Papers in Evolutionary Economic Geography

08.20

Surviving in agglomerations:
Plant evolution and the changing benefits of the local environment

Frank Neffke, Martin Svensson Henning and Ron Boschma

Utrecht University
Urban & Regional research centre Utrecht

Surviving in agglomerations: plant evolution and the changing benefits of the local environment

Frank Neffke[α], Martin Svensson Henning[ß] and Ron Boschma[γ]

Version 20081218.

Abstract

Cities vary with regard to the characteristics of their economic life. A formal model by Duranton and Puga (2001) suggests a division of labour between diversified and specialized cities. Diversified cities (the "nursery cities") provide a fertile environment for search and innovation. Specialized cities, by contrast, are better equipped to facilitate mass-production. In essence, this spurs firms to re-locate as they mature from the exploratory set-up stage to mass-production. In this article, we empirically test the assumptions behind this model by means of survival analysis using Swedish plant level data of over 11 000 plants. More specifically, we investigate the effects of local specialization and local diversity on plant survival at different ages of a plant and for different size categories of plants.

Not all types of local diversity will be of value to a plant. Rather, we expect plants to benefit especially from local diversity in *related* industries. In a similar vein, cities with a large concentration of a broad range of activities in related trades may confer larger benefits than cities with a narrow specialization in the plant's own industry. To quantify the degree of relatedness between industries, we use a new measure, Revealed Relatedness. This serves to identify technological relatedness by measuring economies of scope as implied by the structure of production portfolios of plants.

The findings suggest that regional characteristics strongly influence the chances of a plant to survive. In general, the hypothesized specialization effects are only found when we look at related specialization. Large plants at high stages of maturity form the only exception to this. However, diversity effects are only visible when we take all local diversity into account, not just diversity in related industries. Moreover, it is only young firms that benefit from regional diversity. This indicates that the "nursery city" metaphor holds as much for small, prototype plants as for large mass-production plants.

[α] Urban and Regional Research Centre Utrecht (URU), Section of Economic Geography, Utrecht University, The Netherlands/ Department of Applied Economics, Erasmus School of Economics, Rotterdam, The Netherlands. E-mail: neffke@few.eur.nl

[ß] Department of Social and Economic Geography, Lund University, Sweden. E-mail: martin@keg.lu.se

[γ] Urban and Regional Research Centre Utrecht (URU), Section of Economic Geography, Utrecht University, The Netherlands. E-mail: r.boschma@geo.uu.nl

1. Introduction

Cities are the nuclei of the productive systems of our economies. Not only do people concentrate there. Also most economic activity can be found in or close to urban centers. Moreover, an often reported research finding is that firms in particular industries tend to locate closely together. In the agglomeration externalities literature, this clustering behavior has commonly been ascribed to the Marshallian triplet of localization externalities: labour market pooling, input-output linkages and specialized knowledge spillovers. On the other hand, scholars have also argued that firms benefit from a large diversity in the industrial base of the city in which they are located. As this was famously described in the works of Jane Jacobs, such benefits have come to be known as Jacobs' externalities. However, as cities grow larger, congestion effects become more prominent. It is therefore reasonable to assume that firms experience a trade-off between local diversity and local specialization. After all, the only way a city can be both diversified and, at the same time, accumulate a large concentration of individual industries is to grow very large.

In empirical research, authors have argued that the nature of this trade-off may vary across industries. In a recent article, however, Duranton and Puga (2001), suggest a model in which the outcome of the trade-off depends not on the industry, but rather on the type of activity carried out in a firm. The authors distinguish between two types of activities: exploration and mass-production. The model consists of three core building blocks. First, a firm enters the market with a prototype product, but without a fully developed production technology. To find a suitable production process, it can engage in a trial and error search by imitating one of the locally used production technologies. This is the exploration stage. As soon as the optimal process is found, the firm can start mass-producing its product. Second, localization externalities accrue to a firm if they locate in cities with a high concentration of firms that use the same production process as the firm itself. Third, the overall level of economic activity in the city is a source of congestion effects.

The model incorporates some central beliefs about the production stages of firms and the nature of externalities. The first is that firms move from an exploration (search) stage to an exploitation (mass production) stage. The second assumption is that exploration is best conducted in diversified cities, aptly called "nursery cities". The third premise is that exploitation is best conducted in specialized cities.

In this article, we aim to test these assumptions by investigating plant survival rates. Using a technique originally developed in medical sciences, we determine the age dependence of agglomeration benefits for plant survival in a dataset covering almost 25000 manufacturing plants. Furthermore, as plants that are built for mass-production purposes should in general have a higher starting employment, we also split our sample in small, medium and large plants. In line with the nursery cities model, we find that Jacobs' externalities only contribute to plant survival in the first 15 years of a plants existence. Surprisingly, however, this holds as much for small plants as for large plants. As a matter of fact, the benefits small plants derive from their local environment over time are very similar to the benefits for large plants.

Apart from investigating the link between age dependence and agglomeration externalities, we also venture to overcome the strict dichotomy of specialized versus diversified economic environments. This distinction is not only found in Duranton and Puga's work, but also in the vast majority of the empirical papers on the topic. In the nursery cities model, exploring firms simply try out randomly a random production process they find in their city. However, in

reality, we would expect that firms that are rational enough to locate in diverse cities in order to benefit from knowledge spillovers, also are informed enough to employ a more sophisticated search strategy than just randomly testing any locally available production technology. We are even inclined to argue that firms can be expected to focus their research efforts on a limited set of production processes used in the production of goods related to their own. In such a case, firms would not benefit from just any kind of diversity in the city, but from the presence of plants engaged in technologically related, yet different activities. Similarly, localization externalities can be expected to derive not only from plants in the own industry, but also from plants engaged in technologically related, yet different activities.

To investigate if this is observed in reality, there is a need to specify the *relatedness* between industries in terms of production technology. We will do so by using a method that we have developed elsewhere (Neffke and Svensson Henning 2008). This method estimates relatedness between industries based on co-production patterns in the product portfolios of plants. This method enables us to add a go-between and overcome the artificial diversification/localization dichotomy by adding information about the level of relatedness of local industries.

The empirical consequences of this are startling. Adding a term that captures local concentration of related industrial activity to our regression abolishes any effect of pure (own industry) localization externalities. Apparently, plants do benefit far more from being located close to plants in related industries than being close to real competitors.

The outline of the article is as follows. In section 2, we provide an overview of the literature on agglomeration externalities and a more in depth treatment of the DP model. Section 3 describes the link between agglomeration externalities and plant survival. Here we also put forward some hypotheses to be tested. Section 4 covers a discussion of the data and a presentation of our estimation strategy. In section 5 we turn to the outcomes of the empirical analyses and the robustness checks. Section 6 summarizes and suggests areas for further research.

2. Theoretical background

In the contemporary literature, agglomeration externalities are often divided into two types: localization externalities and Jacobs' externalities.[1] A third type of externality is urbanization externalities. The latter capture the effects of city size. Big cities can often boast high quality amenities and infrastructure, but are also plagued by congestion effects, like pollution and high factor costs. As a consequence, urbanization externalities can just as well represent economies as diseconomies to local firms.

Localization economies refer to the situation where firms benefit from the local presence of other firms belonging to the same industry.[2] Building on Marshall (1920), localization economies result from a large pool of specialized labour, easy access to local supplier and

[1] A third type of externality that is sometimes described are so-called Porter externalities. These externalities derive from fierce local competition that spur innovative activities in a city (see Porter 1990). However, fierceness of competition would ideally be measured in terms of profit margins. Unfortunately, we do not have these data at our disposal and will therefore not consider these types of externalities..

[2] Some authors prefer the term MAR (Marshall-Arrow-Romer) externalities, appealing to a more long-lasting dynamic effect of local specialization. However, empirically, the distinction between static and dynamic agglomeration effects is very demanding in terms of data requirements. The variation in the data must allow an estimate of the precise lag structure of the regressors' effects. Making this distinction is therefore beyond the scope of this paper.

client firms and to unintended local knowledge spillovers in cities with a strong concentration of a particular industry. In formal models, localization economies are often modeled as a result of specialized intermediates producers that compete for the demand of a specific industry on a monopolistically competitive local market. This approach was pioneered by Fujita (1988).

Jacobs' externalities arise when firms benefit from a large number of different industries in the local economy. Most prominently, Jacobs (1969) argued that most innovations result from "adding new work to old work" in cities. The larger the local diversity of ideas, the more new combinations can arise from this. This led Glaeser et al. (1992) to coin the term Jacobs' externalities.[3] A formal treatment of this mechanism was accomplished in 2001 with the article by Duranton and Puga.

By now, a large body of literature has addressed the measurement of different types of agglomeration externalities. In many studies (e.g. Henderson et al. 1995, Henderson 1997, Combes et al. 2004) findings suggest an important role of localization economies. The role of Jacobs' externalities is less well established. However, they seem to be particularly important for young or technologically advanced industries (e.g. Henderson et al. 1995, Neffke et al. 2008). In a meta-analysis of the literature, however, De Groot et al. (2008) show that the findings on agglomeration externalities vary widely across studies. Moreover, the results may to a large extent depend on the industries and time period studied, the geographical area covered and on the specific estimation framework employed. For example, Combes (2000) findings vary greatly across industries, even within the same country and time period, in an economy wide analysis of French local industries.

In light of these divergent outcomes, it is an interesting suggestion that different activities at the *plant level* may benefit from different types of local environments. In the nursery cities model, firms enter the market with a prototype. Before being able to mass-produce their product, firms have to search for the optimal production process. This search is modeled as a trial and error process. Firms can try out any production process that they observe in their city. Although it is possible to relocate in order to discover more alternatives, this is costly. In diversified cities, firms can try out several processes without having to bear the cost of moving to another city. Therefore searching in diversified cities is cheaper compared to searching in specialized cities. Firms will prefer to conduct their exploration in these cities.

Localization economies are present in diversified as well as specialized cities. They are modeled as in Fujita (1988). Final goods producers source intermediates from a sector of intermediates producers. These intermediates can not be traded across cities. The firms in intermediates sectors are engaged in monopolistic competition . The output of each sector is only used in one specific final goods production process. Intermediates enter the final goods producers' production process as in a Dixit-Stiglitz production function. The larger the concentration of firms that make use of a particular set of intermediates, the higher the diversity in intermediates that can be sustained locally, and therefore the higher the localization externalities that accrue to the final goods producers.

To generate the localization effects of comparable size for each of the locally used production processes, diversified cities must grow larger than specialized cities. However, the overall size

[3] As larger cities are usually also more diversified, sometimes, urbanization and Jacobs' externalities are not treated as separate effects. However, in this study, we follow the convention that Jacobs externalities refer to a city with a high degree of diversification, *controlling* for overall city size.

of a city comes at the cost of higher congestion. As soon as firms find the optimal production process, they can upscale their production volumes. Reaching this stage, they do no longer benefit from local diversity. At this moment, firms are faced with a predicament. Their location in big cities now damages them in terms of high congestion costs, but without any economies in exploration to compensate. Therefore, these firms are drawn to specialized cities that strike a better balance between localization, urbanization and Jacobs externalities for them.

The upshot of this is that diversified cities and specialized cities may co-exist, with each focusing on a different task in the economy.[4] More specifically, one may expect that plants that are set-up as small prototype plants should, especially in the early years of their existence, benefit from being located in diversified cities. In contrast, plants that are set-up as large-scale mass-production plants should not benefit at all from Jacobs' externalities, but rather benefit from specialized local environments. In other words, small plants that are still very young are expected to benefit from a large local diversity, while large and already quite mature are better off in a specialized local environment.

However, the way in which firms are proposed to engage in technological exploration appears to be too stylized to be helpful in empirical work. If firms are supposed to be intelligent enough to locate in diversified cities in order to benefit from Jacobs' externalities, surely they must have a more sophisticated research strategy than randomly testing each locally available production process. More realistically, we would expect firms to limit their search to production processes used in related economic activities. This would indicate the importance of something similar to the concept of related variety, introduced by Frenken et al. (2007). A similar kind of reasoning applies to localization externalities. It is not very realistic to expect only plants in the own industry to give rise to localization externalities. One could even argue that plants in the same industry are likely to generate lower knowledge spillovers as they would rather try to prevent any knowledge to leak to their competitors. Plants in related industries, in contrast, may be a source of ideas that are relevant, yet new to an industry. At the same time, firms in related industries may be less protective about knowledge spillovers. Nevertheless, we do not know of any study that addresses this issue.

A main empirical challenge is the measurement of relatedness between different industries. Most existing indicators are either rather ad hoc, like those that assume that two industries are related if they are close to each other in the Standard Industry Classification (SIC) system, or they are biased towards technology intensive industries, like patent based measures. What is needed is a manufacturing wide measure that assesses the degree of relatedness in the production processes used in different industries. In this paper we use a novel approach that does exactly this. In essence, it takes plant portfolios as an expression of the existence of economies of scope. More specifically, the fact that one plant produces products belonging to two different industries is interpreted as an indication of relatedness of the production technologies employed in those industries. Using a database on the product portfolios of a large sample of Swedish manufacturing plants, we arrive at a matrix containing relatedness estimates for the vast majority of industry combinations. The precise method is more involved and is developed in detail in Neffke and Svensson Henning (2008). This measure allows us to estimate the impact of the local concentration of *related* industries on the performance of

[4] This division of labour between cities had already been anticipated in less formal studies. As mentioned before, Jacobs (1969) regards large diversified cities as the breeding ground of new ideas. In the product life cycle location model of Hirsch (1967), the suitability of the national production environment varies with the stages of the product life cycle.

plants. We will refer to the effects of related concentration as *related localization externalities*, as opposed to *pure localization externalities*, which we will use to indicate the effects of a local concentration of activity in the own industry. Similarly, we will make a distinction between *related Jacobs' externalities* and *pure Jacobs' externalities*. Related Jacobs' externalities count the number of related industries with a significant presence in a city, whereas pure Jacobs' externalities count the number of all significant industries in a city, regardless of whether they are related or unrelated. These two types of Jacobs' externalities capture, as it were, the qualitative differences in the local industrial structure, in the sense that they are intended to measure *how many different knowledge fields* are represented in a city. The both types of localization externalities focus on the *concentration* of related and same industrial activity in a city.

3. Estimation framework

In the agglomeration externalities literature, most studies use regional employment growth or regional employment levels (e.g. Glaeser et al. 1992, Henderson et al. 1995). However, a decline in employment does not always result from a decline in productivity and, therefore is not necessarily related to weaker agglomeration externalities. A good example is labour saving investments or a market where demand is relatively inelastic. Under these circumstances, higher productivity simply means that fewer employees produce the same output and, instead of rising, employment may even drop. Plant productivity data (e.g. Henderson, 2003) or plant entry data (Rosenthal and Strange 2003) are then probably more appealing.

Lack of capital data inhibits us to investigate the impacts of agglomeration externalities in line with these approaches. We would have to make the uncomfortable assumption that the capital-labour ratio is constant across plants. Instead, we study agglomeration externalities by estimating their effect on the survival rates of plants. There are preciously few papers in agglomeration externalities that focus on survival rates. Some exceptions are Falck (2007) and Boschma and Wenting (2007). This is surprising because the fact that a plant survives, is a crude, yet very significant performance measure. It is not as volatile as yearly productivity figures, because plants may and build up buffers in years of good fortune and draw upon their reserves in years of bad fortune. Moreover, a survival analysis also explicitly accounts for plants that exit and can therefore be regarded as an interesting complementary approach to the existing analyses in the literature. We choose the plant as unit of analysis, as plants, have one physical location, whereas firms may have different plants in various cities. Such multiple locations severely complicate the task of determining what the relevant local environment is.

Although it is not very common in externality studies, survival analysis has been widely used in the field of industrial dynamics and business studies. Most of these studies have looked at survival of firms or plants with respect to their size and age (Disney et al 2003), pre-entry experience (Thompson 2005), the structure of the market (Buenstorf 2007, Cantner et al 2006), the maturity of the industry (Agarwal and Gort 2002), or combinations of these dimensions (Klepper 2002). A common finding in this literature is that the larger the plant or firm, the lower its survival rate. Moreover, older plants tend to show lower hazard rates. In this paper, we therefore control for these influences. Size is measured by the starting employment of firms. The influence of age is removed from the data using the semi-parametric specification of the Cox model which allows for an unspecified relation between survival and age. To control for market structure and industry effects we use industry dummies.

In some agglomeration externalities studies, corporate and non-affiliated establishments have proven to show differences that may impact the results of our study (e.g. Henderson 2003; Rosenthal and Strange 2003). Henderson argues that corporate plants may draw fewer benefits from the local environment. The reason for this is that plants belonging to larger corporations can use their channels within this corporation to access knowledge and organize supplier and client relations. In a total factor productivity study of American plants, he finds that corporate plants indeed experience lower agglomeration externalities compared to non-affiliated plants. They may therefore be less inclined to engage in local interaction than their non-affiliated counterparts. This suggests that corporate plants are fundamentally different with respect to their externality needs compared to non-affiliated plants. In our analyses, we take this into account by splitting the sample into a corporate and a non-affiliated part.

As we are not interested in the precise shape of the hazard rate with respect to plant age, but rather wish to focus on the externality variables, we choose a Cox proportional hazard model (henceforth referred to as the Cox model). Let $\theta(t, X_i)$ be the hazard rate for a plant of age t with characteristics X_i. The Cox specification now results in:

$$\theta(t, X_i) = \theta_0(t)\exp(\beta' X_i) \tag{1}$$

$\theta_0(t)$ is a function that represents the baseline hazard and will not be specified. A common finding is that the larger the plant, the lower its hazard rate. Therefore, X_i must at least contain a measure of the size of the plant. Moreover, X_i also contains variables describing the local environment at the time of the plant's birth. We discuss these variables in greater detail in the next section.

An important prerequisite for using the Cox model is that the effect of regressors is the same for plants of all ages. This is obviously violated by our prediction that the effect of Jacobs' externalities diminishes as plants grow older. In fact, one might say that the violation of the proportional hazards assumption lies at the heart of our research questions. To solve this issue, we use a method outlined in Hosmer and Royston (2002). This method, originally developed for applications in medical statistics, uses the Aalen linear hazard (Aalen) model as a guide on how to incorporate age-dependent effects in a Cox model.

The hazard function for an Aalen linear hazard model with p regressors is specified as follows:

$$h(t, X_i, \beta(t)) = \gamma_0(t) + \gamma_1(t)x_i^1 + ... + \gamma_p(t)x_i^p \tag{2}$$

In other words, the effect of regressors on the hazard rate is additive, and may vary with plant age, t. The cumulative hazard function associated with the model is:

$$H(t, X_i, B(t)) = \sum_{k=0}^{p} x_i^k G_k(t) \tag{3}$$

Instead of estimating the individual $\gamma_k(t)$'s, it is easier to calculate the cumulative regression coefficients, $G_k(t)$. To get an impression of the time-dependence of the effect of regressor k,

we plot $G_k(t)$ against plant age, t. In such a plot, a proportional hazard in the k^{th} variable should result in a straight line for all values of t. A violation of the proportional hazards assumption would lead to a plot where the slope changes with t. From an inspection of the plots, it is possible to derive the functional shape of the time-dependence in the regressor under scrutiny. The only functional shapes we consider in this model are step functions. At the end of this procedure, we feed the information on the time-dependence back into the Cox model:

$$\theta(t, X_i) = \theta_0(t)\exp\left(\sum_{k=1}^{p}\beta_k(t)x_i^k\right) \quad (4)$$

Where:

$$\beta_k(t) = \alpha_r^k I(A_r^k, t), \text{ with } I(A,t) = \begin{cases} 1 \text{ if } t \in A \\ 0 \text{ elsewhere} \end{cases} \quad (5)$$

4. Data

For the empirical investigation, we use data on Swedish manufacturing plants that were collected by Statistics Sweden. The dataset contains information on about 15,000 individual plants that entered between the years 1970 and 2004. The total number of active plants in this period is around 25,000. Between 1968 and 1989, the sample covers Swedish manufacturing plants with five employees or more that are engaged in manufacturing activities. After that (1990-2004), the data collection regime changes to cover plants with more than 5 employees belonging to *firms* employing at least 10 people. In other words, from 1990 and onwards, the only plants below 10 employees are corporate plants. For the sake of consistency, we focus in the largest part of this text on non-affiliated plants that are larger than 9 employees. We use the complete sample for robustness exercises.

We know to which of Sweden's 278 municipalities each plant in the dataset belongs.[6] The data have been cleaned and checked, both manually and using tailor made algorithms.[7] To avoid problems with left-censored plants, we only study the survival spells of plants that *entered* the database after 1970.[8] We distinguish between corporate and non-affiliated plants using information on the organization identification number. Plants that do not share their organization number with any other plant are called non-affiliated. All other plants are

[5] As we limit ourselves to step functions, the function of age enters multiplicatively in the term between the large brackets. This allows us to transform the regressor values and estimate a Cox model with time varying regressors.

[6] We have merged a few municipalities in order to create consistent definitions over time.

[7] Available on request.

[8] We were able to use a plant identification variable to follow plants over the course of their existence. From 1984 and onwards, the identification variable that had been used in the 1970s was gradually abandoned in favor of a new identification system. Using the years for which both the new and the old identifications numbers were available, we were able to create a consistent identification code for the vast majority of plants. Still, exit rates in 1983 and birth rates in 1984 were slightly higher than expected. We therefore also dropped the spells belonging to plants that entered in 1984 or exited in 1983. In the construction of the variables describing the local environment, we however also picked up contributions from these plants.

corporate plants.[9] All characteristics of a plant are measured at the year in which the plant enters the database. We assume this is also the year in which the plant was created.[10]

In order to measure the effect of agglomeration externalities, we have to define what we mean by the local environment of a plant more carefully. As noted above, for each plant we know in which municipality it is located. However, Swedish municipalities vary enormously in size. In the vast and scarcely populated north, municipalities can cover many thousands of square kilometers. In the much more densely populated south, by contrast, municipalities are limited to a far smaller area (sometimes only small parts of metropolitan areas). Moreover, surely, a municipality that is located at a short distance from the center of the capital city of Stockholm should experience some of the agglomeration externalities that are generated there. To cope with these issues, we determine the position of the largest population core – the largest village, town or city – for each municipality. Typically, there is one clear "municipality capital" surrounded by a couple of smaller villages. Next, we assume that all economic activity takes place in this municipality capital. In general, agglomeration effects should attenuate gradually over distance. Therefore, we base our agglomeration indices on quantities that are generalizations of the well-known population potential. For example, the employment potential of industry *i* in municipality *m* and year *y* is calculated as:

$$E_{i,m,y}^{pot} = \sum_{m' \in M} \left[g(d_{mm'}) \sum_{p \in P_{i,m',y}} E_{p,y} \right] \tag{6}$$

Where

$E_{p,y}$: the employment of plant p in year y,

$P_{i,m,y}$: the set of plants active in industry i and located in municipality m in year y,

M: the set of municipalities in Sweden, and

$g(d_{mm'})$: a function that expresses the attenuation over the road distance between the capitals of m and m' in kilometers, $d_{mm'}$.[11]

We proxy pure localization externalities that accrue to a plant in industry *i* that is located in municipality *m* and was founded in year *y* by the natural logarithm[12] of employment potential[13] of the industry minus the plant's own employment:

---

[9] The use of the organizational number is not unproblematic. Firms may change organizational numbers, and use more than one organizational number for administrative or legal reasons. The separation between corporate and non-affiliated plants by use or organization ID is however the most exact that these data allow.

[10] Given the fact that we only have data on plants of at least five employees, this is obviously only an approximation.

[11] We use the following expression for this attenuation:

$$g(d_{mm'}) = \begin{cases} \exp\left[\frac{\ln 0.01}{90}(d_{mm'} - 10)\right] & if\ d_{mm'} > 10 \\ 1 & if\ d_{mm'} \leq 10 \end{cases}$$

This results in an attenuation for which the contributions of municipalities that are less than 10 km away are counted fully. At longer distances, the distance decay is exponential with parameters such that the employment in a municipality at 100 km contributes for 1% to the overall employment potential of m.

[12] The log-transformation is chosen because it seems reasonable that hazard rates should react to proportional transformation rather than to additive transformations of our regressors.

[13] Henderson (2003) argues that localization effects are best measured by looking at the number of plants in a local industry. The rationale behind this is that each plant represents a different experiment, which constitutes a variation on the industry's production process. Following this line of reasoning, we measure pure and related

(7) $$L_p = E^{pot}_{i,m,y} - E_{p,y}$$

Related localization externalities are defined in an analogous way:

(8) $$RL_p = \sum_{i' \in R_i} E^{pot}_{i',m,y}$$

Where R_i represents the set of industries that are related to industry i, but excludes *i* itself.

To decide which industries should be a part of this set, we make use of the Revealed Relatedness index developed in Neffke and Svensson Henning (2008).[14,15] As explained above, this index of Revealed Relatedness captures the intensity of the plant level economies of scope between industries that are manifest in the product portfolios of manufacturing plants covered by a large Swedish database. Two industries are said to be related if they have a relatedness index of at least 0.14.[16] The maximum of the relatedness index is 1, which would suggest that the production processes used in the involved industries are virtually indistinguishable.

Jacobs' externalities should capture the number of different production processes that are used locally. As we cannot observe production processes, we count the number of industries with a significant local presence instead. We say that an industry has a significant local presence if its employment potential or, following Henderson's argumentation, its number-of-plants potential exceeds a certain threshold. The employment based value is calculated as follows:

(9) $$JAC_{m,y} = \sum_{i \in I} \Xi\left(E^{pot}_{i,m,y}, \xi\right)$$

Where: $\Xi(E,\xi) = \begin{cases} 1 \; if \; E \geq \xi \\ 0 \; if \; E < \xi \end{cases}$ and I is the set of all industries in Sweden.

Similarly, by only counting the number of related industries with a significant local presence, it is possible to construct an indicator of local related diversity:

(10) $$RJAC_{i,m,y} = \sum_{i' \in R_i} \Xi\left(E^{pot}_{i',m,y}, \xi\right)$$

To capture congestion effects we also calculate a population potential for each municipality capital using the procedure in (6).

---

localization externalities using number-of-plants potentials instead of employment potentials. We use number-of-plant based measures in our robustness checks at the end of this paper.

[14] In principle, this matrix of relatedness indices may change over time. However, here we choose the average relatedness between industries in the period 1971-2002, which most closely corresponds to our sampling period. This avoids fluctuations in the set of related industries over time.

[15] We have been able to calculate the Revealed Relatedness index for almost all industry pairs, except for the ones that involved industries with very few plants. Plants in these industries have been ignored in this study.

[16] This number corresponds to a choice of the 3,000 strongest links that were measured between industries. At this level, only very few isolated industries exist and most industries are related to some other industry. Leave out all plants for which no related industries exist, this ensures our sample does not become unnecessarily small. The choice for exactly 3000 links is therefore reasonable but, admittedly, ad hoc.

5. Empirical results

Descriptives and specification details

Tables 1 and 2 contain some general descriptive statistics for our datasets. The entire sample consists of about 14,700 observations (plants). However, if we drop all plants smaller than 10 employees to obtain a consistent sampling definition over time, the number of investigated plants decreases to about 11,500. These very small plants seem to be randomly distributed across the country. The correlation between a dummy representing a size between 5 and 9 employees and each of our agglomeration indicators is always lower than 5%.[17] In what follows, we will focus on the results that were obtained when leaving out all plants under 10 employees.[18] Estimations using the complete sample are not presented. However, to assess the robustness of our outcomes, we have run all analyses using the broader sample as well. We comment on these outcomes where needed.

The set restricted to plants larger than 9 employees contains 2706 corporate and 8829 non-affiliated plants.

Table 1 about here.

The cross-correlations between the regressors are shown in Table 2. Given the sample size, we are not too concerned of multi-collinearity effects. Moreover, any such problems should turn up in our robustness checks. PLANTSIZE is the size of the plant at the time it entered the database. LOC, RLOC, JAC and RJAC are as defined in section four and refer to the values in the foundation year of the plant. As we discussed before, both, the localization and the Jacobs' externalities variables, can be based either on local employment potentials or on local number-of-plants potentials. However, the correlation between pairs of employment and number-of-plants based variables is typically above 95%. This prohibits the estimation of both types of specifications simultaneously. In the main text we calculate the LOC and RLOC indicators using employment potentials. For the JAC and RJAC indicators we use the number-of-plants potentials. This follows the spirit of the nursery cities model most closely. Here, localization externalities arise from a large local demand for intermediates specific to the particular production process used in the industry. Localization externalities in the nursery cities model depend therefore on the overall demand generated by the local industry. This will be correlated with total local employment in the industry. Jacobs' externalities, on the other hand, are associated with the notion of between-industry knowledge spillovers. The more different industries that are active in a city, the greater the variety of locally available production processes will be. However, for these production processes to be sufficiently visible, the industries involved must be of a certain minimum size. Therefore, for Jacobs' externalities, we have to choose a lower limit for when a local presence of an industry is labeled "significant". As each plant in a particular industry represents one experiment with the production process associated with that industry, the number of plants can be regarded as the knowledge spillover potential of the industry.[19] As it is this knowledge spillover potential that

[17] Obviously, this does not hold for the control variable PLANTSIZE. Furthermore, please note that all our regressors contain a log transformation.

[18] Note that the plants below the employment threshold of 10 have been used in the construction of the agglomeration variables. Leaving such plants out would increase the measurement error in these variables. As there is no substantial correlation between the size of a plant and its local environment, we do not believe this procedure leads to spurious results. This belief is confirmed by regressions with agglomeration variables that only use information from plants of at least 10 employees. Outcomes are very similar, but standard errors are somewhat larger.

[19] See Henderson (2003).

also generates Jacobs' externalities, in the main text we use a local number-of-plant potential to decide whether an industry is significantly present in a city. We set this lower limit at least five plants. Later we use different cut-offs, both number-of-plant based and employment based, to assess the robustness of our findings.

Table 2 about here.

Interpretation of the regression tables

In all regressions below, we control for initial plant employment size, and for differences in hazard rates across industries by adding 3-digit industry dummies. The dummies are added to all regression analyses in this article. All variables have been log transformed. This implies that we believe that the effects enter as "hazard elasticities": a δ% increase in the k^{th} variable is supposed to raise the instantaneous probability of exit by $\delta \cdot \beta_p$. This has the attractive property that increasing the population of a town of 10,000 inhabitants by 1,000 has a different effect from adding 1,000 inhabitants to a city with a population of a million. The regression tables report untransformed coefficients with their robust[20] standard errors in brackets. This means that a negative coefficient is associated with a positive effect on survival. In the text, we often discuss findings in terms of hazard ratios. These express the change in hazard rate that is associated with a given increase in the regressor value.

Outcomes

Table 3 summarizes the results of the Cox regressions that assume that the influence of the local environment does not change with the age of a plant. Column (1) uses the observations on both corporate and non-affiliated plants. We start with the traditional set of agglomeration externalities and our control variable for plant size. The effect of plant employment (PLANTSIZE) is strong and has the expected negative sign. A doubling of the initial employment of a plant results in a reduction of the hazard ratio by 14.5%.[21] All three local environment variables have the expected sign. A large population (POPPOT) increases the risk for a plant to exit, with a hazard ratio of about 1.25 for a doubling of the population. Localization externalities are very small and not significant. Doubling the number of active local industries (JAC) is associated with a hazard ratio of about 0.86. Jacobs' externalities therefore raise a plant's survival probability.

In column (2), we add the related localization and related Jacobs' indicators. A large local concentration of *related* industries (RLOC) turns out to significantly contribute to a plant's survival probability, whereas *pure* localization externalities do not have any impact. Moreover, while *pure* Jacobs' effects are significant and beneficial for survival, a large local variety of *related* industries (RJAC) does not matter. However, a problem may be that the related Jacobs' indicator is measuring the diversity in a subset of industries compared to the pure Jacobs' indicator. This means that pure Jacobs' indicator is always at least as large as the related Jacobs' indicator. As a consequence, it may be hard to distinguish between Jacobs'

[20] We control for clustering of residuals on plant ID. Controlling for clustering on municipality or industry yields very similar standard errors.

[21] In fact, the influence of ln(PLANTSIZE) is strongly non-linear, with a decreasing effect for higher values of PLANTSIZE. However, using a non-linear specification does not affect any of the other coefficients and therefore we stick to the simpler linear specification. In later regressions, this non-linearity is expressed in the different coefficients we get for plants of different sizes.

and related Jacobs' effects. We will therefore drop the related Jacobs' externalities indicator from further investigations.[22]

To check whether there are any differences between corporate and non-affiliated plants, we split the sample in two parts. Column (3) is based on the sample of corporate plants and column (4) on the sample of non-affiliated plants. The results are strikingly different. The impact of pure Jacobs' externalities we found in column (2) can be wholly attributed to the non-affiliated sample. By contrast, related localization externalities are strongest in the corporate sample with a point estimate that is four times as large as the one in the sample of non-affiliated plants. The negative effects of a large local population, are again less pronounced in the corporate plants.

Overall, these outcomes suggest that corporate plants indeed interact in different ways with the local environment compared to non-affiliated plants. However, it is not simply the case that corporate plants are isolated from their surroundings. Rather, they seem to have a smaller capacity to exploit the inter-industry knowledge spillovers associated with Jacobs' externalities. As, in the context of the nursery cities model, we are especially interested in the dynamics of Jacobs' externalities we will leave out all corporate plants in the analyses from this point onward.

One problem with the estimations we carried out thus far is that the assumption of proportional hazard rates is violated as indicated by the chi-squared test statistics in the lower part of table 3. In part, this may be caused by a non-constant effect of the industry dummies. However, the outcomes are not only driven by the inclusion of these dummies.[23] Table 4 shows the results of a proportional hazard test with regressor specific test statistics when industry dummies are excluded. The test statistics indicate that the main problem is caused by pure Jacobs' externalities. To investigate the proportional hazards issue further, we make use of the graphs of the Aalen LH model. Figures 1 to 5 contain the graphs of the cumulative regression coefficient for each variable in our model. Each graph shows how the year-on-year compound effect of a covariate on the survival hazard (y-axis) varies with the age of the plants (x-axis). The slope of the graph at a specific age indicates the instantaneous effect of the covariate on plants of a particular age. Each graph contains a solid line representing point estimates over and two dotted lines corresponding to a 95% confidence interval. A practical matter is that for higher ages, the number of plants that are at risk of exiting become very low. As a consequence, in this part of the graphs, the Aalen coefficients are based on only a small number of observations. This makes the Aalen graphs very volatile beyond, say, the age of 25. For this reason, we do not attach much value to the shape of the graphs after this point.

As discussed in section three, of the effect of a regressor changes with age, this results in a departure from a straight line in the Aalen graphs. Starting with the ln(PLANTSIZE) variable, we find a downward sloping line up to the age of 19. This suggests that over this period, the size of the plant has a positive effect on a plant's survival rate. After the age of 19, the line is more or less horizontal, indicating that for mature plants, the size at birth is no longer relevant. Similarly, we can find changes in slopes for ln(LOC), ln(JAC) and ln(POPPOT). RLOC does not seem to undergo any significant changes in slope. On the basis of this visual

[22] In LOC and RLOC, this problem does not arise, as the sets of plants over which both are calculated are mutually exclusive.

[23] Point estimates of significant coefficients shift only marginally (less than 25%) if industry dummies are omitted. However, the effect of pure localization externalities turns significant in column (1). In column (2) and (3) p-values drop below standard significance limits but rise again to the 5% level in column (4).

inspection of the Aalen graphs, we decide to allow the coefficients to change at the following ages:

- ln(PLANTSIZE): 19 years
- ln(LOC): 16 years
- ln(RLOC): no changes
- ln(JAC): 15 years
- ln(POPPOT): 20 years

Table 5 shows the outcomes of a Cox regression with this time-dependence structure specified. Column (1) of table 5 is a repetition of column (4) in table 3 with RJAC omitted. The same regression but now with slopes that are allowed to change at the plant ages specified above is reported on in Column (2). In this specification, the proportional hazards assumption is still violated. However, now this can be wholly attributed to non-proportionalities in the industry dummies, which are not of immediate interest here.[24] Indeed, we find that some of the slopes change substantially with age. Most interesting is the finding that Jacobs' externalities (JAC) improve survival chances only in the early years of a plant's existence. In later years, the point estimate is positive (indicating increased failure rates), but insignificant. A Wald test on equality of slopes shows this difference is significant at levels below 0.1%. The effect of initial employment is only significant in young plants as well. However, here the Wald test comparing young to mature plants gives a p-value of 8%. The population potential (POPPOT) has a strong and significant negative effect on the survival of young plants and no significant effect in mature plants. The difference in slopes is however not significant. If we turn to localization externalities, related localization externalities are modelled as age invariant and turn out to have the usual positive and significant effect on survival rates. The point estimate is very close to the baseline estimates of column (1). Pure localization externalities are never significant.

In the analysis of time dependence, we have to face the possibility that changes in coefficients over time are an artefact of our decision to measure the size of regressors at the time the plant enters our database. If the local environment changes over time, these initial conditions may be less informative for the agglomeration externalities a plant experiences at later ages. This would result in an artificial weakening of the observed externality effects. To investigate this possibility, we also run our analyses with covariates that change over time.
The general story remains the same. There are no significant localization externalities and strong related localization externalities. Pure Jacobs' externalities contribute positively in the early years of a plant and are insignificant later on. The significance of urbanization effects show the same temporal pattern, but are harmful for a plant's survival in younger stages. The main difference with our previous findings is that plant employment now has a strongly increasing positive influence of the survival rates of plants. This is not surprising, as plants are less likely to be closed down at the time they have many employees. A minor difference is that point estimates of related localization externalities drop slightly, and are insignificant if we take all plants into consideration. However, in the analysis of the subsamples of small and medium sized plants to which we turn next, their contribution is significant again. Another minor detail is that the drop in urbanization externalities with rising age now is significant according to our Wald tests.

[24] Outcomes without industry dummies are very similar to the ones shown here. In these estimations, the proportional hazards assumption is never violated at the 10% level.

Apart from studying variations over the age of a plant, we would like to distinguish somehow between mass-production plants and prototype plants. The difficulty is that we do not have any capital-labour ratios to base this distinction on. However, mass-production plants should in general be set up with a larger initial number of employees than prototype plants. Under this assumption the nursery cities model suggests that the coefficients of externalities indicators are different for plants of different sizes. To test this, we divide the sample into three parts, small plants (below 15 employees), medium-sized plants (15-24 employees) and large plants (over 24 employees).[25]

Columns (3) to (5) show the outcomes of regressions based on these sub-samples without age-dependent coefficients. Taking into consideration the standard errors, only the effect of PLANTSIZE is significantly different for plants of different sizes. The differences in the other coefficients can only be regarded as indicative. For example, we see pure localization externalities to rise with an increasing size of the plants and even become significant for the first time in our analyses for the largest plants in our sample. The diseconomies associated with large cities (POPPOT) seem to be more important in large and medium plants than in small plants. This is not entirely unexpected, as rents are higher in big cities.

Columns (6) to (8) contain the results of the analyses with age-dependent coefficients. As we split the sample into small, medium sized and large plants, the sample sizes we use in our regressions drop and results necessarily get less precise. The most striking finding is that pure Jacobs' externalities are significant only for young plants and not for mature plants. This holds for all subsamples and significantly so in the subsamples of medium sized and of large plants.[26] This suggests that all plants benefit from Jacobs' externalities in their early years only. Another interesting outcome is that pure localization externalities are strong and have a positive effect on survival only for large plants at a high age.

Robustness

In the analyses above, we have taken a number of rather ad hoc decisions. To assess the sensitivity of our outcomes to such decisions, we run a large number of alternative specifications. First, we alternate between specifications where localization externalities are measured in terms of number-of-plants potentials and specifications that use employment based indicators. Next, in the construction of JAC and RJAC, we use six different lower limits for the definition of what constitutes a "significant presence" of an industry, corresponding to a minimum of 1, 5 or 10 for the number-of-plants potential, and a minimum of 50, 100 or 250 for the employment potential. We also investigate the influence of omitting industry dummies. Finally, we rerun all regressions on the full sample, including plants under 10 employees. This results in 2*6*2*2=48 different specifications for each regression analysis we have discussed so far. Table 6 and 7 summarize the outcomes of this exercise. The black numbers in the first row of table 6 and 7 represent the percentage of analyses that yield the same sign as the corresponding columns in table 3 and 5. Below these numbers, in grey, we report the percentage of times the outcome was also significant at the 5% level. For example, the entry for ln(EMP) in table 6, column (4) means that 100% of all different robustness specifications yield the negative sign we reported already in table 3, and 100% of all specifications yield both a negative sign and were significant at at least the 5% level. The last eight rows of table 6

[25] As the vast majority of our plants is very small for a sufficiently large sample to remain, we cannot investigate the behavior of very large plants with any reasonable precision.

[26] If we had included the (undersampled) plants with a starting employment between 5 and 9 employees, this result would also hold for the subsample of small plants.

give the percentage of times at which the Wald test for significant changes in slopes was significant at the 5% (black) or at the 10% level (grey).

To a very large degree, the signs and significance of the results of the alternative specifications match the ones we present in the main text. In table 6, the only important departures from our main results are found in coefficients that are not significant in table 3. For example, the RJAC estimates vary relatively widely across specifications. Turning to table 7, we find again primarily corroborations of the reported findings. Signs are for the overwhelming majority of coefficients the same as in table 5. As before, the main differences between specifications are found when variables are not significant. Also the outcomes of the Wald statistics confirm our findings in table 5: low p-values in table 5 correspond to high numbers of rejections of equal coefficients, whereas high p-values usually result in zero percent of such rejections. The overall conclusion we can draw from this exercise is that the main results we reported above are robust to changes in specification.

Discussion

Our estimations show that the local environment has a strong influence on the survival chances of plants. Cities with large populations are associated with high failure rates. This negative impact of population potential is particularly strong for large plants. This is in line with expectations, as plants with many employees need to rent large buildings. Rents are typically very high in big cities. These congestion effects seem to more than nullify any positive impacts of being located close to large markets. This holds for plants of all sizes. However, at more mature stages, the net negative effect of urbanization externalities becomes insignificant.

The distinction between *pure localization* effects generated by nearby plants in the own industry and *related localization* effects which are associated with a large local concentration of employment in related industries, proved fruitful. In general, only related localization effects generate significant benefits in terms of higher survival chances. The effects of pure localization externalities, by contrast, are not significant. The only exception can be found in large plants at a high age. Here, pure localization externalities are associated with extraordinarily large increases in the survival chances of a plant. Large plants may benefit from long-term relationships with local client and supplier firms that are facilitated by a strong local embeddedness in very specialized regions. As these benefits take a long time to materialize, they may indeed only be available to plants at a rather mature stage. The benefits of pure localization effects in large and mature plants remind us of the non-urban industrial complexes of Iammarino and McCann (2006). The nature of (pure) localization externalities in such hierarchical regional complexes can be expected to be very different from the (related) localization externalities in small-scale industrial districts that have been widely studied in economic geography (Asheim 2000).

In the light of the nursery cities model, the most important finding is that only young plants benefit from pure Jacobs' externalities. If young plants engage more in exploration activities than old plants, this can be seen as supporting evidence for the model. However, this age-dependence of Jacobs' externalities is present in plants of all different size classes. This suggests that the role of Duranton and Puga's nursery cities is not limited to small exploratory plants but extends as well to larger plants.

6. Conclusion

In this article, we set out to investigate (1) how agglomeration externalities impact the survival of plants, (2) how the nature of these impacts changes with the evolving maturity of a plant, and (3) how presence of related industries affects the survival of a plant. Overall, we can identify two interesting sets of findings that may proof fruitful ground for further research efforts research.

First, the assumptions of the DP model about changing impacts of agglomeration externalities with plant evolution are by and large empirically justified. The impacts of Jacobs' externalities do indeed change with plant age. Nonetheless, contrary to the expectations, if anything, Jacobs' externalities vary even more with plant age for the medium sized and large plants (that are more likely to be sites of mass-production), than for the small plants (which we would typically associate with exploratory prototype plants). The 'nursery' role of diversified cities is apparently not limited to the prototype development stage, as the DP model suggests. The evidence strongly suggests that all types of plants benefit from local diversity in the years right after they were set up. In these formative stages of a plant's life, economic diversity may be helpful in overcoming all kinds of teething problems. An interesting fact in this matter is that corporate plants do not seem to benefit from local diversity at all. This is not entirely surprising, as corporate plants should be able to solve their start-up problems by drawing on the experience and connections of the mother firm.

Although industry dummies were used in this paper to capture differences across industries in survival chances, we did not specifically study the differences in externality effects across industries. However, the impacts of externalities may change as much with the rising maturity of plants as with the evolution of an industry. For example the findings in Neffke et al. (2008) show that the industry life cycle framework (Klepper 1997), where industries move from exploratory to mature stages, can be used as an alternative in the context of agglomeration externalities. The industry life cycle framework allows for a richer conceptualization of technological change than does the DP model. However, such inquiries are beyond the scope and purpose of this article.

The second main conclusion of the article is that the strict demarcation of pure localization and pure Jacobs' externalities appears to be too artificial. It is useful also to consider externalities that arise in regional environments with many firms in related, but different, activities. According to these arguments, the most important local sources of knowledge a plant can draw on arise from the plants that are engaged in activities that are not precisely the same, yet related to, the own activity. Such plants can be a source of ideas that are different, but still close enough to be relevant. Allowing for such effects severely complicates the picture of the economy, as industries get intertwined in an intricate way. However, it also allows for a deeper understanding of how inter-industry linkages really work. For example, in this paper only related localization externalities show a persistent positive effect on survival chances across all different types of plants. The traditionally much more investigated pure localization effects, by contrast, seem to benefit survival of only a small subset of large and mature plants.

Where this article used a technological relatedness approach, further research may focus on different types of relatedness, such as commonalities in use of labour skills or linkages along the value chain. This should give complementary insights in the relatedness structure of economies.

The concept of agglomeration externalities is an elusive concept, both theoretically and empirically. They are far from being stable across different types of plants, and probably also far from being stable across different industries. Moreover, the fact that industries are not isolated units, but rather related in networks of intertwined activities, is a clear outcome of our analysis of externalities of related agglomeration. But the heterogeneity that seems to be an inherent feature of externalities should not be regarded as problematic. Rather, it should be taken as a challenge for researchers to learn about the ways in which the local interplay between industries takes place over time. As the availability of longitudinal micro data that cover long time periods gradually increases, we will be able to come closer to a thorough understanding of the true nature of varying agglomeration externalities.

References

Agarwal, R and M Gort (2002). Firm and Product Life Cycles and Firm Survival. *The American Economic Review* 92(2):184-190.

Asheim B T (2000). Industrial Districts: The Contributions of Marshall and Beyond. In: G L Clark, M P Feldman and M S Gertler. *The Oxford Handbook of Economic Geography*. Oxford: Oxford University Press.

Bernard A B and J Bradford Jensen (2007). Firm structure, multinationals, and manufacturing plant deaths. *The Review of Economic and Statistics* 89(2):193-204.

Boschma R and R Wenting (2007). The spatial evolution of the British automobile industry. Does location matter? *Industrial and Corporate Change* 16(2):213-238.

Buenstorf G (2007). Evolution on the Shoulders of Giants: Entrepreneurship and Firm Survival in the German Laser Industry. *Review of Industrial Organization* 30:179-202.

Cantner U, K Dreßler and J J Krüger (2006). Firm survival in the German automobile industry. *Empirica* 33:49-60.

Combes P-P (2000). Economic Structure and Local Growth: France, 1984-1993. *Journal of Urban Economics* 47:329-355.

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

De Groot H F, J Poot and M J Smit (2008). *Agglomeration Externalities, Innovation and Regional Growth: Theoretical Perspectives and Meta-Analysis.* Working Papers in Economics 08/01: University of Waikato, Department of Economics.

Disney R, J Haskel and Y Heden (2003). Entry, exit and establishment survival in UK manufacturing. *The Journal of Industrial Economics* LI (1): 91-112.

Duranton G and D Puga (2001). Nursery Cities: Urban Diversity, Process Innovation, and the Life Cycle of Products. *The American Economic Review* 91 (5):1454-1477.

Falck O (2007). Survival chances of new businesses: do regional conditions matter? *Applied Economics* 39:2039-2048.

Frenken K, F van Oort and T Verburg (2007). Related Variety, Unrelated Variety and Regional Economic Growth. *Regional Studies* 41(5):685-697.

Fujita M (1988). A monopolistic competition model of spatial agglomeration. *Regional Science and Urban Economics* 18:87-124.

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

Henderson J V (2003). Marshall´s Scale Economies. *Journal of Urban Economics* 53:1-28.

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

Henderson, J V (1997). Externalities and industrial development. *Journal of Urban Economics* 42: 449-470.

Hirsch S (1967). Location *of Industry and International Competitiveness*. Oxford: Clarendon Press.

Hosmer D W and P Royston (2002). Using Aalen's linear hazards model to investigate time-varying effects in the proportional hazards regression model. *The Stata Journal* 2(4):331-350.

Iammarino S and P McCann (2006). The structure and evolution of industrial clusters: Transactions, technology and knowledge spillovers. *Research Policy* 35(7):1018-1036.

Jacobs J (1969). *The Economy of Cities*. New York: Vintage.

Klepper S (1997). Industry Life Cycles. *Industrial and Corporate Change* 6(1):145-181.

Klepper S (2002). Firm Survival and the Evolution of Oligopoly. *The RAND Journal of Economics* 33(1):37-61.

Marshall A (1920). *Principles of Economics.* London: Macmillan. Original edition 1890.

Neffke F and M Svensson Henning (2008). *Revealed Relatedness: Mapping Industry Space*. Forthcoming in PEEG WP Series.

Neffke F, M Svensson Henning, R Boschma, K-J Lundquist and L-O Olander (2008). *Who Needs Agglomeration? Varying Agglomeration Externalities and the Industry Life Cycle*. Papers in Evolutionary Economic Geography #08.08, Urban and Regional Research Centre Utrecht, Utrecht University.

Porter M E (1990). *The Competitive Advantage of Nations*. London: Macmillan.

Rosenthal S S and W C Strange (2003). Geography, industrial organization, and agglomeration. *The Review of Economics and Statistics* 85(2):377-393.

Thompson, P (2005). Selection and firm survival: evidence from the shipbuilding industry, 1825-1914. *The Review of Economics and Statistics* 87(1): 26-36.

Table 1: variable descriptives.

	Obs	Mean	Std. Dev.	Min	Max
CORP	11535	.2345904	.4237609	0	1
ln(PLANTSIZE)	11535	3.037502	.8437525	2.302585	8.933532
ln(LOC)	11501	4.498378	3.489024	-33.96421	9.939873
ln(RLOC)	11535	7.347622	2.26436	-55.37755	10.83688
ln(JAC)	11535	1.998868	1.084041	0	4.043051
ln(RJAC)	11535	.956228	.8681275	0	3.218876
ln(POPPOT)	11535	11.46406	1.235645	7.656619	13.81222

Table 2: cross-correlations between the regressors.

	ln(PLANTSIZE)	ln(LOC)	ln(RLOC)	ln(JAC)	ln(RJAC)	ln(POPPOT)
ln(PLANTSIZE)	1.0000					
ln(LOC)	0.1004	1.0000				
ln(RLOC)	0.0602	0.4367	1.0000			
ln(JAC)	0.0381	0.4123	0.5087	1.0000		
ln(RJAC)	0.0405	0.3950	0.6543	0.7130	1.0000	
ln(POPPOT)	0.0596	0.4011	0.5111	0.8959	0.6409	1.0000

Variables are as defined in section 3.

Table 3: results of Cox regression where we assume that the influence of the local environment does not change with the age of the plant.

	(1) all	(2) all	(3) CORP	(4) NON-AFFILIATED
ln(PLANTSIZE)	-0.157***	-0.157***	-0.251***	-0.177***
	(0.0154)	(0.0154)	(0.0253)	(0.0235)
ln(LOC)	-0.00622	-0.00268	0.0160	-0.00774
	(0.00400)	(0.00409)	(0.00914)	(0.00471)
ln(JAC)	-0.151***	-0.143***	-0.0207	-0.198***
	(0.0249)	(0.0286)	(0.0626)	(0.0328)
ln(POPPOT)	0.222***	0.236***	0.173**	0.250***
	(0.0226)	(0.0228)	(0.0528)	(0.0255)
ln(RLOC)		-0.0354***	-0.0879***	-0.0234**
		(0.00658)	(0.0180)	(0.00775)
ln(RJAC)		0.0169	0.00355	0.0545
		(0.0256)	(0.0525)	(0.0299)
Industry dummies	yes	yes	yes	yes
Model statistics				
df PH stat	31	33	32	33
PH stat	38.81	40.34	52.42	47.44
Nobs	11501	11501	2698	8803
Log-likelihood	-57434.7	-57424.3	-11285.6	-42426.0

Clustered (on plant identification numbers) standard errors in parentheses),
* p<0.05, **p<0.01, *** p<0.001
Variables are as defined in section 3.
PH statistic is chi-squared distributed under the null-hypothesis of proportional hazards.
All estimations include 3-digit industry dummies.

Table 4: test of Proportional Hazards assumption without industry dummies.

	rho	chi2	df	Prob>chi2
PLANTSIZE	0.00498	0.14	1	0.7115
LOC	-0.00098	0.00	1	0.9466
RLOC	-0.00358	0.06	1	0.8126
JAC	0.02674	3.48	1	0.0623
POPPOT	-0.00512	0.13	1	0.7187
Global Test		10.95	5	0.0523

PH statistic is chi-squared distributed under the null-hypothesis of proportional hazards.
Variables are as defined in section 3.

Table 5: results of Cox regression with age-varying coefficients.

	(1) all	(2) all age-var.	(3) small	(4) medium	(5) large	(6) small age-var.	(7) medium age-var.	(8) large age-var.
ln(PLANTSIZE)	-0.177***		-0.632***	0.0703	-0.0979*			
	(0.0235)		(0.158)	(0.171)	(0.0481)			
ln(LOC)	-0.00767		-0.00712	-0.00583	-0.0183*			
	(0.00472)		(0.00645)	(0.00948)	(0.00909)			
ln(RLOC)	-0.0214**	-0.0220**	-0.0191	-0.0326***	-0.0143	-0.0439***	-0.0327***	-0.0147
	(0.00783)	(0.00777)	(0.0132)	(0.00852)	(0.0180)	(0.0108)	(0.00845)	(0.0181)
ln(JAC)	-0.167***		-0.143***	-0.233***	-0.151*			
	(0.0282)		(0.0389)	(0.0544)	(0.0643)			
ln(POPPOT)	0.249***		0.198***	0.327***	0.294***			
	(0.0255)		(0.0345)	(0.0491)	(0.0613)			
Age-varying variables								
ln(PLANTSIZE) early		-0.185***				-0.143***	0.0517	-0.0993*
		(0.0243)				(0.0182)	(0.175)	(0.0500)
ln(PLANTSIZE) late		-0.00615				-0.170	0.441	-0.0498
		(0.0994)				(0.0932)	(0.987)	(0.183)
ln(LOC) early		-0.00668				-0.000948	-0.00485	-0.0116
		(0.00471)				(0.00537)	(0.00903)	(0.0101)
ln(LOC) late		-0.0240				0.0161	-0.0131	-0.0860***
		(0.0170)				(0.0225)	(0.0272)	(0.0199)
ln(JAC) early		-0.185***				-0.0896**	-0.256***	-0.193**
		(0.0284)				(0.0327)	(0.0550)	(0.0656)
ln(JAC) late		0.0548				-0.120	0.0292	0.162
		(0.0643)				(0.0775)	(0.119)	(0.123)
ln(POPPOT) early		0.253***				0.189***	0.336***	0.303***
		(0.0257)				(0.0289)	(0.0495)	(0.0619)
ln(POPPOT) late		0.170				0.173	0.0513	0.224
		(0.0882)				(0.110)	(0.170)	(0.157)
Industry dummies	yes	yes	yes	yes	yes	yes	yes	yes
Model statistics								
df PH stat	32	36	31	32	32	36	36	36
PH stat	47.57	42.63	27.49	26.86	23.41	39.54	26.34	18.13
Nobs	12378	12378	5980	3436	2962	9572	3436	2962
Log-likelihood	-42427.4	-42418.2	-20719.7	-9903.9	-6695.1	-34868.1	-9900.6	-6689.1
Wald test for change of slopes: p-values								
PLANTSIZE		0.0802				0.783	0.701	0.796
LOC		0.298				0.459	0.747	0.000693
JAC		0.000114				0.679	0.0122	0.00237
POPPOT		0.339				0.878	0.0876	0.596

Clustered (on plant identification numbers) standard errors in parentheses), * p<0.05, **p<0.01, *** p<0.001

Variables are as defined in section 3.

In *age varying* , coefficients are allowed to change slope at the age of: 19 (PLANTSIZE), 16 (LOC), 15 (JAC), 20 (POPPOT).

PH statistic is chi-squared distributed under the null-hypothesis of proportional hazards.

All estimations include 3-digit industry dummies.

Table 6: results of robustness test of table 3.

	(1) all	(2) all	(3) CORP	(4) NON-AFFILIATED
EMP	**100,0%**	**100,0%**	**100,0%**	**100,0%**
	100,0%	100,0%	100,0%	100,0%
LOC	**100,0%**	**93,8%**	**100,0%**	**100,0%**
	75,0%	4,2%	37,5%	68,8%
RLOC		**100,0%**	**95,8%**	**100,0%**
		95,8%	85,4%	79,2%
JAC	**91,7%**	**83,3%**	**35,4%**	**85,4%**
	75,0%	60,4%	0,0%	29,2%
RJAC		**29,2%**	**18,8%**	**50,0%**
		6,3%	10,4%	10,4%
POPPOT	**100,0%**	**100,0%**	**100,0%**	**100,0%**
	100,0%	100,0%	100,0%	100,0%

The table contains the % of times the same sign is obtained as in table 3.
The second (grey) number contains the % of times the sign is the same as in table 3, and the outcome is also significant

Table 7: results of robustness test of table 5.

	(1) all	(2) all age-var.	(3) small	(4) medium	(5) large	(6) small age-var.	(7) medium age-var.	(8) large age-var.
EMP	100,0%		100,0%	100,0%	100,0%			
	100,0%		100,0%	0,0%	66,7%			
LOC	100,0%		100,0%	100,0%	100,0%			
	75,0%		45,8%	0,0%	33,3%			
RLOC	100,0%		100,0%	100,0%	100,0%			
	93,8%		81,3%	91,7%	0,0%			
JAC	83,3%		83,3%	75,0%	83,3%			
	27,1%		66,7%	50,0%	25,0%			
POPPOT	100,0%		100,0%	100,0%	100,0%			
	100,0%		95,8%	100,0%	100,0%			
EMP_E		100,0%				100,0%	100,0%	100,0%
		100,0%				100,0%	0,0%	37,5%
EMP_L		70,8%				100,0%	100,0%	100,0%
		0,0%				0,0%	0,0%	0,0%
LOC_E		100,0%				100,0%	100,0%	100,0%
		66,7%				47,9%	0,0%	0,0%
LOC_L		100,0%				22,9%	100,0%	100,0%
		0,0%				0,0%	0,0%	95,8%
RLOC_E		100,0%				100,0%	100,0%	100,0%
		93,8%				81,3%	91,7%	0,0%
JAC_E		91,7%				89,6%	83,3%	100,0%
		75,0%				66,7%	58,3%	54,2%
JAC_L		95,8%				43,8%	100,0%	100,0%
		37,5%				0,0%	16,7%	25,0%
POPPOT_E		100,0%				100,0%	100,0%	100,0%
		100,0%				95,8%	100,0%	100,0%
POPPOT_L		100,0%				100,0%	29,2%	100,0%
		18,8%				18,8%	0,0%	0,0%
EMP_el		87,5%				0,0%	0,0%	0,0%
		50,0%				0,0%	0,0%	0,0%
LOC_el		0,0%				0,0%	0,0%	87,5%
		0,0%				0,0%	0,0%	75,0%
JAC_el		100,0%				0,0%	83,3%	70,8%
		100,0%				0,0%	83,3%	62,5%
POPPOT_el		0,0%				0,0%	41,7%	0,0%
		0,0%				0,0%	0,0%	0,0%

The table contains the % of times the same sign is obtained as in table 5.
The second (grey) number contains the % of times the sign is the same as in table 3, and the outcome is also significant.
The rows for the tests on equality of slope counts the % of time significantly different at 5% (first) and at 10% (second row).

Figures

Figure 1: Aalen LH model: cumulative regression coefficient for ln(PLANTSIZE).

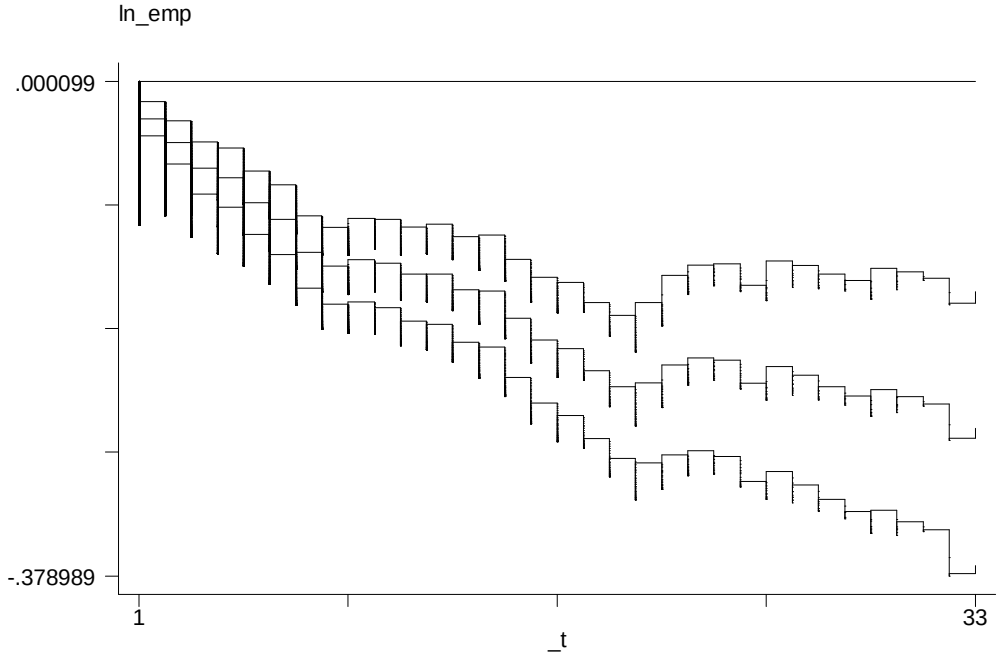

Figure 2: Aalen LH model: cumulative regression coefficient for LOC.

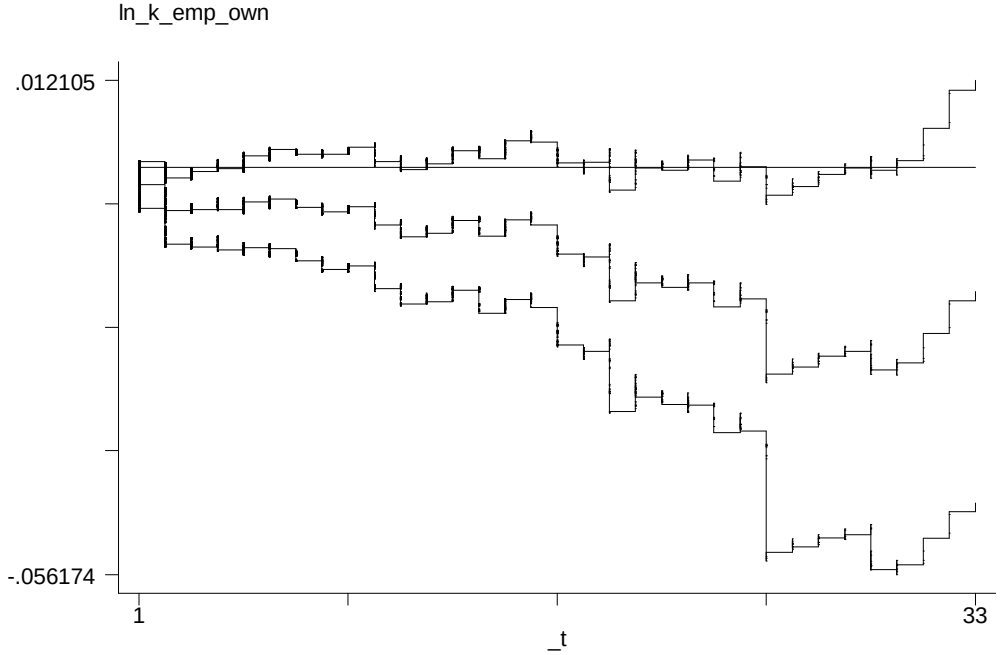

Figure 3: Aalen LH model: cumulative regression coefficient for RLOC.

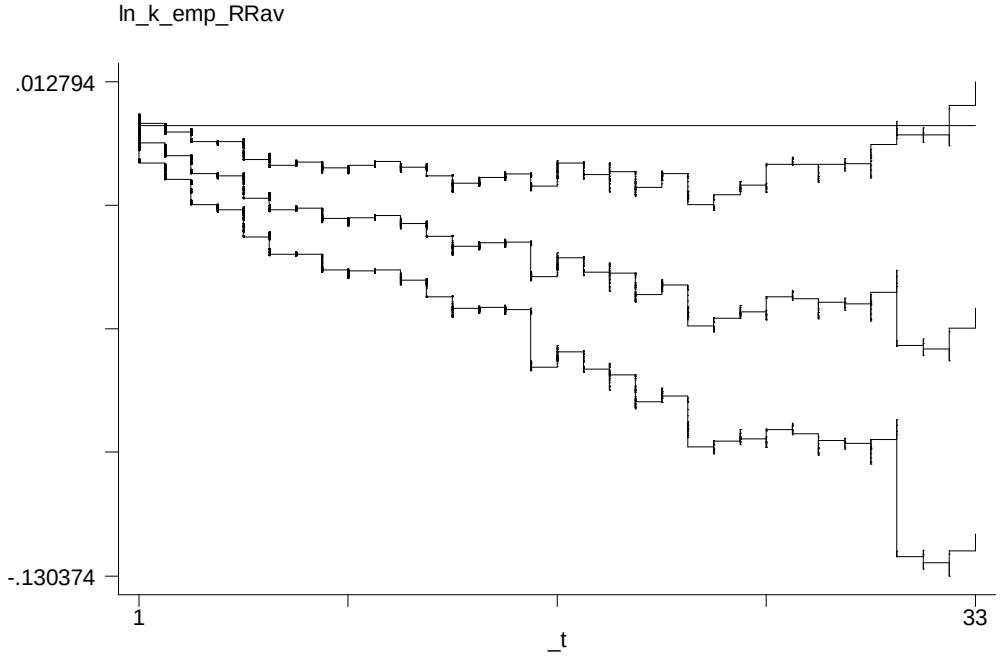

Figure 4: Aalen LH model: cumulative regression coefficient for JAC.

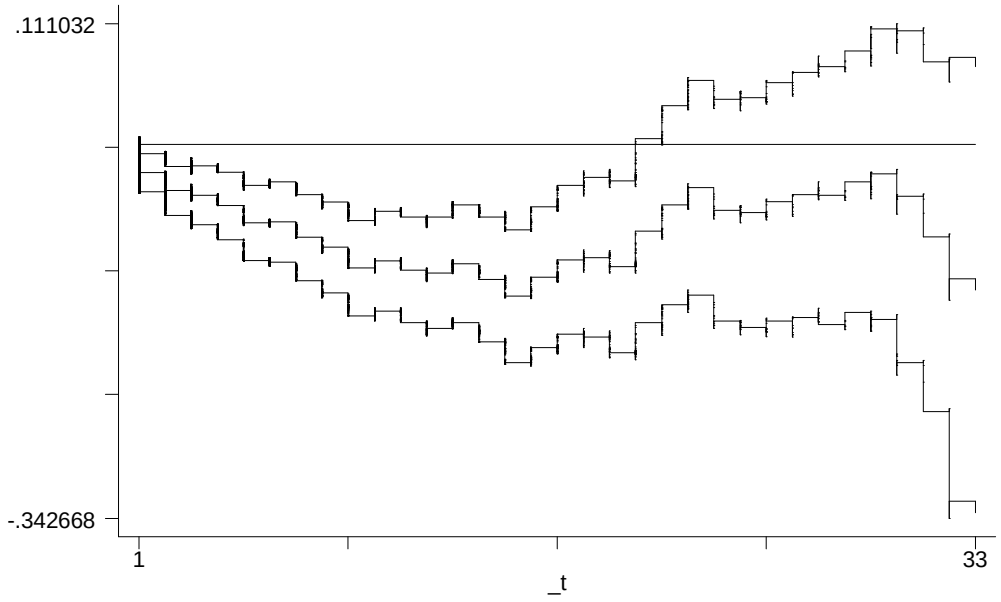

Figure 5: Aalen LH model: cumulative regression coefficient for ln(POPPOT).