New Evidence on the Geographic Concentration of German Industries

Björn Alecke
GEFRA Gesellschaft für Finanz- und Regionalanalysen,
Ludgeristr. 56, 48143 Münster, Germany

Christoph Alsleben[1]
Graduiertenkolleg, University of Dortmund, Germany

Frank Scharr
Hegelstr. 36, 39104 Magdeburg, Germany

Gerhard Untiedt
(corresponding author)
GEFRA Gesellschaft für Finanz- und Regionalanalysen,
Ludgeristr. 56, 48143 Münster, Germany,
untiedt@gefra-muenster.de

Appears in:

Johansson, B., Karlsson C., Stough R.R. (Eds., 2004)

Industrial Clusters and Inter-Firm Networks.

Edgar Elgar Publishing Ltd.

[1] We thank Michael Roos, Christian Bayer and Julia Angerhausen for many helpful comments. We are also grateful for suggestions made by the participants of the *Graduiertenkolleg Workshop*, by Heinz Holländer and Wolfgang Leininger and by the participants of the Uddevalla Symposium 2003 on Entrepreneurship, Spatial Industrial Clusters and Inter-Firm Networks. We acknowledge the financial support of the *Deutsche Forschungsgemeinschaft*.

Abstract

The agglomeration of industries has received much interest both in empirical and theoretical work in recent time. Especially in Germany politicians became inspired by the notion of high-technology industry clusters and German regional policy has seen a wave of initiatives aiming at the formation of such clusters. This paper explores in a systematic way the geographic concentration of German manufacturing industries and relates it to industry characteristics and agglomeration forces proposed by theory. The main finding is that there is no general relationship between agglomeration and R&D or high-technology related business which suggests that hope put in the fast and effective development of "high-tech" clusters might be disappointed.

1 Introduction

With the emergence of the New Economic Geography the issue of spatial concentration of economic activity has received much interest both in economic theory and empirical research. While the New Economic Geography—as well as longstanding concepts such as natural advantages in trade theory and external economies of scale already stressed by Marshall (1920)—has contributed much to our understanding of why firms may tend to cluster together there is still a lack of empirical evidence on the significance and determinants of geographical concentration and its actual relevance for economic policy.

In recent years there has been a fundamental reorientation in regional policy in Germany, presumably inspired much by case study work such as Porter (1990), and its explicit aim has become to promote the formation of high-technology industry clusters and to complement traditional policy measures that support the most backward regions. For example, the "BioRegio" contest set up in 1995 was an initiative that gave financial aid to the three most promising biotechnology clusters in Germany and the "InnoRegio" initiative launched in 1999 allocated funds to the least developed regions in East Germany in order to promote the emergence of business clusters.

Two important questions associated with such policy initiatives are (i) which industries tend to cluster at all and (ii) why do industries cluster? Answering these questions may reveal important leverages for policy initiatives aiming at the promotion of business clusters be it for efficiency or equality reasons.

In this paper we choose Ellison and Glaeser's (1997) index of geographic concentration (EG index) to explore to what degree German manufacturing industries are agglomerated due to natural advantages or spillovers. Lau (1996) did some work in that direction and in a recent contribution Keilbach (2002) explains geographic concentration in a regression analysis but our work is different from these studies on Germany as—to the best of our knowledge—we apply the EG index to German industry data for the first time and have a more recent, more detailed and more comprehensive data set. We contribute to a growing literature that has applied the EG index already to other European countries such as Portugal, Belgium, Ireland (Barrios et al., 2003), the UK (Devereux et al., 1999), France (Maurel and Sédillot, 1999) and Austria (Mayerhofer/Palme, 2001). What distinguishes this paper from that work is that the pattern of geographic concentration is not only described but explained in a regression analysis with agglomeration forces from theory. Additionally, our focus is on high-tech industries as we want to examine the relationship between "innovativeness" of industry and geographic concentration.

This focus is motivated by the fact that German regional policy seems to be obsessed with the idea of "high-tech clusters" and devotes large funds to their promotion all over the country.

2 The Literature on geographic concentration

Geographic concentration of industries goes in hand with industrial specialisation of regions, and in fact the two reflect different approaches to the same phenomenon. There is already a substantial literature on regional specialisation. Amiti (1999) finds that industrial specialisation has increased in EU countries at least for the period of 1980 – 1990. She measures specialisation by a country Gini coefficient and geographic concentration by a locational Gini coefficient as proposed by Krugman (1991). In a similar analysis, Brülhart and Torstensson (1996) and Brülhart (2001) find evidence that specialisation and geographic concentration increased in the EU in the 1980s though there are differences across industries. Midelfart-Knarvik et al. (2000) explain the geographic pattern of industries across EU countries with country and industry characteristics. They find that factor endowment is important which is consistent with traditional trade theory and that Krugman's (1991) forward and backward linkages matter, too. This literature is inspired much by trade theory and by the theory of the "new economic geography" which emphasize the importance of transportation costs and internal increasing returns.

Another strand of the literature focuses on geographic concentration of industry and regional growth. In this field, Marshall's (1920) spillovers rather than the pecuniary externalities of the NEG are believed to be the decisive force behind agglomeration. There is a long-standing debate about the relative importance of intra- vs. inter-industry spillovers and there exists a large body of empirical work both in favour of the former (localisation economies) and the latter (urbanisation economies). Gleaser et al. (1992) and Henderson (1997), among others, find evidence for urbanisation effects in the spirit of Jacobs (1969), while in a more recent contribution Acs et al. (2002) look at high-technology industries and find evidence for neither of them. For Germany, Bode (1998) confirms that urbanisation effects are more important for regional growth than localisation effects.

Regarding the geographic concentration of industry, there is already some work that explores the geographic pattern of industries in depth and with more sophisticated measures (e.g. using plant-level data) than the studies about (cross-country) specialisation or regional growth mentioned above.[2] In most of this work, Ellison and Glaeser's (1997) index of agglomeration is used such as in Devereux et al. (1999) for the UK, Maurel and Sédillot for

[2] For a discussion of various measures of geographic concentration see, for example, Devereux at al. (1999) and Combes and Overman (2003).

France, Mayerhofer and Palme (2001) for Austria and in Barrios et al. (2003) who compare in depth Portugal, Ireland and Belgium. Overall these studies show that resource extractive industries tend to be the most concentrated ones and there is some tentative evidence that high-technology industries are relatively little concentrated. Our work is in the spirit of this strand of literature as we do an analysis of national industries with the focus on “Marshallian” knowledge spillovers.

In the next section we discuss the concentration measure used and describe the agglomeration pattern of German manufacturing industries.

3 Empirical results for Germany

3.1 The measures of concentration

A measure of geographic concentration that has been widely used (e.g. in Brülhart and Torstensson (1996), Audretsch and Feldman (1996) and Amiti (1999)) is the spatial variant of the Gini coefficient introduced by Krugman (1991). A severe disadvantage of the Gini coefficient is, however, that it measures concentration of economic activity both due to internal economies of scale, i.e. the “concentration” within a firm and due to natural advantages or external economies of scale, i.e. concentration resulting from the co-location of independent firms (or plants). In order to be able to distinguish between these two causes of concentration, we use two other measures instead. The first, and the one we put the focus on in this paper, has been proposed by Ellison and Glaeser (1997) (EG) and is derived from an explicit location decision model. The point of departure is the raw concentration of an industry defined as $G_i := \sum_s \left(s_{is} - x_s\right)^2$ where s_{is} is the portion of industry i’s employment located in region s and x_s is the percentage of total employment in that region. Hence, it measures concentration relative to total employment which means that as long as an industry imitates the concentration pattern of aggregate employment it will not be regarded as being concentrated.

The advantage of defining concentration relative to overall employment (as opposed to, for example, population or land area) is that we can take the overall distribution of employment (i.e. cities) as given and do not have to take into account location specific characteristics such as commuting pattern, size and age of the population, soil conditions etc. which certainly determine the distribution of employment. Also, we do not have to take an equal distribution of employment as a benchmark which is clearly no reasonable hypothesis. EG assume that firms choose their location as if dartboards were thrown at the map and that there is an a-priori distribution of firms mimicking the observed pattern in expectation. They show that—given their model of firms’ location decision—$\mathrm{E}(G) = \left(1 - \sum_i x_i^2\right)(\gamma + (1-\gamma)H)$ where γ is a combined measure of the strength of natural advantages and externalities between plants in a broad sense and H is the plant Herfindahl index. Rearranging then yields γ which is the measure of interest. A second advantage is that the model builds on a statistical distribution which allows one to test any observation against a unique null hypothesis. Here, the null is that there is in fact no agglomeration, i.e. plants choose their location in a pure random manner and independently from each other ("dartboard"). In this case, γ = 0 and $\mathrm{E}(G) = \left(1 - \sum_i x_i^2\right) H$.

Nevertheless, there are two important disadvantages with this approach. First, a world with natural advantages and one with externalities between plants are observationally equivalent. We try to overcome this limitation in Section 3 where we relate concentration to agglomeration forces in a regression analysis. Secondly, the EG index can hardly be compared across countries. The reason is that it is "a-spatial" in that it is standardised neither with regard to the number of geographic units under study nor their size. However, in this paper we are primarily interested in the relative concentration of industries, i.e. the ranking. We leave aside this discussion because our approach is to look at industries and not at a region's mix of industries.

Finally, note that because of the "a-spatial" property EG's γ can be used for a variety of aspects of "concentration"; we will use it to measure the concentration of firms belonging to the *same* industry. Thus, this paper examines the existence and strength of *localisation* economies as opposed to *urbanisation* economies which occur *across* industries. When we use the term "cluster" we refer to the agglomeration of an industry. But what is "within" industries and what "across" is mainly a matter of degree. In this paper we focus on three-digit industries but we also look at the concentration of two-digit industries. One might argue that a two-digit industry group contains already a fairly broad range of industries so that one wants to speak of "urbanisation" effects.

As one might worry that the EG index does not depict the reality of a firm's location decision process we choose a similar but simpler measure for comparison, namely a modified version of Devereux's et al. (1999) proposition. They define a measure $\alpha_i = \tilde{G}_i - M_i$ where $\tilde{G}_i = \left(\sum_s s_s^2\right) - \frac{1}{K_i^*}$, $K_i^* = \min(N, K_i)$, $M_i = H_i - \frac{1}{N_i}$, N_i is the number of plants in industry i and K is the number of geographic regions. $\tilde{G}_i$ captures the geographic concentration of employment relative to the uniform share controlling for the maximum number of regions in which employment may be located given that there are (only) N_i plants. To be consistent with the EG index which is relative to total employment, not to a uniform distribution, we use G_i instead of $\tilde{G}_i$. M measures the concentration of employment within firms (Herfindahl index) but relative to a uniform distribution. Then for any given geographic raw concentration *G*, the "internal" concentration of employment is subtracted while controlling for industry size (*N*). Note that unlike the EG index, α is linear in *H;* all else equal, a higher industrial concentration unambiguously decreases geographic concentration. α is positive (but ≤ 1) whenever the distribution of employment (relative to total employment) across regions "exceeds" that across plants, it is zero whenever these are identical and it is negative (but ≥ -1) otherwise.

3.2 The data

The database provides the 1998 distribution of employment at the plant level across the 116 manufacturing industries (including extractive industries) and across German counties (counties). While in their seminal paper EG focus on 4-digit industries and on states as the geographic unit of observation we are only able to use 3-digit industry data but at a much finer geographic level (440 counties as opposed to 51 U.S. states).

Our employment data are not classified but instead contains precise figures for each plant regardless of its size. Therefore, no further improvement in the data is necessary and we directly compute the Herfindahl indices from it. However, the confidentiality of the data means that we are not able to aggregate plants to firms, i.e. determine whether plants are under common ownership. But according to EG's model, firms choose the optimal location for each plant separately, anyway.[3] Further, we are able to group total employment of a plant by education and by occupation (production, management, R&D etc.) which we will make use of when explaining concentration in Section 4. Table 1 gives an overview of the data.

Table 1: Descriptive statistics of manufacturing employment (1998)

Number of 3-digit industries (NACE3)	116
Number of 2-digit industries (NACE2)	27
Number of plants	216,545
Total employment	7,534,781
Average employment per plant	34.8
Geographic units[1)]	440 counties
	225 labour market areas
	97 planning regions

[1)] See Table 11 in the Appendix for a further description.

3.3 How much are industries concentrated?

In EG's simple dartboard model without any spillovers and natural advantages the plants of an industry choose their location randomly. In this case one would have $E(G) = \left(1 - \sum_i x_i^2\right) H =: G_0$. In a first step we test whether *E(G)* is significantly different from

[3] Devereux et al. (1999) aggregate plants that are under common ownership and that are located in the same geographic region. If one assumes that the location of each plant is chosen independently and that a firm may well choose to locate its plants in different places then this procedure seems inconsistent.

G_0 and to our knowledge this is the first formal test for the significance of the agglomeration of German industries. The mean values of G and G_0 are 0.057 and 0.040, respectively, with their difference being highly significant.[4] More precisely, 91 out of the 116 manufacturing industries are significantly more (or less) geographically concentrated than what one would expect if location decisions were pure random.[5] Accordingly, for 25 industries the hypothesis of a pure random location decision cannot be rejected. This is in line with the results of EG and Duranton and Overman (2002) who find for the US and the UK, respectively, that the majority of industries but still not all of them are located in a way different from a random outcome.

The distribution of gamma at the 3-digit-industry level is skewed with mean 0.018 and median 0.006. A striking observation is the large number of industries (75%) that have a γ lower than 0.02 which—as argued in Ellison and Glaeser—can be interpreted as low concentration.[6] We find that only about 10% of all industries have a γ greater than 0.05. We conclude that in Germany slight concentration (at the county level) is widespread while strong concentration is found only in a small subset of industries.

Besides, one can interpret $\phi := \frac{G - G_0}{G}$ as the fraction of raw concentration attributable to some form of spillovers/natural advantage rather than randomness.[7] In Germany, for more than 60% of all industries randomness is at least as important for raw concentration as actual agglomeration of plants (Table 2); in the sub-sample of high-G industries (upper quartile consisting of 29 industries) this share amounts even to 75%. Put differently, for less than half of all industries—and for only few industries with a high raw concentration—natural advantages and/or spillovers play a dominant role in agglomeration. In total, randomness seems to have a bit stronger influence on observed agglomeration than agglomeration forces themselves.

Table 2: Raw concentration attributable to spillovers and/or natural advantage

Range of ϕ	Manufacturing industries	High-G industries
0.00	7%	14%
0.25	28%	28%
0.50	30%	34%
0.75	24%	14%
1.00	11%	10%

Table 3 shows the most and least concentrated industries. Note that the negative gamma of the 15 least concentrated industries is insignificant, i.e. it is presumably zero. What is

[4] The difference is nearly three times larger than the average standard deviation of G.
[5] For these industries the difference between G and G_0 is larger than 1.96 times its standard deviation.
[6] See Ellison and Glaeser (1997), p. 903.

striking is that "high-tech" and "medium-tech" industries are not among the top most concentrated.[8] Rather, they lie in the middle field or even at the lower end of the ranking as Table 9 in the appendix demonstrates. This is much in line with Devereux et al. (1999) for the UK, Maurel and Sédillot (1999) for France and Barrios et al.'s (2003) study comparing Portugal, Belgium and Ireland.

Obviously, resource extractive industries dominate the top group, and our simpler measure α produces fairly the same ranking as γ with the notable exception of *Coke Oven Products* and *Mining of Uranium* (NACE 231 and 120).[9] These two industries consist of only 6 and 2 plants, respectively, each of which is located in a different location so that there is no agglomeration of plants. Hence these industries are underrepresented in the majority of the regions which leads to such a high raw concentration. While the γ indicates that this particular location pattern may well be the outcome of pure random (γ is not significant), α is much less responsive to the high internal concentration of so few plants because it subtracts only the difference between H and the uniform distribution ($1/N$) which is relatively low for these industries.

[7] Note that Ellison and Glaeser (1997), p. 909, use a slightly different expression.
[8] We use a common classification developed by Grupp et al. (2000). See Table 12 in the Appendix for more details.
[9] If the resource related industries are excluded, three out of the nine high-tech industries jump up into the top 15 but one of them still has an insignificant γ.

Table 3: Most and least concentrated manufacturing industries

Rank γ	NACE	γ	H	G	Industry (NACE3)	Sign.[1)]	Rank α
1	112	0.263	0.070	0.314	Service activities incidental to oil and gas extraction, excludi		2
2	131	0.156	0.204	0.327	Mining of iron ores		3
3	335	0.124	0.027	0.147	Manufacture of watches and clocks		6
4	362	0.096	0.010	0.105	Manufacture of jewellery and related articles		7
5	101	0.077	0.045	0.118	Mining and agglomeration of hard coal		10
6	143	0.074	0.097	0.163	Mining of chemical and fertilizer minerals		9
7	132	0.072	0.177	0.235	Mining of non-ferrous metal ores, except uranium and thoriu		8
8	152	0.070	0.026	0.093	Processing and preserving of fish and fish products		14
9	103	0.069	0.044	0.109	Extraction and agglomeration of peat		13
10	263	0.060	0.098	0.151	Manufacture of ceramic tiles and flags		11
11	111	0.049	0.069	0.115	Extraction of crude petroleum and natural gas		16
12	176	0.047	0.012	0.058	Manufacture of knitted and crocheted fabrics		18
13	160	0.041	0.072	0.110	Manufacture of tobacco products		17
14	232	0.041	0.039	0.078	Manufacture of refined petroleum products		19
15	102	0.041	0.050	0.088	Mining and agglomeration of lignite		15
102	222	0.001	0.001	0.002	Printing and service activities related to printing		107
103	281	0.001	0.001	0.002	Manufacture of structural metal products		108
104	292	0.001	0.002	0.003	Manufacture of other general purpose machinery		109
105	158	0.001	0.001	0.001	Manufacture of other food products		112
106	204	0.001	0.009	0.010	Manufacture of wooden containers	no	100
107	159	0.001	0.003	0.003	Manufacture of beverages		111
108	342	0.000	0.008	0.008	Manufacture of bodies (coachwork) for motor vehicles; man	no	113
109	343	0.000	0.014	0.014	Manufacture of parts and accessories for motor vehicles an	no	110
110	311	-0.001	0.057	0.056	Manufacture of electric motors, generators and transformers	no	115
111	316	-0.001	0.021	0.019	Manufacture of electrical equipment n.e.c.	no	116
112	354	-0.001	0.182	0.180	Manufacture of motorcycles and bicycles	no	97
113	231	-0.002	0.263	0.260	Manufacture of coke oven products	no	4
114	341	-0.004	0.046	0.042	Manufacture of motor vehicles	no	114
115	242	-0.005	0.186	0.182	Manufacture of pesticides and other agro-chemical products	no	23
116	120	-0.010	0.654	0.648	Mining of uranium and thorium ores	no	1

[1)] "no" means not significant at the 5% level.

3.4 Industrial scope of agglomeration

As there is concentration *within industries* an interesting question is if there is also concentration at a more aggregated industry level, i.e. at the two-digit industry level (NACE2). Is the concentration of industry *groups* due merely to the concentration of its (sub)industries which would imply that natural advantages and spillovers are industry-specific or is there a *common* effect on the industries of a two-digit industry group? As mentioned in the introduction, when dealing with concrete industry definitions it is a matter of degree at what level of aggregation one wants to speak of "localisation" and "urbanisation" effects. In order to explore this issue we calculate in a first step the degree of concentration at the two-digit industry level for the 25 industry groups that contain more than one sub-industry using EG's γ^c.[10] It reflects how much the location decisions of firms that belong to an industry group are correlated; $\gamma^c = 0$ would indicate that there is no correlation across industries and hence no more agglomeration in the industry *group* than that simply resulting from the concentration of its sub-industries. Table 4 shows the results at the two industry levels.

Table 4: Agglomeration and coagglomeration

County (440)	H	G			γ			α		
		min	av.	max	min	av.	max	min	av.	max
Two-digit	0.040	0.001	0.050	0.648	-0.003	0.004	0.051	0.000	0.014	0.075
Three-digit	0.040	0.001	0.057	0.648	-0.010	0.018	0.263	-0.001	0.029	0.493

When moving to the more aggregate industry definition raw concentration remains nearly unchanged while γ and α decline to less than half their value. Since the magnitude of the co-agglomeration index for industry groups can be interpreted in the same way as the index for industries, the geographic concentration at the two-digit industry level is weaker than at the three-digit level. Table 10 in the appendix presents the results for all 2-digit industry groups.

At the two-digit level there is no concentration in traditional industry groups like automobiles, communication technology, furniture, machinery and rubber which is in line with EG’s findings for the US. Also similarly to the US, there is some co-agglomeration in the textile, metal, lumber and paper industry. However, in absolute terms Germany’s manufacturing industry groups exhibit only little concentration at the county level if one takes 0.05 and 0.02 as an upper and lower benchmark, again.

In a second step we calculate $\lambda := \gamma^c \Big/ \sum_j w_j \gamma_j$ which expresses the agglomeration of the group as a fraction of the weighted average of its industries (γ_i). It indicates that there is no agglomeration attributable to the group as a whole if it is zero and that natural advantages and spillovers are completely group-specific rather than (sub)industry-specific if it is greater than 1. Table 5 shows the distribution of λ. We observe that for nearly all industry-groups there is some degree of co-agglomeration but with about 70% of them having a λ smaller than 0.5. This means that for the majority group-concentration accounts for less than half of the weighted industry-concentration. In contrast, Recycling, Papers and Automobiles seem to share natural advantages or inter-industry spillovers to a high degree but they are not much (or even negatively) concentrated in absolute terms (see also Table 10 in the appendix).

[10] EG extend the model to the co-location of whole industries proposing a measure

$$\gamma^c := \frac{\left[G/\left(1-\sum x_i^2\right)\right] - H - \sum_{j=1}^{r} \gamma_j w_j^2 (1-H)}{1-\sum_{j=1}^{r} w_j^2}.$$

Table 5: Histrogram of λ

Range	Frequency	
-0.2	0	0%
0.0	4	16%
0.2	6	40%
0.4	8	72%
0.6	4	88%
0.8	2	96%
1.0	0	96%
1.2	1	100%

Another way to quantify the relative strength of industry-specific and group-specific agglomeration has been proposed by Maurel and Sédillot (1999). They remark that the concentration of a whole industry group measured by the "simple" γ of the *group* can be written as the weighted average of the γ's of the group members ("intra-industry concentration") and some group-specific component ("inter-industry concentration"). Thus, in addition to comparing agglomerative (γ_j) and co-agglomerative forces (γ^c) one can also express intra-industry agglomeration (γ_j) as a fraction of the group's total concentration (γ^{group}). This ratio ranges from as low as –2% to 134% (see column 6 in Table 10 in the appendix). A fraction of intra-industry concentration greater than 100% corresponds to a negative contribution of the inter-industry component. Communications engineering (NACE 32) on rank 22, for example, is a group whose industries themselves are significantly concentrated but taken together they are rather dispersed.

There is no general relationship between the degree of group-concentration (γ^c) and its magnitude relative to the weighted average of its components (γ_j); the spearman rank correlation is 0.40 and the standard correlation is 0.07. Obviously there are industry groups which are by far more concentrated than the average of their industries but in absolute terms they are only little (or even negatively) concentrated (e.g. Manufacture of motor vehicles, NACE2 34). An implication of this is that one may always want to look at absolute concentration and its source at the same time.

One might worry that the NACE classification misrepresents plants which are difficult to be assigned a single and meaningful industry code. This is most problematic in the field of high-technology related activities where traditional industry codes do not appropriately cover completely new fields of economic activity. Especially Germany's "new economy", characterised by a wave of start-up activity and a boom of the information- and communication industry, is a challenge for the traditional industry classification system. In order to see whether a potentially inappropriate industry definition masks concentration of similar industries, a "high-tech" and "medium-tech" industry group and five groups consisting of closely related, research-intensive industries have been compiled following a common classification by Grupp et al. (2000). (see Table 6, Table 12 in the Appendix gives a list of the NACE3 industries contained).

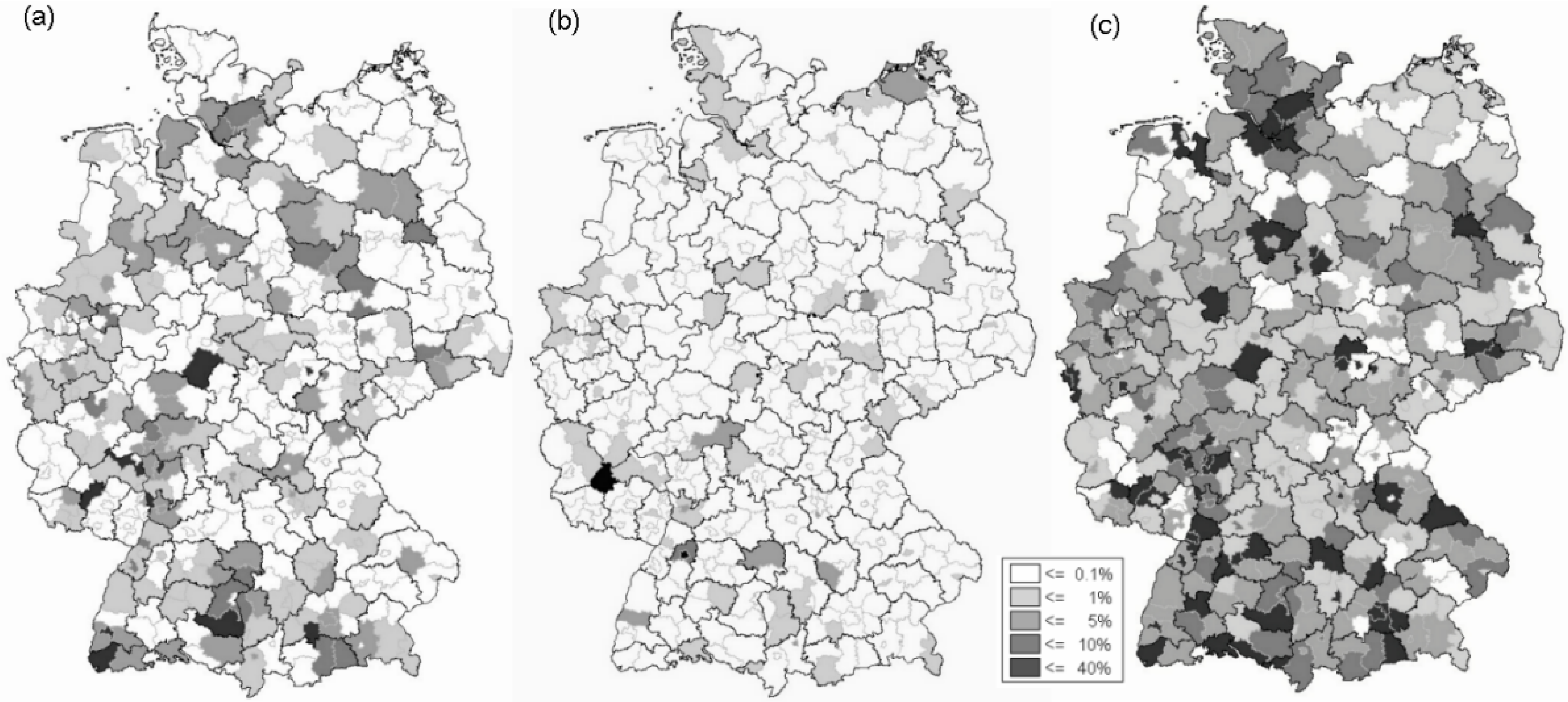
Table 6: Agglomeration of "high-tech" industry groups

Group	G	H	γ^c	Weighted average γ_w	λ
High-tech	0.006	0.004	0.0008	0.0087	0.092
Medium-tech	0.002	0.003	-0.0012	0.0019	-0.646
R&D-intensive Chemicals	0.003	0.017	0.0028	0.0033	0.840
R&D-intensive Manufacture of Machinery		0.001	0.0006	0.0043	0.144
R&D-intensive Manufacture of Electrical Machinery and Apparatus		0.017	-0.0007	0.0006	-1.260
R&D-intensive Automobiles	0.016	0.019	-0.0007	0.0004	-1.789
R&D-intensive Electronic, Optical and Communication Equipment		0.001	0.0005	0.0046	0.117

The result is in contrast to what common wisdom about inter-firm spillovers in the high-technology area suggests. First, both groups have a γ^c close to zero while that of the medium-tech group is even negative. Secondly, they rank only very modestly compared to the standard two-digit manufacturing groups.

We conclude, first, that there is some inter-industry concentration in German manufacturing industries which implies that industries share the benefits of natural advantages and/or spillovers to some degree. But for the very majority agglomeration *within* industries is stronger than *across* industries. It follows that regional policy that is concerned with agglomeration should choose particular industries and not whole industry groups. Secondly, in the high- and medium-tech business not only *industries* but also *industry groups* are not agglomerated much in absolute and relative terms.

Figure 1: Map of Germany - Percentage of total employment: (a) Pharmaceuticals, (b) Tobaco, (c) total "high-tech"



3.5 Geographic scope of agglomeration

The EG index has the property that its expected value is independent of the geographic level provided spillovers are of an all-or-nothing type and natural advantages are not correlated across regions.[11] If any of the two agglomeration forces declines with distance and thus works beyond regions, however, γ reflects the additional probability with which plants locate in the same location. In order to explore whether agglomeration forces exist at a higher geographic level and to account for the fact that administrative boundaries are not necessarily economically relevant we repeat our calculations for Germany's 225 "labour market regions" (LMR) and finally for the 97 planning regions (PR, *Raumordnungsregionen*) both of which represent functional, economically self-contained units of space with regard to commuting and trade patterns.[12]

Figure 1 shows three maps of Germany's counties and planning regions and the distribution of employment in Pharmaceuticals (a), Tabaco (b) and "high-tech" (c) industries as a percentage of total manufacturing employment. Map (a) and (b) illustrate how a little (Pharmaceuticals) and a much concentrated industry (Tobaco) look like. From map (c) two important things can be seen. First, the portion of high-tech employment is highest in big cities such as Hamburg, Berlin, Munich and also in quite a few peripheral regions, mainly in South Germany, but in total high-tech employment can be found almost everywhere in Germany. Secondly, it is very rare that the boundary of a planning region cuts through neighbouring counties with a high portion of high-tech employment. Only in the area of Hamburg, in North Germany, there is a cluster of high-tech employment that extends across at least three planning regions. Hence including counties as well as planning regions as the geographic unit is appropriate here.

Returning to the numbers, Table 7 shows that there is clear tendency of γ to increase with a higher geographic level. For raw concentration, *G*, the same is true but only for the three-digit industries. In fact, there is no rule about how agglomeration changes at a higher geographic level in general. Depending on the distribution of total employment and how the geographic units are aggregated, the degree of concentration and the ranking can—but does not have to—alter substantially. Here the overall ranking, especially the top group, remains nearly unchanged with the notable exception that Coking (NACE 231), which was on rank 113 and had no statistically significant concentration before, jumps into the top 15 of the ranking. The rank

[11] Spatial correlation means that there is a tendency of neighbouring regions to have the same natural endowment.

correlation with the county data is 0.84 for both the LMR and PR but at the more aggregated level more industries are agglomerated only insignificantly.

Table 7: Concentration at the higher geographic level

County (440)	H	G	γ	α
Two-digit	0.040	0.050	0.004	0.014
Three-digit	0.040	0.057	0.018	0.029
LMR (225)	H	G	γ	α
Two-digit	0.040	0.029	0.007	0.016
Three-digit	0.040	0.063	0.025	0.026
PR (97)	H	G	γ	α
Two-digit	0.040	0.047	0.033	0.025
Three-digit	0.040	0.072	0.036	0.039

Dividing the γ‘s at the county level by that of the PR level and taking the median gives a value of 0.517. This means that about 50% of the excess concentration at the PR level stems from the tendency of plants to locate in the same county. First, since a PR on average consists of more than 2 counties we conclude—as EG did for the US—that agglomeration forces within counties are stronger than between counties. Secondly, if we take 0.975 as a benchmark we find that in only five cases concentration at the county level is equal to that at the PR level. For all other industries concentration is higher at the PR level which means that agglomeration forces operate beyond counties.

4 Explaining concentration

The EG index cannot distinguish between the various forces that may drive agglomeration: as noted earlier, any gamma is consistent with a world only with natural advantages, only with spillovers or both. Furthermore, the index captures spillovers in a very broad sense. In a final step we want to determine what forces are actually at work by regressing the EG index on a variety of industry characteristics. We are interested in the existence and magnitude of external effects spurring agglomeration. Based on the considerations of Marshall (1920) literature has established three types of externalities: (1) a pooled market for specialised labour, (2) a pooled market for specialised input services (input sharing) and (3) knowledge spillovers.

[12] EG show that the estimator of γ remains unbiased at higher levels of geographic aggregation. The only concern is the variance of the size of the units which can lead to a bias but which we cannot solve here. See Table 11 in the Appendix for a description of the geographic units used.

4.1 Controls for Marshallian forces

Input sharing. In a world with fixed costs specialisation of firms can lead to a cumulative process of concentration. The more customers an industry which produces a non-tradable service has, the more it can specialise and exploit the increasing returns to scale. This increases productivity and/or the variety of the products which in turn benefits the purchasing industry which is assumed to like variety à la Dixit and Stiglitz (1977). This mechanism may eventually lead to the agglomeration of specialised input producers and specialised purchasing industries (Abdel-Rahman and Fujita (1990)). From the 1998 survey on the cost structure of German manufacturing industries carried out by the German Census Bureau we have detailed data on each industry's components of the total costs. We employ the portion of *technical and industrial services* and the portion of *manufactured inputs* in total shipments as an indicator of how specialised the goods produced are and hence how large gains from sharing inputs could be. Technical and industrial service inputs are likely to be very industry-specific with the largest potential for scale economies and manufactured inputs less special so that we expect a positive sign for both but a much stronger impact of the former.[13]

Labour market pooling. If an industry needs workers with industry specific skills it benefits from locating in an area where the supply of such labour is high because this increases the probability of finding capable personnel (if demand and supply of labour are stochastic). Conversely, specialised workers reduce the probability of being unemployed by moving where demand for their skills is relatively high. All else equal, we should thus observe an industry with specific needs for labour skills to agglomerate (see Helsley and Strange (1990) for a formal model). With the assumption that workers with average skills are relatively immobile and do not need to be mobile because they can find an appropriate job everywhere, it becomes possible to test for the particular effect of specific skills.

We use three alternative measures for the specificity of an industry's labour requirements. The first is the industry's share of employees with a highly specialised occupation. We follow the common definition the German Federal Bureau of Labour (*Bundesanstalt für Arbeit*) and consider "*secondary services*" which includes management, supervision, teaching and R&D (as opposed to "primary services": trading, security, office and general duties). The data are taken from our employment database. The second measure accounts for employees' education. We are able to split up total employment into three groups: *no vocational training, vocational training* and *university degree*. In terms of education the discriminatory power will be highest if we take the first and the third because employees with no vocational training at all are very

[13] Note that Rosenthal and Strange (2003) argue that manufactured inputs are more specialised than services.

unlikely to have a high school degree while those with a university degree must have one. People with a vocational training in contrast, may have very diverse educational backgrounds in real life. We expect a positive coefficient for the university proxy and a zero for the no training proxy if labour market pooling of specialised skills drives agglomeration. Thirdly, we estimate an industry's labour specificity by the deviation from the average national labour composition:

$$labourmix_i = \sum_o \left(x_{io} - \overline{x}_o \right)^2$$

where x_{io} is the percentage of industry i's workforce with occupation o and $\overline{x}_o$ the national average percentage.

The externality we are most interested in is *knowledge spillovers*. Knowledge spillovers imply the idea that when knowledge is created (i.e. research) a significant fraction of it cannot be appropriated but leaks out of a firm. If this knowledge is tacit (which means it cannot be codified) it cannot spread over long distances but requires personal contact and spatial proximity to be transmitted. By their very nature knowledge spillovers are hard to measure directly. We assume that if spatially bounded knowledge spillovers exist between plants then they render a single plant and consequently the respective industry as a whole the more innovative the more concentrated it is (for evidence of knowledge spillovers and different approaches see, for example, Jaffe et al. (1993), Audretsch and Feldman (1996), Anselin and Acs (1997)). Accordingly, we can expect firms to optimise the location of their plants with respect to spillovers to the extent that innovative capacity is crucial for their industry. We take Arrow's (1962) argument that knowledge spillovers are relatively more important in research-intensive industries. Unfortunately, patent data are not available for the NACE industry classification system and data on innovations are available from panel surveys at a highly aggregated level only. The importance of innovation is measured in three other ways. First, we employ Peneder's (1999) dummies specifying whether an industry is R&D intensive and whether it has strong or only little competitive advantages. Secondly, we use a high-tech and medium-tech dummy according to the definition we used already above. Finally, we use an industry's R&D intensity defined as R&D personnel divided by total employment.[14] If knowledge spillovers are an agglomeration force then they should have a positive impact on our concentration measures.

[14] It is correlated with the variable *vocational training + university degree* (labour market pooling) which is quite plausible as R&D is usually carried out by highly educated employees while not all educated employees work in R&D. In fact, Audretsch and Feldman (1996) measure the importance of innovation to an industry also by the portion of high-skilled employees. Both education and a firm's share of R&D personnel are a good proxy each and there are no appropriate instruments available for them.

4.2 Other controls

Transportation costs. The more costly it is to transport a good the more likely a plant cannot exploit the idiosyncratic benefits of a particular location (including those from agglomeration externalities) but has to locate optimally between suppliers and customers to minimise transportation costs (Marshall (1920)). Relatively higher transportation costs of inputs (shipments) induce plants to locate closer to their suppliers (customers). However, it is important to note that this argument is about the colocation of trade partners and has to be distinguished from pure localisation economies. In general it can render an industry either agglomerated *or* dispersed.

With regard to (international) trade the New Economic Geography, however, predicts that industries with higher economies of scale in production technology and lower trade costs are more localised (see, for example, Krugman (1991)). Hummels (2001) shows that for the majority of traded goods "explicit costs", i.e. tariffs and freight, are the most important components in trade costs. Therefore, we proxy the average trade cost of an industry by the inverse of its unit value. From trade data containing both the total weight (tons) and value of goods imported and exported we calculate an average reciprocal unit value as $\frac{1}{UV} = \frac{weight\ imports + exports}{value\ imports + exports}$.[15]

Natural advantages. In principle one needs to account for the possibility that industries are geographically concentrated just because they rely on natural resources such as water or energy sources that are distributed unevenly in space. However, compared to the U.S. for example, Germany is a small country with a relatively even distribution of regional and local power stations so that access to electricity and gas should be fairly the same in all regions. Furthermore, Germany is poor in natural resources and consequently extractive industries are small. In sum, natural advantages should be relevant for only very few industries and we control for them with the help of a *resource extractive dummy* which is assigned to the industries with NACE codes 101 – 145 and 152 (Fish processing).

Size. For any given geographic space a larger but otherwise identical industry will find it more difficult to agglomerate if there are congestion effects. We want to make sure that we capture this effect and consequently control for the size of an industry in terms of total employment.

[15] The portion of actual transportation cost in output (the importance of transportation cost) $\frac{\frac{c}{t}\,weight}{output}$ is then proportional to the reciprocal unit value with $\frac{c}{t}$ assumed to be a constant independent of the industry.

We estimate the model $\gamma = \alpha + \beta \mathbf{X} + \varepsilon$ where $\mathbf{X}$ is a vector of the above industry characteristics. Since we use alternative proxies for knowledge spillovers and labour market pooling we run 9 regressions in total, Table 8 shows the correlation matrix of the variables (see Table 13 in the Appendix for a more detailed description of the variables).

Table 8: Correlation matrix of the variables used in the regressions

		size	dresource	irs	TC	service	manuf	RD_intens	dht	dmt	drd1	drd2	no_train	university	occupation	skilldev
size		1.0														
dresource		-0.2	1.0													
irs		0.3	0.2	1.0												
TC		-0.2	0.8	0.3	1.0											
service	Input sharing	-0.1	0.4	0.5	0.4	1.0										
manuf		-0.2	0.2	0.0	0.1	0.0	1.0									
RD_intens		0.1	0.0	0.2	0.0	0.0	0.1	1.0								
dht	Know-ledge spillovers	-0.1	-0.1	0.0	-0.1	0.0	0.0		1.0							
dmt		0.3	-0.1	0.2	-0.2	-0.1	0.1		-0.1	1.0						
drd1		0.2	-0.2	0.0	-0.1	-0.1	-0.1			0.4	1.0					
drd2		0.1	-0.1	0.2	-0.1	0.0	0.2			0.2	-0.2	1.0				
occupation		0.1	0.0	0.2	0.0	0.0	0.0	1.0	0.5	0.3	0.2	0.6			1.0	
no_train	Labour market pooling	-0.2	-0.1	-0.1	-0.1	0.0	0.1	-0.6	-0.3	-0.2	0.0	-0.3	1.0			
university		0.1	0.0	0.2	-0.1	0.0	0.1	0.9	0.6	0.3	0.2	0.6	-0.6	1.0		
skilldev		0.0	0.4	0.2	0.3	0.3	0.0	-0.1	0.1	-0.2	0.0	0.0				1.0

4.3 Regression results

Before we present our regression results there are two things to note. First, agglomeration theory predicts that plants sensitive to specialised labour, specialised inputs or innovation tend to agglomerate because this will reduce production costs. Especially where we proxy "sensitivity" by cost shares there raises the question of identification. A high share of costs of—say—manufactured inputs indicates susceptibility to sharing inputs and thus a propensity to agglomerate. But this in turn should lower these costs and hence their portion in output. Consequently, what we observe is the equilibrium relationship between industry characteristics and agglomeration which tends to push the regression coefficients towards zero. If we find an insignificant relationship in equilibrium we cannot rule out the possibility that in fact there exists one. On the other hand, if we find a significant relationship we can expect it to be even stronger.[16]

Secondly, an analysis of our data reveals that there are two extreme outliers that lead to a very poor fit of the regression and a distribution of residuals that is almost certainly not normal.

Therefore, we exclude NACE3 335 (Watches), 362 (Jewellery). Both industries are very small (0.08% and 0.23% of manufacturing employment) and are characterised by family-owned, small-scale handcrafts for which the location decision is presumably dominated by family tradition and history and for which our cost proxies do not take effect. After excluding all industries with missing data we are left with 98 observations.

First of all, our control for industry size is highly significant and has the anticipated negative sign in all regressions, that is, bigger industries are less geographically concentrated (see Figure 2 in the Appendix for the regression results). The *resource dummy* is positive and always highly significant and in fact it contributes substantially to the goodness of the regression. *Transportation costs* are negative as expected, highly significant in all regressions and it is one of the most robust explanatory variables. *Internal economies of scale* have the correct sign but are only marginally significant. *Technical and industrial services* has the anticipated sign, is always highly significant and is the most robust variable. *Manufactured inputs* is mostly significant and—somewhat surprisingly—even reduces agglomeration. We conclude that industries that use a higher share of input services tend to agglomerate as theory predicts while the usage of manufactured inputs reduces agglomeration.

The results for labour market pooling are less pronounced. Our proxy for specialised occupations is positive but not significant while those for education (*no vocational training*, *university degree*) are almost always significant both with a positive sign. As low-skilled workers prove to be very immobile we conclude that firms that need them relatively much locate where they are. Apart from that, we note that both workers with no vocational training and those with a university degree represent only a minority of total manufacturing employment (21% and 8%). Based on this one could argue that unemployment insurance is well an issue for the very low skilled, too. We can support this additional argument by replacing the two variables by the industry's share of workers with a *medium* education (vocational training). It is significantly negative implying that those with an average level of education indeed do not need geographic concentration. In addition our measure of an industry's deviation from the national labour mix is positive and significant at the 1%-level. In sum, we interpret this as weak evidence for labour market pooling.

Concerning knowledge spillovers the results are disillusioning. While we found in the previous section that "high-tech" industries belong to the least concentrated industries we now find that even when controlling for other factors, all of the different measures of susceptibility to spillovers are insignificant, which is consistent with that result. In the majority of the

[16] See also Rosenthal and Strange (2003).

regressions the measures are even associated with a negative sign. Especially in the case of our most reliable proxies, namely share of R&D employees and the technology dummies, this is striking.

Before summing up, we want to spend a few comments on agglomeration at the higher geographic level. We noted above that when moving to more aggregate geographic levels there is no rule for the changes in the concentration measure and for Germany we found a higher concentration at the PR level for the majority of the industries. Like Rosenthal and Strange (2001) we find that the agglomeration forces at the higher level are much weaker with many variables turning insignificant (see Figure 3 in the appendix).

5 Discussion and Conclusion

This paper has explored the geographic concentration of German manufacturing industries with the help of Ellison and Glaeser's (1997) concentration index for the first time. Thereby we add to previous empirical work dealing with the concentration in other European countries. The questions we ask is (i) how much plants of an industry are agglomerated and (ii) what factors determine concentration, i.e. we are interested in the pattern and magnitude of localisation economies. The focus is on high-technology related industries motivated by the observation that the idea of "high-tech clusters" is en vogue at the moment and has inspired many policy initiatives.

Concerning the first question we find that 80% of the 116 industries are statistically significantly more concentrated than what would result if location decisions were pure random. However, the degree of concentration is rather low and randomness accounts for almost half of it; only resource related industries exhibit strong concentration and they dominate the group of the top 15. In particular, high-/medium-tech *industries* and *industry groups* are only little concentrated, partly even not significantly so, and rank medium or even lowest. This result does not change when we use an alternative and simpler concentration measure or when we take a more aggregate geographic level.

To answer the second question, we have related *concentration* a variety of industry measures that shall reflect theoretical agglomeration forces in a regression analysis. We find that transportation costs significantly reduce agglomeration which is line with the predictions of the new economic geography. Concerning Marshall's (1920) agglomeration forces we find strong evidence for inputs sharing (specialised service inputs), weak evidence for labour market pooling and no evidence for knowledge spillovers. Neither of our alternative proxies for high-technology or research intensity produces a significant and positive relationship.

Either such spillovers are not limited to knowledge intensive activities (within industries) but instead are much more general than has been assumed so far or they simply do not spur agglomeration. Concerning the first point we should emphasize again that our approach is a top-down one measuring the agglomeration of *industries* and *industry groups*. There are, of course, other methods to explore the geographic distribution of economic activity. Sternberg and Litzenberger (2003), for example, use very similar data but take a regional perspective. For each German county they calculate a "cluster index" per industry. If this index exceeds a critical value, the respective region is considered to host a cluster of the respective industry. Naturally, the hurdle determines *how many* clusters on finds at all. Although using the EG index implies aggregating detailed regional data to one industry measure, the statistical model behind it allows to make more systematic statements than when employing ad hoc criteria of what makes a "cluster".

One can interpret our regression result that R&D intensity has no significant effect on geographic concentration of industries as evidence against localisation economies from knowledge spillovers. If it is "cross-fertilisation" that counts then knowledge may flow between any possible pair of industries and we would hardly observe single-industry (high-tech) clusters on the map, which would be consistent with our finding.

There are two potential explanations for the latter point, namely that spillovers do not necessarily spur agglomeration (at the regional level). The first concerns the range of knowledge spillovers. Spillovers could work at an extremely localised level such as the city-level. Our approach would not necessarily capture such mini-spillovers, and case-studies would perhaps be more appropriate. However, it is important to note that at such a fine geographic level, agglomeration is likely to be due to administrative conditions (e.g. a city's allocation of industrial estate). Alternatively, distance may actually not matter for knowledge flows. Orlando (2002) finds for the U.S. that unlike inter-industry R&D spillovers, intra-industry spillovers do not attenuate by distance. If this were true there would be no need for industries to agglomerate in order to benefit from such spillovers. As Germany is a relatively small country with every major city within one day travel-distance, spatial proximity could be a poor proxy for the importance of personal contact, trust etc. However, in a recent study Bode (2004) shows that in Germany only a very small share of regional knowledge spills over to neighbouring regions due to substantial spatial transaction costs. This indicates that spillovers are indeed spatially bound at the county level which justifies our approach. But a caveat is that we cannot include in our analysis the proximity to public research facilities like he does.

A second explanation for why spillovers may not spur agglomeration has been addressed by Shaver and Flyer (2000). They argue that heterogeneity among firms can lead to asymmetric contributions and benefits from agglomeration externalities and that firms' location choice becomes strategic then. They give empirical evidence that firms with superior technologies, human capital or suppliers have the incentive to locate distant from other firms, especially from firms within their industry, i.e. from direct rivals. Our systematic analysis of manufacturing industries gives some support to their firm-level study.

To conclude, there is no general relationship between agglomeration and R&D or high-technology related business among German manufacturing industries which means that "high-tech" does not make industries agglomerate naturally. This in turn suggests that German regional policy in which much hope is currently put in the fast and effective development of high-tech clusters might see some disappointments.[17]

[17] Note that labour market pooling and input sharing are pecuniary externalities, i.e. unlike knowledge spillovers they operate through markets and hence prices. There is no need for government intervention from an efficiency perspective.

6 References

Abdel-Rahman, H., Fujita, M. (1990), "Product Variety, Marshallian Externalities, and City Sizes", Journal of Regional Science, 30, 165-183

Acs, Z. J., FitzRoy, F. R., Smith, I. (2002), "High-technology employment and R&D in cities: Heterogeneity vs specialization", The Annals of Regional Science, 36, 373-386

Amiti, M. (1999) "Specialisation Patterns in Europe", Weltwirtschaftliches Archiv, 134 (4), 573-593.

Anselin, L., Varga, A., Acs, Z. (1997), "Local Geographic Spillovers between University Research and High Technology Innovations", Journal of Urban Economics, 42, 422-448.

Arrow, K. J. (1962), "Economic Welfare and the Allocation of Resources for Invention", in: Nelson, R. R. "The Rate and Direction of Inventive Activity", Princeton University Press, 609-626.

Audretsch, D. B., Feldman, M. P. (1996), "R&D Spillovers and the Geography of Innovation and Production", American Economic Review, 86 (3), 630-640.

Barrios, S., Bertinelli, L., Strobl, E., Teixeira, A. C. (2003), "Agglomeration Economies and the Location of Industries: A Comparison of Three Small European Countries", CORE Discussion Paper, 2003-67.

Brülhart, M., Torstensson, J. (1996), "Regional Integration, Scale Economies and Industrial Economies", Center for Economic Policy Research Discussion Paper 1435.

Brülhart, M. (2001), "Evolving Geographical Concentration of European Manufacturing Industries", Weltwirtschaftliches Archiv, 137 (2), 215-243

Combes, P.-P., Overman, H. G. (2003), "The spatial distribution of economic activities in the European Union", draft chapter for the forthcoming Handbook of Urband and Regional Economics, Vol. 4

Devereux, M. P., Griffith, R., Simpson, H. (1999), "The Geographic Distribution of Production Activity in the UK", The Institute for Fiscal Studies, Working Paper 26/99

Dumais, G., Ellison, G., Glaeser, E. (1997), "Geographic Concentration as a Dynamic Process" NBER Working Paper

Duranton, G., Overman, H. G. (2002), "Testing for Localisation Using Micro-Geographic Data", CEPR Discussion Paper 3379

Ellison, G., Glaeser, E. L. (1997), "Geographic Concentration in U.S. Manufacturing Industries: A Dartboard Approach", Journal of Political Economy, 105 (5), 879-927

Grupp, H., Legler, H., Jungmittag, A., Schmoch, U. (2000), "Hochtechnologie 2000 - Neudefinition der Hochtechnologie für die Berichterstattung zur technologischen Leistungsfähigkeit Deutschlands", Bericht des FhG-ISI und des NIW im Auftrag des Bundesministeriums für Bildung und Forschung

Helsley, R., Strange, W. (1990), “Matching and agglomeration economies in a system of cities”, Regional Science and Urban Economics, 20, 189-212

Hummels, D. (2001), “Toward a Geography of Trade Costs”, Purdue University, GTAP Working Paper 17

Jaffe, A. B., Trajtenberg, M., Henderson, R. M. (1993), "Geographic Localization of Knowledge Spillovers as Evidenced by Patent Citations", Quarterly Journal of Economics, 108 (3), 577-598

Keilbach, M. (2002), “Determinanten der räumlichen Konzentration von Industrie und Dienstleistungsbranchen. Eine empirische Analyse für Westdeutschland”, Jahrbuch für Regionalwissenschaft”, 22, 61-79

Krugman, P. (1991), “Geography and Trade”, Cambridge, Mass.: MIT Press

Lau, D. (1996), “Local Concentration and International Competitiveness: Some Empirical Evidence for Manufacturing Sectors in Selected European Countries”, Konjunkturpolitik, 42, 181-205

Marshall, A. (1920), “Principles of Economics”, London: MacMillan

Maurel, F., Sédillot, B. (1999), “A measure of the geographic concentration in French manufacturing industries”, Regional Science and Urban Economics, 29, 575-604

Mayerhofer, P., Palme, G. (2001), “Strukturpolitik und Raumplaung in den Regionen an der mitteleuropäischen EU-Außengrenze zur Vorbereitung auf die EU-Osterweiterung. Teilprojekt 6/1: Sachgüterproduktion und Dienstleistungen: Sektorale Wettbewerbsfähigkeit und regionale Integrationsfolgen”, http://www.preparity.wsr.ac.at/public/veroeffentlichungen/at/veroeffentlichungen_a6.html

Midelfart-Knarvik, K. H., Overman, H. G., Venables A. J. (2000), “Comparative advantage and economic geography: estimating the determinants of industrial location in the EU”, CEPR Discussion Paper 2618

Orlando, M. J. (2002), “Measuring R&D spillovers: on the importance of geographic and technological proximity”, Federal Reserve Bank of Kansas City

Peneder , M. (1999), “Intangible Investment and Human Resources. The New WIFO Taxonomy of Manufacturing Industries”, WIFO Working Papers 114/99

Porter, M. (1990), “The competitive advantage of nations”, New York: Free Press

Rosenthal, S. S., Strange, W. C. (2001), “The Determinants of Agglomeration”, Journal of Urban Economics, 50 (2), 191-229.

Shaver, J. M., Flyer, F. (2000), “Agglomeration economies, firm heterogeneity, and foreign direct investment in the United States”, Strategic Management Journal, 21, 1175-1193

Sternberg, R., Litzenberger, T. (2003), “Regional Clusters – Operationalisation and Consequences for Entrepreneurship”, Working Paper No. 2003-02, University of Cologne

7 Appendix

Table 9: The agglomeration of high- and medium-tech manufacturing industries (γ)

Rank γ	NACE	γ	H	G	Industries (NACE3)	Sign.[1]	Rank α
					High-technology industries		
16	296	0.037	0.072	0.105	Manufacture of weapons and ammunition		20
19	233	0.032	0.263	0.285	Processing of nuclear fuel	no	5
23	353	0.027	0.050	0.076	Manufacture of aircraft and spacecraft		25
51	300	0.007	0.035	0.041	Manufacture of office machinery and computers		59
54	322	0.007	0.019	0.025	Manufacture of television and radio transmitters and apparatus fc		61
59	333	0.006	0.124	0.128	Manufacture of industrial process control equipment	no	64
73	321	0.004	0.012	0.016	Manufacture of electronic valves and tubes and other electronic c		81
84	244	0.003	0.018	0.020	Manufacture of pharmaceuticals, medicinal chemicals and botani	no	90
					Medium-technology industries		
32	334	0.015	0.020	0.035	Manufacture of optical instruments and photographic equipment		41
38	352	0.011	0.042	0.052	Manufacture of railway and tramway locomotives and rolling stocl		37
43	315	0.009	0.034	0.042	Manufacture of lighting equipment and electric lamps		52
49	246	0.007	0.010	0.017	Manufacture of other chemical products		56
50	314	0.007	0.046	0.052	Manufacture of accumulators, primary cells and primary batteries	no	42
57	291	0.006	0.006	0.013	Manufacture of machinery for the production and use of mechani		66
62	323	0.005	0.020	0.025	Manufacture of television and radio receivers, sound or video rec		70
68	293	0.005	0.009	0.013	Manufacture of agricultural and forestry machinery		82
72	243	0.004	0.014	0.018	Manufacture of paints, varnishes and similar coatings, printing ink		76
76	294	0.004	0.002	0.006	Manufacture of machine tools		85
82	331	0.003	0.002	0.005	Manufacture of medical and surgical equipment and orthopaedic		93
97	241	0.002	0.071	0.073	Manufacture of basic chemicals	no	98
101	295	0.002	0.002	0.004	Manufacture of other special purpose machinery		106
109	343	0.000	0.014	0.014	Manufacture of parts and accessories for motor vehicles and thei	no	110
111	316	-0.001	0.021	0.019	Manufacture of electrical equipment n.e.c.	no	116
110	311	-0.001	0.057	0.056	Manufacture of electric motors, generators and transformers	no	115
114	341	-0.004	0.046	0.042	Manufacture of motor vehicles	no	114
115	242	-0.005	0.186	0.182	Manufacture of pesticides and other agro-chemical products	no	23

[1] "no" means not significant at the 5% level.

Table 10: The coagglomeration of manufacturing industries (γ^c)

Rank γ^c	γ^c	λ	Rank λ	Rank α	Intra-industry concentration as % of group's concentration	Industry group (NACE2)	# indus-tries	Employ-ment (000's)
1	0.051	0.402	7	1	68%	Extraction of crude petroleum and natural gas; service a	2	6.6
2	0.015	0.382	9	4	93%	Manufacture of coke, refined petroleum products and nu	3	32.0
3	0.007	0.387	8	9	33%	Manufacture of textiles	7	149.5
4	0.005	0.319	11	6	55%	Manufacture of other transport equipment	5	146.2
5	0.003	0.274	12	11	56%	Manufacture of basic metals	5	354.2
6	0.003	0.575	5	22	37%	Manufacture of wood and of products of wood and cork,	5	184.9
7	0.003	0.584	4	10	60%	Publishing, printing and reproduction of recorded media	3	382.2
8	0.003	0.748	2	16	57%	Manufacture of pulp, paper and paper products	2	148.8
9	0.003	1.046	1	21	64%	Recycling	2	34.7
10	0.002	0.137	18	23	65%	Other mining and quarrying	5	66.0
11	0.002	0.513	6	15	30%	Manufacture of chemicals and chemical products	7	489.9
12	0.002	0.208	15	26	49%	Manufacture of other non-metallic mineral products	8	283.6
13	0.002	0.322	10	13	97%	Manufacture of wearing apparel; dressing and dyeing of	3	89.7
14	0.002	0.061	21	7	93%	Tanning and dressing of leather; manufacture of luggag	3	34.3
15	0.002	0.235	14	19	49%	Manufacture of fabricated metal products, except machi	7	796.7
16	0.001	0.080	20	17	89%	Manufacture of furniture; manufacturing n.e.c.	6	287.1
17	0.001	0.166	16	18	65%	Manufacture of medical, precision and optical instrumen	5	394.1
18	0.001	0.237	13	27	43%	Manufacture of food products and beverages	9	732.8
19	0.000	0.157	17	25	91%	Manufacture of rubber and plastic products	2	393.0
20	0.000	0.136	19	24	62%	Manufacture of machinery and equipment n.e.c.	7	1,057.0
21	0.000	-0.143	24	14	-2%	Manufacture of electrical machinery and apparatus n.e.c	6	444.3
22	-0.001	-0.149	25	12	134%	Manufacture of radio, television and communication equ	3	192.4
23	-0.002	0.666	3	20	65%	Manufacture of motor vehicles, trailers and semi-trailers	3	666.7
24	-0.002	-0.027	22	3	101%	Mining of metal ores	2	1.5
25	-0.003	-0.038	23	5	103%	Mining of coal and lignite; extraction of peat	3	102.7

Table 11: Employment, population and size of the geographic units used

Employment	County	LMR	PR
Min	1,675	2,614	9,564
Max	156,256	376,122	398,228
Average	17,125	33,488	77,678
Std.Dev.	17,566	45,282	63,809

Population	County	LMR	PR
Min	35,819	65,336	248,203
Max	3,414,293	3,545,385	3,414,293
Average	185,497	362,551	840,970
Std.Dev.	212,765	483,899	599,511

Area (km²)	County	LMR	PR
Min	36	78	78
Max	3,058	6,427	7,179
Average	810	1,569	3,664
Std.Dev.	593	977	1,483

Note: One reason for the noticeable differences between "population" and "employment" is that we measure manufacturing only which is about 30% of total employment.

Table 12: Industries contained in the hand-compiled industry groups

Group	contains NACE3
High-tech	233, 242, 244, 296, 300, 321, 322, 333, 353
Medium-tech	241, 243, 246, 291, 293, 294, 295, 311, 314, 315, 316, 323, 331, 334, 341, 343, 352
R&D-intensive Chemicals	233, 241, 242, 243, 244, 246
R&D-intensive Manufacture of Machinery	291, 293, 294, 295, 296
R&D-intensive Manufacture of Electrical Machinery and Apparatus	311, 314, 315, 316
R&D-intensive Automobiles	341, 343, 352, 353
R&D-intensive Electronic, Optical and Communication Equipment	300, 321, 322, 323, 331, 332, 333, 334

Table 13: Description of the variables used in the regressions

Variable	Description	Type of spillovers	Regr. #
size	Total industry employment (000's)		
dresource	Resource dummy (set to 1 for NACE3 101-145, 152)		
irs	Internal increasing returns (average establ. size) (000's)		
TC	Transportation cost (tons/€)		
service	Share of services in inputs	Sharing of inputs	
manuf	Share of manufactured goods in inputs		
RD_intens	Share of R&D personnel in total personnel	Knowledge spillovers	A
dht	Dummy "high-tech industry"[1]		B
dmt	Dummy "medium-tech industry"[1]		
drd1	Dummy "R&D intensive industry with few competitive advantages"[2] NACE3: 157, 171, 211, 221, 231, 243, 245, 263, 271, 273, 274, 291, 292, 294, 295, 296, 297, 311, 314, 315, 331, 343, 371		C
drd2	Dummy "R&D intensive industry with strong competitive advantages"[2] NACE3: 232, 233, 241, 242, 244, 246, 247, 265, 300, 312, 321, 322, 323, 332, 333, 334, 341, 353		
occupation	Share of workers with specialised occupation (management, supervision, teaching, R&D)	Labour market pooling	1
no_train	Share of workers with no vocational training		2
university	Share of workers with universitiy degree		
skilldev	Deviation from national labour composition		3

[1] Following Grupp et al. (2000), see Table 12 for a list of the industries contained.
[2] Following Peneder (1999).

Figure 2: Regression results – county level. Dependent variable: γ

Variable	Regression A1	Regression A2	Regression A3	Regression B1	Regression B2	Regression B3	Regression C1	Regression C2	Regression C3
constant	0.0160*** (0.0036)	0.0014 (0.0068)	0.0150*** (0.0033)	0.0166*** (0.0038)	0.0000 (0.0069)	0.0141*** (0.0029)	0.0172*** (0.0040)	0.0003 (0.0069)	0.0131*** (0.0029)
size	-0.0609*** (0.0161)	-0.0494*** (0.0162)	-0.0638*** (0.0159)	-0.0585*** (0.0168)	-0.0491*** (0.0168)	-0.0631*** (0.0166)	-0.0649*** (0.0161)	-0.0539*** (0.0165)	-0.0668*** (0.0159)
dresource	0.0382** (0.0077)	0.0390** (0.0075)	0.0317** (0.0082)	0.0386** (0.0079)	0.0388** (0.0077)	0.0307** (0.0082)	0.0412** (0.0083)	0.0393** (0.0081)	0.0304** (0.0082)
irs	0.0140 (0.0086)	0.0115 (0.0083)	0.0144* (0.0084)	0.0139 (0.0086)	0.0110 (0.0084)	0.0143* (0.0084)	0.0132 (0.0086)	0.0114 (0.0085)	0.0150* (0.0085)
TC	-0.0008** (0.0003)	-0.0007** (0.0003)	-0.0007** (0.0003)	-0.0008** (0.0003)	-0.0007** (0.0003)	-0.0007** (0.0003)	-0.0009** (0.0003)	-0.0007** (0.0003)	-0.0007** (0.0003)
service	0.0866** (0.0292)	0.0845** (0.0283)	0.0752** (0.0291)	0.0861** (0.0295)	0.0837** (0.0287)	0.0743** (0.0294)	0.0878** (0.0290)	0.0845** (0.0284)	0.0746** (0.0291)
manuf	-0.0363* (0.0205)	-0.0412** (0.0200)	-0.0343* (0.0201)	-0.0354* (0.0213)	-0.0400* (0.0208)	-0.0324 (0.0208)	-0.0395* (0.0216)	-0.0399* (0.0211)	-0.0305 (0.0208)
RD_intens	-0.0287 (0.0550)	-0.0403 (0.0362)	-0.0098 (0.0173)						
dht				0.0024 (0.0062)	-0.0005 (0.0062)	-0.0017 (0.0048)			
dmt				-0.0013 (0.0040)	-0.0018 (0.0038)	-0.0016 (0.0035)			
drd1							0.0060* (0.0035)	0.0035 (0.0034)	0.0037 (0.0030)
drd2							0.0032 (0.0054)	-0.0009 (0.0051)	-0.0021 (0.0037)
occupation	0.0137 (0.0492)			-0.0134 (0.0211)			-0.0250 (0.0228)		
no_train		0.0445** (0.0179)			0.0456** (0.0181)			0.0438** (0.0181)	
university		0.0873* (0.0508)			0.0446 (0.0365)			0.0373 (0.0381)	
skilldev			0.0385** (0.0190)			0.0399** (0.0194)			0.0405** (0.0189)
F-statistic	11.79	12.00	12.82	10.40	10.47	11.26	10.98	10.78	11.76
Adjusted R^2	0.47	0.50	0.49	0.46	0.49	0.48	0.48	0.50	0.49

Notes: Standard errors in brackets, * denotes significant at 10% level, ** at 5% level, *** at 1% level

Figure 3: Regression results: planning region level. Dependent variable: γ

Variable	Regression A1	Regression A2	Regression A3	Regression B1	Regression B2	Regression B3	Regression C1	Regression C2	Regression C3
constant	0.0158** (0.0087)	0.0383*** (0.0167)	0.0191*** (0.0081)	0.0176** (0.0091)	0.0419*** (0.0167)	0.0297*** (0.0071)	0.0150 (0.0097)	0.0405*** (0.0171)	0.0303*** (0.0073)
size	-0.1581*** (0.0385)	-0.1675*** (0.0399)	-0.1604*** (0.0389)	-0.1517*** (0.0399)	-0.1613*** (0.0410)	-0.1490*** (0.0411)	-0.1584*** (0.0390)	-0.1691*** (0.0407)	-0.1606*** (0.0402)
dresource	0.0292 (0.0183)	0.0282 (0.0183)	0.0274 (0.0200)	0.0310 (0.0187)	0.0302 (0.0187)	0.0416*** (0.0202)	0.0276 (0.0201)	0.0295 (0.0199)	0.0465*** (0.0209)
irs	0.0581*** (0.0204)	0.0607*** (0.0205)	0.0578*** (0.0205)	0.0581*** (0.0205)	0.0620*** (0.0206)	0.0615*** (0.0208)	0.0589*** (0.0209)	0.0615*** (0.0210)	0.0577*** (0.0216)
TC	-0.0009 (0.0006)	-0.0010 (0.0007)	-0.0009 (0.0007)	-0.0009 (0.0006)	-0.0010 (0.0007)	-0.0011** (0.0007)	-0.0009 (0.0007)	-0.0010 (0.0007)	-0.0013** (0.0007)
service	-0.0339 (0.0695)	-0.0296 (0.0696)	-0.0416 (0.0713)	-0.0339 (0.0699)	-0.0295 (0.0700)	-0.0321 (0.0728)	-0.0350 (0.0701)	-0.0291 (0.0702)	-0.0330 (0.0737)
manuf	-0.0042 (0.0489)	0.0012 (0.0491)	-0.0056 (0.0492)	-0.0037 (0.0505)	0.0023 (0.0508)	-0.0139 (0.0514)	-0.0002 (0.0521)	0.0010 (0.0521)	-0.0262 (0.0525)
RD_intens	-0.0277 (0.1309)	0.0389 (0.0890)	0.1238*** (0.0423)						
dht				0.0110 (0.0147)	0.0134 (0.0150)	0.0297*** (0.0119)			
dmt				-0.0013 (0.0095)	-0.0003 (0.0093)	0.0068 (0.0087)			
drd1							0.0000 (0.0085)	0.0035 (0.0083)	0.0093 (0.0077)
drd2							-0.0032 (0.0132)	0.0020 (0.0126)	0.0186** (0.0094)
occupation	0.1406 (0.1171)			0.0987** (0.0500)			0.1257*** (0.0552)		
no_train		-0.0556 (0.0440)			-0.0605 (0.0441)			-0.0598 (0.0446)	
university		0.0670 (0.1251)			0.0701 (0.0889)			0.0969 (0.0939)	
skilldev			0.0234 (0.0465)			-0.0103 (0.0480)			-0.0059 (0.0478)
F-statistic	5.20	4.73	4.98	4.69	4.32	4.09	4.57	4.21	3.78
Adjusted R^2	0.25	0.25	0.24	0.25	0.25	0.22	0.25	0.24	0.20

Notes: Standard errors in brackets, * denotes significant at 10% level, ** at 5% level, *** at 1% level