Productivity, Age Structure and Monopolistic Competition: An Econometric Study for German Regions.

Stephan Brunow and Georg Hirte
Technische Universität Dresden
stephan.brunow@tu-dresden.de
georg.hirte@tu-dresden.de

November 2008
Manuscript prepared for the 3rd User Conference, Nürnberg

1 Introduction and literature review

Regional productivity is not uniformly distributed over German regions. Similarily, labor market data reveal strong differences in the age structure of the skilled labor force between these regions. These differences will increase during the demographic transition to an elder population currently under way in Germany. This is the point of departure for our study. We want to explore the impact of the age structure on regional productivity in Germany.

There is a rich literature of the effects of age on productivity using individual data. This work is typically motivated using a Mincer (1974) wage equation. Evidence concerning the life-cycle pattern of productivity is rather mixed, though more studies provide evidence of age dependency of individual productivity (see the survey in Skirbekk 2004). This age dependency can be explained by 'learning by doing', 'on the job training' [1] or experience. Thus, productivity will increase. On the other hand, if one gets older, productivity might decline because of obsolescens or depreciation of knowledge. Furthermore, the productivity profile over the life cycle might differ between occupations, industries or regions.

Moreover there are studies on the impact of the age structure of a firms labour force on the productivity of plants. Here, again, evidence on the age dependency is rather mixed (e.g. Haltiwanger et. al. 1999). However, there seems to be more evidence on a humped shaped pattern of productivity of plants with respect to age (e. g. Hellerstein et al. 1999, Hellerstein and Neumark 2004, Ilmakunnas and Maliranta 2005; see, however, Hellerstein and Neumark 1995 with opposite results).

From a regional point of view one could expect a similar distribution of productivity considering age. However, due to aggregation it might be less strong. The literature on regional age dependencies is rather short. As Bönte et al. (2007) show, not just the availability of human capital but also the age pattern of it matters for knowledge based firm startups. Lindh and Malmberg (1999) or Brunow and Hirte (2006) estimate an augmented Mankiew-Romer-Weil model and provide evidence that regional variation of productivity is partially explained by differences in population's age distribution. The work presented here is closely related to earlier work (Brunow and Hirte 2008), where we estimate a Lucas type production function and find a significant inverse u-shaped pattern of regional productivity with respect to age of the skilled labour force. While this study provides first evidence that the age structure is important for productivity on regional level further research is required. Since we use cross section data in that study we were neither able to control for the dynamics of firms or population nor for dynamic effects of externalities such as Marshall-Arrow-Romer (MAR) or Jacobs (1969).

The empirical identification of this kind of externalities starts with the work of Glaeser et al. (1992). Further discussion of identification is due to various work of Henderson (1997, 2003), Henderson et al. (1995) or Combes and Overman (2004). Combes et al. (2004) explore dynamic behaviour of local employ-

[1] See Mincer (1997)

ment and firm growth in France. Further they account for externalities. Their focus is on employment dynamics, firm growth and regional economic structure. However, they do not explore the effects of the age structure and do not focus on regional productivity. For this reasons, we do not follow their analysis.

Instead, we refer to Rosenthal and Strange (2004) who suggest to estimate a production function directly in order to capture productivity effects. On the other hand, they point out the problem to find appropriate information on inputs. However, on the regional level with a high level of aggregation, the problem might be less important considering labour force and land usage. Problems still remain when considering capital and materials. The availibility of data on material usage is still not given on regional level.

Baldwin (1999) develops a regional model of the New-Economic-Geography class which is based on monopolistic competition. This model can be seen as a neoclassical growth model. For empirical work it seems to be reasonable to assume some monopolistic power of firms. We follow Baldwin's theoretical work and derive an augmented regional production function which suits our purpose and can be estimated. Regional productivity depends on infrastructure area, labour and the total number of plants. The underlying theory causes an endogenous regressor, namely the number of plants. Therefore a second equation, the plants equation, is derived to deal with this endogeneity problem. Here, human capital enters the model. We are interested in producitivity differences induced by the age structure. So, we implement the age structure into this second equation. As a consequence, the impact of human capital on regional producitivity enters indirectly the production function via the number of firms.

The structure of the paper is as follows. The next section develops the theoretical framework in order to derive the empirical model. We first consider the production function and then the plant equation. After that we form hypotheses on the expected effect of variables and present some descriptive analysis. Thereafter, we present and discuss the estimates. The paper closes with a conclusion.

2 Theoretical background and empirical motivation

The model presented here follows the conceptual idea of Baldwins (1999) structure of the so-called constructed capital model and can be seen as a variant of Grossman and Helpman (1995). More precisely, we derive an empirically estimatable approach containing two structural equations. The first equation relates to regional production function with labor and the number of plants as input. Further variables such as infrastructure and diversity measures are used to controll for interregional disparities. The second equation is the establishment equation which deals with endogeneity of the number of plants. In the following we derive these simultaneous equations and motivate additional covariates.

2.1 Production function

As usual in the 'new economic geography' literature Baldwin considers two sectors for final consumption goods. There is a monopolistic and a competitive sector, M and C respectively. Each firm uses labour as variable input in production. While wages are the only cost in the competitive sector, there is an additional fixed cost requirement in the monopolistic sector. Total labour market W is separated in the labour requirements of the competitive and monopolistic sector, L_C and L_M, and a research sector H_M.

Total output of the competitive sector C can be written as

$$C = AL_C \tag{1}$$

where L_C is the labour input of the sector and A is the common total factor productivity. The production function of a monopolistic establishment is given by

$$x_i = Al_i$$

where l_i is the labour input of a single plant. Let L_M be the total number of workers and N_M the mass of regional establishments and assume symmetry of all firms in that sector, we can average firms production to

$$x_i = A\frac{L_M}{N_M}. \tag{2}$$

Aggregation of the monopolistic commodities to total sectoral output is, as usual, motivated with a CES-index.

$$M = \left(\sum_i^{N_M} x_i^{\frac{\sigma-1}{\sigma}}\right)^{\frac{\sigma}{\sigma-1}}, \quad \sigma > 1 \tag{3}$$

Here, σ reffers to the elasticity of substitution in production and not in consumption. Inserting (2) in (3) yields total output of the monopolistic sector.

$$M = AL_M N_M^{\frac{1}{\sigma-1}}. \tag{4}$$

Total regional output is the aggregate of the output of the competitive and monopolistic sector. Here, a Cobb-Douglas-Index aggregates these sectoral outputs. Using the production technologies (1) and (4) we derive the regional output which is given by

$$Y = AL_C^{1-\mu}\left[L_M N_M^{\frac{1}{\sigma-1}}\right]^{\mu}, \quad 0 \leq \mu \leq 1 \tag{5}$$

where μ refers to average sectoral shares rather than demand elasticities. This regional production function exhibits increasing returns to scale because of the monopolistic sector. In order to make regions comparable (5) is converted in per-capita terms.

$$\frac{Y}{W} = y = As_C^{1-\mu}s_M^{\mu}N_M^{\mu/(\sigma-1)}. \tag{6}$$

Cameron (1998) derives some similar model starting from a production function where labour and capital is used under constant returns to scale. Then he added further variables to total factor productivity which relate to regional R&D stock and yields a quiet similar model. However, the difference is that the N considered here is the knowledge capital of a region as suggested by e.g. Griliches and Lichtenberg (1984). Because of the transformation in per capita terms urbanisation measures are relaxed, as long as there are no externalities. However, here N_M enters the equation and increases total output with increasing returns to scale. Thus, regional differences in output per capita are due to some externalities driven by the mass of regional monopolistic establishments. If there are agglomeration economies, an estimated parameter should be positive and significant. Cameron (1998) surveys the literature, where N is a measure of the R&D capital stock. The estimated parameter is about 0.05-0.1, depending on the level of aggregation and country specifics. That means that an increase in R&D capital stock rises output by about 5-10%.

The discussion of N has to be done carefully in order to understand what N really measures. First, it is the number of regional establishments within a monopolistic sector. Second it can be seen as some measure of R&D capital if the monopolistic sector is R&D intensive. A third effect of N is that it relates to regional differences in total factor productivity (Cameron 1998). However, N_M is an endogenous variable which we are going to consider in the next subsection.

Taking logarithm of (6) and adding an error term yields the basic regression equation,

$$\ln y = \ln A + \frac{\mu}{\sigma - 1} \ln N + (1-\mu) \ln s_C + \mu \ln s_M + u_y \tag{7}$$

Total factor productivity A might depend on variables of the industry mix, localisation (MAR) and urbanisation measurements (Jacobs)[2]. While static MAR externalities comprise intraindustrial externalities, which are allready captured in N_M, so called Jacobs externalities deal with interindustrial externalities (see Glaeser et al. 1992, Henderson et al. 1995). Baldwins model does not allow considering externalities in the sense of endogenous growth. Martin and Ottaviano (1999) and Baldwin et al. (2001) introduce endogenous growth with the help of dynamic MAR externalities. However, we stick to the original model of Baldwin and insert additional agglomeration effects of the Jacobs type by following Combes et al. (2004). They use the logarithm of the number of industries and a diversity measure of employment between regional industries as urbanisation measures. Accordingly, we adopt their approach. However, considering a production function, total factor productivity might depend more on regional linkages and input-output relations between firms. Thus, we do not use employment diversity but diversity of firms and define DIV_N as the nagtive Herfindahl-index in the following way

$$DIV_N = -\ln \sum_i^I \left(\frac{N_i}{\sum_k^I N_k} \right)^2$$

where I is the number of industries and N_i is the number of employees within industry i[3]. If local

[2] See Combes et al. (2004).

[3] We dropped the time and regional index.

firms are concentrated within a single industry, the diversity measure becomes zero. If establishments are distributed over more than one industry, DIV_N, is positive and increases. The diversity measures relate to Jacobs externalities, especially to industrial linkages and intersectoral relationships within the production function. Thus, the diversity of firms DIV_N controlls for the industry mix and should be included to the regression. Furthermore it allows to capture regional variation in composition of the aggregated sectors.
A second measure, suggested by Combes et al., is. the logarithm of the number of industries located in a region which we call Ind.

As mentioned above, MAR externalities might also exist, and are allready included because of N. In addition we tried various measures suggested by the literature but because of a very high degree of collinearity we could not controll for these.

In order to deal with the special role of Eastern Germany we introduce a dummy variable $East$. Sensitivity analysis reveals that Berlin reaches almost the productivity of the western parts of Germany. Furthermore it is Germany's capitol. So, $East$ does not include Berlin and a $Berlin-$ Dummy is additionally introduced.

The primary study of Aschauer (1989) introduces public capital in a production function. A lot of studies follow Aschauers idea and set the focus on variables such as public infrastructure. Most of these contributions, however, use infrastructure as an input in (private) production. Because there is no price for infrastructure and additionally it can be seen as a public good, we refrain from containing infrastructure as input in private production. We define a ratio of core infrastructure area (road, waterway and train network) to total regional area, which we call *area* and add it to total factor productivity. At least, *area* acts Hicks-neutral to all other input types.

We extend the total factor productivity A_r with these measures and obtain

$$A_r = \exp\left[\beta_0^r + \beta_1 DIV_N + \beta_2 East + \beta_3 Berlin + \beta_4 area + \beta_5 Ind\right]. \tag{8}$$

Nominal gross value added (GVA) is used to approximate productivity Y. In order to avoid unstationarity due to inflation, GVA is deflated by the average inflation rate of Germany[4]. Finally, the production function for the empirical analysis is[5]

$$\begin{aligned} \ln y &= \beta_0^r + \beta_1 DIV_N + \beta_2 East + \beta_3 Berlin + \beta_4 area + \beta_5 Ind \\ &\quad + \beta_6\ s_C + \beta_7 s_M + \beta_8 \ln N_M + u_y \end{aligned} \tag{9}$$

where u_y is the error term and β_0^r is a regional random intercept.

The next section derives the plant equation in order to deal with endogeneity issues of $\ln N_M$.

[4] Thus, we assume that price changes affect all regions by a common factor.

[5] We dropped the regional index r.

2.2 Endogenize the number of firms

Baldwin (1999) suggests that households savings are invested in a riskless asset. Savings are used to fincance a knowledge or research sector which produces new ideas or patents. Each new idea can be seen as a new plant. Hence, saving and rent-seeking affects the regional plant stock. Furthermore, the number of plants can be seen as an approximation of the capital stock. This idea makes the model more tractable for an econometric study since one gets rid of the "traditional" capital stock[6].

Originally, Baldwin assumes a simple production function for innovation

$$Q_N = \frac{1}{a_F} H_F \tag{10}$$

where Q_N is the mass of inventions using H_F workers. a_F is the factor productivity[7]. However, Baldwin does not distinguish various types of workers. The broad literature on firm start-up's, R&D and innovation reveal the crucial role of human capital. We therefore assume that the creation of firms is based on human capital.

The dynamics of the firm stock in Baldwins model is simply given by

$$dN = \frac{1}{a_F} H_F - \delta N \tag{11}$$

where δ is the depreciation rate of the current stock of firms. This function suits well for the theoretical model because it makes the model analytically solvable. However, it is not very flexible for empirical analysis. Therefore we assume

$$dN = \frac{1}{a_F} H_F^{\gamma} - \delta N \tag{12}$$

Since there is no endogenous growth in the long run $dN = 0$ holds[8]. This allows to derive the number of plants as a function of R&D activities and depreciation, given by

$$N = \frac{1}{a_F \delta} H_F^{\gamma}. \tag{13}$$

In a broader sense, the number-of-plants-equation can be seen as a knowledge production function (Audretsch and Feldman 2004, Acs and Varga 2005). So far we derived an equation of the stock of plants which can be used as an instrument equation for the production function, i.e. the endogenous regressor N_M. However, it would be hard to neglect the fact that H_F is exogenous. For that reason (13) is enriched by an additional controll variable, the spatially lagged high-skilled employees $\mathbf{WH}_F$ where W is a standardized binary spatial weights matrix.

[6] Other studies assume that capital flows adjust instantaneously in order to equalize the interest rate. If we additionaly assume that marginal productivity equals the rental price and a rental rate is uniform for all regions, the capital stock can be substituted and disappears.

[7] The higher a_F the less productive is the research sector.

[8] However, since savings affect the capital stock and population might grow, there is an indirect growth process of firms due to population growth.

The productivity parameter of knowledge creation a_F might also be influenced by regional characteristics. Audretsch and Fritsch (1999) provide evidence that firm startups depend on the regional industrial structure and branch specific needs. They controll for human capital, population density, unemployment rate, average manufacturing wage, and taxes. Out theoretical model already includes human capital. Concerning the other variables, we follow Audretsch and Fritsch (1999) and use the unemployment rate *unempl* as an additional covariate. Aus Audretsch and Keilbach (2007) point out, using the unemployment rate controlls for firm start up's out of unemployment as an individual option to get back in work. We do not consider population density because this measure is partially included in the production function. Closely related to the population density is the ratio of settlement area to total regional area (*density*), which we are going to consider as an urbanisation measure in the factor productivity a_F. In addition we enrich a_F by also applying the diversity measure DIV_B to control for urbanisation effects in R&D. It is reasonable to assume that the diversification of employment matters for research activities and not the distribution of plants between monopolistic industries.

Other work on knowledge spillovers and firm creation works out the crucial role of human capital, i.e. the stock of it[9]. They argue that enterpreneurship and knowledge creation depends on a high extend of R&D activities at least in knowledge-based industries. Thus, the usage of H_F should be prefered compared to some normalisation such as the share of human capital input[10] (H_F/W).

Average plant age N_M^{age} might also be a crucial determinant of firm startups. This measure gives information about the tradition of regional economy. Well established regions could have an advantage to introduce new products and ideas on account of experience. Hence, the cost of innovation might be relatively lower and, thus, the productivity of the "research sector" higher in more experienced regions. On the other hand, the older plants of a region are the more likely it might be that they do not see new market specific needs. This situation could lead to underinvestment and less productivity in R&D activities. Also suggested in the literature is the average plant size in terms of employees. However, this approach does not seem to be reasonable in our fashion because average plant size is $(L_M + H_F)/N_M$ and so, it is highly correlated with N_M, the endogenous variable. Therefore we refrain from using it.

The model states that investment takes place where the income stream of a firm relative to cost of invention (Tobins q) is maximised. Assuming that the income stream is higher in a more productive region, we can add $\ln y$ as additional control variable to the regression model.

From 1999 onwards the data collection has changed. Before 1999 there were only firms in the panel with at least a full-time employed person. Now also firms are observed if there is at least one low-payed person. Therefore there is a discrete jump in stock of firms. Hence, a dummy variable $d_{1999} = 1$ if $t \geq 1999$ is included in the regression model.

[9] See e.g. Audretsch et al. (2006), Acs and Armington (2003), or Audretsch and Keilbach (2007).

[10] However, using the share would be in favour with Lucas' (1988) model.

The next task is to implement productivity effects of the age decomposition of human capital. As earlier work suggests, differences in the age composition explains variation in productivity and innovation. Hence, productivity of R&D activities might also be influenced by the age structure, i.e. variations in γ. There is first evidence that the number of firm start ups depend on the age pattern of human capital on the regional level (Bönte et al. 2007). Therefore, we look at the age structure of human capital in R&D activities, in particular, in the creation of plants. We do this by decompose the overall effect of human capital, γ, into the effects of the age cohorts on γ. We consider four age cohorts, i. e. the cohort below the age of 30, the cohort aged 30-44, cohort 45-54 and a cohort aged more than 55 years[11]. Variation in γ is due to variation in cohort shares $s_{Hi} = \frac{No.\ of\ employees\ of\ age\ cohort\ i}{Total\ number\ of\ employees}$,

$$\tilde{\gamma} = \delta_{<30} + \delta_{30/44} s_{30/44} + \delta_{45/54} s_{45/54} + \delta_{55+} s_{55+}$$

Collecting variables we eventually get the empirical equation of the factor productivity a_F

$$a_F = \exp\left[\gamma_0^r + \gamma_4 DIV_B + \gamma_5 N_M^{age} + \gamma_6 density\right]$$

and the logarithm of the plant equation with all its additional covariates

$$\begin{aligned}
\ln N_M &= -\ln\delta - \ln a_F + \gamma \ln H_F + \gamma_1 \ln unempl + \gamma_2 \ln y + \gamma_3 d_{1999} + u_N \\
\ln N_M &= \gamma_1 \ln unempl + \gamma_2 \ln y + \gamma_3 d_{1999} \\
&\quad + \delta_{<30} \ln H_F + \delta_{30/44} s_{30/44} \ln H_F + \delta_{45/54} s_{45/54} \ln H_F + \delta_{55+} s_{55+} \ln H_F \\
&\quad - \gamma_0^r + \gamma_4 DIV_B + \gamma_5 N_M^{age} + \gamma_6 density + u_N
\end{aligned}$$

So far, the we derived a model for the stock of plants of the monopolistic sector. The question of which industry belongs to that sector is considered in the next section.

2.3 Determination of the monopolistic sector

The preceding section introduces the theoretical concept which motivates the regression equations. However, it is still unclear which industry or firm is "monopolistic" and how should N_M be measured. The latter issue is less problematic. Görzig et al. (2007) give evidence for Germany's manufacturing industry that on average each establishment produces one single good. If, however, plants get larger there are less products. The lower bound in that case is 0.7 products per establishment. However, we conclude that each establishment can be seen as a single commodity. So, counting regional establishment stock is a proxy for counting regional product variety. Thus, N_M is measured as the number of regional establishments.

From a New-Economic-Geography viewpoint those industries should be taken as monopolistic, which tend to agglomerate in space. Another possibility is to look at intraindustrial mark-ups. This, however,

[11] Within the data the large group of 30-44 is given, such that there is little space to vary the age structure.

is allmost equivalent to look at agglomaration tendencies because firms will settle down there where intraindustrial mark-ups are higher. This rent seeking argument is in line with the presented and underlying theory of Baldwins model together with Tobins'q. Therefore, our first approach of identification of monopolistic plants is to use the Ellisson's and Glaeser's (1997) Index (EG-Index) in order to identify industries which tend to agglomerate.

Furthermore one might look on interindustrial mark-ups. Then industries with relative high mark-up's compared to other industries would be indicated as monopolistic. This concept needs some further discussion. Some sources of high mark-up's might be the presence of local goods in combination with low regional competition or market entry barriers. However, as long as there is free market entry, mark-up's will disappear over time because of market entry.

Another candidate for relatively high mark-up's might be the presence of fixed cost requirements (possibly together with low demand) for e.g. R&D activities. On firm level capital market imperfections could lead firms to set higher prices such that they finance uncertain R&D on their own. At least, these mark-up's can be interpreted as fixed cost requirement and monopolistic competition would be the market result.

The model presented here assumes fixed cost in terms of operating profits, which is a dividend payed to shareholders. The present value of mark-up's is necessary to cover cost of invention, i.e. wages of the research sector. So, if there exist permanent mark-ups relative to other industries and these are due to some fixed cost requirement for R&D-expenditure, as the model suggests, we can identify establishments as monopolistic if they are more or less human capital intensive. Thus our second approach to identify monopolistic industries is to look at the branch specific human capital input relative to total employment of that particular sector.

After finding the two regression equations and the discussion of the separation[12] of competitive and monopolistic industries we want to form hypotheses on expected values.

2.4 Hypotheses and Estimation Method

The resulting model has two equations with endogenous regressors, i.e. it belongs to the class of simultaneous equation models with Panel data. Both empirical equations are given by

$$\begin{aligned}\ln y &= \beta_0^r + \beta_1 DIV_N + \beta_2 East + \beta_3 Berlin + \beta_4 area + \beta_5 Ind \\ &\quad + \beta_6\ s_C + \beta_7 s_M + \beta_8 \ln N_M + u_y \qquad (14)\\ \ln N_M &= \gamma_o^r + \gamma_1 \ln unempl + \gamma_2 \ln y + \gamma_3 d_{1999} \\ &\quad + \gamma_4 DIV_B + \gamma_5 N_M^{age} + \gamma_6 density. \qquad (15)\\ &\quad + \delta_{<30} \ln H_F + \delta_{30/44} s_{30/44} \ln H_F + \delta_{45/54} s_{45/54} \ln H_F + \delta_{55+} s_{55+} \ln H_F\end{aligned}$$

[12] We use the first year of our panel (1995) for separation to avoid endogeneity of separation and development.

It is worth mentioning that the constant term of the production function and the firm equation can be decomposed in a common and a regional specific effect. Hence, panel analysis should be applied. Furthermore, within the firm equation the common factor includes the log of the depreciation rate which we can not identify.

There are several methods provided to estimate a structural model with Panel data (see Baltagi 2005). We apply the G2SLS estimator of Balestra and Varadharajan-Krishnakumar (1987) for the random effects model and a fixed effects estimator presented in Baltagi (2005). The transformation of the production function in per-capita terms normalises to some extend differences in regional size. N_M, however, remains as a number and captures the externality. Thus, the between variation is important for the identification of the agglomeration forces. That is in favour of the random effects model. On the other hand, N_M might introduce heteroscedasticity considering regions. The fixed effects model demeans N_M within each region. Hence, this type of heteroscedasticity is ruled out. At least, both models have there advantages of identifying and interpreting estimated parameters.

Which sign of estimates we expect, is given in the following. The findings of the literature allow to hypothesize:

- if urbanisation effects are present, we expect β_1, β_5, γ_4, and γ_6 positive and significant.
- because of the still worse position of Eastern Germany we expect $\beta_2 < 0$.
- β_3 is positive (negative) if Berlin performs better (worse) compared to Western Germany, or it might be insignificant.
- Aschauer's debate states that public capital, such as infrastructure makes a region better off. Thus, β_4 should be positive. Literature[13] on infrastructure suggests $0 \leq \beta_4 \leq 0.4$ where most studies conclude a $\beta_4 \leq 0.25$ and for panel data $\beta_4 \leq 0.15$. However, these studies assume infrastructure as an input in production. Hence, these findings are not directly compareable.
- When plants mature they are more established and regional structure is robust, hence $\gamma_5 > 0$. On the other hand, these regions might not identify new market trends and $\gamma_5 < 0$. Thus, depending on the innovation potential and flexibility of matured firms, the parameter might be positive or negative.
- Following Audretsch and Fritsch (1999) we expect a positive effect of the unemployment rate β_6. Further studies conclude that unemployed person are more likeley to start up a firm in order to be

[13] See e.g. Ayogu (1999), Bajo-Rubio et al. (2002), Berechman et al. (2006), Boarnet (1996), Cadot et al. (1999), Calderón und Servén (2002), Stephan (2000), Yamarik (2000) or many others.

better off. However, if we employ the human-capital input share to separate both sectors, we expect a lower value of β_6 or even an insignificant coefficient when we get more restrictive (i.e. when the share of employed human capital increases). The reason for that hypothesis is that unemployed person are less likely to start-up a business in labour intensive branches[14].

- When a region gets more productive ($\ln y$ increases), firms could have an advantage to settle down there and start to produce there. Thus, we would expect a $\gamma_2 > 0$ (agglomeration forces). While Baldwin (1999) fixes wages, one could normalise regional prices and achieve a coherence between regional size (in mass of plants) and wages. Then wages increase when a region gets larger in terms of establishments. Of course, that will increase the cost of innovation of the local research sector and should make a region less attractive for investment. Thus, we would expect $\gamma_2 < 0$. Finally, the effect is unclear so far.

- Considering the age pattern, we expect that the youngest and the oldest are less productive compared to the middle age, that is $\delta_{30/44}$ and $\delta_{45/54}$ are positive and δ_{55+} might be insignificant. $\delta_{<30}$ will be the reference group. However, following Bönte et al. (2007) an increase in the shares of the cohorts of 20-30 and 40-50 relates to innovative start up activities in high-tech industries. Thus the overall effect of the youngest group is unclear.

- Refering to the survey of Cameron (1998) we expect an estimate of N_M about 0.05. Considering β_6 and β_7 we suppose that β_6 will increase when we get more restrictive in the definition of the monopolistic sector because it refers to industry-shares. The parameter of H_F should be positive as long as human capital affects the firm stock. However, because human capital is not only employed in research on firm level, we expect a parameter less than one. The parameter of the spatial lag of H_F is also expected to be positive, since it is also a measure of the human capital potential.

3 Data and Descriptive Statistics

Our research field is Germany. Regional data of gross-added value (GVA), total employment, settlement and infrastructure area are taken from the "Genesis Regional" database provided by the Federal Stadistical Office. Information on firms is taken from the "Betriebshistorikpanel" (BHP), provided by the Institute for Employment Research (Institut für Arbeitsmarkt- und Berufsforschung). The BHP is a 50% sample of all German establishments which employ at least one person. We aggregate NUTS3-regions as suggested by Eckey et al. (1998) to 180 regional labor markets. On average, one labor market region contains 2.4 NUTS3-regions. The regional aggregation is mainly based on commuting flows and

[14] For instance, $s_H = 0.05$ means that we need 5% of employees with a university degree. If only one person is employed with such a degree, we need at least 20 employees to be identified as a monopolistic industry.

should overcome strong spatial interdependence. We consider the time period from 1995 to 2005. The following table 1 describes and correlates the most important variables. Because the definition of some variables depends on the identification problem of the monopolistic sector, this table is prepared for a relatively weak classification, i.e. the share of human capital is 0.05. Being more restrictive strengthens the correlation as expected.

	$\ln y$	$\ln N_M$	$\ln H_F$	DIV_B	DIV_N	$\ln s_C$	$\ln s_M$	$area$	$density$
mean	10.6187	6.9555	7.6599	2.7521	3.2397	-0.5004	-1.0899	-3.0626	-2.1561
s.d.	0.1446	0.9020	1.2139	0.4273	0.1249	0.1252	0.1785	0.3003	0.3717
min	10.1844	5.3660	4.9345	0.5283	2.7690	-1.3760	-2.4607	-3.8264	-2.9871
max	11.0245	9.9971	11.1783	3.6041	3.6555	-0.0998	-0.3663	-2.1800	-0.7482
				Correlation Table					
$\ln N_M$	0.4739	1							
$\ln H_F$	0.2637	0.9039	1						
DIV_B	0.1721	0.4616	0.2953	1					
DIV_N	0.277	0.0511	-0.0452	0.1621	1				
$\ln s_C$	-0.3037	-0.392	-0.5716	0.355	0.0922	1			
$\ln s_M$	0.3952	0.3394	0.4235	-0.2953	-0.0387	-0.9148	1		
$area$	0.5782	0.4558	0.3506	0.2979	0.2575	-0.1713	0.1948	1	
$density$	0.5238	0.6126	0.5381	0.3717	0.1024	-0.2314	0.2147	0.8869	1
Table prepared for sH=0.05, N=1980									

Table 1: Descriptive Statistics and Correlation of important variables

The correlation matrix reveals a strong dependency of human capital and number of plants as suggested by the underlying theoretical framework. Even more, the correlation of the stock of human capital correlates less strong with regional productivity $\ln y$. The positive relationship between human capital and infrastructure area suggests that large (rural) regions attract human capital or that human capital is not land intensive, of course. The same relationship holds for the density measure.

Considering the descriptives of the diversity measure of employment within the monopolistic sector, DIV_B, one can identify a huge range. While the minimum is 0.52, the maximum reaches a level of 3.6. Thus, if there is evidence that diversity matters, an industry mix should be prefered.

Unfortunately the age pattern of human capital is not available using the BHP data. Therefore we approximate it by the age pattern of the employees of firms in the monopolistic sector[15]. Table 2 gives an overview of the age distribution of the various classification of monopolistic firms, the absolut numbers N_M and H_F. Furthermore, the number of valid cases are reported.

The average share of the first cohort (<30 years) is about 20%, that of the second cohort (30-44 years) is over 42%, the third cohort reaches allmost 25% and the oldest cohort of age 55 and older is about 13% in size. The decomposition of the cohorts contains labour market conditions. I.e. the cohort of the oldest compared to the cohort of 45/55 is possibly smaller due to the unavailability of "low" skilled jobs for

[15] E.g. Moretti (2004) approximates the age pattern of employees using the distribution of the population.

	$s_H = 0.05$	$s_H = 0.06$	$s_H = 0.07$	$s_H = 0.08$	$s_H = 0.09$	$s_H = 0.10$	EG_{50}	EG_{75}	EG_{90}
$s_{<30}$	0.200	0.193	0.193	0.189	0.203	0.207	0.207	0.198	0.188
	(0.033)	(0.032)	(0.031)	(0.031)	(0.034)	(0.038)	(0.034)	(0.046)	(0.101)
$s_{30/44}$	0.432	0.430	0.430	0.427	0.430	0.427	0.428	0.445	0.426
	(0.030)	(0.032)	(0.033)	(0.034)	(0.038)	(0.040)	(0.033)	(0.043)	(0.098)
$s_{45/54}$	0.239	0.242	0.242	0.244	0.234	0.232	0.236	0.238	0.244
	(0.033)	(0.033)	(0.033)	(0.033)	(0.033)	(0.032)	(0.032)	(0.042)	(0.079)
s_{55+}	0.129	0.136	0.134	0.139	0.133	0.134	0.129	0.119	0.142
	(0.024)	(0.025)	(0.024)	(0.026)	(0.028)	(0.028)	(0.025)	(0.030)	(0.077)
N_M	1759.0	1447.1	1074.4	939.2	802.7	742.0	1124.0	153.5	40.4
	(2683.8)	(2266.2)	(1541.5)	(1351.5)	(1210.1)	(1117.5)	(1667.4)	(226.1)	(70.0)
H_F	5130.4	4481.7	4173.9	3790.2	3301.7	3027.7	1823.4	420.5	115.1
	(9899.7)	(8480.5)	(7903.5)	(7191.8)	(6522.8)	(6004.5)	(3603.3)	(1063.2)	(466.6)
obs.	1980	1980	1980	1980	1980	1980	1980	1980	1688

s.d. in ()

Table 2: Descriptive Statistics of the age pattern

eldery person. However, human capital is characterised by additional skills and experience. Presumably the "true" share of the oldest group of human capital would exceed 13%. On the other hand, there is the relatively high share of person less than 30 years. In our definition, human capital needs an university degree. Usually such a degree is achieved with an age of about 25. In the share of person under 30 every person is contained older than 17. Thus the "true" share of human capital would be lower than 20%. Comparing the differences in the age cohorts between the different approaches one can not figure out some pattern. So we conclude that independend on the definition of monopolistic and competitive firms, the required labour decomposition in terms of average age seems to be quiet constant.

The picture of the mass of monopolistic firms and the human capital input depends, of course, on the definition. N_M declines rapidly when we get more restrictive in separating the two sectors. As expected, the human capital input does not reduce one by one. Considering the industries identified by the s_H and EG approach, we can not see that firms who tend to agglomerate are human capital intensive. The identification of a monopolistic firm using the Ellison-Glaeser index is as follows: an industry is identified as "monopolistic" if the value of the EG exceeds the i th percentile of all industries. Thus, using the 75th percentile would state that 25 % of all industries are identified as "monopolistic" because they tend more to agglomerate compared to the other 75%. Table 3 reports the total number of identified industries and the common industries in both approaches.

There are for instance 71 branches out of 302 classified as monopolistic considering a human capital input of 7%. From these industries only 14 are identified as monopolistic using the EG - classification. Unfortunately the two concepts of separation of the competitive and monopolistic sectors identify different industries. Thus, those industries which tend to agglomerate are usually not industries which are human capital intensive.

Equally identified industries s_H vs. EG				
		EG_{50}	EG_{75}	EG_{90}
	No. Ind.	**151**	**76**	**32**
$s_H = 0.05$	**114**	56	27	11
$s_H = 0.06$	**89**	41	18	8
$s_H = 0.07$	**71**	32	14	5
$s_H = 0.08$	**61**	28	12	5
$s_H = 0.09$	**51**	23	8	4
$s_H = 0.10$	**44**	20	6	2

Table 3: Comparison of common industries included in both separation approaches

The next section presents the estimation results.

4 Results

The regression model (14) and (15) is estimated using Panel-2SLS. The Hausman test indicates to prefer fixed over random effects and IV estimates in favour of single equation estimation. For that reason we concentrate on fixed effects analysis of the IV estimates. Tables 4 and 5 report the results of the production function (14) and tables 6 and 7 the estimates of the establishment equation for the s_H and EG approach, respectively. Estimates of the random effects models using instrumental variable techniques are attached the appendix.

Instrumental Variable Fixed Effects estimates of the production function $\ln y$						
	$s_H = 0.05$	$s_H = 0.06$	$s_H = 0.07$	$s_H = 0.08$	$s_H = 0.09$	$s_H = 0.10$
$\ln s_C$	0.1064**	0.3568***	0.3676***	0.5669***	0.4897***	0.4288***
	(0.0521)	(0.0692)	(0.0738)	(0.0792)	(0.0904)	(0.0946)
$\ln s_M$	0.1060***	0.1567***	0.1343***	0.1461***	0.0916***	0.0394**
	(0.0278)	(0.028)	(0.0264)	(0.0245)	(0.0188)	(0.0162)
$\ln N_M$	0.0404***	0.0298***	0.0374***	0.0281***	0.0337***	0.0352***
	(0.0062)	(0.0053)	(0.0075)	(0.0078)	(0.008)	(0.0079)
DIV_N	0.2945***	0.2906***	0.2815***	0.2722***	0.2786***	0.2754***
	(0.0213)	(0.0215)	(0.0214)	(0.0215)	(0.0213)	(0.0215)
Ind	-0.0013***	-0.0013***	-0.0012***	-0.0011***	-0.0012***	-0.0013***
	(0.0002)	(0.0002)	(0.0002)	(0.0002)	(0.0002)	(0.0002)
$area$	0.2476***	0.2600***	0.2758***	0.2943***	0.2839***	0.2759***
	(0.0407)	(0.0403)	(0.0407)	(0.0397)	(0.0405)	(0.0404)
$constant$	10.5561***	10.8737***	10.8603***	11.0737***	10.8931***	10.7526***
	(0.1871)	(0.185)	(0.1913)	(0.1873)	(0.1899)	(0.1896)
$obs.$	1980	1980	1980	1980	1980	1980
No. Industries	114	89	71	61	51	44
notes: s.e. in (), *p<0.1 **p<0.05 ***p<0.01						
s_H: critical value for separation of both sectors using the human capital ratio						

Table 4: Instrumental Variable regression results of the production function

Most of the results are significant. Furthermore, a F-test on the fixed effects rejects the hypothesis that there are no individual effects. The correlation of the fixed effects with the explanatory part is about -0.1 for the human capital separation approach. As expected, the industry share of workers in the competitive sector increases and the share of industry of the monopolistic sector on productivity decreases the more restrictive the separation gets. Obviously, for the EG separation a concentration means a reduction of productivity. Thus, industrial concentration does not neccesarily mean higher productivity. On the other hand, human capital intensive branches raise the average productivity. The findings of the parameters for $\ln N_M$ seem to be reasonable for the s_H definition if we compare our findings with that of Cameron (1998). As a robustness check we additionally include the number of firms of the competitive sector. Either it was significant and negative or the estimated parameter was insignificant for the human capital separation and positive and significant for the EG separation. However, the additional regressor did not change the results of the remaining variables. This robustness check supports the findings of the hypothesis that human capital intensive branches increase productivity.

The disadvantage of the fixed effects model is the loss of between variation. The measure of $\ln N_M$ also relates to static Marshall-Arrow-Romer externalities in our case. Even if we show that these externalities are present, we suppress the power of the between variation. A look to the random effects in tables (8) and (9) show that the estimates of $\ln N_M$ are significant and larger compared to the fixed effects model. Thus, MAR externalities matter for interregional comparison.

Considering the urbanisation measures DIV_N and Ind, we find evidence that Jacobs (1969) externalities are present. The more regional firms are dispersed between industries the higher is the average productivity. A reason for this might be regionaly concentrated upstream-industries, which will rise the DIV_N index or linkages between firms. However, if there are too many branches, a reduction of productivity occurs. In particular, a limited labour market and labour competing industries or independend branches worsen regional productivity.

The transportation network, $area$, indicates productivity gains in a more dense area. Even if the parameters are quiet high, they can not be compared directly to findings of the literature, since our variable is measured as a ratio of total regional area and not in per capita or per worker. The settlement density is another measure of urbanisation externalities which gets employed in the literature. Of course, allmost every urbanisation externality is captured within that measure. Adding it to our regression model yields insignificant results and smaller parameters of all measures (DIV_N, $area$, Ind), as expected. The more dense an area gets, the more likely it is to observe a higher number of specialised (monopolistic) firms, the more likely there are various industries and dispersion of firms. Thus, all our measures are captured in the settlement area per capita or per employee. After the interpretation of the productivity function we will now turn to the results of the number of establishment equation.

Most of the estimates are statistically high significant. As in the production function, due to correla-

Instrumental Variable Fixed Effects estimates of the production function $\ln y$			
	EG_{50}	EG_{75}	EG_{90}
$\ln s_C$	0.2186***	0.0873***	-0.4604***
	(0.0279)	(0.0238)	(0.1214)
$\ln s_M$	0.1396***	0.0154***	0.0011
	(0.0129)	(0.0048)	(0.0033)
$\ln N_M$	0.0365***	0.0189	-0.0719***
	(0.0044)	(0.0168)	(0.0144)
DIV_N	0.2173***	0.2638***	0.2112***
	(0.0214)	(0.0222)	(0.024)
Ind	-0.0013***	-0.0008***	0.0002
	(0.0002)	(0.0003)	(0.0002)
$area$	0.2839***	0.3878***	0.3290***
	(0.0387)	(0.0371)	(0.0362)
$constant$	11.0929***	11.0750***	11.1295***
(0.166)	(0.1739)	(0.1532)	
$obs.$	1980	1980	1688
No. Industries	151	76	32

notes: s.e. in (), *p<0.1 **p<0.05 ***p<0.01

EG_{XX}: percentile of the EG-Index for separation

Table 5: Instrumental Variable regression results of the production function

tion of about -0.1 of the individual effects with the explanatory variables the Hausman test prefers the fixed effects model. An F-test indicates that the fixed effects are statistically significant on a high level. Lets consider first the additional controllvariables. The dummy to controll for another counting method of firms, d_{1999}, is positive and significant. Interestingly, it decreases when the separation of both sectors gets more restrictive. This is indeed true because from 1999 onwards firms are also collected when there is only employment of low payed jobs. Usually these jobs or industries are not necessarily human capital intensive and therefore the effect of d_{1999} declines. The unemployment rate seems to be robust and is not much affected by the separation approach. There is a small decline, when s_H increases but this effect does not seem to be very strong. However, regions with low unemployment rates have a higher stock of establishments, which seems to be reasonable. Interestingly is a comparisson between the s_H and EG approach. The unemployment rate affects much more the firm stock considering the EG approach. Because these industries are not neccessarily human capital intensive, this gives some evidence that human capital intensive branches are less affected by differences in unemployment rates. A reason might be the usage of special technology and knowledge. Considering the endogenous variable $\ln y$ we can conclude that an increase in regional productivity reduces the stock of firms. This is in line with the theoretical framework: an increase in productivity will raise wages and makes finally the region less attractive for new firms and hence, one will not observe additional firm startup activities. The only exception is given by the EG_{50} definition. However, there is the half of all industries indicated as monopolistic. Those firms will be included which do not need human capital or fixed cost requirements for start ups. In that case, if wages will raise there might be some firm start up activities. If the stock of regional firms is on average older then there will be less firms. Literature on firm mature reveals that only few firms survive during maturing. Only the most efficent establishments stay in the market. Thus, in our case, the more established a region is in terms of establishment age the more efficient these firms will be and therefore the stock of firms which survived is smaller.

Instrumental Variable Fixed Effects estimates of the establishment equation $\ln N_M$

	$s_H = 0.05$	$s_H = 0.06$	$s_H = 0.07$	$s_H = 0.08$	$s_H = 0.09$	$s_H = 0.10$
$\ln y$	-0.7878***	-1.0186***	0.0607	0.0235	-0.3756**	-0.4779**
	(0.1826)	(0.2264)	(0.1663)	(0.1802)	(0.1756)	(0.1991)
$\ln unempl.$	-0.0734***	-0.0938***	-0.0537***	-0.0759***	-0.0599***	-0.0652***
	(0.0115)	(0.0133)	(0.0097)	(0.0105)	(0.0106)	(0.0115)
$\ln H_F$	0.0037	0.0121	0.0551***	0.0640***	0.0491***	0.0585***
	(0.0187)	(0.0207)	(0.0147)	(0.0151)	(0.0135)	(0.0138)
$s_{30/44} * \ln H_F$	0.0317*	0.0387**	0.0210*	0.0497***	0.0403***	0.0337***
	(0.0162)	(0.0178)	(0.0125)	(0.0133)	(0.0118)	(0.0121)
$s_{45/54} * \ln H_F$	-0.0710***	-0.0737***	-0.0223*	-0.0281**	0.0008	-0.0003
	(0.0169)	(0.0188)	(0.0132)	(0.0134)	(0.0117)	(0.0121)
$s_{55+} * \ln H_F$	0.1851***	0.2440***	0.0503**	0.0522**	0.0986***	0.0829***
	(0.0254)	(0.0289)	(0.0206)	(0.0215)	(0.0182)	(0.0182)
$\ln \mathbf{W} H_F$	0.1079***	0.1274***	0.0601***	0.1011***	0.0742***	0.1038***
	(0.0209)	(0.0233)	(0.0168)	(0.0177)	(0.0169)	(0.0183)
DIV_B	0.0454***	0.0671***	0.015	0.0326**	0.0452***	0.0561***
	(0.0139)	(0.0169)	(0.0127)	(0.013)	(0.0102)	(0.0113)
$density$	0.3310***	0.2809***	0.0232	-0.11	0.1559**	0.1273
	(0.0726)	(0.086)	(0.0634)	(0.0691)	(0.0754)	(0.0844)
N_{age}	-0.0413***	-0.0526***	-0.0513***	-0.0456***	-0.0385***	-0.0361***
	(0.0026)	(0.0027)	(0.0019)	(0.0018)	(0.0017)	(0.0017)
d_{1999}	0.1930***	0.1836***	0.1007***	0.0988***	0.1080***	0.1162***
	(0.0059)	(0.007)	(0.0049)	(0.0055)	(0.0055)	(0.0059)
$constant$	15.5268***	17.5399***	5.7716***	5.2026**	9.9184***	10.5645***
	(2.0773)	(2.5774)	(1.897)	(2.0709)	(2.0618)	(2.34)
obs.	1980	1980	1980	1980	1980	1980
No. Industries	114	89	71	61	51	44

notes: s.e. in (), p*>0.1 p**>0.05 p***>0.01
s_H: critical value for separation of both sectors using the human capital ratio

Table 6: Instrumental Variable regression results of the establishment equation

Considering the urbanisation measures we also find evidence that a broader distribution of employment between industries makes research activities better off (DIV_B). Thus, knowledge transfers seem to exist. The *density* measure is positive and significant what also suggests that dense regions offer better business opportunities. Hence, as it was the case in the production function, Jacobs externalities are present. Considering, however, the EG classification, *density* is significant and negative. Of course, the EG index filters industries which are allready concentrated or tend to concentration. In that case, the regional structure plays a secondary role.

Instrumental Variable Fixed Effects estimates of the establishment equation $\ln N_M$			
	EG_{50}	EG_{75}	EG_{90}
$\ln y$	1.3885***	0.0447	-0.788
	(0.2493)	(0.2476)	(0.5101)
$\ln$ *unempl.*	-0.1189***	-0.1524***	-0.1285***
	(0.0201)	(0.0203)	(0.0385)
$\ln H_F$	0.0277	0.0897***	0.1642***
	(0.0284)	(0.0222)	(0.0351)
$s_{30/44} * \ln H_F$	0.1344***	-0.0527*	-0.1406***
	(0.028)	(0.0286)	(0.0476)
$s_{45/54} * \ln H_F$	-0.1399***	-0.1344***	-0.1211***
	(0.0278)	(0.0253)	(0.0426)
$s_{55+} * \ln H_F$	0.1369***	0.0522	-0.2271***
	(0.0349)	(0.0378)	(0.0651)
$\ln \mathbf{W} H_F$	0.2306***	0.0039	0.0061
	(0.0294)	(0.014)	(0.0169)
DIV_B	0.008	0.0387***	0.0990***
	(0.0147)	(0.0124)	(0.0228)
density	-0.5775***	-0.7135***	-0.5563**
	(0.1336)	(0.1339)	(0.2172)
N_{age}	-0.0236***	-0.0309***	-0.0222***
	(0.0042)	(0.0026)	(0.002)
d_{1999}	0.3889***	0.0564***	0.0204
	(0.0115)	(0.0093)	(0.0142)
constant	-11.2490***	3.0101	10.7376*
	(2.9152)	(2.7993)	(5.7577)
	1980	1980	1688
	151	76	32
notes: s.e. in (), *p<0.1 **p<0.05 ***p<0.01			
EG_{XX}: percentile of the EG-Index for separation			

Table 7: Instrumental Variable regression results of the establishment equation

The human capital input H_F is significant. The value refers to the productivity of the youngest cohort of age less than 30 years. The parameter of the second cohort is positive, which indicates that they are more productive compared to the youth. Obviously, the third cohort has a negative coefficient which is significant in most cases. This result is not in line with findings of the literature. Usually the peak in productivity lies between 40 and 55 years of age. The results might be driven by the definition of the age cohorts. The shares are constructed on the age distribution of all workers and not just of that of human capital. Only if obsolescence and depreciation of human capital starts quiet early for non-human capital workers, then this negative parameter might be interpreted as such. On the other hand the estimate of the oldest cohort is quiet large and positive, which is also not that much in line with the literature. However, considering labour market conditions, person with high skills are more likely to be in the labour force and thus, the relative share of workers of age 55 and older with human capital in the share s_{55+} would be higher than in other age cohorts. Therefore the parameter is positive and significant.

As it was the case in the production function, the between variation is neglected in the fixed effects model. Referring to the random effects, the estimated parameters of $\ln H_F$ and the age cohorts is different. First, the youth is more productive, when regional comparison is done. The second cohort is still more productive compared to the youth. As it was the case in the fixed effects model, the cohort of 45/54 is still less productive. Obviously, the random effects model reveals that the oldest group is less productive compared to the youngest which is the opposite result of the fixed effects model. The interpretation could be that it is good to have a more equaly distributed age structure (what the fixed effects model states) but too much of the oldest workers compared to other regions worsens that region.

Beside the problem of the class of 45/54, we find evidence that the age pattern affects the productivity parameter of human capital. Not only the human capital of the own region but also the distribution of human capital is a significant determinant for the number of establishments, since the parameter for $\mathbf{W}H_F$ is positive and significant.

Considering the *East* and *Berlin* dummy variables reveals, what we find in earlier work (see Brunow/Hirte 2008). While Western Germany performs better than the eastern parts, Berlin is more likely as the West.

After the analysis of the regression results we want to conclude.

5 Conclusion

This paper analyses the effect of human capital and the age pattern on productivity in Germany (1995-2005). The ideas of the theoretical model of Baldwin (1999) was taken in order to derive an empirical model. This model exhibits increasing returns to scale in production due to the number of plants within a region. A second equation derives the number of plants in order to deal with this endogenous regressor. The theory suggests that firms need some research to exist. The theory of firm start ups and knowledge creation suggests, that research activities and R&D are human capital intensive. Thus, we implement human capital to our model. I.e. we assume that human capital explains the number of regional establishments. In our analysis human capital is defined as the stock of person with an university degree.

The theory separates two types of industries - a competitive and a monopolistic sector. Only the latter branch exhibits increasing returns. In order to separate both branches we use the Ellison Glaeser index and the share of human capital to total industrial employment. While we can not find increasing returns using the Ellison Glaeser index, there is evidence that human capital intensive branches offer externalities. In addition we vary the productivity parameter of human capital by the age structure. We find evidence that the youngest cohort is less productive compared to the second cohort of age 30/44. Unfortunatelly the cohort of 45/54 yields negative or insignificant results, indicating that they are less productive compared to the youth. However, this result might be driven by the definition of the age structure. The oldest cohort of 55+ is the most productive one, which can be motivated by a labour market argument.

In addition we controll for urbanisation effects. The more employment spreads across industries, the more productive a region is as long as there are not too much branches. Infrastructure and dense regions increase productivity. Eastern Germany reaches almost 80% of the productivity of Western Germany and Berlin seems to achieve the same productivity as the West. Most findings are consistent with findings of the literature.

Currently, we do not consider migration on account of the low level of migration in our data base. To overcome potential endogeneity problems we add the stock of human capital of the surrounding regions as an explanatory variable to the regression. However, we want to deal with the migration issue in future work.

6 Appendix

Instrumental Variable Random Effects estimates of the production function $\ln y$						
	$s_H = 0.05$	$s_H = 0.06$	$s_H = 0.07$	$s_H = 0.08$	$s_H = 0.09$	$s_H = 0.10$
$\ln s_C$	-0.0617	0.1774***	0.2463***	0.4315***	0.3270***	0.2706***
	(.0451)	(.0639)	(.0686)	(.0746)	(.0822)	(.0864)
$\ln s_M$	0.0548**	0.1069***	0.1080***	0.1198***	0.0713***	0.0271*
	(.0263)	(.0267)	(.0253)	(.0237)	(.0176)	(.0152)
$\ln N_M$	0.0645***	0.0471***	0.0672***	0.0609***	0.0619***	0.0661***
	(.0055)	(.0048)	(.0069)	(.0074)	(.0075)	(.0075)
DIV_N	0.2602***	0.2630***	0.2574***	0.2558***	0.2561***	0.2532***
	(.0190)	(.0195)	(.0195)	(.0198)	(.0195)	(.0197)
Ind	-0.0012***	-0.0008***	-0.0011***	-0.0010***	-0.0011***	-0.0012***
	(.0002)	(.0002)	(.0002)	(.0002)	(.0002)	(.0002)
$area$	0.0605***	0.0840***	0.0746***	0.0902***	0.0802***	0.0751***
	(.0179)	(.0188)	(.0197)	(.0198)	(.0198)	(.0198)
$East$	-0.2077***	-0.2041***	-0.2067***	-0.1921***	-0.2045***	-0.2033***
	(.0123)	(.0133)	(.0137)	(.0139)	(.0135)	(.0135)
$Berlin$	-0.1047*	-0.0667	-0.0907	-0.0715	-0.0811	-0.0849
	(.0123)	(.0652)	(.0674)	(.0684)	(.0673)	(.0673)
$constant$	9.8185***	10.1303***	10.0834***	10.2364***	10.1237***	10.0042***
	(.1144)	(.1226)	(.1257)	(.1264)	(.1229)	(.1230)
$obs.$	1980	1980	1980	1980	1980	1980
No. Industries	114	89	71	61	51	44
notes: s.e. in (), *p<0.1 **p<0.05 ***p<0.01						
s_H: critical value for separation of both sectors using the human capital ratio						

Table 8: Instrumental Variable regression results of the establishment equation

Instrumental Variable Random Effects estimates of the production function $\ln y$			
	EG_{50}	EG_{75}	EG_{90}
$\ln s_C$	0.1857***	0.0450*	-0.2715**
	(.0283)	(.0267)	(.1179)
$\ln s_M$	0.1125***	-0.0101**	-0.0059
	(.0125)	(.0051)	(.0036)
$\ln N_M$	0.0458***	0.1425***	-0.0068
	(.0041)	(.0204)	(.0150)
DIV_N	0.2030***	0.2758***	0.1905***
	(.0199)	(.0227)	(.0216)
Ind	-0.0009***	-0.0026***	0.0003
	(.0002)	(.0004)	(.0003)
$area$	0.1054***	0.0606**	0.1257***
	(.0193)	(.0243)	(.0203)
$East$	-0.1877***	-0.1825***	-0.2164***
	(.0138)	(.0157)	(.0146)
$Berlin$	-0.0493	-0.1811**	-0.016)
	(.0699)	(.0821)	(.0729)
$constant$	10.4251***	9.7897***	10.3622***
	(.1094)	(.1364)	(.1082)
obs.	1980	1980	1688
No. Industries	151	76	32
notes: s.e. in (), *p<0.1 **p<0.05 ***p<0.01			
EG_{XX}: percentile of the EG-Index for separation			

Table 9: Instrumental Variable regression results of the establishment equation

Instrumental Variable Random Effects estimates of the establishment equation $\ln N_M$						
	$s_H = 0.05$	$s_H = 0.06$	$s_H = 0.07$	$s_H = 0.08$	$s_H = 0.09$	$s_H = 0.10$
$\ln y$	0.0138	0.7924***	0.8489***	1.1387***	0.2628	0.2298
	(.1863)	(.2319)	(.2134)	(.2397)	(.1948)	(.2089)
$\ln unempl.$	-0.0944***	-0.1228***	-0.0732***	-0.0973***	-0.0957***	-0.1057***
	(.0130)	(.0164)	(.0139)	(.0157)	(.0132)	(.0138)
$\ln H_F$	0.2497***	0.3223***	0.2545***	0.2614***	0.2170***	0.2095***
	(.0192)	(.0224)	(.0194)	(.0209)	(.0159)	(.0157)
$s_{30/44} * \ln H_F$	0.0356*	0.0630***	0.0223	0.0625***	0.0363**	0.0383**
	(.0186)	(.0221)	(.0181)	(.0201)	(.0150)	(.0149)
$s_{45/54} * \ln H_F$	-0.1300***	-0.1684***	-0.0835***	-0.0946***	-0.0479***	-0.0471***
	(.0192)	(.0231)	(.0190)	(.0200)	(.0149)	(.0149)
$s_{55+} * \ln H_F$	0.0677**	0.0487	-0.0571**	-0.0782**	0.0349	0.03
	(.0281)	(.0338)	(.0285)	(.0307)	(.0220)	(.0212)
$\ln \mathbf{W} H_F$	0.0324*	0.0420**	0.0091	0.0503**	0.0450***	0.0755***
	(.0180)	(.0199)	(.0190)	(.0208)	(.0169)	(.0175)
DIV_B	0.0734***	0.0622***	0.0269	0.0324*	0.0592***	0.0670***
	(.0149)	(.0195)	(.0175)	(.0188)	(.0126)	(.0132)
density	0.6475***	0.4956***	0.3819***	0.2308***	0.5441***	0.5199***
	(.0619)	(.0710)	(.0697)	(.0785)	(.0709)	(.0755)
N_{age}	-0.0351***	-0.0502***	-0.0469***	-0.0441***	-0.0321***	-0.0306***
	(.0030)	(.0034)	(.0028)	(.0029)	(.0023)	(.0022)
d_{1999}	0.1766***	0.1620***	0.0766***	0.0708***	0.0768***	0.0817***
	(.0068)	(.0087)	(.0072)	(.0082)	(.0070)	(.0073)
constant	6.5517***	-2.4982	-2.7123	-6.6819**	3.1363	3.1758
	(2.0690)	(2.5682)	(2.3907)	(2.7057)	(2.2404)	(2.4054)
obs.	1980	1980	1980	1980	1980	1980
No. Industries	114	89	71	61	51	44
notes: s.e. in (), *p<0.1 **p<0.05 ***p<0.01						
s_H: critical value for separation of both sectors using the human capital ratio						

Table 10: Instrumental Variable regression results of the establishment equation

Instrumental Variable Random Effects estimates of the establishment equation lnN_M			
	EG_{50}	EG_{75}	EG_{90}
$\ln y$	1.4673***	-0.1047	0.4067
	(.2627)	(.2458)	(.4095)
$\ln unempl.$	-0.1336***	-0.2065***	-0.2232***
	(.0206)	(.0212)	(.0361)
$\ln H_F$	0.1951***	0.1549***	0.1597***
	(.0276)	(.0234)	(.0347)
$s_{30/44} * \ln H_F$	0.1136***	-0.0494	-0.0759
	(.0291)	(.0310)	(.0472)
$s_{45/54} * \ln H_F$	-0.2269***	-0.1978***	-0.1214***
	(.0293)	(.0275)	(.0430)
$s_{55+} * \ln H_F$	0.1025***	0.0477	-0.1824***
	(.0373)	(.0409)	(.0651)
$\ln \mathbf{W} H_F$	0.1374***	0.0174	0.0363**
	(.0251)	(.0145)	(.0158)
DIV_B	0.0285*	0.0908***	0.1511***
	(.0156)	(.0134)	(.0233)
density	0.1208	0.4035***	0.1701
	(.1011)	(.1021)	(.1608)
N_{age}	-0.0207***	-0.0214***	-0.0192***
	(.0044)	(.0027)	(.0021)
d_{1999}	0.3571***	0.0081	-0.0435***
	(.0125)	(.0103)	(.0137)
constant	-10.8116***	6.6061**	-0.5688
	(2.9871)	(2.7243)	(4.6020)
obs.	1980	1980	1688
No. Industries	151	76	32
notes: s.e. in (), *p<0.1 **p<0.05 ***p<0.01			
EG_{XX}: percentile of the EG-Index for separation			

Table 11: Instrumental Variable regression results of the establishment equation

7 Literature

1. Acs, Z., C. Armington (2003): The geographic concentration of new firm formation and human capital: Evidence from the cities, Working Paper No. CES 03-05, *Center for Economic Studies,* Washington, D.C.

2. Acs, Z., A. Varga (2005): Entrepreneurship, Agglomeration and Technological Change, *Small Business Economics*, 24, 323-334.

3. Audretsch, D.B., M.P. Feldman (2004): Knowledge Spillovers and the Geography of Innovation, *Handbook of Regional and Urban Economics*, 4, 2713-2739

4. Audretsch, D.B., M. Keilbach, E. Lehmann (2006): *Enterpreneurship and economic growth,* Oxford University Press, New York

5. Audretsch, D.B., M. Keilbach (2007): The localisation of enterpreneurship capital: Evidence from Germany, *Papers in Regional Science*, 86 (3), 351-365

6. Audretsch, D.B., M. Fritsch (1999): The Industry Component of Regional New Firm Formation Processes, *Review of Industrial Organization*, 15, 239-252

7. Ayogu, M. D. (1999): Before Prebendalism: A Positive Analysis of Core Infrastructure Investment in a Developing Fiscal Federalism, *African Development Review* 11 (2), 169-198

8. Bajo-Rubio, Ó., C. Díaz-Roldán, M. D. Montávez-Garcés (2002): Optimal Endowments of Public Investment: An Empirical Analysis for the Spanish Regions. *ERSA Conference Papers 2002/14*

9. Baldwin, R.E. (1999): Agglomeration and endogenous capital, *European Economic Review* 43, 253-280

10. Baldwin, R.E., P. Martin, G. Ottaviano (2001): Global income divergence, trade and industrialization: the geography of growth take-off, *Journal of Economic Growth* 6, 5-37

11. Balestra, P., J. Varadharajan-Krishnakumar (1987): Full-information estimations of a system of simultaneous equations with error component structure, *Econometric Theory*, 3, 223-246.

12. Baltagi, B.H. (2005): *Econometric analysis of panel data*, Wiley College, Chichester

13. Berechman, J., D. Ozmen, K. Ozbay (2006): Empirical analysis of transportation investment and economic development at state, county and municipality levels, *Transportation* 33 (6), 537-551

14. Blanchard, O.J. (1985): Debt, Deficits, and Finite Horizons, *Journal of Political Economy*, 93 (2), 223-247

15. Boarnet, M. G. (1996): The Direct and Indirect Economic Effects of Transportation Infrastructure, *The University of California Transportation Center*, Working Paper No. 340

16. Bönte, W., O. Falck, S. Heblich (2007): Demography and Innovative Enterpreneurship, *Jena Economic Research Papers*, 2007-084

17. Brunow, S., G. Hirte (2006): Age structure and regional economic growth, *Jahrbuch für Regionalwissenschaft (Review of Regional Research)* 26, 3-23

18. Cadot, O., L. Röller und A. Stephan (1999): A Political Economy Model of Infrastructure Allocation: An Empirical Assessment, *Wissenschaftszentrum Berlin,* FS IV 99-15

19. Calderón, C., L. Servén (2002): The Output Cost of Latin America's Infrastructure Gap, *Central Bank of Chile*, Working Paper No. 186

20. Cameron, G. (1998): Innovation and Growth: a survey of empirical evidence, *mimeo*

21. Ciccone, A. (2002): Agglomeration effects in Europe, *European Economic Review*, 46, 213-227

22. Combes, P.-P., T. Magnac, J.-M. Robin (2004): The dynamics of local employment in France, *Journal of Urban Economics* 56, 217-243

23. Combes, P.P., H.G. Overman (2004): The Spatial Distribution of Economic Activities in the European Union, *Handbook of Regional and Urban Economics*, 4, 2845-2909

24. Eckey, H.F., K. Horn, W. Stock (1998): Überprüfung des Zuschnitts der Arbeitsmarktregionen für die Neuabgrenzung des Fördergebiets 1999 im Rahmen der Gemeinschaftsaufgabe „Verbesserung der regionalen Wirtschaftsstruktur", Kassel 02/1998

25. Eckey, H.F. , R. Kosfeld, M. Türck (2007): Regionale Entwicklung mit und ohne räumliche Spillover-Effekte, *Jahrbuch für Regionalwissenschaft (Review of Regional Research)*, 27 (1), 23-42

26. Ellison, G., E.L. Glaeser (1997): Geographic Concentration in U.S. Manufacturing Industries: A Dartboard Approach", *Journal of Political Economy,* 105 (5), 889-927

27. Glaeser, E.L., H. D. Kallal, J.A. Scheinkman, A. Shleifer (1992): Growth in Cities, *Journal of Political Economy* 100 (6), 1126-1152

28. Görzig, B., M. Gornig, R. Pohl (2007): Spezialisierung und Unternehmenserfolg im verarbeitenden Gewerbe Deutschlands, *Vierteljahrshefte zur Wirtschaftsforschung,* 76 (3), 43-58

29. Greene, W.H. (2003): *Econometric Analysis*, Prentice Hall, New York, 5.ed.

30. Griliches, Z., F. Lichtenberg (1984): R&D and Productivity Growth at the Industry Level: Is There Still a Relationship?, in Griliches, Z. ed. *R&D, Patents and Productivity*, Chicago: University of Chicago Press

31. Grossman, G.M., E. Helpman (1995): *Innovation and Growth in the global economy*, 5. ed., MIT Press, Cambridge (Mass.)

32. Haltiwanger, J. C., J.I. Lane, J.R. Spletzer(1999): Productivity differences across employers: the roles of employer size, age, and human capital, American Economic Review, Papers and Proceedings 8**9**, 94-98

33. Henderson V., A. Kuncoro, M. Turner (1995): Industrial Development in Cities, *Journal of Political Economy*, 103 (5) 1067-1090

34. Henderson V. (1997): Externalities and Industrial Development, *Journal of Urban Economics,* 42, 449-470

35. Hellerstein, J.K. and D. Neumark (1995): Are earnings profiles steeper than productivity profiles? Evidence form Israeli firm-level data, Journal of Human Ressources 30, 89-112

36. Hellerstein J K, Neumark D, Troske K R (1999): Wages, productivity, and workers characteristics: evidence from plant-level production functions and wage equations, Journal of Labor Economics 17, 409-446

37. Hellerstein, J.K. and D. Neumark (2004), Production function and wage equation estimation with heterogeneous labor: evidence from a new matched employer-employee data set, NBER Working Paper 10325.

38. Henderson V. (2003): Marshall's scale economies, *Journal of Urban Economics,* 53, 1-28

39. Henderson V., A. Kuncoro, M. Turner (1995): Industrial Development in Cities, *Journal of Political Economy, 103 (5), 1067-1090*

40. Hirte, G., S. Brunow (2008): The Age Pattern of Human Capital and Regional Productivity, *Dresden Discussion Paper Series in Economics*, No. 01/08

41. Ilmakunnas P. and M. Maliranta (2005): Technology, Labour Characteristics and Wage-Productivity Gaps, Oxford Bulletion of Economics and Statistics 67, 623-645

42. Jacobs, J. (1969): *The Economy of Cities*, Random House

43. Lindh, T., B. Malmberg (1999): Age structure effects and growth in the OECD, 1950-1990, *Journal of Population Economics*, 12, 399-431

44. Lucas, R.E. (1988): On the mechanics of economic development, *Journal of Monetary Economics* 22, 3-42

45. Lucas, R.E. (2001): Externalities and Cities, *Review of Economic Dynamics*, 4, 245-274

46. Martin, P., G. Ottaviano (1999): Growing locations: industry location in a model of endogenous growth, *European Economic Review* 43, 281-302

47. Mincer, J. (1997): The Production of Human Capital and the Live Cycle of Earnings: Variations on a Theme, *Journal of Labor Economics* 15 (1) Part 2, 26-47

48. Moretti, E. (2004): Estimating the social return to higher education: evidence from longitudinal and repeated cross-sectional data, *Journal of Econometrics*, 121, 175-212.

49. Romer, P.M. (1990): Endogenous Technological Change, *Journal of Political Economy*, 98 (5) Part 2

50. Rosenthal, S.S., W.C. Strange (2004): Evidence on the nature and sources of agglomeration economies, *Handbook of Regional and Urban Economics*, 4, 2119-2171

51. Skirbekk V. (2004) Age and Individual Productivity: A Literature Survey, in: Feichtinger G. (ed.) Vienna Yearbook of Population Research, Vienna: Verlag der Österreichischen Akademie der Wissenschaften.

52. Stephan, A. (2000): Regional Infrastructure Policy and Its Impact on Productivity: A Comparison of Germany and France, *Wissenschaftszentrum Berlin,* Discussion Paper FS IV 01-02

53. Yamarik, S. (2000): The effect of public infrastructure on private production during 1977-96. Department of Economics, *The University of Akron*, mimeo