A non parametric survival analysis of Business demography dynamics in Portugal[1]

Alcina Nunes[2]
Elsa de Morais Sarmento[3]

1. Introduction

This article addresses the post-entry performance of Portuguese firms, using survival and hazard functions, along a period of eighteen years, from 1985 to 2007. The method follows the "Manual on Business Demography Statistics" (OECD/Eurostat, 2007), so as to assure international comparability with other datasets, such those from the Entrepreneurship Indicators Programme (OECD/Eurostat 2008 and 2009). In the exercise, we use a sub-set of the *Quadros de Pessoal* dataset (Employment Administrative Records by the Portuguese Ministry of Labour and Social Security), where only active employer enterprises are considered. The survival analysis is then disaggregated in different dimensions, namely sectors, regions and size class.

This article proceeds as follows. In Section 2, we present some basic concepts. In section 3, we briefly describe the Kaplan-Meier survival function and the Nelson-Aalen hazard rate that underlie our analysis. In Section 4, we perform a survival analysis for the economy as a whole. In Sections 5, 6 and 7 we break down the analysis by economic sectors, regions and start-up size classes, respectively. Section 8 concludes.

2. Basic Concepts

Following OECD/Eurostat (2007), three basic definitions will be used. First, we adopt the concept of "employer enterprises", which consists of the population of active enterprises, with at least one paid employee. Hence, the analysis focuses on a specific subsample of the universe of Portuguese enterprises. Second, a "birth" amounts to the "creation of a combination of production factors with the restriction that no other enterprises are involved in the event". This means that a birth occurs only when an enterprise starts activity. Births do not include entries into the population which result from break-ups, spit-offs, mergers, restructuring of enterprises or reactivations of units which are dormant within a period of two years. Our population consists of enterprises that have at least one paid employee in its birth year and also of enterprises that, despite existing before the year in consideration, were below the one employee threshold. An employer enterprise entry is thus counted in the dataset as a birth of an employer enterprise after it recruits its first employee, while complying with the above mentioned requisites. Third, an employee enterprise "death" occurs when an employer enterprise stops having employees. Deaths do not include exits from the population due to mergers, take-overs, break-ups or restructuring of a set of enterprises. Moreover, deaths do not include exits from a sub-population if it results from a change of activity. Therefore, a death can occur because the enterprise ceases to trade or because it shrinks below the one employee threshold. The manual recommends waiting for two years after the reference period to allow for reactivations, before deaths are calculated.

3. Survival and hazard functions

This section provides an analysis of new firm survival, which draws extensively on the survival analysis literature in industrial economics.

---

[1] This work reflects the opinions of the authors and not of the Ministry of Economics, Innovation and Development.
The authors would like to thank Gabinete de Estratégia e Planeamento of the Portuguese Ministry of Labour and Social Security (in the person of Dr. Manuel João) for the provision of data and the helpful assistance.
[2] Gabinete de Estratégia e Estudos do Ministério da Economia e da Inovação; Escola Superior de Tecnologia e Gestão do Instituto Politécnico de Bragança, Portugal.
[3] Gabinete de Estratégia e Estudos do Ministério da Economia e da Inovação; Departamento de Economia e Gestão, DEGEI, Universidade de Aveiro, Portugal.

The survivor function reports the probability of a firm of surviving beyond time *t* (the moment of observation), that is the probability that there is no failure event (a "death") prior to *t*. The function is equal to one at time *t*=0 and decreases towards zero as time (*t)* goes to infinity. Considering *T* a non-negative variable, denoting the time to a failure event ("death"), in this case given by the time taken by an enterprise to exit the market from the moment of entry. The survivor function is thus represented by:

$$S(t) = 1 - F(t) = \Pr(T > t)$$

With $F(t) = \Pr(T \leq t)$ being the cumulative distribution function.

The hazard function or the conditional failure rate is the instantaneous rate of failure. It is the (limiting) probability that the faileure event ("death") event occurs in a given interval, conditional upon the subject having survived to the beginning of that interval, divided by the width of the interval:

$$h(t) = \lim_{\Delta t \to 0} \frac{\Pr(t + \Delta t > T > t \mid T > t)}{\Delta t} = \frac{f(t)}{S(t)},$$

Where $f(t) = \frac{dF(t)}{dt} = \frac{d\{1 - S(t)\}}{dt} = -S'(t)$ is the density function.

The hazard rate measures the rate at which risk is accumulated and can vary from zero (no risk at all) to infinity.

The integral from 0 to *t* of the hazard rates is known as the cumulative hazard function ($H(t)$). It records the number of times failures were observed over a given time period.

In practice, to estimate the survivor function, $S(t)$, that is the probability of survival past time *t* or, equivalently, the probability of failure after *t*, the non-parametric Kaplan-Meier estimator was applied. For a dataset with observed failure times, $t_1, ..., t_k$, where *k* is the number of distinct failure times observed in the data, the Kaplan-Meier estimate at any time *t* is given by:

$$\hat{S}(t) = \prod_{j | t_j \leq t} \left(\frac{n_j - d_j}{n_j} \right)$$

Where n_j is the number of enterprises at risk at time t_j and d_j is the number of failures at time t_j . The product is done for all the failure periods, departing from time *t*.

The most common estimator for the cumulative hazard rate is the non-parametric Nelson-Aalen estimator, which is defined by the sum of the instantaneous ratio of the failures over the number of enterprises at risk. This estimator is thus given by:

$$\overline{H}(t) = \sum_{j | t_j \leq t} \frac{d_j}{n_j}$$

4. Survival and hazard functions for the all economy

In Table 1, we estimate hazard duration and survival functions for the Portuguese economy as a whole. The survival function shows the probability of survival, considering that the firm has been active during a certain period, and the hazard function shows the probability of "death" throughout a given period of time.

According to this table, In Portugal, during the period from 1987 to 2005, approximately 86% of all the employer enterprise births remained active after one year of activity. That means that around 14% of all enterprises, died before they their first year of activity was completed. These results are in line with the OECD´s estimates, where around 60% to 80% of newly born enterprises survive beyond the first two years of activity, and only around 40% to 50% of total enterprises survive beyond the seventh year of activity. Eurostat (2009) also reported for the whole business economy, that roughly half of the enterprises survive during their first 5 years.

This data also reveals that after six years of activity, almost 50% of the Portuguese enterprise population was still active. In Portugal, the estimated median duration of a new born enterprise lies between 5 and 6 years (Figure 1). After 18 years of activity, only 22% of employer enterprise start-ups were still alive or equivalently, almost 78% had already exited the market.

Table 1 - Life Table for Employer Enterprise Births, 1987-2005

Time	Observations	Deaths	Censured Observations	Kaplan-Meier		Nelson Aalen	
				Survivor Function	Failure Function	Hazard Rate	Cumulative Hazard Rate
Years	nº	nº	nº	%	%	%	%
				P(S)	100-P(S)	P(D)	$\sum P(D)$
1	451.041	63.088	24000*	86,0%	14,0%	14,0%	14,0%
2	364.233	46.351	22000*	75,1%	24,9%	10,9%	26,7%
3	295.786	32.973	28000*	66,7%	33,3%	8,4%	37,9%
4	235.002	23.655	24000*	60,0%	40,0%	6,7%	47,9%
5	187.102	17.353	19000*	54,4%	45,6%	5,6%	57,2%
6	150.840	12.966	12000*	49,7%	50,3%	4,7%	65,8%
7	125.525	10.059	11000*	45,8%	54,2%	4,0%	73,8%
8	104.121	7.735	9.613	42,4%	57,6%	3,4%	81,2%
9	86.773	6.089	7.943	39,4%	60,6%	3,0%	88,3%
10	72.741	5.068	7.491	36,6%	63,4%	2,8%	95,2%
11	60.182	4.172	11000*	34,1%	65,9%	2,5%	102,2%
12	45.130	3.037	6.150	31,8%	68,2%	2,3%	108,9%
13	35.943	2.422	5.626	29,7%	70,3%	2,2%	115,6%
14	27.895	1.681	5.546	27,9%	72,1%	1,8%	121,7%
15	20.668	1.133	4.733	26,4%	73,7%	1,5%	127,1%
16	14.802	805	5.361	24,9%	75,1%	1,4%	132,6%
17	8.636	490	4.418	23,5%	76,5%	1,4%	138,2%
18	3.728	228	3.500	22,1%	77,9%	1,4%	144,4%

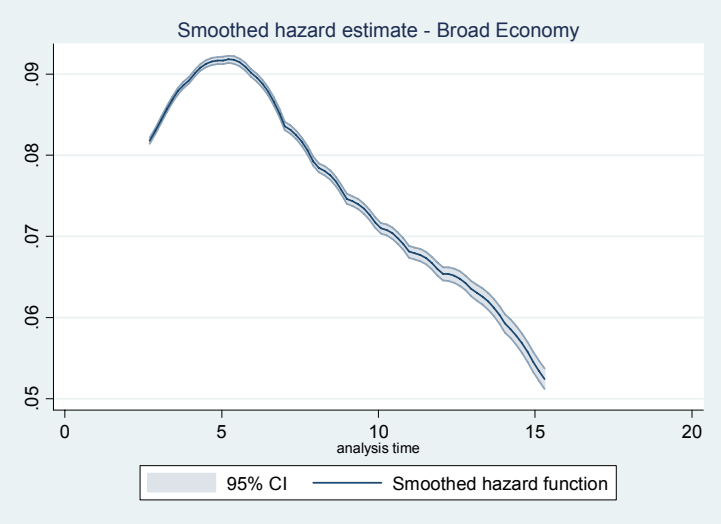
Source: Own calculations based on *Quadros de Pessoal*, GEP, MTSS.
Notes: * Approximate values.

Probability of survival calculated according to the Kaplan-Meier survival function is based on the following formula: $\hat{S}(t) = \prod_{j|t_j \le t} \left(\frac{n_j - d_j}{n_j} \right)$.

The Nelson Aalen Hazard Rate, or the risk associated to the probability of death is calculated according to the following formula: $\overline{H}(t) = \sum_{j|t_j \le t} \frac{d_j}{n_j}$, described in section 2.1.

In Figure 1, we depict the smoothed hazard estimate or unconditional hazard function for the total economy. This function exhibits an inverted U-shape, with a maximum around the sixth year of activity (Figure 1). This means that, after a firm enters the market, the conditional probability of failure increases continuously until the sixth year. After the sixth year, the hazard rates decline steeply. Such pattern is similar to that found in other economies, such as Italy (Audretsch et al., 1999), the UK (Bhattacharjee, 2005), Germany (Wagner, 1994), UK, Italy and the US (Bartelsman et al., 2005) and Spain (López-Garcia and Puente, 2006). In all these cases, the maximum of the unconditional hazard function is reached before the sixth year, indicating that Portuguese firms keep on failing for a longer period, before the hazard rate starts declining.

Figure 1 – Smoothed hazard estimate for the total economy, 1987-2005 (%)



Source: Own calculations based on *Quadros de Pessoal*, GEP, MTSS.

5. Survival and hazard functions across regions

Table 2 presents the results for the non-parametric estimation, for each of the seven Portuguese NUTII regions. This framework explores the relationship between age and the regional hazard of exit.

In line with the results shown previously for the total economy, over 85% of newly born employer enterprises remain active during their first year of activity in all regions. The one-year survival rate varies from a low of 85% in the Açores, to a high of 87,5% in the Centro region, meaning that the new born enterprises died more prematurely in Açores than in other Portuguese regions.

Table 2 also reveals that the survival gap between the two extreme regions grows systematically with time. Within 6 years of activity, the region Norte is the only one with less than 50% of enterprise survival probability, lagging behind all other regions in terms of enterprise survival.

On the other hand, Centro has a higher survival rate than the economy's average. It is the region where more firms manage to survive longer throughout the period considered in this study.

There are also clear disparities between regions, in particular between Norte and Centro, in terms of median duration survival. At the end of the analysis period, Norte is the region that presents the lowest survival rate, with only 20,7% of the firms' population managing to survive after eighteen years of activity. In Centro, in turn, 27,4% of active start-ups are still alive after 18 years.

Table 2 - Survival Table for Employer Enterprise Births by NUTII region, 1987-2005

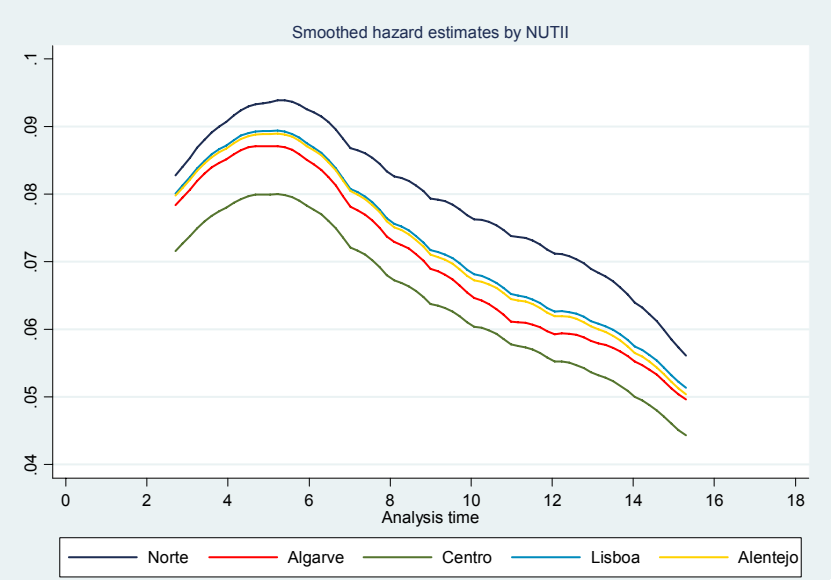
Time	Norte	Centro	Lisboa e Vale do Tejo	Alentejo	Algarve	Açores	Madeira
1	85,6%	87,4%	85,5%	85,8%	85,6%	85,1%	86,1%
2	75,1%	77,7%	75,1%	75,5%	75,5%	74,2%	76,0%
3	66,5%	70,1%	67,0%	67,0%	67,7%	67,0%	68,3%
4	59,8%	64,0%	60,5%	60,4%	61,2%	59,9%	61,3%
5	54,1%	58,9%	55,0%	54,9%	55,8%	54,5%	56,2%
6	49,4%	54,5%	50,4%	50,7%	51,1%	50,5%	51,6%
7	45,3%	50,7%	46,6%	46,9%	47,2%	46,7%	47,5%
8	41,7%	47,5%	43,2%	43,4%	44,2%	43,7%	44,6%
9	38,7%	44,5%	40,2%	40,5%	41,1%	41,2%	41,7%
10	35,8%	41,9%	37,6%	37,7%	38,5%	38,9%	38,6%
11	33,0%	39,5%	35,1%	35,2%	36,2%	36,3%	36,6%
12	30,5%	37,4%	32,8%	33,0%	34,0%	33,9%	34,3%
13	28,1%	35,3%	30,8%	31,0%	32,0%	31,3%	31,7%
14	26,4%	33,4%	29,0%	29,3%	30,2%	29,4%	29,9%
15	24,8%	31,8%	27,4%	27,8%	29,0%	28,2%	28,2%
16	23,2%	30,4%	26,1%	26,2%	27,8%	26,4%	26,9%
17	21,9%	28,9%	24,6%	24,9%	25,4%	25,4%	26,6%
18	20,7%	27,4%	22,9%	23,2%	23,9%	23,8%	25,4%

Source: Own calculations based on *Quadros de Pessoal*, GEP, MTSS.

The median duration of firms at the regional level (Figure 2), is below seven years for most regions, except for Centro (around the eight year).

The disparities among the Portuguese regions are confirmed by equality tests. Both Log-rank and Wilcoxon (Breslow) tests allow for the rejection of the hypothesis of survival equality among regions[4].

Figure 2 – Smoothed hazard estimate by NUTII, 1987-2005 (%)



Source: Own calculations based on *Quadros de Pessoal*, GEP, MTSS.

6. Survival and hazard functions across size classes

A general finding in the literature is that most firms start small, live small and die small. According to Eurostat (2009), Portugal has the highest share of enterprises births in the 1 to 4 employees' size class. Small firms in Portugal are also being created at a faster pace than larger firms, gaining share in both enterprise and employment (Sarmento and Nunes, 2009).

[4] The hypothesis being tested considers that there are no subgroup differences in survivor functions. We find the probability that the observed differences occur by chance is below 0,0. This piece of evidence is not included in the present work, but is available at request.

We find that smaller firms exhibit the lowest survival probability (Table 3). More than 15% of micro firms with fewer than 5 employees "die" in the first year of activity (only around 85% manage to survive), whereas large firms with over 250 employees, have a much higher survival rate, of 93,9%. Differences between size classes are significant. Conditional on overcoming the first ten years, the smallest sized firms are the only ones to have a survival probability below 50% (31% for the 1 to 4 size class). Over time, the gap between the smallest and the largest firms' survivor rates widens. The bigger the firm, the higher the probability of survival.

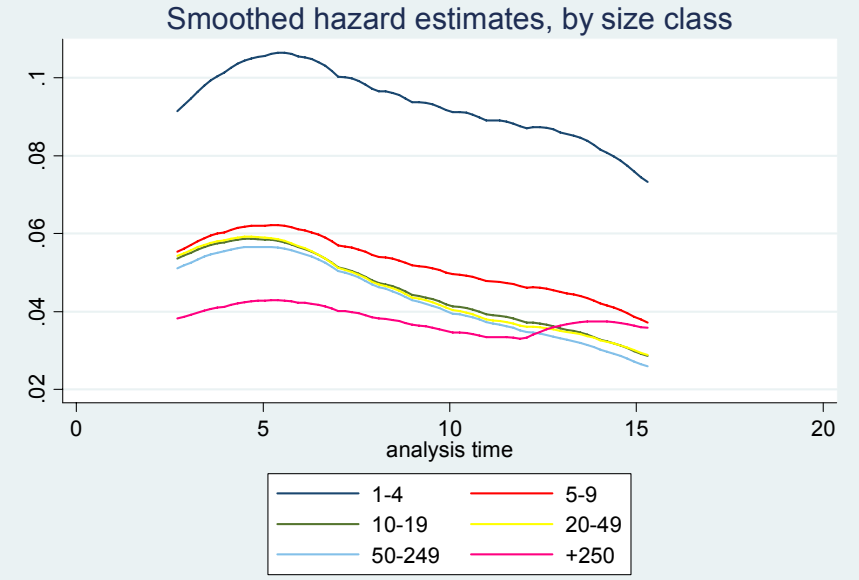
Table 3 - Survival Table for Employer Enterprise Births by size class, 1987-2005

Time	1-4	5-9	10-19	20-49	50-249	+250
1	84,9%	90,0%	90,1%	89,8%	92,0%	93,9%
2	73,4%	82,6%	82,8%	82,7%	84,7%	87,6%
3	64,3%	76,5%	76,9%	76,6%	78,6%	82,8%
4	57,0%	71,2%	71,7%	71,3%	73,2%	78,9%
5	51,0%	66,7%	67,2%	66,8%	68,9%	74,6%
6	45,8%	62,9%	63,5%	62,9%	64,8%	71,9%
7	41,3%	59,5%	60,6%	60,3%	62,0%	70,7%
8	37,5%	56,5%	58,0%	57,7%	59,7%	69,1%
9	34,1%	53,8%	55,7%	55,4%	57,1%	65,8%
10	31,0%	51,1%	53,5%	53,2%	55,0%	62,4%
11	28,2%	48,6%	51,4%	51,3%	52,8%	60,6%
12	25,7%	46,2%	49,4%	49,5%	51,1%	57,6%
13	23,4%	43,9%	47,5%	47,4%	49,1%	55,9%
14	21,4%	42,0%	45,8%	46,0%	47,8%	54,7%
15	19,7%	40,5%	44,4%	44,5%	46,4%	54,0%
16	18,2%	39,0%	42,9%	43,5%	44,9%	52,2%
17	16,7%	37,4%	41,9%	41,9%	44,0%	50,6%
18	15,1%	35,8%	40,8%	40,4%	42,9%	43,8%

Source: Own calculations based on *Quadros de Pessoal*, GEP, MTSS.

Differences in hazard rates across firm size classes are particularly evident in the early stages of a firm's life (Figure 3). The regional disparity, observed in the previous section, is also confirmed among different size classes. The equality tests performed allow the acceptance of the hypothesis that firms present distinct survive performances according to their dimension. The largest size class reveals some deterioration in its survival capacity after the 12th year of activity, depicted by the "overshooting" of the hazard estimation function.

Figure 3 – Smoothed hazard estimate by size class, 1987-2005 (%)



Source: Own calculations based on *Quadros de Pessoal*, GEP, MTSS.

7. Survival and hazard functions across broad sectors

Our analysis now turns to the question of whether failure rates vary according to industry membership. Table 4 shows survival rates at different lifetimes across broad sectors for a period of ten years (after 1995 due to the start of European System of Accounts of 1995, and up to 2006 due to the problems of compatibility with Classification of Economic Activities Revision 3, introduced in 2007).

Enterprises operating in the construction sector have the lowest survival probabilities over all this time period and show the greatest survival gap between the first and its tenth year of activity (a decrease of 55,1 p.p.). Its hazard peak is reached within the first 4 years of activity (Figure 4), but survival tends to decline faster than in other sectors. On the other hand, the agriculture sector has had the highest survival rates up to the fourth year of activity.

Table 4 – Survival table for employer enterprise births by broad sectors, 1995-2006

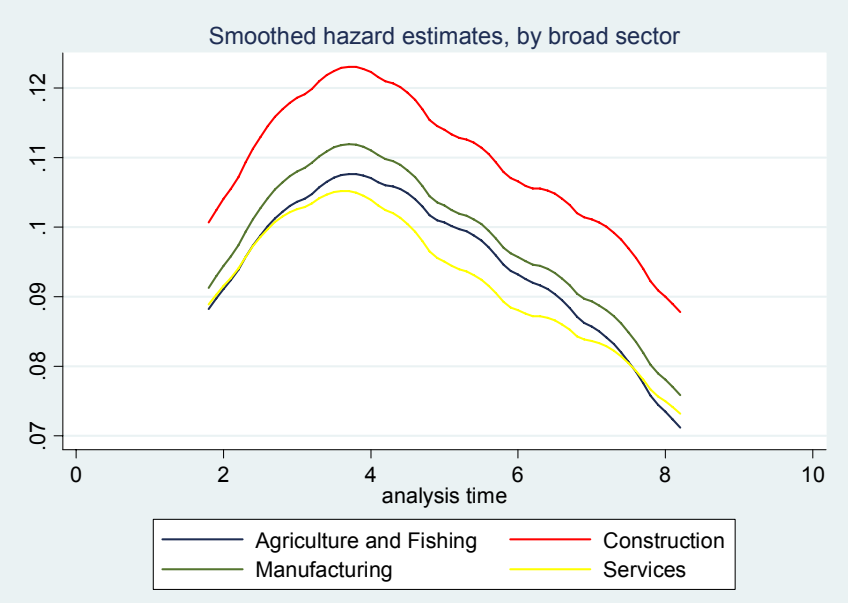
Time	Agriculture and Fishing	Construction	Manufacturing	Services
1	86,5%	84,8%	86,4%	85,9%
2	76,1%	73,3%	75,7%	75,4%
3	67,9%	63,8%	66,9%	66,9%
4	60,3%	56,0%	59,4%	60,1%
5	54,3%	49,4%	52,8%	54,5%
6	49,3%	44,2%	47,7%	49,8%
7	44,8%	39,8%	43,9%	45,7%
8	40,9%	36,3%	40,4%	42,3%
9	38,8%	33,1%	37,2%	39,1%
10	36,1%	29,7%	34,2%	35,7%

Source: Own calculations based on *Quadros de Pessoal*, GEP, MTSS.

The smoothed hazard estimate shows that in the service sector, the probability of "death" increases steeply within the first three years, but the hazard peak is reached sooner than in other sectors. Following this point, an increase in age, brings about a flatter probability of failure at the lower end of the distribution (Figure 4).

The existence of disparities among the Portuguese regions is also confirmed by the equality tests performed. Both Log-rank and Wilcoxon (Breslow) tests allow for the rejection of the hypothesis of survival equality across broad sectors.

Figure 4 – Smoothed hazard estimate by broad sectors (%)



Source: Own calculations based on *Quadros de Pessoal*, GEP, MTSS.

8. Final remarks

In our analysis, we find that around 25% of enterprises entering the market fail within the first 2 years of activity and that more than 50% fail within a period of six years. We have also found that the instantaneous probability of exit is monotonically decreasing with firm size and that, after entry, the conditional probability of failure increases continuously up to the sixth year of activity.

Breaking down by region, sector and class dimension, we identify statistically significant disparities. As for the regional dimension, it is worth noting the disparities in terms of median duration survival, in particular between Norte and Centro. Within the first 6 years of activity, the Norte is the only region registering less than 50% of enterprise survival probability, lagging behind all other regions in terms of enterprise survival, while Centro is the region where firms survive longer throughout the period considered. We also observed that the survival gap between the Norte and Centro has been systematically increasing during the period .

As for the firm dimension, we found a significant relationship between size and chance of survival. This is particularly observable for new start-ups, who face the greatest uncertainty regarding market conditions (this accords to Jovanovic, 1982, who stresses post-entry learning as a fundamental determinant of firm performance and survival).

At the sectoral level, we find that firms in the construction sector exhibit the highest risk of failure. Firms in the service sector, in turn, display the highest survival rates. The services sector also exhibits a tendency for the hazard peak to be reached sooner, which means that chances of survival relating to firm age, start increasing sooner than in other broad sectors.

BIBLIOGRAPHY

Audretsch, D., E. Santarelli, and M. Vivarelli (1999) "Start-up size and industrial dynamics: some evidence from Italian manufacturing," International Journal of Industrial Organization, Elsevier, vol. 17(7), pages 965-983, October.

Bhattacharjee, A. (2005) "Models of Firm Dynamics and the Hazard Rate of Exits: Reconciling Theory and Evidence using Hazard Regression Models," Econometrics 0503021, EconWPA.

Bartelsman, E.J., Scarpetta, S. and Schivardi, F. (2005) "Comparative Analysis of Firm Demographics and Survival: Evidence from Micro-level Sources in OECD Countries", Industrial and Corporate Change, Vol 14, No. 3, pp 365–391.

Eurostat (2009) "Business Demography: employment and survival", Statistics in focus, No. 70/2009

Eurostat and OECD (2007) "Eurostat/OECD Manual on Business Demography Statistics".

Jovanovic, B. (1982) "Selection and evolution of industry", Econometrica, 50 (3), pp. 649-670.

López-Garcia, P. and Puente, S. (2006) "Business demography in Spain: determinants of firm survival", Documentos de Trabajo No. 608, Banco de España.

OECD/Eurostat (2009), "Measuring Entrepreneurship A Collection of Indicators 2009 Edition", OECD-Eurostat Entrepreneurship Indicators Programme.

OECD/Eurostat (2008), "Measuring Entrepreneurship: A digest of indicators", OECD-Eurostat Entrepreneurship Indicators Programme.

Sarmento, E. and A. Nunes (2010) “Getting smaller: size dynamics of employer enterprises in Portugal” in “Cadernos Sociedade e Trabalho”, Ministério do Trabalho e da Solidariedade Social (forthcoming).

Wagner, J. (1994) “The Post-Entry Performance of new Small Firms in German Manufacturing Industries”, The Journal of Industrial Economics XLII., 125-131.