Review of Industrial Organization **17:** 249–265, 2000.

Firms’ Growth, Size and Age: A Nonparametric Approach

JOSÉ C. FARIÑAS and LOURDES MORENO*
Universidad Complutense and PIE-FEP, Plaza Marqués de Salamanca, 8, 28006 Madrid, Spain
E-mail: farinas@funep.es

Abstract. This paper offers empirical evidence of firm failure rates as well as the mean of the distribution of realized growth rates, distinguishing between the sample of non-failing firms and the sample of all firms, failing and non-failing. Attention is directed at identifying a set of characteristics, in particular the size and age of firms, systematically related to the patterns of firm growth and exit, using a panel of Spanish manufacturing firms. The two main contributions of the paper are the use of nonparametric techniques and the analysis of issues ignored in other studies like the regression-to-the-mean bias and the measurement of learning effects. We find evidence that failure rates and the mean growth rate of successful firms decline with size and age. When failing firms are integrated, there are no significant differences in the mean growth rate across the age and size of firms. Regression-to-the-mean does not prove to be a substantial factor behind the negative relationship between size and growth of surviving firms.

Key words: Age, exit, firm growth, learning, size.

I. Introduction

This paper examines the patterns of firm exit and growth, paying attention to traditionally major issues in theoretical and empirical research on firm growth. Do small firms grow faster than large firms? Do young firms grow faster than old firms? For several decades the conventional wisdom has been that expected growth rates are independent of firm size, a property known as Gibrat’s law. Theoretical works that followed this path took the law as an assumption or as a desirable implication; in particular Simon and Bonini (1958) assume that Gibrat’s law holds for firms above the minimum efficient size level.

More recent theoretical literature on industry evolution has emphasized the importance of learning as far as firm growth and changes in market structure are concerned. Different models emphasize different factors determining firm dynamics: Jovanovic (1982), the learning process about innate efficiencies; Ericson and

---

* The authors are particularly grateful to Miguel A. Delgado, Jordi Jaumandreu, Mark J. Roberts and two anonymous referees for helpful comments and suggestions. An earlier version of this paper was presented at the 24th Annual EARIE Conference in Leuven and at the Jornadas de Economía Industrial in Madrid. This research has been partially funded by the project CICYT SEC497-1368.

Pakes (1995), success in R&D expenditures; Cabral (1995), sunk cost in building production capacity. Another relevant distinction within this literature is the difference between active learning – firms actively invest to learn – as in Ericson and Pakes's model, and passive learning – just staying in business increases the knowledge of firms about their innate efficiency – as in Jovanovic and Cabral models, a distinction examined in Pakes and Ericson (1988).

Early empirical literature on firm growth took current size as the only explanatory variable. More recent empirical studies that take learning models as reference have related firm growth to size and age. Within this empirical literature, which cannot be easily summarized, given the substantial amount of empirical work that has been done, one of the major themes lay in dealing with the problem of sample selection censoring due to the exit of firms. Two different techniques have been applied. The first technique, used by Evans (1987) and Hall (1987), applies a standard sample selection model where growth is treated as a latent variable combined with a failure rule that generates a censored sample of firm growth rates. The second technique, used by Dunne et al. (1989), concentrates on the estimation of conditional growth rates in two groups of firms: the sample of non-failing firms and the sample of all firms, failing and non-failing.

In this paper we examine the patterns of firm growth and exit using a panel of Spanish manufacturing firms. The main features and contributions of our paper can be summarized as follows. First, we use the approach taken by Dunne, Roberts and Samuelson (DRS) to deal with the problem of sample selection censoring. Therefore, the paper examines firm failure rates as well as the mean of the distribution of realized growth rates distinguishing between the sample of non-failing firms and the sample of all firms including failing ones. Second, we use two different nonparametric approaches to summarize the across size and age variation in growth and exit rates. The first one is based on a standard regression model and the second one is based on kernel regression estimators. Both approaches complement each other and allow us to estimate population parameters of the distribution of realized growth and failures rates. Apart from the use of nonparametric techniques, the econometric approach used includes novelties that distinguish our paper from other contributions: firstly, we admit firm-level heterogeneity and autocorrelation of arbitrary form in each firm's disturbance of the growth and exit equations; secondly, a linear probability approach is taken to model the probability of failure. Third, the paper examines the size-growth relationship under alternative size measures in an attempt to deal with the regression-to-the-mean fallacy.

The remainder of this paper is divided into five sections. The second section outlines a set of testable predictions based on Jovanovic's model, and describes the empirical model and the estimation procedures. The third describes the data set. The fourth section reports the basic empirical results. Section five provides additional results on the size-growth relationship under alternative size measures which allow us to determine how substantial the regression-to-the-mean bias is. Finally, conclusions are made in section six.

II. Empirical Model

The empirical model in this paper is based on the learning model of Jovanovic (1982), which offers an explicit role for age and size in explaining firm growth. The key feature of the model is the presumption that firms have an initial distribution of efficiency levels that they do not know before entering the industry. Over time, firms obtain information about their cost through production. More efficient firms grow and survive. Less efficient firms learn about their relative inefficiency and choose to exit.

The previous selection model predicts that the age and size of firms will affect failure rates. Firstly, for any given size of firm, its probability of exit is a decreasing function of firm age. Older firms have more precise estimates about their innate efficiency and this reduces the likelihood of failure. Secondly, the probability of failure decreases with firm size. This occurs because the larger the firm is, the less likely it is to exceed the failure boundary. Two additional predictions concern the relationship between growth and age-size characteristics. Firstly, the proportional rate of growth of a firm conditional on survival is decreasing in size. Bounded efficiency is the rationale for this negative relationship. Secondly, for a given size the proportional rate of growth is smaller the older the firm. Diminishing returns to learning is the rationale for this inverse relationship, since older firms have less scope for additional efficiency gains.

Before formulating an empirical model that allows us to study how firm growth varies according to age and size characteristics, a distinction has to be made between potential and observed growth rates. This distinction is important to account for sample selection due to exit. As we mention in the introduction, we follow the method proposed by DRS to permit an investigation of age as well as size effects. Let the random variable g_t' denote the potential proportional growth rate of a firm between t and $t+1$, $g_t' = (q_{t+1} - q_t)/q_t$, where q_{t+1} denotes optimal size at $t+1$ for a non-failing firm. Let $j(g'|x)$ denote the probability density function of potential growth rates for a firm conditional on a set of characteristics x.

Jovanovic's learning model shows that there is a minimum critical size q_{t+1}^* that can be expressed as a critical growth rate $g_t^* = (q_{t+1}^* - q_t)/q_t$ and defines the rule for the firm to decide to stay or to exit from business. If the firm's potential growth rate is equal to or greater than g_t^*, $g_t' \geq g_t^*$, realized growth, g_t, and potential growth coincide:$g_t = g_t'$. If potential growth is below g_t^*, $g_t' < g_t^*$, the firm exits the market and realized growth is $g_t = -1$.

The probability distributions of realized growth conditional on x reflect the interaction between the conditional distribution of potential growth rates $j(g'|x)$ which is not observable, and the failure condition. What can be measured directly is the density function of growth rates conditional on survival, $h(g|x)$, and the density of realized growth rates for all firms in operation at t, labelled $f(g|x)$, which includes firms that exit before $t+1$. According to the previous failure condition, the relationship between $h(g|x)$ and $f(g|x)$ is particularly simple: the function $f(g|x)$ is obtained by truncating the tail of the distribution $h(g|x)$ below g^* and replacing it with a mass point at -1 representing the proportion of firms that fail between t and $t+1$.

There is a relationship between the means of $h(g|x)$ and $f(g|x)$. Let F be an indicator function, taking the value 1 when a firm exits the market and 0 otherwise. Then the conditional mean growth of all firms, including those that exit, can be written as,

$$E(g|x) = E(g|x, F = 1)P(F = 1|x) + E(g|x, F = 0)P(F = 0|x), \qquad (1)$$

where $P(F = 1|x)$ and $P(F = 0|x)$ are the conditional probabilities that a firm does fail or not. Notice that $E(g|x,\ F = 1) = -1$. Therefore, the conditional mean growth rate of all firms (failing and non-failing) equals the conditional mean growth rate of successful firms weighted by the conditional probability of success, minus the probability of failure.

Equation (1) indicates that the mean growth rate of successful firms given the characteristics x, $E(g|x,\ F = 0)$, is an accurate indicator of the conditional mean growth rate of all firms, $E(g|x)$, if and only if the conditional probability of exit, $P(F = 1|x)$, is constant. On the contrary, if as the selection model of industry predicts, the probability of failure diminishes with size and age, the mean growth rate of all plants can increase in those characteristics. This offsetting effect implies that large firms can grow at a higher rate than small firms even though large surviving firms grow at a lower rate than small ones.

An empirical model is now developed to examine the variation of $E(g|x)$, $E(g|x,\ F = 0)$ and $E(F|x) \equiv P(F = 1|x)$ across the size and age characteristics of firms. We parameterize the model in the following way:

$$y_{it} = \sum_{j=1}^{90} \alpha_j D_j + \sum_{k}^{24} \gamma_k D_k^{sa} + \sum_{m}^{2} \beta_m D_m^c + \epsilon_{it}, \qquad (2)$$

where y_t represents either g_{it}, the growth rate of firm i between t and $t+1$, or F_{it}, a discrete variable which adopts the value 1 if firm i fails from period t to $t+1$ and 0 otherwise. D_j are dummy variables that represent interactions among industries and time periods, and are used in the analysis to control for these effects. D_m^c are also control variables which identify firms with a current size category different from its initial one, and D_k^{sa} are dummy variables which classify firms according to their size and age and allow us to estimate the variation across these categories. Finally, ϵ_{it} is a random error.

The dummy variables D_j ($j = 1, \ldots, 90$) represent interactions among industry and time variables which are defined for each of the 18 industries in each of the 5 time periods. Since we are using data on firms from diverse manufacturing industries, some control for differences in growth across the heterogeneous group of industries at different points in time is needed. The set of dummy variables

D_k^{sa} (k = 1, ..., 24) represents combinations of 5 size and 5 age categories (see the next section for a more precise definition of these variables). The selection models, taken here as reference, predict nonlinear effects of size and age on growth as well as interactions between size and age. The reference category corresponds to the youngest and smallest firms which are omitted in the set of variables D_k^{sa} and therefore the estimated means are measured as deviations from the omitted category.

Jovanovic's model assumes no fixed factors. Therefore, the firms' information on their efficiency is fully reflected in their current size and age, because there is no difference between current size and optimal long-run size, given present information. If fixed factors or adjustment costs are important, observed size may not be fully adjusted to the desired optimal level. In practice, size adjustments may be slow due to adjustment costs. To capture these fixed elements, the remaining dummy variables D_m^c (m = 1, 2) assign firms into different groups according to whether the initial size at $t - 1$ is greater or lower than the current size at t.

Two distinguishing features concerning econometric procedures should be mentioned. The first one refers to the assumptions made about the disturbance term in Equation (2). It is assumed that ϵ_{it} is a random disturbance with zero mean and independently distributed across firms but without imposing restrictions on each firm's autocovariance structure. In particular we admit heteroscedasticity and autocorrelation of arbitrary form in each firm's disturbance. Under these assumptions, robust covariance matrix estimators of the White type have been computed (see Arellano, 1987).

The second feature concerns the estimation of conditional expectation of the probability of failure $E(F|x)$ by Equation (2). We use a linear probability model to estimate the conditional expectation of exit for two reasons. First, the coefficients obtained provide a direct measure of the variation in probabilities across the firms' size-age categories. And second, only dummy variables are included among the independent variables so that the linear probability model and the non-linear model produce consistent estimators. Because the linear probability model is a conditionally heteroscedastic regression model with known conditional variances, it has been estimated by Feasible Generalized Least Squares (see Maddala, 1983). Since the errors are correlated, robust standard errors of the type mentioned are also computed for FGLS estimators.

The loss of information due to the grouping procedure used for size and age in the standard regression method can be avoided examining the conditional expectation of firm growth and exit with nonparametric methods based on Kernel regression estimators. As in the standard regression methods, we are interested in estimating the conditional expectations of firm growth and failure rates: $E(y|x)$. The conditional expectation, $m(x) = E(y|x)$, is estimated by:

$$\hat{m}(x) = \frac{\sum_{i=1}^{n} K_h(x - x_i)y_i}{\sum_{i=1}^{n} K_h(x - x_i)}, \tag{3}$$

where y_i represents either g_i, the growth rate of firm i, or F_i, a discrete variable which adopts the value 1 if firm i fails; and x_i represents either s_i, the log of size of firm i, or a_i, the log of age of firm i. The function K_h is a Gaussian Kernel with bandwidth h:

$$K_h(x - x_i) = \frac{1}{h}(2\pi)^{-1/2} \exp\left(-\frac{(x - x_i)^2}{2h}\right), \tag{4}$$

and the bandwidth number chosen is $h = cn^{-1/5}$, for a suitable choice of c based on a process of trial and error (see Delgado and Robinson, 1995).

III. Data

The sample of firms used is drawn from the Encuesta sobre Estrategias Empresariales (ESEE), an annual survey of the activity and business strategies of Spanish manufacturing firms carried out by the Ministry of Industry. The base year for this annual survey is 1990. In this year 2,188 firms participated in the survey according to the following sampling fractions: all firms with more than 200 employees were asked to participate, and the rate of participation reached 67.6% of the population of firms within that size category. Firms that employed between 10 and 200 employees were chosen by a random sampling scheme, and the sampling fraction was 3.9% of the number of firms in the population. The sampling scheme was conducted for each manufacturing NACE class (two digit) level and the following size categories defined in terms of the number of employees: 10–20; 21–50; 51–100; 101–200; and more than 200 employees. In subsequent years the survey maintained its initial characteristics, minimising attrition and annually incorporating newly created firms selected with the same sampling criteria as in the base year. The technical details about the method and definitions under which the survey is carried out can be found in the Ministry of Industry (1992).

Selective sampling is one of the characteristics of our database. In particular, for studying the relation of growth, g, to size, x, the data set has oversampled large firms: there is a random sampling fraction of 0.04 for firms of $x \leq 200$ employees and a fraction of 0.7 for firms of $x > 200$ employees. As Goldberger (1991) suggests, the estimation of equations like (2) will serve to estimate the population relation between g and x. The conditional pdf of growth given size, denoted $f(g|x)$ and $h(g|x)$ in Section 2, is not affected by this selective sampling procedure on x, and the conditional expectation function $E(g|x)$ will also serve as an estimator of the population relation between growth and size (see, Goldberger, 1991, pp. 145–147). However, to estimate the parameters of interest, we can also take advantage of the characteristics of the data set and establish an explicit relationship between our sample data and the population of Spanish manufacturing firms. The regression coefficients and test statistics can also be estimated, uprating the observations of firms by their sampling fractions. In the next section, we also include some results obtained by this approach.[1]

The set of firms included in the estimation was surveyed from 1990 to 1995. They form a non-balanced panel data set, due to the exit and entry of firms into the database. From t to $t + 1$, we are able to observe non-failing firms and exits. Entries of new firms into the market are integrated in the database in subsequent years ($t + 2$, $t + 3$, etc.). Therefore, we are not able to have precise information on entry rates, but our data set has a significant fraction of observations, around 15% of the total number of observations used in the estimation, which corresponds to very young firms less than 5 years old. From this cohort we are able to observe the post-entry performance of new firms.

The methodology employed uses multiple observations of firms which are representative of the population of Spanish manufacturing firms classified according to industry and size categories. We emphasise that these categories coincide precisely with groups used to define the dummy variables included in regression model (2), which examines the across size and age variation in growth and exit processes. The definition of variables is made according to the following criteria:

The discrete variable F_{it} assigns 1 to firms that fail between t and $t + 1$ and 0 if they survive. Apart from firms going into bankruptcy and firms that have decided to close, exits include firms whose primary economic activity changes to a non-manufacturing industry, and also disappearing companies through mergers by absorption by another company.

Firm growth g is defined as the discrete rate of growth between t and $t + 1$ of the total number of persons employed, which is measured at the end of the year. For firms with $F = 1$, total employment at the end of the year t equals zero and $g = -1$.

The age of the firm is computed as the difference between the calendar year at t and the birth-year reported by the firm. Five age classes are considered: less than 5 years, 6–10 years, 11–25 years, 26–50 years and more than 50 years.

Firm size is measured by employment. The number of employees in period t is divided into the following five classes: firms with less than 20 employees, firms with 21–50 employees, firms with 51–200 employees, firms with 201–500 employees and firms with more than 500 employees.

Firms are classified, according to the principal economic activity they report, into one of the eighteen two-digit NACE-CLIO R44 manufacturing industries.

A dummy variable, D^c, compares the firm's employment size class at t with its previous year size class. In particular, this variable identifies firms with an initial size at $t - 1$ greater or lower than the current size at t.

The initial sample obtained from the data set includes 10,196 observations on 1,971 firms over the period 1990–1995. The panel of firms is unbalanced since many firms entered and exited after 1990. Discrete rates of growth have to be defined and initial size considered so that the sample is reduced for each firm up

---

[1] We thank an anonymous referee for bringing this issue and suggestion to our attention.

Table I. Firm growth and exit rates disaggregated by current size (average annual rates in % and number of observations in parentheses)

	Size (number of employees)					
	< 20	21–50	51–200	201–500	> 500	Total
Growth rate of surviving firms	4.2	−1.4	−2.7	−2.9	−4.5	1.4
	(2179)	(1586)	(986)	(1443)	(667)	(6861)
Exit rates	7.3	6.8	5.5	2.9	2.3	6.8
	(172)	(116)	(57)	(43)	(16)	(404)
Growth rate of all firms	−3.4	−8.1	−8.0	−5.7	−6.7	−5.5
	(2351)	(1702)	(1043)	(1486)	(683)	(7265)

to a maximum of five cross-sections with a total number of 8,225 observations. Since we can distinguish between internal growth and growth through mergers, we focus on internal growth. Mergers and acquisitions are excluded from the previous sample. The number of observations excluded for this reason is 807, around 2.0% of the observations in each year. Finally, 153 observations have also been excluded for various errors due to firms' reporting bad data. The final number of observations amounts to 6,861 for the sample of non-failing firms and to 7,265 for the sample of all non-failing and failing firms.

To illustrate the differences between the two samples, Table I presents estimates of the average growth, distinguishing between non-failing and all firms, and failure rates disaggregated according to size. For firms that do not fail, the values decline with size and the same pattern is present for exit rates, indicating that small firms fail more often than large firms. Both tendencies interact in the sample of all firms, where smaller firms do not present any clear difference in their growth with respect to larger firms. Similar results are obtained for the variation of exit rates and growth rates across the age of firms.

IV. Empirical Results on Firm Growth and Exit Rates

This section reports results on the estimation of Equation (2). The empirical model examines how the mean growth rate and the exit rate vary across size and age of firms, controlling for industry and year categories. First column of Table II reports the regression coefficients for firm failure rates. All coefficients measure deviations from the mean rate of the exit of a firm less than 5 years old and with less than 20 employees. For this class of firms the average annual exit rate during the period and through all industries is 8.5%.

Two patterns emerge from the estimates. First, within any age categories, the failure rate tends to decline with increases in a firm's current size. In general, there is a downward trend in the failure rate with increasing firm size, particularly for age groups under 5 years and 26–50 years. For firms that are in age groups 6–10, 11–25 and more than 50 years old, the downward trend is also present but it is not monotonic. The second pattern is that the probability of exit decreases with age for a given size; however, this effect is less apparent and it is not monotonic.

The pattern of age and size effects on failure rates can be examined more closely by comparing the estimated model with several restricted versions of this model. All hypotheses tests are based on Wald statistics using robust variance estimates.[2] Table III reports in the second column the hypotheses tests that a firm's current size, age and interactions size-age have no effect on the failure rate. They indicate that size and age, as well as interactions size-age, have significant effects on the failure rates. The existence of significant effects means in this context that there are differences in the failure rates across size, age and interactions of size and age. In particular, failure rates are lower for larger firms and the same relationship is present though less pronounced for older firms. Other inferences on failure rates concern the importance of both industry and year effects.

In addition to the effects already mentioned, a control variable that identifies firms with a different current size category, t, with respect to previous year size category, $t-1$, is introduced. The estimates reported in Table II indicate that firms with an initial size category at $t-1$ greater than current size at t have failure rates nearly 2 percentage points greater than firms that remain at the same size category. This difference is not statistically significant. Neither is there evidence of lower exit rates for firms with current size greater than previous year size. We can interpret these findings as evidence of fixed factors to be unimportant in affecting the firm's exit rate process.

According to models of passive learning there is a failure boundary or critical growth rate that gives the firm the rule to decide whether to stay or to exit from the market. The prediction that the probability of exit is a decreasing function of age is true if the failure boundary is fixed with respect to age. However the failure boundary may decrease with age and this in turn increases the probability of failure as the firm ages (Pakes and Ericson, 1988).

Figure 1 presents the estimate of the conditional expectation of firm failure rates for a given age. These values have been estimated by a Kernel regression estimator and are defined over the whole range of the age of firms. Failure rates decline steeply over firms with ages under 20 years old; for firms within the range of 20–40 years old, the exit rate flattens; and finally, for firms over 40 years old, no relationship in failure rates is observed. Empirically, therefore, we do not observe the failure rate always declining with age. This result can have the interpretation that learning effects last no more than an average of 20 years.

Second and third columns of Table II report the regression coefficients for the mean growth rate of firms. One column corresponds to the sample of surviving

---

[2] Wald-test are based on the expression $W(q) = (Rb)'[RV_0(b)R']^{-1}(Rb)$, where R is the corresponding matrix of restrictions tested, b is the vector of the estimated coefficients, $V_0(b)$ is the robust covariance matrix of the unrestricted model, and q the number of restrictions.

Table II. Regression coefficients for annual firm failure rates and mean annual growth rates (standard errors in parentheses)

			Mean annual growth rates			
	Failure rates		Surviving firms		All firms	
Intercept[a]	0.085	(0.011)[b]	0.073	(0.015)[b]	−0.022	(0.019)
Age/size (t):						
≤ 5 years						
21–50 employees	−0.005	(0.019)	−0.082	(0.020)[b]	−0.069	(0.028)[b]
51–200	−0.015	(0.020)	−0.111	(0.032)[b]	−0.101	(0.042)[b]
201–500	−0.030	(0.020)	−0.089	(0.024)[b]	−0.037	(0.033)
>500	−0.074	(0.012)[b]	−0.147	(0.023)[b]	−0.061	(0.027)[b]
6–10 years						
< 20 employees	−0.022	(0.013)	−0.045	(0.017)[b]	−0.024	(0.023)
21–50	−0.008	(0.015)	−0.079	(0.018)[b]	−0.067	(0.025)[b]
51–200	−0.044	(0.015)[b]	−0.085	(0.021)[b]	−0.036	(0.031)
201–500	−0.041	(0.015)[b]	−0.094	(0.024)[b]	−0.040	(0.028)
> 500	−0.044	(0.020)[b]	−0.086	(0.024)[b]	−0.025	(0.037)
11–25 years						
< 20 employees	−0.017	(0.013)	−0.030	(0.019)	−0.004	(0.024)
21–50	−0.037	(0.012)[b]	−0.084	(0.017)[b]	−0.046	(0.022)[b]
51–200	−0.028	(0.014)	−0.099	(0.018)[b]	−0.063	(0.024)[b]
201–500	−0.048	(0.013)[b]	−0.087	(0.016)[b]	−0.020	(0.022)
> 500	−0.044	(0.016)[b]	−0.129	(0.022)[b]	−0.071	(0.028)[b]
26–50 years						
< 20 employees	−0.022	(0.016)	−0.053	(0.020)[b]	−0.023	(0.028)
21–56	−0.020	(0.015)	−0.085	(0.019)[b]	−0.039	(0.025)
51–200	−0.035	(0.013)[b]	−0.104	(0.017)[b]	−0.058	(0.023)[b]
201–500	−0.056	(0.011)[b]	−0.105	(0.015)[b]	−0.037	(0.020)
> 500	−0.055	(0.012)[b]	−0.114	(0.017)[b]	−0.047	(0.023)[b]
> 50 years						
< 20 employees	−0.016	(0.023)	−0.090	(0.022)[b]	−0.075	(0.041)
21–50	−0.050	(0.019)[b]	−0.128	(0.022)[b]	−0.074	(0.031)[b]
51–200	−0.027	(0.021)	−0.118	(0.022)[b]	−0.071	(0.032)[b]
201–500	−0.028	(0.016)	−0.113	(0.017)[b]	−0.066	(0.024)[b]
> 500	−0.051	(0.015)[b]	−0.118	(0.017)[b]	−0.052	(0.024)[b]
Size $(t - 1) >$ size (t)	0.018	(0.012)	0.009	(0.019)	−0.013	(0.022)
Size $(t - 1) <$ size (t)	0.012	(0.010)	−0.024	(0.015)	−0.029	(0.021)

[a] The reported intercept is the simple average value of the 90 estimated industry-time terms.
[b] Significant at the 0.05 level.
Number observations: 6861 (surviving firms), 7265 (all firms and failure rates).

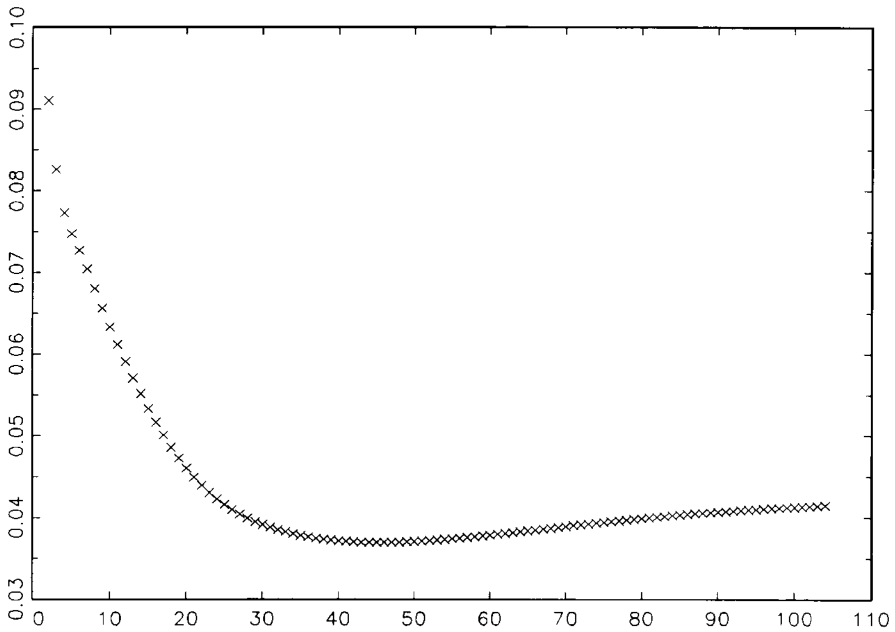

Figure 1. Nonparametric estimate of failure rates conditional on age: all firms.Note: Horizontal axis represents firms' age in number of years.

firms while the other refers to the full sample of firms including nonsurviving ones. The empirical regularity found in most previous studies is confirmed here, since the mean growth rates of non-failing firms decrease with size. In virtually all age categories, growth declines monotonically with size. When coefficients are examined for a given size category, mean growth rates are decreasing with age although this relationship is less pronounced for the largest category of firms with over 500 employees. The hypothesis test statistics reported in Table III (see second column) indicate that size and age have significant effects on growth patterns. Age-size interactions are also important. In addition, industry and year effects also have a significant influence on growth.

The regression coefficients for the mean growth rates over all firms are also reported in Table II. The size and age patterns in the sample of all firms contrast sharply from the patterns observed in the non-failing sample. In the former, the size pattern is not uniform at all, and the differences in growth rates across the size of firms are not statistically significant. This pattern of no relationship between expected growth and size appears because the reduction in the failure rate with increased size and the reduction in the growth rate of non-failing firms with increased size compensate each other. The net effect of age on firm growth is similar to the effects of size. The hypothesis test statistics reported in Table III confirm that size, age and interactions size-age effects are not significant.

Table III. Hypothesis test statistics

Hypothesis (no effect)	Number of restrictions	Exit rate	Growth rate (surviving firms)	Growth rate (all firms)
Size effects	20	93.8[a]	97.4[a]	30.9
Age effects	20	38.8[a]	44.1[a]	18.1
Interaction size-age effects	16	35.5[a]	26.3[a]	12.4
Industry effects	85	350.9[a]	208.3[a]	168.3[a]
Years effects	72	201.8[a]	333.1[a]	222.0[a]
Change of size category effect	2	3.8	3.0	2.1

[a] Significant rejection at the 0.05 level.

The regression coefficients and test statistics of Tables II and III have also been estimated using as weights for each observation the square root of the inverse of the sampling fraction that corresponds to its size class and industry. All comments already made are consistent for both sets of results and the hypotheses testing procedure leads to the same basic conclusions[3]

To conclude this section, we present the estimates obtained by the Kernel regression estimators. Figure 2 presents the estimators of the conditional expectations of firm growth for a given size within different age groups. To compare them to the regression coefficients reproduced in Table II, we distinguish between the sample of surviving firms and the sample of all firms. Results, as can be seen, coincide basically with the two patterns already observed in Table II. The first one is that in the sample of surviving firms, growth rates diminish with size in all age classes, excluding the group of firms more than 50 years old. The absence of any relationship between size and growth for the group of mature firms (over 50 years old) is consistent with Jovanovic's theory, if technology is Cobb–Douglas with decreasing returns to scale. Second, in the sample of all firms, failing and non-failing, growth rates have either an increasing, decreasing and/or highly nonlinear relationship with size depending on the particular age class.

V. Size and Growth: Negative Relationship or Regression-to-the-Mean?

Regression results reported in Table II indicate that the mean growth rate of successful firms declines with size. A possible interpretation of this finding is that firms are simply regressing to their mean size. If this were the case, the finding suffers from a statistical weakness known as the regression fallacy or regression-to-the-mean bias. Davis et al. (1996) have argued that most longitudinal studies of the relationship between employer size and growth suffer this statistical pitfall. In

---

[3] This set of alternative results is available upon request.

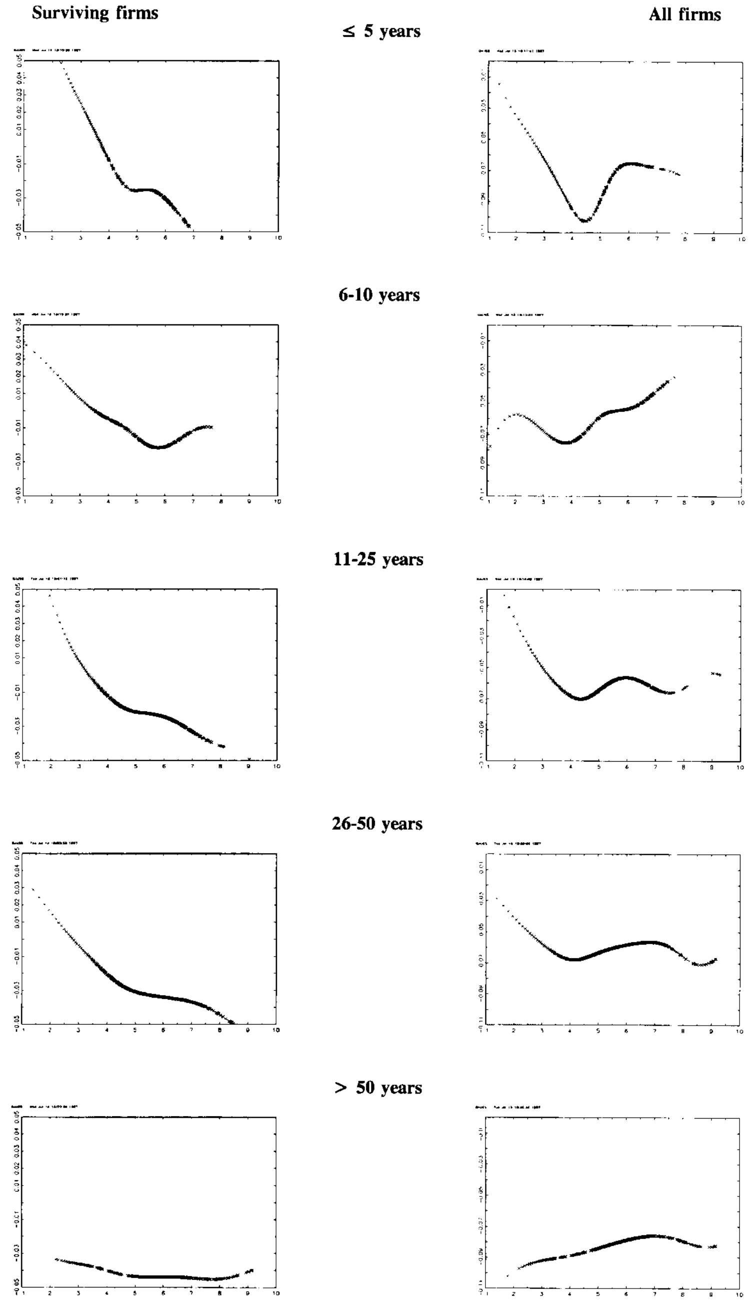

Figure 2. Nonparametric estimates of mean growth conditional on size and for given age groups. Note: Horizontal axis represents firms' log size. Size is measured as the number of employees.

this section we provide additional results, examining the size-growth relationship under alternative size measures.

The regression-to-the-mean bias may arise whenever measurement error introduces transitory fluctuations in observed size.[4] If size is measured in the base year (current period) with error, firms that have a transitorily low size due to measurement error will on average seem to grow faster than those with a transitorily high size. The bias also arises whenever firms experience transitory fluctuations in size. With an optimal long-run size due to intrinsic characteristics, as in Jovanovic's model, and given the existence of transitory fluctuations around that equilibrium value, the fallacy arises if firms are classified into size classes defined in terms of the base year (current period). Both phenomena, measurement error and transitory fluctuations, are important features of longitudinal data on firms, especially when the time period is short, and transitory rather than permanent components of firms' growth are likely to dominate the result.

In the previous section, the results were obtained following the standard practice of classifying continuing firms and firm exits by their base-year size. In our case, the base year corresponds to the current period, which is the initial year of the time interval over which the growth rate is defined. To appreciate the extent of the regression-to-the-mean bias, Table IV reports two additional sets of coefficients and test statistics that emerge under alternative size measures.

The first column of Table IV reproduces coefficients obtained from a regression where current firm size is still used. Given that the fallacy arises because firms can migrate between size categories from one year to the next, the simplest way to avoid the regression fallacy is to control boundary-crossing firms. This is done by including two dummy variables which identify firms crossing between size classes. The first dummy variable assigns 1 to firms with an employment size class at t lower than size class at $t + 1$ and 0 otherwise. The second dummy assigns 1 to firms with an employment size class at t greater than size at $t + 1$ and 0 otherwise. Results presented in the second column of Table IV follow Davis et al.'s (1996) suggestion of measuring size by average firm size. Average size is defined as equal to the simple average of a firm's employment at the current year (t) and next year ($t + 1$).[5]

The results obtained by means of the two approaches point to basically the same conclusion: the mean growth rates of non-failing firms decline steeply over the first three size intervals and then flatten over the rest of the intervals. This regularity basically reflects the type of relationship between growth and size already reported in Table II for non-failing firms. Our analysis does not find any substantial bias in favor of a greater growth of small business with respect to large firms when size is measured by using a base-year measure. We interpret these results as an indication

---

[4] See Hart (1995) for a more general discussion of this issue.

[5] As detailed in Hart (1995), other estimation techniques involve the use of grouping methods or instrumental variables and estimates based on both the direct and the reverse regressions as bounds for the true regression.

Table IV. Mean annual growth rates: regression coefficients for surviving firms with alternative measures of firm's size (standard errors in parentheses)

	Current firm size[c]		Average firm size[d]	
Intercept[a]	0.037	(0.014)[b]	0.036	(0.014)[b]
Age/size (t):				
≤ 5 years				
21–50 employees	−0.019	(0.018)	0.020	(0.021)
51–200	−0.070	(0.025)[b]	−0.067	(0.027)[b]
201–500	−0.038	(0.021)	−0.042	(0.025)
> 500	−0.074	(0.022)[b]	−0.090	(0.021)[b]
6–10 years				
< 20 employees	−0.034	(0.015)[b]	−0.038	(0.016)[b]
21–50	−0.027	(0.016)	−0.011	(0.017)
51–200	−0.042	(0.018)[b]	−0.028	(0.027)
201–500	−0.055	(0.020)[b]	−0.054	(0.022)[b]
> 500	−0.042	(0.021)[b]	−0.023	(0.021)
11–25 years				
< 20 employees	−0.018	(0.018)	−0.024	(0.016)
21–50	−0.037	(0.015)[b]	−0.021	(0.018)
51–200	−0.065	(0.016)[b]	−0.064	(0.017)[b]
201–500	−0.038	(0.014)[b]	−0.043	(0.016)[b]
> 500	−0.066	(0.018)[b]	−0.070	(0.020)[b]
26–50 years				
< 20 employees	−0.047	(0.018)[b]	−0.044	(0.016)[b]
21–56	−0.043	(0.016)[b]	−0.033	(0.020)
51–200	−0.074	(0.015)[b]	−0.077	(0.016)[b]
201–500 *	−0.056	(0.014)[b]	−0.063	(0.015)[b]
> 500	−0.060	(0.015)[b]	−0.070	(0.016)[b]
> 50 years				
< 20 employees	−0.100	(0.018)[b]	−0.066	(0.022)[b]
21–50	−0.046	(0.018)[b]	−0.079	(0.020)[b]
51–200	−0.099	(0.020)[b]	−0.096	(0.021)[b]
201–500	−0.061	(0.015)[b]	−0.072	(0.016)[b]
> 500	−0.073	(0.016)[b]	−0.076	(0.016)[b]
Size ($t-1$) > size (t)	0.382	(0.033))[b]		
Size ($t-1$) < size (t)	0.238	(0.011)[b]		
Hypothesis test statistics: (number of restrictions in parentheses)				
Size effects (20)	56.4[b]		65.9[b]	
Age effects (20)	64.9[b]		62.4[b]	
Interaction size-age effects (16)	28.1[b]		28.5[b]	
Industry effects (85)	168.3[b]		205.8[b]	
Year effects (72)	233.5[b]		337.3[b]	

[a] The reported intercept is the simple average value of the 90 estimated industry-time terms.

[b] Significant at the 0.05 level or significant rejection at the 0.05 level.

[c] Current size means the size at initial year (t) of the time interval between t and $t+1$ over which the growth rate is calculated.

[d] Average firm size equals the simple average of the firm's employment at t and $t+1$.

Number of observations: 6.861.

of the fact that in the sample of surviving firms, the relationship between size and growth is affected by additional factors other than simply the tendency of firms to regress to their mean size.

VI. Conclusions

This paper presents an empirical analysis based on the theoretical model of Jovanovic (1982) which relies on firm heterogeneity, reflected through size and age differences, and on market selection to generate patterns of firm growth and failure. The panel of firms used in this study is representative of the population of Spanish manufacturing firms stratified according to size and industry categories during the period 1990–1995. This data base allows us to estimate firm failure rates according to the size and age of firms, and the conditional mean on size and age characteristics of the distribution of growth rates for two groups of firms: all firms in operation at the beginning of each period and all firms that do not fail.

Our main findings are the following: (1) failure rates decline with size and age of firms; (2) the mean growth rate of successful firms declines with size and age; (3) when failing firms are integrated into the analysis, there are no significant differences in the mean growth rate across the age and size of firms. Two additional findings complete our summary. Regression-to-the-mean does not prove to be a substantial factor behind the negative relationship between the size and growth of surviving firms. The result can be interpreted as an indication of the fact that this observed negative relationship is affected by factors like bounded efficiency, market selection or learning mechanisms, and not simply by the tendency of firms to regress to their mean size. Finally, our results are consistent with a large body of evidence (see Mata, 1993; Audretsch, 1995; Hart and Oulton, 1996; Geroski, 1998; and many others) that emphasizes evolutionary and learning effects, which point out the existence of threshold sizes, as well as ages below which smaller or younger surviving firms grow faster.

The results can be compared with empirical evidence available on Gibrat's law and related literature on stochastic growth models as surveyed by Sutton (1997). Hall (1987), Evans (1987a, 1987b) and Dunne and Hughes (1994) show that the relationship between growth and size is not constant but rather decreasing, even after controlling for the selection of firms out of the sample. Instead, we find that the relationship is decreasing for surviving firms but rather constant for all firms including failing ones. Our results are closer to evidence in Dunne et al. (1989), which report a nondecreasing relationship between expected growth and size for plants owned by multiplant firms when failing units are included. These differences in the pattern of results could be related with differences in methodology. The first group of studies deals with the sample selection bias by Heckman's two-stage procedure which implies that the distribution of potential growth is treated as a latent-variable distribution. The second methodology estimates parameters of the distribution of realized growth rates for the sample of surviving firms and the sample of all firms. The analysis of these differences may be a useful direction for research.

References

Arellano, M. (1987) ‘Computing Robust Standard Errors for Within-Groups Estimators’, *Oxford Bulletin of Economics and Statistics*, **49**, 431–434.

Audretsh, D. (1995) *Innovation and Industry Evolution*. Cambridge, MA: The MIT Press.

Cabral, L. (1995) ‘Sunk Costs, Firm Size and Firm Growth, *The Journal of Industrial Economics*, **43**, 161–172.

Davis, S., J. Haltiwanger, and S. Schuh (1996) ‘Small Business and Job Creation: Dissecting the Myth and Reassessing the Facts’, *Small Business Economics*, **8**, 297–315.

Delgado, Miguel A., and P. M. Robinson (1995) ‘Nonparametric and Semiparametric Methods for Economic Research’, in L. Oxley, D. A. George, and S. Sayer (eds.), *Surveys in Econometrics*. Oxford: Basil Blackwell.

Dunne, P., and A. Hughes (1994) ‘Age, Size, Growth and Survival: UK Companies in the 1980’s’, *The Journal of Industrial Economics*, **42**, 115–140.

Dunne, T., M. Roberts, and L. Samuelson (1989) ‘The Growth and Failure of US Manufacturing Plants’, *Quarterly Journal of Economics*, **96**, 671–698.

Ericson, R., and A. Pakes (1995) ‘Markov-Perfect Industry Dynamics: A Framework for Empirical Work’, *Review of Economic Studies*, **62**, 53–82.

Evans, D. S. (1987a) ‘Tests of Alternative Theories of Firm Growth’, *Journal of Political Economy*, **95**, 657–674.

Evans, D. S. (1987b) ‘The Relationship between Firm Growth, Size and Age Estimates for 100 Manufacturing Industries’, *Journal of Industrial Economics*, **35**, 567–581.

Geroski, P. A. (1998) ‘An Applied Econometrician’s View of Large Company Performance’, *Review of Industrial Organization*, **13**, 271–293.

Goldberger, A. (1991) *A Course in Econometrics*. Cambridge, MA: Harvard University Press.

Hall, B. (1987) ‘The Relationship between Firm Size and Firm Growth in the US Manufacturing Sector’, *Journal of Industrial Economics*, **35**, 583–606.

Hart, P. E. (1995) ‘Galtonian Regression across Countries and the Convergence of Productivity’, *Oxford Bulletin of Economics and Statistic*, **57**, 287–293.

Hart, P. E., and N. Oulton (1996) ‘Growth and Size of Firms’, *The Economic Journal*, **106**, 1242–1252.

Jovanovic, B. (1982) ‘Selection and Evolution of Industry’, *Econometrica*, **50**, 649–670.

Maddala, G. S. (1983) *Limited-Dependent and Qualitative Variables in Econometrics*. Cambridge: Cambridge University Press.

Mata, J. (1993) ‘Firm Entry and Firm Growth’, *Review of Industrial Organization*, **8**, 567–578.

Ministry of Industry (1992) *Un panorama de la industria española*. Madrid: MYCIT.

Pakes, A., and R. Ericson (1988) ‘Empirical Implications of Alternative Models of Firm Dynamics’, University of Wisconsin, Working Paper 8803.

Simon, H., and C. Bonini, C. (1958), ‘The Size Distribution of Business Firms’, *American Economic Review*, **48**, 607–617.

Sutton, J. (1997) ‘Gibrat’s Legacy’, *Journal of Economic Literature*, **35**, 40–59.