

Roos, Michael

Conference Paper

How important is geography for agglomeration?

42nd Congress of the European Regional Science Association: "From Industry to Advanced Services - Perspectives of European Metropolitan Regions", August 27th - 31st, 2002, Dortmund, Germany

Provided in Cooperation with:

European Regional Science Association (ERSA)

Suggested Citation: Roos, Michael (2002) : How important is geography for agglomeration?, 42nd Congress of the European Regional Science Association: "From Industry to Advanced Services - Perspectives of European Metropolitan Regions", August 27th - 31st, 2002, Dortmund, Germany, European Regional Science Association (ERSA), Louvain-la-Neuve

This Version is available at:

<https://hdl.handle.net/10419/115863>

Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

Terms of use:

Documents in EconStor may be saved and copied for your personal and scholarly purposes.

You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.

If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.

How important is geography for agglomeration?

Michael Roos

University of Dortmund

Department of Economics

M.Roos@wiso.uni-dortmund.de

Abstract

The economic geography literature distinguishes between two types of reasons for economic agglomeration. Regional concentration of economic activity can be attributed to "first nature" meaning geographic advantages and disadvantages given by nature or to "second nature" meaning agglomeration economies by the interaction of economic agents. Several recent studies tried to estimate the relative importance of the two types of explanation. Most of these studies seem to exaggerate the importance of natural advantages because of loose definitions of geography. We describe geography by a small set of non-economic variables and estimate their importance for agglomeration in Germany. We find that about one third of the agglomeration of economic activity can be attributed to geography.

Keywords: economic geography, first nature, second nature, agglomeration

JEL codes: R12, O52

1 Introduction

Economic activity such as consumption and production is not evenly distributed in space. Instead, both households and firms agglomerate at certain locations with the consequence that some locations are densely populated while the population density in others is very low. There are two fundamentally different approaches to explain this phenomenon of agglomeration. The first and most obvious type of explanation is that regions with a large population and many firms must have some "natural advantage" relative to other regions of the same size. According to this reasoning, nature endows all places with specific features, some of

which favor and others which hamper life and production at that particular place. High or low population densities and the presence of certain industries can then be explained by an accidental accumulation of favorable or unfavorable natural features. Favorable characteristics of a location might be the availability of fertile soil, mineral resources, and navigable rivers, whereas unfavorable ones are, for instance, a mountainous surface, an unhealthy climate or geographic remoteness.

The second approach to explain agglomerations is less straightforward. It poses that agglomerations are due to "agglomeration economies". What sounds like a tautology means that for some reasons it must be advantageous for households and firms to be there where many other households and firms are, more or less irrespective of the particular geographic location. What matters in these theories is the physical interaction between economic agents among themselves rather than the interaction between agents and nature. Agglomeration advantages might arise because of knowledge and information spillovers, economies of intraindustry specialization, labor market economies or economies of scale in industry-specific public services (Richardson 1995). Another reason for agglomeration economies are market size effects, which recently regained a lot of popularity in Krugman's New Economic Geography (Fujita et al. 1999). One salient feature of the New Economic Geography is that it completely abstracts from natural advantages and disadvantages, which Krugman labels "first nature" (Krugman 1993). Krugman's theory shows that agglomerations can be explained by "second nature" alone, i.e. by man-made agglomeration economies due to increasing returns to scale and transportation costs, for instance.

Of course, both types of causes are complements rather than substitutes in explaining real-world agglomerations. This has already been seen by Alfred Marshall (Kim 1999). One important question is how much of the observed agglomeration of population and industries can be attributed to each type. Several empirical studies tried to estimate the relative importance of "first nature" in explaining agglomerations¹ (Ellison/Glaeser (1999), Kim (1999b), Rosenthal/Strange (2001) Gallup et al. (1999)). Using U.S. state level data, Ellison/Glaeser (1999) regress a measure of agglomeration in the 4-digit manufacturing industries on several sets of variables designed to measure cost advantages. Among these are labor costs and labor qualification and the size of the consumer market. They find that "at least half of the observed geographic concentration is due to natural advantages" (p. 316). Kim (1999) regresses

¹The term "first nature" does not appear in the studies cited. We use it as a summary term for everything that does not fall under Krugman's definition of "second nature" as given above.

value added in twenty 2-digit manufacturing industries on factor endowments U.S. states, i.a. on labor and capital. His result is that "[t]he unweighted average of the adjusted- R^2 for the twenty industries are 0.86 in 1880 and 0.83, 0.78 and 0.74 in 1900, 1967 and 1987 respectively" (p. 8). One critical point is how these studies define "first nature". Ellison/Glaeser (1999) admit that they "use the term 'natural advantage' fairly broadly" (p. 311). The problem is that neither the regional endowments with mobile factors such as labor and capital nor the prices of these factors are really exogenous. If, for example, a region has a large endowment of skilled labor and thus a low relative price of this factor, this is not really a natural advantage that can be attributed to "first nature". There might be a reverse causation running from the presence of a particular industry in a region to the region's endowment with labor or capital. Apart from the econometric endogeneity problems, such an estimation does not really tell us anything about the relative importance of "first" and "second nature".

This problem is partly solved by the approach of Gallup et al. (1999). They analyze how income levels, income growth, and population density on a global scale are related to "geography". For all countries with more than one million inhabitants, the authors perform growth regressions, in which they use variables describing the location, the spatial distribution of the population, and the prevalence of malaria in addition to some standard variables. They find that most of these variables are significant and improve the fit of the baseline growth regression without geography variables. Again, it can be objected that neither the population distribution within a country nor the prevalence of malaria are exogenous. In their regression of population density on geography, however, most regressors seem exogenous. Distances to the coast and waterways and several measures of elevation, soil quality, availability of water and climate serve as independent variables. In the international sample used² these factors explain 73% of the observed variability of the population density³. Yet it might be that this estimation exaggerates the importance of first nature. First, Gallup et al. use a large number of regressors, i.a. 36 dummies for ecozones and six dummies for elevations. Second and more importantly, on a global scale one would expect a priori that geography explains much of the uneven distribution of population. Many of the world's regions are not simply unfavorable but even hostile for human settlements, such as deserts, polar and boreal areas or very high altitudes. On the other hand, very high popula-

²For the population density regression, Gallup et al. divide the Earth's land areas in about 14,000 1° by 1° cells.

³Measured as unadjusted R^2 .

tion densities, especially if they are observed in single cities or whole countries might be rather due to historical reasons than due to specific geographic advantages. An explanation for the fact that, especially in Africa or the Americas, many large cities lie at the coast could be that European settlers (or conquerors) founded cities where they first arrived (see Ades/Glaeser, 1995, who discuss the case of Buenos Aires). And the high population density in China and maybe other countries as well might be due to institutional particularities rather than (or in addition to) geography (see Landes, 1998, ch. 2). Although Gallup et al. (1999) is an interesting study it does not help to assess the relative importance of the competing explanations for agglomeration, since not only the geographic, but also the institutional, historical, cultural, and economic conditions are too diverse in the large, aggregated sample.

In this paper, we also regress agglomeration measures on geography variables. However, we address the aforementioned problems in several ways. First, we choose only independent variables for which a good case of exogeneity can be made. Second, we restrict the number of regressors strongly in order to preclude overfitting. Third, we only look at the agglomeration of population and industry only within one country (Germany) and not across countries all over the world in order to control for potentially very different socio-economic conditions. Our ultimate goals are to assess the relative importance of first nature relative to second nature and to learn how the two are interrelated. In order to achieve this, we analyze how much of the observed variance can be explained by a predefined set of geography variables and check which of these variables are significant. This gives us an indication how first nature might matter and how much is left to second nature. Accordingly, we regard this piece of work not as a test of one theory against another but rather as explorative work levelling the path to further theoretical work.

In Section 2, we discuss our approach in greater detail, explain the choice to variables, and describe the data. In Section 3, we present and discuss the results. We conclude the paper in Section 4.

2 Method and data

In simple OLS estimations, we regress several measures of agglomeration on a number of variables capturing geographic features which can be expected to favor or to hamper the evolution and existence of population or industry concentrations. In this section, we describe the regions used in this analysis, give some descriptive statistics of the independent

variables⁴, and motivate the choice of regressors. We also discuss some econometric issues.

2.1 Regions

For the analysis, we use recent data (1994 and 1997) of 97 German regions. We use data on the level of "Raumordnungsregionen". These regions are aggregates of the 443 counties (Kreise). They are functional entities aimed to provide comparable information to political decision makers⁵. Table 1 provides some summary data on the aggregate level of states (Bundesländer). Although the regions are designed to be comparable, there is remarkable variation in their area. This is due to the need not to make the functional units larger than political entities. The table shows that four regions are identical with states: Berlin, Bremen, Hamburg, and Saarland. The first three are city states which have only small territories and are not likely to be functional units, since there is a lot of commuting between them and the nearby regions (BfLR 1996). Another particularity is that, with an average size of 4895 km², regions in East Germany (the former GDR) are larger than their counterparts in the West with only 3542 km² on average.

Table 1: Aggregated data on population, value added, and regions

⁴More information about the variable definitions and data sources is provided in the appendix.

⁵In the EUROSTAT system of European territories, "Raumordnungsregionen" are in between the levels NUTS-3 and NUTS-2. For a description of the regions, see BfLR (1996).

State	Population 1997 1000	Value added 1994 Mio. DM	Regions per state #	Ø Area of region km ²
Baden-Württemberg	10397	461148	12	2979
Bayern	12066	540233	18	3920
Berlin	3426	139300	1	889
Brandenburg	2573	59488	5	5896
Bremen	674	30589	1	327
Hamburg	1705	119351	1	755
Hessen	6032	311197	5	5811
Mecklenburg-Vorpommern	1808	39404	4	5793
Niedersachsen	7845	291073	13	3648
Nordrhein-Westfalen	17975	708662	13	2620
Rheinland-Pfalz	4018	139324	5	3969
Saarland	1081	39780	1	2570
Sachsen	4522	100811	5	3682
Sachsen-Anhalt	2702	60000	4	5111
Schleswig-Holstein	2757	100674	5	3146
Thüringen	2478	53677	4	4044
Germany	82057	3194711	97	3681

Source: BBR 1999, Statistik regional 1998

2.2 Agglomeration

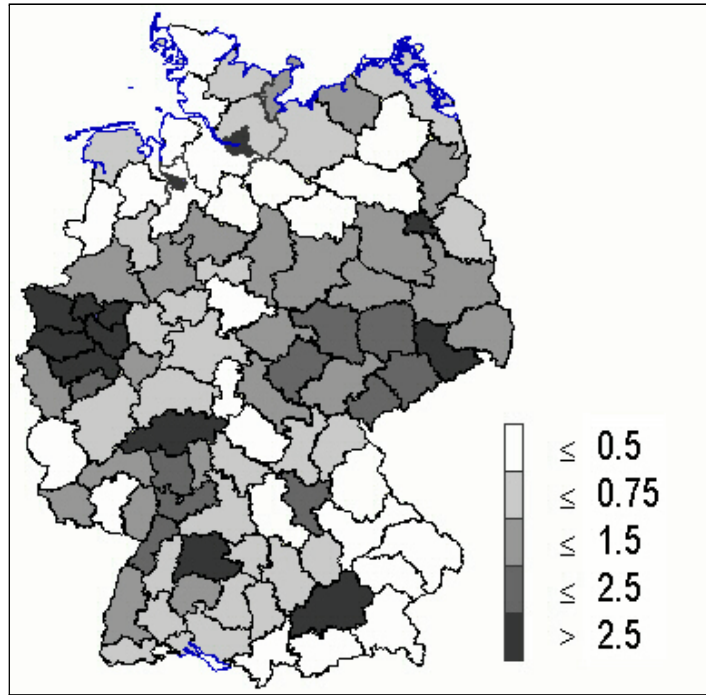
As a measure of the agglomeration of economic activity we use value added per square kilometer. Table 2 shows that value added per square kilometer is much lower in the regions that belonged to the former GDR than in the western regions. In order to obtain a measure of agglomeration which is independent of scale, we divide each region's value added per km² by the average value added density in the West and in the East respectively.

Table 2: Value added per km² in 1994

	all $\frac{\text{MioDM}}{\text{km}^2}$	west $\frac{\text{MioDM}}{\text{km}^2}$	east $\frac{\text{MioDM}}{\text{km}^2}$	west w/ city $\frac{\text{MioDM}}{\text{km}^2}$	city $\frac{\text{MioDM}}{\text{km}^2}$
Mean	13.74	16.83	3.20	11.87	135.99
Stdd	24.91	27.59	1.84	11.41	36.77
Med	6.34	7.57	2.66	7.38	156.34
Min	0.97	2.37	0.97	2.37	93.54
Max	158.08	158.08	7.85	57.69	158.08
Obs	97	75	22	72	3

Source: Statistik regional 1998

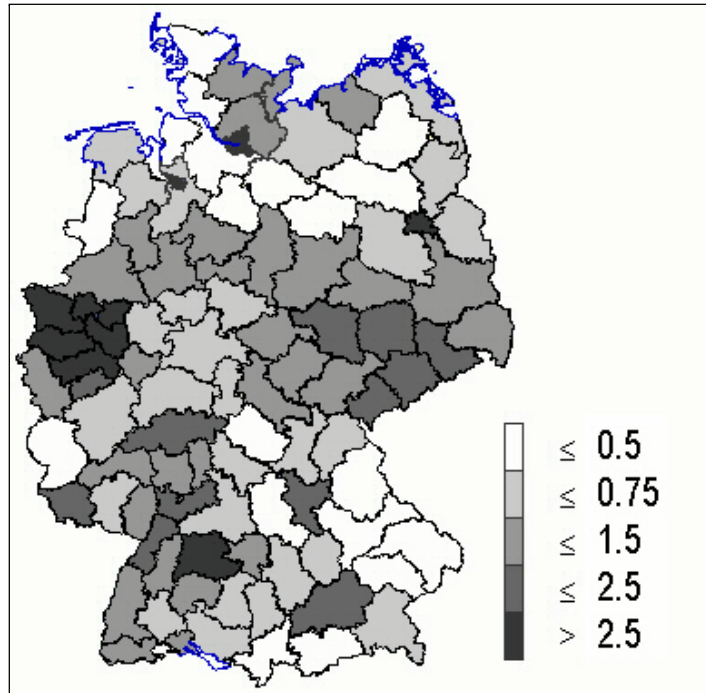
Thus we get the relative value added density with respect to the average density of that part of Germany to which the regions belongs. Figure 1 shows the result. Not surprisingly the city states have the highest value added density because of their small areas.



Relative value added per km² in 1994

Except for the city states one can recognize four economic centers in the western part and one in the eastern part of Germany. In the West, economic activity is heavily agglomerated in the Rhein-Ruhr area, the Rhein-Main area, and the regions around Stuttgart and Munich. In East Germany, the value added density is highest in the south with a high concentration in the Dresden area. Roughly speaking, the density is very low with less than half of the average in the North and in the South.

It is interesting to compare the concentration of value added with the spatial concentration of population. Figure 2 shows the relative population density. The overall pattern is very similar. Accordingly the raw correlation between the relative value added per km² and the relative population density is about 96%. However, population is a bit less concentrated than value added. There are both fewer regions with a population density less than half of the average and fewer with a density larger than 2.5 time the average.



Relative population density in 1997

Of course, this similarity between the patterns is not random. There are close linkages between the distributions of population and economic activity. Households and firms interact on product and labor markets and if these markets are spatially segmented, we expect that economic activity takes place where people live and vice versa. An interesting question is whether the location of households determines the distribution of firms or whether households reside where firms are. Actually, it is probably impossible to disentangle the location choices neatly, but conceptually it might be useful to have an idea how the causality runs in a specific context. In Germany, there are no different ecozones that make a particular place especially favorable or unfavorable to live there. On the contrary, all over the country a moist, temperate climate prevails. Accordingly, the spatial distribution of the population is not predetermined by the interaction of climate and geography as it is in other countries. For example in dry and hot countries, there is a natural reason for humans to live at rivers and in the tropics living in high elevations is healthier (more comfortable) than in low ones. Thus the population distribution in Germany is likely to be determined either by the past or current distribution of production. We therefore believe that if we find that geography has an influence on the spatial distribution of economic activity, it works through cost and other advantages for firms.

In order to explore the relation between geography and economic activity a bit more in detail, we also look at each region's sectoral distri-

bution of employment. This might give us an indication how geography influence the regional specialization. Unfortunately, there is a tradeoff between disaggregation in the geographic and in the sectoral dimension. At fine levels of geographic disaggregation, the employment data are very incomplete so that we only use the division of the economy in five broad sectors. We use the share of employment in each of the five sectors agriculture, forestry, and fishery ("agriculture"), mining, industry, and construction ("industry"), trade, transportation, and telecommunications ("trade"), other services ("services"), and government, private households, and nonprofit organizations ("nonprofit"). Table 3 describes the data.

Table 3: Summary statistics of employment shares in 1996

	agriculture	industry	trade	services	nonprofit
	sector	share	of	all	employed
Mean	0.037	0.353	0.185	0.211	0.215
Stdd	0.021	0.061	0.026	0.036	0.036
Med	0.036	0.354	0.182	0.208	0.205
Min	0.005	0.195	0.129	0.145	0.124
Max	0.097	0.495	0.262	0.353	0.329
Obs	97	97	97	97	97

Statistik regional 1998

2.3 Geography

Although many geographic features such as soil and water quality, prevailing wheather conditions, or fauna and flora might influence the spatial distribution of households and firms, we restrict the number of variables describing "geography" to those we regard as the most important ones. Roughly speaking, these can be summarized as resource endowment and location. The underlying idea is that geography affects the location choice of firms through its effect on transportation costs to input and to output markets. This is in line with the traditional location theory based on the work of Weber (see Schätzl 1998) and Krugmans New Economic Geography. However, location matters not only in the physical geography but also in the political geography. Closeness to political power is sometimes seen as an advantage for firms (Ades/Glaeser (1995), Funck (1998)). Firms, industries, and individuals can influence the political process better if they are close to political decision makers. On the other hand, a national border is often seen as a disadvantage for the local economy for several reasons (McCallum (1995), Mathias (1980)). In the following, we will present the variables chosen to capture these effects and derive hypotheses on the expected signs of the

coefficients to be estimated.

Since agriculture is a fairly small sector in the German economy, the only natural resources we look at are mineral resources. Germany is poorly endowed with minerals, with coal as the only exception. Both bituminous coal and lignite are available in large quantities. In 1994, Germany was the world's largest producer of lignite (207 Mio. tons, compared to 90 Mio. tons in Russia and 75 Mio. tons in the United States, Fischer 1996) and the ninth largest producer of bituminous coal. Although much of the coal is used for the generation of electricity, coal and coal products such as tar and gas are also important inputs to the steel and the chemical industry. Therefore, the availability of coal in a region is an advantage for these industries because they can economize on transportation costs when they are close to the coal mines. In our regressions, we use two dummies, *bit* and *lig*, that indicate the presence of bituminous coal or lignite in a region. We prefer dummies to output quantities or other possible variables because we want to measure the exogenous **endowment**, not the endogenous **output**. We expect the dummies to have positive effects on agglomeration.

We examine three features of the physical geography that might have an effect on agglomeration. First, we use the dummy *sea* to indicate if a region lies at the seashore. Gallup et al. (1999) find that the seashore is a locational advantage because of the importance of ports for international commodity trade. However, not every coastal region also has a port and without a port, the seashore is a natural frontier restricting the mobility of goods and factors. The total effect seems ambiguous to us. Second, we examine if rivers favor agglomeration. Historically, rivers were the cheapest and fastest ways for the transportation of goods and even today bulk goods are often carried on waterways. Since Germany has a dense net of rivers and canals and our regions are relatively large, we concentrate on the largest rivers only. Otherway, almost every region would get a river dummy. The dummy *river* is one only if the region has a river which is navigable for boats of more than 1350 tons. The river dummy should have a positive coefficient. Our last variable describing the physical geography is *alt* measuring the highest altitude in a region. Although it would be preferable to use the average instead of the maximum altitude, it is easier to get data for the latter. However, we think that a region's maximum elevation is a good proxy for the average one because the highest hills and mountains rarely stand alone in otherwise flat area, but are mostly parts of mountain ranges. We would expect elevation to be negatively correlated with agglomeration because mountains both make transportation costly and limit the space available for settlements and industrial production.

The last group of variables broadly describes the political geography. Germany is a federation of states (Bundesländer) which have quite a lot of legislative and executive power which is geographically concentrated in the capital of each state. For several reasons it might be beneficial for firms to locate in or close to capitals, e.g. in order to influence the legislative process, to have direct access to information about government investment plans or to provide goods and services to the public sector. Therefore, we expect the coefficient of the capital dummy `cap` to have a positive sign. Another characteristic of the political geography is the national frontier. McCallum (1995) shows that the Canada-U.S. border is a significant hindrance to trade despite the similarity of these countries in terms of language, culture, and institutions. The same might be true for Europe, where these factors differ considerably between almost all countries. These differences might also restrict factor mobility across the frontier even if it is allowed as in the countries of the European Union. But border regions might be disadvantaged for other reasons as well. Mathias (1980) argues for the Saarland region that before and after the Second World War investment in infrastructure has been lower than in non-border regions for strategic reasons and that this underprovision of infrastructure was harmful for regional development long after the war. Similar arguments might be plausible for other border regions, too. Therefore, the border dummy `bord` should have a negative coefficient. Our last variable does not really belong neither to the physical nor to the political geography. We nevertheless suppose that the geographical location of a region might be relevant. Activities for which transport costs and/or travel time are important might be concentrated in centrally located regions as hubs of logistics or administration. We describe the location of a region by the average distance to all other regions meaning that a region is centrally located if the average distance is small. The coefficient of the variable `loc` is expected to be negative.

2.4 Econometric issues

Although the econometric model is very simple, there are some potential problems to be considered. In order to avoid multicollinearity, we must be careful which regressors to use together in our regressions. Table 4 shows the correlations between the independent variables. Not surprisingly, both `loc` and `alt` are highly correlated with `bord` and `sea`. Therefore, we include either `loc` and `alt` or `bord` and `sea` in the regression, but not both pairs.

Table 4: Correlation matrix of independent variables

	loc	alt	bord	sea	riv	bit	lig	cap	east
alt	.087	1							
bord	.494	.327	1						
sea	.411	-.372	-.064	1					
riv	-.105	-.115	-.053	.003	1				
bit	-.053	-.133	.143	-.105	-.153	1			
lig	-.041	.097	.086	-.154	.073	-.115	1		
cap	-.011	-.024	-.101	.024	.133	-.001	.082	1	
east	.097	-.143	.039	.021	-.004	-.151	.268	.128	1
city	.056	-.184	-.125	.295	.092	-.050	-.073	.096	-.097

The second potential problem is heteroscedasticity because of the different sizes of the regions. We check for heteroscedasticity using the White test and estimate White’s heteroscedasticity-consistent standard errors. We also perform the Jarque-Bera test on normality of the estimation residuals.

Another problem related to the size of regions is that outliers might influence the estimation results strongly. The city states are likely to bias the results because they have very high value added densities simply due to their small areas. We correct for this using the city state dummy *city*. Although the Eastern regions are larger than the ones in the West (see Table 1), which might be one reason for their lower population and value added density, this does not cause a problem here because we correct for the lower densities in the east by using the relative densities anyway. We nevertheless introduce the dummy *east* in our regressions with the sectoral employment shares in order to correct for potential systematic differences due to the past of the East German economy.

3 Results

3.1 Value added

Table 5 contains our main results⁶. In columns (1) and (2), we have the results of our basic regressions with all regressors that are not strongly correlated. In both specifications, the R^2 of 0.55 is fairly high for such simple regressions. Geography as described by seven variables, of which at least five are dummies, explains more than half of the observed variation of the relative value added density in German regions.

Table 5: Influence of geography on value added density

⁶The results of regressions with the population density as dependent variable (not shown) are very similar. This is not surprising due to the high correlation between population density and value added density.

	(1)	(2)	(3)	(4)	(5)	(6)
	$\ln(\frac{vva}{km^2})$	$\ln(\frac{rva}{km^2})$	$\ln(\frac{rva}{km^2})$	$\ln(\frac{rva}{km^2})$	$\ln(\frac{rva}{km^2})$	$\ln(\frac{rva}{km^2})$
const	3.284*	-0.452***	4.535***	-0.431***	4.528***	-0.428***
	(1.945)	(0.085)	(1.701)	(0.075)	(1.703)	(0.075)
city	2.843***	2.769***	2.729***	2.856***		
	(0.192)	(0.245)	(0.199)	(0.285)		
bit	1.086***	1.069***	1.001***	0.976***	1.001***	0.966***
	(0.257)	(0.242)	(0.253)	(0.270)	(0.253)	(0.273)
lig	0.368*	0.374*	0.392*	0.364	0.390*	0.350
	(0.210)	(0.206)	(0.211)	(0.220)	(0.210)	(0.218)
riv	0.176	0.179				
	(0.127)	(0.130)				
sea		-0.333**		-0.342**		-0.409***
		(0.162)		(0.150)		(0.156)
ln(alt)	0.066					
	(0.050)					
cap	0.666***	0.668***	0.714***	0.719***	0.738***	0.773***
	(0.188)	(0.188)	(0.191)	(0.184)	(0.200)	(0.189)
bor		-0.157				
		(0.141)				
ln(loc)	-0.734**		-0.870***		-0.869***	
	(0.315)		(0.293)		(0.292)	
R ²	0.55	0.55	0.54	0.53	0.35	0.34
F	65.29	29.49	53.55	26.93	12.44	9.43
# Obs.	97	97	97	97	94	94
p(JB)	0.470	0.446	0.366	0.384	0.355	0.409
p(White)	0.167	0.789	0.150	0.496	0.078	0.314

White standard errors in parenthesis; asterisks designate significance levels: * 10%, ** 5%, *** 1%; p(JB): empirical significance level of the Jarque-Bera test statistic; p(White): empirical significance level of the White test statistic

The Jarque-Bera test and the White test indicate that the standard errors of the coefficient estimates are reliable. Of the resource variables, the dummy for bituminous coal is highly significant and has a large positive coefficient. Lignite is also positive but far less precisely estimated. The estimates in columns (1) and (2) are very similar. The only statistically significant variable describing the physical geography is the seashore dummy. It has a negative coefficient and is significant at 5%. Elevation and waterways are not relevant. In our last group of variables, only the border dummy is not significant. The capital dummy coefficient is precisely estimated and positive, as expected. The location measure coefficient is negative and significant at the 5% level. To sum up our first results, all coefficients have the expected signs, except for the elevation

measure, which is not significant, however. In contrast to the result in Gallup et al. (1999), we find that the seashore has a negative influence and rivers have no impact on the agglomeration of economic activity.

In order to see whether the inclusion of the insignificant variables influences the estimation of the other coefficients, we run the regressions again without those variables. The results in columns (3) and (4) show, that the estimates are quite stable. Except for `bit` all estimates increase in absolute value. In column (4), `lig` becomes insignificant. The location measure estimate is now reliable with an error probability of 1%.

As a second robustness check, we drop the city state observations from our sample because they might influence the results strongly (see Table 2). Columns (5) and (6) contain the results of the regressions without the city state observations and without the insignificant variables from the first regressions. The most important result of this exercise is that the R^2 drops to about one third. Since we included the city state dummy as a control and not as an explanatory variable, one third is a better estimate of geography's impact on agglomeration than one half. The estimates in (5) are almost identical to the ones in (3). The major differences are that `sea` and `cap` again are higher in absolute value in column (6) relative to (4). By far the most important variable is the bituminous coal dummy. Bituminous coal deposits increase a region's relative value added density by about 170%⁷ and a state capital by more than 100%. The seashore decreases the value added density by about 30% and a one percent increase in the average distance to all other regions leads to a decrease of approximately 0.87 percent.

The results of this section can be summarized as follows: Geography explains about one third of the observed variation in relative value added density in Germany. The single most important natural advantage is endowment with bituminous coal, whereas Germany's other important natural resource, lignite, does not favor agglomeration considerably. The only important feature of the physical geography the seashore, which is a disadvantage. Rivers or mountains do not influence the agglomeration of economic activity in general. Agglomeration is higher in regions with a state capital, but it is not affected by national frontiers. Agglomeration is lower if a region is less centrally located.

3.2 Employment

In order to explore the influence of geography on agglomeration in greater detail, we repeat our regressions with the sectoral employment shares as

⁷In semilogarithmic equations, the dependent variable changes by $(exp(\beta)-1)*100$ percent if the dummy changes from zero to one, where β is the dummy coefficient.

dependent variables⁸. This exercise shows whether certain geographic features cause regional specialization. Table 6 contains the results of the regression with the regressors `alt` and `loc` and Table 7 shows the estimates with `sea` and `bor`. In general, the results are quite similar, although the fit is slightly better in the regressions in Table 6. The R^2 lies between 0.27 and 0.52 in Table 6 and between 0.25 and 0.45 in Table 7. In both tables, the R^2 is lowest for the "nonprofit" sector and highest for "trade, transportation, and telecommunication". There is no case in which the sign of the coefficients of the statistically significant variables is different in both tables. The coefficients' signs of `alt` and `sea` are exactly opposite, as it should be, and the effects of `loc` and `bord` go into the same directions, too.

The employment share in "agriculture, forestry, and fishery" is about 2.4 percentage points lower in regions with bituminous coal and about 1 percentage point lower in regions with lignite deposits. Given that land is a major input in the agricultural sector, these findings are not surprising because the opportunity cost of using land as an input to agriculture rises if mineral resources can be extracted. It is also lower - by 1.5 percentage points - in capital regions, but higher in peripheral regions, as indicated by the positive coefficients of `loc` and `bord`. None of the physical geography variables `riv`, `alt`, and `sea` affects employment in agriculture in a systematic way.

Table 6: Influence of geography on sectoral employment

⁸In absolute levels, not in logs.

	(1)	(2)	(3)	(4)	(5)
	agriculture	industry	trade	services	nonprofit
const	-0.179*** (0.064)	0.462** (0.181)	0.424*** (0.062)	0.123 (0.097)	0.171 (0.106)
east	0.004 (0.004)	0.004 (0.010)	-0.017*** (0.004)	-0.014* (0.008)	0.023*** (0.008)
city	-0.036*** (0.006)	-0.035 (0.045)	0.019 (0.021)	0.071** (0.031)	-0.018 (0.025)
bit	-0.023*** (0.005)	-0.002 (0.016)	0.019*** (0.006)	0.017* (0.009)	-0.011 (0.008)
lig	-0.009* (0.005)	-0.011 (0.013)	0.012** (0.005)	0.015 (0.011)	-0.006 (0.008)
riv	<0.001 (0.004)	-0.010 (0.010)	0.010** (0.004)	0.014** (0.006)	-0.014* (0.007)
ln(alt)	-0.001 (0.002)	0.026*** (0.005)	-0.012*** (0.002)	-0.001 (0.003)	-0.011*** (0.003)
cap	-0.015*** (0.004)	-0.055*** (0.014)	0.013** (0.006)	0.038*** (0.010)	0.020** (0.009)
ln(loc)	0.040*** (0.010)	-0.043 (0.029)	-0.030*** (0.010)	0.014 (0.016)	0.019 (0.018)
R ²	0.37	0.41	0.52	0.40	0.27
F	12.52	8.37	12.70	5.73	6.19
# Obs.	97	97	97	97	97
p(JB)	0.026	0.891	0.826	0.225	<0.001
p(White)	0.970	0.958	0.635	0.024	0.999

White standard errors in parenthesis; asterisks designate significance levels: * 10%, ** 5%, *** 1%; p(JB): empirical significance level of the Jarque-Bera test statistic; p(White): empirical significance level of the White test statistic

Although the R^2 of the "industry" regressions are fairly high, only few regressors are statistically significant. Both tables show that employment in industry is about 5 percentage points lower in regions with a capital. In Section 2.3 we mentioned several arguments, why firms might want to locate in or close to capitals as centers of the political process. However, in order to exert influence on politicians or to have good access to information, it is not necessary to produce in capitals. It suffices to be there, e.g. with the headquarter. The headquarter of industrial firms, on the other hand, is likely to have only a relatively small part of the total employment. The only other variables that are statistically significant are the sea dummy and the elevation measure alt, which indicate that at the seashore the employment share in industry is less than in the inland. Interestingly, the resource dummies bit and lig are not significant. This is surprising because "industry" also comprises employment in mining.

However, only 1.8% of the employed subject to compulsory insurance contributions in the industry sector work in coal mining. In addition, we argued in Section 2.3 that coal deposits might matter for agglomeration because downstream industries could economize on transport costs if the located where the coal is mined. This argument is not supported by the "industry" regression. The strong correlation between value added density and coal deposits found in Section 3.1 cannot be explained by simple arguments creating a direct linkage between coal and industry production. At least, such arguments are not valid today, even though they probably were important in the past. In general, "industry" does not seem to be affected much by geography. In unreported regressions of the employment shares in all 2-digit industries almost no variable was significantly different from zero and our specification statistics indicated very poor model fits.

Table 7: Influence of geography on sectoral employment

	(1)	(2)	(3)	(4)	(5)
	agriculture	industry	trade	services	nonprofit
const	0.039*** (0.003)	0.381*** (0.009)	0.177*** (0.004)	0.196*** (0.005)	0.208*** (0.005)
east	0.006 (0.004)	-0.005 (0.011)	-0.013*** (0.004)	-0.014* (0.007)	0.026*** (0.008)
city	-0.030*** (0.007)	-0.055 (0.035)	0.030** (0.014)	0.064** (0.028)	-0.009 (0.032)
bit	-0.025*** (0.004)	-0.031 (0.020)	0.035*** (0.007)	0.021** (0.010)	0.001 (0.007)
lig	-0.011** (0.005)	-0.014 (0.012)	0.014*** (0.005)	0.017 (0.011)	-0.005 (0.008)
riv	-0.001 (0.004)	-0.015 (0.011)	0.013*** (0.004)	0.014** (0.006)	-0.012 (0.008)
sea	0.005 (0.005)	-0.082*** (0.014)	0.025*** (0.006)	0.017* (0.010)	0.035*** (0.011)
cap	-0.014*** (0.004)	-0.044*** (0.015)	0.006 (0.007)	0.036*** (0.011)	0.015 (0.009)
bor	0.010** (0.004)	0.008 (0.010)	-0.014*** (0.004)	-0.003 (0.006)	-0.001 (0.007)
R ²	0.31	0.37	0.45	0.42	0.25
F	10.96	6.41	14.53	5.52	6.96
# Obs.	97	97	97	97	97
p(JB)	0.123	0.916	0.541	0.387	<0.001
p(White)	0.918	0.968	0.548	0.048	0.912

White standard errors in parenthesis; asterisks designate significance levels:
* 10%, ** 5%, *** 1%; p(JB): empirical significance level of the Jarque-Bera

test statistic; $p(\text{White})$: empirical significance level of the White test statistic

The employment shares in the sector "trade, transportation, and telecommunication" are well explained by geography. All independent variables are significant at least in one of the two regressions, and most in both. The most unexpected result is that the employment share is positively correlated with coal deposits, and especially strongly with bituminous coal endowments. Since we cannot think of any direct linkage between coal mining and employment in the "trade" sector, we conclude that the results must be explained by another factor. As in the case with "industry" employment, it seems plausible that in the past, coal caused the agglomeration, which is still present in coal regions although coal itself is not important anymore. The population density in the German coal regions is almost twice as high as the average density of all regions (648 inhabitants per km^2 compared to 329 inhabitants per km^2)⁹. This high population density is likely to be the cause of the high employment share in the "trade" sector because of the need of retailers to be close to the final customer and wholesaler to be close to retailers. Urban transportation is also important in densely populated areas. The river dummy *riv*, which was not statistically significant in any of the value added regressions, has a positive effect on the "trade" employment share. This supports the hypothesis that rivers are still important transportation ways. The argument that waterways play an important role in commodity trade is also supported by the relatively large, positive coefficient of the *sea* dummy (and the negative one of *alt*). In capital regions, the "trade" employment share is higher as well, although this comes out only in one of the regressions. In the regression in Table 7, it might be that the variables *city*, *sea*, and *cap* interact because Berlin is a capital and the other city states Hamburg and Bremen have sea access. Notice that the *city* dummy is significant in Table 7 but not in Table 6. Finally, both *loc* and *bor* affect the "trade" employment share negatively. This is consistent with the hypothesis in Section 2.3.

For the "service" sector, the picture is similar to the "trade" sector with the exceptions that neither *lig* nor *bor* nor *loc* have coefficients different from zero. But now, in both tables the capital dummy is significant and has a large coefficient¹⁰. It means that capitals are not only political but also service centers. The explanation why the *bit* dummy is positive is probably related to the one given for the "trade" case, i.e. the high population density in these regions might be the reason. Again, the dummies *riv* and *sea* have positive coefficients. Together

⁹It is even larger than the density in capital regions, which is 629 $\frac{\text{inhabitants}}{\text{km}^2}$.

¹⁰In capital regions the employment share in the service sector is by one standard deviation higher than in non-capital regions.

with the observation that the city dummy is positive and very large, this suggests a strong linkage between the service sector and the trade sector. Historically, sites at rivers and/or with access to the sea have ever been commercial centers which specialized in services supporting long-distance trade like insurance and banking services.

It is not surprising that employment in the "nonprofit" sector is not well explained by geography¹¹. Employment in the public sector and in nonprofit organizations is determined by political considerations rather than by geographical features. Accordingly, at least in Table 6 the capital dummy seems to be positive. Why rivers should have a negative and the seashore should have a positive impact on employment in this sector is difficult to rationalize.

In this subsection, we analyzed the influence of geography on regional specialization rather than on agglomeration. However, the results point to ways how one could explain the results from the agglomeration regressions. The first interesting result is that industry as a whole does not seem to be positively influenced by geographical characteristics. None of the analyzed features favors a region in a way that it has a particularly large share of employment in the industrial sector. This is different for the "trade, transportation, and telecommunication" sector. Rivers, access to the sea, and central locations help a region to specialize in activities of this sector. This corresponds to the obvious explanation that geography affects the economy by transportation costs. It is less obvious why the results of the service sector regressions are so similar to the trade sector regressions. Exactly for the service sector we would not expect transport cost advantages to matter much. We interpret our results as evidence for strong linkages - historical and current ones - with the trade sector. The large coefficient of capitals, however, indicates that there are also other mechanisms leading to a regional concentration of services. It is very interesting that bituminous coal has a positive influence on the employment shares in "trade" and "services" but not in the other sectors. A priori we would not have expected coal regions to be specialized in trade or services but rather in industry. Our results suggest that the coal regions on a macro level underwent a structural change from the regions specialized in industry they were in the past to trade and service regions. Presumably, coal mining caused the agglomeration of population, which somehow remained stable even after the coal industry had lost its economic significance for other industries. This agglomeration of population might be the reason for the high employment shares of the

¹¹Notice that the Jarque-Bera test indicates that the null hypothesis of normally distributed residuals is rejected at the 0.1 percent error level. Therefore, the standard errors of the coefficients are hardly reliable.

"trade" and the "service" sector. The same reasoning might be valid for the relationship between agglomeration and geography. In the value added regressions, the bituminous coal dummy was by far the most important variable, which stands in sharp contrast the actual economic importance of coal to the German economy today.

4 Conclusions

We have analyzed the influence of geography on the agglomeration of economic activities. Looking at overall agglomeration instead of the agglomeration of certain industries at very disaggregated levels, we take a different approach to the ones present in the literature. We regress the relative value added density in each region on a set of geography variables and find that approximately one third of the variation of this agglomeration measure can be attributed to geography. Of our geography variables, bituminous coal, capitals, and a central location have positive effects on agglomeration, whereas the seashore's effect is negative. Lignite deposits, mountains, rivers, and national borders have no influence on agglomeration.

Using regional employment shares in five broad sectors, we find that industry location as a whole is hardly determined by geography. However, the employment shares in "trade, transportation, and telecommunication" and in "services" are well explained by geography, which explains between 40 and 52% of the observed variation. We regard the good fit of these regressions as evidence that both transport costs and history matter for the explanation of macro level agglomerations. In addition to these aspects that are important in the theoretical literature on agglomeration, we also find that political factors play an important role, which has been largely ignored by the theoretical literature. In almost all regressions, the capital dummy is significant at the 1% level.

The strong significance of bituminous coal in most of our regressions is another interesting result. The large coefficient estimates are likely to exaggerate the coal's current economic importance by far and seem to reflect its past significance. The same might be true, although to a smaller extent, for the river dummy in the "service" regressions. The specialization of river cities on services could be a heritage from pre-industrial times when long-distance trade relied heavily on waterways. Cities at big rivers had a natural advantage in trade and gained economic and political power as commercial centers.

The result that both geography and history matter for today's spatial distribution of economic activities will not surprise neither geographers nor historians. However, it is not reflected in most economic theories explaining agglomeration and regional specialization. Accordingly, these

theories cannot explain where agglomerations will arise and are likely to exaggerate the sensitivity of regions to major structural changes due to the change of some economic parameters. On the other hand, economic theories help to explain why agglomerations form at locations without any natural advantage and why agglomeration do not dissolve after the natural advantage has vanished. Our regression results suggest that "first nature" does not explain more than one third of the agglomeration of economic activities in Germany. This leaves ample room for "second nature" theories what most economic theories are.

References

- Ades, A. K., E. L. Glaeser (1995), "Trade and Circuses: Explaining Urban Giants", *Quarterly Journal of Economics* 110, 195-227.
- BfLR (1996): Bundesforschungsanstalt für Landeskunde und Raumordnung, "Neuabgrenzung von Raumordnungsregionen nach den Gebietsreformen in den neuen Bundesländern", Arbeitspapier 5/1996.
- Ellison, G., E. L. Glaeser (1997), "Geographic Concentration in U.S. Manufacturing Industries: A Dartboard Approach.", *Journal of Political Economy* 105, 889-927.
- Ellison, G., E. L. Glaeser (1999), "The Geographic Concentration of Industry: Does Natural Advantage Explain Agglomeration?", *American Economic Review Papers and Proceedings* 89, 311-316.
- Fischer Weltalmanach 1997 (1996), Fischer, Frankfurt.
- Fujita, M., P. Krugman, A. Venables (1999), "The Spatial Economy - Cities, Regions, and International Trade", MIT Press, Cambridge, MA.
- Funck, R. H. (1995), "Competition among Locations: Objectives, Instruments, Strategies, Perspectives", in: *Urban agglomeration and economic growth*, H. Giersch (ed.), Springer, New York and Heidelberg.
- Gallup, J. L., J. Sachs, J., A. Mellinger (1999), "Geography and Economic Development", *International Regional Science Review* 22, 179-232.
- Kim, S. (1999a), "Urban Development in the United States, 1690-1990", NBER Working Paper No. 7120.

- Kim, S. (1999b), "Regions, Ressources, and Economic Geography: Sources of U.S. Regional Comparative Advantage, 1880-1987.", *Regional Science and Urban Economics* 29, 1-32.
- Krugman, P. (1993), "First nature, second nature, and metropolitan location", *Journal of Regional Science* 33, 129 -144.
- Landes, D. (1998), "The wealth and poverty of nations: Why some are so rich and some so poor", Norton, New York.
- Mathias, K. (1980), "Wirtschaftsgeographie des Saarlandes", ???.
- McCallum, J. (1995), "National Borders Matter: Canada-US Regional Trade Patterns", *American Economic Review* 85, 615-623.
- Richardson, H. W (1995), "Economies and Diseconomies of Agglomeration", in: *Urban agglomeration and economic growth*, H. Giersch (ed.), Springer, New York and Heidelberg.
- Rosenthal, S., W. C. Strange (2001), "The Determinants of Agglomeration", *Journal of Urban Economics*, electronic version, doi:10.1006/juec.2000.2206, available online at [http: www.idealibrary.com](http://www.idealibrary.com).
- Schätzl, L. (1998), "Wirtschaftsgeographie 1 - Theorie", Schöningh, Paderborn.
- Venables, A. (1999), "But Why Does Geography Matter, and Which Geography Matters?", *International Regional Science Review* 22, 238-241.

Appendix

Data description and sources

Regions: The spatial unit of analysis is the "Raumordnungsregion".

We assigned a center to each region, which was either the central city ("Oberzentrum") as defined by the Bundesforschungsanstalt für Landeskunde und Raumordnung (see BfLR 1996) or the biggest city when no central city was available. The area data for the regions was obtained from the CD-ROM "INKAR - Indikatoren und Karten zur Raumentwicklung - Ausgabe 1999" of the Bundesamt für Bauwesen und Raumordnung. The 97 regions with their code numbers and their centers are: 1: Schleswig-Holstein Nord (Flensburg), 2: Schleswig-Holstein Süd-West (Itzehoe), 3: Schleswig-Holstein Mitte (Kiel), 4: Schleswig-Holstein Ost (Lübeck), 5: Schleswig-Holstein Süd (Elmshorn), 6: Hamburg (Hamburg), 7: Westmecklenburg (Schwerig), 8: Mittleres Mecklenburg/Rostock (Rostock), 9: Vorpommern (Stralsund), 10: Mecklenburgische Seenplatte (Neubrandenburg), 11: Bremen (Bremen), 12: Ost-Friesland (Wilhelmshaven), 13: Bremerhaven (Bremerhaven), 14: Hamburg-Umland-Süd (Stade), 15: Bremen-Umland (Delmenhorst), 16: Oldenburg (Oldenburg), 17: Emsland (Meppen), 18: Osnabrück (Osnabrück), 19: Hannover (Hannover), 20: Südheide (Celle), 21: Lüneburg (Lüneburg), 22: Braunschweig (Braunschweig), 23: Hildesheim (Hildesheim), 24: Göttingen (Göttingen), 25: Prignitz-Oberhavel (Neuruppin), 26: Uckermark-Barnim (Eberswalde), 27: Oderland-Spree (Frankfurt/Oder), 28: Lausitz-Spreewald (Cottbus), 29: Havelland-Fläming (Potsdam), 30: Berlin (Berlin), 31: Altmark (Stendal), 32: Magdeburg (Magdeburg), 33: Dessau (Dessau), 34: Halle/Saale (Halle/Saale), 35: Münster (Münster (Westfalen)), 36: Bielefeld (Bielefeld), 37: Paderborn (Paderborn), 38: Arnsberg (Arnsberg), 39: Dortmund (Dortmund), 40: Emscher-Lippe (Gelsenkirchen), 41: Duisburg/Essen (Essen), 42: Düsseldorf (Düsseldorf), 43: Bochum/Hagen (Bochum), 44: Köln (Köln), 45: Aachen (Aachen), 46: Bonn (Bonn), 47: Siegen (Siegen), 48: Nordhessen (Kassel), 49: Mittelhessen (Gießen), 50: Osthessen (Fulda), 51: Rhein-Main (Frankfurt/Main), 52: Starkenburg (Darmstadt), 53: Nordthüringen (Nordhausen), 54: Mittelthüringen (Erfurt), 55: Südthüringen (Suhl), 56: Ostthüringen (Gera), 57: Westsachsen (Leipzig), 58: Oberes Elbtal/Osterzgebirge (Dresden), 59: Oberlausitz-Niederschlesien (Görlitz), 60: Chemnitz-Erzgebirge (Chemnitz), 61: Südwestsachsen (Zwickau), 62: Mittelrhein-Westerwald (Koblenz), 63: Trier (Trier), 64: Rheinhessen-Nahe (Mainz), 65: Westpfalz (Kaiserslautern), 66: Rheinpfalz

(Ludwigshafen), 67: Saar (Saarbrücken), 68: Unterer Neckar (Mannheim), 69: Franken (Heilbronn), 70: Mittlerer Oberrhein (Karlsruhe), 71: Nordschwarzwald (Pforzheim), 72: Stuttgart (Stuttgart), 73: Ostwürttemberg (Aalen), 74: Donau-Iller (BW) (Ulm), 75: Neckar-Alb (Tübingen), 76: Schwarzwald-Baar-Heuberg (Villingen-Schwenningen), 77: Südlicher Oberrhein (Freiburg im Breisgau), 78: Hochrhein-Bodensee (Konstanz), 79: Bodensee-Oberschwaben (Ravensburg), 80: Bayerischer Untermain (Aschaffenburg), 81: Würzburg (Würzburg), 82: Main-Rhön (Schweinfurt), 83: Oberfranken-West (Bamberg), 84: Oberfranken-Ost (Bayreuth), 85: Oberpfalz-Nord (Amberg), 86: Industrieregion Mittelfranken (Nürnberg), 87: Westmittelfranken (Ansbach), 88: Augsburg (Augsburg), 89: Ingolstadt (Ingolstadt), 90: Regensburg (Regensburg), 91: Donau-Wald (Passau), 92: Landshut (Landshut), 93: München (München), 94: Donau-Iller (BY) (Memmingen), 95: Allgäu (Kempten), 96: Oberland (Garmisch-Partenkirchen), 97: Südostoberbayern (Rosenheim).

Value added: We obtained gross value added ("Bruttowertschöpfung (unbereinigt)") from the CD-ROM "Statistik regional - Ausgabe 1998" of the German statistics offices. This data is available on the county level, so we had to aggregate the data. We calculated the value added density for each region dividing the regional value added by the area of each region. In order to get the relative value added density, we divided each density by the average density of East and West Germany respectively. We treated Berlin as a West German region.

Employment shares: The CD-ROM "Statistik regional" contains data on the estimated regional total employment in five sectors in 1996.

Coal: We assigned dummies to the regions with bituminous coal or lignite deposits, which are depicted on the map in "Dierke Weltatlas 1996, 4. Auflage, S. 56 Nr. 2. The bituminous coal regions are: 35, 39-41, 43, 45, and 67. The lignite regions are: 22, 27, 28, 32-34, 42, 44, 48, 51, 57, 59, 85, and 96.

Rivers: We only considered rivers, but ignored canals. The river classification as navigable for boats of more than 1350 tons was obtained from the "Dierke Weltatlas 1984 (genaue Quelle suchen)". The following regions have a river dummy of one: 5-7, 11, 13-15, 19, 21, 25, 26, 28, 31-33, 36, 41, 42, 44, 46, 51, 52, 57, 58, 62-64, 66, 68-70, 72, 77, 78, 80-83, 90, 91.

Capitals: Germany consists of 16 federal states. We did not treat the city stated Hamburg and Bremen as having a capital. Although Berlin is a city state, too, we counted it as a capital because it is the federal capital.

Seashore: Although not lying directly at the coast, Hamburg and Bremen were assigned sea dummies, because they have major sea-ports.

Altitude: We measured the highest elevation of each region (in meters) using the CD-ROM "Route 66 (genaue Quelle nachschlagen)". Table 8 contains summary statistics.

Location: Our measure of centrality is the average direct distance from each region's center to all other centers in kilometers. Table 8 contains summary statistics.

Table 8: Summary statistics

	Mean	Stdd	Min	Max
Altitude	623	537	6	2962
Location	321	54	239	457