

Ensaio

A Semi-Parametric Survival Analysis of Business Demography Dynamics in Portugal

Alcina Nunes²³
Elsa de Morais Sarmento²⁴

Abstract: We address the post-entry performance of new Portuguese firms by investigating the structural characteristics of the hazard and survival functions, using semi-parametric survival analysis. It is based on the application of the entrepreneurship definitions and methodology of the “Manual on Business Demography Statistics” (OECD/Eurostat, 2007) to the *Quadros de Pessoal* dataset. The main contribution of this work is the application of a recent internationally comparable methodology for entrepreneurship to provide a multidimensional overview of firm and survival dynamics over a period of eighteen years, while guaranteeing international comparability. By correcting heterogeneity with a different specification model, we obtain stronger magnitudes of the hazard ratios found previously.

1. Introduction

This work addresses the post-entry performance of new Portuguese firms by investigating the structural characteristics of the hazard and survival functions, using semi-parametric survival analysis.

It is based on the application of the entrepreneurship definitions and methodology of the Manual on Business Demography Statistics (OECD/Eurostat, 2007) to the *Quadros de Pessoal* dataset (Employment Administrative Records by the Portuguese Ministry of Labour and Social Security), which is the main data source in Portugal for the universe of employer enterprises. Accordingly, the analysis focuses on a specific subsample, consisting on the population of active enterprises only, with at least one paid employee. These are the so-called “employer active enterprises”. Entrepreneurship performance indicators were then calculated, following the work of the Entrepreneurship Indicators Programme (OECD, 2008). This allowed the computation of a comprehensive array of entrepreneurship indicators on employer enterprise and survival dynamics in Portugal, over a period of 18 years, disaggregated in dimensions such as sectors, regions and size classes.

Most empirical studies on regional variations in entry and exit rates at the international level are either based on survey data like the Global Entrepreneurship Monitor (Acs et al., 2008), business data, business registration data (Klapper et al., 2008; Klapper et al., 2009) or a mix of the previous (Baterlsman et al., 2005a; Baterlsman et al., 2005b). Moreover, most only take into account the manufacturing sector. There is scarce evidence of studies on entrepreneurial activity that encompass simultaneously all sectors, regions and countries. Portugal is somehow an exception, where extensive research has been done in firm dynamics using mostly *Quadros de Pessoal* (Mata and Portugal, 1994; Mata et al., 1995; Mata, 1993; Mata and Machado, 1996; Baptista et al., 2008; Cabral, 2007; Cabral and Mata, 2003; Baptista and Carias, 2007; Baptista and Mendonça, 2007).

Following a brief description of the dataset and core definitions, the next section presents a semi-parametric analysis of survival in Portugal, where estimations for the total economy and broad sectors are

²³ Escola Superior de Tecnologia e Gestão do Instituto Politécnico de Bragança, Portugal.

²⁴ 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 for the provision of data and the helpful assistance.

provided, as well as estimation results which take heterogeneity into account. Finally, the last section concludes.

2. Dataset and Core Definitions

The *Quadros de Pessoal* (Employment Administrative Records) is an annual survey conducted in Portugal by the Portuguese Ministry of Labour and Social Security (*Gabinete de Estratégia e Planeamento do Ministério do Trabalho e da Segurança Social*), which provides a rich and comprehensive matched employer-employee dataset. It is of the utmost importance for most research purposes concerning Portuguese labour market analysis, characterisation of labour market qualification structure, as well as for the study of both employer and employee characteristics and linkages, in several areas of scientific research, namely in the entrepreneurship research field. The empirical literature on entrepreneurship refers explicitly, the importance of working with such a linked employee-employer database.

Indeed, linked firm-level data is fundamental to answer questions about the relationships between entrepreneurial determinants and entrepreneurial performance, at several levels, since it allows to follow individual firms for a particular period of time observing their overall characteristics and related changes: identification, location, main activity, legal identity and year of legal birth, stock capital, turnover and number of establishments and employees. The availability of longitudinal datasets is also extremely relevant for a time-series analysis of entrepreneurship, in terms of the performance and survival of specific cohorts of newly created firms over time.

The entrepreneurship database obtained from the *Quadros de Pessoal*, following the Eurostat/OECD (2007) methodology, consists of an annual average of 215,903 active employer enterprises over the period 1985-2007, with an annual average of 36,803 births and 23,743 deaths.

The survival analysis provided in the following sections, will take place over this new entrepreneurship dataset, where only *real* births and deaths are accounted for. Thus, the core measure of births reflects the concept of employer enterprise birth. A birth amounts to the “creation of a combination of production factors with the restriction that no other enterprises are involved in the event” (Eurostat/OECD, 2007). A birth occurs when an enterprise actually 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. This population thus 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. The employer enterprise birth rate is based on a ratio where the numerator follows the above definition for employer enterprise births, while the denominator is the population of active enterprises with one or more employees during the reference period.

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. We have tried to identify those situations in order to remove them from the population, according to Eurostat/OECD’s methodology. 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.

The churn rate is an indicator which is used for the measurement of turbulence. It is viewed as an economy’s ability to expand and adjust its structure of production to the market’s changing needs and is given by the sum of birth and death rates (Eurostat/OECD, 2007).

3. A Semi-Parametric Survival Analysis

Both seminal and most recent literature agrees that size affects the survival rates of new firms (Mata et al., 1995; López-García and Puente, 2006), which has generated one of the most striking stylized facts in the literature of industry dynamics (Audretsch and Mahmood, 1994). Concerning the concept of initial firm size (corresponding to the size at the time of a firm's birth), several studies have reported that the probability of firm exit from the market decreases with initial size. Large firms experience higher survival rates than smaller units. According to the literature, there are several reasons behind this. The most prevalent relate to the efficient scale needed to operate efficiently in a market, to the capital intensity production technology, to the firms' capacity to access financial markets and to the management ability of small entrepreneurs.

Regarding the first reason, Audretsch and Mahmood (1994) have considered that larger firms are more likely to be closer to the necessary minimum efficient scale to operate efficiently in a market. Frequently, entering small avoids big losses, as firms expecting a good performance usually start up larger. Even if larger firms find themselves to be less efficient than they had expected, they may become smaller before they do exit the market (Mata and Portugal, 2004). Additionally, larger firms diversify more than smaller ones, which also contributes to reduced market risks.

Moreover, the stock of capital accumulated by firms should also be considered. Small firms are in general less capital intensive, so variable costs represent a larger share of capital costs. Despite the different cost structure of small firms allowing extra flexibility to market fluctuations, it does not prevent them to be the first to exit the market in more severe periods of economic downturn.

Thirdly, internal financial constraints and internal capital markets imperfections are also commonly pointed out as reasons for the smaller size of entrants. Firms enter small not because they choose to, but because new firms underinvest as they are financially constrained, which leads to a negative impact on firms' survival probabilities (López-García and Puente, 2006). In the presence of market instability, all this also accrues to these new smaller firms not being able to sustain their market positions for long periods of time (Mata and Portugal, 2004).

The last reason pointed out previously relates to the entrepreneur management ability. It is often considered that due to intrinsic reasons and incentives, smaller firms employ less able managers, who can more easily abandon the market. In fact, being an entrepreneur has higher opportunity costs when the economy's wages grow, and lower quality managers are more likely to miscalculate their true value label (Mata and Portugal, 2004). Geroski, Mata and Portugal (2003) quoting Lucas (1978), refer that worse management capabilities often translates into bigger costs, for any given firm size, those leading to firms choices towards operating at a smaller scale.

Mata, Portugal and Guimarães (1995) and Geroski, Mata and Portugal (2003) underline the previous observations relating to the importance of initial firm size in explaining the survivor probability of firms. However, they argue that current size is a better predictor of failure than initial size. After controlling for initial size, measuring current size amounts to measuring firm performance. According to them, the fact that a firm has grown in the past, signals that it has been performing well and therefore its probability of exit is low. Moreover, Mata, Portugal and Guimarães's (1995) findings indicate that after controlling for size differences, past growth matters for survival, suggesting a partial adjustment process of firm size in the post-entry period. Although accepting their arguments López-García and Puente (2006) highlight the fact that current size could be endogenous to the firm dynamics, since firms that are about to abandon the market, grow smaller before exiting and vice-versa.

Next, we present an overview of the theoretical foundations of the Cox Proportional Hazard Model.

3.1. Modeling with the Cox Proportional Hazard Model

The statistical representation of the relation between the survival time of a firm and specific variables is known as the hazard rate model of the duration of the life of a firm. According to the model a given firm j faces a hazard rate (h_j) that is a function of a baseline hazard rate (h_0), which all firms face, transformed by a set of explanatory variables (X) through a vector of parameters (β). The hazard rate model can be written in the form $h_j(t) = f(h_0(t), \phi(X, \beta))$.

In this model, two firms with the same birth date will face a different hazard function if, and only if, their other characteristics are different. By definition, the model seems a natural solution to understand the temporal pattern of survival and to identify the covariates that could be related significantly to survival. Additionally, it is also a good solution for working with longitudinal datasets, characterized by right censored data and other types of selection issues.

An empirical application of the model implies the specification of a functional form for the hazard function. One of the most common options is the proportional hazard model: $h_j(t) = h_0(t)\phi(X, \beta)$. In this specification, the shape of the hazard function is the same for all individuals, and variations in the explanatory variables will translate into parallel displacements of this function, thereby affecting only the scale of the hazard function and not its shape. Given the fact that the hazard is a conditional probability and, therefore, must be positive, a convenient functional form for $\phi(X, Y)$ is exponential.

Hence the hazard a subject j faces is written in the following form: $h_j(t) = h_0(t)e^{(X \cdot \beta)}$. Note that this particular functional form offers the advantage of a very convenient interpretation of the estimated coefficients, since $\beta = \frac{\partial \ln \phi(X, \beta)}{\partial X}$. This means that the coefficient of one explanatory variable is the constant proportional effect of a unit increase of this variable on the conditional probability of exiting.

The assumption made for the functional form of $\phi(X, Y)$ is widely accepted, the same does not happening for the functional form of the baseline hazard, since different parametric specifications of the hazard function display different duration dependence behaviours. Positive (negative) duration dependence implies that the likelihood of failure (exit) at time t , conditional on the duration up to t , is increasing (decreasing) in t . *A priori* it is not obvious which distribution is most appropriate even when economic theory provides some clues concerning the way the baseline hazard varies over time. In case of doubt, one line of action to consider is to make no assumption about the functional form of the baseline hazard. Such a method was first suggested by Cox (1972) and the resulting models are called semi-parametric. Cox (1972) also suggested that the proportional hazard model could be easily extended, to account for time varying covariates. This is what we will approach next.

The model incorporates the main features of discrete duration models, as described by Lancaster (1990), where the logarithm of the probability that a firm exits at time t given that it survived in $t-1$ is explained by a series of explanatory covariates X_{t-1} plus a set of parameters identifying the baseline hazard function, according to the following specification:

$$\log h(t | x_t, x_0) = \lambda_t + \beta x_t + \gamma x_0, \text{ for } t = 1, \dots, k$$

The use of the partial likelihood function does not require that $h_0(t)$ must be specified, which allows the estimation of β and γ and avoids the risk of misspecifying the baseline hazard function. The model described previously, considers two types of heterogeneities that may cause exit, and that need to be considered: current heterogeneities between firms, that is heterogeneities based on differences that exist in period t , and heterogeneities that occur from differences that existed in the moment when firms were created ($t=0$). Heterogeneities due to differences in founding conditions include those conditions that are cohort specific, i.e., which take a common value for all firms in the same cohort, such as macroeconomic or industry-wide factors and those which are firm-specific (Baptista and Mendonça, 2007).

In our case (and in López-García and Puente's, 2006) the survival is a continuous phenomenon, but the available information is reported annually in the month of October, transforming time in a discrete variable. To circumvent this, we have grouped the data, by creating 11 interval specific dummy variables (one for each spell year at risk) and shall be using a discrete hazard model. The most common discrete time representation of an underlying continuous time Cox proportional hazard model is the complementary log-log (cloglog model), which is what will be used in the following estimations. The major advantage of using the hazard model is that each firm contributes several times to the likelihood function, each time it is at risk.

3.2. Explanatory Variables

In the estimation framework, we have considered seven explanatory variables and sector and year dummies (Table 1). The first explanatory variable is the firm start-up size. It is measured by the logarithm of the number of employees at the firm's year of birth. We expect a negative influence on the hazard rate, that is, larger start-ups should face a reduced risk of survival. The second variable relates to the number of employees reported at the year of measurement. Besides these two firm characteristics, the specific conditions of the industry are also likely to affect firm survival (López-García and Puente, 2006). Among the measures of firm dynamics, it is important to control for industry entry and growth rate and its degree of competition. Thus, the third variable has to do with the firm's entry rate. New firms are more likely to live longer if they enter expanding industries or industries with low entry activity (Mata, Portugal and Guimarães, 1995).

Another important industry characteristic is the degree of competition intensity, which is measured through the Herfindhal-Hirschman Index (HHI). Highly concentrated industries may allow suboptimal scale of new firms and therefore provide more opportunities for survival after entry. On the other hand, according to the industrial organization literature, highly concentrated industries might as well represent a higher potential for incumbent's collusion and therefore display a more aggressive behaviour towards new entries (Mata and Portugal, 1994, López-García and Puente, 2006).

By definition, at start-up there is no post-entry growth. The effect of growth can only be perceived as firms age and current size shifts from initial size. At any time after start-up, current size can be viewed as initial size plus the change in size which occurred. As size is measured in logs, this change is the cumulative growth rate since start-up. Therefore, after controlling for the effect of start-up size, the coefficients associated with the current size offer an estimate of the effect of post-entry growth (Mata, Portugal and Guimarães, 1995).

Turbulence is a measure of firm dynamics and is given by the sum of birth and death rates (Eurostat/OECD, 2007). It is the pace at which firms are starting up and closing down and is a common measure of the level of entrepreneurial activity in an economy.

Table 1 -Explanatory Variables Considered in the Model

| Variable | Definition | Measurement |
|---------------------|---|---|
| Start-up Size | Number of employees at the birth year of the firm. | Logarithm of the number of employees. |
| Current Size | Number of employees at the current year. | Logarithm of the number of employees. |
| Industry Entry Rate | Industry entry rate calculated for sectors defined at a 2-digit CAE level. | Logarithm of the industry entry rate, defined as the number of entrants divided by the total number of firms in industry. |
| Concentration (HHI) | Herfindhal-Hirschman Index (HHI) calculated for industries at a 2-digit CAE level. | Logarithm of the HHI. |
| Growth | Logarithmic difference of industry employment in two consecutive periods. | Logarithm of the number of employees at year t minus the logarithm of the number of employees at year $t-1$. |
| Entry Rate X Growth | Interaction variable, defined as the product of entry and growth. | Product of logarithms. |
| Turbulence | Sum of entry and exit rates calculated for sectors defined at a 2-digit CAE level. | Sum of logarithms of the industry entry rate with the industry exit rate. |
| Sector Dummies | Dummies for 4 broad sectors: Agriculture, Construction, Manufacturing and Services. | ----- |
| Year Dummies | Dummies for each current year. | ----- |

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

Note: * The literature has shown that there is a non linear effect of the start-up size on survival, which is normally accounted for via a log transformation. The specification is reasonable given that the value of the likelihood increases.

There may well be differences in survival rates between industries beyond those captured by the industry-specific variables. For this reason, industry dummy variables are also included in the analysis. Finally, since the overall state of the economy has long been indicated as an important force driving firms out of business, we also include year dummies, so as to control for the macroeconomic environment (López-García and Puente, 2006; Mata, Portugal and Guimarães, 1995).

3.3. Estimation Results for the Total Economy

Table 2 shows the cloglog regression results for the total economy, using different specifications of the model. In the second column (Model 1), the model estimates the proportional hazard function for firm and industry start-up conditions. The idea is to avoid the introduction of possible endogenous effects with the presence of variables such as the current size or the industry's annual growth. These variables are then included in the model presented in column 3 (Model 2). Both models control for broad industry dummies and for macroeconomic effects through year dummies. The year dummies values have also been introduced but the values are not shown, as usually no clear pattern can be discernible from the estimated coefficients (Mata, Portugal and Guimarães, 1995).

The values presented below are the hazard ratios, that is, the ratio of the hazard rate when the variable increases by one unit. A hazard ratio over one implies that an increase in the given explanatory variable increases the probability of exit and, correspondingly, a hazard ratio below one means that an increase in the variable decreases the hazard probability.

As argued in the literature and mentioned previously, the start-up size of a firm improves the changes of survival. Smaller firms are the most likely to exit, probably because they are the least efficient. The coefficient of the start-up size in the first estimated model is illustrative. However, when we observe the second model, the effect of a firms' current size seems to be predominant. When introducing the sum of the start-up and the current size (by denoting S_0 and S_t the initial and current size, respectively, and by α and β the correspondent coefficients, the effect of size is expressed by $\alpha S_0 + \beta S_t$), it becomes evident that the current size improves the chances of survival and that the initial size does not. This result is consistent with the results of Mata, Portugal and Guimarães (1995). According to the authors, firms that have started smaller and have experienced faster post-entry growth, face a higher probability of survival. Indeed, our overall effect is line with the previous authors' results.

The results for the first model also indicate that in industries characterized by high entry rates, at the moment of birth, post-entry survival is more difficult. Firms that experience more competition from entrants, have a higher probability of failure. A higher entry rate combined with fast growth rates for any given industry generates, in general, a shorter duration of firms (Mata, Portugal and Guimarães, 1995 and Gort and Klepper, 1982). This somehow expected piece of evidence can be also drawn from our results. It might seem easier to enter the market in earlier stages of the product life-cycle, when markets are expanding, but it becomes particularly to secure survival.

So far, all our results have stressed the literature's conclusions. However, we draw different conclusions from the effect of industry growth results (Model 2). What we would expect is that firms operating in fast growing industries, would suffer from a smaller probability of failure (since they can more easily be established the market without harming the competitors), but our results show otherwise.

To help clarifying this result, it should be pointed out that industries in the early stages of their life-cycles usually register both high rates of entry and exit (Agarwal and Gort, 1996 and Baptista and Karaoz, 2007). In general, industries with higher than average entry rates, also exhibit higher than average exit rates (Cabral, 2007), due to birth and death rates being highly correlated across industries, corroborating the idea that "entry barriers are exit barriers" (Mata et al., 1995). The combined effect of entry and growth could explain this unexpected effect of industry growth on survival probabilities. Industries experiencing higher growth rates are also more turbulent, registering high rates of entry and also of exit (the "revolving door" at work), thus decreasing the likelihood of survival.

Table 2 - Estimation Results

| Variable | Model (1) | Model (2) |
|--|----------------------|----------------------|
| Log of Start-up Size | 0.692 *** (0.002) | 1.339 *** (0.007) |
| Log of Current Size | ----- | 0.588*** (0.002) |
| Industry (2 digit) Start-up entry rate | 1.353 *** (0.013) | 1.42*** (0.018) |
| Start-up Industry HHI (2 digit) | 0.986 *** (0.001) | 0.881*** (0.001) |
| Industry Growth (log) | ----- | 1.326*** (0.018) |
| Growth x Entry rate | ----- | 1.966** (0.017) |
| Turbulence | ----- | 4.195 *** (0.193) |
| <u>Sector Dummies</u> | | |
| Agriculture | (a) | (a) |
| Construction | 1.130 *** (0.012) | 1.405 *** (0.021) |
| Manufacturing | 1.123 *** (0.013) | 1.567 *** (0.024) |
| Services | 0.951 *** (0.010) | 1.115*** (0.016)5 |
| Year Dummies | Yes | Yes |
| Number of firms | 447772 | 447772 |
| LR X2 | 42744.42 *** | 46506.55 *** |
| Log likelihood | -655716.05 | -423334.96 |

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

Note: (a) refers to the reference sector.

The year dummies "yes" means that they have been included in the estimation. Standard deviation is shown in brackets and *, **, *** means, respectively, 10, 5 and 1% level of significance.

Heterogeneity is also found in the main model. By correcting with a different specification model, we obtain stronger magnitudes of the hazard ratios found previously (Appendix).

4. Conclusions

We find that firms that start small and experience faster post-entry growth, face a higher probability of survival. Firm's current size dimension is also extremely important to determine the probability of survival, particularly in the Services sector. A higher entry rate combined with fast growth rates for any given industry also generates a shorter duration of firms. It might seem easier to enter the market in earlier stages of the product life-cycle, when markets are expanding, but it becomes particularly difficult to survive. Firms that experience more competition from entrants, also face higher probabilities of failure.

However, we find a different result from the literature, for the effect of industry growth in survival rates. Firms operating in industries which are growing faster seem to suffer from a higher probability of failure. The combined effect of entry and growth can also contribute to explaining this unexpected effect of industry growth on survival probabilities. This has to do with turbulence and the high rates of entry and exit verified in Portugal throughout this period. Industry growth, by sector, reinforces this view, and the

interaction between high industry growth rates and entry rates at the start-up moment, seems to unfold the reasons for the negative impact on firms' survival. By correcting heterogeneity, we obtain stronger magnitudes of the hazard ratios.

References

- Acs, Z., S. Desai and Klapper L. (2008) "What does "Entrepreneurship" Data Really Show?, a Comparison of the Global Entrepreneurship Monitor and the World Bank Group Datasets", World Bank, Policy Research Working Paper Series 4467.
- Audretsch, D. (1995) "Innovation, Growth and Survival", *International Journal of Industrial Organization*, n. 13, p. 441-457.
- Audretsch, D. B. and Mahmood T. (1994) "The Rate of Hazard Confronting New Firms and Plants in U.S. Manufacturing", *Review of Industrial Organisation*, vol. 9, p. 41-56.
- Baptista, R., Escária, V. and Madruga, P. (2008) "Entrepreneurship, Regional Development and Job Creation: the Case of Portugal." *Small Business Economics*, n. 30, p. 49-58.
- Baptista, R. and Carias, C. (2007) "Job Creation and Destruction by Small Businesses in the Portuguese Regions", IN+, presented at the European Regional Science Association, ERSA 2007.
- Baptista, R. and Mendonça, J. (2007) "Short and Long Term Survival of New Businesses: the Role of Human Capital", IN+, Draft version of 2007.
- Bartelsman, E.J., Scarpetta, S. and Schivardi, F. (2005a) "Comparative Analysis of Firm Demographics and Survival: Evidence from Micro-level Sources in OECD Countries", *Industrial and Corporate Change*, vol 14 (3), p. 365–391.
- Bartelsman, E.J., Haltiwanger, J. and Scarpetta, S. (2005b) "Measuring and Analyzing Cross Country Differences in Firm Dynamics," (co-authored with Eric Bartelsman and Stefano Scarpetta), in *Producer Dynamics: New Evidence from Micro Data*, (Dunne, Jensen and Roberts, eds.) NBER/University of Chicago Press.
- Brito, P. and Mello, A. (1995) "Financial Constrains and Firm Post-Entry Performance", *International Journal of Industrial organization*, n.13, p. 543-566.
- Cabral, L. (2007) "Small firms in Portugal: A selective Survey of Stylized Facts, Economic Analysis and Policy Implementation", *Portuguese Economic Journal*, vol. 6 (1), p. 65-88.
- Cabral, L. and Mata, J. (2003) "On the Evolution of the Firm Size Distribution: Facts and Theory", *The American Economic Review*, vol. 93(4), p. 1075 - 1090.
- Cox, D.R. (1972) "Regression Models and Life Tables", *Journal of the Royal Statistical Society, Series B*, vol 34(2), p. 187-202.
- Dunne, T., Roberts, M. J. and Samuelson, L. (1989) "The Growth and Failure of U.S. Manufacturing Plants," *The Quarterly Journal of Economics*, MIT Press, vol. 104(4), pages 671-98, November.
- Eurostat/OECD (2007) "Eurostat/OECD Manual on Business Demography Statistics".
- Geroski, P.A., Mata, J. and Portugal, P. (2003) "Founding Conditions and the Survival of New Firms", DRUID Working Paper n. 07-11.
- Gort and Klepper (1982) "Time Paths in the Diffusion of Product Innovations", *Economic Journal*, n. 92, p. 630-653.
- Jovanovic, B. (1982) "Selection and evolution of industry", *Econometrica*, vol. 50(3), p. 649-670.

Klapper, L., R. Amit and M. Guillén (2008), "Entrepreneurship and firm formation across countries" The World Bank Policy Research Working Paper Series 4313

Klapper, Leora, A. Lewin and J. M.Q. Delgado (2009) "The impact of business environment on the business creation process", The World Bank Policy Research Working Paper Series 4937

Lancaster, T. (1992) *The Econometric Analysis of Transition Data*, Cambridge University Press.

López-García, P. and Puente, S. (2006) "Business demography in Spain: determinants of firm survival", Documentos de Trabajo n. 608, Banco de España.

Mata, J. (1993) "Entry and type of entrant: evidence from Portugal", *International Journal of Industrial Organization*, vol 11(1), p. 101-122.

Mata, J. and Machado, J. (1996) "Firm start-up size: a conditional quantile approach", *European Economic Review*, vol 40, p. 1305-1323.

Mata, J. and Portugal, P. (1994) "Life Duration of New Firms", *The Journal of Industrial Economics* vol 42(3), p. 227-245.

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*, n.13, p. 459-481.

APPENDIX

As most of the research in the field, our analysis also relies on observed firm characteristics, which does not account for possible firm-specific unobserved characteristics, such as the quality of the entrepreneurial project or the human capital of managers. According to Dolton and van der Klaauw (1995), the effects of unobserved individual heterogeneity are not so important when the baseline hazard is non-parametric, as in our case. However, we observe that controlling for unobserved heterogeneity results in more consistent and unbiased estimates of the covariates' coefficients.

An alternative specification of the cloglog model is shown in Table 3, so as to evaluate the importance of controlling the heterogeneity among firms. We used the most straightforward and common method (such as the one used in López-García and Puente, 2006) to incorporate individual unobserved heterogeneity. This method consists of the inclusion into the specification of the proportional hazard model described above, of a random variable, following a Gamma distribution with unit mean (so that the baseline hazard can be interpreted again as the hazard of the mean individual if explanatory variables are re-scaled conveniently) and some positive variance δ^2 . The hazard function is now represented as $h_j(t) = h_0(t)e^{(x \cdot \beta)v_j}$, where v_j is the value of the random variable for the individual j .

| Estimation results with Heterogeneity Correction | |
|--|-------------------------------------|
| Variable | Model 2 Heterogeneity correction |
| Log of Start-up Size | 1.349 *** (0.008) |
| Log of Current Size | 0.429 *** (0.003) |
| Industry (2digit) Start-up entry rate | 1.401 *** (0.026) |
| Start-up Industry HHI (2 digit) | 0.985 *** (0.002) |
| Industry Growth (log) | 1.023 *** (0.008) |
| Growth x Entry rate | 1.043 *** (0.007) |
| Turbulence | 6.130*** (0.276) |
| <u>Sector Dummies</u> | |
| Agriculture | (a) |
| Construction | 1.551 *** (0.028) |
| Manufacturing | 1.740 *** (0.033) |
| Services | 1.168 *** (0.020) |
| Year Dummies | Yes |
| Number of firms | 447772 |
| Log likelihood | -382627.76 |
| LR test of Gamma variance. | |
| $\chi^2(1)$ | 38478.9 *** |

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

Note: (a) refers to the reference sector.

The year dummies "yes" means that they have been included in the estimation. Standard deviation is shown in brackets and *, **, *** means, respectively, 10, 5 and 1% level of significance.

Thus, Table 3 shows the hazard ratios of the most complete model (Model 2), while controlling for heterogeneity. We can infer that this specification is preferable to the previous ones. The p-value associated to the likelihood test of Gamma variation is zero, indicating statistically significant heterogeneity. This fact does not change the sign of the impact of the explanatory variables. It only changes the magnitude of the hazard ratios. It is now possible to observe higher magnitude ratios.