attractive. But this evaluation lacks other important information. How can this trend of spatial clustering be explained in terms of local geographical patterns? Are there different regimes of quality-of-life clusters? Do these "nice places" all cluster in southern Germany? If there are any outliers, where are they situated? These questions cannot be answered by means of global measures of autocorrelation. For this purpose, the ESDA uses the tool Moran scatterplot (Anselin 1996) and analyses local indicators of spatial association, LISA (Anselin 1995).

3.6 Local Indicators of Spatial Association

In the context of the analysis of local patterns of spatial associations, there are two main statistics under discussion (Getis/Ord 1996, Kosfeld/Eckey/Türck 2007):

a. G_i and G_i^* (Getis/Ord 1992, Ord/Getis 1995),
b. Local Moran (Anselin 1995).

Both statistics aim at identifying significant spatial clustering of high or low values in the observations. This paper concentrates on the local Moran, because it has some advantages over the G_i statistics. Firstly, there is a difference in the range of spatial regimes that are identified by both statistics. While the G_i statistic focuses on the significant clustering of high values ("hot spots") or low values ("cold spots"), the local Moran examines the similarity in value constellations. Therefore, the results of this statistic divide the observations into four groups: high-high together with low-low combinations as spatial clusters and high-low together with low-high combinations as spatial outliers. Since this analysis focuses on the entire range of spatial clusters and spatial outliers, the local Moran is the measure of choice.
Secondly, the G_i statistics have been developed for data sets in which there is no global spatial autocorrelation. However, the previous section indicated a high degree of spatial autocorrelation in the quality of life. In addition, the definition of a LISA requires that the local statistics sum to form a global measure of correlation. Moran's I can be calculated as

the arithmetic mean of the local Moran values for all observations, but this does not hold for the G_i statistic. This characteristic of the local Moran can be used for an initial assessment of the local structure of the spatial regimes. Therefore, the Moran scatterplot (Anselin 1996) can be plotted.

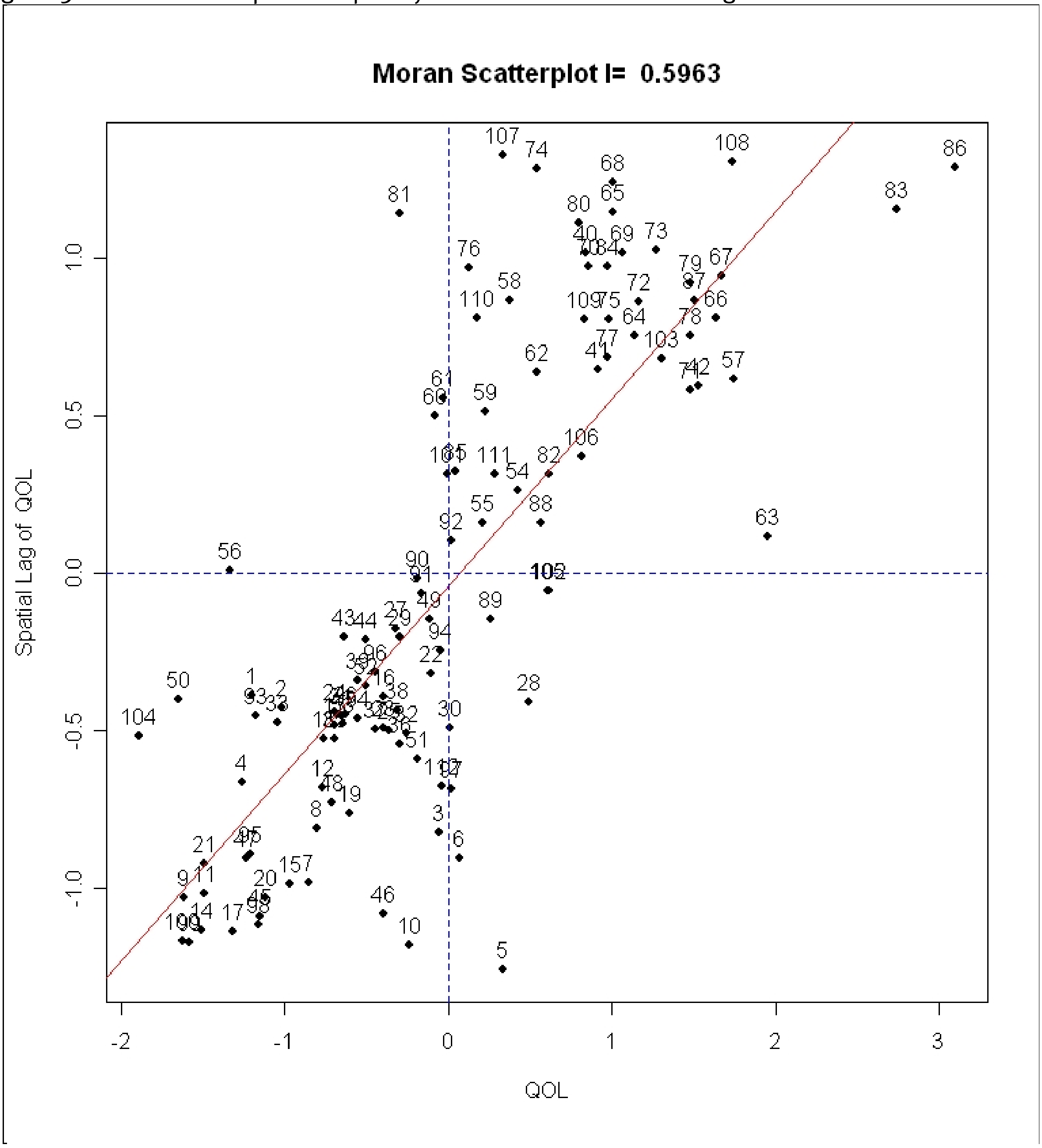
Formula 3 lists the calculation of the local Moran (Beaumont/Ertur/Le Gallo 2004):

$$I_i = \frac{(x_i-\mu)}{m_0} \sum_j w_{ij}(x_j - \mu) \ with\ m_0 = \sum_i (x_i - \mu)^2/n, \quad (3)$$

with x as the value of the observation and μ as the mean. The summation over j is conducted only for those regions that are specified as neighbors in the referring spatial weights matrix. A positive (negative) value of the local Moran indicates a clustering of similar (dissimilar) values. Taking the initial values of the core observations into consideration, the previously described spatial regimes can be identified. The statistical inference for the local Moran must be based on a conditional permutation approach, which inevitably leads to pseudo-significance levels. Since the occurrence of global spatial autocorrelation interferes with the local dependencies between different regions, the p-values of the local Moran are somewhat overestimated. Therefore, the overall significance level of 5% might be too generous. Here, 9999 permutations were used in all calculations for the local Moran and the significance levels of 1% and 0.1% were also considered (Beaumont/Ertur/Le Gallo 2004).

A first step in analyzing local patterns of spatial association is to take a look at the Moran scatterplot (Figure 5). In this graph, the value for quality of life for each regional labor market is plotted against the weighted average of its surrounding 5-nearest neighbors. This scatterplot of individual values and spatial lags can be divided into four quadrants. Because Figure 5 is based on standardized values, the slope coefficient can be interpreted as the Moran's I for global autocorrelation.

Figure 5: Moran scatterplot for quality of life in West German regional labor markets

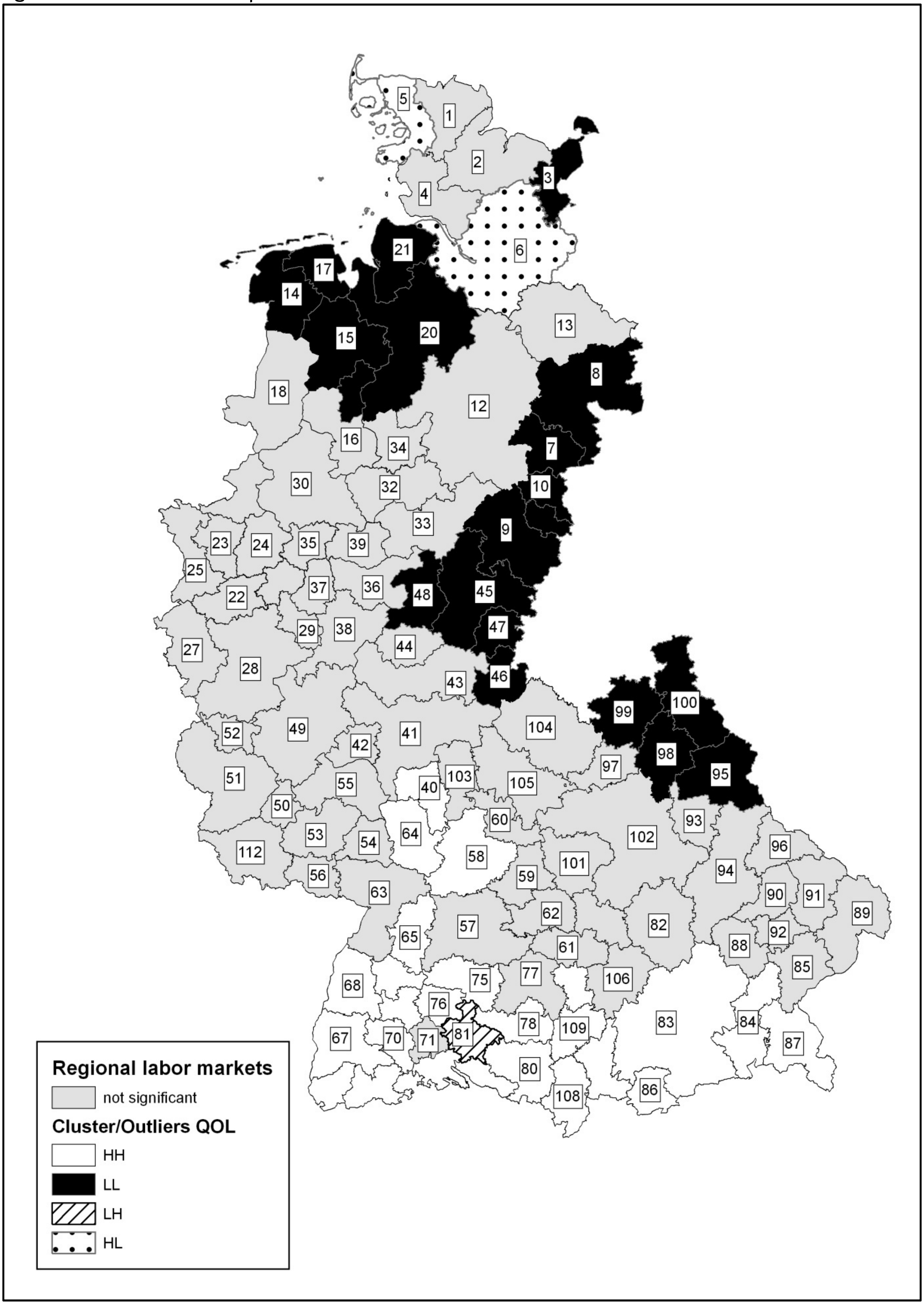


As the analysis of global association already suggests, there is a strong incidence of clustering. Almost all regions are positioned in the high-high (upper right) or low-low (lower left) quadrants. This strengthens the first set of results which yielded a high spatial autocorrelation for quality of life. A visual examination of the Moran scatterplot provides more insight into the hypothesis of a north-south divide. As depicted in Figure 5, the regional IDs belonging to labor markets in the north tend to be in the low-low quadrant. However, the points for southern labor markets appear mainly in the high-high cluster quadrant. Here Munich (83) and Garmisch-Partenkirchen (86) constitute extreme values within the group of high-high clusters.

In addition to the identification of cluster tendencies, the Moran scatterplot can also depict spatial outliers. A few regional labor markets can be found in the high-low (lower right) or low-high (upper left) quadrant. Therefore, Hamburg (6), North Friesland (5) together with Cologne/Bonn (28) and Passau (89) can be classified as spatial outliers. In contrast to their neighbors, they yield a relatively high value for quality of life. Pirmasens (56) and Sigmaringen (81), however, have a relatively low quality of life compared to the surrounding regions.

Although the analysis of the Moran scatterplot gives useful insights into the local structure of spatial association, there is no accompanying assessment of statistical significance. For this purpose, the individual significance of each local Moran coefficient is calculated on the basis of 9999 permutations. The inference is based on the pseudo-p-value of 0.05 (Table 4 in the appendix). Figure 6 highlights the significant spatial clusters and outliers (Anselin/Syabri/Kho 2006). It should be noted that the LISA cluster map colors only the centers of the clusters. Because the local Moran is calculated in combination with the particular neighborhood, according to the spatial weights matrix, the spatial range of the clusters should be seen in the broader context of the region, that is, including the neighboring regions which are not highlighted (Anselin 2005).

Figure 6: LISA cluster map



The cluster map contradicts the hypothesis of a clear north-south divide. The spatial structure of the clusters of quality of life is more differentiated. Indeed, some low-low clusters - such as Luebeck (3) - are located in the more northern regions of Germany. However, the bulk of the low-low clusters (13 from 19) is situated along the former inner-German frontier in the middle of Germany. This “dismal” and somewhat pessimistic evaluation is also confirmed by recent population and employment projections (van Suntum/Rusche 2007). Against this twofold geographical concentration of areas with a relatively low quality of life, there is a clear socioeconomic strength in the southern German regions in terms of QOL. All high-high clusters can be found in the south of Bavaria and Baden-Wuerttemberg and in the triangle of Darmstadt/Heilbronn/Mannheim. Concerning the spatial outliers, there are only two significant high-low combinations. Both are in the north of Germany: Hamburg and North Friesland. They can be characterized as “islands” of relatively high quality of life. As the second type of spatial outliers, a low-high combination, there is only a significant local Moran for Sigmaringen. Compared with its neighbors in the south of Germany, it has a relatively low QOL. When reducing the necessary p-value to a level of 0.01, the identified north-mid-south divide becomes even more obvious.

This concludes the Exploratory Spatial Data Analysis. The results reveal useful insights into the spatial structure and distribution of the quality of life. However, the conclusions from the ESDA can be used further in a regression framework which augments the econometric specifications.

4 Augmentations to Econometric Modeling

Related to econometric modeling, spatial dependence is an important phenomenon that has to be accounted for in the regression specifications. A significant spatial autocorrelation in the data leads at least to some degree, to inefficient estimators and at worst, they are also biased. Therefore, it is necessary in spatially referenced regressions, to test for spatial dependence in the residuals (Florax/Nijkamp 2004). If there is a significant correlation, the remedy can be analyzed further by a series of LM-tests. They assist in the decision as to which of the standard spatial econometric approaches has to be applied. These two basic models are the spatial lag and the spatial error model. In both cases, a spatially lagged variable is included in the (maximum likelihood) regression specification. It is added into the first model as an additional exogenous variable, in the second, as part of the disturbance term (Eckey/Kosfeld/Türck 2006b).

In a regression set up, the ESDA results help to construct a more appropriate regression model. An example of efficient augmentations that can be derived from the ESDA, is the influence of economic fundamentals on the quality of life. From a theoretical point of view, the rate of employment and the disposable income of households play an important role in the evaluation of the quality of life in the various labor markets. The rate of employment approximates the supply of labor, and the disposable income mirrors the economic performance of each regional labor market. These labor market effects were not an explicit component in the initial determination of quality of life. Buettner and Ebertz used survey data to represent labor market effects. The coefficients of these variables displayed a strong impact of labor-market-related variables on the quality of life. Thus, the analysis of income and employment should also confirm these initial assessments.

For issues of model specification, it is important to mention that a significant spatial autocorrelation is not inevitable. More specifically, it is often just a case of an improperly specified regression. In this interpretation, spatial autocorrelation is generated, due to an omitted variable problem and is not a ubiquitous issue in geographically based analysis. A more exact definition of the spatial dependencies can therefore assist in reducing the regression complexity (McMillen 2003). In this article, the ESDA results are used to construct dummy variables for all significant high-high and low-low regions. These regions are highly spatially intercorrelated. Because the significant clusters represent the extremes in the differences of quality of life, the dummy variables can be interpreted as instruments of the non-economic (dis-)amenities that exert an impact on the quality of life.

Table 3 lists the results of two competing regression models. The first is a simple OLS estimation of the economic fundamentals of quality of life. The second OLS estimation is augmented with the information gained from the ESDA - the two dummy variables are included. For a better interpretation of the coefficients, the calculations are based on a log-log model.

Table 3: OLS regressions with and without augmentation based on ESDA results

	OLS I			OLS II		
	Coefficients	t-value	p-value	Coefficients	t-value	p-value
Constant	3.223	7.511	< 0.01	4.622	13.119	< 0.01
Rate of Employment	0.126	1.234	0.220	0.249	3.237	< 0.01
Disposable Income	0.685	5.142	< 0.01	0.229	2.069	0.041
Dummy HH	-	-	-	0.120	6.710	< 0.01
Dummy LL	-	-	-	-0.105	-5.584	< 0.01
Overall Model Fit	adj.R^2=0.295 F=24.259***			adj.R^2 = 0.609; F=44.183***		
Test for spatial autocorrelation	Moran's I = 0.497*** Robust LM (lag) = 10.886*** Robust LM (err) = 1.044			Moran's I = 0.049		
Goodness of fit	Schwarz criterium = -195.826 Akaike criterium = -203.981			Schwarz criterium = -254.374 Akaike criterium = -267.967		

Notes: *** significant at a level of 0.01

For the first regression, the statistical tests indicate a high and significant spatial autocorrelation in the regression residuals. Statistical inference from the coefficients is therefore biased. On the basis of the LM-tests a decision has to be made as to which spatial regression model fits best. In this case, a spatial lag model would be the appropriate choice. In contrast, the second, augmented regression is not influenced by any significant spatial autocorrelation in the residuals. The implementation of the ESDA results therefore enables an identification of the most important drivers of spatial dependence. The coding scheme of dummy variables was useful in absorbing these effects ex ante.

The regression coefficients of the second OLS estimation are all significant and have the expected signs. In general this rather straightforward regression model achieves high explanatory power. Compared to the first regression, the goodness of fit of the model increases with respect to all listed measures given in Table 3.

Both the rate of employment and the level of disposable income impact strongly on the quality of life. Given that the geographical best and worst locations are encoded in the dummy variables, it is evident that location does matter. It should be noted that both cluster types indicate a similar absolute strength of influence on the quality of life. The relative impact on the QOL is smaller for these approximations of extreme sets of (dis-)amenities than for the labor market data. In sum, the distribution of quality of life in regional labor markets is therefore influenced mainly by two factors: economic prosperity and location.

This short econometric excursion demonstrates that the results of an ESDA are not just useful in providing insights into the spatial distribution of variables. The study also serves as an auxiliary tool for determining the structure of the econometric model.

5 Conclusions

This research has extended the existing research on the quality of life in German regions by analyzing its spatial structure. In order to do so an Exploratory Spatial Data Analysis was conducted, which enables the user to explore georeferenced data with respect to statistically significant global and local spatial associations in the variables. In the first step of the analysis, it became clear that, before commencing the spatial research, it is necessary to investigate which spatial aggregation level is appropriate for the particular research question. If a level is chosen with too small a scale, the dependencies and results can be biased, due to the modifiable areal unit problem. The visually drawn hypothesis of a clear north-south divide cannot be confirmed overall. Because there is a high spatial autocorrelation between quality of life in regional labor markets, the study does not reveal a schematic north-south development. Most labor markets with relatively low values for quality of life are situated in the former inner-German frontier in central Germany. The results for the local spatial associations confirm this conclusion. In the same context, it became clear that the regions with the highest quality of life are found solely in southern Germany.
In addition, the transfer of the main ESDA results to an econometric specification constitutes a methodological improvement in two ways. Firstly, it raises the explanatory power of the regressions and secondly it avoids the occurrence of spatial autocorrelation in the residuals. The implementation of the detected cluster structures helped to approximate the extremes of the geographically based (dis-)amenities in the labor markets.

References

Anselin L (2005) Exploring Spatial Data with GeoDa™: A Workbook, Urbana

Anselin L (1996) The Moran Scatterplot as an ESDA tool to assess local instability in spatial association. In: Fisher M et al. (Hrsg.) Spatial Analytical Perspectives on GIS, London

Anselin L (1995) Local Indicators of Spatial Association – LISA. In: Geographical Analysis, 27, 2: 93-115

Anselin L (1988) Spatial Econometrics: Methods and Models, Dordrecht

Anselin L, Syabri I, Kho Y (2006) GeoDa: An Introduction to Spatial Data Analysis. In: Geographical Analysis, 38: 5-22

Anselin L, Syabri I, Smirnov O (2002) Visualizing Multivariate Spatial Correlation with Dynamically Linked Windows. In: REAL Working Paper, no. 02 T-8, University of Illinois at Urbana-Champaign

Baumont C, Ertur C, Le Gallo J (2004) Spatial Analysis of Employment and Population Density: The Case of the Agglomeration of Dijon 1999. In: Geographical Analysis 36, 2: 146-176

Bode E (2008) Delineating Metropolitan Areas using land prices. In: Journal of Regional Science, 48, 1: 131-163

Büttner T, Ebertz A (2007a) Lebensqualität in den Regionen: Erste Ergebnisse für Deutschland. In: ifo Schnelldienst, 60, 15: 13-19

Büttner T, Ebertz A (2007b) Quality of life in the regions: results for German Counties. In: Annals of Regional Science, Online First, DOI 10.1007/s00168-007-0204-9

Dall'erba S (2005) Distribution of regional income and regional funds in Europe 1989-1999: An exploratory spatial data analysis. In: Annals of Regional Science 39: 121-148

Eckey H-F, Kosfeld R, Türck M (2007) Regional Convergence in Germany: A Geographically Weighted Regression Approach. In: Spatial Economic Analysis, 2, 1: 45-64

Eckey H-F, Kosfeld R, Türck M (2006a) Abgrenzung deutscher Arbeitsmarktregionen. In: Raumforschung und Raumordnung, 64, 4: 299-309

Eckey H-F, Kosfeld R, Türck M (2006b) Räumliche Ökonometrie. In: WiST 10: 548-554

Fischer G et al. (2007) Standortbedingungen und Beschäftigung in den Regionen West- und Ostdeutschlands. Ergebnisse des IAB-Betriebspanels 2006. In: IAB Forschungsbericht Nr. 5/2007

Florax R, Nijkamp P (2004) Misspecification in Linear Spatial Regression Models. In: Kempf-Leohnhard K (Hrsg.) Encyclopedia of Social Measurement, San Diego

Getis A (2008) A History of the Concept of Spatial Autocorrelation: A Geographer's Perspective. In: Geographical Analysis, 40: 297-309

Getis A, Aldstadt J (2004) Constructing the Spatial Weights Matrix Using a Local Statistic. In: Geographical Analysis, 36, 2: 90-104

Getis A, Ord J K (1996) Local Spatial Statistics: An Overview. In: Longley P (Hrsg.) Spatial Analysis, 261-277

Getis A, Ord J K (1992) The Analysis of Spatial Association by Use of Distance Statistics. In: Geographical Analysis, 24, 3: 189-206

Griffith D A (1995) Some Guidelines for specifying the geographical weights matrix contained in spatial statistical models. In: Arlinghaus S L, Nystuen J D (Hrsg.) Practical handbook of spatial statistics: 66-82

Guillain R, Le Gallo J (2006) Measuring Agglomeration: An Exploratory Spatial Analysis Approach Applied to the Case of Paris and its Surroundings. REAL Working Paper, no. 06 T-10, University of Illinois at Urbana-Champaign

Kosfeld R, Eckey H-F, Dreger C (2006) Regional Productivity and Income Convergence in the Unified Germany, 1992-2000. In: Regional Studies, 40: 755-767

Kosfeld R, Eckey H-F, Türck M (2007) LISA (Local Indicators of Spatial Association). In: WiST, 2007, 3: 157-163

Le Gallo J, Ertur C (2003) Exploratory spatial data analysis of the distribution of regional per capita GDP in Europe, 1980-1995. In: Papers in Regional Science: The Journal of the RSAI 82, 4: 175-201

Magrini S (2004) Regional (Di)Convergence. In: Henderson J V, Thiesse J F (Hrsg.) Handbook of Regional and Urban Economics, 4: 2741-2796

McMillen D P (2003) Spatial Autocorrelation or model misspecification? In: International Regional Science Review, 26, 2, 208-217

Michels W, Rusche K (2008) Abgrenzung von Wohnungsmarktregionen mit Hilfe von Arbeitsmarktverflechtungen. In: Materialien zum Siedlungs- und Wohnungswesen, Band 43

Möller J (2008) Regional variations in the price of building land: a spatial econometrics approach for West Germany. In: Annals of Regional Science, Online First, DOI 10.1007/s00168-007-0207-6

Niebuhr A (2001) Convergence and the Effects of Spatial Interaction. In: Jahrbuch für Regionalwissenschaft, 21: 113-133

Openshaw S (1984) The Modifiable Areal Unit Problem, Norwich

Openshaw, S/Taylor P J (1979) A million or so correlation coefficients: three experiments on the modifiable areal unit problem, in: Wrigley, N (ed.) Statistical Applications on the Spatial Sciences, Norwich, 127-144

Ord J K, Getis A (1995) Local Spatial Autocorrelation Statistics: Distributional Issues and an Application. In: Geographical Analysis, 27, 4: 286-306

Patacchini E, Rice P (2007) Geography and Economic Performance: Exploratory Spatial Data Analysis for Great Britain. In: Regional Studies 41, 4: 489-508

Roback J (1982) Wages, rents, anf the quality of life. In: Journal of Political Economy 90, 6, 1257-1278

Rosen S (1979) Wages-based indexes of urban quality of life, in: Mieszkowski P, Straszheim M (eds) Current issues in urban economics. Baltimore

Suntum U van/Rusche, K (2007) Regionale Beschäftigung und demografischer Wandel. in: Wirtschaftsdienst, 1/2007, 48-53

Tobler W R (1979) Cellular Geography. In: Gale S, Olsson G (Hrsg.) Philosophy in Gegography: 379-386

Uhde N (2008) Validity of Scoring Methods in the Presence of Outliers. In: SSRN Working Paper Series, No. 1113049

Appendix

Table 4: Results for Local Moran

RAM-ID	RAM-Name	QOL	Local Moran	Cluster	pseudo-p-value
1	Flensburg	142.5	0.469	LL	0.179
2	Kiel	146	0.435	LL	0.167
3	Lübeck	164.5	0.049	**LL**	**0.023**
4	Dithmarschen	141.5	0.836	LL	0.052
5	Nordfriesland	172	-0.418	**HL**	**0.001**
6	Hamburg	166.85	-0.058	**HL**	**0.014**
7	Braunschweig	149.25	0.840	**LL**	**0.009**
8	Wolfsburg	150.25	0.649	**LL**	**0.027**
9	Göttingen	134.66	1.665	**LL**	**0.003**
10	Goslar	161	0.286	**LL**	**0.001**
11	Osterode	137	1.520	**LL**	**0.004**
12	Hannover	150.85	0.522	LL	0.058
13	Lüneburg	152.33	0.336	LL	0.134
14	Emden	136.66	1.710	**LL**	**0.002**
15	Oldenburg	147	0.958	**LL**	**0.007**
16	Osnabrück	158	0.156	LL	0.187
17	Wilhelmshaven	140.33	1.503	**LL**	**0.002**
18	Emsland	151	0.401	LL	0.104
19	Vechta	154	0.461	**LL**	**0.034**
20	Bremen	144.14	1.155	**LL**	**0.005**
21	Bremerhaven	137	1.374	**LL**	**0.010**
22	Düsseldorf	163.6	0.034	LL	0.240
23	Duisburg	158	0.195	LL	0.127
24	Essen	152.25	0.306	LL	0.161
25	Krefeld	158.66	0.181	LL	0.121
26	Wuppertal/Hagen	153.5	0.283	LL	0.153
27	Aachen	159.5	0.056	LL	0.362
28	Köln/Bonn	175.12	-0.203	HL	0.180
29	Gummersbach	160	0.059	LL	0.338
30	Münster	165.75	-0.003	HL	0.129
31	Borken	153	0.297	LL	0.149
32	Bielefeld	160.66	0.131	LL	0.122
33	Höxter/Paderborn	145.66	0.493	LL	0.135
34	Minden	155	0.255	LL	0.145
35	Dortmund	152.33	0.364	LL	0.105
36	Hochsauerlandkreis	160	0.160	LL	0.099
37	Märkischer Kreis	157	0.222	LL	0.132
38	Siegen	159.66	0.135	LL	0.156
39	Soest	155	0.187	LL	0.225
40	Darmstadt	181.66	0.851	**HH**	**0.013**
41	Frankfurt am Main	183.11	0.590	HH	0.072
42	Wiesbaden	195	0.911	HH	0.083
43	Gießen	153.33	0.130	LL	0.330
44	Marburg	156	0.105	LL	0.330
45	Kassel	143.5	1.256	**LL**	**0.003**
46	Fulda	158	0.431	**LL**	**0.003**
47	Hersfeld	142	1.112	**LL**	**0.014**
48	Waldeck	152	0.516	**LL**	**0.042**
49	Koblenz	163.42	0.017	LL	0.381
50	Birkenfeld	134	0.661	LL	0.166
51	Trier	162	0.112	LL	0.089

52	Daun	156	0.178	LL	0.214
53	Kaiserslautern	153.25	0.309	LL	0.131
54	Ludwigshafen	173.83	0.112	HH	0.277
55	Mainz	169.6	0.033	HH	0.346
56	Pirmasens	140	-0.010	LH	0.491
57	Stuttgart	199.16	1.082	HH	0.078
58	Heilbronn	172.75	0.321	**HH**	**0.029**
59	Schwäbisch Hall	170	0.117	HH	0.124
60	Main-Tauber	164	-0.043	LH	0.132
61	Heidenheim	165	-0.018	LH	0.107
62	Ostalb	176	0.345	HH	0.075
63	Karlsruhe	203	0.232	HH	0.370
64	Mannheim	187.5	0.863	**HH**	**0.044**
65	Pforzheim	185	1.161	**HH**	**0.007**
66	Freudenstadt	197	1.326	**HH**	**0.034**
67	Freiburg	197.66	1.580	**HH**	**0.019**
68	Ortenau	185	1.257	**HH**	**0.003**
69	Rottweil	186	1.082	**HH**	**0.015**
70	Schwarzwald-Baar	182	0.831	**HH**	**0.014**
71	Tuttlingen	194	0.860	HH	0.088
72	Konstanz	188	1.006	**HH**	**0.029**
73	Lörrach	190	1.304	**HH**	**0.012**
74	Waldshut	176	0.695	**HH**	**0.004**
75	Reutlingen	184.5	0.792	**HH**	**0.039**
76	Zollernalb	168	0.120	**HH**	**0.017**
77	Ulm	184.33	0.669	HH	0.061
78	Biberach	194	1.120	**HH**	**0.047**
79	Bodenseekreis	194	1.369	**HH**	**0.020**
80	Ravensburg	181	0.892	**HH**	**0.008**
81	Sigmaringen	160	-0.336	**LH**	**0.006**
82	Ingolstadt	177.5	0.196	HH	0.224
83	München	218.16	3.167	**HH**	**0.003**
84	Rosenheim	184.33	0.950	**HH**	**0.020**
85	Altötting	166.5	0.015	HH	0.220
86	Garmisch-Partenkirchen	225	3.993	**HH**	**0.002**
87	Traunstein	194.5	1.307	**HH**	**0.025**
88	Landshut	176.5	0.091	HH	0.344
89	Passau	170.66	-0.039	HL	0.391
90	Straubing	162	0.003	LL	0.498
91	Deggendorf	162.5	0.010	LL	0.456
92	Dingolfing	166	0.002	HH	0.387
93	Amberg	143	0.533	LL	0.144
94	Regensburg	164.75	0.011	LL	0.298
95	Weiden	142.33	1.083	**LL**	**0.012**
96	Cham	157	0.140	LL	0.244
97	Bamberg	166	-0.013	HL	0.055
98	Bayreuth	143.33	1.298	**LL**	**0.004**
99	Coburg	135.2	1.861	**LL**	**0.001**
100	Hof	134.5	1.892	**LL**	**0.001**
101	Ansbach	165.5	-0.002	LH	0.229
102	Nürnberg	177.41	-0.033	HL	0.466
103	Aschaffenburg	190.66	0.889	HH	0.063
104	Schweinfurt	129.4	0.978	LL	0.107

105	Würzburg	177.25	-0.034	HL	0.466
106	Augsburg	181.33	0.306	HH	0.193
107	Kaufbeuren	172	0.441	**HH**	**0.001**
108	Kempten	199	2.271	**HH**	**0.002**
109	Memmingen	181.5	0.668	**HH**	**0.037**
110	Günzburg	169	0.143	**HH**	**0.042**
111	Donau-Ries	171	0.088	HH	0.231
112	Saarbrücken	164.83	0.028	LL	0.054

bold: significant at 0.05