DENSITY DEPENDENCE AND THE SURVIVAL OF FIRMS A MULTI-AGENT APPROACH

Pedro Campos[1], Pavel Brazdil[2]
LIACC-NIAAD
Faculty of Economics – University of Porto
R. de Ceuta, nº 118, 6º, 4050 - Porto - Portugal
[1]pcampos@fep.up.pt; [2]pbrazdil@liacc.up.pt

KEYWORDS

Multi-Agent Simulation, Density Dependence Model, Genetic Algorithms.

ABSTRACT

We have used a Multi-Agent Based Simulation to simulate a representative set of firms for a specific type of industries and geographical locations and observed the interaction among them, with special attention to their geographical location and mortality. We defined a variant of the density dependence model to set up the dynamics of the firms in the simulation. By comparing some real and simulated results, and adjusting the chosen parameters with the help of a Genetic Algorithm, we obtain estimates for density limits and other intrinsic parameters not observable in reality.

INTRODUCTION

What are the determinants of the survival of firms? What explains their births and deaths? To investigate the survival of organizations some researchers use the density dependence model, where vital rates of birth and death of firms are dependent on the size of the population, the population density. Spatial components (geographical barriers, localized resource environments, etc.) affect also the evolutionary dynamics of organizational populations in a nonlinear way.
Based on this idea and due to its relevance for organizational survival, we focused our research on certain parameters explained in detail later on (eg. density of firms, age and size), and the impact on mortality of organizations using a Multi-Agent based simulation. We used a calibration process, based on Genetic Algorithms to find the combination of these determinants that minimize the distance between real and simulated aggregate behaviour. We compared vital rates (birth and death and growth rates) of simulated data with real data from Portuguese Textile Industry from 1997 to 2003. The region of Ave has been chosen because it is a geographical cluster where textile industries are concentrated in Portugal. To analyse the impact of different variables in the organizational mortality we applied survival analysis models and examined the regression coefficients.

The remaining of the paper is structured as follows: in the next section we introduce the density dependence model and discuss the determinants of the survival of organizations. In the following sections we present some related works and describe the simulation and calibration processes. Conclusions and future work are presented at the end of the paper.

DENSITY DEPENDENCE

The field of Organizational Ecology (OE) establishes a relation between the society outside organizations and the internal life of organizations (“firm” and “organization” are terms used indistinctively in this paper). OE was introduced by Stinchcombe (1965), Carroll and Hannan (1989, 1992), Hannan and Freeman (1984), among others, who received motivation from sociology and ecology. Recent research determined that density (the number of organizations in a population) affects rates of organizational founding and mortality and formal models of density dependence have been developed. In the density dependence model, vital rates of birth and death of firms are dependent on the population density
According to this model, the survival of a firm depends on the number of firms that are located on its neighbourhood. In most situations, this neighbourhood is not strictly geographic but an “industrial neighbourhood”, where firms are considered nearby if they share the same market or resources. In OE, the notion of rational behaviour or profit maximization is not accepted as the main driving motivation for firms. Rather, this role is taken by forces of natural selection and organizational inertia (Wissen, 2004). This theory offers a comprehensive way of studying the evolution of populations of firms. We will introduce some variations in the model, as it seems to be weak in two main aspects: first, the size of the organization is not taken into account in theory, whereas clearly large and small firms have quite distinct characteristics. Second, firms differ not only with respect to size and economic activity, but also with respect to geographical location, that is, spatial heterogeneity (Wissen, 2004). The density dependence model is a single industry model, which means that it ignores the relationship with the environment, including the relationship with other industries. For those reasons, we will give special

attention to the geographical location, the size of the firm and the inter-industry relationships in our analysis.
Two basic forces are responsible for the dependency between the population density and founding (birth) and failure (death): legitimation and competition. Legitimation refers to the degree that a new organizational form is known and accepted in society. The founding rate is proportional to the level of legitimation whereas the mortality rate is inversely proportional to the level of legitimation (opposite associations are observed with respect to competition). The join effects of legitimation and competition explain to a large degree the specific S-shaped form of growth rates over time, because as the population grows, legitimation increases and competition is still very restricted, so that the growth rate increases. At a certain level, the maximum level of legitimation is reached, and competition starts to increase fast. Consequently, the growth rate decreases fast to zero or even becomes negative.

RELATED WORKS

Many authors, (as Albino 2003; Zhang 2003) use Multi-Agent Systems (MAS) for modelling clusters of organizations. Epstein and Axtell (1996) explain that the use of MAS is appropriate when interactions among (heterogeneous) agents are complex and nonlinear and when the space is crucial. Considering the impact of density on the mortality of organizations, Carroll and Hannan's (1989) use simulation models to conclude that density at the time of founding may change the mortality rate of adult organizations. Lomi and Larsen (1998) used cellular automata in a Game of Life-like simulation, based on an Ecological Dynamics of local structure in a nearest neighbour model to simulate organizational populations. The main goal of their work was to study the effects of density at the time of founding on firms growth, decline and death. Here we are concerned with an application to a real industrial context, and assign more relevance to some concepts such as size, location and age. We have compared real and simulated results, adjusting some parameters with the help of a Genetic Algorithm and obtained estimates for density limits and other intrinsic parameters not observable in reality. The observation of the simulated aggregate behaviour and its comparison with reality validates the model and helps to identify the main determinants of organizational survival.
This kind of comparison with "real conditions of organizations" within the density dependence model is original in the scope of the works concerning the simulation of organizations.

TECHNIQUES, MODELS AND DATA.

The Multi-Agent System

Our system MASOS – Multi Agent System for Organizational Survival - is composed by a landscape, agents, and a rule. The landscape is represented by a graph that interconnects the places where firms can be located. We focused our environment in four regions of North-West of Portugal where textile industry is concentrated (showed in Fig.2). One of them, Ave, is often named as an Industrial District or cluster and is the centre of our study.

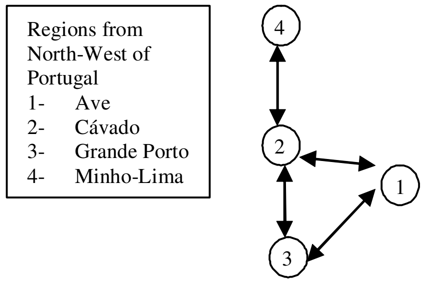

Figure 1: North-western region of Portugal represented as a graph where arrows connect contiguous areas representing the four regions of the study.

Each firm is represented by an agent with a set of attributes: location, age and size. There is one rule in MASOS that connects the agents and the environment which is responsible for the determination of the firms' birth and death processes regarding the variant of the density model presented before.

Variant of the density model

To determine the birth and death of a firm, we have based our representation on a variant of the density dependence model. According to this variant, in every region, a firm can get born or die, depending on a function which is based on the number of firms that are located in its neighbourhood. For that purpose, we have calculated the density, D_j^t, involving a particular region, j, at time t, as the average number of firms per square kilometre that are placed in the region or in the nearest neighbour regions at the period, t:

$$D_j^t = \sum_{i=1}^{K_j} \frac{f_i^t}{Area_i} \quad (1)$$

In this formula K_j is the number of regions that belong to the neighbourhood of region j; $Area_i$ is the area of the regions involving j, measured in square kilometres and f_i^t corresponds to the number of firms existing in the region i at time t (i=1, 2, …, K_j), and is computed as follows:

$$f_i^t = \sum_{j=1}^{N_i} \delta_{ij}^t \qquad (2)$$

where N_i represents the total number of firms in the region i and δ_{ij}^t is the state (dead - coded as 0 - or alive – coded as 1) of the j^{th} firm of the region i at time t. For example, if at time 2 the number of firms around a particular firm in the region of Ave (region 1), is 2345, then $f_1^2 = 2345$. The level of neighbourhood is computed according to the distance between regions. In our simulation, density will affect the processes of founding and failure in the following manner: if the number of alive organizations in the neighbourhood of a specific firm belongs to the survival interval [DS_l ; DS_u], then the organization will have more probability to stay alive. Otherwise it will have more probability to die by the effect of “overcrowding” or “solitude”, depending on whether the number is greater than DS_u or lower than DS_l (where DS_l and DS_u are respectively the lower and upper bounds of the density survival interval). The same idea applies to the process of founding where DB_l and DB_u respectively are the lower and upper bounds of the density founding interval. Moreover, considering what was said about the association between density and legitimation, DS_u may give us an idea about the maximum levels of legitimation for a particular type of industry.

We will consider that age and size are also determinants of the firm’s survival. Following Mata, Portugal and Guimarães (1995), size is an important determinant of the chances of survival and it exerts a negative effect on the instantaneous failure rate. So, we will consider that larger firms have more survival chances. Size will be measured by the number of employees. Also, experimental studies have shown that infant firms are more exposed to death (Carroll and Hannan, 1989). Thus, we will define S_l and A_l as the lower size and age thresholds above which an organization is more likely to survive. Therefore, assuming independence among density, size and age, the probability that a firm maintains in state 1, i.e., the survival for the firm i in the region j at time t, named P($Survival_j$), can be formulated as :

$$P(Survival_j) = P(\delta_{ij}^t = 1 \mid D_j^t, S_i^t, A_i^t) =$$
$$= P(\delta_{ij}^t = 1 \mid D_j^t) \times P(\delta_{ij}^t = 1 \mid S_i^t) \times P(\delta_{ij}^t = 1 \mid A_i^t) \qquad (3)$$

where S_i^t represents the size of the firm i at time t and A_i^t is the Age of the organization i (measured in years) at the time t. Figure 3 shows how the interval limits are associated with the probability of survival.

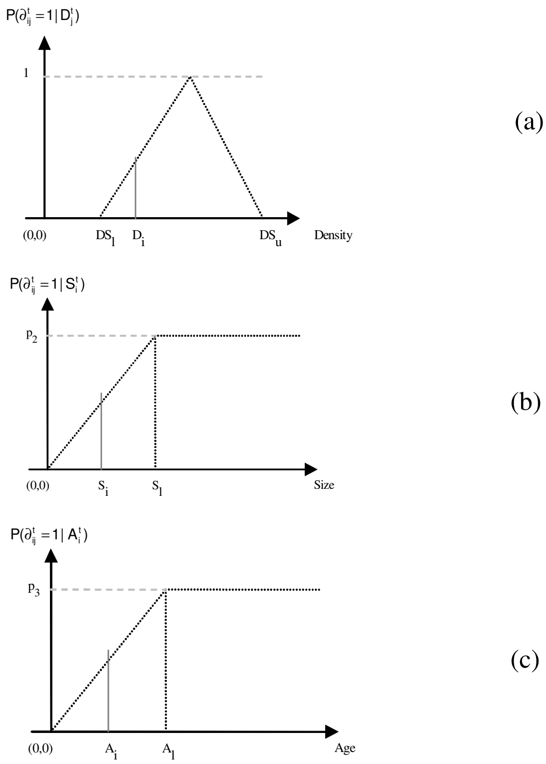

Figure 2: Probability of survival and interval limits for Density Survival (a), Size (b) and Age (c) and illustrative values for D_i, S_i and A_i. (at time t)

The parameters of the MAS (DS_l, DS_u, DB_l, DB_u, S_l, and A_l) need to be determined. They will be defined randomly at the beginning of the iterations but are optimized by calibration.

Preparation of the simulation

Real information is introduced in the simulation, in order to represent reality in a feasible way: distances between regions; a attraction level, created by the authors, measuring the degree of industrialization and road access for industry location based on physical infrastructures, taxes, knowledge capital, inter-industry linkages, etc; number of firms in the textile industry as well as number of employees in that sector; resident population and vital (birth and death) rates for textile firms. The main data source for this work was the Firms and Plants Central File (FCEE) collected by INE - the Portuguese National Statistics Institute. We note that there is no information about density interval limits, and age and size thresholds. We prepared the simulation to generate the same conditions for birth and death of firms that exist in the four regions in the North-West of Portugal in the beginning of the study: the same number of firms in 1997 was distributed within the four regions according to the attraction levels; the age of each firm was set to 1 in first year; size was randomly generated, according to real size observed in each location.

To compare the impact of certain variables on the survival of firms, we used a Cox Proportional Hazard Model. Cox model (Cox, 1984) is used to determine the level of importance of some variables (X_{jt}) in the mortality of a cohort with time t. This model avoids the need to estimate ho(t), the formal distribution of the data (Cox, 1984, Mata, J. Portugal, P. and Guimarães, P., 1995).

$$h_j(t, X_{jt}) = h_0(t)\exp(X_{jt}\beta) \qquad (4)$$

Calibration of the final solution

The calibration process uses a Genetic Algorithm (GA). GAs can be seen as adaptive algorithms based on simulated evolution (Goldberg, 1989, Mitchell, 1997) using an appropriate metaphor of selection, crossover and mutation from Biology. In the context of the GA, a population of hypotheses is created in which each element of the population (hypothesis) contains the combination of parameters (DS_l, DS_u, DB_l, DB_u, S_l, and A_l) represented as genes. The GA then tries to search for the hypothesis that maximizes the fitness function value, by minimizing the distance between real and simulated output data. The output data are the simulated birth and death rates for the region of Ave. Fitness function contains a measure of the distance (Manhattan Distance) between the values of the vital rates produced from synthetic data and real observed data. We select half of the best fitted elements on each iteration and crossover carry out exchanging their genes. We also introduced a mutation rate of 10% in the selected population. We ran several arrangements of hypotheses and iterations of the whole algorithm, most of them containing 20 hypotheses and 10 iterations.

Simulation outline

The simulation algorithm considers the following steps:

1) Start-up: a fixed number of initial firms are randomly assigned to the regions taking into consideration the attraction level of each region.
2) Randomly define the parameters of the model: upper and lower boundaries for survival and founding, and lower limits for age and size.
3) Cycle for the MAS and hypotheses testing. Repeat until a predetermined threshold for the best fitted hypothesis is reached or a fixed number of iterations is attained (the one that happens first).

- *(if it is not the first iteration of the model): Apply Selection, Crossover and Mutation for the set of h hypotheses, and produce a new generation of h hypotheses. Runs the MAS so many times as the number of different hypotheses*
  - *Cycle for the generations of firms (using the parameters of a specific hypothesis):*
    - *(3.1) Produce N generations of firms; Determine in which 'regions' a new firm will be born; Determine which firms will die; Firms that survive can grow (increment their size);*
    - *(3.2) Compare vital rates (birth, death) produced by the population of firms from simulated data (at the region aggregate level) with those from real data and calculate the fitness function value to obtain the most fitted hypothesis.*

4) Return the best fitted hypothesis

Results and Validation

As we are firstly more concerned with the qualitative behaviour of the model, we started by observing the regularity of the simulated evolution of the number of firms in the four regions. Figure 3 shows the evolution of the population density in two of the regions for the textile industry. As the data from 1997 was the base for the simulation, we were curious about the evolution path of textile industry produced by the MAS. The figure shows simulated values for Ave (represented as "S Ave") and Minho-Lima ("S Minho-Lima)" in doted lines. Real values for those regions are represented in solid lines. Ave has the most part of textile industries in the region (as expected) and presents similar values as observed in reality. Even if the simulated evolution paths are not exactly the same as the real paths for every region, we believe that the MAS represents the reality in a coherent manner. Although this similarity may represent a validation of the simulation model in a certain way, we are more interested in analyzing further aspects of the survival of organizations that can be estimated from the simulation, rather than searching for a perfect fit between simulation and reality. In any case we need to be aware of the problem of overfitting the data.

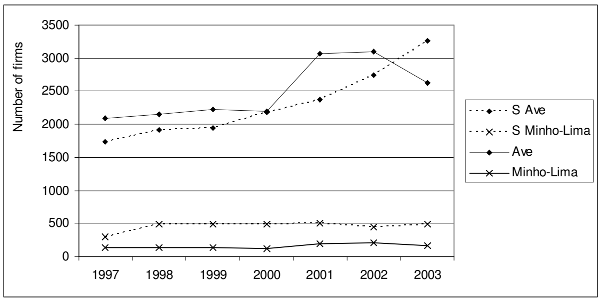

Figure 3: Real and simulated evolution of the number of textile firms in two regions of Portugal

We have also analyzed the Regression coefficients from a Cox proportional Hazard Model, estimate the impact of the covariates in the survival of organizations (for the Cox Regression, the cohort of firms that has been considered was produced by the most fit hypothesis). We have compared the sign of the coefficients with those obtained in other works (Mata and Portugal, 1994; Mata, Portugal and Guimarães, 1995; Carroll and Hannan 1989; Carroll and Hannan, 1992). The second column in Table 1 shows the predicted signs (obtained from literature - the same signs have been predicted in all studies) of the association between firm survival and the covariates.

Covariate	Predicted Sign	Coef	se(coef)	p
Contemporaneous Density	-	-0.02688	0.00883	0.0023
Density at Founding	+	227.835	0.08537	0.0000
Size	-	-0.00412	0.00139	0.0031

Table 1: Coefficients of variables obtained from Cox Regression.

All the variables have a significant impact on the survival of firms, which can be seen from the p-values in last column (the standard error of the coefficients are represented as se(coef)). Carroll and Hannan (1989) explain that the density at founding has substantive implications on the survival of organizations. As we can see in Table 1, density at the time of a firm founding has a positive impact on the mortality of organizations while contemporaneous density (the number of firms existing at the moment of the death) has a corresponding negative effect. As expected, the effect of the size on firm survival is also negative, confirming what was said before and helping to validate the simulation model. Since the system seems to perform consistently when compared to reality (for comparable data), we tried to extract other useful information from the simulation using a calibration process. Our aim was to get an idea about the parameters for textile industry in the region of Ave. After running the simulation for a given number of iterations, the output values listed below correspond to the best combination of values (hypotheses), i.e. those which minimize the distance function used in this optimization process. These solutions of MASOS cannot be confirmed, because there is no available data for this kind of analysis. So a question arises whether this set of parameters is unique or, alternatively, whether there are more solutions. For that purpose we have analysed the outputs of several runs of the MAS simulation and retained the 10 best fitted hypotheses that constitute the solutions of the calibration process. We have found some small variations, as showed in Table 2.

DS_l	Dsu	DB_l	DB_u	S_l	A_l	Average Age
2	10	4	24	2	2	17,57
2	10	3	22	3	2	22.21
4	9	3	25	4	2	18.2
4	9	4	25	2	2	20.20
5	13	4	24	1	2	22.02
3	10	3	25	5	2	20.99
3	10	3	25	5	2	20.99

Table 2: Density interval limits (multiplied by 10) and Age and Size limits obtained from the best fitted solutions after GA calibration

We recall that DS_l and DS_u respectively are the lower and upper bounds of the density survival interval and DB_l and DB_u are respectively the lower and upper bounds of the founding (or birth) interval. A_l and S_l are the size, and age thresholds. These solutions define the hypothetical limits of the survival and founding intervals for the region of Ave between 1998 and 2003 (1997 was the starting year). Region of Ave seems to be very tolerant to newborn firms since estimated density birth interval limits lies between 0.3 and 2.5, which means that there is good receptivity for incoming firms in the period of analysis. New firms are almost always accepted but only some of them will survive: small size or infant firms are most likely to fail and older and larger firms will prevail. So, in the region of Ave, density survival (DS) levels of textile industries, measured in number of industries per square kilometre, stand more or less between 0.9 and 1.3 for the period of analysis. These values may constitute a type of legitimation limit that is implicitly established in the region. We have also obtained the ages and size limits (because small and younger firms are less likely to survive). We have found an age limit (A_l) of 2 and a size limit (S_l) that stands between 1 and 5.

CONCLUSIONS AND FUTURE WORK

Simulation with Multi-Agent Systems enables us to explore the situations where interactions are complex and nonlinear. In this work we focused our research on the density and studied its impact on the mortality of organizations using a Multi Agent based simulation. Data from the Portuguese textile industry have been used and we have tried to find out the density interval limits for this industry in the region of Ave, one of the most important textile clusters in Portugal. We have compared vital rates (birth and death rates) of simulated data with real data from Portuguese Textile Industry from 1997 to 2003. The real evolution path of the number of firms seemed quite similar when compared to the simulated evolution. Applying Survival Analysis, we could confirm the qualitative properties of the association among variables. Finally we have calibrated the MAS to identify the main determinants of organizational survival using a Genetic Algorithm and found the combination of these determinants that minimize the distance between real and simulated aggregate behaviour. The number of industries per square Kilometre stands more or less between 0.9 and 1.3 for the period of analysis We have also concluded that these firms are less likely to survive over the age of 2 and for a size interval that stands between 1 and 5.

In the future, we plan to compare the results with those obtained by some alternative techniques to provide a qualitative evaluation of the proposed model. Econometric techniques, such as regression analysis, for example, are often used to build up models where we can clearly distinguish between explanatory and dependent variables. We intend to compare the results of our method with classical techniques and conclude about the pertinence of our choice. Time series analysis is also suitable for modelling data, since we are going to generate sets of time indexed values. Another possibility could be Particle Swarm Optimization, particularly useful for our applications, as these are nonlinear optimization problems.

AUTHOR BIOGRAPHIES

Pedro Campos was born in Porto, Portugal, and obtained a degree in Applied Mathematics from Portucalense University and further post-graduate qualification (equivalent to M.Sc.) at the Faculty of Economics of the University of Porto (FEP). He is a Ph.D. Student in Managerial Sciences at FEP and member of LIACC - Laboratory of Artificial Intelligence and Computer Science.

Pavel Brazdil has obtained a Ph.D. in 1981 at the University of Edinburgh. He is a Full Professor at the Faculty of Economics of University of Porto and he is the coordinator of the group NIAAD of the Laboratory of Artificial Intelligence and Computer Science (LIACC).

REFERENCES

Albino, V., et al. 2003. “Coordination Mechanisms based on Cooperation and Competition within Industrial Districts: An Agent-based Computational Approach”. Journal of Artificial Societies and Social Simulation, vol. 6, no. 4.

Carroll, G., Hannan, M. 1989. “Density Delay in the Evolution of Organizational Populations: A model and Five Empirical Tests”, Admnistrative Science Quaterly, 34, 411-430.

Carroll, G., Hannan, M. 1992. “The Demography of Corporations and Industries”. Princeton University Press, Pinceton, New Jersey

Cox, D.R. and Oakes, D. 1984, “Analysis of Survival Data”. Chapman and Hall. Cambridge

Epstein, J., Axtell, R. 1996. “Growing Artificial Societies, Social Science from the bottom up”. Brookings Institution Press, The MIT Press.

Goldberg, D. 1989. “Genetic Algorithms in Search, Optimization, and Machine Learning”

Hannan, M. and Freeman J. 1984, “Structural Inertia and Organizational Change”. American Economic Review, 49, 149-164.

Lomi, A., Larsen, E. R. 1998. “ Density Delay and Organizational Survival: Computational Models and empirical Comparisons”. Computational & Mathematical Organization Theory, 3, 4, 219-247.

Mata, J. Portugal, P. and Guimarães, P. 1995. “The Survival of New plants: start-up conditions and post-entry evolution”. International Journal of Industrial Organization, 13, 459-481.

Mitchell, T. 1997. “Machine Learning”. Mc Graw Hill.

Stinchcombe, A. 1965. “Social structure and organizations”. The handbook of Organizations, Chapter 4, March, J. (Eds.), Rand McNally & Company, 142-193.

Wissen, L. (2004). “A Spatial Interpretation of the Density Dependent Model in Industrial Demography”. Small Business Economics, 22, 253-264.

Zhang, J. 2003 “Growing Silicon Valley on a Landscape: an agent-based approach to high-tech industrial clusters”. Journal of Evolutionary Economics, 13, 529-548.