PAPERS on Economics & Evolution



# 0523

# A Stochastic Theory of Geographic Concentration and the Empirical Evidence in Germany

by

**Thomas Brenner**

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

ISSN 1430-4716



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

# *A Stochastic Theory of Geographic Concentration and the Empirical Evidence in Germany*

**Thomas Brenner***
Max-Planck-Institute for Research into Economic Systems
Evolutionary Economics Unit
Kahlaische Str. 10
07745 Jena, Germany
brenner@econ.mpg.de

Preliminary version

* I want to thank Tom Brökel and Guido Bünstorf for helpful comments and discussions. The usual disclaimer applies.

*ABSTRACT*. A stochastic model of the evolution of the firm population in a region and industry is developed. This model is used to make predictions about the expected probability distribution of the firm number in regions and their dynamics. Data on the spatial distribution of firms in Germany is used to check the predictions and estimate the parameters of the model. This is done for 196 industries separately.
*KEYWORDS*: geographic concentration, industry dynamics, local clusters, empirical methodology.
*JEL classification*: C12, L60, R12

## 1. Introduction

Geographic concentration has attracted much attention in economics and geography in recent year. This was mainly triggered by some outstanding or exemplary cases of such geographic concentration, especially Silicon Valley and industrial districts in the Third Italy (see, e.g., Becattini 1990 and Saxenian 1994). Many different explanations for the existence of such geographic concentration have been put forward. Examples are the concept of industrial districts (see, e.g., Marshall 1920, Becattini 1990, Vou & Wilkinson 1994, van Dijk 1995, Markusen 1996 and Pietrobelli 1998), the concept of local clusters (see, e.g., Porter 1990 and 1994), and the concept of innovative milieux (see, e.g., Camagni 1995 and Maillat 1998). All these concepts have in common that they argue that firms benefit from other firms co-located in the same region in some way. They differ in terms of what mechanisms cause the benefits. Proposed mechanisms are, for example, spillovers, a joint development and use of human capital in the region, the attraction of suppliers and service firms to the region, and cooperation. Independent of these details, all of these mechanisms lead to the fact that regions with already a large number of firms (in a certain industry) become attractive for further firms (of this industry). In New Economic Geography models have been developed that are based on this mechanism (see, e.g., Krugman 1991 and 1996 and Keilbach 2000). Further theoretical models of geographic concentration that are based on the same kind of local mechanisms have be put forward (see Maggioni 2002 and Brenner 2004).

However, the models of New Economic Geography are not developed to fit them to empirical data and thus check their adequateness to describe actual geographic concentration. The models of Maggioni (2002) and Brenner (2004) make very general predictions which are tested by these authors. Maggioni mainly states that during the emergence of local clusters the economic activity follows an s-curve and proves this empirically for the U.S.A. and Italy (see Maggioni 2002). Brenner deduces from his theoretical model that industry-specific economic activity should be distributed bimodally among regions and shows that this holds empirically for at least around half of the manufacturing industries in Germany (see Brenner 2004). Although in both approaches industries are studied separately, these approaches are not able to estimate the industry-specific strength of geographic concentration.

Ellison and Glaeser (1997) offer such a measure. They start from the assumption that firms are randomly distributed in space and calculate a value that measures the deviation of the empirical data from this assumption (or, to be precise, something similar to it), the so-called Ellison-Glaeser index (see Ellison & Glaeser 1997). Hence, Ellison and Glaeser came so far somehow nearest to measure the strength of geographic concentration in different industries. However, they do not provide an explicit theory of how this geographic concentration comes about and, thus, they are not able to measure the strength of the underlying forces. They discuss the different forces that might be relevant and they estimate in a later paper to what extent the observed geographic concentration can be explained by local resources (see Ellison & Glaeser 1999). However, different kinds of deviations from a random location of firms might lead to the same Ellison-Glaeser index. This is caused by the fact that Ellison and Glaeser do not model the mechanisms that they believe to be responsible for the geographic concentration they observe. They only develop a counter-model.

This paper adds to the Ellison-Glaeser approach by studying one potential cause of geographic concentration in detail: local self-augmenting processes. A similar approach is taken by Bottazzi, Dosi, Fagiolo and Secchi (2005) who develop a stochastic model of firm location in which they include a dependence of location choices on the decisions already made by other firms. We develop this model further, mainly by omitting the restrictive assumption that the influence of other firms on a location choice is linear.

Above it has been argued that all concepts of local clustering are based on some kind of local self-augmenting processes. Therefore, a stochastic theory of firm location is developed here on the basis of the existence of such local self-augmenting processes. Empirical data from Germany is used to fit the parameters of the model and to check whether the model is able to describe the empirical data. This offers two results. First, it is examined whether the empirical spatial distribution of each industry can be explained by a theory that is based on local self-augmenting processes. Second, the strength of the local self-augmenting processes is a parameter of the model. Through fitting the model to the empirical data we obtain a measure of the strength of the local self-augmenting processes. We obtain such a measure for each industry that is studied. In total 196 3-digit industries in Germany are studied.

The paper proceeds as follows. In the next section a stochastic model of firm location is developed. The empirical data and the applied method for fitting the model to the empirical data is described in Section 3. Section 4 contains the empirical results and some discussion. Section 5 concludes.

## 2. Model

We develop a stochastic model of firm location in the following. Although geographic concentration might also show up in the form of the spatial concentration of employees of a industry, we focus on the geographic distribution of firms or firm sites here. This means that we neglect the fact that different locations might offer different possibilities to grow for firms. Only the number of firms or firm sites in each location is studied here. This restriction is shared with the previous works by Ellison and Glaeser (1997) and Bottazzi, Dosi, Fagiolo and Secchi (2005).

### 2.1. Basic assumptions

Before we can go into the details of modelling the location choices of firms, we have to decide about the temporal order in which firms make these decisions. Ellison and Glaeser (1997) do not have to care about this. In their model the choices of location of different firms are independent of each other. If firms make their location choices dependent on the location of other firms the temporal order of these choices matters. Bottazzi, Dosi, Fagiolo and Secchi (2005) assume that there is a given number of firms but that at any time firms can exit or enter the market. They calculate the equilibrium spatial distribution that results from a permanent exiting and entering of firms. An alternative approach would be to assume that all existing firms have decided one after the other about their location in the past and stay in their locations for ever.

In reality relocation happens rarely and the observed spatial distribution of an industry shows quite some dependence on the original location of firms. Furthermore, the number of exits and entries typically decreases if an industry becomes mature (see Klepper 1997). However, there is, at least, some continuous dynamics in firm location in the form of relocation, exits and entries. Hence, assuming all firms to locate once for ever seems not to fit the empirical observations. We decide to build the model on the same assumption as Bottazzi, Dosi, Fagiolo and Secchi (2005). This means that we assume that either a sufficient number of relocations or a sufficient number of exits and entries occur in each industry, so that the geographic distribution of an industry changes permanently.

The basic assumption that is necessary for the analysis below is that in each region there is always a positive probability that an existing firm exits or moves to another region and that in each region there is always a positive probability that a new firm is started or moves into the region. The probabilities might be very small but they should be above zero. This implies that the stochastic model that is developed here is ergotic. A system is ergotic if all possible states of the system can be reached with a positive probability independent of the history of the system, even if it might take very long to reach them. The ergotic characteristic of a system implies that the probability to find the system in a certain state at a randomly picked point in time is exactly the same as the probability to find the system in this state after an infinitely long time (see Haken 1983). The probability to find the system in a certain state after an infinitely long time is given by the stable stationary probability distribution of the stochastic dynamics. It can be calculated similar to the equilibrium of a deterministic model. However, it also describes the situation at any point in time that is temporarily not too near to the specific initial conditions. All industries that are studied here are included in the industry classification since more than 10 years and therefore exist for quite some time already. This means that we do not have to assume the system to be in an equilibrium state and can, nevertheless, make predictions on the basis of the stationary states of the system.

From a modelling point of view, the relocation of one firm leads to exactly the same dynamics in the spatial distribution of firms as the joint event of an exit and an entry. Hence, we only consider exits and entries in the following stochastic model.

### 2.2. Modelling entries

Let us start with modelling entries. To this end, we consider one region and ask the question of how likely a firm enters the industry under consideration in this region. This can, for example, been done by thinking of an evaluation of the option to start a firm in the industry and region under consideration. One could also think of an expected utility or the expected profits that can be made by founding such a firm. Several factors play a role.

First, the situation in the industry matters. Start-ups are more likely to occur in industries that have on average high profit rates and low entry barriers. However, this effects all regions similarly and has mainly an impact on the total number of entries. We can model this by a multiplicative factor $\varsigma_i$ that determines how many start-ups are appearing in industry $i$.

Second, local conditions influence the profitability of founding a firm in a certain region. Many influential conditions could be named, such as the available human capital, wage rates, transport costs, the availability of natural resources, public research institutions, or land prices. Some are connected to the feasibility to run a firm in the region, others are connected to the costs of operating a firm or the ability to enter new markets. For a first empirical analysis we consider only one factor in this paper. However, the model allows for an easy inclusion of many other factors and it is intended to conduct further studies with more factors in the future. The factor that is included here is the number of people that are available to be employed in a region. This number has two different impacts on the likelihood of a firm founding in a region. On the one hand, the founder of a firm usually lives already before founding a firm in the same region. Hence, the number of people in a region determines the number of potential founders. The more people are in a region, the more people can found a firm. On the other hand, the availability of potential employees influences the possibilities to run a firm.

In the former case the likelihood of entries would have to be multiplied by the number of people in the region. In the later case the evaluation of founding a firm would depend on the number of people in the region. We tested both kinds of models and found that the former kind of modelling fitted the empirical data better. The number of people who are employed in region r ($\in \{1, ..., 97\}$) is denoted by $w_r$ here. We use this number as a proxy for how many people life in a region who might found a firm. The structure of German regions is such that regions with a high number of employees are the big cities while regions with a low number of employees are usually rural areas. Since the number of employees in a region is only a proxy for the number of potential founders and since this number might have also other effects, such as attracting firms to the regions, we chose a very flexible modelling. Hence, we assume that the probability of an entry in region $r$ and industry $i$ is given by

$$p_{i,r,entry} = \varsigma_i \cdot w_r^{\sigma_i} \tag{2.1}$$

where $\sigma$ is a parameter that denotes to what extent the number of employees in the region matters. This means that the model contains a measure for the importance of the size of the region on the location decisions of firms. If location choices are made independent of the size of the region, $\sigma = 0$ should hold. If entries are even more frequent in rural areas, $\sigma$ should be negative.

Third, we consider local self-augmenting processes, as they are argued to be responsible for the existence of local clusters. Basically, local self-augmenting processes mean that it is the more likely that firms are founded in a region or move to a region the more firms are located there already (see also Bottazzi, Dosi, Fagiolo & Secchi 2005). This means that the likelihood or evaluation $v_{i,r}$ of founding a firm in industry $i$ and region $r$ increases with the number of firms $n_{i,r}$ that are already located in region $r$ and belong to industry $i$. It is unclear what functional form this dependence has. Bottazzi, Dosi, Fagiolo and Secchi (2005) assume that the probability of an firm entry increases linearly with the number of firms located there. We want to be more flexible here. Since all the processes mentioned in the literature on local clusters, such as human capital accumulation, spillovers, social interaction and so on, might be involved, we choose a general formulation:

$$v_{i,r} = \nu_{S,i} \cdot n_{i,r}^{\rho_{S,i}} \tag{2.2}$$

where $\rho_{S,i}$ is a parameter that determines the shape of the impact of the already existing firms and $\nu_{S,i}$ is a parameter that determines how much co-location with other firms matters. Both parameters might differ between industries. If $\rho_{S,i} = 1$ we obtain the assumption by Bottazzi, Dosi, Fagiolo and Secchi (2005). If $\rho_{S,i} < 1$ the impact of firms already present in a region decreases with their number. This means that there are diminishing returns to co-location. If $\rho_{S,i} > 1$ the opposite holds: the more firms are located in one region the more another firm benefits from each of them.

Fourth, if $\rho_{S,i} > 1$, the attractiveness of a region would increase without limit for an increasing number of firms in the region. However, it is well known that an increasing number of firms in a region, at least from a certain threshold onwards, increases wages, rents and so on in this region. This means that the costs of running a firm there increase and the evaluation of the region decreases. Hence, we can expect to have also a negative impact of the firm number $n_{i,r}$ on the evaluation $v_{i,r}$ of founding a firm in this region. Again we use a very general formulation and expand the evaluation function by another term:

$$v_{i,r}(n_{i,r}) = \nu_{S,i} \cdot n_{i,r}^{\rho_{S,i}} - \phi_{S,i} \cdot n_{i,r}^{\eta_{S,i}} \tag{2.3}$$

where $\eta_{S,i}$ is a parameter that determines the mathematical shape of this negative impact and $\phi_{S,i}$ determines its strength. If there is no such negative impact, $\phi_{S,i} = 0$ can be expected. If $\eta_{S,i} = \rho_{S,i}$, the two terms on the right-hand side of Equation (2.3) can be merged to one. In general, it can be expected that the negative impact of a high number of firms in a region dominates the positive impact for very large numbers of firms, while the positive impact should be stronger than the negative one for smaller numbers of firms. Mathematically this implies $\eta_{S,i} > \rho_{S,i}$.

Equation (2.3) defines the evaluation of founding a firm in a region. Further factors could be added, but in this paper we limit the considered factors to those included in Equation (2.3). Now, we have to determine how potential founders decide on the basis of this evaluation. A logit approach is used here. In a logit approach it is assumed that people vary in their evaluation of decision options because of mistakes or individual preferences and characteristics (see Mc Fadden 1981). Hence, the objectively determined evaluation $v_{i,r}$ represents only the average of the evaluations in a population of individuals. If the evaluations are Gumbel-distributed within this population, as it is assumed in the logit approach (see McFadden 1981), the probability that one individual chooses option $(i, r)$ of a set of options $\mathcal{R}$ is given by

$$\frac{\exp[\mu \cdot v_{i,r}(n_{i,r})]}{\sum_{q\in\mathcal{R}}\left(\exp[\mu \cdot v(q)]\right)} \; . \tag{2.4}$$

$\mu$ is a parameter that determines the variation of the evaluation within the population. This means that $\mu$ denotes the heterogeneity of the individuals. The higher $\mu$, the less different are the individuals in their evaluation of the expected profits from the different options. $\mathcal{R}$ denotes the set of all other activities that each individual can take instead of founding a firm in region $r$. This set includes founding a firm in other regions as well as founding firm in another industry or not founding a firm at all. The evaluation can be interpreted as a consideration of the expected profits from founding a firm in comparison to the opportunity costs. However, the mathematical characteristics of the model allow us to ignore the details of the alternative opportunities. The denominator of the fraction on the right-hand side of Equation (2.4), which includes all these alterative opportunities, is the same for all options $r$ and can be replaced by a constant $c$. Thus, we obtain for the probability that a firm entries in region $r$ (including the effects discussed above leading to Equation (2.1):

$$p_{i,r,entry}(n_{i,r}) = \frac{\varsigma_i}{c} \cdot w_r^{\sigma_i} \cdot \exp[\mu \cdot (\nu_{S,i} \cdot n_{i,r}^{\rho_{S,i}} - \phi_{S,i} \cdot n_{i,r}^{\eta_{S,i}})] \; . \tag{2.5}$$

The alternative opportunities only influence the value of $c$. This approach is also adequate for a situation in which a firm moves from one region to another. In this case, the firm has also to make a decision between several possible regions, which can be modelled as described above.

### 2.3. Modelling exits

Let us now come to modelling exits. As mentioned above, we do not model the size of firms. This implies that we are not able to distinguish the likelihood to exit between firms of different size. All firms in a region are treated as being identical. Hence, we simply assume that each firm in a region has the same probability to exit the market. This probability is denoted by $p_{i,r,exit}$.

The situation of modelling is somewhat different in the case of exits compared to entries. If a firm exits because of bankruptcy, there is no decision made. If, however, a firm site disappears in a region because it moves to another region, the above approach would be applicable.

Let us start with a discussion of a firm or firm site that is really closed down. If we assume that all firm in a region and industry are similar, the probability of such an event in a region and industry is proportional to the number $n_{i,r}$ of firms or firm sites that might exit. In contrast to entries, the probabilities of exits seems not to depend on the number of employees in a region or the size of the region. However, the arguments about location economies should also hold for exits. In regions with a lot of firms of the same industry they should benefit from each other and should therefore have a lower risk of exiting. The same arguments as above hold. However, the functional form of the depends on the number of existing firms is less clear. There is no empirical evidence or theoretical argument that could be used.

Thus, we assume the following in order to keep the model as simple as possible: We assume that the number of existing firms influences the probability of exiting in an exponentially increasing and decreasing way, similar to the model for entries:

$$p_{i,r,exit}(n_{i,r}) = n_{i,r} \cdot \exp[\mu \cdot (\nu_{E,i} \cdot n_{i,r}^{\rho_{E,i}} - \phi_{E,i} \cdot n_{i,r}^{\eta_{E,i}})] \ . \tag{2.6}$$

The parameters $\nu_{E,i}$, $\rho_{E,i}$, $\phi_{E,i}$ and $\eta_{E,i}$ have the same characteristics as the parameters $\nu_{S,i}$, $\rho_{S,i}$, $\phi_{S,i}$ and $\eta_{S,i}$, respectively. Their interpretation is, however, somewhat different. Here they determine the dependence of the likelihood of an exit on the number of existing firms, while above they determine the evaluation of a region on the basis of the number of existing firms. Furthermore, if $\nu_{S,i}$ is positive, $\nu_{E,i}$ can be expected to be negative. What makes entries more likely should make exits less likely. The same holds for $\phi_{S,i}$ and $\phi_{S,i}$.

Equation (2.6) has the advantage that it fits also a process in which a firm decides to move to another location. In this case the second term on the right-hand side of Equation (2.6) can be interpreted as an evaluation of the option to leave the region.

### 2.4. Dynamics

Above we have set up the probabilities of exits and entries in a region $r$ and industry $i$. We now set up a Markov chain for describing the dynamics in one region. To this end, we assume that two events, such as exits or entries, never occur at exactly the same time. This assumption is supported by the fact that we might make the time unit of the modelling as small as we like without a consequence for the analysis below. This implies that only three things can happen with the number of firms $n_{i,r}$ in a region at any point in time: the number might remain constant, might increase by one, or might decrease by one. The probabilities for these changes are given by the probabilities of exits and entries. We obtain a Markov chain. A

Markov chain is characterised by three characteristics. First, the state of the system, here the region, can be denoted by one natural number, the number of firms $n_{i,r}$ in our case. Second, this number changes by maximally one each point in time. Third, the probabilities for such changes only depend on the current state, but not on the history of the system.

The dynamics of a Markov chain can be described by the so-called Master-equation (see Weidlich 1991). To this end, we set up a stochastic formulation of the state of a region $r$. We define $P_{i,r}(n,t)$ as the probability to find $n$ firms in region $r$ and industry $i$ at time $t$.

In order to describe the dynamics of the system, let us assume that at time $t$ region $r$ actually contains $n$ firms of industry $i$. This implies that three states can be reached at time $t+1$. It might be that nothing happens, implying that at time $t+1$ there are still $n$ firms in region $r$. It might be that one firm enters. This happens with a probability given by $p_{i,r,entry}$ and implies that $n+1$ firms are located in region $r$ at time $t+1$. Finally, it might be that one firm exits (given that $n \geq 1$). This happens with a probability given by $p_{i,r,exit}$. Such an event leads to $n-1$ firms located in region $r$ at time $t+1$.

Now, we can ask how $P_{i,r}(n,t+1)$ looks like, given that we know all probabilities $P_{i,r}(n,t)$ at time $t$. There are three event that might lead to a certain number of firms $n$ in region $r$ at time $t+1$. First, there might have been $n-1$ firms in this region at time $t$ and one new firm has been founded. This happens with a probability given by $P_{i,r}(n-1,t) \cdot p_{i,r,entry}(n-1)$. Second, there might have been $n$ firms in region $r$ at time $t$ and no firm has entered or exited. The probability for such an event is $P_{i,r}(n,t) \cdot (1 - p_{i,r,entry}(n) - p_{i,r,exit}(n))$. Third, there might have been $n+1$ firms in region $r$ at time $t$ and one firm has exited. This happens with a probability of $P_{i,r}(n+1,t) \cdot p_{i,r,exit}(n+1)$. Hence, we obtain:

$$\begin{aligned} P_{i,r}(n,t+1) \quad = \quad & P_{i,r}(n-1,t) \cdot p_{i,r,entry}(n-1) \\ & + P_{i,r}(n,t) \cdot (1 - p_{i,r,entry}(n) - p_{i,r,exit}(n)) \\ & + P_{i,r}(n+1,t) \cdot p_{i,r,exit}(n+1) \ . \end{aligned} \tag{2.7}$$

Rearranging Equation (2.7) we obtain:

$$\begin{aligned} P_{i,r}(n,t+1) - P_{i,r}(n,t) \quad = \quad & p_{i,r,entry}(n-1) \cdot P_{i,r}(n-1,t) \\ & + p_{i,r,exit}(n+1) \cdot P_{i,r}(n+1,t) \\ & - (p_{i,r,entry}(n) + p_{i,r,exit}(n)) \cdot P_{i,r}(n,t) \ . \end{aligned} \tag{2.8}$$

This equation is called the master-equation and it completely defines the stochastic dynamics of the system. The first two terms on the right-hand side of Equation (2.8) are the probability inflows into state $n$, while the last term on the right-hand side of Equation (2.8) describes the probability outflow from state $n$.

## 2.5. Probability prediction

As mentioned above, the model that has been set up above is ergodic. This implies that the stationary probability distribution $P_{i,r,st}(n)$ also denotes the likelihood that the system is found in state $n$ at any randomly chosen time (Haken 1983). Hence, the stationary probability distribution $P_{i,r,st}(n)$ is the prediction for a cross-sectional empirical study.

In the case of an ergotic Markov chain the stationary probability distribution can be calculated with the help of detailed balance (Weidlich 1991). To this end, we have to define the probability flow $w_{i,r,n\to\tilde{n}}$ from one state $n$ to another state $\tilde{n}$. The probability flow is zero for all states $n$ and $\tilde{n}$ that satisfy $\mid n-\tilde{n} \mid \geq 2$ because we assumed that never two events occur at the same time. Hence, only two cases have to be considered: $\tilde{n}=n-1$ and $\tilde{n}=n+1$. From Equation (2.8) we obtain:

$$w_{i,r,n\to n-1} = p_{i,r,exit}(n) \tag{2.9}$$

and

$$w_{i,r,n\to n+1} = p_{i,r,entry}(n)\ . \tag{2.10}$$

In the case of detailed balance the stationary probability distribution is given by (see Weidlich 1991)

$$P_{i,r,st}(n) = \prod_{m=0}^{n-1} \frac{w_{i,r,m\to m+1}}{w_{i,r,m+1\to m}} \cdot P_{i,r,st}(0) \tag{2.11}$$

where $P_{r,st}(0)$ is determined by the condition

$$\sum_{n=0}^{\infty} P_{i,r,st}(n) = 1\ . \tag{2.12}$$

Inserting Equations (2.5), (2.6), (2.9), and (2.10) into Equation (2.11) we finally obtain

$$P_{i,r,st}(n) = P_{i,r,st}(0) \cdot \prod_{m=0}^{n-1} \frac{\varsigma_i \cdot w_r^{\sigma_i} \cdot \exp\left[\mu \cdot \nu_{S,i} \cdot m^{\rho_{S,i}} - \mu \cdot \phi_{S,i} \cdot m^{\eta_{S,i}}\right]}{c \cdot (m+1) \cdot \exp\left[\nu_{E,i} \cdot m^{\rho_{E,i}} - \phi_{E,i} \cdot m^{\eta_{E,i}}\right]}\ . \tag{2.13}$$

This equation can also be written in the form:

$$P_{i,r,st}(n) = P_{i,r,st}(0) \cdot \prod_{m=0}^{n-1} \frac{\varsigma_i \cdot w_r^{\sigma_i} \cdot \exp\left[\mu \cdot \nu_{S,i} \cdot m^{\rho_{S,i}} - \mu \cdot \phi_{S,i} \cdot m^{\eta_{S,i}} - \nu_{E,i} \cdot m^{\rho_{E,i}} + \phi_{E,i} \cdot m^{\eta_{E,i}}\right]}{c \cdot (m+1)}\ . \tag{2.14}$$

We realise that this equation now contains four quite flexible functions for the influence of the actual number of firms in a region and industry on the further development of this number. It can be argued that a positive and a negative impact are present which cannot be represented in one term. However, four terms seem to be too many. For several industries we checked whether three instead of two such terms improve the fit to the data significantly (likelihood ratio test) and whether one instead of two such terms describe the empirical data similarly well. Both has been not the case. Hence, two terms seem to be adequate. Therefore, we reduce the above equation to

$$P_{i,r,st}(n) = P_{i,r,st}(0) \cdot \prod_{m=0}^{n-1} \frac{C_i \cdot w_r^{\sigma_i} \cdot \exp\left[\nu_i \cdot m^{\rho_i} - \phi_i \cdot m^{\eta_i}\right]}{(m+1)} \,. \tag{2.15}$$

We also defined $C_i = \frac{\varsigma_i}{c}$ and removed $\mu$. Mathematically this does not reduce the generality of the model. This fact and the reduction to two terms has some interpretational consequences.

First, if we fit the model to empirical data we are not able to disentangle the effect of the overall founding activity in the industry $\varsigma_i$ and the impact of alternatives to founding a firm $c$. All that we get is a measure $C_i$ that is a mixture of two very different influences and therefore impossible to interpret. Second, we will not be able to estimate that value of $\mu$ which determines to what extend people decide for the best option.

Third, we will not be able to figure out to what extend the existence of firms in a region influences entries or exits. The impact on entries enters the equation in the same way as the impact on exits. All we get is information about the combined influence on entries and exits.

All these limitations are not caused by the mathematical simplification that we conducted above. They are caused by the fact that we only have data about the existing firms at a certain point in time. As a consequence, we fit the predicted distribution (2.15) to the empirical data. This does for logical reasons not allow to disentangle the impacts discussed above. To disentangle the impact on entries and exits, data on entries and exits would be needed. This is planned for the future.

Unfortunately, it is impossible to simplify Equation (2.15) into a mathematical form that does not contain a $\prod$ or $\sum$. Therefore, it is also not possible to calculate $P_{r,st}(0)$ in a simple form. The stationary probability distribution can only be treated numerically.

However, this does not restrict the empirical analysis. Equation (2.15) describes the theoretical prediction for finding a certain number of firms $n$ in a certain region $r$ and industry $i$ dependent on six parameters: $C_i$, $\sigma_i$, $\nu_i$, $\rho_i$, $\phi_i$, and $\eta_i$. Examples of such a theoretically predicted distribution are given in Figure 1.

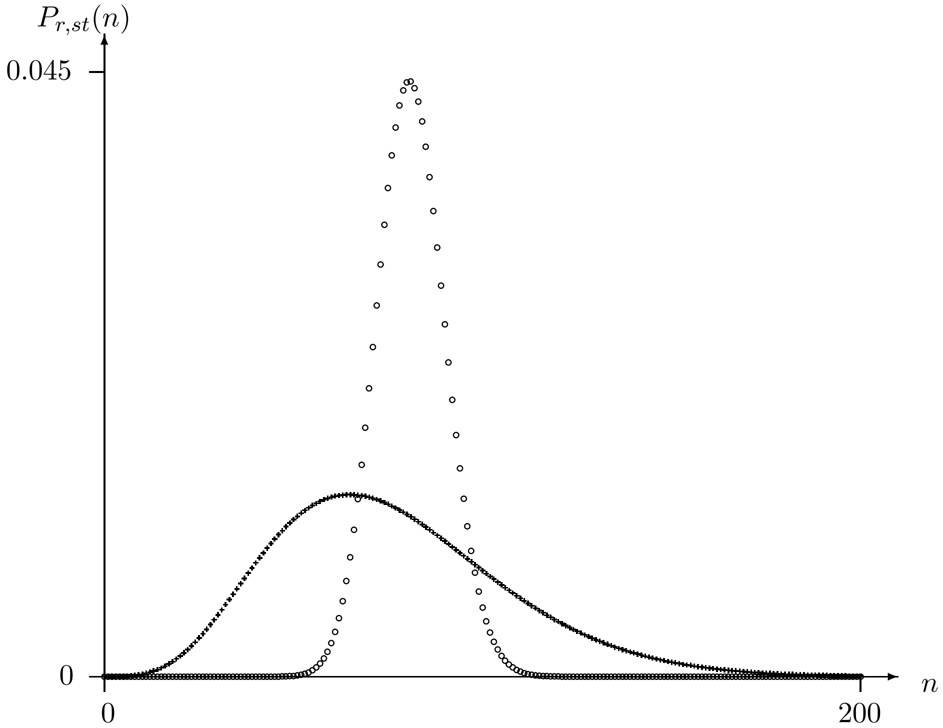


**Figure 1:** Theoretically predicted probability distribution for the number of firms in a region with average size according to the complete model (+) and the reduced model (o). The parameters of the model are fitted to the industry 'manufacture of plastic products'.

## 3. Empirical data and method

### 3.1. Empirical data

The data that are used here have been collected by the German Federal Institute for Labour. The dataset contains the number of firm sites for each 3-digit industry[1] and each of the 97 'Raumordnungsregionen'[2] in Germany for the 30th of June in 2003. The number of firm sites is the number of firm sites at which at least one person is employed. Firms that consist only

[1] Industries are classified according to the WZ93-classification, which has been the standard classification of industries in Germany at the considered time.

[2] 'Raumordnungsregion' are the kind of regions used in German statistics that come most nearest to labour market areas. They take into account the commuting of people. However, they do not split the 440 German administrative regions ('Kreise') and do not include areas from different states,

of the owner are not included in the data. This does not matter much in most industries, but might lead to a bias in a few industries, especially some service industries in which self-employed, one-person firms are frequent. If a firm has several sites in the same municipality, it is counted as one firm site.

The study that is conducted here is restricted to 196 of the 222 industries on the 3-digit level. We exclude all industries that contain less than 100 firm sites because the distribution of less than 100 firm sites seems to be not very representative for the underlying mechanisms. However, none of the excluded industries is of importance in the discussion of local clusters in the literature. Furthermore, we exclude those industries that represent single households and general state expenditure. 10 of the considered industries belong to agriculture and mining, 92 are manufacturing industries, and 94 are service industries. The industries are denoted by $i$ $(\in \{1, 2, ..., 198\}$. The regions are denoted by $r$ $(\in \{1, 2, ..., R\}$, $R = 97)$. The number of firms sites in each region and industry is denoted by $n_{i,r}$.

### 3.2. Fitting the model to the data

Above a clear stochastic prediction has been deduced from the model: Equation (**??**). This prediction defines for each region the likelihood that each number of firm sites should be expected to occur in reality. Therefore, fitting the parameters is straight-forward. We search for the parameter set that causes the actual numbers of firms in each region to be predicted with the highest probability. The actual numbers of firm sites in each region $r$ and industry $i$ are given by $n_{i,r}$. Hence, the likelihood of the occurrence of the actual site numbers according to the above model is given by

$$L_i(C_i, \sigma_i, \nu_i, \rho_i, \phi_i, \eta_i) = \prod_{r=1}^{97} \left( P_{r,st}(0) \cdot \prod_{m=0}^{n_{i,r}-1} \frac{C_i \cdot e_r^{\sigma_i} \cdot \exp\left[\nu_i \cdot m^{\rho_i} - \phi_i \cdot m^{\eta_i}\right]}{(m+1)} \right) , \quad (3.1)$$

where $P_{i,r,st}(0)$ is still defined by Equation (2.12).

The parameters have to be chosen such that the likelihood $L(C_i, \sigma_i, \nu_i, \rho_i, \phi_i, \eta_i)$ is maximised. To this end an Evolutionary Strategy (see Rechenberg 1973) is used[3]. By this, we

which makes them different from what real labour market areas would be in some cases. Nevertheless, they are to most adequate spatial unit available in Germany.

[3] Before the likelihood has been maximised a number of pre-studies have been conducted to find the setting of the Evolutionary Strategy that leads to the fastest convergence. Finally, an Evolutionary Strategy with a population of 60 individuals is used. For the next generation the individuals are picked with a probability proportional to their fitness, which equals the likelihood $L(C_i, \sigma_i, \nu_i, \rho_i, \phi_i, \eta_i)$.

obtain for each industry $i$ the parameter values $\hat{C}_i$, $\hat{\sigma}_i$, $\hat{\nu}_i$, $\hat{\rho}_i$, $\hat{\phi}_i$, and $\hat{\eta}_i$ for which the model describes the actual distribution of firm sites among the German regions best. The maximum likelihood value $\hat{L}_i$ is also obtained.

### 3.3. Checking the adequateness of the model

The model can be fit to the data in the way described above even if it is completely wrong. Therefore, we examine the goodness of the model with the help of the Komolgorov-Smirnov test. The problem that we face is that the model predicts a different probability distribution over the number of firm sites for each region because the distribution depends on the number of employees $e_r$ in the region. The predicted distributions cannot be tested separately because we have for each region only one realisation.

To solve this problem, the distributions for the regions are aggregated. This means that we calculate the probability predicted by the model that in a randomly picked region a certain number of firm sites is found. A predicted probability distribution over the numbers of firm sites results. This is compared with the help of the Komolgorov-Smirnov test with the aggregated actual distribution of the numbers of firm sites in the regions. The result tells us whether the model developed above is able to describe the actual spatial distribution of firm sites.

We do not only want to test whether the model is adequate but also whether the specification of the interaction between the location decisions of firms is necessary and adequate. Therefore, we will check whether the ability of the model to describe the actual data improves significantly in comparison to the same model without considering the interdependence of location choices. Such a model is obtained if we set $\nu_i = 0$ and $\phi_i = 0$. This means that only an entry rate that depends on the size of the region and an exit rate proportional to the number of firms in a region are considered. This model, called reduced model in the following, can also be fitted to the empirical data, which is done with the same procedure as the fit of the complete model. Only two parameters, $C_i$ and $\sigma_i$, have to be fitted in this case. Again the Komolgorov-Smirnov test is used to check whether this model is able to explain the actual spatial distribution of firms. If the reduced model fails the Komolgorov-Smirnov test, while the complete model passes the test, we have evidence that the interdependence between location decisions is necessary for the explanation of the actual spatial distribution of the considered industry.

With a probability of 1% crossovers are build and all parameter values of the individuals of the new generation are moved by a normally distributed random value that has for 20 individuals a variance of 10% of the actual value and for 40 individuals a variance of 5% of the actual value. It has been found that only occassionally minor improvements are detected after 500 generations. Therefore, 500 generations are run for each fit.

To test whether the complete model explains the empirical data significantly better than the reduced model, we use a likelihood ratio test. Fitting the parameters of the reduced model, we obtain a maximum likelihood value, which we call the maximum likelihood value $\hat{L}_i^{(r)}$ of the reduced model. This can be compared to the maximum likelihood value that is obtained for the complete model. We calculate the log-likelihood-ratio

$$\log\left(\frac{\hat{L}_i}{\hat{L}_i^{(r)}}\right) \tag{3.2}$$

and check this ratio for significance. If it is significant, the inclusion of an interdependence between location decisions increases the ability of the model to describe the empirical data significantly (for an example of how the two models differ in their predictions can be seen in Figure 1).

## 4. Results and discussion

### 4.1. Adequateness of model

Before the fitted parameters of the model are presented and discussed, we want to know how well the above developed model describes reality. A Komolgorov-Smirnov test is used. In total 196 industries are tested. For only 13 of these industries the Komolgorov-Smirnov test rejects the above model on the basis of the empirical data:

- Farming of animals,
- Growing of crops combined with farming of animals (mixed farming),
- Textile weaving,
- Manufacture of other products of wood, manufacture of articles of cork, straw and plaiting materials,
- Manufacture of lighting equipment and electric lamps,
- Retail sale of second-hand goods in stores,
- Restaurants,
- Sea and coastal water transport,
- Other supporting transport activities,
- Letting of own property,
- Real estate activities on a fee or contract basis,
- Administration of the State and the economic and social policy of the community, and
- Activities of other membership organizations.

This means that the model chosen above is strongly confirmed. Furthermore, there is no obvious difference between the different industries, not even between agriculture & mining, manufacturing and service industries. The model describes all of them equally well. In the following analysis only those industries are considered that are well described by the model because only for these industries the results can be trusted. However, this only restricts the considered number of industries from originally 196 to 183.

A second check of the model is based on a log-likelihood ratio test. We obtain even stronger confirmation from this test. Of the 198 industries only in two cases we do not find that the complete model describes the actual spatial distribution significantly better than the reduced model. These industries are the renting of automobiles and the manufacture of sports goods. For all other industries we can state that the locations of firm sites are not independent from each other. However, this does not imply that in these industries economies of location play a role or even that firms prefer to be located near to each other. An alternative explanation is that the firms of one industry are attracted by the same local factors that are given to a different extent in different regions. Examples for such local factors are public research institutes, natural resources and human capital. The approach taken here does not allow us to distinguish between agglomeration caused by the availability of local resources and agglomeration caused by economies of location and clustering forces. What we can state is that a model that includes interdependencies of location decisions describes the actual spatial distribution of firms significantly better than a model without such interdependencies and that this holds for all industries. Hence, as Ellison and Glaeser (1997), we show that a model in which firms locate randomly in space misses relevant mechanisms of firm location. We go beyond Ellison and Glaeser's work by offering a model that is able to predict the actual spatial firm distribution adequately, at least for most industries.

The failure of a model that assumes random locations on the basis of the sizes of the region can be shown with the help of the Komolgorov-Smirnov test industries are too many to present the result for each of them. Therefore, we put the industries together in broader classes and present for each class the number of industries for which the model proves to be adequate. This gives quite an overview of what kind of industries are well described by the model and what kind of industries the models fail to capture. The results of the Komolgorov-Smirnov test for both models, the complete and the reduced, are given in Table 1. The reduced model, which does not assume any interdependence in location decisions, fails to describe the actual spatial distribution of firms for nearly half of the analysed industries. It does very poorly in the mining industries, where it is rejected for all industries. 36 of the 92 analysed manufacturing industries are well described by the reduced model. Most adequate is the reduced model in case of the service industries. It is rejected only for 25 of the 94 analysed service industries.

### 4.2. Parameters of the model

The parameters that are obtained by fitting the model to the empirical data give us some information about the mechanisms that cause the spatial distribution of firm sites. Therefore, we will discuss the values obtained for the various parameters in the following. An overview on the parameter values is given in Table 2 and 3.

As mentioned above, the parameter $\hat{C}_i$ cannot be interpreted precisely. It somehow denotes the amount of founding activity in the industry. Hence, it is no surprise that some service industries, in the classes of construction, wholesale, hotels and restaurants, and eduction, are characterised by very high values of $\hat{C}_i$. Besides this, no evident differences between the industries are observed. Of more interest is the parameter $\hat{\sigma}_i$. In the model is was assumed that the number of employees in a region increases the number of firm founding there. However, we did not assume a linear dependence but choose a more flexible dependence of the form:

$$p_{i,r,entry} \propto e_r^{\sigma_i} \ . \tag{4.1}$$

Of course, the expectation was that $\hat{\sigma}_i$ varies somehow around one if the size of the region (in form of the number of employees) really matters and around zero if the size of the region has no impact. For most industries the former should hold. However, Table 2 shows that $\hat{\sigma}_i < 1$ holds for all industries and that for most industries $\hat{\sigma}_i$ is much nearer to zero than to a value of one. This would imply that the size of regions does not matter, at least not very much for most industries.

To investigate into this matter further, the reduced model is quite helpful. The reduced model has also been fitted to the empirical data and contains also a parameter $\hat{\sigma}_i$, which we denote by $\hat{\sigma}_i^{(r)}$ in the following. The difference to the complete model is that the reduced model does not contain the interdependence between location decisions. In Figure 2 the values of $\hat{\sigma}_i$ and $\hat{\sigma}_i^{(r)}$ are plotted for all industries. Three observations are of interest. First, the two values are correlated. This means that if the size of regions plays an important role according to the reduced model, it also plays a role according to the complete model. This means that $\hat{\sigma}_i$ measures somehow the same in the two models. That is what should be expected and tells us that including interdependencies between location decisions has not messed up things.

Second, the values of $\hat{\sigma}_i^{(r)}$ are much higher than the values of $\hat{\sigma}$ and range around a value of one, as we originally expected. Excluding the interdependence between location decisions, the

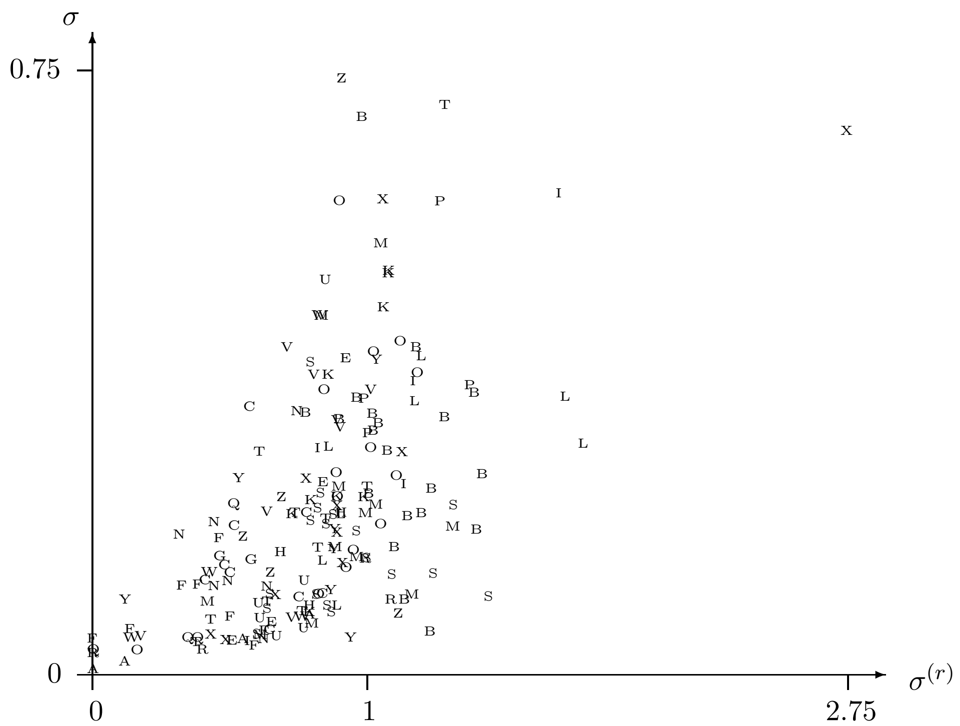


**Figure 2:** Values of the fitted parameters $\sigma$ and $\sigma^{(r)}$ for each industry. The values are marked by letters that correspond to the groups of industry defined in Table 2.

number of entries really depends linearly on the size of regions, on average. Some industries are more likely to be found in large regions while other are more likely located in small regions, but on average the number of entries matches the size of regions. This matching disappears if we include the interdependence between location decisions. This means that quite some part of the fact that more firms enter in larger regions can be explained by the fact the firms tend to locate where other firms of the same industry have already located. In other words, most start-ups and new firm sites do show up in regions with a large population not because there are more employees that could found a firm, but because the conditions are more favourable in places where already other firm sites of the same industry are located. The latter happens to be the case very often in regions with large populations. The size of a region matters directly only to a limited extent but matters indirectly because the first firms are more likely to be started there and others follow. Empirically this shows up in the fact that most firm sites, at least for many industries, are found in the large regions but that not every large region contains a large number of firms of a considered industry. As a consequence, fitting the model to the empirical data leads to the above result.

Third, not all industries show the same behaviour. While most industries have a value of $\hat{\sigma}_i^{(r)}$ that is approximately 1 and a value of $\hat{\sigma}_i$ that ranges between 0 and 0.75, there are some exceptions. The first, and very obvious, exception is the industry of scheduled air transport with $\hat{\sigma}_i^{(r)} = 2.74$. This industry is very strongly concentrated in those regions that have the highest number of employees, meaning the very urban regions. This seems not to be very surprising for this kind of industry. It is more surprising that this industry is the only one that shows such a strong concentration in the population agglomeration.

Besides this, there is a number of industries with a very low value of $\hat{\sigma}_i^{(r)}$. If we consider all industries with $\hat{\sigma}_i^{(r)} < 0.2$ we obtain

- Foresting, logging and related service activities,
- Fishing, fish farming and related service activities,
- Extraction and agglomeration of peat,
- Processing and preserving of fish and fish products,
- Manufacture of dairy products,
- Saw-milling and planing of wood; impregnation of wood,
- Manufacture of agriculture and forestry machinery,
- Manufacture of watches and clocks,
- Building and repairing of ships and boats, and
- Camping sites and other provision of short-stay accomodation.

All these industries are clearly rooted in rural areas. In addition, if we look at those industries with $0.2 < \hat{\sigma}_i^{(r)} < 0.5$ most of them belong to the above classes of food products etc. (F), mining and quarrying (Q), non-metallic products (N), and textiles (T). Service industries are nearly not found in this list, except such industries as building and repairing of ships and boats and camping sites and other provision of short-stay accomodation. Most service industries have values of $\hat{\sigma}^{(r)}$ around or above one because they are concentrated in urban areas.

The parameters $\hat{\nu}_i$, $\hat{\rho}_i$, $\hat{\phi}_i$ and $\hat{\eta}_i$ describe the benefits that are obtained by firms from being located in regions where already other firms of the same industry are located. Table 3 gives an overview on these parameters.

The first interesting result is that the parameters $\hat{\nu}_i$ and $\hat{\phi}_i$ are never negative. While the parameters $\hat{\rho}_i$ and $\hat{\eta}_i$ are restricted to positive values during the fitting of the parameters, no such restriction is imposed on the parameters $\hat{\nu}_i$ and $\hat{\phi}_i$. Hence, they might become negative. If they are both positive, the co-location with other firms of the same industry has as well a positive impact, $\hat{\nu}_i n_{i,r}^{\hat{\rho}_i}$, as a negative impact, $\hat{\phi}_i n_{i,r}^{\hat{\eta}_i}$ (see Equation 2.3). For $\hat{\nu}_i > 0$ and $\hat{\phi}_i < 0$ co-location would only have a positive impact. For $\hat{\nu}_i < 0$ and $\hat{\phi}_i > 0$ it would only have a negative impact. Both seem not to be the case.

Comparing the sectors – agriculture & mining, manufacturing and service – there are little obvious differences in the values of the parameters. It can be observed that $\hat{\rho}_i$ is rarely above 1 and that $\hat{\rho}_i > 1$ never holds for mining or service industries. $\hat{\rho}_i$ determines how the benefits from co-locating with other firms of the same industry increase with the number of these other firms. If $\hat{\rho}_i < 1$, each additional firm adds less to the benefits than the firms already there. If $\hat{\rho}_i > 1$, each additional firm adds more than the already existing firms to the benefits of co-location. Hence, $\hat{\rho}_i > 1$ means that their are increasing returns to co-location, at least until they are out-weighted by the negative term $\hat{\phi}_i n_{i,r}^{\hat{\eta}_i}$. Such increasing returns to co-location are only observed for the following three industries:

- Manufacture of bricks, tiles and construction products, in baked clay,
- Manufacture of weapons and ammunition, and
- Manufacture of accumulators, primary cells and primary batteries.

The parameters are $\hat{\nu}_i$, $\hat{\rho}_i$, $\hat{\phi}_i$ and $\hat{\eta}_i$ are quite related. Figure 3 depicts these relationships. It can be observed that the parameters $\hat{\nu}_i$ and $\hat{\phi}_i$ are positively correlated and that $\hat{\nu}_i > \hat{\phi}_i$ holds for almost all industries. Hence, the positive impact of one firm in a region on the decision of the next firm is stronger than its negative impact in almost all industries. This usually changes for larger numbers of firms because in most industries $\hat{\eta}_i > \hat{\rho}_i$ holds, so that the negative impact becomes stronger than the positive impact for large numbers of firms.

## 5. Conclusion

In this paper a model of co-location has been developed. This model goes beyond the modelling approaches in the literature. Therefore, the first aim was to show that this model is able to describe adequately the spatial distribution of most industries in Germany. This was done very successfully. 183 of 196 studied industries can be described adequately with this model.

The model is fitted to empirical data in this paper. This opens up the possibility to conduct many different analysis. Only some of these analysis are conducted here. The questions that are addressed here are those about the impact of the size of a region on the location decision of firms and the strength and shape of the co-location forces.

We find that the population in a region matters for the likelihood that the first firms in

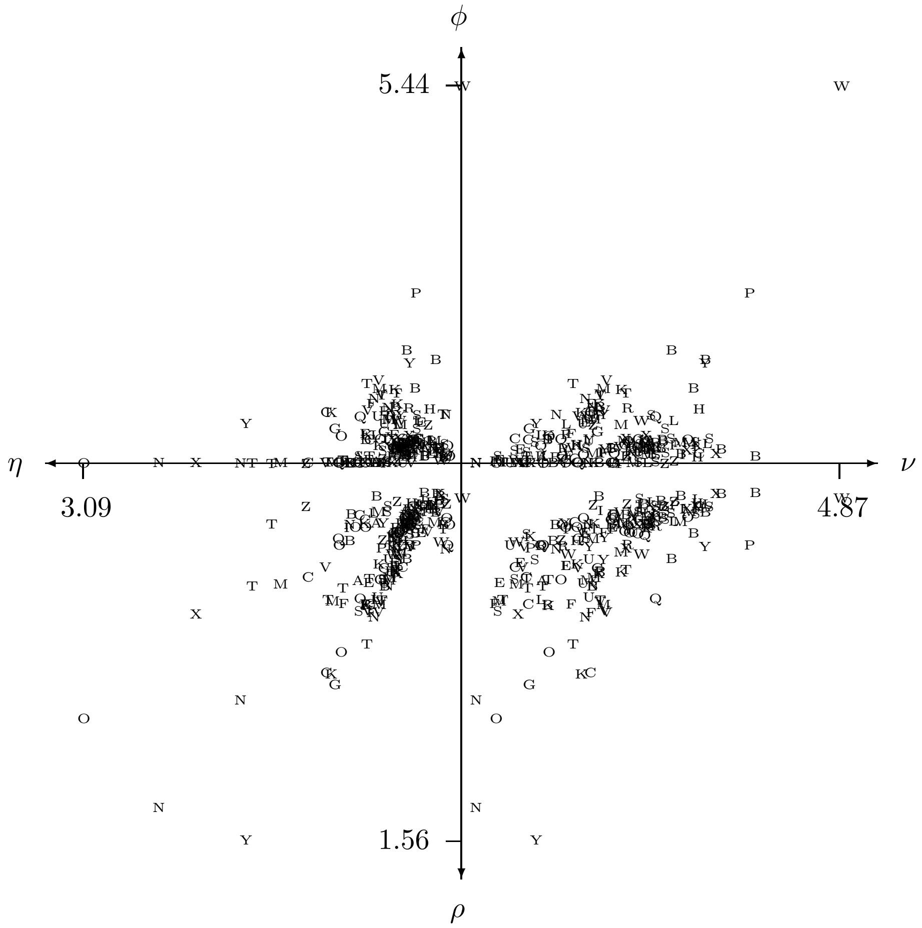


**Figure 3:** Values of the fitted parameters $\nu$, $\rho$, $\phi$ and $\eta$ for each industry (each quarter of the diagramm depicts two parameters, so that each industry is represented by four marks). The values are marked by letters that correspond to the groups of industry defined in Table 2.

an industry are founded in this region. However, firms that start later or firm sites that are established later locate rather near to existing firm sites in the industry than in regions with a large population.

Furthermore, we find that in almost all industries co-location matters, although to a varying degree. We also find that in most industries there is a positive effect of co-location with a small number of other firms but a negative effect of co-location with many other firms.

These are some first results obtained from the new model of the spatial industry dynamics that is developed here. However, the model allows for many more analysis. Just to mention two of them, local resources could be easily integrated into the model and analysis and data on entries and exits would allow to also check the dynamic characteristics of the model. Hence, we hope for other researcher to join in using the model to obtain a better understanding of spatial industry dynamics.

## References

Becattini, G. (1990). The Marshallian Industrial District as a Socio-economic Notion. In: (F. Pyke, G. Becattini & W. Sengenberger (Eds.)) *Industrial Districts and Inter-firm Co-operation in Italy*, pp. 37-51. Geneva: International Institute for Labour Studies.

Bottazzi, G., Dosi, G., Fagiolo, G. & Secchi, A. (2005). *Sectoral and Geographical Specificities in the Spatial Structure of Economic Activitie*. Sant' Anna School of Advanced Studies, Pisa.

Brenner, T. (2001). *Self-organisation, Local Symbiosis of Firms and the Life Cycle of Localised Industrial Clusters*. Papers on Economics & Evolution #0103, Max-Planck-Institute Jena.

Brenner, T. (2004). *Local Industrial Clusters: Existence, Emergence and Evolution*. London: Routledge.

Camagni, R.P. (1995). The Concept of Innovative Milieu and Its Relevance for Public Policies in European Lagging Regions. *Papers in Regional Science* **74**, 317-340.

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

Ellison, G. & Glaeser, E.L. (1999). The Geographic Concentration of Industry: Does Natural Advantage Explain Agglomeration?. *American Economic Review* **89**, 311-316.

Haken, H. (1983). *Synergetics*. Berlin: Springer Verlag.

Keilbach, M. (2000). *Spatial Knowledge Spillovers and the Dynamics of Agglomeration and Regional Growth*. Heidelberg: Physica Verlag.

Klepper, S. (1997). Industry Life Cycles. *Industrial and Corporate Change* **6**, 145-181.

Krugman, P. (1991). Increasing Returns and Economic Geography. *Journal of Political Economy* **99**, 483-499.

Krugman, P. (1996). *The Self-Organizing Economy*. Cambridge: Blackwell Publishers.

Maggioni, M. (2002). *Clustering Dynamics and the Location of High-Tech-Firms*. Heidelberg: Physica-Verlag.

Maillat, D. (1998). From the Industrial District to the Innovative Milieu: Contribution to an Analysis of Territorised Productive Organisations. *Recherches Economiques de Louvain* **64**, 111-129.

Markusen, A. (1996). Sticky Places in Slippery Space: A Typology of Industrial Districts. *Economic Geography* **72**, 293-313.

Marshall, A. (1920). *Principles of Economics*. London: Macmillan.

Pietrobelli, C. (1998). The Socio-Economic Foundations of Competitiveness: An Econometric Analysis of Italian Industrial Districts. *Industry and Innovation* **5**, 139-155.

Porter, M.E. (1990). *The Competitive Advantage of Nations*. New York: The Free Press.

Porter, M.E. (1994). The Role of Location in Competition. *Journal of the Economics of Business* **1**, 35-39.

Rechenberg, I. (1973). *Evolutionsstrategie: Optimierung technischer Systeme nach Prinzipien der biologischen Evolution*. Stuttgart: Frommann-Holzboog Verlag.

Saxenian, A.-L. (1994). *Regional Advantage*. Cambridge: Harvard University Press.

van Dijk, M.P. (1995). Flexible Specialisation, The New Competition and Industrial Districts. *Small Business Economics* **7**, 15-27.

Vou, J.-I. & Wilkinson, F. (1994). Competition and Co-operation: Toward Understanding Industrial Districts. *Review of Political Economy* **63**, 259-278.

Weidlich, W. (1991). Physics and Social Science - the Approach of Synergetics. *Physics Reports* **204**, 1-163.

| class of industries | Number of industries in this class of industries | Industries not rejected by the Komolgorov-Smirnov test complete model | reduced model |
|---|---|---|---|
| Agriculture, hunting, forestry and fishing | 6 | 4 | 1 |
| Mining and quarrying | 4 | 4 | 0 |
| Food products, beverages and tobacco | 8 | 8 | 1 |
| Textiles and leather | 11 | 10 | 3 |
| Wood and wood products | 5 | 4 | 1 |
| Pulp, paper, publishing and printing | 4 | 4 | 3 |
| Chemicals, rubber, plastics and petroleum | 9 | 9 | 4 |
| Non-metallic, mineral products | 7 | 7 | 2 |
| Metals and metal products | 12 | 12 | 3 |
| Machinery and equipment | 7 | 7 | 4 |
| Electrical and optical equipment | 15 | 14 | 9 |
| Transport equipment | 8 | 8 | 5 |
| Manufacturing n.e.c. | 6 | 6 | 1 |
| Manufacturing total | 92 | 89 | 36 |
| Electricity, gas and water supply | 4 | 4 | 2 |
| Construction | 5 | 5 | 2 |
| Wholesale, retail trade and repair | 19 | 18 | 15 |
| Hotel and restaurants | 5 | 4 | 2 |
| Transport, etc. | 12 | 10 | 7 |
| Financial intermediation | 5 | 5 | 3 |
| Business activities, etc. | 22 | 20 | 19 |
| Education | 4 | 4 | 2 |
| Health and social work | 3 | 3 | 3 |
| Leisure activities and beauty services | 8 | 8 | 8 |
| Public administration and organisations | 7 | 5 | 6 |
| Service total | 94 | 86 | 69 |
| total | 196 | 183 | 106 |

**Table 1:** Number of industries in each class of industries and the number of industries for which the two models are not rejected by the Komolgorov-Smirnov test (significance level: 0.05).

| | | Range of parameter | |
|---|---|---|---|
| | | $\hat{C}_i$ | $\hat{\sigma}_i$ |
| A | Agriculture, etc. | 0.42-2.37 | 0.007-0.08 |
| Q | Mining and quarrying | 0.12-0.90 | 0.03-0.22 |
| F | Food products, etc. | 0.27-2.21 | 0.04-0.18 |
| T | Textiles and leather | 0.23-1.79 | 0.04-0.70 |
| W | Wood and products | 0.21-2.22 | 0.01-0.45 |
| P | Paper, etc. | 0.55-1.12 | 0.30-0.59 |
| K | Chemicals, etc. | 0.32-1.59 | 0.04-0.51 |
| N | Non-metallic products | 0.55-1.67 | 0.04-0.33 |
| M | Metals and products | 0.24-2.22 | 0.04-0.54 |
| Y | Machinery and equipment | 0.61-1.35 | 0.04-0.40 |
| O | Electrical and optical equip. | 0.22-1.09 | 0.05-0.59 |
| V | Transport equipment | 0.51-1.64 | 0.04-0.41 |
| U | Manufacturing n.e.c. | 0.87-1.15 | 0.11-0.50 |
| | Manufacturing total | 0.13-2.22 | 0.01-0.70 |
| G | El., gas and water supply | 0.51-1.88 | 0.11-0.34 |
| C | Construction | 0.30-41.0 | 0.05-0.21 |
| S | Wholesale, etc. | 0.27-30.6 | 0.05-0.39 |
| R | Hotel and restaurants | 0.31-14.5 | 0.02-0.15 |
| X | Transport, storage and communication | 0.20-0.63 | 0.04-0.68 |
| I | Financial intermediation | 0.27-0.93 | 0.04-0.60 |
| B | Real estate, renting and business services | 0.37-1.68 | 0.05-0.70 |
| E | Education | 0.44-9.81 | 0.04-0.40 |
| H | Health and social work | 0.35-2.53 | 0.08-0.21 |
| L | Leisure activities and beauty services | 0.33-1.61 | 0.08-0.40 |
| Z | Public administration and organisations | 0.29-2.60 | 0.06-0.75 |
| | Service total | 0.20-41.0 | 0.02-0.75 |
| | total | 0.12-41.0 | 0.007-0.75 |

**Table 2:** Classes of industries and the ranges for parameters $\hat{C}_i$ and $\hat{\sigma}_i$ that are obtained for these classes.

| | | Range of parameter | | | |
|---|---|---|---|---|---|
| | | $\hat{\nu}_i$ | $\hat{\rho}_i$ | $\hat{\phi}_i$ | $\hat{\eta}_i$ |
| A | Agriculture, etc. | 1.03-2.43 | 0.14-0.49 | 0.02-0.34 | 0.05-0.85 |
| Q | Mining and quarrying | 1.01-2.50 | 0.29-0.56 | 0.01-0.68 | 0.11-1.01 |
| F | Food products, etc. | 0.43-2.40 | 0.26-0.62 | 0.03-0.87 | 0.11-0.96 |
| T | Textiles and leather | 0.52-2.32 | 0.20-0.75 | 0.0002-1.16 | 0.15-1.71 |
| W | Wood and products | 0.67-4.87 | 0.14-0.38 | 0.05-5.44 | 0.01-0.58 |
| P | Paper, etc. | 0.89-3.72 | 0.17-0.45 | 0.01-2.46 | 0.30-0.65 |
| H | Chemicals, etc. | 0.87-2.52 | 0.20-0.87 | 0.01-1.06 | 0.36-1.07 |
| N | Non-metallic products | 0.18-2.20 | 0.22-1.43 | 0.006-0.94 | 0.12-2.48 |
| M | Metals and products | 0.46-2.92 | 0.19-0.59 | 0.009-1.08 | 0.23-1.49 |
| Y | Machinery and equipment | 0.95-3.14 | 0.24-1.56 | 0.005-1.45 | 0.42-1.77 |
| E | Electrical and optical equip. | 0.44-2.91 | 0.22-1.06 | 0.007-0.42 | 0.10-3.09 |
| V | Transport equipment | 0.78-2.04 | 0.20-0.62 | 0.01-1.21 | 0.19-1.12 |
| O | Manufacturing n.e.c. | 0.62-1.57 | 0.33-0.50 | 0.009-0.58 | 0.48-0.64 |
| | Manufacturing total | 0.18-4.87 | 0.14-1.56 | 0.0002-5.44 | 0.01-3.09 |
| G | El., gas and water supply | 0.82-1.95 | 0.21-0.87 | 0.008-0.75 | 0.75-1.26 |
| C | Construction | 1.45-1.93 | 0.23-0.25 | 0.01-0.08 | 0.13-0.79 |
| S | Wholesale, etc. | 0.48-3.21 | 0.13-0.61 | 0.02-0.70 | 0.09-0.83 |
| R | Hotel and restaurants | 1.02-2.50 | 0.25-0.35 | 0.10-0.80 | 0.40-0.57 |
| X | Transport, etc. | 0.72-2.43 | 0.17-0.63 | 0.01-0.42 | 0.22-2.18 |
| I | Financial intermediation | 1.62-1.83 | 0.19-0.27 | 0.01-0.17 | 0.17-0.92 |
| B | Business activities, etc. | 1.10-3.79 | 0.12-0.59 | 0.008-1.64 | 0.14-0.92 |
| E | Education | 0.49-2.40 | 0.24-0.50 | 0.05-0.60 | 0.32-0.76 |
| H | Health and social work | 1.47-3.06 | 0.16-0.32 | 0.08-0.79 | 0.26-0.49 |
| L | Leisure activities and beauty services | 1.03-3.18 | 0.15-0.57 | 0.01-0.57 | 0.22-0.72 |
| Z | Public administration and organisations | 1.29-2.75 | 0.15-0.32 | 0.0006-0.55 | 0.11-1.27 |
| | Service total | 0.48-3.79 | 0.12-0.87 | 0.0006-1.64 | 0.11-2.18 |
| | total | 0.18-4.87 | 0.12-1.56 | 0.0002-5.44 | 0.01-3.09 |

**Table 3:** Classes of industries and the ranges for the parameters $\hat{\nu}_i$, $\hat{\rho}_i$, $\hat{\phi}_i$ and $\hat{\eta}_i$ that are obtained for these classes.