Relative Innovative Capacity of German Regions: Is East Germany Still Lagging Behind?

Michael Berlemann and Vera Jahn

Introduction

Most economists will agree that innovation, in the form of new products and processes for example, is a fundamental prerequisite for economic prosperity (see Dohse 2004). Innovation is also the key to solving many of the world’s most pressing social challenges such as improving the quality of the physical environment or health care. The innovative capacity of a region is rooted in its microeconomic environment and depends, among others, on factors such as the intensity of scientists and engineers in the regional workforce and the degree of protection of intellectual property (see Baumert, Buesa and Heij 2010).

Since the microeconomic environment can be influenced by economic policy measures, politicians have always been interested in monitoring innovative performance. A prominent example is the development in East and West German regions. While in the early years after German Reunification East and West Germany converged considerably in terms of per capita GDP, this process stagnated in the late 1990s (Berlemann and Thum 2005). A large share of the gap in economic performance between the two parts of Germany can be attributed to differences in total factor productivity (Berlemann and Wesselhöft 2012). In order to be able to close the remaining gap, East Germany needs productive and innovative enterprises. However, various studies such as Dohse (2004) or Eickelpasch (2009) came to the result that East Germany is still lagging behind West Germany in innovative capacity due to lower R&D expenses and smaller average firm size, for example. As a result, German politics has engaged in various programs to increase East German innovative capacity, see Belitz, Fleischer and Stephan (2001); BMBF (2009 and 2010).

Little official data on innovations is available to date. To some extent, this informational gap is closed by survey studies (see Rammer et al. 2013). However, firms have little incentive to publish information on their true innovative capacity. Newly developed processes are typically kept secret in order to take advantage of new technologies as long as possible (Moser 2013). Moreover, surveys often do not allow for the study of innovation activity on a highly disaggregated level due to the scarcity of surveyed enterprises.

The primary aim of this paper is to deliver empirical evidence on the relative innovative capacity of German regions on the NUTS 3-level. Special attention is devoted to the question of whether East German regions, in fact, still perform worse in terms of innovation activity. In the absence of reliable data on innovations, we follow the existing literature (Schmookler 1962 and 1966; Sokoloff 1988; Moser and Voena 2012) by using patents as a proxy for innovative activity. As Griliches (1990, 1661) argues in his survey, patent statistics have several advantages: “they are available; they are by definition related to inventiveness, and they are based on what appears to be an objective and only slowly changing standard”. However, using patents as indicators of innovative activity also has several caveats. Firstly, not all innovative activity is patented. Several reasons may account for that fact. In some countries patent laws do not exist at all (Moser 2013). And even if patent laws exist, inventors may decide not to patent their intellectual property due to the costs incurred for example, and opt for other available means of protecting intellectual property instead such as secrecy or the absence of any commercial interest. Due to the fact that not all inventions are patented, patents as an absolute measure of innovative activity will underestimate the true level of innovative activity. Secondly, the share of innovations that inventors choose to patent is subject to changes over the course of time. As Moser (2013)
reports, secrecy was a highly effective means of protecting mid-nineteenth-century improvements in chemicals because, at that time, competitors were unable to reverse engineer them. As science progressed, the possibilities of reverse engineering grew considerably, thereby increasing the incentive to patent new chemical innovations. Whenever the share of patented innovations changes over the course of time, comparisons along the time-axis obviously become problematic. Thirdly, the share of innovations that inventors choose to patent also varies considerably between the different sectors (Levin et al. 1987; Cohen, Nelson and Walsh 2000; Harhoff and Hoisl 2006). Thus, regions or countries with differing sector structures are hardly comparable on the basis of simple patent counts.

In order to deal adequately with the three problems mentioned, we proceed as follows. Firstly, we refrain from trying to uncover the true level of innovative activity in German regions. Since we have no reliable information on the share of patented innovations in all innovations, we concentrate on constructing a measure of relative innovative capacity. Secondly, since we have no reliable information on the development of the share of patented innovations in all innovations over the course of time, we refrain from including the time dimension in our analysis. Instead, we focus on a cross section of data. For reasons of data availability, we choose 2008 as a sample year. Thirdly, when constructing our measure of innovative capacity we correct for differences in regions’ sector structure.

The remainder of the paper is organized as follows: in the next section we outline the construction of an adequate indicator of relative innovative capacity on a regional level. We subsequently turn to a description of the patent used and firm data. Based on these data, we then calculate the indicator and discuss the results. The paper ends with some conclusions.

**Indicator construction**

Our approach of constructing an adequate indicator of relative innovative capacity of German NUTS 3-regions is based on the idea of comparing the expected number of patents per region with the number that actually occur. Whenever a region generates more (or less) patents than an imaginary German region with the same sector structure, the referring region turns out to be overly (or insufficiently) innovative.

Formally, our indicator of relative innovative performance of a region $i$, $R_i$, can be calculated as follows:

Let $I$ be the number of regions, $J$ the number of sectors, $P_{i,j}$ the number of patents in region $i$ and sector $j$ and $N_{i,j}$ the number of firms in region $i$ and sector $j$.

Factual patent density in region $i$ is then given by

$$D_i = \sum_{j=1}^{J} \frac{P_{i,j}}{N_{i,j}}$$

with $N_i$ being the number of firms in region $i$, i.e.

$$N_i = \sum_{j=1}^{J} N_{i,j}$$

Whenever firms within the same sector perform similarly in terms of generating innovations in all regions, patent density is expected to vary one to one with the structure of the regional economy. Expected patent density is thus given by

$$D'_i = \sum_{j=1}^{J} \frac{1}{N_i}$$

with $D'_j$ being average patent density in sector $j$ over all regions $i$, i.e.

$$D'_j = \sum_{i=1}^{I} \frac{P_{i,j}}{N_{i,j}}$$

We then define relative innovative performance of region $i$ as

$$R_i \equiv D_j - D'_j$$

Positive values of $R_i$ go along with overly innovative regions, while negative values indicate underperforming regions.

In order to construct the indicator of relative innovative capacity described we require two sorts of data: patent data and data on the regional structural composition of the economy.

**Patent data**

Adequate patent data have to fulfil several requirements. Firstly, the data should be representative of the innovative activity of German firms. Secondly, the patent data must be available on the regional level, in our case on the German NUTS 3-level. Thirdly, we need information on how patents are distributed among the regional economies’ sectors.
In principle, three sources of patent data could be used to study the relative innovative performance of German regions. These sources are related to the three typical ways that German inventors can choose to go when securing their intellectual property via patenting. Firstly, when an applicant is interested in securing the patent only for the German market, s/he can contact the German Patent Office (Deutsches Patent- und Markenamt, DPMA). Secondly, if s/he is also interested in patenting his/her invention in other European countries, s/he can send his application to the European Patent Office. After receiving an application, the European Patent Office transmits the application to the addressed national patent offices. Thirdly, whenever an applicant also (or only) seeks patenting outside Europe, s/he can take advantage of the rules set out in the Patent Cooperation Treaty (PCT), which was signed in June 1970 and entered into force in January 1978. While the treaty was initially signed by only 18 states, nowadays almost 150 states take part in the agreement. After an international patent application is filed with the appropriate patent office (Receiving Office), the application enters an international phase. Throughout this phase an authorized International Search Authority (ISA) conducts a systematic investigation of patentability. In the next step the application is forwarded to the responsible national patent offices, which then initiate the regular patenting procedures.

In a first step we compare the patent applications filed at the European Patent Office and PCT applications. Both sorts of data can be extracted from the REGPAT Database (January 2013 edition), maintained by the OECD. Both databases allow for tracking of the applicants back to the regional NUTS 3-level. Moreover, both databases use the International Patent Classification (IPC) to attribute the applications to different sectors. As a result, both databases are easily comparable. Interestingly enough, we find a very high correlation coefficient of 0.98 between the two databases on the regional level. The correlation coefficient for applications to the IPC groups is only slightly smaller (0.96).

In a second step we compare the patent data from the European Patent Office with the data from the German Patent Office. However, the German data is only available on a federal state level (NUTS 1). A comparison on this comparatively high regional level of aggregation nevertheless leads to a high correlation coefficient of 0.98. The comparison of sector distribution is conducted on the three-digit-IPC-level and again delivers a comparatively high level of correlation (0.89).

Overall, we conclude that focusing on the patent dataset of the European Patent Office does not lead to biased results and we employ this data in our subsequent analysis.

The OECD database offers a list of applications to the European Patent Office featuring applicants’ names and addresses. Moreover, the REGPAT Database provides lists of the priority years and International Patent Classifications (IPC) of inventions. In order to construct a dataset including all necessary information we merged these lists via the unique application identification number. We then extracted applications with priority year 2008 and a German applicant’s address. Doing so left us with a list of 57,287 entries on patent applications. However, since many of these entries are related to the same inventions (e.g. because of

---

1 The REGPAT Database is available from the OECD on request.
3 The priority year is the year of the first filing of a patent application to any patent office in the world. It is thus most closely connected to the date of invention (OECD 2009). We therefore follow the advice issued by the OECD (2008 and 2009) and use the priority date in order to reflect innovative drive.
4 Observations with wrong or missing values were deleted from the dataset.
5 The choice of our sample year is based on the availability of data on the sector structure on the NUTS 3-level. This sort of data, which will be described in the subsequent section, was only available for the year 2008.
joint patent applications), the underlying number of patents is considerably smaller.

In the next step, we transform the data from the generated list of patent application entries to patent count data on the regional and sector level. This task is not easy to solve because of two reasons. Firstly, it is very possible that multiple applicants, coming from different regions, jointly apply for the same patents. Secondly, an innovation can belong to more than one IPC classification and therefore to different sectors of the economy. And of course, both cases can occur together.

In order to illustrate the applied, quite complex procedure we use an imaginary example. In Table 1 we show the structure of the dataset derived from the OECD REGPAT Database. The first column of the dataset contains the unique patent application identification number (Appln_id) for an innovation. The second column identifies the applicant via a unique identification number (Person_id). The third column reports the region of the applicant’s office residence (Reg_code). In a number of cases, the applicant’s address could not be allocated to a single NUTS-3-region. In this case the dataset includes the same application various times, once for every involved region. In these comparatively rare cases an equal share of the invention has to be attributed to all regions involved. Column four reports what share of the invention is attributed to the referring NUTS 3-region. Often firms cooperate in research and development, which leads to joint inventions and joint patent applications. Column five reports the share of an invention which can be attributed to the applicant named in column two (App_share). As shown in the example, all applicants receive the same share of the patent. In by far the most cases, applicant shares for the same patent add up to one. However, this is not the case whenever at least one of the inventors comes from abroad (see e.g. the case described by Appln_id 6). Column six reports the IPC for every entry in the dataset (IPC). Whenever an invention belongs to different IPCs at the same time, the dataset contains one entry for every IPC classification.

In order to deal with the case of multiple IPCs for the same invention, we calculate and report the share of an invention, which is attributed to the referring IPC class (IPC_share), in column seven. As in the case of applicants from different regions we attribute the same share of the invention to every IPC class involved. Finally, each entry in our dataset has to be attributed to the IPC class of the referring region. The share of each entry is given by Final_share, which is calculated by multiplying the Reg_share, App_share and IPC_share. It is worth noting that the sum of final shares adds up to one over all entries with the same application identification number (Appln_id) when-

<table>
<thead>
<tr>
<th>Appln_id</th>
<th>Person_id</th>
<th>Reg_code</th>
<th>Reg_share</th>
<th>App_share</th>
<th>IPC</th>
<th>IPC_share</th>
<th>Final_share</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>10</td>
<td>DEA34</td>
<td>1.00</td>
<td>1.00</td>
<td>B65D</td>
<td>0.33</td>
<td>0.33</td>
</tr>
<tr>
<td>1</td>
<td>10</td>
<td>DEA34</td>
<td>1.00</td>
<td>1.00</td>
<td>E06B</td>
<td>0.33</td>
<td>0.33</td>
</tr>
<tr>
<td>1</td>
<td>10</td>
<td>DEA34</td>
<td>1.00</td>
<td>1.00</td>
<td>F21V</td>
<td>0.33</td>
<td>0.33</td>
</tr>
<tr>
<td>2</td>
<td>20</td>
<td>DE133</td>
<td>1.00</td>
<td>1.00</td>
<td>G01B</td>
<td>1.00</td>
<td>1.00</td>
</tr>
<tr>
<td>3</td>
<td>20</td>
<td>DE133</td>
<td>1.00</td>
<td>1.00</td>
<td>H05K</td>
<td>1.00</td>
<td>1.00</td>
</tr>
<tr>
<td>4</td>
<td>20</td>
<td>DE133</td>
<td>1.00</td>
<td>1.00</td>
<td>D06M</td>
<td>1.00</td>
<td>1.00</td>
</tr>
<tr>
<td>5</td>
<td>30</td>
<td>DE724</td>
<td>1.00</td>
<td>0.50</td>
<td>A61J</td>
<td>1.00</td>
<td>0.50</td>
</tr>
<tr>
<td>5</td>
<td>40</td>
<td>DE300</td>
<td>1.00</td>
<td>0.50</td>
<td>A61J</td>
<td>1.00</td>
<td>0.50</td>
</tr>
<tr>
<td>6</td>
<td>50</td>
<td>DEF02</td>
<td>1.00</td>
<td>0.50</td>
<td>F01K</td>
<td>1.00</td>
<td>0.50</td>
</tr>
<tr>
<td>7</td>
<td>60</td>
<td>DEA16</td>
<td>0.50</td>
<td>1.00</td>
<td>G06F</td>
<td>1.00</td>
<td>0.50</td>
</tr>
<tr>
<td>7</td>
<td>60</td>
<td>DEA17</td>
<td>0.50</td>
<td>1.00</td>
<td>G06F</td>
<td>1.00</td>
<td>0.50</td>
</tr>
<tr>
<td>8</td>
<td>70</td>
<td>DEE05</td>
<td>1.00</td>
<td>0.33</td>
<td>E21B</td>
<td>0.50</td>
<td>0.17</td>
</tr>
<tr>
<td>8</td>
<td>70</td>
<td>DEE05</td>
<td>1.00</td>
<td>0.33</td>
<td>C25D</td>
<td>0.50</td>
<td>0.17</td>
</tr>
<tr>
<td>8</td>
<td>80</td>
<td>DE21H</td>
<td>0.50</td>
<td>0.33</td>
<td>E21B</td>
<td>0.50</td>
<td>0.08</td>
</tr>
<tr>
<td>8</td>
<td>80</td>
<td>DE21H</td>
<td>0.50</td>
<td>0.33</td>
<td>C25D</td>
<td>0.50</td>
<td>0.08</td>
</tr>
<tr>
<td>8</td>
<td>80</td>
<td>DE212</td>
<td>0.50</td>
<td>0.33</td>
<td>E21B</td>
<td>0.50</td>
<td>0.08</td>
</tr>
<tr>
<td>8</td>
<td>80</td>
<td>DE212</td>
<td>0.50</td>
<td>0.33</td>
<td>C25D</td>
<td>0.50</td>
<td>0.08</td>
</tr>
</tbody>
</table>

Source: Own calculation.

---

3 This is due to the fact that German postcodes do not always coincide with NUTS 3-regions.

4 While in the original dataset the IPC classification is available on the eight-digit-level, we aggregated the IPCs of the inventions to the four-digit-level, which led to a reduction of the dataset to 36,675 observations.
ever all applicants are residing in Germany. Summing up the final shares over the entire dataset leads to the number of 22,340.16 patent applications to the European Patent Office in the priority year 2008 which can be attributed to applicants located in Germany.

In order to get a first impression of the regional distribution of patents we add up the final shares for each German NUTS 3-region. The results are shown in Figure 1. It is easily visible that the average number of patent applications in West German regions is larger than in East Germany. While the average number of patent applications per NUTS 3-region in West Germany is 64.6, it amounts to only 14.8 in East Germany. A Welch-t-test indicates that this difference is statistically significant on the 99 percent confidence level.

In both East and West Germany there is a huge degree of inter-regional variation in patent applications. However, these variations can hardly be attributed to differences in innovative activity since (i) the NUTS 3-regions differ considerably in size and are thus likely contain very different numbers of enterprises and (ii) the sector structure of the regions differs heavily. In order to generate an unbiased picture of innovative activity in German regions, it is therefore necessary to correct for the number of economically active enterprises and the sector structure of the referring regions.

**Firm data**

In order to control for the number of enterprises and the structure of the regional economy (in the necessary depth), detailed data on the NUTS 3-level is necessary. However, such data is not available from official statistics for secrecy reasons. We therefore make use of the firm database of Creditreform. Creditreform is the largest company information service in Germany. Organized into 130 local branches, Creditreform collects data on all economically active firms in Germany. Besides information on the location, sector classification, legal status, date of foundation, trade register details, sales and employees, the database also includes information on firms’ balance sheets and creditworthiness. For our study, Creditreform provided us access to its entire firm database for the year 2008. As of 31 December 2008 the database contained 3,954,721 economically active enterprises.

Figure 2 shows the distribution of enterprises contained in the Creditreform Database on German NUTS 3-regions. In the Creditreform Database, enterprises are attributed to German regions according to the location of the company. Whenever a firm is active in various sectors, the Creditreform Database reports the industrial sector of an enterprise in which the company generates the biggest turnover.

The number of firms in the Creditreform Database is considerably larger than that in the German Sales Tax Statistics 2008. However, since various professions (e.g., journalists, advocates or physicians), as well as small businesses, are exempted from the sales tax, this does not come as a surprise.

---

*Figure 1: Number of patent applications to the European Patent Office by NUTS 3-regions in Germany, priority year 2008*  
Source: Own calculations based on the OECD REGPAT Database.
to their headquarters. This procedure is quite useful for our purposes since patent applications in general are also filed by the headquarters of the inventing firms.

The Creditreform Database also includes information on the industrial sector in which a firm generates its largest turnover. We use this information to calculate the sector structure on the NUTS 3-level. The dataset contains sector information in the NACE Rev. 2 classification on the two-digit-level. In Figure 3 we show how the firms in the Creditreform Database distribute over different sectors. For visualization reasons we aggregate the sector information to the one-digit-level in the figure.

While the patent data employed make use of the IPC classification mentioned previously, the firm data from the Creditreform Database uses the NACE Rev. 2 nomenclature. In order to make both classifications compatible, we use the concordance table provided by Lybbert and Zolas (2012) to transform the IPC classification into the NACE Rev.2 nomenclature (two-digit-level).

Results

In this section we use the datasets described to shed some light on the question of which regions perform best in terms of relative innovative capacity. In a first step we refine the picture on innovation performance given in Figure 1 by calculating and illustrating the number of patent applications to the European Patent Office per enterprise. Figure 4 shows the results for German NUTS 3-regions for patent applications with priority year 2008. A comparison of Figure 1 and Figure 4 indicates that correcting for the number of firms tends to increase the differences between East and West Germany. The average number of patent applications in West German regions amounts to 0.005, while the figure is only 0.001 in East Germany. Applying a Welch-t-test to the data again delivers the result that the difference is statistically different from zero on the 99 percent confidence level.

In the next step we study whether patent applications per enterprise differ between different industrial sectors. As Figure 5 clearly depicts, there are, in fact, considerable differences in the number of patent applications between sectors. By far highest number of patent applications per enterprise is found in the mining and quarrying sector (B). Above-average values can also be found for the manufacturing sector (C), the sector covering activities of households as employers, undifferentiated goods- and services-producing activities of households for own use (T) and the sector concerned

11 We make no attempt at weighting the enterprises by sales or employees, since this information is only available for a subsample of all enterprises in the dataset.

12 For roughly 5 percent of the firms in the Creditreform Database (218,317 cases) no sector classification was available. The referring enterprises are summarized in the group 'no sector information'.

The lowest number of patent applications per enterprise can be detected in the sector covering wholesale and retail trade, the repair of motor vehicles and motorcycles (G), the real-estate sector (L) and the ‘other services’ sector (S). In the light of this comparatively strong variation between sectors, the sector structure should be taken into account when judging regions’ relative innovative capacities. As the indicator of relative regional innovative capacity described earlier accounts for this sector variation, we make use of this indicator in the following.

In Figure 6 we show the resulting values for the indicator of relative regional innovative capacity. The displayed patterns imply that the strong difference between East and West German regions in patenting activity is much less pronounced after correcting for both the number of enterprises and the industrial structure of the regional economy. The average indicator value in West Germany (−0.011) is only slightly higher (i.e. less negative) than its East German counterpart (−0.013). However, the difference is still significant on the 95 percent confidence level.

When studying Figure 6, one might also hypothesize that urban districts tend to perform systematically better than rural districts in terms of innovation. An inspection of Figure 7, which summarizes the ten least and most innovative regions, substantiates this speculation. Only one out of the ten least performing regions is an urban district (Landshut UD). And only two of the best performing regions are rural districts (Erlangen-Höchstadt RD and Heidenheim RD). In fact, the average number of patents per enterprise in urban regions (−0.008) turns out to be significantly higher than that in rural regions (−0.013). This difference is significant on the 99 percent confidence level.

Conclusions

In this paper we deliver empirical evidence on the relative innovative performance of German NUTS 3-regions, based on patent applications at the European Patent Office. The constructed indicator, which can be downloaded from the internet page of the authors, indicates that there is a considerable variance in innovative capacity on the regional level. This holds true even after controlling for region size and industrial structure. We find rural regions to be systematically less innovative than urban areas. However, we only find a slight difference between East and West German regions when controlling for the regional sector structure. Thus, given the prevailing industrial structure there is little difference between East and West German innovative capacity.

However, the prevailing disadvantageous industrial structure of East Germany nevertheless lowers the number of absolute inventions. As a result, the num-

http://www.hu-c-h.de/berlemann/.
The number of patents per enterprise is considerably lower in East Germany. This fact probably contributes to the failure of East Germany to reach West Germany’s level of per capita GDP.

References


BMBF (2009), Forschung und Innovation für Deutschland: Bilanz und Perspektive, Bonn and Berlin: Bundesministerium für Bildung und Forschung.


---

**Figure 6**

Relative regional innovative capacity by NUTS-3-regions in Germany, 2008

Source: own calculations based on the OECD REGPAT Database and the Creditreform Database.

---

**Figure 7**

Ten least and most innovative German NUTS-3 regions 2008

The dark bars stand for rural districts, the lighter ones for urban districts.

Source: Own calculations based on the OECD REGPAT Database and Creditreform Database.