# The Impact of Aging and Technological Relatedness on Agglomeration Externalities: A Survival Analysis

Frank Neffke (Department of Applied Economics, Erasmus University Rotterdam, The Netherlands)
Martin Henning (Department of Social and Economic Geography, Lund University, Sweden)
Ron Boschma (SERC, URU, Section of Economic Geography, Utrecht University, Utrecht, The Netherlands)

**November 2009**

# The Impact of Aging and Technological Relatedness on Agglomeration Externalities: A Survival Analysis

**Frank Neffke**[*]
**Martin Henning**[**]
**Ron Boschma**[***]

**November 2009**

\* Department of Applied Economics, Erasmus University Rotterdam, The Netherlands
\*\* Department of Social and Economic Geography, Lund University, Sweden
\*\*\* SERC, URU, Section of Economic Geography, Utrecht University, Utrecht, The Netherlands

Abstract

We study localization, urbanization, and Jacobs' externality effects on plant survival in Sweden (1970-2004). We focus on two questions: (1) do agglomeration externalities change with the age of plants? (2) using new information about the relatedness among industries, what is the role of technological relatedness among local industries? We find that agglomeration externalities affect survival chances of plants. This effect, however, differs between corporate and non-affiliated plants. Furthermore, we find that Jacobs' externalities benefit only young plants, whereas urbanization externalities harm plants at all ages. Localization externalities are insignificant, while the presence of related industries substantially increases survival rates.

# 1: Introduction

It is often argued that both industrial diversity and regional concentration of an industry are beneficial for the economic performance of local firms. Ideally, therefore, a city would host both a large number of different industries and show large concentrations of each of these industries. Unfortunately, such an ideal city would be extremely large and thus plagued with substantial congestion effects. Firms therefore experience a trade-off between local diversity and local specialization. Economic geographers have long suspected that this trade-off depends on the activity that is carried out in a plant. More specifically, it is often argued that diversified cities are important in the development of new technologies by promoting technological spillovers between industries, whereas specialized cities play a larger role in the exploitation of existing technologies, a suspicion that was formalized in a recent article by Duranton and Puga (2001). Moreover, framing the question of what constitutes the ideal industrial mix for a city as a trade-off between specialization and diversification overlooks the fact that many industries are linked to each other technologically. Therefore, the potential of inter-industry spillovers in a city may depend strongly on the technological linkages between the local industries.

Since the seminal article by Glaeser et al. (1992), the links between the geographical concentration of economic activity and the performance of local firms and industries have been studied extensively in the field of agglomeration externalities. Most authors investigate how the local environment affects such performance indicators as productivity or employment growth. By contrast, there are very few papers which study how different types of agglomeration externalities influence plants' survival rates. However, the fact that one plant survives while others do not is an important signal of the performance of a plant. To gain more insights into the relations between agglomeration externalities and plant survival, we focus on how three traditional types of agglomeration externalities – localization, Jacobs', and urbanization externalities – affect the survival rates of plants.

Our main concern is twofold. First, as young plants are more likely than old plants to use and further develop new technologies, we expect that the effect

of such externalities depends on the age of the plant. Therefore, one question we raise is how the impact of agglomeration externalities changes as plants grow older. Such an analysis requires age-dependent parameters that are not compatible with the proportional hazards models commonly used in survival analysis. Thus, to determine which agglomeration benefits prolong the lives of plants in different stages of their existence, we use a framework that combines the Aalen linear hazards model with Cox proportional hazards regressions. In contrast to common proportional hazards estimation, this methodology allows us to investigate the age dependence of parameters and therefore to assess how agglomeration externalities change with the age of a plant.

The second question we raise is how important a nearby presence and diversity of related industries are for the success of a plant. We therefore extend the traditional set of agglomeration externalities with two categories that capture the diversity and concentration of related economic activity, which we call related Jacobs' and related localization externalities, respectively. To determine which industries are related technologically, we apply a novel indicator developed by Neffke and Svensson Henning (2008) that extracts information on the relatedness between industries from co-production patterns in the product portfolios of plants. This incorporation of the technological structure of the economy into the study of agglomeration externalities constitutes the second important aspect of our article.

We conduct our analyses using a dataset covering almost 25,000 manufacturing plants in Sweden. In line with the Duranton and Puga's (2001) nursery cities model, we find that Jacobs' externalities only contribute to plant survival in the first fifteen years of a plant's existence. After this age, plants no longer benefit from being located in diversified cities. As mass-production plants are typically larger than prototype plants, we expected that larger plants might be set up from the outset as mass-production facilities. Surprisingly, however, these outcomes did not differ much when we focused only on small, medium, or large plants. As a matter of fact, the benefits that small plants derive from their local environment over time are very similar to the benefits for large plants.

Our findings also confirm the importance of related industries in the local economy. Although local related Jacobs' effects have no impact, adding a term that captures related localization externalities to our regression eclipses the effect of pure localization externalities. Apparently, plants benefit far more from being located close to plants in related industries than being close to real competitors.

The remainder of the article is structured as follows. In section 2, we discuss the theoretical background with an emphasis on the nursery cities model. Section 3 describes the empirical literature on survival analysis and agglomeration externalities. Section 4 presents the estimation framework. The data and construction of variables are discussed in section 5. In section 6, we turn to the outcomes of the empirical analyses and the robustness checks. Section 7 summarizes and concludes the article.


## 2: Theoretical background

Agglomeration externalities are costs or benefits that firms derive from being located close to other economic actors. In the contemporary literature, agglomeration externalities are often divided into three types: urbanization externalities, localization externalities, and Jacobs' externalities.[1] Urbanization externalities capture the effects of city size. Big cities often boast high quality amenities and infrastructure, but they are also plagued by congestion, resulting in pollution and high factor costs. As a consequence, urbanization externalities can just as well represent economies as diseconomies to local firms. Localization externalities are benefits that firms derive from the local presence of other firms belonging to the same industry.[2] Building on Marshall (1920), these benefits are thought to accrue

---

[1] As larger cities are usually also more diversified, urbanization and Jacobs' externalities are sometimes 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.

[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

from a large pool of specialized labor, from easy access to local supplier and client firms, and from local knowledge spillovers between firms in the same industry. Jacobs' externalities arise when firms benefit from the presence of a large number of different industries in the local economy. Jacobs (1969) argued that most innovations result from "adding new work to old" 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" to capture the inter-industry spillover benefits of local diversity.

By now, a large body of literature has studied the different types of agglomeration externalities. In many articles (*e.g.,* Henderson *et al.*, 1995; Henderson, 1997; Combes *et al.*, 2004), findings suggest that localization externalities play an important role. 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 the present article, inspired by Duranton and Puga's (2001) *nursery cities* model, we suggest that both the *industry level* and the *plant level* may determine which types of local environments generate the largest benefits.

According to Duranton and Puga (2001), firms enter the market with a prototype. Before being able to mass-produce their product, firms have to search for the optimal production process. The way they do this is by imitating locally available production technologies. The advantage of diversified cities is that firms can more easily imitate several different processes without relocating. Jacobs' externalities thus lower costs involved in technological search processes. Consequently, firms prefer to perform their exploration activities in these cities.

Localization externalities are modeled as in Fujita (1988) and are, in principle, present in diversified as well as specialized cities. However, to generate localization externalities for each of the locally used production

data must allow an estimate of the precise lag structure of the regressors' effects. Making this distinction is therefore beyond the scope of this chapter.

processes that are comparable in size to the ones in specialized cities, diversified cities must grow very large. This comes at the cost of substantial congestion effects. Once firms finish their exploration efforts and no longer benefit from local diversity, they face a predicament. Their current location in big cities imposes high congestion costs 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. The upshot of this is that diversified cities and specialized cities may co-exist, but diversified cities act as incubators for new firms (so-called nursery cities), whereas specialized cities are more adequate environments for mass-production.[3]

This model is very appealing in its simplicity. However, the way in which firms are proposed to engage in technological exploration appears to be too stylized for empirical work. If managers of 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 *related variety* or related diversity, as discussed by Frenken *et al.* (2007). Boschma and Iammarino (2009) show that the local economies of Italian regions indeed grow faster when their industries exhibit a high degree of coherence.

A similar kind of reasoning applies to localization externalities. It is not realistic to expect that only plants in the same industry 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 knowledge from leaking 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 such related industries may

---

[3] This division of labor between cities had already been anticipated in less formal studies. For instance, Jacobs (1969) regards large diversified cities as the breeding ground of new ideas. Similarly, 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.

be less protective about knowledge spillovers, as they are not direct competitors. In this way, the diversity and concentration in related industries may prove to be an important asset for regional economies. In an international trade context, Hidalgo *et al.* (2007) have shown that the technological position of the current export portfolio of a country vis-à-vis other product categories largely determines into which product categories a country can successfully diversify. Similarly, Neffke and Svensson Henning (2008) show that regions more readily diversify into industries that are related to their current local industries than into technologically unrelated industries.

## 3: Agglomeration externalities and survival analysis

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, it is not necessarily related to weaker agglomeration externalities. Good examples are labor saving investments or markets where demand is relatively inelastic. Under these circumstances, higher productivity simply means that fewer employees produce the same output, and employment may even drop instead of rising. Plant productivity data (*e.g.,* Henderson 2003) and plant entry data (Rosenthal and Strange, 2003) are thus more appealing. A disadvantage of such data is that they are very sensitive to cyclical economic movements. Moreover, of a more practical order, we lack data on capital stocks. We would therefore have to make the uncomfortable assumption that the capital-labor ratio is constant across plants. Instead, our investigations rely on survival analysis. As noted by Bernard and Jensen (2007), "plant shutdown is one of the few unambiguous observed signals of plant performance" (p. 193). In contrast to productivity or employment growth, plant survival is less volatile as it is less affected by short term economic shocks. Moreover, the relation between plant survival and agglomeration externalities has so far received only very scarce attention of the field.

There are preciously few papers on agglomeration externalities that focus on survival rates. This is surprising because, as argued above, the fact that a plant survives is a crude yet very significant performance measure. More importantly, 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. Some exceptions are Falck (2007) and Boschma and Wenting (2007). Both studies find that the local environment significantly affects survival. Falck finds that the number of new businesses in the same region and industry raise the survival chances of a new establishment. Boschma and Wenting conduct a survival analysis of the British automobile industry between 1895 and 1968. They find that the local presence of related industries, such as bicycles and coach making, has contributed significantly to the survival of the automobile producers. On a different account, Dumais et al. (2002) show that the relatively stable agglomeration patterns of industries is the result of a set of countervailing forces. On the one hand, new establishment formation and growth in existing establishments leads to a more even distribution of economic activity across space, whereas establishment closure leads to higher levels of concentration.

In some agglomeration externalities studies, corporate and non-affiliated establishments have proven to be affected differently by local factors (*e.g.,* Henderson, 2003; Rosenthal and Strange, 2003). In a total factor productivity study of American plants, Henderson finds that corporate plants do indeed experience lower agglomeration externalities than non-affiliated plants. The explanation the author provides is that plants belonging to larger corporations can use their channels within the corporation to access knowledge and organize supplier and client relations. Therefore, corporate plants may rely less on the local environment than non-affiliated plants do. This suggests that corporate plants are fundamentally different from non-affiliated plants with respect to their externality needs. In our analyses, we take this into account by splitting the sample into a corporate and a non-affiliated part.

Although it is not very widespread in externality studies, survival analysis has been used intensively in the fields 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 (Cantner *et al.,* 2006; Buenstorf, 2007), the maturity of the industry (Agarwal and Gort, 2002), or combinations of these dimensions (Klepper, 2002). Bernard and Jensen (2007) investigate the influence of a wide variety of plant and firm characteristics. Like Henderson, these authors stress the influence of corporate relations: they show that, after controlling for plant characteristics, plants that belong to multi-plant or multinational firms are more likely to close down. A robust finding in this literature is that the larger and older the plant, the higher its survival rate.

This article aims to deepen our understanding of the relationship between agglomeration externalities and plant survival. Our main research goals are threefold. First, we want to investigate the effect of agglomeration externalities on the survival rates of plants. Second, we want to assess how these effects depend on the age of a plant. Third, we want to find out whether related industries give rise to important agglomeration effects, both in terms of related diversity and of related localization effects.

## 4: Estimation framework

In this article, we base our estimations on the semi-parametric Cox proportional hazards model (Cox, 1972, henceforth, simply Cox model). This model is widely used in the analysis of survival spells. In the Cox model, we estimate the influence of covariates on the hazard rate of an individual, which in our study will be a plant. In an informal way, the hazard rate at age $t$ can be thought of as the rate at which plants exit, given that they have survived up to age $t$ (*e.g.,* Greene, 2000, pp. 937-950). In the Cox model, the hazard rate is specified as a function of the age of a plant and some plant characteristics. Let $\theta(t, X, \beta)$ be the hazard rate for a plant of age $t$ with $k$ different characteristics that are summarized in matrix $X$. $\beta$ is a vector of parameters. The Cox specification now results in:

(1)     $\theta(t, X, \beta) = \theta_0(t) \exp(\beta' X)$

$\theta_0(t)$ is an unspecified function that represents the baseline hazard, capturing the direct impact of plant age on plant survival. In our application, this is an attractive feature. Previous research has clearly shown that age is an important factor in determining the survival rate of plants. However, although we would like to know the effect of plant age on our agglomeration parameters, we are not interested in the relation between age and survival *per se*. Moreover, as this relation may not be a simple linear one, as shown in Falck (2007), a parameterization of this crucial variable might induce serious model misspecification. Instead, the coefficients in the Cox model are obtained from maximum partial likelihood estimation, which only uses information on the order in which plants exit, removing the importance of the exact time scale. A second advantage of the Cox model is that parameter estimates can be expressed in the intuitively convenient form of hazard ratios. For instance, a hazard ratio of 2 indicates that by increasing the corresponding variable by 1, the associated plants are being shut down at twice the rate compared to the baseline situation.

An important prerequisite for using the Cox model is that the effect of covariates 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 uses the Aalen linear hazard model (Aalen, 1989; henceforth, simply Aalen model) as a guide on how to incorporate age-dependent effects in a Cox model.

The hazard function for an Aalen model with $k$ covariates is quite different from the Cox model. It is not multiplicative but additive, and it is specified as follows:

$$(2) \qquad h(t, X, \gamma(t)) = \gamma_0(t) + \gamma_1(t)x_1 + \ldots + \gamma_k(t)x_k$$

As in the Cox model, the coefficients $\gamma_1$ to $\gamma_k$ link the change of the baseline hazard rate, $\gamma_0$, to a one unit change in the corresponding covariate. However, unlike the Cox model, the effects of the covariates now may be

different for different ages of a plant. We can derive the cumulative hazard rate by integration:[4]

$$(3) \quad H(t, X, \Gamma(t)) = \int_0^t \left[ \sum_{p=0}^{k} \gamma_p(u) x_p \right] du$$

$$= \sum_{p=0}^{k} x_p \left[ \int_0^t \gamma_p(u) du \right]$$

$$= \sum_{p=0}^{k} x_p \Gamma_p(t)$$

The $\Gamma_p(t)$'s are called cumulative regression coefficients. Instead of estimating the individual $\gamma_p(t)$'s, it is easier to calculate these cumulative regression coefficients (Hosmer and Lemeshow, 1999, pp. 338). The $\Gamma_p$'s should be regarded as empirical functions that describe the impact of the corresponding covariates. More specifically, the slope of these functions gives information about the influence of the covariate at a particular age (Aalen, 1989). If, at age $t$, the slope is positive, the covariate raises the hazard rate and is therefore associated with a negative effect on plant survival. Analogously, negative slopes indicate a positive effect on plant survival, and horizontal slopes suggest that the covariate has no impact on the survival of a plant.

To get an impression of the age-dependence of the effect of the $p^{\text{th}}$ covariate, we plot $\Gamma_p(t)$ against plant age, $t$. In such a plot, a proportional hazard in the $p^{\text{th}}$ variable should result in a straight line for all values of $t$. A violation of the proportional hazards assumption would lead to a plot with non-linearities and a slope that changes with $t$. From an inspection of the plots, it is possible to derive the functional form of the age-dependence in the covariate under scrutiny. For the sake of simplicity, we propose piecewise linear functional forms. In other words, we allow the slope to change at certain break points, but between two breakpoints it remains constant. At the end of this procedure, we feed the information on the age-dependence back into the Cox model of equation (1), but now the coefficients take

---

[4] In this notation we take $x_p$ to be a vector of ones.

different values for different sets of $t$, as indicated by the argument $t$ of the $\beta$'s: [5]

$$(4) \qquad \theta\big(t, X, \beta(t)\big) = \theta_0(t)\exp\left(\sum_{p=1}^{k}\beta_p(t)x_p\right)$$

As the inspection of the Aalen graphs will only lead to a maximum of one change of slope per variable, we can express $\beta_p(t)$ as:

$$(5) \qquad \beta_p(t) = \begin{cases} \beta_p^h & \text{if } t \in t_p^h \\ \beta_p^l & \text{if } t \in t_p^l \end{cases}$$

$t_p^l$ and $t_p^h$ are sets of $t$ that correspond to low and high values of $\beta_p(t)$ respectively.

## 5: Data

For the empirical investigations, we use data on Swedish manufacturing plants that were collected by Statistics Sweden.[6] The dataset contains employment and industry information for 25,000 individual plants that were active somewhere between 1970 and 2004. From 1970 to 1989, the sample covers Swedish manufacturing plants with five employees or more. In 1990, the data collection regime changes to cover plants with more than five employees belonging to *firms* employing at least ten people. Plants with fewer than ten employees are thus only reported if they are part of a larger firm. In other words, from 1990 and onwards, the only plants below ten employees are corporate plants. In this article, we mainly focus on non-affiliated plants with at least ten employees, but we use the complete sample for robustness exercises. As we do not know the age of plants that entered before 1971, we

---

[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] The data have been cleaned and checked, both manually and using tailor made algorithms. Detailed descriptions are available from the authors.

only study the survival spells of the about 11,500 plants that entered in 1971 or thereafter.[7]

In order to measure the effect of agglomeration externalities, we have to define what we mean by the local environment of a plant. In terms of geography, we know in which of Sweden's 277 municipalities a plant is located.[8] However, Swedish municipalities vary enormously in size. In the vast and scarcely populated north, municipalities can cover many thousands of square kilometers, whereas in the much more densely populated south, 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 centre of the capital city of Stockholm should experience some of the agglomeration externalities that are generated there. In fact, it is reasonable to assume that agglomeration effects attenuate gradually over distance.

To cope with these issues, we base our agglomeration externality indicators on a number of geographical potentials. This works as follows. In Swedish municipalities, typically there is one clear "capital" agglomeration, surrounded by a couple of smaller population cores. We first determine the position of these "municipality capitals". Next, we assume that all economic activity in a municipality takes place in its core. This reduces the geography of Sweden to 277 points. For each of these points, we calculate a number of 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:

---

[7] We were able to use a plant identification variable to follow plants over the course of their entire 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 externality variables describing the local environment, we however picked up contributions from plants regardless of their entry year.

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

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

Where:

$E_{\pi,y}$ :        the employment of plant $\pi$ 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

$g(d_{mm'})$ :        a function that expresses the attenuation over the distance over road between the capitals of $m$ and $m'$ in kilometers, $d_{mm'}$.[9]

As all variables are measured at the plant's birth, which we take to be the year that the plant enters our database, the subscript $y$ is equal to the entry year of a plant. In the robustness analysis, however, we also investigate what happens if it is taken into consideration that covariates change over time.

A main empirical challenge in this article is the measurement of relatedness between different industries. Most existing indicators are either *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. We need a manufacturing wide measure that assesses the degree of relatedness in the production processes used in different industries. In this article we use the Revealed Relatedness (RR) index that was developed by Neffke and Svensson Henning (2008). In this methodology, 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.

In essence, this RR index takes plant portfolios as an expression of the existence of production level economies of scope. The procedure consists of three steps. First, for each pair of industries, the number of co-occurrences (i.e., the number of plants that produce products from both industries at the same time) is counted. Next, this number is compared to a prediction of the number of co-occurrences based on some overall industry characteristics. This effectively corrects for the fact that some industries attract many co-occurrences because they are very profitable or simply very large compared to other industries. Using a database on the product portfolios of a large sample of manufacturing plants, Neffke and Svensson Henning arrive at a matrix containing relatedness estimates for the vast majority of industry combinations.[10,11] The maximum of the RR index is 1, which would indicate that the production processes used in the involved industries are virtually indistinguishable. In this study, we call two industries related if they have an RR index of at least 0.14, which corresponds to selecting the 3,000 strongest links in the matrix.[12] On the bases of this procedure, we can decide which industries are related to each other. This allows us to construct a new variant of both localization and Jacobs' externalities.

For the traditional localization externalities, we will use the term *pure localization externalities.* We proxy these pure localization externalities for a plant $\pi$ in industry $i$ located in municipality $m$ and founded in year $y$ by the employment potential of the industry minus the plant's own employment:

(7) $\qquad LOC_\pi = E_{i,m,y}^{pot} - E_{\pi,y}$

---

[10] The RR index was available 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 article.

[11] In principle, the RR matrix may change over time. However, to avoid fluctuations in the set of related industries over time, we use the average relatedness between industries in the period 1971-2002, which corresponds most closely to our sampling period.

[12] At this level, most industries are related to at least one other industry. As we leave out all plants for which no related industries exist, this prevents reducing our sample unnecessarily. The choice for exactly 3,000 links is therefore reasonable but, admittedly, *ad hoc.*

Related localization externalities are defined in an analogous way, but they measure the employment potential in related industries:

(8) $\qquad RLOC_{\pi} = \sum_{i' \in R_i} E_{i',m,y}^{pot}$ .

Here $R_i$ represents the set of industries that are related to industry $i$ according to our definitions above but excludes $i$ itself.

To measure Jacobs' externalities we need a variable that captures 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. An industry can be said to have a significant local presence if its number-of-plants potential exceeds a certain threshold, $\xi$. In the main text, we report on analyses where this threshold is set equal to five plants. The variable for traditional or *pure Jacobs' externalities* is thus calculated as follows:

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

where $\Xi(X, \xi) = \begin{cases} 1 \text{ if } X \geq \xi \\ 0 \text{ if } X < \xi \end{cases}$, $I$ is the set of all industries in Sweden, and

$PLANT_{i,m,y}^{pot}$ is the number-of-plants potential in industry $i$, municipality $m$, and year $y$. Similarly, by only counting the number of *related* industries with a significant local presence, it is possible to construct an indicator for *related Jacobs' externalities*:

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

Alternatively, the threshold, $\xi$, may be set in terms of an industry's local employment potential. There are two reasons why we prefer a number-of-plants based threshold. The first reason is that, in the nursery cities framework, diverse places are thought to have a large inter-industry knowledge spillover capacity of a location. According to Henderson (2003), each plant can be thought of as a specific experiment with the production technology in an industry. Therefore, the capacity of a local industry to

generate knowledge spillovers may be better proxied by looking at the number of plants than the number of employees. By contrast, the localization externalities in the nursery cities model arise from a large local demand for intermediates specific to the particular production process used in the industry. Localization externalities depend, therefore, on the overall demand generated by a local industry, which will be most correlated with the industry's total local employment. Secondly, a more technical reason to prefer a number-of-plants based value is that it reduces the correlation with the population potential. After all, to fill in the jobs in a municipality that has many industries with a high employment potential, the population potential must be high as well. Therefore, this variable will be less co-linear with our urbanization variable. Be this as it may, in our robustness analyses, we check whether our results also hold for indicators based on employment potentials.

Finally, our variable to measure urbanization externalities is simply the population potential of the municipality of a plant: $POPPOT_{m,y}$.

In terms of plant characteristics, we know from the literature that the employment of a plant contributes significantly to its survival. This effect is controlled for in all our analyses by the inclusion of the variable *PLANTSIZE* that is equal to the number of employees in the plant. In sum, our matrix *X* in equation (1) contains the variables PLANTSIZE, LOC, RLOC, JAC, RJAC, and POPPOT that enter the analysis log transformed.

## 6: Empirical results

*Descriptive statistics and specification details*

Tables 1 and 2 contain some general descriptive statistics for our datasets. As we presuppose that corporate plants behave differently from non-corporate plants, we distinguish between them by use of an organization identification number. Plants that do not share their organization number with any other plant are called non-affiliated. All other plants are corporate plants.[13] The set restricted to plants with at least ten employees contains 2706 corporate and 8829 non-affiliated plants.

The entire sample consists of about 14,700 observations (plants that entered after 1970). However, as explained above, we drop all plants smaller than ten employees to obtain a consistent sampling definition over time. This reduces the number of investigated plants to about 11,500. The plants that are lost in this operation seem to be randomly distributed across the country. The correlation between a dummy representing a size between five and nine employees and each of our agglomeration indicators is always lower than 5 percent. In what follows, we will focus on the results that were obtained when leaving out all plants under ten employees.[14]

---

[13] The use of the organization number is not unproblematic. Firms may change organization numbers and use more than one organization number for administrative or legal reasons. The distinction between corporate and non-affiliated plants by use of organization IDs is, however, the best our data allow.

[14] Note that the plants below the employment threshold of ten 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 ten employees. Outcomes are very similar, but standard errors are somewhat larger.

The cross-correlations between the covariates are shown in Table 2. Given the sample size, in principle we have no reason to be overly concerned about multicollinearity effects. Moreover, if multicollinearity is an issue and our regression analyses run into numerical problems, small changes in our estimation equation should have large effects on coefficients. Such problems should therefore be easily detected when we experiment with different specifications of our externality variables in our robustness checks.

*Interpretation of the regression tables*

In all regressions below, we control for differences in hazard rates across industries by adding 3-digit industry dummies. Furthermore, all variables have been log transformed. This implies that the coefficients should be interpreted as "hazard elasticities": a $\delta$ % increase in the $p^{\text{th}}$ variable raises the exit rate by $\exp(\delta\beta_p)$. The regression tables report untransformed coefficients with their robust[15] 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 increase (or decrease) in the rate with which plants exit that is associated with a given change in the covariate value.

*Outcomes*

Table 3 summarizes the results of the Cox regressions where it is assumed that the influence of agglomeration externalities is constant across the entire lifetime 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. The doubling of the initial employment of a plant results in a

---

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

reduction of the hazard ratio by 14.5 percent.[16] 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 have a small positive, yet not significant effect on survival. By contrast, Jacobs' externalities are significant. Doubling the number of significant 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 (LOC) still do not have any impact. A mirror image is found for diversity effects: while *pure* Jacobs' effects are significant and beneficial for survival, a large local variety of *related* industries (RJAC) does not matter.

A problem may arise due to the fact that RJAC and JAC are related to each other by construction. Where RJAC counts the number of significant *related* industries in a city, JAC counts the *overall* number of significant industries in the city. Moreover, the variables RJAC and RLOC are also possibly interacting. After all, if our RJAC variable indicates that there are a large number of industries active in the city, the sum total of employment in these industries, RLOC, will also be large. We therefore ran some experiments with different covariate specifications to investigate the coefficient of RJAC.

First, we replace JAC with a variable that counts the number of unrelated significant local industries (say, UJAC). In this specification, the industry counts of UJAC and RJAC are carried out over mutually exclusive sets: the set of related industries and the set of unrelated industries. This eliminates the problem that JAC and RJAC are correlated by design. The result of this is that the coefficient of RJAC indeed changes. Its point estimate drops below zero, indicating that related diversity has a positive effect on survival.

---

[16] 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. Therefore, we proceed using the simpler linear specification. In later regressions, this non-linearity is manifest in the different coefficients we get for plants of different sizes.

However, this effect is not significant. Next, we rerun the analysis without RLOC, allowing the RJAC variable to pick up all effects of the presence of related industries. Even now, however, the effect of RJAC remains insignificant. We thus conclude that RJAC does not influence survival in any significant way. We therefore return to our initial specification but drop the related Jacobs' externalities indicator, RJAC.

Next we split our sample to check whether there are any differences between corporate and non-affiliated plants. Column (3) is based on the sample of corporate plants, and column (4) covers the sample of non-affiliated plants. The results are indeed 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 almost 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, contrary to our expectations, 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. Because, 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.

In the lower part of Table 3, we see that there is, in fact, a problem with the analyses we ran thus far. The reported chi-squared test statistics indicate that the assumption of proportional hazard rates is violated. This may be due to non-constant effects of the industry dummies. Table 4 shows the results of a proportional hazard test with covariate specific test statistics when industry dummies are excluded. The hypothesis of proportional

hazards is still rejected.[17] The test statistics indicate that the main problem is caused by pure Jacobs' externalities.

To investigate this issue further, Figures 1 to 5 graph the cumulative regression coefficient for each variable in an Aalen model that contains the same covariates as before.[18] To rehearse the interpretation of the graphs: each graph shows how the year-on-year compound effect of a covariate on the hazard rate (y-axis) varies with the age of the plants (x-axis). The slope of the graph indicates the instantaneous effect of the covariate on hazard rates at a particular age, so departures from a straight line indicate changes in the effects of the covariate. Each graph contains a solid line representing point estimates and two dotted lines corresponding to a 95 percent confidence interval. A practical matter is that, for higher ages, the number of plants that run the risk to exit becomes very low. As a consequence, at high ages, Aalen coefficients are based on only a small number of observations, and the graphs become very volatile. For this reason, we do not attach much value to the shape of the graphs after an age of 25 years.

---

[17] In fact, although these variables exert significant effects, they do not seem to confound other variables substantially. Point estimates of significant coefficients shift only marginally (less than 25 percent) if industry dummies are omitted. Furthermore, the effect of pure localization externalities turns significant in columns (1) and (4).

[18] As the Aalen model estimates a separate coefficient for each age for all of the covariates, identification is based on the plants that are still at risk of exiting at a certain age. As soon as there is too little variation in one of the covariates, the Aalen model stops producing estimates. Due to the small number of plants in some of our 3-digit industries, the required variation in the corresponding industry dummy is already lost at the age of 22. As we are only interested in the changes in the coefficients of our main covariates, we drop the industry dummies in the Aalen models. However, if we include industry dummies, the graphs are very similar to the ones we depict here.

Starting with the ln(PLANTSIZE) variable, we find a downward sloping line up to the age of 19 years. This suggests that, over this period, the size of the plant has a positive effect on a plant's survival rate. After 19 years, the line is more or less horizontal, indicating that for mature plants the initial size is no longer relevant. Similarly, we can find changes in slopes for ln(LOC), ln(JAC), and ln(POPPOT). ln(RLOC) does not seem to undergo any significant changes in slope. On the basis of this visual 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 age-dependence structure specified. Column (1) of Table 5 is a repetition of column (4) in Table 3 but with the variable RJAC omitted. The same regression, though 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, this can now be wholly attributed to non-proportionalities in the industry dummies, which are not of immediate interest here.[19]

In line with the Aalen graphs, we find that some of the slopes change substantially with age. The initial employment (PLANTSIZE) only contributes significantly to the survival of young plants. Mature plants do not seem to benefit from higher initial employment levels, but a Wald test comparing the coefficients of young and mature plants shows that the difference in slope is only significant at an 8 percent level. 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, however, is not significant. If we turn to localization externalities, we find that pure

---

[19] 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 percent level.

localization externalities are not significant at any age. Related localization externalities are modeled as age invariant and turn out to have the usual positive and significant effect on survival rates, with a point estimate that is very close to the baseline estimates of column (1). The most interesting finding is, however, that Jacobs' externalities (JAC) improve survival chances only in the early years of a plant's existence. At higher ages, the point estimate is positive (indicating increased failure rates) but insignificant. A Wald test on equality of slopes shows this change of slopes is significant at any conventional level.

A possible explanation for the change of coefficients is that this is an artifact of our decision to measure the size of covariates at the time of birth. If the local environment changes over time, these initial conditions may be less informative for the agglomeration externalities a plant experiences as it grows older. This would result in an artificial weakening of the observed externality effects at higher plant ages. To investigate this possibility, we rerun our analyses with covariates that change over time.

The general picture remains the same. There are no significant localization externalities, but there are strong related localization externalities. Pure Jacobs' externalities contribute positively in the early years of a plant and are insignificant later on. Again, urbanization effects are only significant and negative in the early years and turn insignificant thereafter.[20] The main difference with our previous findings is that the positive influence of plant employment on the survival rates of plants is now strongly increasing. This is not surprising, as the current plant size should carry great weight in the decision to close a plant. A minor difference is that point estimates of related localization externalities drop slightly, and they are insignificant if we take all plants into consideration. However, when we analyze subsamples of small and of medium sized plants below, the effect of RLOC turns significant again.

In their model, Duranton and Puga suggest that young firms are involved in exploration activities, whereas mature firms focus on mass-production.

---

[20] A minor difference is that this drop is significant now.

Because plants that were set up for mass-production can be expected to be larger than prototype plants, we may find 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).[21]

Columns (3) to (5) show the outcomes of regressions based on these sub-samples without age-dependent coefficients. Taking the standard errors into account, only the effect of PLANTSIZE differs significantly between plants of different sizes. The differences in the other coefficients can solely be regarded as indicative. However, pure localization externalities seem to rise with an increasing size of the plants, and in our large plants sample they are even significant for the first time in our analyses. 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 large plants need to rent bigger spaces, and the higher rents in big cities should affect them more strongly.

Columns (6) through (8) contain the results of the analyses with age-dependent coefficients. Parameter estimates get less precise as a consequence of the smaller number of observations after dividing our sample into subsamples. The most striking finding is that pure Jacobs' externalities are significant only for young plants and not for mature plants across all subsamples, especially in the subsamples of medium sized and large plants.[22] This suggests that plants of any size benefit from Jacobs' externalities in their early years only. Another interesting outcome is that pure localization externalities have a strong positive effect on the survival of large plants at a high age. This suggests that, after several years, large plants start extracting important benefits from being located close to other plants in their industry.

---

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

[22] 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.

*Robustness*

To conduct the analyses above, we had to take a number of *ad hoc* decisions. Below we assess the sensitivity of our outcomes to these decisions by introducing small variations in the estimation specification. In particular, we alternate between specifications where localization externalities are measured in terms of number-of-plants potentials and those where they are measured in terms of employment based indicators. Next, we introduce six alternative lower limits for the definition of what constitutes a "significant presence" of an industry when calculating JAC and RJAC.[23] 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 48 different specifications for each regression analysis we have discussed so far. Tables 6 and 7 summarize the outcomes of this exercise. The upper rows in Table 6 and in the upper part of Table 7 show the percentage of analyses that yield the same sign as the corresponding columns in Table 3 and 5. Directly below each of these rows, in italics, we report the percentage of times the outcome was also significant at the 5 percent level. For example, in the rows belonging to RLOC and column (3) of Table 6, we can see that 95.8 percent of all 48 different robustness specifications yield the same negative sign we find in Table 3, and 85.4 percent of all 48 specifications yield both a negative sign and were significant at least at the 5 percent level. The bottom part of Table 7 gives the percentage of times that the Wald test for changes of slopes was significant at the 5 percent level.

To a very large degree, the signs and significance of the results of the alternative specifications match the ones we presented 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 point estimates of the RJAC estimates show quite some variation in their signs,

_____

[23] These are 1, 5, or 10 when we look at the local number-of-plants potential, and 50, 100, or 250 for the local employment potential.

but they are also hardly ever significant. Turning to Table 7, we find again primarily corroborations of the reported findings. Signs for the overwhelming majority are the same as in Table 5. As before, the main differences between specifications are found when variables are not significant. In the bottom part of Table 7, moreover, we see that our important findings concerning the changes in slopes are very robust as well. For variables that significantly changed slopes in Table 5 (as indicated by low p-values for the Wald statistic), we also find that other specifications indicate a rejection of equal slopes at the 5% level. High p-values, which would suggest that there is no indication that slopes change, usually result in zero percent of such rejections. Overall, we can conclude that the main results we reported above are quite robust.

*Discussion*

Overall, we find a strong and robust impact of local agglomeration variables on the survival rates of plants. In terms of the traditional urbanization, localization, and Jacobs' externalities, we find that only Jacobs' externalities increase survival probabilities. By contrast, a large local population leads to a higher failure rate of plants. This is consistent with recent theoretical work by Melitz and Ottaviano (2008) about the connection between market size and fierceness of competition. The authors show that an increase in market size results in lower profit margins. As large cities represent large local plants, urbanization externalities would increase the competition between local plants and raise their exit rates.[24] Another potential explanation is that the higher rents in large cities constitute an important congestion effect. This interpretation is supported by the fact that the negative effects of urbanization externalities are particularly strong for medium and large plants.

A puzzling finding is that localization externalities do not provide any benefits, except for the most mature and largest plants. One may speculate that old plants benefit from localization effects because they have had the time and the clout to structure a local cluster of firms to their own advantage. However, due to the small number of observations in this

---

[24] We thank professor Duranton for pointing this out.

category, this may also be a statistical artifact. The more general result that localization externalities are not important for most plants contrasts with a main finding by Dumais et al. (2002), which states that lower exit rates reinforce the concentration of industries in a limited number of regions. However, our pure localization externalities are more narrowly defined than the concentrations of broad 3-digit industry classes used by Dumais and his co-authors. In fact, if we look at the combined effect of pure and related localization externalities, we find this indeed strongly reduces the hazard rate of local plants. In general, however, it is the local activity in *related* industries, not in the same industry, that is responsible for the lion's share of these benefits. A plausible explanation for this finding is that plants in related industries can give access to ideas that are new for a plant but can be easily adapted to use in their own industry.

Interestingly, the effects of agglomeration externalities are very different for corporate than for non-affiliated plants. Most strikingly, corporate plants do not benefit at all from Jacobs' externalities, but they do benefit from related localization externalities and substantially more than their non-affiliated counterparts. Intuitively, corporate plants can substitute external linkages by internal linkages and therefore draw less on the local environment. This would explain the absence of Jacobs' externalities, but the strong effect of related localization externalities suggests that this is not the whole story. The difference between corporate and non-affiliated plants may, therefore, constitute an interesting point of departure for future research.

In light of the nursery cities model, the most important finding undoubtedly 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 to larger plants as well. Economic diversity may not only support explorative activity but also help overcome all kinds of teething problems encountered by any newly founded plant.

# 7: Conclusion

We set out to investigate (1) how agglomeration externalities impact the survival of plants, (2) how the presence of related industries affects the survival of a plant, and (3) how the nature of these impacts changes with the evolving maturity of a plant.

Our finding in answer to the first question depends on the type of the firm. More specifically, corporate plants and non-affiliated plants experience different agglomeration benefits. We focused our analyses on the non-affiliated plants, which constitute by far the largest group in our sample. If we look only at traditional categories of agglomeration externalities, we find that pure Jacobs' externalities increase survival rates substantially, pure localization externalities have no significant influence, and urbanization externalities lead to higher failure rates.

In an investigation of the changes of the influence of agglomeration externalities over time, we find that the assumption of the nursery cities model that young plants benefit from local diversity is empirically justified. The benefits of Jacobs' externalities indeed drop as plants grow older. However, contrary to expectations that explorative activities in small plants should benefit most from the inter-industry spillovers associated with Jacobs' externalities, if anything, these externalities vary even more strongly with plant age for medium sized and large plants than for small plants. The "nursery" role of diversified cities is apparently not limited to the prototype development stage. In the other agglomeration externalities, such changes of slope are not as apparent.

Next, we extended the analysis by looking at diversity and concentration in related industries. The local *diversity* of related industries does not yield any significant benefits. However, the influence of a concentration of related industries, which we labeled related localization externalities, contributes greatly to the survival of local plants. Accordingly, the most important sources of knowledge for a plant are plants that are engaged in activities that are not precisely the same as their own activities but still are related to them. Such plants are a source of ideas that are novel but still close enough to existing practices to be relevant.

Taking into consideration relatedness linkages severely complicates the picture of a local economy, as industries get intertwined in an intricate way. However, it also gives rise to a deeper understanding of how plants interact at the regional level. For example, whereas the traditionally much more investigated pure localization effects seem to lead to higher survival rates only in a small subset of plants that are both large and mature, related localization externalities show a persistent positive effect on survival chances across all different types of plants. In fact, the consistency with which these industries generate externalities suggests that they constitute one of the prime assets of a city.

This finding indicates that, in order to better understand regional economies, we need to probe deeper into the links that exist between industries. This article has focused on technological linkages. However, non-technological local client-supplier interaction may also contribute to a region's success. Comparing and combining the role of both types of relatedness may be a fruitful area of research. In this article, we have also pooled all industries together to draw inferences. Industry specific idiosyncrasies were supposed to be captured completely by industry fixed effects. However, the maturity of an industry has often been argued to influence the benefits that firms can derive from agglomeration externalities (e.g., Glaeser *et al.*, 1992; Audretsch and Feldman, 1996; Neffke *et al.*, 2008). Therefore, another interesting question is whether there is an additional effect of the overall maturity of an industry that has not yet been captured by the age structure of its plants.

# Literature

Aalen, O.O. (1989), A Linear Regression Model for the Analysis of Life Times, *Statistics in Medicine*, 8: 907-925.

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

Audretsch, D.B., Feldman, M.P. (1996), Innovative Clusters and the Industry Life Cycle, *Review of Industrial Organization*, 11: 253-273.

Bernard, A.B., Jensen, J.B. (2007), Firm Structure, Multinationals, and Manufacturing Plant Deaths, *Review of Economics and Statistics*, 89(2): 193-204.

Boschma, R.A., Iammarino, S. (2009), Related Variety, Trade Linkages, and Regional Growth in Italy, *Economic Geography*, forthcoming).

Boschma, R.A., Wenting, R. (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., Dreßler, K., Krüger, J.J. (2006), Firm Survival in the German Automobile Industry, *Empirica* 33: 49-60.

Combes, P.-P., Magnac, T., Robin, J.-M. (2004), The Dynamics of Local Employment in France, *Journal of Urban Economics*, 56: 217-243.

Cox, D.R. (1972), Regression Models and Life-Tables. *Journal of the Royal Statistical Society, Series B (Methodological)*, 34(2): 187-220.

Disney, R., Haskel, J., Heden, Y. (2003), Entry, Exit and Establishment Survival in UK Manufacturing, *Journal of Industrial Economics,* 1: 91-112.

Dumais, G., Ellison, G., Glaeser, E.L. (2002), Geographic Concentration as a Dynamic Process, *Review of Economics and Statistics,* 84(2): 193-204.

Duranton, G., Puga, D. (2001), Nursery Cities: Urban Diversity, Process Innovation, and the Life Cycle of Products, *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., Van Oort, F.G., Verburg, T. (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., Kallal, H.D., Scheinkman, J.A., Schleifer, A. (1992), Growth in Cities, *Journal of Political Economy*, 100(6): 1126-1152.

Greene, W.H. (2000) *Econometric Analysis* (4th edition). New Jersey: Prentice Hall.

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

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

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

Hidalgo, C.A., Klinger, B., Barabási, A.-L., Hausmann, R. (2007), The Product Space Conditions the Development of Nations, *Science*, 317: 482-487.

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

Hosmer, D.W., Lemeshow, S. (1999) *Applied Survival Analysis, Regression Modelling of Time to Event Data*. New York: John Wiley & Sons.

Hosmer, D.W., Royston, P. (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.

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

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*. 8th edition. London: Macmillan and Co.

Melitz, M.J., Ottaviano, G.I.P. (2008), Market Size, Trade, and Productivity, *Review of Economic Studies*, 75: 295-316.

Neffke, F.M.H., Svensson Henning, M. (2008) Revealed Relatedness: Mapping Industry Space. PEEG Working Paper Series #08.19.

Neffke F.M.H., Svensson Henning, M., Boschma, R.A., Lundquist, K.-J., Olander, L.-O. (2008) Who Needs Agglomeration? Varying Agglomeration Externalities and Industry Life Cycle. PEEG Working Paper Series #08.08.

Rosenthal, S.S., Strange, W.C. (2003), Geography, Industrial Organization, and Agglomeration, *Review of Economics and Statistics* 85(2): 377-393.

Thompson, P. (2005), Selection and Firm Survival: Evidence from the Shipbuilding Industry, 1825-1914, *Review of Economics and Statistics*, 87(1): 26-36.

# Tables and Figures

**Table 1:**     Descriptive statistics of covariates

|  | Obs | Mean | Std. dev. | Min | Max |
|---|---|---|---|---|---|
| CORP | 11535 | 0.2346 | 0.4238 | 0.0000 | 1.0000 |
| ln(PLANTSIZE) | 11535 | 3.0375 | 0.8438 | 2.3026 | 8.9335 |
| ln(LOC) | 11501 | 4.4984 | 3.4890 | -33.9642 | 9.9399 |
| ln(RLOC) | 11535 | 7.3476 | 2.2644 | -55.3776 | 10.8369 |
| ln(JAC) | 11535 | 1.9989 | 1.0840 | 0.0000 | 4.0431 |
| ln(RJAC) | 11535 | 0.9562 | 0.8681 | 0.0000 | 3.2189 |
| ln(POPPOT) | 11535 | 11.4641 | 1.2356 | 7.6566 | 13.8122 |

Variables are as defined in section 5.4. CORP is a dummy variable that takes value 1 for corporate plants and 0 for non-affiliated plants.

**Table 2:** Cross-correlations between covariates

|  | 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 5.

**Table 3:** Cox regressions of plant survival rates assuming age-invariant effects

| | (1) all | (2) all | (3) CORP | (4) NON-AFFILIATED |
|---|---|---|---|---|
| ln(PLANTSIZE) | -0.157*** | -0.157*** | -0.251*** | -0.177*** |
| | (0.015) | (0.015) | (0.025) | (0.024) |
| | | | | |
| ln(LOC) | -0.006 | -0.003 | 0.016 | -0.008 |
| | (0.004) | (0.004) | (0.009) | (0.005) |
| | | | | |
| ln(JAC) | -0.151*** | -0.143*** | -0.021 | -0.198*** |
| | (0.025) | (0.029) | (0.063) | (0.033) |
| | | | | |
| ln(POPPOT) | 0.222*** | 0.236*** | 0.173** | 0.250*** |
| | (0.023) | (0.023) | (0.053) | (0.026) |
| | | | | |
| ln(RLOC) | | -0.035*** | -0.088*** | -0.023** |
| | | (0.007) | (0.018) | (0.008) |
| | | | | |
| ln(RJAC) | | 0.017 | 0.004 | 0.055 |
| | | (0.026) | (0.053) | (0.030) |
| | | | | |
| Industry dummies | yes | yes | yes | yes |
| | | | | |
| Model statistics | | | | |
| df PH stat | 31 | 33 | 32 | 33 |
| PH stat | 38.8 | 40.3 | 52.4 | 47.4 |
| Nobs | 11501 | 11501 | 2698 | 8803 |
| Log-likelihood | -57434.7 | -57424.3 | -11285.6 | -42426 |

Clustered (on plant identification numbers) standard errors in parentheses, * $p<0.05$, **$p<0.01$, *** $p<0.001$. Variables are as defined in section 5. 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 in age-invariant model

|  | rho | chi$^2$ | df | Prob>chi$^2$ |
|---|---|---|---|---|
| **ln(PLANTSIZE)** | 0.0050 | 0.14 | 1 | 0.7115 |
| **ln(LOC)** | -0.0010 | 0.00 | 1 | 0.9466 |
| **ln(RLOC)** | -0.0036 | 0.06 | 1 | 0.8126 |
| **ln(JAC)** | 0.0267 | 3.48 | 1 | 0.0623 |
| **ln(POPPOT)** | -0.0051 | 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 5. Model has been estimated without industry dummies

**Table 5:**  Cox regressions 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*** (0.024) | | -0.632*** (0.158) | 0.070 (0.171) | -0.098* (0.048) | | | |
| ln(LOC) | -0.008 (0.005) | | -0.007 (0.006) | -0.006 (0.009) | -0.018* (0.009) | | | |
| ln(RLOC) | -0.021** (0.008) | -0.022** (0.008) | -0.0191 (0.013) | -0.033*** (0.009) | -0.014 (0.018) | -0.019 (0.013) | -0.033*** (0.008) | -0.015 (0.018) |
| ln(JAC) | -0.167*** (0.028) | | -0.143*** (0.039) | -0.233*** (0.054) | -0.151* (0.064) | | | |
| ln(POPPOT) | 0.249*** (0.026) | | 0.198*** (0.035) | 0.327*** (0.049) | 0.294*** (0.061) | | | |
| **Age varying variables** | | | | | | | | |
| ln(PLANTSIZE) early | | -0.185*** (0.024) | | | | -0.600*** (0.160) | 0.052 (0.175) | -0.099* (0.050) |
| ln(PLANTSIZE) late | | -0.006 (0.099) | | | | -1.699 (1.007) | 0.441 (0.987) | -0.050 (0.183) |
| ln(LOC) early | | -0.007 (0.005) | | | | -0.007 (0.007) | -0.005 (0.009) | -0.012 (0.010) |
| ln(LOC) late | | -0.024 (0.017) | | | | -0.008 (0.027) | -0.013 (0.027) | -0.086*** (0.020) |
| ln(JAC) early | | -0.185*** (0.028) | | | | -0.150*** (0.039) | -0.256*** (0.055) | -0.193** (0.066) |
| ln(JAC) late | | 0.055 (0.064) | | | | -0.030 (0.099) | 0.029 (0.119) | 0.162 (0.123) |
| ln(POPPOT) early | | 0.253*** (0.026) | | | | 0.199*** (0.035) | 0.336*** (0.050) | 0.303*** (0.062) |
| ln(POPPOT) late | | 0.170 (0.088) | | | | 0.262 (0.137) | 0.051 (0.170) | 0.224 (0.157) |
| industry dummies | yes | yes | yes | yes | yes | yes | yes | Yes |
| **model statistics** | | | | | | | | |
| Chi2 stat. | 47.6 (32) | 42.6 (36) | 27.5 (31) | 26.9 (32) | 23.4 (32) | 27.4 (35) | 26.3 (36) | 18.1 (36) |
| # plants model | 8803 | 8803 | 4686 | 2333 | 1784 | 4686 | 2333 | 1784 |
| log likelihood | -42427.4 | -42418.2 | -20719.7 | -9903.9 | -6695.1 | -20717.5 | -9900.6 | -6689.1 |
| **Wald tests for change of slope: p-values** | | | | | | | | |
| ln(PLANTSIZE) | | 0.080 | | | | 0.283 | 0.701 | 0.796 |
| ln(LOC) | | 0.298 | | | | 0.987 | 0.747 | 0.001 |
| ln(JAC) | | 0.000 | | | | 0.211 | 0.012 | 0.002 |
| ln(POPPOT) | | 0.339 | | | | 0.639 | 0.088 | 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 5. PH statistic is chi-squared distributed under the null-hypothesis of proportional hazards. All estimations include 3-digit industry dummies. Chi2 stat. contains the test statistic for the proportional hazards assumption with d.o.f. in parentheses.

**Table 6:**      Robustness of results in Table 3

|  | (1) ALL | (2) ALL | (3) CORP | (4) NON-AFF. |
|---|---|---|---|---|
| ln(PLANTSIZE) | 100.0% | 100.0% | 100.0% | 100.0% |
|  | *100.0%* | *100.0%* | *100.0%* | *100.0%* |
| ln(LOC) | 100.0% | 93.8% | 100.0% | 100.0% |
|  | *75.0%* | *4.2%* | *37.5%* | *68.8%* |
| ln(RLOC) |  | 100.0% | 95.8% | 100.0% |
|  |  | *95.8%* | *85.4%* | *79.2%* |
| ln(JAC) | 91.7% | 83.3% | 35.4% | 85.4% |
|  | *75.0%* | *60.4%* | *0.0%* | *29.2%* |
| ln(RJAC) |  | 29.2% | 18.8% | 50.0% |
|  |  | *6.3%* | *10.4%* | *10.4%* |
| ln(POPPOT) | 100.0% | 100.0% | 100.0% | 100.0% |
|  | *100.0%* | *100.0%* | *100.0%* | *100.0%* |

The table contains the percentage of times the same sign is obtained as in Table 3. The second number (in italics) contains the percentage of times when both the sign is the same as in Table 3 and the outcome is also significant.

**Table 7:** Robustness of results in Table 5

| | (1)<br>ALL | (2)<br>TV<br>ALL | (3)<br>SMALL | (4)<br>MEDIUM | (5)<br>LARGE | (6)<br>TV<br>SMALL | (7)<br>TV<br>MEDIUM | (8)<br>TV<br>LARGE |
|---|---|---|---|---|---|---|---|---|
| ln(PLANTSIZE) | 100.0% | | 100.0% | 100.0% | 100.0% | | | |
| | *100.0%* | | *100.0%* | *0.0%* | *66.7%* | | | |
| ln(LOC) | 100.0% | | 100.0% | 100.0% | 100.0% | | | |
| | *75.0%* | | *45.8%* | *0.0%* | *33.3%* | | | |
| ln(RLOC) | 100.0% | 100.0% | 100.0% | 100.0% | 100.0% | 100.0% | 100.0% | 100.0% |
| | *93.8%* | *93.8%* | *81.3%* | *91.7%* | *0.0%* | *81.3%* | *91.7%* | *0.0%* |
| ln(JAC) | 83.3% | | 83.3% | 75.0% | 83.3% | | | |
| | *27.1%* | | *66.7%* | *50.0%* | *25.0%* | | | |
| ln(POPPOT) | 100.0% | | 100.0% | 100.0% | 100.0% | | | |
| | *100.0%* | | *95.8%* | *100.0%* | *100.0%* | | | |
| ln(PLANTSIZE)<br>early | | 100.0%<br>*100.0%* | | | | 100.0%<br>*100.0%* | 100.0%<br>*0.0%* | 100.0%<br>*37.5%* |
| ln(PLANTSIZE)<br>late | | 70.8%<br>*0.0%* | | | | 100.0%<br>*0.0%* | 100.0%<br>*0.0%* | 100.0%<br>*0.0%* |
| ln(LOC) early | | 100.0%<br>*66.7%* | | | | 100.0%<br>*47.9%* | 100.0%<br>*0.0%* | 100.0%<br>*0.0%* |
| ln(LOC) late | | 100.0%<br>*0.0%* | | | | 77.1%<br>*0.0%* | 100.0%<br>*0.0%* | 100.0%<br>*95.8%* |
| ln(JAC) early | | 91.7%<br>*75.0%* | | | | 89.6%<br>*66.7%* | 83.3%<br>*58.3%* | 100.0%<br>*54.2%* |
| ln(JAC) late | | 95.8%<br>*37.5%* | | | | 43.8%<br>*0.0%* | 100.0%<br>*16.7%* | 100.0%<br>*25.0%* |
| ln(POPPOT) early | | 100.0%<br>*100.0%* | | | | 100.0%<br>*95.8%* | 100.0%<br>*100.0%* | 100.0%<br>*100.0%* |
| ln(POPPOT) late | | 100.0%<br>*18.8%* | | | | 100.0%<br>*18.8%* | 29.2%<br>*0.0%* | 100.0%<br>*0.0%* |
| Wald tests for change of slope: p-values | | | | | | | | |
| ln(PLANTSIZE) | | 87.5% | | | | 0.0% | 0.0% | 0.0% |
| ln(LOC) | | 0.0% | | | | 0.0% | 0.0% | 87.5% |
| ln(JAC) | | 100.0% | | | | 0.0% | 83.3% | 70.8% |
| ln(POPPOT) | | 0.0% | | | | 0.0% | 41.7% | 0.0% |

The table contains the percentage of times the same sign is obtained as in Table 5. The second number (in italics) contains the percentage of times when both the sign is the same as in Table 5 and the outcome is also significant. The rows for the tests on equality of slopes count the percentage of times that slopes are significantly different at 5% confidence level.

**Figure 1:** Aalen graph of cumulative regression coefficient for ln(PLANTSIZE)

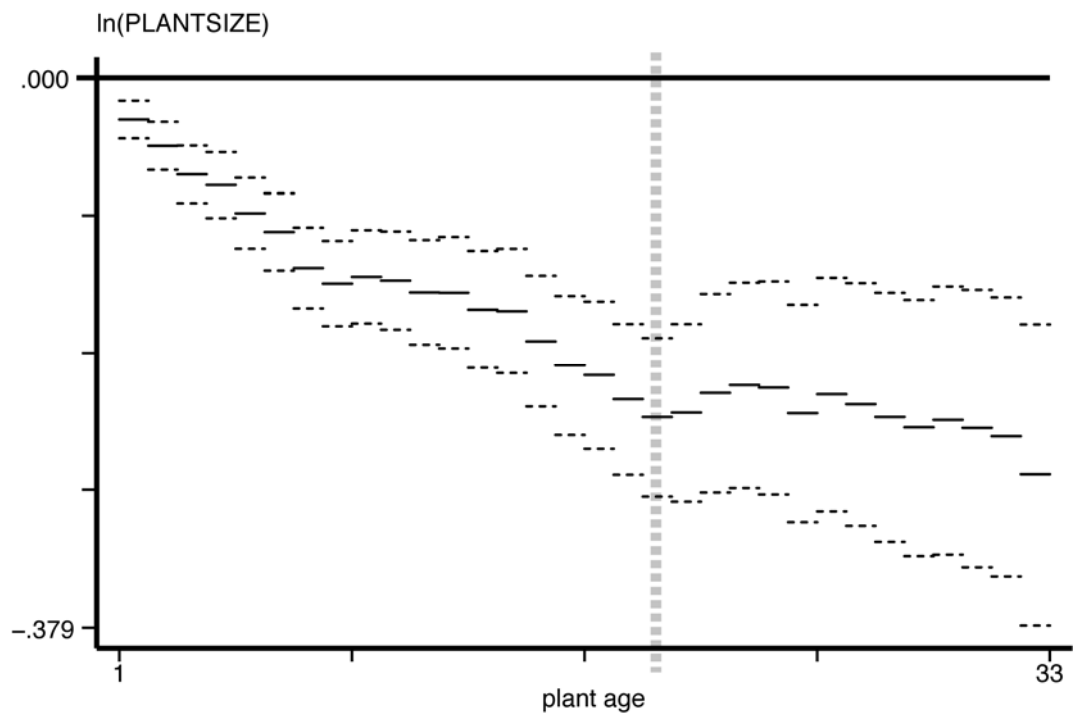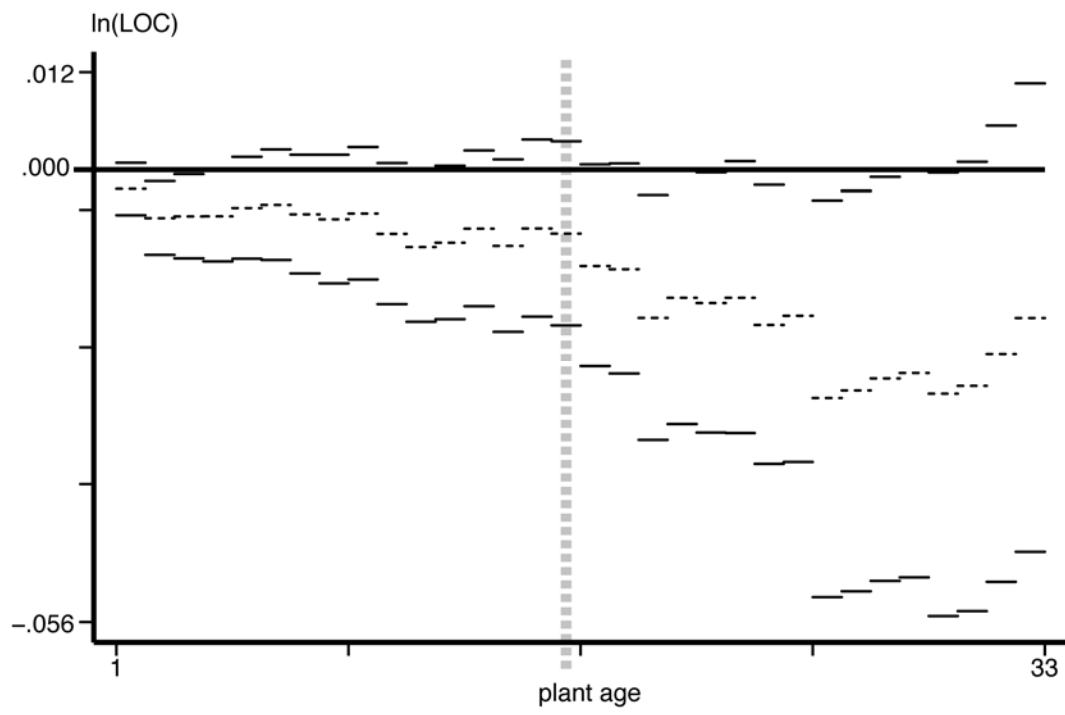**Figure 2:** Aalen graph of cumulative regression coefficient for ln(LOC)

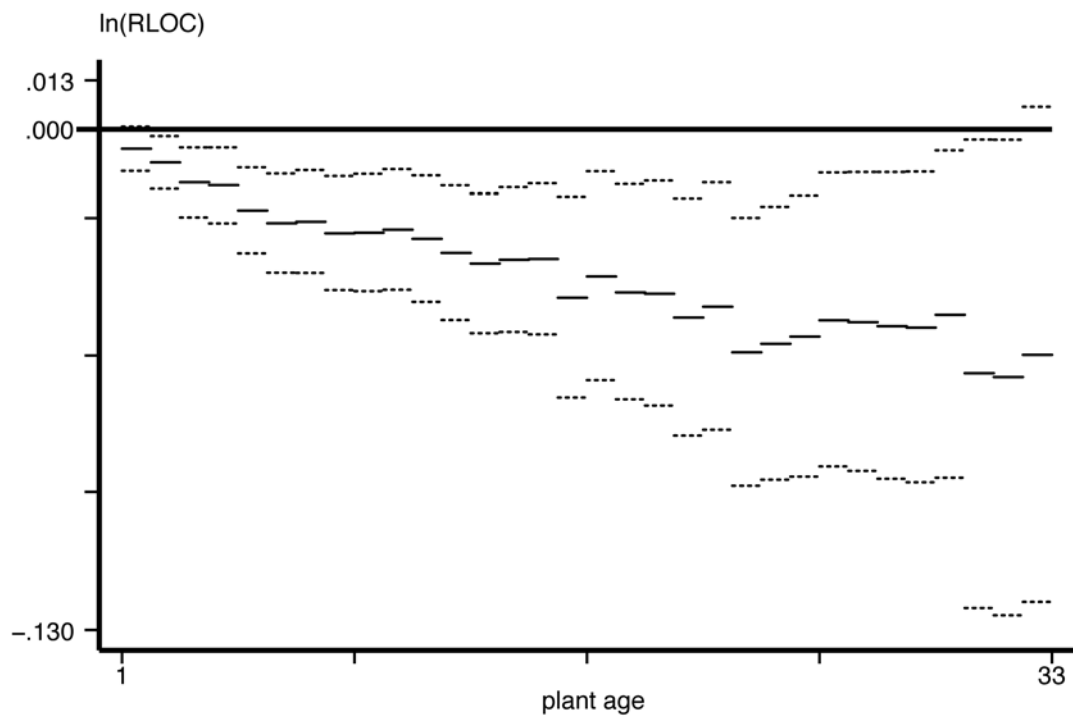**Figure 3:** Aalen graph of cumulative regression coefficient for ln(RLOC)

**Figure 4:** Aalen graph of cumulative regression coefficient for ln(JAC)

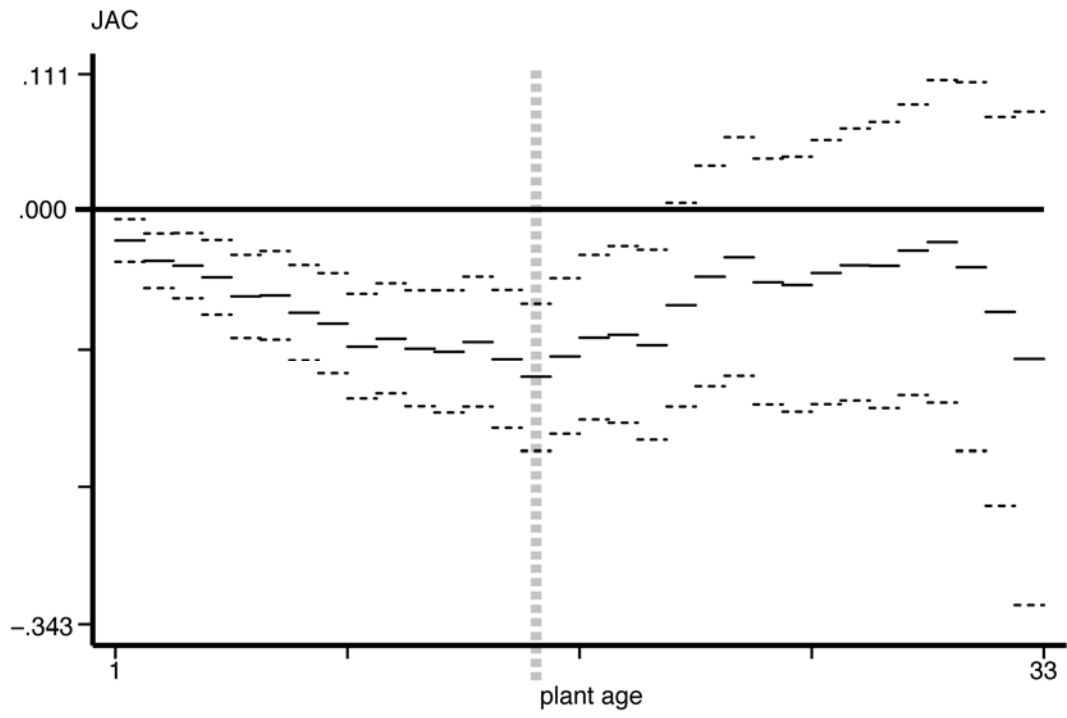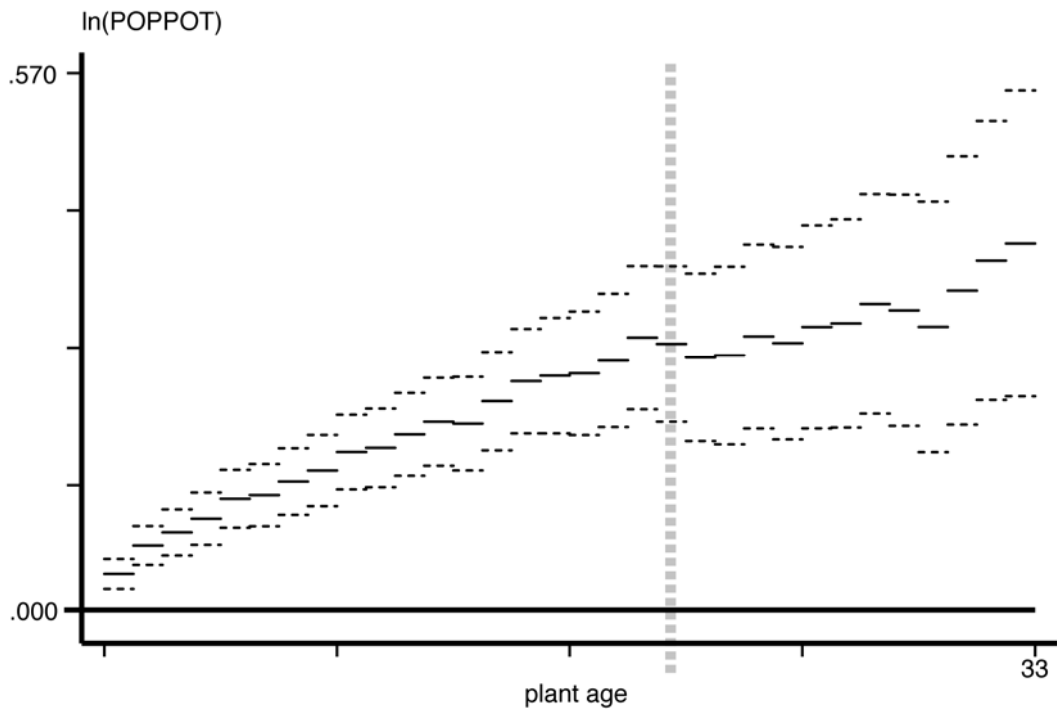**Figure 5:** Aalen graph of cumulative regression coefficient for ln(POPPOT)