

Brunow, Stephan

**Conference Paper**

## Regional industrial firm growth under monopolistic competition and human capital externalities

50th Congress of the European Regional Science Association: "Sustainable Regional Growth and Development in the Creative Knowledge Economy", 19-23 August 2010, Jönköping, Sweden

**Provided in Cooperation with:**

European Regional Science Association (ERSA)

*Suggested Citation:* Brunow, Stephan (2010) : Regional industrial firm growth under monopolistic competition and human capital externalities, 50th Congress of the European Regional Science Association: "Sustainable Regional Growth and Development in the Creative Knowledge Economy", 19-23 August 2010, Jönköping, Sweden, European Regional Science Association (ERSA), Louvain-la-Neuve

This Version is available at:

<http://hdl.handle.net/10419/119022>

**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.*

# Regional industrial firm growth under monopolistic competition and human capital externalities

*Manuscript prepared for the 50th European Congress of the  
Regional Science Association International*

19th - 23rd August 2010, Jonkoping, Sweden

Stephan Brunow

Technische Universitaet Dresden

Stephan.Brunow@tu-dresden.de

## **Abstract**

Typically firm formation is explained by labour productivity measures such as wages or GDP per capita. Those measures, however, are misleading in the presence of agglomeration forces because those variables might be biased because of agglomeration rents. This is the departure of our research. This paper derives from a New Economic Geography perspective an empirical estimatable approach to consider regional industry-specific firm growth. In this model no wage measures appear as explanatory variable. The crucial determinant is the real market access based on households maximisation problem. The model is flexible enough to consider various industries without making any assumption on whether they relate to monopolistic competition or perfect markets. The model suits to adress various research questions. We start with the first question whether the suggested crucial variables explain firm formation and are significant. These variables also relate to agglomeration and dispersion forces and therefore this approach is a test of the underlying NEG model.

# 1 Introduction

Firm growth and firm formation is often seen as a crucial factor of economic growth and development. From a policy perspective firm growth should maintain regional labour demand, rise local income and welfare or reduce unemployment rates. Then one policy aim is to foster steady (regional) firm formation. However, in the presence of agglomeration forces and i.e. positive externalities industrial concentration might occur. This in turn make few regions better off while other regions loose. Regional disparities due to urbanisation (Jacobs 1969) and location (Marshall-Arrow-Romer externalities) externalities result which is not in line with the policy perspective. On the other side strong dispersion effects such as competition or in the presence of (high) transportation cost weaken agglomeration forces such that firms and industries are equally distributed over regions. Both mechanism are well known in the literature and especially addressed in the New Economic Geography (NEG). Therefore, testing the empirical relevance of externalities and the NEG is essential to make useful policy. There is a large body of literature which aims to identify of those effects. Main contributions relating to the identification of externalities are due to the work of Glaeser et al. (1994) and various work of Henderson (1995, 2003). Intraindustrial spillover effects are a fundamental part of the NEG. Typically a so-called nominal wage equation is estimated<sup>1</sup> to test the relevance of the NEG. Rosenthal and Strange (2004) summarize and discuss possible ways to measure agglomeration forces, i.e. the dependent variables. One way is to consider firm formation which we are going to adress in the following.

The branch of firm growth literature growth typically use wage levels and GDP per capita as crucial explanatory variables because they relate to labour productivity and should therefore be determinants for start-up activities. The mentioned nominal wage equation of the NEG world contains an agglomeration rent on wages. Furthermore, NEG models typically assume a homogenous distributed labour productivity in space. Then differences in wages are not due to productivity but agglomeration rents. Therefore, using labour productivity measures such as wages might be misleading. One can not be sure whether one measures labour productivity or agglomeration rents.

This is the departure for our research. We focus on industry-specific regional firm growth while we avoid to use labour productivity measures as fundamental explanatory variable. We derive a model which is flexible enough to consider competitive sectors as well as localisation industries in a NEG sense. The conceptual ideas are taken from Baldwin (1999), Baldwin et al. (2001) and Martin and Ottaviano (1999). The resulting model states that not GDP per capita but a complex GDP per firm variable explains firm growth. The approach features agglomeration and dispersion forces and is quiet flexible to address various

---

<sup>1</sup>See Hanson (2000), Brakman et al. (2002), Mion (2003) and Niebuhr (2006).

questions. We want to start with the first question whether the suggested explanatory variable(s) explain firm formation. Indirectly this is a test of the underlying NEG models.

The paper is organised as follows. The next section outlines the theoretical background and derives the basic theoretical equation of regional industry-specific firm growth. Section 3 contains the empirical specification, the introduction to the data base and motivates additional control variables. After that the estimation results of a random coefficient model are presented and discussed. The paper closes with a conclusion.

## 2 Theoretical Consideration

Determinants of firm entry and firm formation are frequently addressed in the literature. Besides others, regional unemployment, human capital, branch-specific needs, labour productivity and urbanisation and location externalities explain firm establishment. Our model explicitly considers location externalities and is based on theoretical work of Baldwin (1999), Baldwin et al. (2001) and Martin and Ottaviano (1999). Their work derive models of neoclassical or endogenous growth based on the ideas of the NEG literature. This chapter derives the main specification for the following empirical model.

The economy consists of households which consume a variety of composite goods  $C_i$  from different branches or industries  $i$ . The temporal utility function is of type Cobb Douglas with nested CES-subutility functions for each industry. The parameters  $\alpha_i$  and  $\sigma_i$  label industry specific elasticities. The utility function of a representative household in region  $s$  is given by

$$U_s = \prod_{i=1}^I C_{is}^{\alpha_i}; \quad C_{is} = \left( \sum_{n=1}^{N_i^w} (x_{ni}^{rs})^{\frac{\sigma_i-1}{\sigma_i}} \right)^{\sigma_i/(\sigma_i-1)}; \quad \sum_{i=1}^I \alpha_i = 1, \quad 0 \leq \alpha_i \leq 1 \quad \sigma_i > 1.$$

$x_{ni}^{rs}$  is the  $n$ -th commodity of a particular industry  $i$  with  $N_i^w$  producers world wide. It might be produced within the home region  $s$  or imported from any other region  $r$ . Households maximise utility facing a budget constraint with an expenditure level  $E_s$ . Then Marshallian demand of  $x_{ni}^{rs}$  can be easily derived<sup>2</sup> and it is given by

$$x_{ni}^{rs} = \alpha_i \frac{(p_{ni}^{rs})^{-\sigma_i}}{P_{is}^{1-\sigma_i}} E_s$$

$p_{ni}^{rs}$  is the consumer price of that good in  $s$  and  $P_{is}$  is the perfect consumer price index of that industry  $i$  in region  $s$ .

There might be various distinct products or producers within an industry  $i$ . They might offer ho-

---

<sup>2</sup>See Brakman et al. (2001).

mogenous or heterogenous commodities. Within the theoretical NEG framework sector assignment for monopolistic and competitive markets is given in advance. For an empirical investigation this is not very handy. The crucial point here is whether households can distinguish products or not. The advantage of the CES index is that it allows to consider homogenous goods in the case of an infinite substitution elasticity<sup>3</sup>  $\sigma_i$ . Thus, we allow various producers to produce a homogenous good - households would consume the product with the lowest price. Then a competitive sector results<sup>4</sup>.

World demand  $x_{ni}^r$  of a single product  $n$  manufactured in region  $r$  is simply the sum of  $x_{ni}^{rs}$  over all regions. For the sake of simplicity we utilise the concept of iceberg transportation cost  $T_{rs}$ . We introduce the freeness of trade<sup>5</sup>, which is in a range of  $0 \leq \phi_{rs} = T_{rs}^{1-\sigma_i} \leq 1$ , and finally find gross world demand  $\bar{x}_{ni}^r$  of a single firm, which is given by

$$\bar{x}_{ni}^r = \alpha_i (q_{ni}^r)^{-\sigma} \sum_s^R \phi_{rs} \frac{E_s}{P_s^{1-\sigma}}, \quad (1)$$

where  $q_{ni}^r$  is the mill price of that producer. Each firm faces world demand as long as trade is not prohibitive. So far we find gross world demand  $\bar{x}_{ni}^r$  based on household utility maximisation. Now we want to consider the firms maximisation problem to supply that quantity.

Following the NEG literature there is a variable input requirement of labour proportional to output. Let  $y_i = \frac{1}{b} l_{in}$  the production technology of a representative firm where  $l_{in}$  is the labour requirement of the  $n$ -th firm. It is worth to note that labour productivity is constant and equalised over all regions. Labour earns the exogenous wage rate  $w_i^r$ . There might be a fixed cost requirement  $\pi_{ni}^r$  to produce at all. Maximising (zero) profits with respect to quantity yields the pricing rule  $q_{ni}^r$ . We allow some price setting opportunity for each firm, such that  $q_{ni}^r$  depends on  $\bar{x}_{ni}^r$ . The resulting mill price is determined by a mark-up on marginal cost and depends on  $\sigma_i$ .

The theoretical models assume that workers are regionally immobile but they can choose the industry in which they work. This makes the model from a theoretical point tractable and allows to normalise wages. The result is that nominal wages  $w_r$  are equalised over all industries within a region. Assuming that there is at least one sector where no transportation cost occur and which is of the homogenous producer type allows to equalise wages of this sector interregionally. Then, all nominal wages are normalised and are equal over all regions and industries. With some normalisation, it yields the pricing rule  $q_{ni}^r = w_r = 1$ ,

<sup>3</sup>For simplicity we assume that  $\sigma_i$  is a constant within industry and therefore common for all firms within that market.

<sup>4</sup>Let  $y_i = \frac{1}{b} l_{ik}$  the production technology of a potential competitive market where  $l_{ik}$  is the labour requirement of the  $k$ -th firm. Total labour requirement  $L_i$  equals  $N_i l_{ik}$ . Substitution in the CES function of that particular industry yields  $C_i = \frac{1}{b} L_i N_i^{\frac{1}{\sigma_i-1}}$ . Taking  $\lim_{\sigma_i \rightarrow \infty} C_i$  yields  $C_i = \frac{1}{b} L_i$  which is the standard formula of the competitive sector in the NEG world.

<sup>5</sup> $\phi_{rs}$  tends to be Zero when trade cost increase. It takes the value One when trade is totally free.

With that pricing rule and zero profits and market clearing (equation (1)) we can derive a coherence between fixed cost  $\pi_{ni}^r$  and output  $\bar{x}_{ni}^r$ , which is given by<sup>6</sup>

$$\pi_{ni}^r = \frac{1}{\sigma_i} \bar{x}_{ni}^r = \frac{\alpha_i}{\sigma_i} \sum_s^R \phi_{rs} \frac{E_s}{P_{is}^{1-\sigma_i}}.$$

It is worth to note that the mark-up on marginal cost to cover  $\pi_{ni}^r$  disappears in the case of  $\sigma_i \rightarrow \infty$ , which is the case of a homogenous good (competitive market). Thus, firms fixed cost recovery depends on the world distribution of expenditures, prices and trade freeness<sup>7</sup>.  $\alpha_i E_s / P_{is}^{1-\sigma_i}$  is a measure of real expenditures  $e_{si}$ . The sum term is called the market potential or market access. In the next step we focus on  $P_{ir}$ , the (unobservable) price index. Within the empirical literature this price index is often assumed to be constant over all regions because data on regional prices is typically not available. It follows that nominal rather than real expenditures are considered. Nominal market access and expenditure is frequently used in empirical studies that investigate implications of NEG<sup>8</sup>. However, the price index can be derived with the help of the expenditure function explicitly. Doing so and using the introduced simplifications we find a coherence between  $P_{ir}$  and the regional distribution of firms of that industry<sup>9</sup>, namely

$$P_{ir} = \left[ \sum_s^R \phi_{rs} N_{si} \right]^{\frac{1}{1-\sigma_i}}.$$

The industry-specific regional price index is an average price depending on trade cost and firms distribution. Substitution of this price index in  $\pi_i^r$  yields

$$\pi_i^r = \frac{\alpha_i}{\sigma_i} \sum_s^R \phi_{rs} \frac{E_s}{\sum_k^R \phi_{ks} N_{ki}} = \frac{\alpha_i}{\sigma_i} \sum_s^R \phi_{rs} e_s. \quad (2)$$

Within an industry firms fixed cost (operating profit) depends solely on the spatial distribution of expenditures and firms. If transportation cost rise demand of other regions will decrease ( $\phi_{rs}^i = T_{rs}^{1-\sigma_i} \rightarrow 0$ ). If they are infinite large supply takes place only in the home region. However, if a region has a high stock of firms and the surrounding regions have also a high stock of firms then the denominator goes up and demand and  $\pi$  of a firm reduces. This pushes firms to other regions (market crowding, dispersion force). If a firm is far away from concentrated regions the denominator gets smaller because of the bias due to  $\phi$  and this will rise  $\pi$  (protection against competition). On the other hand, being closed to bigger markets

<sup>6</sup>See Baldwin et al. (2001).

<sup>7</sup>Every firm within an industry and region faces the same problem such that we drop the index for the  $n$ -th firm in the following.

<sup>8</sup>See Niebuhr (2004).

<sup>9</sup>See Baldwin et al. (2001) for details.

( $E$  of of the own and surrounding regions rise) let rise  $\pi$  what makes it profitable to be closer to the bigger market (home market effect, agglomeration force). Unfortunately, both effects can not be separated because of the double sum. The strength of agglomeration and dispersion forces depends, besides others, on trade cost. This crucial parameter is common within the NEG literature and in theoretical work typically analysed within bifurcation diagrams. Thus, to identify both forces at work we will have to consider various trade cost levels.

The operating profit  $\pi_{ir}$  is almost unobservable although the explanatory part is.  $\pi_{ir}$  is therefore unfeasible for an empirical model as dependent variable. From now on  $\pi_{ir}^*$  labels this unobservable variable. However, this operating profit is essential for firms location decision. Firms would settle down<sup>10</sup> in a region where  $\pi_{ir}^*$  offers the highest discounted income stream  $PV(\pi_{ir}^*)$ . It is amazing that a firm settles down where it has to pay the highest fixed cost: the model assumes that each firm has to be invented by a research sector. Cost of invention have to be covered before running the business. Households saving is invested in a riskless asset. This capital is used to pay for invention and innovation. After the establishment of a single firm the firm has to pay a dividend to shareholders, which are the households. This is  $\pi$  and therefore it is worth to settle down where  $\pi$  is maximised. Following Baldwin et al. (2001) the present value can be calculated by the depreciation rate  $\delta_i$ , the time preference of households  $\theta$  and the (long run) growth rate of the world firm stock  $g_{Ni}$  in the case of endogenous growth,

$$PV(\pi_{ir}^*) = \frac{\pi_{ir}^*}{\delta_i + \theta + g_{Ni}}.$$

Thus, we observe a discrete firm entry in a region where  $\pi_{ir}^*$  offers the highest return and covers cost of invention (Tobins  $q$ ). In a neoclassical model  $g_{Ni}$  is zero in the long-run. In the short-run this does not necessarily be true such that  $g_{Ni} > 0$  and it declines steadily until it is zero.

One new firm needs  $a_{Fi}$  units of labour  $H_i$  of a research sector. Because of the normalisation of wages  $a_{Fi}$  represents the replacement cost of Tobins  $q$  in the case of neoclassical growth. Endogenous growth is driven by  $a_{Fi} = c * 1/N_i^w$ , which follows a learning curve (Romer 1990).  $c$  is some other constant capturing other effects. Thus,

$$a_F \stackrel{\text{Tobins } q}{=} \frac{\pi_{ir}^*}{\delta_i + \theta + g_{Ni}}.$$

If this condition holds, we observe firm start-ups. This equation solves for the long run firm stock or the growth rate of it. In both cases of neoclassical and endogenous growth it depends on the world population. Changes in world population leads to a proportional change of the firms stock or the growth

---

<sup>10</sup>It is assumed that established firms can not relocate.

rate. Solving for  $g_{Ni}$  leads to

$$g_{Ni} = \frac{\pi_{ir}^*}{a_F} - (\delta_i + \theta).$$

Per definition, the world growth rate  $g_{Ni}$  can be decomposed in regional parts,  $g_{Ni} = dN_i/N_i^w = \sum_r^R dN_{ir}/N_i^w$ . Substitution yields

$$\begin{aligned} \frac{dN_{i1} + dN_{i2} + \dots + dN_{iR}}{N_i^w} &= \frac{\pi_{ir}^*}{a_F} - (\delta_i + \theta) \\ \frac{dN_{ir}}{N_i^w} &= -\frac{\sum_{s=1; s \neq r}^R dN_{is}}{N_i^w} + \frac{\pi_{ir}^*}{a_F} - (\delta_i + \theta) \end{aligned} \quad (3)$$

As expected, the regional growth rate is higher the higher  $\pi_{ir}^*$  is. Firm formation in other regions limit the own firm entry process because of the negative sign. Even in the case of neoclassical growth we still observe firm entry to replace depreciated firms. In the presence of strong agglomeration forces we might observe relocation of industries even in the neoclassical case. Therefore equation (3) suits for both types of model. Substitution of the known part of  $\pi_{ir}^*$ , equation (2) gives the basic model from a theoretical perspective.

$$\frac{dN_{ir}}{N_i^w} = -\frac{\sum_{s=1; \forall s \neq r}^R dN_{is}}{N_i^w} + \frac{\alpha_i}{a_F \sigma_i} \sum_s^R \phi_{rs} e_{is} - (\delta_i + \theta) \quad (4)$$

The regional firm entry rate is higher the higher nearby real expenditures  $e_s$  are. The sum term  $\sum \phi_{rs} e_{is}$  is a measure of the real regional-specific market potential. Bergmann and Sternberg (2007) state that agglomeration forces are directly linked to regional demand. Since  $\pi_{ir}^*$  relates to demand, this approach features those effects by a microeconomic foundation. However, Bergmann and Sternberg notice that the identification of agglomeration forces is typically captured by local wages<sup>11</sup> or per capita income<sup>12</sup>. Here the crucial variable is some measure based on firms profitability and market access.

The firm entry process is based on innovation. In the case of a competitive sector those cost and therefore  $\pi_{ir}$  are Zero in the long-run. Furthermore,  $\sigma_i$  tends to be infinity. However, in the short run there might be an additional premium as long as demand exceeds supply. The market potential is therefore a valid instrument to capture firm entry processes.

If firms innovation is costly, then labour input in the research sector is relevant. Human capital is in the literature accepted and seen as a motor of innovative processes. Neither the derived model relies on that assumption nor it takes measures of human capital directly into account. This gives some flexibility in the empirical analysis.  $\pi_{ir}^*$  has to increase the higher research cost are to fulfill Tobins  $q$ . As it was noted

<sup>11</sup>See Berglund and Brännäs (2001) or Gerlach and Wagner (1994).

<sup>12</sup>See Carree (2002) or Ritsilä and Tervo (2002).



earlier  $\pi_{ir}$  gets larger when  $\sigma_i$  takes relative lower values ( $>1$ ). In that case monopoly power of single firms raises. However free market entry reduces monopoly power such that monopolistic competition results. Therefore we conclude that small  $\sigma_i$  estimates relate to higher  $\pi_{ir}$  and this in turn relates to more human capital intensive research rather than to monopoly power price setting. This hypothesis we do not test directly but we will focus on it as a partial result of the estimates. We start with the first stage and ask whether the real market potential relates to firm formation.

In the next section we want to consider equation (4) empirically and derive the empirical specification.

### 3 Empirical Consideration, data and hypothesis

This section focus on applied data, forms hypotheses, introduces the concept of measuring the market potential and motivates further control variabls. Furthermore it derives the empirical model. Before we turn to the more technical part of the empirical model we first focus on data sources. The Establishment History Panel of the Institute of Employment Research IAB of Germany is used to get information on the number of firms and other firm-specific and regional related information of German regions. It is a 50% sample of all German establishments who employ at least one person. In the year 1999 the data collection changed. I.e. more firms were registered in the sample. This discrete step would highly overestimate firm entries between 1998 and 1999 and therefore we consider the time period from 1999 to 2005. We use the industry classification wz73 on a two digit level. We limit the sample and drop the public sector ( $wz73 > 70$ ) but keep corporate services. Furthermore we drop industries which are based on natural resources (i.e. mining). In total we consider 70 distinct industries. Regional data, esp. GDP is taken from the GENESIS regional data base of the Ferderal Statistical German office. Regions are classified following Eckey (2001). This classification aggregate NUTS3 level to 180 labour market regions where the main criteria for aggregation is based on commuting flows. On average 2.4 NUTS3 regions collapse to one labour market region. This aggregation overcomes strong local spatial autocorrelation due to a common labour market. The agregation further captures local industry-specific linkages to some degree.

The last chapter derives a model of regional industry-specific firm growth  $g_{ir}$  based on the spatial distribution of real expenditures, i.e. the real market potential within an industry, namely

$$g_{ir} \equiv \frac{dN_{ir}}{N_i^w} = -\frac{\sum_{s=1; \forall s \neq r}^R dN_{is}}{N_i^w} + \frac{\alpha_i}{a_F \sigma_i} \sum_s^R \phi_{rs} e_{is} - (\delta_i + \theta).$$

On the left hand side there is the number of new established regional firms within an industry normalised by the world firm stock of that industry. This is not the regional (industry-specific) growth rate. Using

the regional growth rate defined as  $dN_{ir}/N_{ir}$  implies an infinite high growth rate for empty regions because  $N_{ir}$  would be zero. Thus, we loose all firm entries when we consider the regional firm growth rate. To reveal market crowding or home market effects it is essential to focus on empty regions as well. However, in the definition  $dN_{ir}/N_i^w$  all available potential empty regions are available and enter the data set. Thus, with 180 regions and 70 industries the data set contains 12600 cases.

$dN_{ir}$  counts all firm entries which take place between 1999 and 2005.  $N_i^w$  is the mass of firms in 1999. This computation rules out any time information. All variables on the right hand side relate to the year 1999.

On the right hand side the first term captures the growth perspective of all other regions. This can be seen as a spatial lag dependence. Assuming a spatial weighting matrix  $W$  we might transform the first part to  $\rho W g_i$ . Griffith (2006) outlines the problem of model misspecification when  $\rho$ , the autoregressive parameter, is negative. Our model however suggests a negative sign. The parameter bundle of the second term is industry specific. From a theoretical point of view the estimates should be positive in the case of monopolistic competition or research intensive industries and insignificant for the homogenous good case ( $\sigma_i \rightarrow \infty$ ). It should not be negative.

The sum term  $\sum_s^R \phi_{rs} e_{is}$  needs to be discussed as well. First of all we have to construct  $e_{is} = E_s / \left( \sum_r^R \phi_{rs} N_{ir} \right)$  and second the sum of  $e_{is}$ . Starting with the former. Assuming a constant saving rate over all regions  $E_r$  can be approximated by total regional GDP. Since we relate  $\pi_{ir}^*$  to household demand we refrain from using gross value added. The consumption expenditure of GDP allocated to a specific industry is scaled by  $\alpha_i$ , the parameter of the utility function.

Unfortunately  $\phi_{rs}$  is industry-specific since it contains  $\sigma_i$  in its calculation, which is a problem. However, we know that  $\phi_{rs}$  is per definition in a range between Zero and One which gives the strategy to approximate  $\phi_{rs} = T_{rs}^{1-\sigma_i}$ : If one assumes that closer regions have higher  $\phi_{rs}$  – values compared to further regions (that means lower trade cost), we might use a distance based weighting matrix  $W$  with elements between 0 and 1 to deflate nominal expenditures<sup>13</sup> with the distance-weighted firm distribution. We employ various weighting matrices which contain the value 1 on the main diagonal<sup>14</sup> to calculate  $e_{ir}$ .

To compute  $\phi_{rs}$  we follow the approach of Bröcker (1989). An element of the weights matrix  $W$  is then given as

$$\phi_{rs} = \exp(-d_{rs}\tau),$$

where  $d_{rs}$  labels the distance and  $\tau$  is a distance decay parameter. This distance decay parameter  $\tau$

<sup>13</sup>See Niebuhr (2004).

<sup>14</sup>The 1 values are necessary such that also the stock of firms of the home region enter the calculation of  $e_{is}$ .

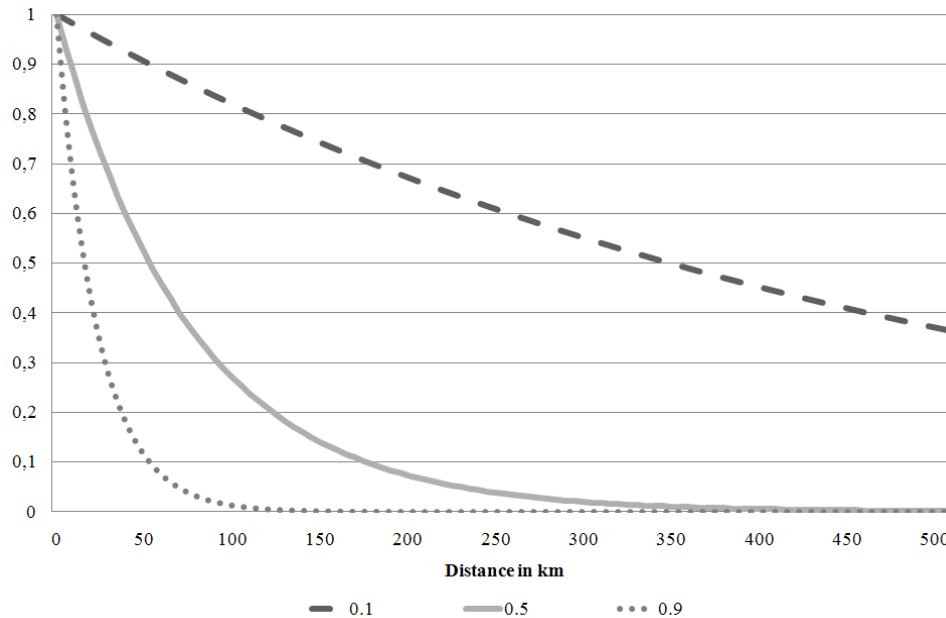


Figure 1: Coherence between  $\phi$  and distance between regions due to various  $\gamma$  values

depends on the average distance of neighboring regions and a normalised distance decay parameter  $\gamma$ , which is between 0 and 1.  $\gamma$  describes the influence of the distance on regional dependence. The lower  $\gamma$  the slower is the reduction of interregional interdependencies with distance. The link between  $\gamma$  and  $\tau$  is

$$\gamma = 1 - \exp^{-\tau D},$$

where  $D$  is the average distance of all regions to their respective neighbors (see Niebuhr 2001). In our case  $D$  is 53 km and  $\gamma$  is chosen to be 0.1, 0.5, and 0.9 to capture the range of two extreme and one moderate decay value<sup>15</sup>. The resulting weighting matrices are labeled by  $W_{0.1}$ ,  $W_{0.5}$  and  $W_{0.9}$ , respectively. The  $W$  matrices are not row standardised because the calculation / deflation is not based on an average but on a potential.

These matrices relate to transportation and trade cost of commodities. Baldwin and Forslid (2000) show theoretically in a NEG model which is closed to our approach that lowering trade cost encourages agglomeration forces. Thus we hypothesise that  $W_{0.1}$  is in favour of agglomeration while  $W_{0.9}$  relates more to dispersion. Coherence between  $\phi$  and distance between regions due to various  $\gamma$  values

<sup>15</sup>Given  $\gamma$  and  $D$  solves for  $\tau$  which enters the calculation of  $\phi_{rs}$ .

Figure 1 presents the coherence between  $\gamma$ , distance and the resulting  $\phi$  value. The key feature here is that for  $\gamma = 0.9$  the weight  $\phi_{rs}$  is about 0.1 when the distance is 50 km. That is, the  $W_{0.9}$  matrix basically considers the home region while values of other regions are highly discounted (high trade cost). On the other extreme for  $\gamma = 0.1$  (low trade cost), the resulting weight  $\phi$  does not decline much with distance. For a distance of about 350 km the weight is still 0.5. The moderate  $W_{0.5}$  matrix gives a weight  $\phi = 0.5$  for a distance between regions of about 50 km. That is that neighbouring regions enter with the half weight of the own region within the calculation.

So far we computed real expenditures  $e_{is}$ . Now we are going to consider the sum-term of  $e_{is}$ . If only real expenditures of the home region are relevant,  $\phi_{rs} \forall r \neq s$  is zero. Then only  $e_{ir}$  enters the regression model. To make the model more tractable we let the logarithm  $\ln e_r$  to enter the regression equation. The resulting model is labeled as *Home*. The coefficient  $\alpha_{1i}$  should be industry specific as discussed above.

If  $\phi_{rs} \neq 0$  we can consider various specifications. Instead of using  $\ln e_{ir}$  we might use  $\ln \bar{W}e_{ir}$  as true market potential<sup>16</sup> to approximate the sum term  $\sum_s^R \phi_{rs} e_{is}$ . The resulting model is labeled as *Potential*. However, this approach does not allow to distinguish increases in home or foreign demand. Therefore an alternative is to split the part into  $\alpha_{1i} \ln e_{ir} + \alpha_{2i} \ln W e_{ir}$  with the coefficients  $\alpha_{1i}$  and  $\alpha_{2i}$  to estimate<sup>17</sup>. The model is labeled as *split*. Again,  $\alpha_{1i}$  should be positive in the presence of agglomeration forces. It might be insignificant as well but not negative. The effect of  $\alpha_{2i}$  captures the effect of increases of real market potential of other regions. One might expect a positive sign. However, if the effect of other regions of  $\ln e_{is}$  is dominant over the potential  $\ln W e_{is}$ , the effect of  $\alpha_{2i}$  could be insignificant or negative. because it is worth to settle down in those other regions.

The last term ( $\delta_i + \theta$ ) of equation 4) relates to the constant term and is again industry specific.

Collecting terms give the basic regression approaches relating to the theoretical model. First, omitting the lag-regressive term, yields

$$g_{ir} = \alpha_{1i} \ln e_{ir} + \mu_i + \varepsilon_{ir} \quad (\text{Home})$$

$$g_{ir} = \beta_{1i} \ln \bar{W}e_{ir} + \mu_i + \varepsilon_{ir} \quad (\text{Potential})$$

$$g_{ir} = \alpha_{1i} \ln e_{ir} + \alpha_{2i} \ln W e_{ir} + \mu_i + \varepsilon_{ir} \quad (\text{Split})$$

<sup>16</sup>  $\bar{W}$  labels a weighting matrix where the elements of the main diagonal are set to One.

<sup>17</sup> The elements of the main diagonal of  $W$  are in that case zero and therefore the bar disappears.

Second, all these models can be enriched by  $\rho W g_i$  which give

$$g_{ir} = \rho W g_i + \alpha_{1i} \ln e_{ir} + \mu_i + \varepsilon_{ir} \quad (\text{Lag-Home})$$

$$g_{ir} = \rho W g_i + \beta_{1i} \ln \bar{W} e_{ir} + \mu_i + \varepsilon_{ir} \quad (\text{Lag-Potential})$$

$$g_{ir} = \rho W g_i + \alpha_{1i} \ln e_{ir} + \alpha_{2i} \ln W e_{ir} + \mu_i + \varepsilon_{ir} \quad (\text{Lag-Split})$$

where  $\mu_i$  is an industry specific effect and  $\varepsilon_{ir}$  an error term.

The equation still contains  $a_F$  which gives some freeness for additional variables controlling for other (productivity) effects. As it was mentioned in the theoretical model considering solely  $e_{ir}$  gives evidence whether the market potential influences firm formation. It does not allow to focus on agglomeration and dispersion forces directly. Therefore we add the share of established firms in the particular region  $s_{Ni} = N_{ir}/N_i^w$ . If the home market effect dominates, then the effect of  $s_{Ni}$  on  $g_{ir}$  should be positive. It also captures competition effects (Porter externalities). If  $s_{Ni}$  rises, regional competition rises. Then the market crowding effect dominates and the estimate of  $s_{Ni}$  should be negative.

We further add the variables *empty* and *monopol* to distinguish whether there is no or at least one firm of that industry established in that region. An empty region might be of advantage that it gives monopoly power to a new-commer. An incumbent might have some monopoly power and therefore higher profits which attracts other firms to enter the market in that region.

A next variable captures diversity effects (Jacobs (1969) externalities). The variable *DIV* relates to intra-regional diversity of employment between industries and is computed by a negative Herfindahl index as Combes et al. (2004) suggest. The more diverse employment is the higher is *DIV*. Another urbanisation measure is the log of the total number of regional industries  $\ln ind$ .

We further add the average industry-specific firms age and its square value. An established region might have lower growth perspectives since the industry is fostered and firms are already very productive such that new commers face strong competition. On the other hand a higher average age might attract new firms because of agglomeration forces.

Since our research field is Germany a dummy variable *East* indicates whether the region is part of East Germany.

The literature on firm-startups often uses population density as a measure of urbanisation and agglomeration forces<sup>18</sup>. The denser a region is one will expect higher firm formation. We capture those effects explicit in  $e_{ir}$ . In denser regions typically total *GDP* is higher because there is a higher stock of

<sup>18</sup>See Armington and Acs (2002), Audretsch and Fritsch (1999), Brixy and Grotz (2007), Fritsch and Falck (2007), Reynolds et. al. (1994), Sternberg and Bergmann (2003) or Sutaria and Hicks (2004).

wage earners. Thus, demand for products and  $e_{ir}$  increase in dense areas. Therefore we omit density measures.

The data set can be seen as a panel data set where the industries relate to the cross-sectional units and the regions relate to the time dimension. An industry-specific effect is already included. The general model reads as

$$y_{ir} = \rho W y + \beta_i x_{ir} + \gamma z_{ir} + \mu_i + \eta_r + \varepsilon_{ir},$$

where  $\eta_r$  is a regional specific effect. There is a set of parameters to estimate which are industry-specific ( $\beta_i$ ) and others which are common for all industries ( $\gamma$ ). Because the model relates to industries we expect that  $\mu_i$  is correlated with some of the explanatory variables of  $x_{ir}$  or  $z_{ir}$ . Therefore we demean the data on industry level such that  $\mu_i$  disappears. We refrain from demeaning over single regions<sup>19</sup> because the regional effect does not have to be relevant for every industry. We assume that it is a pure random effect. The model simplifies to

$$y_{ir} = \rho W y + \beta_i x_{ir} + \gamma z_{ir} + \eta_r + \varepsilon_{ir}.$$

The model still have the industry specific parameter  $\beta_i$  which is unaffected by data transformation. Fritsch and Müller (2004) give evidence that regional firm growth rates are industry-specific. Different branches exhibit distinct pattern and determinants of growth. This supports our approach of industry-specific parameter estimates. We assume that those parameters are randomly distributed over industries<sup>20</sup>.

The estimation strategy is as follows. We estimate a mixed or random coefficient model<sup>21</sup> by restricted maximum likelihood (Thompson 1962) without spatial dependence over all industries and use the full power of the panel structure. The resulting model reads as

$$y_{ir} = \beta_i x_{ir} + \gamma z_{ir} + \eta_r + \varepsilon_{ir}.$$

We can apply all three models (5) to (5) adding the other variables and use the whole sample of 12600 cases. However, because of the omitted spatial lag the estimates could be inconsistent. Therefore we reestimate a spatial lag model for every of the 70 industries. Then, however, we can not estimate the

<sup>19</sup>This would yield a two-way-fixed effects model.

<sup>20</sup>Adding dummy variables for a random slope parameter for each of 70 industries would introduce strong collinearity and therefore we stick to the random coefficient approach.

<sup>21</sup>See e.g. Searle, Casella and McCulloch (1992), McCulloch and Searle (2001) or Raudenbush and Bryk (2002) for an overview.

regional specific effects  $\eta_i$ . This model reads as

$$y_{ir} = \rho W y + \beta_i x_{ir} + \gamma_i z_{ir} + \varepsilon_{ir},$$

where we get an industry-specific estimate  $\beta_i$  and  $\gamma_i$  for every dependent variable. In that case we consider 180 observation per industry and the definition of (5) to (5). The next section outlines and discusses the estimation results.

## 4 Estimation Results

We derived a model of industry-specific firm growth which is outlined in the previous sections. In that section we present the estimation results. First we focus on the models without spatial lag dependence. Tables 1 to 3 report the estimation results of the models (5) to (5), respectively. All tables are organised as follows. The models are estimated with and without regional random effects  $\eta_i$ . The  $W_x$  row in the upper part relate to the applied distance decay function to determine how distance is weighted in the computation of real industry-specific expenditure  $e_{ir}$  and the sum term  $\sum_s^R \phi_{rs} e_s$ . For simplicity we apply the same weighting matrix to determine  $e_{ir}$  and the sum term.

In the next section of the table the estimated parameters are presented. The block after relates to the standard deviation of the random coefficients.  $sd(\ln e_{ir})$  and  $sd(\ln \bar{W} e_{ir})$  relate to the random slope distribution of  $\beta_i$ ,  $sd(region)$  to the distribution parameter of the region-specific effect  $\eta_r$ . As robustness we also allow other variables to be randomly distributed. However, those estimates were insignificant which means that they are not random but have a common estimate for all industries. The reported Wald test relates to the jointly-significance of all estimated parameters, the likelihood-ratio test LR r.c. tests the significance of regional random coefficients  $\eta_r$ .

First we want to give an overall picture of the estimates. All three models indicate that the parameters are jointly significant. In most cases the regional effects  $\eta_i$  are significant. Because the application of different  $W$  definitions relate to different questions, i.e. how far each firms market potential is and because the different model definitions aim to analyse the sum term  $\sum \phi_{rs} e_s$  there is no best model. However, going through the estimates allows to draw an overall picture. First we want to explore common variables.

Increases in the share of firms  $s_{Ni}$  decreases the growth perspective of a region indicating that stronger competition and market crowding effects are present. The parameter is quiet robust and significant. The estimate lies in a range of -0.2306 to -0.2098. The *empty* variable supports this finding since in all models the parameter is positive and highly significant. Thus, empty regions attract new firms. The effect of

Table 1: Regional industry-specific firm growth with random slope for model (home)

<i>W</i> to approximate $\phi_{rs}$	without regional random effects			with regional random effects		
	$W_{0.1}$	$W_{0.5}$	$W_{0.9}$	$W_{0.1}$	$W_{0.5}$	$W_{0.9}$
$\ln e_{ir}$	0.3465*** (0.1113)	0.2650*** (0.0999)	0.2876*** (0.0894)	0.3478*** (0.114)	0.2614*** (0.1012)	0.2926*** (0.0901)
$s_{Ni}$	-0.2291*** (0.003)	-0.2241*** (0.0029)	-0.2121*** (0.003)	-0.2306*** (0.0031)	-0.2257*** (0.003)	-0.2130*** (0.0031)
<i>empty</i>	0.5700*** (0.0753)	0.5668*** (0.0752)	0.4615*** (0.0769)	0.5618*** (0.0753)	0.5583*** (0.0752)	0.4560*** (0.0769)
<i>monopol</i>	-0.4211*** (0.0972)	-0.4238*** (0.0976)	-0.3125*** (0.0989)	-0.4321*** (0.0972)	-0.4327*** (0.0976)	-0.3179*** (0.0989)
<i>East</i>	0.0129 (0.0598)	-0.0897 (0.0568)	-0.0848 (0.0581)	0.0143 (0.0661)	-0.089 (0.0623)	-0.0839 (0.0616)
<i>DIV</i>	0.1008 (0.1042)	0.0926 (0.1053)	0.1707 (0.1062)	0.0999 (0.1157)	0.0875 (0.116)	0.1735 (0.1128)
<i>firm age</i>	0.0499*** (0.0059)	0.0489*** (0.0059)	0.0437*** (0.0061)	0.0502*** (0.0059)	0.0490*** (0.006)	0.0439*** (0.0061)
<i>firm age</i> <sup>2</sup>	0.0008 (0.0005)	0.0007 (0.0005)	0.0015*** (0.0005)	0.0008 (0.0005)	0.0008 (0.0005)	0.0015*** (0.0005)
$\ln ind$	0.2292 (0.5115)	0.9089** (0.4227)	1.1349*** (0.3633)	0.2869 (0.5674)	1.0107** (0.4609)	1.1555*** (0.382)
<i>constant</i>	-1.0243 (2.0691)	-3.7481** (1.7141)	-4.6742*** (1.4751)	-1.2572 (2.2951)	-4.1595** (1.869)	-4.7571*** (1.5507)
<i>sd</i> ( $\ln e_{ir}$ )	0.8171** (0.074)	0.7654*** (0.0704)	0.6780*** (0.0683)	0.8180** (0.074)	0.7664*** (0.0705)	0.6792*** (0.0684)
<i>sd</i> ( <i>Residual</i> )	2.5832*** (0.0163)	2.6058*** (0.0165)	2.6387*** (0.0167)	2.5789*** (0.0164)	2.6018*** (0.0166)	2.6361*** (0.0168)
<i>sd</i> ( <i>region</i> )				0.1512*** (0.0431)	0.1466*** (0.0451)	0.1167*** (0.0539)
Wald $\chi^2$	6606.99***	6996.73***	6497.61***	6503.44***	6830.35***	6260.32***
LR r.c. $\chi^2$				4.04**	3.39**	1.38

N=12600, s.e. in (); \* p<0.1, \*\* p<0.05, \*\*\* p<0.01

regions with only one firm established (*monopol*) is not clear. While it is negative and significant and negative for model (5) and (5) it is insignificant for model (5).

Focussing on the urbanisation measures *DIV* and  $\ln ind$  reveals no additional firm growth is to expect because of employment diversity. However, the total number of regional established industries increases firm formation within an industry. This gives evidence for industrial linkages who affect firm formation and industrial linkages and needs.

The average age of all established firms and especially the squared value allow to draw an interesting picture, since the influence is u-shaped. Both parameters exhibit in all models a positive sign. The age variable was mean centered with Eastern and Western Germany means before the industry-specific centering was done. Therefore the values can take negative values. Computing the minimum of the



Table 2: Regional industry-specific firm growth with random slope for model (potential)

<i>W</i> to approximate $\phi_{rs}$	without regional random effects			with regional random effects		
	$W_{0.1}$	$W_{0.5}$	$W_{0.9}$	$W_{0.1}$	$W_{0.5}$	$W_{0.9}$
$\ln We_{ir}$	-0.2655 (0.2689)	0.1705 (0.1508)	0.3416*** (0.1213)	-0.2854 (0.2969)	0.159 (0.1557)	0.3321*** (0.1219)
$s_{Ni}$	-0.2190*** (0.0029)	-0.2208*** (0.003)	-0.2162*** (0.003)	-0.2216*** (0.003)	-0.2237*** (0.0031)	-0.2173*** (0.0031)
<i>empty</i>	0.5961*** (0.077)	0.6064*** (0.077)	0.5212*** (0.0773)	0.5836*** (0.077)	0.5935*** (0.0769)	0.5185*** (0.0773)
<i>monopol</i>	-0.3681*** (0.1002)	-0.3734*** (0.0999)	-0.3441*** (0.0996)	-0.3755*** (0.1002)	-0.3814*** (0.0999)	-0.3475*** (0.0996)
<i>East</i>	-0.1492** (0.0621)	-0.0603 (0.0643)	-0.0225 (0.0615)	-0.1500** (0.0697)	-0.0609 (0.0719)	-0.024 (0.0645)
<i>DIV</i>	-0.0771 (0.1045)	-0.0859 (0.1031)	0.0584 (0.105)	-0.0783 (0.1178)	-0.0898 (0.1171)	0.0532 (0.1109)
<i>firm age</i>	0.0419*** (0.0061)	0.0435*** (0.0061)	0.0426*** (0.0061)	0.0421*** (0.0061)	0.0438*** (0.0061)	0.0426*** (0.0061)
<i>firm age</i> <sup>2</sup>	0.0011** (0.0005)	0.0012** (0.0005)	0.0014*** (0.0005)	0.0011** (0.0005)	0.0012** (0.0005)	0.0015*** (0.0005)
$\ln ind$	2.8379*** (0.3133)	2.7462*** (0.3236)	2.0041*** (0.3425)	2.9627*** (0.3465)	2.8884*** (0.3604)	2.0756*** (0.3577)
<i>constant</i>	-11.5626*** (1.2718)	-11.2186*** (1.3124)	-8.2176*** (1.3895)	-12.0668*** (1.4062)	-11.7928*** (1.4616)	-8.5071*** (1.4512)
<i>sd</i> ( $\ln We_{ir}$ )	1.0679 (0.3466)	0.9892 (0.1297)	0.8745 (0.0994)	1.0943 (0.3416)	0.995 (0.1299)	0.876 (0.0995)
<i>sd</i> ( <i>Residual</i> )	2.6932*** (0.017)	2.6788*** (0.0169)	2.6585*** (0.0168)	2.6880*** (0.0171)	2.6732*** (0.017)	2.6562*** (0.0169)
<i>sd</i> ( <i>region</i> )				0.1694*** (0.0447)	0.1752*** (0.0436)	0.1114*** (0.0582)
Wald $\chi^2$	7117.06***	7060.89***	6593.3***	6744.85***	6670.82***	6386.08***
LR r.c. $\chi^2$				4.84**	5.57***	1.05

N=12600, s.e. in (); \* p<0.1, \*\* p<0.05, \*\*\* p<0.01

quadratic function reveals an average age of -25 years. This is below the minimum of the data. That states that the firm entry rate increases the older - on average - regional established firms are. This result gives two distinct interpretation: first it could indicate the replacement of depreciated firms. Second it could indicate agglomeration forces: It might be more easy to enter a market where older incumbants are. In other words, incumbants establish new associated firms. Unfortunately, both interpretations relate to different questions and can not be separated at this stage.

Now we are going to consider the crucial variables which relate to  $e_{ir}$ . Estimates of  $\ln e_{ir}$  appear in the models (5) and (5) where the values are a bit lower in the latter model. They are always positive and significant. This is in line with the suggested theory. When firms profit and market potential increase then a firm would settle down in that regions. In those cases we conclude that  $\sigma_i$  is relatively smaller

and thus we identify industries which are not truly competitive. Focussing on the  $W_{0.9}$  columns of table 2 supports this finding. Since for  $\gamma = 0.9$  mainly the home region is relevant, the estimates of  $\ln \bar{W}e_{ir}$  are very closed to those of the models focused on above. However, considering the estimates of  $\ln \bar{W}e_{ir}$  for the other weighting matrices does not allow to draw a picture so far. The positive effect of the home demand disappears. If we consider the estimates of model (5) explains the insignificant result of the 2 model. Here,  $\ln e_{ir}$  is positive and  $\ln We_{ir}$  is negative and both estimates are highly significant. The parameter of  $\ln We_{ir}$  is treated as non-random. As a robustness test we allow for randomness of that parameter but could not find a significant result. If the potential of foreign demand in a region increases then regional growth perspective decreases. This is definitely not in line with agglomeration forces. Apparently, the home demand effect of other regions dominates the effect of agglomeration. This conclusion is independent of the definition of  $W$  such that it does not depend on different level of trade cost. Going thru the random coefficients industry by industry for the random slopes of  $e_{ir}$  – estimates shows that most of them do not differ in sign but in its value. This indicates that trade cost do not change the general picture.

To conclude. Increases of real home market potential and firms profit increases incentives to establish new firms in that region. This is the home market effect. However, increases in the market potential in a region of foreign demand does not speed up regional firm growth. Thus regional demand dominates firm formation and not <world demand>. Other effects such as strong competition and the possibility of monopoly power in empty regions give evidence of strong dispersion forces. The work of Brühlhart et al. (2004) conclude strong dispersion forces. Comparing estimates between the different weighting schemes support Baldwins and Forslids (2000) theoretical finding that lower trade cost strengthen agglomeration forces. The estimates of the  $e_{ir}$  variables are higher.

The next task is to focus on the spatial lag models for each industry separately. With 70 valid industries, three different models to represent the sum term  $\sum \phi_{rs}e_{is}$  and three weighting matrices there would be too much parameters to present. Therefore we only draw an overall picture. The lag-regressive parameter of the spatial lag  $\rho$  is, as expected, in almost every case negative. There are few cases where  $\rho$  is positive, contradicting with the theory. However, this result can be explained by the fact that a positive  $\rho$  occurs basically for  $W_{0.9}$ , where distance is highly discounted. Then the interpretation is that nearby regions have the same sign of the growth perspective  $g_{ir}$ . However, those estimates are seldom.

The majority of  $\rho$  estimates is insignificant although negative. It stands out that a negative and significant  $\rho$  mainly occurs for  $W_{0.1}$ . In that case distance is weakly discounted. This supports the underlying theory: every region is relevant and it indicates some relocation mechanism. One might

expect for relocation that in cases where  $\rho$  is negative and significant also estimates of the  $e_{ir}$  variables are positive and significant. This general statement does not hold.

There might be another coherence between parameter estimates. One might expect that in agglomeration industries  $\rho$  is significant and the estimate of  $s_{Ni}$  is insignificant or positive significant. Again, this conclusion can not be drawn.

To summarize: In about 2/3 of all estimates  $\rho$  is insignificant, in very few cases it is positive significant and when it is negative and significant no general picture occurs. In cases where it is significant about a half is weekly significant on a 10% level. The majority of insignificant  $\rho$  estimates indicates that the potential inconsistency of the omitted spatial lag is a minor problem. So, the estimated parameters of the random coefficient models seem to be valid.

The theoretical model assumes that firms has to be invented. It was stated that innovation is based on human capital. We can draw at this stage an interesting picture: Industries with a positive random slope for  $\ln e_{ir}$  engage at least 5% of their employees with persons which hold an university degree. Industries with a negative random slope typically do not employ person with university degree. Of course, in both cases are few exceptions. Thus, the empirical effidence is in line with the theoretical thoughts. However, the linkage between parameter estimates and human capital input (here: employment of person which hold an university degree) should be explored in more detail.

The estimates identify significant agglomeration and dispersion forces. They further support the crucial influence of real market access on firm formation. We can conclude that basically demand of the home region fosters regional firm growth. We further find strong dispersion forces, i.e. a high share of firms within the home region lowers significant the firm growth rate. As a first outlook we can state that industries which seem to be less competitive are more research intensive.

To sum up we can state that the market potential and real expenditure are crucial variables for firm formation and that agglomeration and dispersion forces are relevant. The estimates support therefore basic elements of the NEG literature.

## 5 Conclusion

This paper develops an empirical approach to test fundamental forces of the New Economic Geography literature. Based on theoretical work of Baldwin (1999), Baldwin et al. (2001) and Martin and Ottaviano (1999) a theoretical model is developed to explain regional industry-specific firm formation. While in the literature on that topic frequently labour productivity measures are applied this paper works out a market

access measure based on firms expected average profit. The theoretical model is then tested empirically. Germany is chosen as research field. In total, we consider 70 different industries based on a two digit level. Every firm who enters the market between 1999 and 2005 is counted. We find evidence that firms will settle down in regions where profit and real expenditure gets higher. This is in line with the idea of agglomeration forces. However, we also find dispersion and competition forces. Regions with a high share of firms of a particular industry face a significant lower firm growth rate.

As a first insight we can state that industries who are more affected by market access and which seem to be less competitive (following the interpretation of the theoretical model) are those industries who are more human capital intensive. This is in line with the theoretical model as well because those industries should be more research intensive. However, at this stage this is just a first draw and should be considered in more detail.

We conclude that the market potential and real expenditure are crucial variables to explain firm formation and that agglomeration and dispersion forces are relevant. The estimates support basic elements of the NEG literature.

## 6 Literature

1. Armington, C., Z. J. Acs (2002): The determinants of regional variation in new firm formation, *Regional Studies* 36, 33–45.
2. Audretsch, D.B., M. Fritsch (1999): The Industry Component of Regional New Firm Formation Processes, *Review of Industrial Organization* 15, 239-252.
3. Baldwin, R.E. (1999): Agglomeration and endogenous capital, *European Economic Review* 43, 263-280.
4. Baldwin, R.E., R. Forslid (2000): The Core-Periphery Model and Endogenous Growth: Stabilizing and Destabilizing Integration, *Economica* 67, 307-324.
5. 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.
6. Berglund, E., K. Brännäs (2001): Plants' entry and exit in Swedish municipalities. *The Annals of Regional Science* 35, 431–448.
7. Bergmann, H., R. Sternberg (2007): The changing face of entrepreneurship in Germany. *Small Business Economics* 28, 205–221.
8. Brakman, S., H. Garretsen, C. van Marrewijk (2001): *An Introduction to Geographical Economics: Trade Location and Growth*, Cambridge University Press: Cambridge.
9. Brakman, S., H. Garretsen, M. Schramm (2002): New Economic Geography in Germany: Testing the Helpman-Hanson Model, *HWWA Discussion Paper No. 172*, Hamburg.
10. Brixy, U., R. Grotz (2002): Räumliche Differenzierungen von Betriebsgründungsintensität und Überlebenschancen in Westdeutschland 1983 bis 1997. *Raumforschung und Raumordnung* 60, 100–122.
11. Bröcker, J. (1989): Determinanten des regionalen Wachstums im sekundären und tertiären Sektor der Bundesrepublik Deutschland 1970 bis 1982. Florentz, München (Schriften des Instituts für Regionalforschung der Universität Kiel, vol. 10.)
12. Brühlhart, M., M. Crozet, P. Koenig (2004): Enlargement and the EU Periphery: The Impact of

- Changing Market Potential. *The World Economy* 27 (6), 853-875.
13. Carree, M. A. (2002): Does unemployment affect the number of establishments? A regional analysis for the United States. *Regional Studies* 36, 389-398.
  14. Combes, P.-P., T. Magnac, J.-M. Robin (2004): The dynamics of local employment in France, *Journal of Urban Economics* 56, 217-243.
  15. Eckey, H.F. (2001): Der wirtschaftliche Entwicklungsstand in den Regionen des vereinigten Deutschlands, *Discussion Paper no. 20/01*, Universität Kassel.
  16. Fritsch, M., O. Falck (2007): New business formation by industry over space and time: A multi-dimensional analysis. *Regional Studies* 41, 157-172.
  17. Gerlach, K., J. Wagner (1994): Regional differences in small firm entry in manufacturing industries: Lower Saxony, 1979-1991. *Entrepreneurship and Regional Development* 6, 63-80.
  18. Glaeser, E.L., H. D. Kallal, J.A. Scheinkman, A. Shleifer (1992): Growth in Cities, *Journal of Political Economy* 100 (6), 1126-1152.
  19. Griffith, D.A. (2006): Hidden negative spatial autocorrelation, *Journal of Geographical Systems* 8, 335-355.
  20. Hanson, G.H. (2000): Market Potential, Increasing Returns, and Geographic Concentration, *University of Michigan and NBER*, (revised version of NBER Working Papers 6429).
  21. Henderson, V. (2003): Marshall's scale economies, *Journal of Urban Economics* 53, 1-28.
  22. Henderson, V., A. Kuncoro, M. Turner (1995): Industrial Development in Cities, *Journal of Political Economy* 103 (5), 1067-1090.
  23. Jacobs, J. (1969): *The Economy of Cities*, Random House: New York.
  24. Martin, P., G. Ottaviano (1999): Growing locations: industry location in a model of endogenous growth, *European Economic Review* 43, 281-302.
  25. McCulloch, C.E., S.R. Searle (2001): *Generalized, Linear, and Mixed Models*, New York: Wiley.
  26. Mion, G. (2003): Spatial Externalities and Empirical Analysis: The Case of Italy, *Journal of Urban Economics* 56: 97-118.
  27. Niebuhr, A. (2001): Convergence and the effects of spatial interaction, *Jahrbuch für Regionalwissenschaft* 21: 113-133.
  28. Niebuhr, A. (2004): Spatial Effects of European Integration: Do Border Regions Benefit Above Average?, *HWWA Discussion Paper* 307, Hamburg.
  29. Niebuhr, A. (2006): Market access and regional disparities: New economic geography in Europe, *The Annals of Regional Science* 40 (2), 313-334.
  30. Raudenbush, S.W., A.S. Bryk (2002): *Hierarchical Linear Models: Application and Data Analysis Methods*. 2nd ed., Thousand Oaks, CA: Sage.
  31. Reynolds, P., D. J. Storey, P. Westhead (1994): Cross-national comparisons of the variation in new firm formation rates, *Regional Studies* 28, 443-456.
  32. Ritsilä, J., H. Tervo (2002): Effects of unemployment on new firm formation: Micro-level panel data evidence from Finland, *Small Business Economics* 19, 31-40.
  33. Rosenthal, S.S., W.C. Strange (2004): Evidence on the Nature and Sources of Agglomeration Economies, in *Handbook of Regional and Urban Economics* 4, Ch. 49.
  34. Searle, S.R., G. Casella, C.E. McCulloch (1992): *Variance Components*, New York: Wiley.
  35. Sternberg, R., H. Bergmann (2003): Global Entrepreneurship Monitor: Unternehmensgründungen im weltweiten Vergleich, *Länderbericht Deutschland 2002*.
  36. Sutaria, V., D. A. Hicks (2004): New firm formation: Dynamics and determinants, *Annals of Regional Science*, 38, 241-262.
  37. Thompson, W.A. (1962): The problem of negative estimates of variance components, *Annals of Mathematical Statistics* 33, 273-289.

Table 3: Descriptive statistics of model variables

<i>W</i> to approximate $\phi_{rs}$	without regional random effects			with regional random effects		
	$W_{0.1}$	$W_{0.5}$	$W_{0.9}$	$W_{0.1}$	$W_{0.5}$	$W_{0.9}$
$\ln e_{ir}$	0.3170***	0.2261**	0.2387***	0.3173***	0.2239**	0.2452***
	0.1109	0.0992	0.0892	0.1129	0.1005	0.0899
$\ln We_{ir}$	-0.1457***	-0.1436***	-0.1432***	-0.1442***	-0.1434***	-0.1438***
	0.0175	0.0171	0.0152	0.0175	0.0171	0.0152
$s_{Ni}$	-0.2242***	-0.2201***	-0.2098***	-0.2253***	-0.2216***	-0.2106***
	0.0031	0.003	0.003	0.0031	0.003	0.0031
<i>empty</i>	1.1897***	1.2012***	1.2419***	1.1778***	1.1928***	1.2391***
	0.1058	0.1064	0.1129	0.1059	0.1065	0.113
<i>monopol</i>	-0.0984	-0.0966	0.051	-0.1094	-0.1043	0.0474
	0.1045	0.1048	0.1059	0.1045	0.1049	0.1059
<i>East</i>	-0.0088	-0.1342**	-0.1483**	-0.0079	-0.1334**	-0.1474**
	0.0597	0.0569	0.0583	0.0645	0.0621	0.0623
<i>DIV</i>	0.0902	0.072	0.1383	0.089	0.0682	0.1425
	0.1039	0.1051	0.1059	0.1128	0.1151	0.1135
<i>firm age</i>	0.0619***	0.0607***	0.0591***	0.0620***	0.0608***	0.0593***
	0.0061	0.0061	0.0062	0.0061	0.0061	0.0063
<i>firm age</i> <sup>2</sup>	0.0010*	0.0010*	0.0017***	0.0010*	0.0010**	0.0017***
	0.0005	0.0005	0.0005	0.0005	0.0005	0.0005
$\ln ind$	0.6541	1.4410***	1.6271***	0.699	1.5246***	1.6436***
	0.5127	0.4262	0.3658	0.5556	0.4618	0.3869
<i>constant</i>	-1.9119	-5.1022***	-5.9731***	-2.1015	-5.4407***	-6.0368***
	2.0663	1.717	1.4764	2.24	1.862	1.5627
<i>sd</i> ( $\ln e_{ir}$ )	0.8133**	0.7592***	0.6749***	0.8139**	0.7601***	0.6760***
	0.0736	0.0699	0.068	0.0737	0.07	0.0681
<i>sd</i> ( <i>Residual</i> )	2.5763***	2.5987***	2.6296***	2.5730***	2.5949***	2.6267***
	0.0163	0.0164	0.0166	0.0164	0.0165	0.0167
<i>sd</i> ( <i>region</i> )				0.1316***	0.1419***	0.1253***
				0.0469	0.0456	0.0506
Wald $\chi^2$	6712.1***	7106.37***	6631.3***	6626.93***	6946.78***	6358.76***
LR r.c. $\chi^2$				2.43*	3.06**	1.84*

N=12600, s.e. in (); \* p&lt;0.1, \*\* p&lt;0.05, \*\*\* p&lt;0.01