Spatial modeling of firm survival

Giuseppe Arbia§, Giuseppe Espa*, Diego Giuliani*, Rocco MiccioloÀ
§*Università Cattolica del Sacro Cuore, Rome, Italy. Email: Giuseppe.arbia@rm.unicatt.it and College of William & Mary, Williamsburg (US)*
***Department of Economics and Management, University of Trento, Italy.*
À*Department of Psychology and Cognitive Sciences, University of Trento, Italy.*

Keywords: *Agglomeration, Firm survival, Spatial econometrics.*
JEL classification codes: *C21 · D92 · L60 · O18 · R12*

Preliminary version. Do not cite or quote without written permission of the authors.

1. Introduction and overview

The presence of knowledge spillovers and shared human capital is at the heart of the Marhall-Arrow-Romer externalities (MAR) hypothesis (see Marshall, 1890, Arrow, 1962, Romer, 1986) which represented the fount of a flood of scientific contributions produced in the last decades in the field of firm formation, agglomeration, growth and survival. According to the MAR hypothesis, similar firms located close by increase the chance of human interaction, labor mobility and knowledge exchange which in turn has an effect on firm creation, development and survival. Most of the earlier empirical contributions on knowledge externalities, mainly due to data limitation, considered data aggregated at a regional level leading to contrasting empirical results (Mansfield, 1995, Henderson, 2003; Rosenthal and Strange, 2003). In particular, the role of agglomeration economies has been considered to explain firm entry at a regional level and its effects on the growth of regional employment and regional production (Glaeser *et al*., 1992; Henderson, 2003). In most of this literature we can identify two major shortcomings. The first pertains the statistical sphere and concerns the way in which the agglomeration theories are empirically tested, the second concerns the main focus of these studies. More specifically the first limitation refers to the fact that, with data aggregated at a regional level, conclusions are based on the arbitrary definition of jurisdictional spatial units so that aggregating them in a different way we can obtain different (and often, contrasting) evidences: this is the essence of the so-called Modifiable Areal Unit Problem (Arbia, 1989). Furthermore theoretical models of new firm formation and firm exit are generally grounded on the behavior of the individual economic agent (see Hopenhayn, 1992; Krueger, 2003; Lazear, 2005) and they can be tested empirically on regional aggregates only under the unrealistic assumption of a homogeneous firm behavior within the region. The second limitation is constituted by the fact that, somewhat surprisingly, while concentrating on the effects of agglomeration on firm creation and growth, the literature has, conversely, largely ignored its effects on firm survival. These are the issues our paper seeks to contribute.

The search for a statistical solution to the MAUP has lead, in recent years, to an increasing interest towards the use of micro-data and establishment-level information thus eradicating at its deep roots the problem of defining *a-priori* the level of geographical partition and thus providing more robust evidences. This approach allows the use firm-level variability in spatial concentration in order to test if industrial localization and concentration influences firm demography and, also, to reconcile the contrasting empirical evidences found at an aggregated level. In this respect, a series of recent papers that introduced the use of point pattern analysis methods (generated mainly in the ecological and epidemiological literature, see e. g. Ripley, 1977; Diggle, 2003) are of tremendous potential impact in this area in order to measure geographical patterns of industries and their effect on firm creation, growth and survival (e. g. Arbia *et al.,* 2010; 2012; 2013; Marcon and Puech, 2010). The use of establishment-level data, however, until now, has been traditionally concentrated on firm creation (see Helfat and Lieberman, 2002 for a review) while only few papers are devoted to study industrial localization effects on firm exit, survival and bankruptcy. Some remarkable examples are those reported in Staber (2001), Folta *et al*. (2006), Sahver and Flyer (2000) and De Silva and McComb (2012).

The present paper aims at contributing to the existing literature by answering to some of the open methodological questions reconciling the literature of Cox proportional hazard with that on point pattern and thus capturing the true nature of spatial information. In particular our interest is in modeling the effects of spatial concentration and interaction on the probability of firm survival by incorporating both geographical variables and spatial interaction effects. We present a methodological advance with respect to the current literature (e. g. De Silva and McComb, 2012) in that we suggest to model the probabilities of firm exit of individual firms (with explicit consideration of their location), with a Cox proportional hazards model with micro-founded spatial covariates which takes into account both the spatial interactions among firms and the potential effects deriving from agglomeration. We also present some empirical results based on a recently released database on Italian firm demography managed by the Italian National Institute of Statistics (ISTAT), created in accordance with the procedures suggested by OECD and Eurostat. This database overcomes a series of inaccuracies due to non-demographic events (such as changes of activity, mergers, break-ups, split-off, take-over and restructuring) and to obtain a more realistic picture of firm demography with respect to that obtained examining traditional microdata drawn from Business Register.

In view of achieving our aim the present paper is divided into 3 more sections. Section 2 will be devoted to a brief summary of the state-of-the-art empirical methods to analyze survival data. Section 3 presents the methodology and shows the results of an empirical application to the case of start-up firms in the health and pharmaceutical sector during the years 2004-20008 in Italy. Finally Section 4 contains some concluding remarks.

2. A brief summary of survival analysis techniques

Survival data analysis (or of ñfailure time dataò) concentrates on the analysis of data corresponding to the time elapsed between a time origin and the time of occurrence of an event of interest (a ñfailureò). The topic is extremely challenging because, when the time elapsed represents the response variable of a model, standard statistical methods cannot be employed.

Perhaps the most striking feature of survival data is the presence of censoring, i.e. the presence of incomplete observation, when the follow-up time is shorter than that necessary for an event to be observed. Censoring can arise due to time limits and other restrictions and makes it very hard to calculate even the simplest descriptive statistics like, e. g., the mean or the median survival times. Furthermore, survival data often exhibit a positively skewed distribution with a high degree of asymmetry which makes the normal distribution hypothesis unreasonable.

Failure times can be considered empirical realizations of a positive random variable *T*. In what follows we consider *T* as a continuous random variable characterized by a probability density function *f(t)* and a (cumulative) distribution function $F(t) = \Pr[T \le t] = \int_0^t f(x)dx$. The survival function $S(t)$ can therefore be defined as the complementary function of $F(t)$:

$$S(t) = 1 - F(t) = \Pr[T > t] \tag{1}$$

is the probability of surviving at *t* (Marubini e Valsecchi, 1995).

Apart from the three above mentioned functions $(f(t), F(t), S(t))$, two more functions $(\lambda(t), \Lambda(t))$ are of interest when dealing with survival data. The first function, say $\lambda(t)$, is called the hazard function, and represents the instantaneous failure rate for an individual surviving to time *t*

$$\lambda(t) = \lim_{\Delta t \to 0^+} \frac{\Pr\{t \le T < t + \Delta t \mid T \ge t\}}{\Delta t}; \tag{2}$$

In Equation (2) $\lambda(t)\mathrm{d}t$ thus represents the probability that the event of interest occurs in the infinitesimal interval $(t, t+\mathrm{d}t)$, given survival at time t (Marubini e Valsecchi, 1995).
The second function, say $\Lambda(t)$, is called the *cumulative hazard function* and represents the integral of the hazard function: $\Lambda(t)=\int_0^t \lambda(x)dx$. It is easy to show that $\lambda(t)=f(t)/S(t)$ and that, therefore, $\Lambda(t)=-\log S(t)$. From this relationship it is immediate to verify that $\Lambda(t)$ diverges so that $\lambda(t)$ is not a conditional density function.
Parametric survival models are commonly specified by defining a plausible functional form for $\lambda(t)$ from which $S(t)$ and $f(t)$ can be derived. The simplest distribution (which plays a central role in the analysis of survival and epidemiological data) is the exponential distribution (Marubini e Valsecchi, 1995) which assumes the hazard function to be constant through time $(\lambda(t)=\lambda)$. The basic model can be then extended to include regression variables, which enable to investigate the role of selected covariates taking into account the effect of confounding factors.
If Y is a continuous response, regression models are commonly used to model its expectation $E(Y)$. Since in the exponential distribution the expectation is $1/\lambda$, an alternative way (Glasser, 1967) is to model the hazard as:

$$\lambda(t,\mathbf{x})=\lambda_0 \cdot \exp(\mathbf{b}'\mathbf{x}) \tag{3}$$

In Equation (3) $\mathbf{x}$ is a vector of k covariates including a constant term and $\mathbf{b}$ is a vector of unknown regression parameters to be estimated. Since $\mathbf{b}'\mathbf{x}=b_0+b_1x_1+\ldots+b_kx_k$, the term $\lambda_0=\exp(b_0)$ represents the failure rate in the reference category (that is when $\mathbf{x}=\mathbf{0}$).
It is important to note that the model specified as in Equation (3) relies on two basic assumptions: (i) the hazard function is independent of the values of the covariates and (ii) the covariates act in a multiplicative way on the baseline hazard. Therefore, if we consider two individuals characterized by covariate vectors $\mathbf{x}_1$ and $\mathbf{x}_2$ respectively, the hazard ratio:

$$\frac{\lambda(t,\boldsymbol{x}_2)}{\lambda(t,\boldsymbol{x}_1)}=\exp[\boldsymbol{b}'(\boldsymbol{x}_2-\boldsymbol{x}_1)] \tag{4}$$

is independent on time. For this reason model (3) is called a *proportional hazard model*.
In a seminal paper (Cox, 1972) introduced a regression model which is currently the most widely used regression model in the analysis of censored survival data. In the Cox model, the hazard function depends on both time and covariates, but through two separate factors:

$$\lambda(t,\mathbf{x})=\lambda_0(t)\cdot \exp(\mathbf{b}'\mathbf{x}) \tag{5}$$

In Equation (5), the baseline hazard $\lambda_0(t)$ is arbitrary defined (although it is assumed to be the same for all individuals), while the covariates act in a multiplicative way on the baseline hazard. In this sense, the Cox model is a semi-parametric model where the hazards are proportional, since the hazard ratio, given by

$$\frac{\lambda_0(t)\exp(\boldsymbol{b}'\boldsymbol{x}_2)}{\lambda_0(t)\exp(\boldsymbol{b}'\boldsymbol{x}_1)} = \exp[\boldsymbol{b}'(\boldsymbol{x}_2 - \boldsymbol{x}_1)] \tag{6}$$

is independent of time. An important difference from the parametric model (3) is in the form of the linear predictor $\mathbf{b}'\mathbf{x} = b_1 x_1 + ... + b_k x_k$ which does not include an intercept term. In terms of the inferential strategy, the parameter estimators of a Cox model and the significance tests are usually based on the partial likelihood technique (Cox, 1975).

It is important to observe that the Cox proportional hazard model has been widely used in the empirical literature to model firm survival (e. g. De Silva and McComb, 2012). However, no attempt has been made thus far to explain how spatial interactions among firms affect their survival probabilities and the effects on survival deriving from spatial agglomeration. In the next section of this paper, we will improve the explicative power of a Cox model by taking explicitly into account spatial information while modeling the hazard function and the survival probabilities. We will illustrate the use of a *spatially-augmented Cox proportional hazard model* by means of an empirical application based on pharmaceutical and medical devices.

3. A spatially-augmented Cox proportional hazard model: the survival of pharmaceutical and medical device manufacturing start-up firms in Italy

In this section we illustrate the use of a Cox proportional hazards model based on micro-founded spatial covariates in order to assess the effects of agglomeration externalities generated by incumbent firms on the survival of start-up firms. The novelty of the present study lies in the way the micro-founded spatial covariates are built and formalized. We claim that our approach allows to overcome the methodological pitfalls met by the agglomeration measures typically used in the current literature while uncovering the problem of firm survival.

3.1 Description of the dataset

The empirical exercise focuses on 3,217 start-up firms of pharmaceutical and medical device manufacturing industry, located in Italy, that started their activity in the period 2004-2008. In order to assess the effects of agglomeration externalities on the survival of these firms, we also use the data about the 10,572 incumbent firms of the same industry, born before 2004 and still surviving in 2009. This dataset is a subset of an internationally comparable database on Italian firm demography built up and managed by the Italian National Institute of Statistics (ISTAT), in accordance with the procedures suggested by OECD and EUROSTAT and based on the statistical information contained in the National Business Registers.

The Business Registers collect yearly a large set of information on the date of registration (i.e. firm entry) or deregistration (i.e. firm exit) for each business unit. However, this information does not purely represent firm demography, as registration and deregistration may also depend on non-demographic events such as changes of activity, mergers, break-ups, split-off, take-over and restructuring. Even if much of literature on firm demography regularly makes use of data extracted from the Business Registers without any controls for the influence of non-demographic aspects, it should be noted that a simple observation of data from Business Register does not allow to properly compare firm demography at an international level due to a series of inconsistencies like different definitions, different units of observation, different national legal systems and so on. In this paper, we specifically exploit data on the true firm entries and exits to remove some of these inconsistencies. For each firm, the database currently contains, for the period 2004-2009, information about firm code, sub-sector of activity (according to the NACE classification), number of firm's employees, legal status (according to the current classification), firm's birth (if occurred in the period 2004-2009), termination date (in case the exit occurred before 2009) and the precise spatial location (in terms of GMT longitude and latitude coordinates).

3.2 Definition of the micro-founded spatial covariates

In the empirical literature based on firm-level data (*e.g.* Staber et al., 2001; Ferragina and Mazzotta, 2014 among others) agglomeration externalities effects on firm exit are typically assessed by regressing the probability (or hazard) of firm default on locational measures, such as industry specialization indices. Then the statistical significance, sign and magnitude of the associated estimated regression parameters are used to assess the empirical evidence indicating whether agglomeration externalities play a significant role in firm survival.

In this paper, however, we argue that the locational measures commonly used by researchers (such as the Locational Quotient or the Ellison-Glaeser index (Ellison and Glaeser, 1997) may not be adequate for at least three reasons.

First of all, these measures are calculated on regional aggregates built on arbitrarily defined spatial units (such as provinces, regions or municipalities) and, hence, they introduce a statistical bias arising from the discretional definition of space (*i.e.* the so-called *modifiable areal unit problems* bias, see Arbia, 1989). As an evidence of this effect, Beaudry and Schiffauerova (2009) reviewed the relevant regional science literature and found that the emergence and intensity of agglomeration externalities are strictly dependent on the level of spatial aggregation of data.

Secondly, the dependent variable (namely the hazard rate) is defined at a firm level while the locational measures are defined at a region level. As a consequence the regression model will be necessarily based on the implicit assumption that firmõs behavior is homogeneous within each region, which is certainly too restrictive in many empirical situations.[1]

Thirdly, the locational measures commonly employed in the literature do not provide any indication of their statistical significance, and therefore we cannot conclude in a conclusive way whether we are dealing with high or low spatial concentration.

To solve the above problems, we develop a firm-level distance-based measure of spatial concentration to be included in the Cox proportional hazard model thus taking into account the presence of spatial effects in firm survival.

[1] For a comprehensive discussion of the weaknesses of the region-level locational measures, see Duranton and Overman (2005) and Combes et al. (2008).

Furthermore, unlike the regional-level locational measures that can only detect the presence of externalities at regional level, the firm-level measures we propose here allow us also to clearly identify which firms benefit from MAR externalities testing, for example whether the small firms benefit from this effect relatively more than the big firms .
In order to build up a set of variables to capture MAR externalities, we rely on the well-established idea in the literature (Glaeser et al., 1992) that the degree of specialization of an industry matters more than its size. The rationale behind this hypothesis is that the degree of specialization can be seen as a proxy of the intensity and density of interaction among firms (Beaudry and Schiffauerova, 2009). In what follows we build up a firm-level distance-based measure of spatial concentration able to capture the start-up firmĝs potential for Marshall externalities generated by incumbents, founded on Getis local *K*-function (Getis, 1984). A local *K*-function is a statistical measure allowing to assess spatial interactions among geo-referenced locations. Indeed, in the context of micro-geographic data, which are identified by maps of point events (as represented by their longitude/latitude coordinates), Getis local *K*-function can be seen as an explorative tool that summarizes the characteristics of a spatial distribution of point events relative to the location of a given point event. In our particular case the event of interest is represented by the presence of start-up firms in a particular location and our modelling frameworks aims at testing statistically if a given individual start-up firm is more likely to be localized in a clustering situation. For any given start-up firm *i*, located in a given geographic area, the local *K*-function can be defined as follows:

$$K_i(d) = E\left[\sum_{j \neq i} I(d_{ij} \leq d)\right] \Big/ \lambda \qquad (7)$$

where the term d_{ij} is the Euclidean distance between the *i*th start-up firm and *j*th incumbent locations, $I(d_{ij} \leq d)$ represents the indicator function such that $I = 1$ if $d_{ij} \leq d$ and 0 otherwise, and λ represents the mean number of firms per unitary area (a parameter called *spatial intensity*). Therefore, $\lambda K_i(d)$ can be interpreted as the expected number of further incumbent firms located up to a distance *d* of the *i*th start-up firm. The local *K*-function quantifies the degree of spatial interaction between the *i*th start-up firm and all other incumbent firms at each possible distance *d*, and hence can be exploited to develop a micro-based measure of spatial concentration.
Henderson (2003) established that both the number of firms and employment level in a region are key determinants of the generation of spillovers within the region. For this reason, we introduce weights in Equation (7) to account for the number of employees of each firm. In this way we obtain the following weighted version of the local *K*-function:

$$WK_i(d) = E\left[\sum_{j \neq i} e_i e_j I(d_{ij} \leq d)\right] \Big/ \lambda \mu^2 \qquad (8)$$

In expression (8), the terms e_i and e_j denote the number of employees of *i*-th and *j*-th incumbent firm respectively, and μ the mean number of employees per firm. As a consequence, the term $\lambda\mu^2 WK_i(d)$ can be interpreted as the mean of the sum of the products formed by the number of employees of the *i*th start-up firm and the number of employees of all other incumbent firms located up to a distance *d* of the *i*th start-up firm.
Turning now to the inferential aspects, following Getis (1984) and Penttinen (2006), a proper unbiased estimator of $WK_i(d)$ for a study area with *n* firms is given by

$$\widehat{WK}_i(d) = \left(\sum_{j \neq i}^{n} e_i e_j w_{ij} I(d_{ij} \le d) \right) \Big/ (n-1)\hat{\lambda}\hat{\mu}^2 \tag{9}$$

where $\hat{\lambda}$ is the estimated spatial intensity[2] and $\hat{\mu}$ is the mean number of employees per firm computed on the *n* observed firms. Due to the presence of edge effects arising from the bounded nature of the study area, an adjustment factor, say w_{ij}, is introduced thus avoiding potential biases in the estimates close to the boundary[3]. The adjustment function w_{ij} expresses the reciprocal of the proportion of the surface area of a circle centred on the *i*th start-up firm's location, passing through the *j*th incumbent firm's location, which lies within the area *A* (Boots and Getis, 1988).
As a final step, we use the function expressed in Equation (9) to obtain a measure of spatial concentration with a clear benchmark value allowing to assess if the *i*th start-up firm is located in an agglomerated industrial area. The most popular approach in the literature (see *e.g.* Beaudry and Schiffauerova, 2009) has been to refer to a relative benchmark, in which an industry in a region is considered as geographically concentrated (or dispersed) if it is overrepresented (or underrepresented) within the region with respect to the entire economy. A relative measure allows to control for the presence of spatial heterogeneity in the study area and hence it is able to identify spatial concentration due to the genuine spatial interactions amongst economic agents (see *e.g.* Haaland et al., 1999 and Espa et al., 2013).
In light of these considerations, a firm-level relative measure spatial concentration for the health and pharmaceutical new established economic activities, can be defined as:

$$RS_i(d) = \widehat{WK}_{i,sector}(d) \Big/ \widehat{WK}_{i,all}(d) \tag{10}$$

where $\widehat{WK}_{i,sector}(d)$ is the weighted local *K*-function estimated on the incumbent firms belonging to the same health and pharmaceutical sub-sector of activity of the *i*th start-up firm and $\widehat{K}_{i,all}(d)$ is the weighted local *K*-function estimated on all incumbent firms of the entire health and pharmaceutical

[2] $\hat{\lambda} = n/|A|$, where *A* is the study area and $|A|$ denotes its surface.

[3] Firms located near the boundary of the study area may be close to unobserved firms located outside the study area. Neglecting this circumstance may lead to a biased estimate.

industry. If, at a given distance *d*, $RS_i(d)$ tends to be equal to 1 then the *i*th start-up firm is located in an area with a spatial extension of *d* where economic activities are randomly and independently located from each other, implying absence of spatial interactions. When, at a given distance *d*, the functional expressed in Equation (10) is significantly greater than 1, then the *i*th start-up firm is located in a cluster with a spatial extension of *d* where the incumbent firms of its sub-sector of activity are more concentrated then all incumbent firms of the dataset, implying presence of spatial concentration. For example, a value of $RS_i(d) = 2$ indicates that amongst the incumbent firms located within the distance *d* from the *i*th firm, the level of economic activity of incumbent firms belonging to the *i*th firmôs sub-sector is two times the level of economic activity of incumbent firms of the entire health and pharmaceutical industry. On the other hand, when at a given distance *d*, $RS_i(d)$ is significantly lower than 1, the *i*th start-up firm is located in a dispersed area, where the incumbent firms of its sub-sector of activity are less concentrated then all incumbent firms of the dataset, implying presence of spatial dispersion. For example, $RS_i(d) = 0.5$ indicates that amongst the incumbent firms located within the distance *d* from the *i*th start-up firm, the level of economic activity of incumbent firms belonging to the *i*th start-up firmôs sub-sector is half the level of economic activity of incumbent firms of the entire health and pharmaceutical industry.

The functional expressed in Equation (10) represents a relative measure of spatial concentration since the benchmarking value of random localization is represented by the spatial distribution of all the economic activities and hence a specific sub-sector exhibits spatial concentration (or dispersion) if its spatial distribution is more concentrated (or dispersed) than the spatial distribution of the economic activities as a whole. Therefore, it represents a micro-geographic firm-level version of the location quotient and hence a proper measure to assess the working of MAR externalities.

In order to evaluate the statistical significance of the values of $RS_i(d)$ a proper inferential framework needs to be introduced. Since the exact distribution of $RS_i(d)$ is unknown, its variance cannot be evaluated theoretically and no exact statistical testing procedure can be adopted. As a consequence, in order to test the null hypothesis of absence of spatial interactions, that is $RS_i(d) = 1$, we may base our conclusions on Monte Carlo simulated confidence envelopes (Besag and Diggle, 1977). In practice, we generate *n* simulations in each of which the *m* incumbent firm locations are randomly labelled with the observed *m* sub-sector of activity ómarkersô. Then, for each simulation, we calculate a different $RS_i(d)$ function. We are then able to obtain the approximate $n/(n+1) \times 100\%$ confidence envelopes from the highest and lowest values of the $RS_i(d)$ functions calculated from the *n* simulations under the null hypothesis. Finally, if the observed $RS_i(d)$ falls, at the given distance *d*, outside the envelopesð upward or downwardð this will indicate a significant departure from the null hypothesis of absence of spatial interactions.

3.3 Definition of the control variables

We considered two control variables, namely the number of employees in each firm and their legal status. More specifically, the number of employees are measured as the annual mean number of firm's employees classified into 3 categories, namely: firms with only 1 employee (2,496 firms in our database representing the 77.6% of all firms in our database), firms with a number of employees between 2 and 5 (681 firms, the 21.2%) and firms with more than 5 employees (40 firms, the 1.2%). In terms of the legal status the 3,217 start-up firms included in our dataset belong to three main categories, namely: sole trader (2,454 start-up firms, representing a percentage of 75.4% of the whole dataset), partnerships (365 start-up firms, the 11.3%) and companies (412 start-up firms, the 12.8%).

3.4 Results

Table 1 contains the Kaplan-Meier estimates of survival probability of the 3,217 start-up firms included in our database observed in the period 2004-2008. After one year from the entry around 94% of firms still survive, while after four years around 1 firm out of 4 exit. In the end, after five years of observation, the estimated survival probability is around 72%. The graph reported in Figure 1 shows that during the first five years of activity the propensity to exit tends to be constant over time.

Table 1. Kaplan-Meier estimates of survival probability of the 3,217 pharmaceutical and medical device manufacturing start-up firms in Italy, period 2004-2008

time	firms at risk	firm exits	survival probability	standard error	lower 95% CI	upper 95% CI
1	3217	198	0.938	0.00424	0.930	0.947
2	1595	109	0.874	0.00712	0.860	0.888
3	917	64	0.813	0.00990	0.794	0.833
4	496	32	0.761	0.01289	0.736	0.787
5	219	12	0.719	0.01690	0.687	0.753

Figure 1. Kaplan-Meier survival curves for the 3,217 pharmaceutical and medical device manufacturing start-up firms in Italy, period 2004-2008

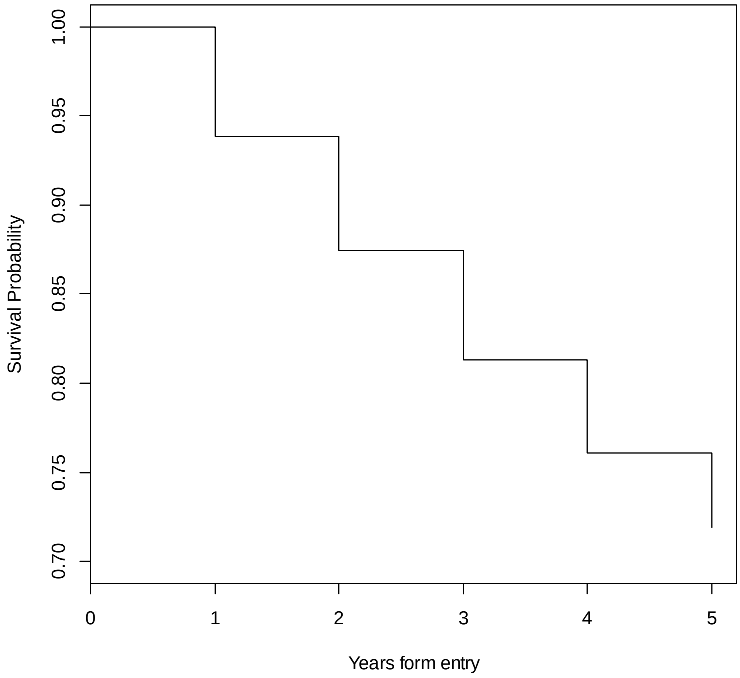

Table 2 contains the results of the spatially-augmented proportional hazard model estimations using the micro-founded spatial covariate. In particular, we considered the measure $RS_i(d)$ included in the model as a categorical variable classified within three categories, namely *Dispersion* (when $RS_i(d)$ is significantly lesser than 1, the baseline category), *Independence* (when $RS_i(d)$ = 1) and *Concentration* when $RS_i(d)$ is significantly greater than 1.

Three different models have been estimated according to the three values of distance *d* at which $RS_i(d)$ has been computed. We selected the values of *d* = 5, 15 and 60 kilometres as they are the most frequent distances amongst all the observed bilateral distances between the start-up and incumbent firms.

Table 2. Cox proportional hazard estimates for the 3217 pharmaceutical and medical device manufacturing start-up firms in Italy, period 2004-2008

	d = 5 km	*d* = 15 km	*d* = 60 km
$RS_i(d)$- *Independence*	0.3905***	0.4412***	0.4880***
	(0.1268)	(0.1288)	(0.1312)
$RS_i(d)$- *Concentration*	0.3056**	0.3954***	0.4672***
	(0.1289)	(0.1332)	(0.1466)
Employees - (1-5)	-0.1539	-0.1637	-0.1749
	(0.1320)	(0.1314)	(0.1310)
Employees - (>5)	-1.2264*	-1.2232*	-1.1961*
	(0.7160)	(0.7159)	(0.7158)
Legal status - partnerships	0.3495*	0.3682**	0.3780**

	(0.1792)	(0.1791)	(0.1786)
Legal status - companies	0.5215***	0.5150**	0.4874**
	(0.200)	(0.2005)	(0.2008)
Wald χ^2	22.21***	25.35***	27.15***

*** Denotes statistical significance at the 1% level, ** denotes statistical significance at the 5% level, and * statistical significance at the 10% level. Standard errors are in parentheses.

Table 2 shows that, for the three distances considered, the coefficients associated with the $RS_i(d)$ variable are positive and strongly significant. According to the way $RS_i(d)$ has been formalized, this implies that start-up firms located in a dispersed area (i.e. located relatively far from the incumbent firms belonging to the same sub-sector) will tend to have a lower hazard of exit. In other words, geographic proximity to incumbent firms increases the risk of failure, thus highlighting the presence of negative MAR externalities. This evidence suggests that in the industry of pharmaceutical and medical devices there is a prevalence of competitive over co-opetitive behaviors amongst economic agents.

4. Conclusion

In this paper we have introduced a spatially-augmented Cox proportional hazards model which includes micro-founded spatial covariates to assess the effects of agglomeration externalities generated by incumbent firms on the survival of start-up firms. We showed that this empirical methodology can assess properly the endogenous effects of interaction among economic agents on firm exits while overcoming the problems present in the various agglomeration measures typically used in the literature. By way of illustration, an application to the firms of pharmaceutical and medical device manufacturing industry, located in Italy in the period 2004-2009, has been conducted. The analysis indicates that the firm exit phenomenon is significantly affected by the spatial interactions of the neighboring competitors. The firm exit, albeit important, is only an aspect of the entire phenomenon of business demography. In order to study the spatial dynamics of firm demography thoroughly, the new firm formation and growth phenomena should be also considered along with the firm death phenomenon. The aim of developing a methodology to empirically assess all these sub-processes is left to future studies.

List of references

Arbia G. (1989) *Spatial data configuration in statistical analysis of regional economic and related problems*, Kluwer Academic Publisher, Dordrecht.

Beaudry C., Schiffauerova A. (2009) Who's right, Marshall or Jacobs? The localization versus urbanization debate, *Research Policy*, 38, 2, 318ï 337.

Boots B.N., Getis. A. (1988) *Point Pattern Analysis*, Sage Scientific Geography Series Vol. 8, London, UK.

Besag J., Diggle P.J. (1977) Simple Monte Carlo tests for spatial pattern, *Applied Statistics*, 26, 327ï 333.

Combes P.-P., Mayer T., Thisse J.F. (2008) *Economic geography: the integration of regions and nations*, Princeton University Press, Princeton.

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

Cox D.R. (1975) Partial likelihood, *Biometrika*, 62, 269ï 276.

De Silva D.G., McComb R.P. (2012) Geographic concentration and high tech firm survival, *Regional Science and Urban Economics*, 42, 691ï 701.

Duranton G., Overman H.G. (2008) Exploring the detailed location patterns of UK manufacturing industries using microgeographic data, *Journal of Regional Science*, 48, 213ï 243.

Ellison G., Glaeser E.L. (1997) Geographic concentration in U.S. manufacturing industries: A dartboard approach, *Journal of Political Economy*, 105, 889ï 927.

Espa G., Arbia G., Giuliani D. (2013) Conditional vs unconditional industrial agglomeration: disentangling spatial dependence and spatial heterogeneity in the analysis of ICT firmsô distribution in Milan, *Journal of Geographical Systems*, 15, 31ï 50.

Ferragina A.M., Mazzotta F. (2014) Local Agglomeration Economies: What Impact on Multinational and National Italian Firmsô Survival?, *Procedia - Social and Behavioral Sciences*, 110, 8ï 19.

Folta T.B., Cooper A.C., Baik Y.S. (2006) Geographic cluster size and firm performance, *Journal of Business Venturing*, 21, 217ï 242.

Getis A. (1984) Interaction modelling using second-order analysis, Environment and Planning A, 16, 173ï 183.

Glaeser E.L., Kallal H.K., Scheinkman J.A., Shleifer A. (1992) Growth in cities, *Journal of Political Economy*, 100, 1126ï 1152.

Glasser M. (1967) Exponential survival with covariance *Journal of the American Statistical Association*, 62, 561ï 568.

Haaland J.I., Kind H.J., Midelfart-Knarvik K.H., Torstensson J. (1999) *What determines the economic geography of Europe?*, Centre for Economic Policy Research, Discussion paper, 2072.

Henderson J.V. (2003) Marshallôs Scale Economies, *Journal of Urban Economics*, 53, 1ï 28.

Mansfield E. (1995) Academic research underlying industrial innovations: sources, characteristics, and financing, *Review of Economics and Statistics*, 77, 1, 55ï 65.

Marcon E., Puech F. (2010) Measures of the geographic concentration of industries: improving distance based methods, Journal of Economic Geography, 10, 5, 745ï 762.

Marubini E., Valsecchi M.G. (1995) *Analysing Survival Data from Clinical Trials and Observational Studies*, Wiley, Chichester.

Penttinen A. (2006) Statistics for Marked Point Patterns, *The Yearbook of the Finnish Statistical Society*, 2006, 70ï 91.

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

Shaver M.J., Flyer F. (2000) Agglomeration economies, firm heterogeneity, and foreign direct investment in the United States, *Strategic Management Journal*, 21, 1175ï 1193.

Staber U. (2001) Spatial proximity and firm survival in a declining industrial district: the case of knitwear firms in Baden-Wurttemberg, *Regional Studies*, 35, 329ï 341.