

Broekel, Tom; Brenner, Thomas

**Working Paper**

## Local factors and innovativeness: an empirical analysis of German patents for five industries

Papers on Economics and Evolution, No. 0509

**Provided in Cooperation with:**

Max Planck Institute of Economics

*Suggested Citation:* Broekel, Tom; Brenner, Thomas (2005) : Local factors and innovativeness: an empirical analysis of German patents for five industries, Papers on Economics and Evolution, No. 0509, Max Planck Institute of Economics, Jena

This Version is available at:

<https://hdl.handle.net/10419/31853>

**Standard-Nutzungsbedingungen:**

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

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

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

**Terms of use:**

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

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

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



# 0509

**Local Factors and Innovativeness – An Empirical  
Analysis of German Patents for Five Industries**

**by**

**Tom Broekel  
Thomas Brenner**

The *Papers on Economics and Evolution* are edited by the  
Evolutionary Economics Group, MPI Jena. For editorial correspondence,  
please contact: [evopapers@econ.mpg.de](mailto:evopapers@econ.mpg.de)

ISSN 1430-4716

© by the author

Max Planck Institute of Economics  
Evolutionary Economics Group  
Kahlaische Str. 10  
07745 Jena, Germany  
Fax: ++49-3641-686868

# *Local Factors and Innovativeness - An Empirical Analysis of German Patents for Five Industries*

Tom Brökel and Thomas Brenner

Max Planck Institute of Economics  
Evolutionary Economics Group  
Kahlaische Str. 10  
07745 Jena, Germany

---

## Abstract

A growing body of work emphasizes the role that the spatial component plays in the innovation process. These perspectives brought the region's infrastructure and its endowment with crucial factors into the focus of research. Given that these factors do significantly influence the innovativeness of local firms, it is important to identify precisely which regional characteristics matter.

The aim of this paper is to identify a number of key influences out of a multitude of structural factors that are thought to influence the firm's innovation activity. We examine more than eighty variables that approximate the financial, geographical and social-economic factor endowment of a region.

The variables are tested with a linear and log - linear model. The two staged procedure examines the variable's bivariate correlation with patent data of five industries. Based on these outcomes multivariate regression models are applied in the second stage. The results for the different models are compared and their advantages and disadvantages are discussed. We find a strong impact of economic agglomeration, extramural science institutions and human capital. In the case of human capital, especially the graduates at the technical colleges are collocated with high regional innovativeness. Furthermore, significant differences are observed for the five industries and for using the two models.

# 1 Introduction

Recently, the spatial component of innovation has attracted much attention through the discussion of concepts such as the *diffusion of innovation*, *innovative milieus*, *regional innovation systems* and *technological infrastructure*. They raise the question of why some regions are more innovative than others and find the answer among other things in the regions infrastructure and their endowment with crucial factors. Especially the concept of ‘technological infrastructure’ (Feldman and Florida, 1994) puts the focus on the agglomeration of a geographically defined infrastructure. The spatial proximity of its components “. . . promotes information transfer and spill-overs that lower the costs and reduce the risks associated with innovation” (Feldman and Florida, 1994, p. 214).

Given the influence of regional factors on the innovation performance of local firms, it is important to identify precisely which regional characteristics or factors matter for their innovativeness. Here the literature offers a wide range of studies. A variety of variables, approximating a regions endowment and characteristics, have been found to be crucial, such as human capital, firm size, etc. However the literature taking into accounts all of these factors on an equal base is rather thin.

We show that the approach of an isolated investigation of single factors, or just small groups of factors, bears the risk of spurious correlations and the overestimation of their influence. Since the integration of a firm into its regional surroundings is complex and single connections are hard to isolate, a broader view is needed.

Two of the few studies that include a larger bandwidth of variables, which potentially influence the innovation activity of firms, are incorporated here. The study based on German regions by Weibert (1999) uses a linear model to investigate a multitude of factors, whereas the study by Feldman and Florida (1994), uses a log - linear model to investigate a somewhat smaller number of factors making up the ‘technological infrastructure’. Both models are based on different implications and reveal different results. Weibert (1999) finds employees with a professional training, the number of apprenticeship training posits as well as agglomeration crucial. In contrast, Feldman and Florida (1994) detects an impact of the ‘technological infrastructure’ containing: agglomeration of firms in related industries, university R&D, industrial R&D and business-service firms.

Therefore, we go beyond these existing studies by describing the region by a broad bandwidth of financial, geographical and social-economic variables, such as, e.g., GDP per capita, graduates of universities and venture capital companies. Thus, more than eighty variables approximating certain characteristics of the environment are included in the study. Further, we use Weibert’s *linear model* and the *log - linear model* from Feldman and Florida (1994). In order to find the relevant variables we employ for both models a bivariate correlation analysis combined with a multivariate regression.

Following a common approach for measuring the innovativeness of firms in specific regions, patent data is used as a dependent variable in the analysis. Brenner and Greif (2006), Malerba et al. (2000) and Pavitt (1984) find differences in the innovativeness of firms between industries. Therefore, we study five industries separately: in-organic chemicals (IN\_CH), organic chemicals (OR\_CH), automobiles (CAR), electronics (ELEC) and a mixed branch of optics, medical devices, computer engineering and measurement engineering (OPTIC).

Our study reveals that the results depend on which underlying model is used. Furthermore, we find differences between the studied industries. Nevertheless, economic agglomeration variables are undoubtedly crucial for innovation activities in the investigated industries. Similarly, application-orientated research institutions such as the Fraunhofer Society as well as human capital provided by technical colleges are highly correlated with high regional innovativeness. The paper proceeds as follows. In the next section we discuss some theoretical considerations and describe the independent variables as well as the respective data sources. Section 3 discusses the spatial units, the use of patent data and the statistical method that is applied. The bivariate correlation results are given and discussed in section 6. The different multivariate regression models are introduced in section 7. Their results are presented and compared with the bivariate correlation analysis. Section 8 concludes.

## 2 Theoretical considerations

Based on the observation that the innovative activity differs among regions that are part of the same national innovation system, the question of this disparities' source is raised. The concept of *regional innovation systems* claims that the regions specific social-economic environment has a crucial impact on the firms innovativeness located in it. For example the region's specific endowment with economic factors, such as human capital, the existence of certain institutions, the activity of local authorities, as well as the interaction between actors in networks are considered here.

Although there is a basic agreement on the importance of the regional level there is no commonly accepted definition or precise understanding of the regional innovation system's elements. A variety of studies proclaim different regional actors and factors to be crucial for the firms' innovative performance. This results in a multitude of concepts and studies of how to describe and explore the 'speciality' of regions hosting very innovative companies.

While empirical-oriented studies naturally favor measurable factors, it has to be mentioned that there are also very prominent ideas that shift the focus onto the not directly measurable interactions of actors or sector-specific cultural foundations. For example the 'innovative milieu' approach is built around the idea of a shared cultural base in terms of social-economic problems and solutions. Here, only the coherence of the production systems, the culture and the most important actors allow such a milieu and its positive impact on the innovation activity of local

firms to develop (Weibert, 1999). Similarly the concept of ‘social capital’ draws the attention to the structure of relations between and among actors (Coleman, 1988).

Unlike these approaches, we assume that the structure of regional innovation systems can be described and analyzed on the basis of measurable factors. Although the impact of the firms surrounding environment is acknowledged, there is no common definition or list of the relevant elements of regional innovation systems. Thus, the question is raised: Who are the relevant actors and what are the important influencing factors in a regional innovation system?

For example the presence of certain sized firms (Brenner and Greif, 2006; Stenke, 2000); science institutions (Soete et al., 2002; ISI, 2000), various kinds of human capital (Soete and Stephan, 2003; Fröderer et al., 1998), financial resources (OECD, 2000; Peter, 2002), spillover, cooperation and networks (Pittaway et al., 2003) and many factors more, have been identified as important elements in a *regional innovation system*. All these studies revealed an important impact of the specific aspects examined.

However, most of the studies analyze the elements of the innovation process separately. They are also detached from other elements, in that these studies barely control for such factors as agglomeration, which can be measured by population density. An approach considering more than a single or a small number of factors is the ‘technological infrastructure’ by Feldman and Florida (1994). They look at the agglomeration of a geographically defined infrastructure that influences the creation and diffusion of innovation. It enhances the capacity for innovation as their respective regions develop and specialize in particular technologies and industrial sectors, by providing sources and networks of knowledge as well as expertise and technical resources (Feldman and Florida, 1994).

To apply this concept and analyze its influence on the local firms innovation performance, an understanding of the ‘technological infrastructure’ is needed. Feldman and Florida (1994) defines it as a) agglomeration of firms in related industries, b) university R&D, c) industrial R&D and d) business-service firms (Feldman and Florida, 1994, S. 211). Further, they include control variables for e) the population of the federal state (USA), f) an index of geographic concentrations and g) industry sales as proxy for demands for innovations generated within an industry. The functions describing variables b), c) and d) also include the number of corporate headquarters, federal-funded research and development centres and the stock of receipts for general management and consulting services.

Although this model includes a comparatively large number of variables describing the regions relevant characteristics, we find this model still too narrow. This is because there are the many other studies mentioned previously, which find additional variables to be crucially important. Therefore, we take up the approach by Weibert (1999) that includes a larger bandwidth of variables and combine it with some elements from the Feldman and Florida (1994) model.

Therefore, in order to identify the relevant variables out of a large pool of factors, all of them are treated equally in the beginning and no precise model is developed. Obviously, because of

its character as a cross section analysis, this study cannot empirically reveal the causal relationships between and among the elements as well as with the innovation output, nevertheless this is done by a theoretical discussion.

To measure the local firms innovativeness, we follow a common approach using patent data as a dependent variable for the analysis. It will be briefly described in the following section.

## 3 The dependent variables

### 3.1 Patent data

As an approximation of a region's innovativeness we use the patent applications per 100,000 inhabitants of the year 2000, published by the *Deutsches Patent- und Markenamt* (German Patent Office) in the year 2002 (Greif and Schmiedl, 2002). The spatial arrangement of patent applications relates to the residence of inventors. As regional units, the German *Raumordnungsregionen* are chosen, which divide the country into 97 separate regional units. These regions are formed on a basis that takes into account the geographic nature of such economic functions as commuter flows. With respect to patent data the place of residence and the work place of an inventor belongs most likely to the same *Raumordnungsregion*. Thus, the chosen spatial unit offers an adequate way to analyze the link between the social-economic endowment of such a region to its patent data.

The advantages and disadvantages of using patent applications to indicate innovations have been discussed at great length (see e.g. Feldman and Florida, 1994; Röhl, 2000; Malerba and Orsenigo, 1996). Therefore, we refrain from discussing this issue, except for one problem: the ambient conditions of a region may have been influenced by the economic success of innovations in the past. Because the time lag between innovations and their effect on the economic surroundings is unknown, it cannot be excluded that innovations may influence the economic situation of a region. This effect is not explicitly included in the analysis here, but will be part of the discussion and evaluation of the results.

Aware of the strengths and weaknesses of the information provided by patents, we will use them as dependent variables in this study. Based on the categorization of the patent data, we use the sum of patent applications from corporations and private persons. This excludes the patent applications of public science institutions, that potentially bias the results. This bias results from the trivial fact that the more science institutions are located in a region the more patents of this category will be applied for.

### 3.2 The industries

Next to the spatial dimension of innovation process, there is also a technological one. Because every industry has its own pattern of innovation activity, it is reasonable to differentiate the

analysis and compare a variety of industries. The categorization of industries used here is based on the *International Patent Classification* (IPC) as shown in Table 1.

The industries have been chosen because the patent data can be clearly assigned [see][Greif

Code	Name	IPC
IN_CH	in-organic chemistry	12
OR_CH	organic chemistry & petroleum processing industry	5, 13, 14, 15
CAR	automobile-, ship- & aircraft building industry	10
ELEC	electronics	30
OPTIC	optics, medical devices, computer engineering and measurement engineering	26, 27

Table 1: Industries, Codes and IPC

and Potkowik (1990) and all five industries are relatively R&D intensive (Audretsch, 1998; ISI, 2000). This guaranties a sufficient number of patents in every region and non-biased results caused by a ambiguous assignment of patent data to industries.

According to Malerba and Orsenigo (1996) industries can be categorized by the characteristics of their technology base. They found that for the innovation process in certain industries a ‘widening pattern’ corresponding to Schumpeter’s hypothesis of *creative destruction*. In these industries, new firms are found to be the most innovative. As such they just enter the market and drive the relatively ‘old’ firms out of the market.

In contrast, some industries reveal a ‘deepening pattern’, in the sense of Schumpeter’s hypothesis, the *firm internal accumulation of knowledge*. Here only a few large firms dominate the market and generate most innovations. They benefit from past innovations that are now being part of the firm’s internal knowledge stock.

Following Malerba and Orsenigo (1996), a technological regime of the ‘widening pattern’ type is more likely to be relevant for OPTIC and IN\_CH. On the other hand the ‘deepening pattern’ is more likely to be found for CAR, OR\_CH and ELEC. This because according to Malerba et al. (2000), an increasing importance of external sources of technical knowledge is associated with a ‘widening’ pattern, so that public research should be relatively more important for the innovativeness of firms belonging to IN\_CH and OPTIC.

Pavitt (1984) classifies the industries by the organization of their innovation activity and the characteristic of the base-technology’s change. His classification takes the source, the kind and the effect of the innovation into account. Applying this categorization to the industries studied here, ELEC and OR\_CH are *science based*, implying strong connections to public science institutions. Thus, all studied industries except CAR can be expected to show a strong connection to public research institutions, although for different reasons.

The *production intensive* industries, here CAR and OPTIC, are divided into two subcategories: *scale intensive*, this accounts for CAR, and *specialized suppliers*, which holds for OPTIC. The accumulation of ‘tacit knowledge’, i.g. not codified knowledge, is crucial for *scale intensive*



industries (Pavitt, 1984). Therefore, firms benefit from being collocated with the emitters of knowledge, because such knowledge diffuses mainly via face-to-face contacts (Audretsch, 1998). For firms of the *scale intensive* industry, the most important source of technological know-how are suppliers and consulting engineers, whereas for the *specialized suppliers*, customers and users are more crucial (Pavitt, 1984). This means that the innovation activity of CAR should be positively affected by the agglomeration of firms and industries. The innovativeness of OPTIC firms should be positively related to public science institutions because they are not only partners for R&D, but also users of their products.

## 4 The independent variables

In this section the 80 independent variables are presented. For an easier orientation and later interpretation, they are sorted into different subsections where the same contexts is shared.

The variables used are listed in Table 3 in the appendix. In a cross-section analysis, time lags between the effects of a change in the independent variable on the dependent variable are observed to be a problem. Patent applications are of short-term and medium-term stability in volume as well as in spatial and sectorial structure. Additionally, they can be used in empirical analyses with economic data even in greater temporal differences (Greif and Schmiedl, 2002). Nevertheless, we use social-economic data from the year 2000 if possible. Some exceptions can be found in Table 3 in the Appendix. The variables will be discussed in detail in the following subsections.

### 4.1 Agglomeration

It is widely accepted that firms located in larger cities have advantages in generating innovations (Koschatzky, 2001). Agglomerations are advantaged because of the higher average utilization of their public infrastructure. In addition, the availability of trained workers, the spatial proximity to potential cooperation partners and a better access to relevant markets work in favour of firms located in agglomerations (Koschatzky, 2001).

Spillovers between corporations, institutions and other participants in the innovation process have a spatial component. The exchange of ‘tacit knowledge’ is bound to face-to-face contact (Audretsch, 1998). It is trivial, therefore that the likelihood of benefiting from such spillovers is higher in agglomerations.

On the other hand, agglomerations also have disadvantages. Along with a high population density are a lower disposability of industrial estates and higher costs of labour and taxes (Nerlinger, 1998; Rohr-Zänker and Müller, 2003). This evidence suggests that it is only up to a certain level of agglomeration where the advantages exceed the disadvantages. Beyond this critical level, the disadvantages will be dominant.

To account for such an influence it is common to use dummy variables (see e.g. Brenner and

Greif, 2006). As a very rough approximation for agglomeration, the dummies R1 for regions containing agglomerations (*Agglomerationsräume*), R2 for regions with cities (*verstädterte Räume*) and R3 for rural areas (*ländliche Räume*), are taken into consideration.

Besides these dummies, we include a number of variables that either measure the concentration of economic activity directly, such as the population density (POP\_DEN), nature related area per inhabitant (NATURE), density of employees in employment (EMPL\_DEN), density of work places (WORK\_DEN), type of region (REG\_TYPE), or measure something that is clearly related to the density of economic activity, such as persons per household (PERS\_HH), collection rate of communities for trade tax (COLL\_COMU) and price per m<sup>2</sup> for building land (BUILD).

To test for differences in the patent activity in the ‘Neuen Länder’ the dummy EAST is included in the study. A number of studies on the innovativeness of firms in those regions have found that private R&D is still significantly below the level of West Germany (see e.g. Felder and Spielkamp, 1998; Rammer and Czarnitzki, 2003; Röhl, 2000). Even with massive public research projects in this region, its lack of network structures, the poor capital market (especially for risky investments) the ‘brain drain’ and the lack of big corporations’ headquarters are likely to negatively influence firm innovation activity in East-Germany.

## 4.2 Human capital

In the literature it is frequently claimed that human capital is crucial to the innovation process (see e.g. Fröderer et al., 1998; Soete et al., 2002). The lack of highly qualified workers is especially a substantial constraint for innovations (Kugler, 2001).

It is unknown whether school leavers (SCHOOL), undergraduates (UNGRAD), graduates (GRAD\_...) and apprentices (APPR and APPR\_DEN) are going to stay in the region after their education. In general, highly qualified persons are less mobile than less qualified persons (Stenke, 2000). It can be argued that a person will first look for employment in the region. If they have to choose between equal jobs in different regions, they will prefer the job offered closest to their current location. Hence, the potential human capital approximates the human capital pool from which regional firms can select their employees from.

The amount of potential human capital effects, due to selection mechanisms and competition, the quality of the local firms employees. However, in Germany we have to take into account that the foundation of universities was used to advance rural areas and regions lacking in infrastructure (Blume and Fromm, 2000). Consequently politics distorted the above selection mechanisms. This leads us to expect a lower influence of universities than predicted by economic theory.

According to the German educational system, graduates are broken down into graduates of universities and the graduates of technical colleges. Because technical colleges are more often specialized in the same technological fields as local corporations (Beise and Stahl, 1999), a

stronger effect on the innovation activity can be expected for them.

Relatively more difficult is the task of estimating the importance of the various education fields. Naturally, math & natural science and engineering graduates are included into the study, because they are the researchers and engineers of the future. Their influence will likely depend on the industry analyzed. Furthermore, we include graduates of economics. They are especially crucial during the founding process of a business, where services and consultants play an important role (Nerlinger, 1998).

In addition, apprentices (APPR and APPR\_DEN) are also important for the diffusion and generation of knowledge (Soete et al., 2002). the number of school leavers with a qualification for university entrance (SCHOOL) is included as well, because the long-term development of the human capital potential is partly determined them.

While the above concentrate more on the potential human capital still being educated, the following variables deal with the actual human capital participating in the innovation process.

We split human capital up into different social-economic groups. It can be argued that especially young employees (YOUNG\_...) are relevant for innovation activity. They are representing creativity and ideas coming directly from the educational system into corporations, the so-called 'knowledge transfer via heads'<sup>1</sup> (ISI, 2000).

The share of female employees (FEMALE) and the share of potential employees in work (EMPLOY), as well as the rate of unemployment (UNEMP), are approximating the extent to which the endowment of a regions human capital is used. The potential impact of the share of employees with a professional training (EMP\_TRAINED), the share of employees with low qualification (EMP\_LOW) and with high qualification (EMP\_HIGH) are further measures of the quality of local human capital. Commuting furthermore effects a region's human capital. The capability of a region to absorb human capital from other regions increases the quality of the over-all human capital through selection and accumulation processes (Stenke, 2000). To take this into account, the numbers of commuters moving into the region (COM\_IN) and leaving it (COM\_OUT) are included. The number of courses at adult education centres (AD\_EDU) represent the opportunities for off-the-job training, i.e. the approximation of the change of the in-the-job human capital quality.

All variables indicating a high quality of human capital in a region are expected to have positive influences, although it has to be acknowledged that these variables are often strongly related to agglomeration effects.

It has to be pointed out, that universities and technical colleges not only provide qualified human capital, but they can also be as research institutions and cooperation partners. It is difficult to separate those two functions from each other and measure independently their impact on the patent activity of local firms. Both effects need to be considered when interpreting the results. Nevertheless, we tried by including the number of faculties and the number of graduates at uni-

---

<sup>1</sup>Translated by the author.

versities and technical colleges, to differentiate between these different roles in the innovation process.

### 4.3 Presence of universities UNI and technical colleges TC

Universities (including technical colleges) and extramural science institutions are major players in the innovation process. Together, these represent approximately one third of the over-all R&D capacities in Germany (ISI, 2000). Furthermore, they represent the core of formal and informal regional networks (Soete et al., 2002). Besides this, they also offer support for business foundations, consultants, use of laboratory equipment and are an important source for spin-offs (ISI, 2000).

Spillovers from universities and technical colleges to local firms are generated through various mechanisms, such as cooperation, master theses, internships, movement of employees, and informal contacts between employees. Adding to that, universities function as knots in formal and informal networks, easing up the flow of knowledge and information (Dybe and Kujath, 2000).

In order to measure the presence of universities and technical colleges, the numbers of faculties from different scientific fields are totaled for every region (...F\_UN and ...F\_TC). The data is segregated for universities and technical colleges because we aim to also understand the difference in the impact of these two kinds of institutions. Usually universities spend a larger share of their budget on research, varying between 30 percent for clinics and 60 percent for engineering faculties. For certain faculties of technical colleges this share is as low as 5 percent (Beise and Stahl, 1999).

At the same time, technical colleges have a stronger local focus and are more application oriented. Technical colleges “have gained a reputation for down-to-earth research and applicable engineering know-how, compensating the shortcomings of universities which are oriented towards basic research” (Beise and Stahl, 1999, p. 5).

To sum up, universities and technical colleges are expected to have a positive influence on the innovation output of local firm, because of their combined function as supplier of qualified human capital and as research institutions. But there is the danger of overestimating these effects. The difference in the impact of universities and technical colleges is likely to depend upon the studied industry. There is evidence that industries, dominated by larger firms, are more likely to cooperate with universities, while industries characterized by the ‘widening pattern’ will probably gain more from the collocation with technical colleges (Beise and Stahl, 1999).

### 4.4 Extramural science institutions

Naturally, extramural science institutions (*esi*) are generating many patents themselves. As mentioned in Section 3, those patents are not included in this study. Their role as partners in

cooperation, as emitters of a wide range of spillovers, as suppliers and customers are the focus here. A survey by Nicolay and Wimmers (2000) stated that 82 percent of innovative cooperation had contact to *esi*. Independent from the size of firms, two thirds of them were referring to institutions located in the same region.

Another aspect is the role they play in the foundation process of small innovative firms. Those firms can share the laboratories of the *esi* and find partners, customers or suppliers in these institutions. Adding to that, *esi* are themselves emitters of spin-offs, that are likely to be located close by (Nerlinger, 1998).

Similar to universities and technical colleges, different industries prefer to cooperate with different institutions. Firms with a high R&D intensity are more likely to cooperate with *esi*, while firms having a middle R&D intensity cooperate rather with universities and technical colleges (Spielkamp and Vopel, 1998). For the industries included here, this implies a stronger relationship between *esi* and ELEC as well as OPTIC, because they show a higher R&D intensity (Pavitt, 1984).

The structural factor *extramural science institutions* consists of the numbers of research-fellows in the different regional research institutions in a region. We consider the ‘big four’ institutions in Germany: the *Helmholtz Association* (HGF), the *Max Planck Society* (MPG), the *Fraunhofer Society* (FHG) and the *Leibnitz Association* (LEIB). It is important for the later interpretation of the results to note that the MPG and the FHG are concentrated in the southwest of Germany and that they are often located next to universities (ISI, 2000). While the FHG focuses on applied research, the MPG is more into basic research (Beise and Stahl, 1999). This leads to the forecast that the MPG will have a stronger relationship to the industries that depend more on basic research, such as OPTIC, OR\_CH and IN\_CH, while the FHG is more likely to be important for CAR and ELEC.

The HGF consists of 14 large scale institutes all over Germany. Because of the limited data, the expected positive influence might not be significant. The institutes of the LEIB have been part of programs to help regions lacking in infrastructure, especially regions in the former GDR (ISI, 2000). According to the argument in Section 4.1 this spatial correlation with factors that influence innovation negatively might dominate the results for the variable LEIB.

## 4.5 Economy structure

The variables pooled in this subsection reflect inter-industry economic structures. Patents are mainly generated in the secondary sector (manufacturing), i.e. if a region has a higher stake in the tertiary sector it has a disadvantage in generating relatively high numbers of patents. Therefore, we include the share of these two sectors (EMP\_SEC2 and EMP\_SEC3) and the gross added value per employee in these two sectors (VA2 and VA3) in the analysis. The share of employees in the business service sector (EMP\_SERVICE) measures the local availability of business services, such as consultants, financial services, etc. We expect them to positively

influence the patent activity of local firms (Preißl and Wurzel, 2001). The number of business foundations per inhabitant (FOUND) symbolizes an active and subserve economic climate in the region which reduces the risks and increases the probability of realizing innovations. Besides this, the foundings is one way of turning an invention into an innovation, by entering the market with an idea or a new product. Thus good conditions for the foundation of businesses are an incentive for doing research in the first place and, thus, generating inventions. Acs and Audretsch (1992) find a positive relationship between innovation activity and business registrations in regions, so that we expect a strong positive effect for all industries.

The turnover per employee (TURN) is a variable accounting for various factors. It reflects the market success of regional products. Therefore, the turnover is a variable measuring the ability of local firms to turn their inventions into innovations and selling them. The same holds for the turnover in foreign countries per employee (TURN\_ABROAD). “Exports activities seem to be one of the major determinants of a firm’s propensity to patent” (Encaoua et al., 2000, p. 325). Both turnover variables are also affected by firm size and agglomeration (Röhl, 2000), although they are expected to have positive impacts on patent activity.

#### 4.6 Financial facilities

R&D usually requires investments. Especially small- and medium-sized enterprises (SME) need external financial resources. The survey on the financial situation of SMEs by von Reden and Struck (2002) reveals that SMEs rank reserved granting of credits highest among external constrains to innovation.

Furthermore, in Germany the universal bank principle is practiced, implying a lack of necessary know-how (Peter, 2002) and a focus on securities and personal financial circumstances rather than on business ideas or personal qualifications (Kugler, 2001). Loans by suppliers and customers are often bound to long term relationships (Pfirrmann et al., 1997), which are usually not available for young firms. Meanwhile the transaction costs for finding and convincing external private persons to invest into the business are normally high and therefore such sources are only of limited relevance for young firms.

This signifies that entrepreneurs will first use their private or family’s resources (Nerlinger, 1998) and the resources provided by public programs. Especially the programs by the KfW Bankengruppe (KfW banking group) are relevant. The multitude of regional and national programs makes it difficult to find acceptable data. The variable public subsidiaries per inhabitant (ECO\_SUP) tries to add up a variety of such programs.

Venture capital (VC) is becoming increasingly important in Germany. In addition to financial resources, VC-firms provide guidance, consulting and expertise. These firms gain through contracting a partial control over the management and this kind of capital is therefore called ‘informed capital’ (Baltzer, 2000). In the USA evidence was found that firms partly financed by VC are growing faster and have lower rates of bankruptcy (Soete and Stephan, 2003). They

apply more often for patents, but do not have a higher share of R&D resources (OECD, 2000). In Germany the market for venture capital is small in comparison with the Anglo-Saxon countries. Thus, its importance should not be overestimated. But its remarkable growth during recent years justifies including the number of offices of venture capital companies who are members of the *German Venture Capital Association e.V.* (VC) in the study, which we expect to have a positive influence on the patent activity.

Other variables belonging to this structural factor are the GDP per inhabitant (GDP) and wage & salaries in manufacturing industry per employee (WAGE), accounting for the ability of the region to provide private capital for entrepreneurs and investments. The investments per inhabitant (INV) is a variable standing for diverse factors. On the one hand, large investments are only possible if there are ways of financing them. On the other hand, investments are determined by other factors, such as the economic structure. The public debt per inhabitant (DEBT) gives a hint of the financial situation of communities, being important financiers for infrastructure and local networks.

The relation between financial facilities and the innovativeness of a region is subject to the causality problem. This holds especially for public finances and VC-corporations. If a region's firms are very innovative, a lot of spin-offs or start-ups are likely to occur and the attraction of the potential market for VC-corporations improves. The same holds for public finance. Because innovative firms are more likely to generate profits, the income of communities through tax payments will rise. Therefore, these relationships should be analyzed by a time-series study, which is not done here.

#### 4.7 Regions attractiveness

To attract highly qualified workers and keep them in the region, attractive occupations and pleasant surroundings in the sense of soft location factors, have to be offered (Stenke, 2000). Often they are difficult to measure and are overlaid by 'hard' factors (Grabow et al., 1995). Nevertheless their impact can be considerable (Weibert, 1999). The choice variables for inclusion in the analysis are difficult because it is often not the quantity, but the subjectively perceived quality that is relevant. Weibert (1999) uses in his study the number of overnight stays per 1,000 inhabitants in a region as a proxy. But this number is varying strongly, therefore the more stable endowment with beds in hotels, motels and resorts is used here (BEDS).

The recreation area per inhabitant (RECRE) is self-explanatory as well as the places in kindergartens per inhabitant (KINDER), whereby KINDER is strongly correlated with the dummy for the 'Neue Länder', indicating a significant difference in the social infrastructure between the two parts of Germany (Brenke et al., 2002). The number of automobiles per inhabitant (CARS) is surprisingly negatively correlated to the population density ( $r = -0.419^*$ ) and the rate of

unemployment ( $r = -0.642^{**}$ ).<sup>2</sup> This leads to the conclusion that in urban regions public transport is a substitute for the holding of private automobiles, as well as a lower automobile density in the ‘Neue Länder’.

Further variables are the rate of persons moving into the region (MOVE\_IN) and leaving it (MOVE\_OUT) in terms of changing residence. It has to be pointed out that both variables are strongly correlated ( $r = 0.945^{**}$ ).

After having examined the theoretical base and the proclaimed effects of most of the variables used, we now introduce the empirical model.

## 5 Procedure of empirical analysis

Empirically, this study describes which specific factor endowment correlates with high innovativeness of firms in a region. It contributes to the existing research insights on the collocation of regional structural elements with firm innovativeness. Following Weibert (1999), this will be achieved in a two-stage procedure: first by a bivariate correlation analysis followed by a multivariate regression investigation. In contrast to Weibert (1999), we test his *linear model* in comparison to the *log - linear model* used by Feldman and Florida (1994). Both models are introduced in Section 5.1.

This section presents the methods used in this study. First, an overview over the steps taken is given. Then, each step is described in detail.

In the first step two data sets are created, one containing the original data, the other being the transformed (logarithmized) data, as described in Section 5.2. Next the correlation matrices, using Pearson’s Correlation Coefficient are produced. Its results are presented in Section 6. This section also contains the first implications that are derived from the bivariate analysis using the original data (*linear model*) as well as the transformed data (*log - linear model*).

This is followed by a pre-selection of independent variables included into the multivariate regressions, described in Section 7.1. Next the multivariate regressions are applied on the pre-selected data testing the relative importance of the variables. Its outcomes are presented in section 7.3. A discussion follows.

### 5.1 Introduction of the models

The following section describes the procedure of the empirical analysis. The empirical investigation is divided into two parts. First, the variables are examined separately. Their standalone relationship to the dependent variable is investigated with the Pearson’s Correlation Coefficient  $r$ . In addition, we focus on strong correlations among the independent variables. It is acknowledged that this type of analysis is endangered by spurious correlations. But, as it is shown in

---

<sup>2\*\*</sup> indicates a significance level of 0.01 whereas \* stands for a level of 0.05. No star (\*) states an insignificant result.



the theoretical part, each of the variables has a potential influence on innovation activity. In addition, critical high correlation with other independent variables are also taken into account. The second part of the study is the multivariate regression analysis. Its aim is a more differentiated insight into the relationships between the independent variables and patent data. Therefore the multivariate regression analysis investigates the relative importance of different independent variables on the local firms innovation activity (Weibert, 1999). In Section 7 some more comments are made on this analysis.

Before the empirical analyses are conducted, some statements need to be made about the functional relationship between the dependent and independent variables relevant for the bivariate and multivariate case. While Weibert (1999) is just assuming linear relationships between the variables, Jaffe (1989) and Feldman and Florida (1994) orientate on a classical economic Cobb-Douglas production function and suggest a non-linear relationship, that can be transformed through linearization into a log - linear model. This well-known linear regression is referred to as the *linear model* in the following. The multivariate regression model used later on is constituted as:

$$Y = d + b_1X_1 + b_2X_2 \dots + b_nX_n + u_t \quad , \quad (1)$$

with  $X_1 \dots X_n$  as  $n$  independent variables,  $Y$  as dependent,  $b_1 \dots b_n$  as regression coefficients,  $d$  as constant,  $t$  as number of the variables values and  $u_t$  as statistical error term with the usual characteristics (Eckey et al., 2001).

As mentioned, in a single independent variable case this model assumes a linear relationship with the dependent variable and an additive connection of the independent variables in the multivariate case.

In contrast, the power function used by Jaffe (1989) and Feldman and Florida (1994) is widely used in economic theory, especially as a production function (Samuelson and Nordhaus, 1998). It allows us to model decreasing marginal products of the independent variables as well as increasing marginal products, whereas the *linear function* only accounts for constant marginal products. In a single variable form, it is written as:

$$Y = d * X^c \quad . \quad (2)$$

The parameter  $c$  determines the functional form of the relationship between the independent and dependent variable. If the parameter  $c$  equals 1 a linear function is expressed, otherwise a non-linear relationship is revealed. The parameter  $d$  is a constant.

The *power function* as a multivariate regression model is written as:

$$Y = d * X_1^{c_1} * X_2^{c_2} \dots * X_n^{c_n} + u_n \quad , \quad (3)$$

with  $c_1 \dots c_n$  as the  $n$  regression coefficients.

We assume such a functional relationship implies the following differences to the linear model.

First, instead of searching for the best fitting linear model, non-linear relationships are investigated. The best fit is achieved by varying the power coefficient, which can be understood as the elasticity of that variable. That is the percentage change of the dependent variable if the independent variable changes by one percent. As a second difference to the *linear model*, the independent variables are connected multiplicatively in the multivariate form.

To test such models, the standard OLS-estimation does not work. The tests for such non-linear regression models are iterative estimation procedures that minimize the sum of the squared model residuals, (see e.g. Bühl and Zöfel, 2002). Although those procedures are implemented in the software, they are seldom used. It is common to transform those models into a log - linear relationship, (see e.g. Jaffe, 1989). For our study this has an additional advantage: the ‘stepwise’ mechanism can be applied, as described in more detail in Section 7.

## 5.2 The transformation of the non-linear regression model

It is obvious that the power function can be transformed into a function that is linear in its parameters by using a logarithmic transformation (see Hartung et al., 2001; Hippmann, 1994), that can be analyzed by the OLS-estimation. Unfortunately this is connected to two problems. First, all values of the variables need to be positive. This is done by a transformation of the values that “eliminates zero values yet preserves the relative ranking of . . . observations” (Feldman and Florida, 1994, p. 223). The changed variables will be marked with a  $\tilde{\phantom{x}}$  and the transformation appears as:

$$\tilde{X} = \log[10(1 + X)] . \quad (4)$$

Second, the logarithmic transformation implies that the error term is logarithmic normal distributed. Although this is theoretically quite doubtful, it is usually assumed for practical reasons (see Tiede, 1987; Chatterjee and Price, 1995):

$$Y = d * X_1^{c_1} * X_2^{c_2} \dots * X_n^{c_n} * u_t . \quad (5)$$

Accepting the above the *power function* is transformed to:

$$\ln(\tilde{Y}) = \ln(d) + c_1 * \ln(\tilde{X}_1) + c_2 * \ln(\tilde{X}_2) \dots + c_n * \ln(\tilde{X}_n) + \ln(u_t) . \quad (6)$$

This will be called *log - linear model* in the following. In the analysis, the variables are first transformed in the such a way, then they are tested with the correlation analysis and afterwards with the multivariate regression analysis, Section 7. Of course this transformation is not applied to dummy variables as R1, R2, R3 and EAST.

Having the models described, the results of the bivariate correlation analysis can be presented and discussed.

## 6 Bivariate correlation analysis

### 6.1 Bivariate results for IN\_CH

Here Pearson's Correlation Coefficient ( $r$ ) is used to analyse the data, thereby revealing the standalone relevance of the variables, investigating the five industries separately. But not only the correlation between the *independent variables* and the patent data is examined, also important correlations between the *independent variables* are shown. In the following both results, for the *linear model* as well as for the *log - linear model* are discussed.

In the following we will briefly discuss the most interesting findings for each industry, accounting for the two models.

In the *linear model* the industry IN\_CH shows a very broad scope of variables having a strong correlation to the patent data, see table 4. For five variables a  $r$  above 0.6 is reached. The two strongest variables are the engineering and economics graduates of technical colleges, GRAD\_TC\_ENG and GRAD\_TC\_ECO, implying a significant importance of technical colleges for this industry. These two variables are quite highly correlated with each other ( $r = +0.681^{**}$ ). Furthermore, GRAD\_TC\_ENG and GRAD\_TC\_ECO are highly correlated with proxies for agglomeration, as the population density, POP\_DEN, and the prices for building land, BUILD, ( $r > +0.6$ ), which are themselves strongly related to the patent data of this industry. The same holds for the variables for sales tax, S\_TAX, venture capital, VC, and with a negative relationship the number of apprentices, APPR, which are next in the ranking. This negative relationship results from APPR's highly negative correlation with the agglomeration variable POP\_DEN ( $r = -0.722^{**}$ ).

The *log - linear model*, see also table 4, reveals strong negative correlations with the apprentices, APPR, and with the type of region, REGTYPE, (where a high value for REGTYPE implies a region dominated by rural structures). In general, the results for the *log - linear model* outline the positive impact of agglomeration (POP\_DEN, BUILD), which are often in line with industrial agglomeration as approximated by work place density, WORK\_DEN and COLL\_COMU. Interestingly, the *log - linear model* finds no strong relationships between patent data and the number of graduates, the most important variable in the *linear model*. The only variable that stands out is the number of medical faculties at universities and technical colleges, MED\_F, although it has to be acknowledged that university faculties dominate this variable. A certain concentration of this variable towards agglomerations ( $r = 0.455^{**}$ ) as well as a correlation with the faculties for math & natural science at universities ( $r = 0.630^{**}$ ) might explain this finding, but very likely there are other connections that are not intuitive at this point.

This leads to the conclusion that the innovation activity of IN\_CH seems to rely on human capital, especially technicians, as well as financial resources and industrial agglomeration. This human capital aspect is reflected by the strong share of mechanical technology in the in-organic chemistry industry. Furthermore, good economic surroundings mainly achieved in agglom-

eration areas supports the innovation process. Nevertheless the underlying model drastically influences the results.

## 6.2 Bivariate results for OR\_CH

The *linear models* results for OR\_CH, displayed in Table 5, are weaker in terms of the correlation coefficients than the ones for IN\_CH. Nevertheless, they basically show the same structure. Here, compared to IN\_CH, the relatively lower weight of the economics graduates at technical colleges, GRAD\_TC\_ECO, is depicted. This implies that this industry relies more on engineers and technicians rather than economic surroundings. Nevertheless the financial infrastructure, VC, S\_TAX and GDP, are relevant.

Furthermore, the *log - linear model* again finds a dramatically weaker relationship with the graduates, e.g.  $r = +0.351^{**}$  with GRAD\_TC\_ENG, and reveals higher coefficients for economic and population agglomeration. The most important variables are the share of apprentices, APPR, and the prices for building land, BUILD, with the same relations as for IN\_CH. The importance of the financial infrastructure represented by TAX, P\_TAX and VC is similar to the findings of the *linear model*. Interestingly the founding activities, FOUND, appear in the top ten with a positive coefficient, implying strong spin-off activities or a comparatively small firm size structure of this industry leading to many new businesses. This goes in line with the importance of the surroundings, such as venture capital, for this industry.

Concluding OR\_CH seems to rely stronger on a supporting regional environment than IN\_CH, while the importance of human capital is only revealed in the *linear model*.

## 6.3 Bivariate results for CAR

In contrast to the *linear model* findings for the chemistry industries, CAR has a stronger connection to the university graduates (especially engineers) as shown in Table 6. This relationship is not surprising and reflects the expectations in Section 3.2 as well as the findings of a ‘deepening pattern’ by Malerba et al. (2000). It is underlined by the strong correlation with Fraunhofer institutes, FHG. Surely the strong concentration of the Fraunhofer Society in Southwest Germany plays a role, where also the CAR industry is mainly located.

In addition, the third variable being significantly correlated to the patents of CAR is BUILD, indicating economic attractive regions. This is not surprising, as CAR is one of the most important industries in Germany. Due to the collocation of CAR’s firms with suppliers and the tendency to concentrate in one region, the demand for building land will rise. A similar argument can be put forward for the variables connected to tax revenues, P\_TAX and TAX.

The comparatively strong correlation with VC is probably a result of the double-sided concentration to the southwest. The CAR industry is dominated by large corporations with big research facilities, not needing external venture capital.

Most of the findings of the *linear model* are affirmed by the *log - linear model*. Here again the graduates vanish out of top ranks as do the extramural science institutions, here the FHG and MPG. Instead, the negative impact of the location in the East German federal states is revealed to larger extent through the negative correlation with the compensating subsidiaries of communities, COMP, and the rate of unemployment, UNEMP.

The results for CAR show a strong effect of human capital for the *linear model*, but in contrast to the chemistry industry, university graduates are more relevant. In addition, the strong weight of FHG suggests a crucial impact of cooperation with applied science research institutions, both extramural and with universities (*linear model*). In contrast, the *log - linear model* suggests a stronger impact of large-scale firms, which create high public incomes.

## 6.4 Bivariate results for ELEC

A slightly different picture is revealed for ELEC. The *linear model*, see Table 7, shows strong relationships with agglomeration and human capital. The R&D facilities of ELEC are heavily concentrated in economically attractive regions, such as agglomeration areas, as the strong positive correlation with BUILD reveals ( $r = +0.722^{**}$ ). This is caused by the strong focus to agglomerations, such as Munich.

Venture capital companies are also located in Munich, finding in young innovative firms of the electronics industry a lucrative market. Such firms might be spin-offs or start-ups, having found a niche to compete against the big players of this industry, especially Siemens. To consult and support such firms, specific know-how is needed, that is usually not provided by the German universal banks, but by highly specialized venture capital firms. It is, however, not clear if the growth of the venture capital market relies on the innovation activity vice versa.

As mentioned in Section 3.2 universities and technical colleges are important partners and suppliers of human capital in this industry. Especially the graduates of engineering at technical colleges, GRAD\_TC\_ENG, and university math & natural science graduates GRAD\_UN\_NAT, and university engineering graduates, GRAD\_UN\_ENG, show strong correlations. Because those variables are not highly correlated with each other, the question for substitution effects between technical colleges and universities arises. The strong correlation with the aggregated number of math & natural science faculties at universities and technical colleges, NAT\_F, might be a hint for such an effect.

Surprisingly, the MPG as well as the institutes of the FHG have high coefficients, despite differences in their focus. They are not strongly correlated with each other ( $r = +0.474^{**}$ ), so substitution or complementation effects might take place.

As for the other industries, the *log - linear model* basically reveals only the influence of agglomeration, and the weak presence of this industry in the East German regions. All other variables vanish out of the top ten ranks. In general the results are, except for the higher coefficients very similar to the ones for the other industries.

Summing up, a mixture of science and education institutions, economic surroundings and agglomeration effects, which all seem to have an impact on the innovativeness of ELEC firms, are relevant. Again the underlying model dramatically influences the results.

## 6.5 Bivariate results for OPTIC

The results for the OPTIC industry look very similar to those for the ELEC industry, see Table 8. The variables BUILD, VC, MPG, AB\_TC\_ENG and FHG find their way to the top of the bivariate correlation coefficient ranking for the *linear model*. As a major difference to the results for ELEC the variables order changed slightly, but the interpretation stays the same as for ELEC. This is also true for the *log - linear model*.

Summarizing the bivariate correlation analysis: Although a number of interesting variables having a strong relationship with the patent data are identified for the different sectors, the underlying model dramatically influences the results. The impact of human capital and extramural science institutions, are especially important in the *linear model* vanish in the *log - linear model*. This effect is partly caused by the different treatments of 'zero-values'. While they obtain a value above zero due to the transformation (see Section 5.2) in the *log - linear model* they are zero in the *linear model*. Although this effect decreases the correlation coefficients for variables with a certain number of 'zero - values', as e.g. the graduates and the number of scientists employed by extramural science institutions, it is not explaining the tremendous change in the results. Running tests comparing the correlation coefficients of the *linear model* and the *log - linear model* for the variables including and excluding 'zero-values' shows an increase in the coefficient for the *log - linear model* if the 'zero-values' are excluded. Nevertheless the variables related to human capital and public research will not appear in the top ten for the *log - linear model* even if the 'zero-values' are excluded.

Additional, the correlation analysis has shown that only certain factors play a role for the innovation output of firms. All together, only 30 of the 80 factors show up in the Tables 4 to 8 in the Appendix. These are derived mainly from the lists of variables that describe agglomeration, financial facilities, human capital, and extramural science institutes. Variables that describe the region's attractiveness do not appear at all, while only a few variables representing the employees and general economy structure seem to be important.

# 7 Multivariate regressions analysis

## 7.1 Pre-selection of variables

The aim of the multivariate regression analysis is to find the best combination of independent variables that explain the patent data. As all variables are potentially influential and might be

interrelated, building up a theoretical model in advance is impossible due to its complexity. Nevertheless a statistical analysis is able, but taking some statistical restrictions into account, to find the best-fitting model and the combination of variables with the greatest explanatory power. By examining this best-fitting model further insights are gained on relevant variables describing highly innovative regions.

Before multivariate regression models can be applied, the multitude of variables need to be reduced. This is due to the ratio between the number of observations<sup>3</sup> and the number of variables tested by the multivariate regression. In comparison to the number of the variables values, there have to be clearly fewer variables included into the regression analysis (Backhaus et al., 2000). Therefore, the amount of 80 variables need to be reduced to an acceptable level.

To achieve this in a consistent manner, the selection method by Weibert (1999) is used here. As such the results of the bivariate correlation analysis are used as a basis. Initially all the variables, that have low explanatory power are excluded. It is difficult to define the respective threshold, but a level of an absolute  $r = 0.4$  seems to be reasonable in consideration of the correlation coefficients absolute range from 0 to approximately 0,75. This cuts the number of variables approximately by half. Subsequently all variables being correlated by  $r = 0.8$  or above to another independent variable are eliminated such that the variable with the higher correlation to the patent data will remain in the analysis. Through this, strongly correlated variables with the same theoretical background are eliminated except for the one with the strongest explanatory power (Weibert, 1999). The procedure has to be conducted separately for the *linear model* and the *log - linear model* leading to various selected variables. An exception is the number of faculties. They are highly correlated with the number of graduates in the *linear model*, and to a little lower, but relatively larger extent in the *log - linear model* as well, leading to their ex-ante exclusion in both models. The same holds for the variables representing a sum of other independent variables, such as GRAD\_ECO, the sum of GRAD\_FH\_ECO and GRAD\_UN\_ECO. Applying this selection mechanism, the number of variables that are tested by the multivariate regressions differ for each industry and model, ranging between 11 and 23 as shown in Table 9 (*linear model*) and table 10 (*log - linear model*). Before presenting the results, the models and the ‘stepwise’ mechanism are explained in more detail.

## 7.2 The ‘stepwise’ mechanism

During the ‘stepwise’ regression mechanism, the software tests the variables step by step, whereas the variable maximizing a certain quality factor stays in the model. The criterion used is the partial correlation coefficients significance level. A variable is included if the significance level is lower than a certain value, here 0.05. If the value is exceeding 0.1, the variable will be excluded (Backhaus et al., 2000). Thereby, this the regression filters out the best fitting combi-

<sup>3</sup>The number of observations equals the number of regional units, here 97 ‘Raumordnungsregionen’.

nation of variables from the pre-selected variables.

The need of transforming the non-linear model into the *log - linear model* is obvious here, because this mechanism does not work for non-linear relationships. Due to the transformation, this selection mechanism can be applied to the *linear model* as well as the *log - linear model*. Nevertheless, this selection mechanism is not applicable for all 80 independent variables. Therefore, we used the mentioned pre-selection that reduced the number to the maximum of 23 variables, which seems applicable here.

This procedure for the *log - linear model* has one major disadvantage. While it is easy to initially transform the data and then test the transformed data with a linear regression, the results only account for the transformed data, even if the parameters are transformed back. This makes it impossible to directly compare the results<sup>4</sup> of the *log - linear model* with those of the *linear model* testing the unchanged data (Hippmann, 1994).

Nevertheless, there are two ways to achieve a possible comparability between their results. One way is to calculate the residuals obtained by the re-transformed *log - linear model*<sup>5</sup>. Then the correct *adj. R<sup>2</sup>* can be calculated, (see e.g. Hippmann, 1994). Although this way is correct, it will not deliver the optimal parameters for this data and selected variables, but still the optimal parameters of the *log - linear model*.

The second way is to take the resulting parameters from the *log - linear model* and include them with the selected variables as starting values for a non-linear regression based on iterative estimation procedures. This optimizes the parameters for the original data and deliver a *adj. R<sup>2</sup>* comparable to the results of the *linear model*. This method is applied here.

Unfortunately, two problems occur. First, the variables that are pre-selected in their transformed form might be highly correlated in their untransformed form (larger than  $r = 0.799^{**}$ ). If such a high correlation is observed in the non-linear regression, the variable with the lower linear bivariate correlation coefficient is excluded.

The second problem is related to the significance of the parameters. Since the *linear model* contains only significant<sup>6</sup> variables (due to the 'stepwise' mechanism), the same needs to hold for the non-linear regression. To achieve this, insignificant variables are manually excluded. Of course this does not apply to the constant.

Summarizing this rather complex procedure, first there is the pre-selection relying on the bivariate correlation analysis done. Next the *linear regression* is applied to the original data. Therefore, its results are presented unchanged. For the *log - linear model* the variables are transformed. This data is analyzed by a standard OLS-regression and the 'stepwise' mechanism. But its results have to be 'retested' with a non-linear regression using the original data. Then possible insignificant variables are excluded to obtain the final results.

---

<sup>4</sup>The goodness of fit and the values of the parameters.

<sup>5</sup>Developing the original power function out of the parameters from the log - linear model and applying it to the original data.

<sup>6</sup>To the two-tailed level of 0.05.



## 7.3 General results for industries and functions

industry	linear model		log - linear model	
	variables	parameters	variables	parameters
IN_CH	COM_IN	+0.150*	APPR	-1.964**
	APPR	-0.188**	FOUND	+1.614**
	GRAD_TC_ECO	+0.008**	REGTYPE	-0.516**
	FOUND	+1.270*		
	adj. R <sup>2</sup>	<b>0.645**</b> +	adj. R <sup>2</sup>	<b>0.751**</b>
OR_CH	FOUND	+34.434**	FOUND	+3.6698**
	GDP	+3.873**	REGTYPE	-1.2618**
			TURN_ABROAD	+1.2093**
	adj. R <sup>2</sup>	<b>0.435**</b> +	adj. R <sup>2</sup>	<b>0.661**</b>
CAR	VC	+6.008**	P_TAX	+3.482**
	VA3	+3.161**	BUILD	+4.623**
			GRAD_UN_NAT	+1.899**
			EMP_TRAIND	+26.352**
	adj. R <sup>2</sup>	<b>0.323**</b> +	adj. R <sup>2</sup>	<b>0.664**</b>
ELEC	MPG	+0.039**	BUILD	+1.301**
	VA3	+1.993**	TAX	+1.071**
	GRAD_TC_ENG	+0.067**	WAGE	+1.491*
	FHG	+0.118*	UNEMP	-0.353*
	adj. R <sup>2</sup>	<b>0.740**</b> +	adj. R <sup>2</sup>	<b>0.766**</b>
OPTIC	MPG	+0.041**	BUILD	+0.955**
	GDP	+1.826**	UNEMP	-0.727**
	GRAD_TC_ENG	+0.055**	EMP_HIGH	+1.401**
	FHG	+0.154*		
	VC	+2.013*		
	adj. R <sup>2</sup>	<b>0.796**</b> +	adj. R <sup>2</sup>	<b>0.841**</b>

+ Results are obtained with variance robust estimates.  
\*\* Indicates a two-tailed significance level of 0.01 whereas \* stands for a level of 0.05. No star (\*) states an insignificant result.

Table 2: Results of the multivariate regressions

Before the results are presented, some remarks on the data have to be made. Due to the dependent variable, patents per 100,000 inhabitants, is per definition a continuous and non-negative value, including some 'zero-values', a tobit-regression model seems to be adequate (Wooldridge, 2003). But because there are only five 'zero-values' in IN\_CH, one in ELEC and non in all the rest of the industries (dependent variables), we will stay with the OLS linear regression model. In addition, and even more convincing, using a tobit-regression we cannot control for heteroscedasticity, which turned out to be present for the *linear model*.

To deal with heteroscedasticity, which is identified by the White's test, a robust variance estimator is used. The test and the robust variance estimators both go back to White (1980) and

are implemented in the software used<sup>7</sup>. Results obtained with the robust variance estimators are marked with an “+”.

Diagnostics for multicollinearity are done by the variance inflation factors (VIF) for each of the independent variables, (see Greene, 2000). According to informal rules of thumb applied to the VIF, there is evidence of multicollinearity if the largest VIF is greater than 10, (see Chatterjee and Price, 1995). We follow this rule here.

For a cross section analysis, autocorrelation is usually not relevant, because the variables are not systematically ordered (Greene, 2000, p. 525). The results of the different multivariate regressions are presented in Table 2.

The general higher goodness of fit for the *log - linear model* is initially quite noticeable. One explanation might be that in the non-linear model fewer problems with respect to multicollinearity and heteroscedasticity occur. The latter especially strongly influences the *linear model*. The observed heteroscedasticity in the linear models might be a sign of non-linearity in the data (Backhaus et al., 2000).

In general, agglomeration variables are more important in the *log - linear model* than in the *linear model*, while it is the other way around for graduate's variables and extramural science institutions. The same has been observed in the bivariate correlation analysis. The lesser importance of graduate variables and the extramural science institutions in the *log - linear model* might be caused by the fact that these variables take a value of zero for many regions. In a multiplicative model, such as the *log - linear model*, this implies that the number of patents in all these regions is predicted to be zero as well. This does not fit the data. Hence, the inclusion of these variables might even lessen the fit in the case of the *log - linear model*.

This leads us to conclude that neither of the two models is adequate to describe the innovation output of regions. The data seems to be of non-linear character, which causes problems with heteroscedasticity in the *linear model* and causes the *log - linear model* to offer the better fit. However, variables representing universities, graduates and extramural science institutions do not seem to have a multiplicative effect. Their effect seems rather to be additive. Hence, a combined model would probably be more adequate. Nevertheless, with such a model it would be impossible to conduct the analysis in the way we did here using 80 independent variables. An alternative option is to include both the linear and the *log - linear* analysis, in the interpretation. In the following section the results for each industry will be presented.

#### 7.4 Multivariate results for industry IN\_CH

Two variables are included in both the *linear* and the *log - linear* models. These are the share of employees that are apprentices, APPR, with a negative coefficient and the number of business foundations, FOUND, with a positive coefficient. The APPR coefficient is in all analyses

<sup>7</sup>STATA 8. For further discussion, (see Greene, 2000, p. 462-465).

negative due to its negative linear correlation with agglomeration variables, e.g.  $r = -0.722^{**}$  with the population density. This is also confirmed by the appearance of the regions type, REGTYPE, in the log - linear model. REGTYPE takes its lowest value for agglomerations, so that the negative sign indicates higher innovation outputs in agglomerations. The variable FOUND indicates the collocation of this industry's firms with attractive economic surroundings, and it is going in line with the 'widening pattern' implying that small young firms lead innovations (Malerba and Orsenigo, 1996). This is also supported by the significance of the amount of commuting in the region, COM\_IN, in the linear model.

Industries with a 'widening pattern' are more likely to cooperate with local technical colleges (Beise and Stahl, 1999). The human capital provided in this matter seems to be more important than the institution's function as a cooperation partner and an emitter of 'tacit knowledge', because technical colleges spend a much lower share of their budget on research projects (Beise and Stahl, 1999). This is confirmed by the lower bivariate linear correlation of the patent data with the number of engineering faculties at technical colleges, ENG\_F\_TC ( $r = +0.416^{**}$ ), than with the number of engineering graduates at technical colleges, GRAD\_TC\_ENG ( $r = +0.6596^{**}$ ). The strong linear correlation with other graduate numbers, e.g.  $r = +0.681^{**}$  with GRAD\_TC\_NAT, also underlines the importance of human capital for this industry.

In addition, the over-all the results for IN\_CH confirm the prediction based on the industry's inherent characteristics and the strong reliance on qualified human capital.

## 7.5 Multivariate results for industry OR\_CH

The results for OR\_CH are weaker, meaning that the  $R^2$  is overall comparatively small. This is probably caused by the industry's classification that may not be homogeneous enough. The very high coefficient for the number of new business registrations, FOUND, in the *linear model* seems to be critical, but it is partly confirmed by its coefficient in the *log - linear model* indicating a cubic relationship. Thus, an attractive economic surrounding seems to be important for OR\_CH as well, although the classification by Malerba and Orsenigo (1996) imply a 'deepening pattern' of innovation that is a lower importance of small firms for the innovation progress. The applicability of this pattern is affirmed by the influence of the share of sales in foreign countries, TURN\_ABROAD, in the *log - linear model*, hinting to a concentration of R&D activities in large companies. They are more likely to generate a larger share of their revenues in foreign countries.

In the same model the variable REGTYPE (the type of the region) shows the expected negative coefficient. A certain concentration to agglomerations is true for OR\_CH as well. Although the variable GDP, being part of the *linear model*, suggests that economic agglomeration has a positive impact mainly via the financial situation of the region. But combined with the importance of FOUND the surroundings for OR\_CH seem to be very accommodating to new business foundations. This may result from the biotechnology industry being at least partly included into

this industry.

As mentioned previously the results are comparatively weak, being possibly caused by an ill-defined industry in terms of the underlying IPC. The heterogeneous results are therefore difficult to interpret, but good economic surroundings and an export orientation seem to be most important here.

## 7.6 Multivariate results for industry CAR

For CAR no variable is part of both models. However the *linear model* shows a very low *adj. R<sup>2</sup>* of only 0.332\*\*, being the lowest for all models. Since this industry is dominated by large companies, (see Brenner and Greif, 2006), venture capital is not very relevant. Also the inclusion of the value added in the tertiary sector, VA3, seems only partly reasonable. CAR is a typical manufacturing industry, but the results show an relation with the service sector. Taking also the venture capital into account, the *linear model* points at a strong business service sector developing in the ‘shadow’ of the manufacturing industry.

More interesting are the results for the *log - linear model*. Aside from the coefficients for P\_TAX, which indicates the financial effect of large manufacturing plants on the region’s financial situation and the building land prices, BUILD, indicating economic attractive and active regions, the number of math & science graduates at universities are also included. Actually, the engineering graduates seem to be more appropriate here, but it has to be acknowledged that both are highly log - linear correlated ( $r = +0.732^{**}$ ). Unfortunately, this is also true for the university economics graduates,  $r = +0.923^{**}$ , leading universities in general being significant for this industry.

The last variable included in the *log - linear model* is the share of employees with professional training, EMP\_TRAIND, have to be mentioned. The strong positive coefficient is due to very low values, between 0 and 1. Nevertheless, its importance is not surprising, since it shows that the manufacturing dominated industry are driven by mainly engineering-based innovations.

Overall, the regional surroundings of highly innovative regions in this industry are formed by its large companies. But human capital, especially coming from universities as well as workers with a professional training, are related to this industry’s research capacities.

## 7.7 Multivariate results for industry ELEC

In the two models for ELEC no variable is included in both and the structures are quite different too, although the *adj. R<sup>2</sup>* are relatively equal. In the *linear model* extramural science institutions play an important role. This confirms the findings of Pavitt (1984), categorizing ELEC as a science based industry, and neglecting the implications by Malerba and Orsenigo (1996) because the ‘deepening pattern’ for ELEC would rather suggest a lesser importance of external knowledge sources.

However, it has to be acknowledged that both extramural science institutions relevant here, the MPG and FHG, show a strong concentration in the south-west of Germany, a pattern also true for ELEC. Albeit the application-oriented research of the FHG results in spillovers to firms of this industry. In contrast, the strong importance of the MPG, which is more oriented towards basic research, is somewhat surprising. But cooperation between firms of ELEC and institutes of the MPG are very likely and both parties benefit from such relations.

This it also true for the engineering graduates from technical colleges, GRAD\_TC\_ENG, that have a reputation of having a relatively stronger practical education than their counterparts from the universities. In addition, engineering subjects are especially important for ELEC firms. We have not evaluated to what extent the allocation of technical colleges differs from that of universities, but it seems that technical colleges are relatively more spatially spread. Besides this, there are a greater number of engineering graduates at technical colleges (23, 302) than at universities (15, 932).

VA3 controls for agglomeration, which has a positive impact here as well. It also accounts for active economic surroundings, especially in the business service sector.

In contrast to the *linear model* the *log - linear model* shows a stronger impact of agglomeration variables, headed by the building land prices, BUILD. The mentioned concentration of that industry in Southwest Germany, explains the positive signs of the coefficients for TAX and WAGE and the negative sign for the unemployment rate, UNEMP. Here a time series investigation needs to deliver insights into the causal relationship.

To sum up, the classification by Pavitt (1984) is affirmed. Human capital provided by technical colleges and extramural science institutions play an important role for innovation output, for industry ELEC. Their impact is much more crucial than it is for the previous mentioned industries. Innovation activity in ELEC also seems to be concentrated in agglomerations.

## 7.8 Multivariate results for industry OPTIC

Both models have the highest *adj. R<sup>2</sup>* for OPTIC compared to all the other industries. For the *linear model*, the results are very similar to ELEC, so that the results are just briefly summarized here. A noticeable difference is the importance of VC. In contrast to the results for CAR, here the impact of venture capital seems more reasonable. The rather small and medium sized firms of OPTIC and the ‘widening pattern’ of innovation relevant for this industry (see Malerba and Orsenigo, 1996) will profit from cooperation with venture capital, which not only providing financial capital but also important knowledge (Baltzer, 2000).

The *log - linear model* shows that agglomeration BUILD has a strong positive impact and that the unemployment rate UNEMP, which also controls for the location disadvantages in the ‘Neue Bundesländer’ ( $r = +0.817^{**}$  between EAST and UNEMP) has a negative impact. This indicates the concentration of this industry in Southwest Germany and it’s under-representation in the eastern parts, especially in terms of R&D capacities. The relevance of well-educated human

capital for this industry is represented by the share of highly qualified employees EMP\_HIGH. Similar to the ELEC industry, the firms of OPTIC are relying on extramural science institutions as well as good financial conditions and trained human capital. In addition, innovation activity is concentrated in economically attractive agglomerations.

## 8 Conclusion

The aim of the study was to find key characteristics of German regions that are endowed with highly innovative firms, while taking a very broad set of characteristics into account. This has been done for the 97 German 'Raumordnungsregionen' and five industries. The analysis was divided into two stages. First, the bivariate correlation analysis provided the basis for the interpretation of later results and for the pre-selection mechanism for the multitude of structural variables that are needed for the multivariate regression models at the second stage. It was shown that while the technological dimension of the innovation process is relevant the geographic component is also crucial. For the five industries different variables revealed a relationship with the patent data used as dependent variables.

For ELEC and OPTIC, especially extramural science institutions, the presence of technical colleges (in terms of the number of their graduates) and economic agglomeration variables showed a significant impact.

On the other hand, the two chemical industries accounted for more heterogeneous results. Nevertheless, a fairly strong relationship to agglomeration and especially to the number of firm foundations was revealed. Furthermore, IN\_CH seems to be more concentrated in agglomerations than OR\_CH. Also, in the bivariate analysis, it has a stronger connection to technical colleges and their engineering and economic faculties, whereas in the multivariate models the economic faculties seem to be especially crucial. But because all the graduates variables are strongly interconnected, it is difficult to separate their influence. Also it has to be acknowledged that the weak results for OR\_CH could be caused by an inadequate definition of that industry in terms of the aggregated IPC categories.

For the CAR industry economic agglomeration variables showed highly significant coefficients. Besides this, a strong connection with universities is found to be crucial, in contrast to those industries that are more oriented towards technical colleges. In the bivariate *linear* analysis the engineering graduates particularly revealed a strong influence, which is not fully confirmed by the multivariate models. Here the *log - linear model* points towards the higher importance of the math and university natural science graduates. As mentioned before, this is due to the strong interrelatedness of the different graduate variables.

As for the different models and methods applied, important implications have been established. While generally the multivariate regression results confirmed the results of the bivariate correlation analysis, strong differences have been revealed comparing the *linear model* and the *log -*

*linear model.*

The bivariate correlation analysis showed that variables with a large variance or containing a non-trivial number of 'zero-values' generally showed weaker results in the *log - linear model*. This was especially true for the number of university and technical college graduates, the number of their faculties and for the scientists of extramural science institutions. The *log - linear model* seems to be incapable of showing the real importance of these variables, so that agglomeration variables dominate the results.

At the same time, the results for the *log - linear model* obtained a higher *adj. R<sup>2</sup>*. Not surprisingly, heteroscedasticity did not show up using the transformed data. Hence, in general, the use of a *log - linear model* seems to be more adequate. A trade-off between obtaining better statistical results and being able to adequately include the important variables connected with graduates and research institutes results.

Besides the different aspects relevant for choosing an adequate model in first place, some key factors of regions that are endowed with highly innovative firms have been obtained. In general, the application-based focus of extramural science institutions as well as technical colleges is found to be highly important for the innovation activity of firms across most industries. Also economic agglomeration variables and the founding activity revealed an over-average impact, although the causal direction could be from the innovation performance to these factors and not the other way around.

As for the variables, although already 80 independent variables have been investigated, it has to be acknowledged that one important set of variables is not included in this study: the population of firms and employees in the industry under consideration. This factor has been tested as relevant for innovativeness in (Brenner and Greif, 2006). The combination of both data sets is planned for future research.

All in all we elaborated on the basic approach to evaluating specific factors separately. Here, analyzing the region's relevant factor endowment, involves high risks of overestimating the specific factors effect. While the biggest problem remains the selection of relevant variables to include in the models, the use of non-linear regression analysis offers an additional tool providing deeper insights into the structure of the data. Furthermore, the joint analysis of diverse industries showed the importance of the technological dimension of the innovation process.

This work marks a starting point for further research. It gives a first insight into the coherence between the endowment of a region with a crucial set of factors and the innovativeness of local firms. The analysis lacks a dynamic component, which would answer to what extent the variation of the identified factors have an impact on the firm's ability to generate patents over time. Another subject for further investigation might be the relationship between the different factors used in this work as independent variables.

## References

- Acs, Z. and Audretsch, D. (1992). *Innovationen durch kleine Unternehmen*. Wissenschaftszentrum Berlin für Sozialforschung, Berlin.
- Audretsch, D. (1998). Agglomeration and the location of innovative activity. *Oxford Review of Economic Policy*, Vol. 14, No. 2, Oxford.
- Backhaus, K., Erichson, B., and Plinke, W. (2000). *Multivariate Analysemethoden*. Springer - Verlag Berlin, Heidelberg.
- Baltzer, K. (2000). *Die Bedeutung des Venture Capital für innovative Unternehmen*. Shaker Verlag, Aachen.
- Beise, M. and Stahl, H. (1999). Public research and industrial innovations in germany. *Research Policy*, Vol. 28, No. 4, pages 397–422.
- BLK (2003). *Studien- und Berufswahl*. Bund-Länder-Kommission für Bildungsplanung und Forschungsförderung (BLK), BW Bildung und Wissen Verlag und Software GmbH, Berlin.
- Blume, L. and Fromm, O. (2000). Regionalökonomische Bedeutung von Hochschulen. *Kas-seler Wirtschafts- und Verwaltungswissenschaften, Band 11, Deutscher Universitäts Verlag, Wiesbaden*.
- Brenke, K., Geppert, K., Hopf, R., Pfeiffer, I., Spieß, C. K., Vesper, D., and Wagner, G. G. (2002). Bausteine für die Zukunft Berlins. *DIW-Wochenbericht 10/02, Deutsches Institut für Wirtschaftsforschung Berlin, Berlin*.
- Brenner, T. and Greif, S. (2006). The dependence of innovativeness on the local firm population an empirical study of german patents. *Industry and Innovation, Vol. 13, No. 1, forthcoming*.
- Bühl, A. and Zöfel, P. (2002). *SPSS 11 Einführung in die moderne Datenanalyse unter Windows*. Pearson Education Deutschland GmbH, München, 8. edition.
- BVK (2003). *Sitz der Mitglieder*. Bundesverband Deutscher Kapitalbeteiligungsgesellschaften e.V, <http://www.bvk-ev.de>.
- Chatterjee, S. and Price, B. (1995). *Praxis der Regressionsanalyse*. R. Oldenbourg Verlag, München, 2. edition.
- Coleman, J. (1988). Social capital and the creation of human capital. *American Journal of Sociology, Vol. 94, Supplement: Organizations and Institutions: Sociological and Economic Approaches to the Analysis of Social Structure*, pages 95–120.
- Dybe, G. and Kujath, H. J. (2000). *Hoffnungsträger Wirtschaftscluster*. Institut für Regionalentwicklung und Strukturplanung, Berlin.
- Easystat (2002). *Easystat- Statistik regional, CD-Rom*. Statistische Ämter der Länder des Bundes, Wiesbaden.
- Eckey, H.-F., Kosfeld, P., and Dreger, C. (2001). *Ökonometrie*. Betriebswirtschaftlicher Verlag Gabler, Wiesbaden, 2. edition.



- Encaoua, D., Hall, B., and Laisney, F. (2000). *The Economics and Econometrics of Innovation*. Kluwer Academic Publishers, Boston.
- Felder, J. and Spielkamp, A. (1998). *Innovationsstrategien und Forschungsaktivitäten ost-deutscher Unternehmen*. Zentrum für Europäische Wirtschaftsforschung GmbH, Mannheim.
- Feldman, M. P. and Florida, R. (1994). The geographic sources of innovation: Technological infrastructure and product innovation in the United States. *Annals of the Association of American Geographers*, Vol. 84, No. 2, pages 210–229.
- Fraunhofer-Gesellschaft (2000). *Befristete und Unbefristete Wissenschaftler*. Fraunhofer Gesellschaft, München.
- Fröderer, K., Krey, K., and Palme, K. (1998). Innovation und Mittelstand, eine Umfrage bei 1871 Unternehmen. *Beiträge zur Gesellschafts- und Bildungspolitik, Instiut der deutschen Wirtschaft, Köln*.
- Grabow, B., Henckel, D., and Hollbach-Grömig, B. (1995). Weiche Standortfaktoren. *Schriften des Deutschen Instituts für Urbanistik, Vol. 89, Stuttgart*.
- Greene, W. H. (2000). *Econometric Analysis*. Prentice Hall International (UK), London.
- Greif, S. and Potkowik, G. (1990). *Patente und Wirtschaftszweige*. Carl Heymanns Verlag KG - Köln, Berlin, Bonn, München.
- Greif, S. and Schmiedl, D. (2002). *Patentatlas 2002 Dynamik und Strukturen der Erfindungstätigkeit*. Deutsches Patent- und Markenamt, München.
- Hartung, J. et al. (2001). *Statistik für Wirtschafts- und Sozialwissenschaften: Deskriptive Statistik*. R. Oldenbourg Verlag, München, 2. edition.
- Helmholtz-Association (2000). *Programme - Zahlen - Fakten*. Helmholtz Association of National Research Centres.
- Hippmann, H.-D. (1994). *Statistik für Wirtschafts- und Sozialwissenschaftler*. Schäfer-Pöschel Verlag, Stuttgart.
- INKAR (2002). *Aktuelle Daten zur Entwicklung der Städte, Kreise und Gemeinden, CD-Rom, Berichte, Vol. 14*. Bundesministerium für Bauwesen und Raumplanung, Berlin.
- ISI (2000). *Endbericht an das BMBF Regionale Verteilung der Innovations- und Technologiepotentialen in Deutschland und Europa*. Fraunhofer Institut für Systemtechnik und Innovationsforschung, München.
- Jaffe, A. (1989). Real effects of academic research. *American Economic Review*, Vol. 79, No. 5, pages 957–970.
- Koschatzky, K. (2001). *Räumliche Aspekte im Innovationsprozess*. Wirtschaftsgeographie, Vol. 19, LIT, Münster.
- Kugler, F. (2001). Gründungsforschung in der Region Südthüringen. *Kurzbericht zum AIF-Projekt, Fachhochschule Schmalkalden, Schmalkalden*.

- Leibniz-Gemeinschaft (2000). *Gesamtpersonal 2000*. Wissenschaftsgemeinschaft Gottfried Wilhelm Leibniz e.V., Bonn.
- Malerba, F. and Orsenigo, L. (1996). Schumpeterian patterns of innovation are technology-specific. *Research Policy*, Vol. 25, No. 3, pages 451 – 478.
- Malerba, F., Orsenigo, L., and Breschi, S. (2000). Technological regimes and schumpeterian patterns of innovation. *Economic Journal*, Vol. 110, No. 463, pages 388 – 410.
- Max-Planck-Society (2000). *Jahrbuch 2001*. Max-Planck-Society, Munich.
- Nerlinger, E. (1998). *Standorte und Entwicklungen junger innovativer Unternehmen*. Nomos Verlagsgesellschaft, Baden-Baden.
- Nicolay, R. and Wimmers, S. (2000). *Kundenzufriedenheit der Unternehmen mit Forschungseinrichtungen*. Deutscher Industrie- und Handelstag (DIHT), Berlin, Bonn.
- OECD (2000). A new economy? The changing role of innovation and information technology in growth. *Information Society, OECD*.
- Pavitt, K. (1984). Sectoral patterns of technical change: towards a taxonomy and a theory. *Research Policy*, Vol. 13, No. 6, pages 343 – 373.
- Peter, V. (2002). Institutionen im Innovationsprozess. In *Technik, Wirtschaft, Politik*, Vol. 46. Schriftenreihe des Fraunhofer - Insituts für Systemtechnik und Innovationsforschung, Physica -Verlag, München.
- Pfiffmann, O., Wupperfeld, U., and Lerner, J. (1997). Venture capital and new technology based firms. In *Technology, Innovation and Policy*, Vol. 4. Series of the Fraunhofer Institute for Systems and Innovation Research (ISI), München.
- Pittaway, L., Robertson, M., Munir, K., Denyer, D., and Neely, A. (2003). Networking and innovation: a systematic review of the evidence. *International Journal of Management Reviews*, Vol. 5 / 6, No. 3&4, pages 137–168.
- Preißl, B. and Wurzel, U. (2001). Zur Rolle der privaten und öffentlichen Forschungseinrichtungen in europäischen Innovationssystemen. *DIW-Wochenbericht 30/01, Deutsches Insitut für Wirtschaftsforschung Berlin, Berlin*.
- Rammer, C. and Czarnitzki, D. (2003). Innovationen und Gründungen in Ost-deutschland. In *Studien zum deutschen Innovationssystem, Nr. 15-2003*. Zentrum für Europäische Wirtschaftsforschung GmbH, Mannheim.
- Röhl, K.-H. (2000). Die Eignung der sächsischen Agglomerationsräume als Innovations- und Wachstumspole für die wirtschaftliche Entwicklung des Landes. *Diskussionsbeiträge aus dem Institut für Wirtschaft und Verkehr Nr. 1/2000, Technische Universität Dresden, Fakultät für Verkehrswissenschaft, Dresden*.
- Rohr-Zänker, R. and Müller, W. (2003). *Projektbericht: Perspektive Ländlicher Räume und die Rolle der Raumordnung*. Institut für Entwicklungsplanung und Strukturforschung GmbH an der Universität Hannover, Hanover.
- Samuelson, P. and Nordhaus, W. (1998). *Volkswirtschaftlehre*. Wirtschaftsverlag Carl Ueberreuter, Frankfurt, Wien, 15. edition.

- Soete, B. and Stephan, A. (2003). Nachhaltiges wirtschaftliches Wachstum durch Innovationen: Die Rolle von kleinen und mittleren Unternehmen. *DIW-Wochenbericht 38/03, Deutsches Institut für Wirtschaftsforschung Berlin, Berlin.*
- Soete, B., Wurzel, U., and Drewllo, H. (2002). Innovationsnetzwerke in Ostdeutschland: Ein noch zu wenig genutztes Potential zur regionalen Humankapitalbildung. *DIW-Wochenbericht 16/02, Deutsches Institut für Wirtschaftsforschung Berlin, Berlin.*
- Spielkamp, A. and Vopel, K. (1998). Mapping innovative clusters in national innovation systems. In *Discussion Paper No. 98-45*. Zentrum für Europäische Wirtschaftsforschung GmbH.
- StBA (2000). *Absolventen der Universitäten und Fachhochschulen nach Fächergruppeneordnet*. Statistisches Bundesamt, Wiesbaden.
- Stenke, G. (2000). Grossunternehmen in innovativen Milieus, das Beispiel Siemes / München. *Kölner Forschungen zur Wirtschafts- und Sozialgeographie, Vol. 54.*
- Tiede, M. (1987). *Statistik*. R. Oldenbourg Verlag, München.
- von Reden, E. and Struck, J. (2002). Nutzen und Wirkung öffentlicher Förderprogramme für Wachstumsunternehmen. In *Wertorientiertes Start - Up Management*, pages 413 – 428. Verlag Franz Vahlen, München.
- Weibert, W. (1999). *Regionale Determinanten der Innovation*. Nomos Verlagsgesellschaft, Baden-Baden.
- White, H. (1980). A heteroscedasticity-consistent covariance matrix estimator and a direct test for heteroscedasticity. *Econometrica, Vol. 48*, pages 817–838.
- Wooldridge, J. (2003). *Introductory econometrics a modern approach*. South Western Publ., Mason, 2 edition.

## A Appendix

Variables, symbols and their sources		
patent data	variables	sources
patents per inhabitant in-organic chemicals	IN_CH	(Greif and Schmiedl, 2002)
patents per inhabitant organic chemicals	OR_CH	(Greif and Schmiedl, 2002)
patents per inhabitant electronic industry	ELEC	(Greif and Schmiedl, 2002)
patents per inhabitant automobile-, ship- and aircraft-building industry	CAR	(Greif and Schmiedl, 2002)
patents per inhabitant optics, medical devices, computer engineering and measurement engineering	OPTIC	(Greif and Schmiedl, 2002)

<b>agglomeration</b>	<b>variables</b>	<b>sources<sup>8</sup></b>
population density	POP_DEN	(INKAR, 2002)
persons per household	PERS_HH	(INKAR, 2002)
nature related area per inhabitant	NATURE	(INKAR, 2002)
collection rate of communities for trade tax per inhabitant aggregated for 'Raumordnungsregionen'	COLL_COMU	(Easystat, 2002) [#]
density of employees in employment	EMPL_DEN	(INKAR, 2002)
density of work places	WORK_DEN	(INKAR, 2002)
type of region (classification by German Federal Office for Building and Regional Planning)	REGTYPE	(INKAR, 2002)
price per $m^2$ for building land	BUILD	(INKAR, 2002)
region contains agglomerations (Agglomerationsräume)	R1	(INKAR, 2002)
region contains cities (Verstädterte Räume)	R2	(INKAR, 2002)
rural areas (ländliche Räume)	R3	(INKAR, 2002)
region is part of former GDR	GDR	(INKAR, 2002)
<b>presence of universities &amp; technical colleges</b>	<b>variables</b>	<b>sources</b>
engineering faculties at universities & technical colleges	ENG_F	(BLK, 2003)
math & natural science faculties at universities & technical colleges	NAT_F	(BLK, 2003)
economics & business administration faculties at universities & technical colleges	ECO_F	(BLK, 2003)
medicine faculties at universities & technical colleges	MED_F_G	(BLK, 2003)
engineering faculties at universities	ENG_F_UN	(BLK, 2003)
math & natural science faculties at universities	NAT_F_UN	(BLK, 2003)
economics & business administration faculties at universities	ECO_F_UN	(BLK, 2003)
medicine faculties at universities	MED_F_UN	(BLK, 2003)
engineering faculties at technical colleges	ENG_F_TC	(BLK, 2003)
math & natural science faculties at technical colleges	NAT_F_TC	(BLK, 2003)
economics & business administration faculties at technical colleges	ECO_F_TC	(BLK, 2003)
medicine faculties at technical colleges	MED_F_TC	(BLK, 2003)
<b>structure of employees reported to social security</b>	<b>variables</b>	<b>sources</b>
share of employees at age 20 - 25	YOUNG_25	(INKAR, 2002)
share of employees at age 25 - 30	YOUNG_30	(INKAR, 2002)
share of employees at age 20 - 30	YOUNG	[#]
share of female employees	FEMALE	(INKAR, 2002)
share of employees fit to work	EMPLOY	(INKAR, 2002)
share of employees with low qualification	EMP_LOW	(Easystat, 2002)
share of employees with high qualification	EMP_HIGH	(Easystat, 2002)
share of employees with a professional training	EMP_TRAIND	(INKAR, 2002)
rate of unemployment	UNEMP	(INKAR, 2002)
rate of commuters into the region	COM_IN	(INKAR, 2002)
rate of commuters out of the region	COM_OUT	(INKAR, 2002)
number of courses at adult education centres per 10,000 inhabitants	AD_EDU	(INKAR, 2002)

<sup>8</sup>Variables marked with a # are results of personal calculation based on already listed variables.

<b>economy structure</b>	<b>variables</b>	<b>sources</b>
share of employees secondary sector	EMP_SEC2	(INKAR, 2002)
share of employees tertiary sector	EMP_SEC3	(INKAR, 2002)
share of employees in business service	EMP_SERVICE	(Easystat, 2002)
number of business foundations per capita	FOUND	(INKAR, 2002)
sales per employee	TURN	(INKAR, 2002)
sales in foreign countries per employee	TURN_ABROAD	(INKAR, 2002)
gross value added secondary sector	VA2	(INKAR, 2002)
gross value added tertiary sector	VA3	(INKAR, 2002)
<b>financial facilities</b>	<b>variables</b>	<b>sources</b>
number of venture capital companies and branch offices	VC	(BVK, 2003)
GDP per inhabitant	GDP	(INKAR, 2002)
wage and salaries in manufacturing industry per employee	WAGE	(INKAR, 2002)
investments per inhabitant	INV	(INKAR, 2002)
communities debts per inhabitant	DEBT	(Easystat, 2002)
general compensation subsidiaries of communities per inhabitant (Ausgleichseinnahmen der Gemeinden)	COMP	(INKAR, 2002)
public subsidiaries (Wirtschaftshilfen) per inhabitant	ECO_SUP	(INKAR, 2002)
overall tax gains per inhabitant	TAX	(INKAR, 2002)
income tax gains per inhabitant	IN_TAX	(INKAR, 2002)
trade tax gains per inhabitant	P_TAX	(INKAR, 2002)
sales tax gains per inhabitant	S_TAX	(INKAR, 2002)
<b>potential human capital</b>	<b>variables</b>	<b>sources</b>
school leaver with qualification for university entrance	SCHOOL	(INKAR, 2002)
undergraduates per 1,000 inhabitants	UNGRAD	(INKAR, 2002)
economics & social science graduates per inhabitant	GRAD_ECO	(StBA, 2000)
math and natural science graduates per inhabitant	GRAD_NAT	(StBA, 2000)
engineering graduates per inhabitant	GRAD_ENG	(StBA, 2000)
economics & social science graduates per inhabitant at technical colleges	GRAD_TC_ECO	(StBA, 2000)
math and natural science graduates per inhabitant at technical colleges	GRAD_FC_NAT	(StBA, 2000)
engineering graduates per inhabitant at technical colleges	GRAD_TC_ENG	(StBA, 2000)
economics & social science graduates per inhabitant at universities	GRAD_UN_ECO	(StBA, 2000)
math & natural science graduates per inhabitant at universities	GRAD_UN_NAT	(StBA, 2000)
engineering graduates per inhabitant at universities	GRAD_UN_ENG	(StBA, 2000)
share of employees (reported to social security) that are apprentices	APPR	(INKAR, 2002)
density of apprenticeship places	APPR_DEN	(INKAR, 2002)
<b>extramural science institutes</b>	<b>variables</b>	<b>sources<sup>9</sup></b>
employees Fraunhofer-Gesellschaft	FHG	(Fraunhofer-Gesellschaft, 2000)
employees Max Planck Society	MPG	(Max Planck Society, 2000)
employees Helmholtz Association	HEL	(Helmholtz-Association, 2000)
employees Leibniz Gemeinschaft	LEIB	(Leibniz-Gemeinschaft, 2000)
sum of employees extramural R&D Institutes	SCIENCE	[#]

<sup>9</sup>Variables marked with a # are results of personal calculation based on already listed variables.

regions attractiveness	variables	sources
recreation area per inhabitant	RECRE	(INKAR, 2002)
overnight stay beds for tourists per inhabitant	BEDS	(INKAR, 2002)
places in kindergarten per inhabitant	KINDER	(INKAR, 2002)
rate of moving in	MOVE_IN	(INKAR, 2002)
rate of moving away	MOVE_OUT	(INKAR, 2002)
automobiles per inhabitant	CARS	(INKAR, 2002)
fertility rate	FERT	(INKAR, 2002)

Table 3: Variables, symbols and their sources

IN_CH	linear	sig.	IN_CH	log - linear	sig.
GRAD_TC_ENG	+0.660	**	APPR	-0.616	**
GRAD_TC_ECO	+0.642	**	POP_DEN	+0.580	**
S_TAX	+0.621	**	REGTYPE	-0.571	**
VC	+0.619	**	WORK_DEN	+0.557	**
APPR	-0.613	**	EMP_TRAIND	-0.552	**
ECO_F	+0.596	**	VC	+0.532	**
P_TAX	+0.593	**	BUILD	+0.508	**
EMP_TRAIND	-0.575	**	COLL_COMU	+0.490	**
BUILD	+0.563	**	MED_F	+0.487	**
POP_DEN	+0.554	**	S_TAX	+0.482	**

Table 4: Bivariate correlations IN\_CH

OR_CH	linear	sig.	OR_CH	log - linear	sig.
GRAD_TC_ENG	+0.526	**	APPR	-0.678	**
S_TAX	+0.518	**	BUILD	+0.663	**
APPR	-0.504	**	REGTYPE	-0.660	**
VC	+0.500	**	EMP_TRAIND	-0.647	**
GDP	+0.499	**	POP_DEN	+0.646	**
VA3	+0.491	**	WORK_DEN	+0.625	**
P_TAX	+0.484	**	FOUND	+0.570	**
POP_DEN	+0.469	**	VC	+0.563	**
TAX	+0.468	**	GDP	+0.556	**
FOUND	+0.459	**	TAX	+0.549	**

Table 5: Bivariate correlations OR\_CH

CAR	linear	sig.	CAR	log - linear	sig.
GRAD_UN_ENG	+0.570	**	P_TAX	+0.734	**
FHG	+0.558	**	TAX	+0.727	**
BUILD	+0.531	**	VA3	+0.702	**
P_TAX	+0.509	**	BUILD	+0.690	**
VC	+0.507	**	GDP	+0.690	**
TAX	+0.487	**	COMP	-0.690	**
GDP	+0.470	**	IN_TAX	+0.681	**
GRAD_TC_ENG	+0.467	**	UNEMP	-0.629	**
MPG	+0.461	**	APPR	-0.611	**
INCOME	+0.452	**	S_TAX	+0.587	**

Table 6: Bivariate correlations CAR

ELEC	linear	sig.	ELEC	log - linear	sig.
BUILD	+0.723	**	BUILD	+0.753	**
VC	+0.682	**	TAX	+0.695	**
GRAD_TC_ENG	+0.663	**	P_TAX	+0.689	**
MPG	+0.659	**	APPR	-0.675	**
FHG	+0.635	**	COMP	-0.660	**
NAT_F	+0.611	**	IN_TAX	+0.649	**
GRAD_UN_NAT	+0.601	**	POP_DEN	+0.648	**
S_TAX	+0.580	**	VA3	+0.644	**
GRAD_UN_ENG	+0.577	**	EMP_TRAIND	-0.644	**
EMP_TRAIND	-0.574	**	GDP	+0.624	**

Table 7: Bivariate correlations ELEC

OPTIC	linear	sig.	OPTIC	log - linear	sig.
BUILD	+0.726	**	BUILD	+0.751	**
VC	+0.724	**	POP_DEN	+0.702	**
MPG	+0.712	**	APPR	-0.680	**
GRAD_TC_ENG	+0.677	**	EMP_TRAIND	-0.672	**
FHG	+0.668	**	P_TAX	+0.664	**
NAT_F	+0.645	**	TAX	+0.653	**
GRAD_UN_NAT	+0.636	**	S_TAX	+0.651	**
GRAD_UN_ENG	+0.617	**	WORK_DEN	+0.627	**
S_TAX	+0.601	**	COMP	-0.611	**
NAT_F_UN	+0.592	**	WAGE	+0.606	**

Table 8: Bivariate correlations OPTIC

IN_CH	OR_CH	CAR	ELEC	OPTIC
GRAD_TC_ENG	GRAD_TC_ENG	GRAD_UN_ENG	BUILD	BUILD
GRAD_TC_ECO	S_TAX	FHG	VC	VC
S_TAX	APPR	BUILD	GRAD_TC_ENG	MPG
VC	VC	P_TAX	MPG	GRAD_TC_ENG
APPR	GDP	VC	FHG	FHG
EMP_TRAIND	POP_DEN	GRAD_TC_ENG	GRAD_UN_NAT	GRAD_UN_NAT
BUILD	FOUND	MPG	S_TAX	GRAD_UN_ENG
TAX	EMP_TRAIND	GRAD_UN_NAT	GRAD_UN_ENG	S_TAX
GRAD_UN_ECO	TURN_ABROAD	EMP_DEN	EMP_TRAIND	EMP_TRAIND
REGTYPE	GRAD_TC_ECO	APPR	GRAD_TC_ECO	GRAD_TC_ECO
FOUND	REGTYPE	VA3	APPR	APPR
COM_IN	COM_IN	IN_TAX	VA3	REGTYPE
COLL_COMU	WAGE		GRAD_TC_NAT	GDP
GRAD_TC_NAT			EMP_DEN	GRAD_TC_NAT
EMP_HIGH			EMP_HIGH	EMP_DEN
WAGE			FOUND	EMP_HIGH
EMP_SERVICE			WAGE	WAGE
SCHOOL				FOUND

Table 9: Variables included in multivariate linear regression

IN_CH	OR_CH	CAR	ELEC	OPTIC
TURN_ABROAD	APPR	P_TAX	BUILD	BUILD
APPR	BUILD	BUILD	ECO_SUP	POP_DEN
POP_DEN	REGTYPE	YOUNG_30	NATURE	APPR
REGTYPE	UNGRAD	UNEMP	POP_DEN	EMP_TRAIND
EMP_TRAIND	EMP_TRAIND	EMP_TRAIND	S_TAX	P_TAX
VC	VC	S_TAX	COLL_COMU	S_TAX
BUILD	GRAD_TC_NAT	EMP_LOW	APPR	REGTYPE
COLL_COMU	POP_DEN	NATURE	EMP_TRAIND	WAGE
S_TAX	WAGE	VC	UNEMP	TURN_ABROAD
GRAD_UN_ECO	TURN_ABROAD	WAGE	REGTYPE	GRAD_UN_NAT
VA3	EMP_HIGH	APPR	WAGE	NATURE
	S_TAX	POP_DEN	TAX	EMP_DEN
	FOUND	TURN_ABROAD	TURN_ABROAD	VC
	COLL_COMU	REGTYPE	RECRE	UNEMP
	GRAD_UN_ECO	COM_IN	YOUNG_30	FOUND
	GDP	EMP_DEN	FOUND	RECRE
	EMP_SERVICE	FEMALE	EMP_LOW	COLL_COMU
	GRAD_UN_NAT	GRAD_UN_NAT	GRAD_UN_NAT	EMP_LOW
	MPG	ECO_SUP	EMP_DEN	ECO_SUP
	LEIB		VC	GRAD_UN_ENG
	ECO_SUP			MPG
	COM_IN			EMP_HIGH
	PERS_HH			GRAD_TC_ECO
				GRAD_TC_ENG

Table 10: Variables included in multivariate non-linear regression