

Coad, Alex

Article

Investigating the exponential age distribution of firms

Economics: The Open-Access, Open-Assessment E-Journal

Provided in Cooperation with:

Kiel Institute for the World Economy – Leibniz Center for Research on Global Economic Challenges

Suggested Citation: Coad, Alex (2010) : Investigating the exponential age distribution of firms, Economics: The Open-Access, Open-Assessment E-Journal, ISSN 1864-6042, Kiel Institute for the World Economy (IfW), Kiel, Vol. 4, Iss. 2010-17, pp. 1-30, <https://doi.org/10.5018/economics-ejournal.ja.2010-17>

This Version is available at:

<https://hdl.handle.net/10419/32772>

Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

Terms of use:

Documents in EconStor may be saved and copied for your personal and scholarly purposes.

You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.

If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.



<http://creativecommons.org/licenses/by-nc/2.0/de/deed.en>

Investigating the Exponential Age Distribution of Firms

Alex Coad

Max Planck Institute of Economics, Jena

Abstract While several plots of the aggregate age distribution suggest that firm age is exponentially distributed, we find some departures from the exponential benchmark. At the lower tail, we find that very young establishments are more numerous than expected, but they face high exit hazards. At the upper tail, the oldest firms are older than the exponential would have predicted. Furthermore, the age distribution of disaggregated industries (such as the international airline industry) is less regular and can display multimodality. Although we focused on departures from the exponential, we found that the exponential was a useful reference point and endorse it as an appropriate benchmark for future work on industrial structure.

JEL L20, L25, L11

Keywords Age distribution; exponential distribution; firm size distribution; survival

Correspondence Alex Coad, Max Planck Institute of Economics, Evolutionary Economics Group, Kahlaische Strasse 10, D-07745 Jena, Germany; e-mail: coad@econ.mpg.de

1 Introduction

A very large literature has focused on the firm size distribution.¹ In fact, there is even a JEL classification code that specifically seeks to accommodate research on the firm size distribution.² Among this vast body of literature, some authors seek to determine the best functional fit to the empirical firm size distribution, generally focusing on the lognormal, the Pareto, and the generalized beta as the most suitable candidates. Other theoretical models seem to focus on generating the empirically-observed size distribution as one of the main ‘reality checks’ of their model’s predictions. Why has the empirical firm size distribution received so much attention? We suggest a few reasons here. First, the size distribution gives a useful summary representation of the structure of industries and economies, that allows for comparisons across samples and over time. Second, the size distribution displays a smooth, regular shape at the aggregate level that matches closely to theoretical densities. Third, the familiar right-skewed shape of the aggregate size distribution is remarkably robust across datasets and is referred to as a stylized fact of industry structure. Fourth, data on firm size is relatively easy to obtain. Fifth, the regular shape of the size distribution provides an explanandum that has inspired theoretical models.³

In contrast to the size distribution, the age distribution has barely been investigated. In this paper, we suggest that the age distribution is a useful summary representation of the structure of industries, that it displays a regular shape that is robust across datasets and is a close match to the exponential distribution. The age distribution is also observed to make a useful contribution to theoretical modelling of the firm growth process.

One reason why the aggregate age distribution has not been investigated in previous empirical work is presumably because of the difficulty in obtaining data on the ages of large samples of firms. Headd and Kirchoff (2009, p. 548) comment on “the dearth of information by business age” and write that “Simply stated, industrial organization and small business researchers are deprived of firm-age data.” The following reasons can help explain why data on firm age is harder to obtain than data on variables such as firm size. First, while the size distribution

¹ For a recent survey see de Wit (2005).

² This JEL code is L11: Production, Pricing, and Market Structure; Size Distribution of Firms. This information is taken from the February 2009 update to the JEL classification system (see http://www.aeaweb.org/journal/jel_class_system.php).

³ For example, Gibrat’s (1931) celebrated model of firm growth.

is constructed from current data (i.e. a firm's current size), the age distribution is constructed from historical data concerning a firm's initial conditions (i.e. a firm's year of founding). Furthermore, in the construction of administrative datasets firms are required to give information on variables such as sales and employment (e.g. for tax reasons), while this requirement does not exist for age data. While we acknowledge that data on firm age has not always been easy to obtain, we also observe that the situation has been improving, which leads us to consider that the firm age distribution will receive more attention in future research.

Although the prior literature has not focused on the age distribution *per se*, many studies have focused on the related issue of firms' survival rates. An early contribution coined the term 'liability of newness' to describe how young organizations face higher risks of failure (Stinchcombe, 1965). More recently, however, authors have referred to the 'liability of adolescence' (Bruderl and Schussler, 1990; Fichman and Levinthal, 1991) to explain why firms face an initial 'honeymoon' period in which they are buffered from sudden exit by their initial stock of resources. Still others have identified liabilities of senescence and obsolescence (Barron et al., 1994) according to which older firms are expected to face higher exit hazards. As such, the literature on firm age and survival has given conflicting predictions, and scholars who are not familiar with the subtleties of these conflicting concepts may not have a clear idea about the age structure of firms in an industry.

Furthermore, the existing literature on firm survival has often focused on tracking small samples of firms in specific industries.⁴ In this vein, some studies have provided evidence that there are distinct periods of high entry, and also high exit (i.e. 'shakeout'), at specific stages in the life cycle of some industries and submarkets.⁵ While we acknowledge that detailed analysis of specific industries has been a fruitful field of research, in this paper we complement the existing literature by focusing on the age distribution at the aggregate level. In the absence of detailed information on the survival histories of specific age cohorts, it may be preferable to focus on the cross-sectional age distribution at a point in time. Instead of focusing on mortality rates over time for small samples of firms, the

⁴ For example, Delacroix and Carroll (1983) on Argentinian and Irish newspapers, Barron et al. (1994) on credit unions in New York City, Klepper (2002) on the automobile, tyre, television, and penicillin industries in the US, and Thompson (2005) on the iron and steel shipbuilding industry in the US.

⁵ See for example Klepper and Thompson (2006) on the US laser industry, Guenther (2009) on the German machine tools industry, and Buenstorf and Klepper (2009) on the US tyre industry.

age distribution corresponds to a snapshot of accumulated mortality rates for all firms from all age cohorts combined.

The age distribution can be linked to many areas of industry structure and dynamics, and we anticipate that it will be of interest to a wide range of economists working on both empirical and theoretical issues. For example, the age distribution might shed light on the structure of the age of technology used in an industry, and also the degree of adoption of general purpose technologies throughout the economy, if firms are assumed to be characterized by the capital vintage of the period in which they enter (as in the theoretical model in Salter (1960)). Furthermore, to the extent that organizations remain fundamentally inert once they are founded (Hannan and Freeman, 1984), the age distribution can elucidate the variety of different types of organization operating in an industry. Given that banks sometimes condition their loans according to firm age (Winker, 1999), there might be a commercial interest in the aggregate firm age distribution. Finally, an understanding of the aggregate age distribution can shed light on the representativeness of firm-level databases. With the benefit of hindsight, it appears that some previous work investigating how firm performance varies with age has apparently been undertaken on datasets in which very young firms are noticeably under-represented.⁶

The aim of this paper is to draw attention to the exponential age distribution, and also to discuss cases in which the empirical age distribution drifts away from the exponential. Even in these situations, however, we argue that the exponential is a helpful reference point. Section 2 presents the theoretical interest in the exponential age distribution, and Section 3 shows how empirically-observed age distributions from a number of different aggregate datasets seem to match well to the exponential case. Section 4 investigates the age distribution of young establishments in the US. Section 5 investigates the age distribution of the world's oldest firms. Section 6 presents a disaggregated analysis of the international airlines sector. Section 7 concludes.

⁶ For instance, Fagiolo and Luzzi (2006, p. 31) observe a lognormal age distribution, which leads us to suspect that young firms (younger than the modal age) are under-represented. Similarly, Huergo and Jaumandreu (2004b,a) investigate the age distribution of firms in their database and observe a bimodal distribution, with the lower mode corresponding to the 5-8 years age category, and the upper mode corresponding to the residual 37+ category. In comparison to these previous studies, the present paper provides a benchmark that can be used to gauge the extent of sample selection and the under-representation of very young firms in databases.

2 Theoretical Modelling

We now demonstrate how the age distribution has been of interest in theoretical models. In particular, an exponential age distribution is assumed in the following simple model of the firm size distribution. In this mathematical model, a Gibrat growth process is shown to give a lognormal firm size distribution (FSD) within cohorts, which is then combined with an exponential distribution of firm age to give a Pareto firm size distribution that has often been observed at the aggregate level.⁷ The basic mathematical model (i.e. integrating a lognormal distribution over an exponential distribution to obtain a Pareto) was previously used by Huberman and Adamic (1999) to explain the number of web pages on internet sites, before being brought into economics by Reed (2001), who focused mainly on explaining the distributions of earnings and city sizes. Coad (2010) applies this model to explain the Pareto firm size distribution, and presents some preliminary analysis on the firm age distribution. Although the age distribution is explicitly modelled here, it can also be seen as an implicit feature of earlier models of industry evolution that have mechanisms of firm entry and exit.⁸

Let x_t be the size of a firm at time t , and let ε_t be random variable representing an iid idiosyncratic, multiplicative growth shock over the period $t - 1$ to t , with mean $\bar{\varepsilon}$ and variance σ^2 . We have

$$x_t - x_{t-1} = \varepsilon_t x_{t-1} \quad (1)$$

which can be developed to obtain

$$x_t = (1 + \varepsilon_t)x_{t-1} = x_0(1 + \varepsilon_1)(1 + \varepsilon_2) \dots (1 + \varepsilon_t) \quad (2)$$

It is then possible to take logarithms in order to approximate $\log(1 + \varepsilon_t)$ by ε_t to obtain⁹

$$\log(x_t) \approx \log(x_0) + \varepsilon_1 + \varepsilon_2 + \dots + \varepsilon_t = \log(x_0) + \sum_{s=1}^t \varepsilon_s \quad (3)$$

⁷ A Pareto firm size distribution has been reported by many authors, such as Axtell (2001) and Gaffeo et al. (2003) (see also de Wit (2005) for a survey). However, a Pareto firm size distribution is not uncontroversial. Some scholars argue that the Pareto is only a good fit for the upper tail of the firm size distribution (Marsili, 2005; Growiec et al., 2008).

⁸ Such as the model in Simon (1955); see de Wit (2005) for a survey.

⁹ This logarithmic approximation is only justified if ε_t is 'small' enough (i.e. close to zero), which can be reasonably assumed by taking a short time period (Sutton (1997)).

In the limit, as t becomes large, the $\log(x_0)$ term will become insignificant, and we obtain:

$$\log(x_t) \approx \sum_{s=1}^t \varepsilon_s \quad (4)$$

Central Limit Theorem implies that $\log(x_t)$ is normally distributed, which means that firm size (i.e. x_t) is lognormally distributed:

$$P(x_t) = \frac{1}{x_t \sqrt{2\pi\sigma^2 t}} e^{-\left(\frac{(\ln x_t - \bar{\varepsilon}t)^2}{2\sigma^2 t}\right)} \quad (5)$$

This lognormal firm size distribution corresponds to firms of the same age, within the same cohort.¹⁰ In an extension of the model, however, we need no longer assume that t has the same value for all firms. Instead, we suggest that t is itself a random variable. It seems reasonable to assume the distribution of firm age to be exponentially distributed. If t is exponentially distributed, we have:

$$P(t) = \lambda e^{-\lambda t} \quad (6)$$

Figure 1 shows what an exponential distribution looks like on linear axes (left) and also with a logarithmic y-axis (right).

In order to obtain the mixture of these two distributions, we apply the following rule: if the distribution of a variable a , $p(a, b)$, depends on a parameter b which in turn is distributed according to its own distribution $r(b)$, then the distribution of a is given by $p(a) = \int r(b) \cdot p(a, b) db$ (Adamic and Huberman, 1999; Huberman and Adamic, 1999).

This gives us the following:

$$P(x_t) = \int \lambda e^{-\lambda t} \cdot \frac{1}{x_t \sqrt{2\pi\sigma^2 t}} e^{-\left(\frac{(\ln x_t - \bar{\varepsilon}t)^2}{2\sigma^2 t}\right)} dt \quad (7)$$

and, as in Adamic and Huberman (1999), this can be developed to yield:

$$P(x_t) = C \cdot x_t^{-\beta} \quad (8)$$

¹⁰ An interesting and recent strand of literature has investigated how the firm size distribution for young cohorts of firms evolves over time as these cohorts grow older (Cabral and Mata, 2003; Angelini and Generale, 2008; Cirillo, 2010). These studies generally observe that the firm size distribution for very young cohorts is particularly skewed to the right, but that $\log(\text{size})$ becomes less skewed and more symmetric (i.e. size approaches the log-normal) as cohorts grow older. The model presented here assumes that the size distribution within cohorts is lognormal, which therefore might not be entirely appropriate for cohorts of very young firms.

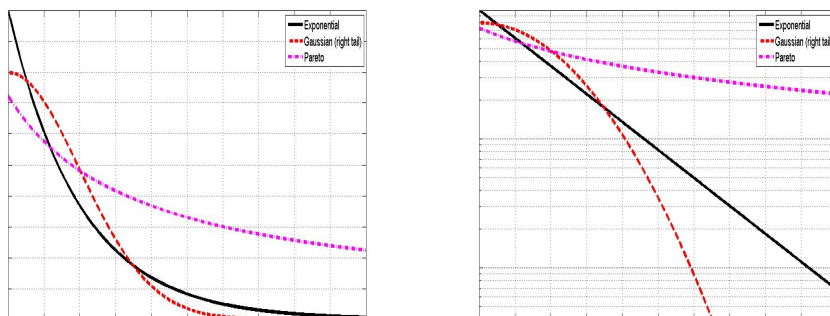


Figure 1: Exponential distribution plotted alongside a Gaussian right tail and a Pareto distribution, on linear axes (left) and with a logarithmic y-axis (right).

where C is a constant and is given by $C = \lambda / \sigma (\sqrt{(\bar{\epsilon} / \sigma)^2 + 2\lambda})$. The exponent β is in the range $[1, \infty]$ and is determined by $\beta = 1 - \frac{\bar{\epsilon}}{\sigma^2} + \frac{\sqrt{(\bar{\epsilon}^2 + 2\lambda\sigma^2)}}{\sigma^2}$. When the mean growth rate is close to 0%, $\bar{\epsilon}$ will be close to 1. As a result, if λ is small (implying that the exponential decay is relatively weak, i.e. that it is not uncommon to find firms with an age much greater than one)¹¹, and if σ is small (which is not implausible either), then the exponent β will be close to Zipf's value of 1, which has been observed in empirical work on US firms (Axtell (2001)).

3 The Aggregate Age Distribution

The scant empirical evidence on the age distribution suggests that the exponential distribution is a valid heuristic.¹² Figures 2 and 3 shows the age distribution for Indian small scale businesses and also for Spanish firms. Figure 2 shows that, even in a sample of small businesses, these firms have very different ages. Most firms are relatively young, but some are extremely old. Figure 4 shows the age

¹¹ This condition is trivial since the duration t of a Gibrat-type 'shock' can be made arbitrarily short.

¹² Cook and Ormerod (2003) suggest a power law distribution for age, but this is not based on any direct observation of data, but observation of aggregated data and speculation about the disaggregated structure generating the aggregated data. Fagiolo and Luzzi (2006, p. 31) observe a log-normal age distribution, but as discussed above this may simply be an artifact of their database, in which very young firms are apparently under-represented.

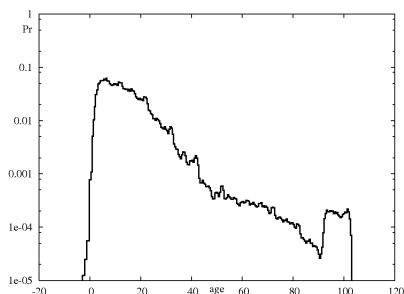


Figure 2: Kernel density of the age distribution of Indian small scale businesses in 2003.
Source: Coad and Tamvada (2008).

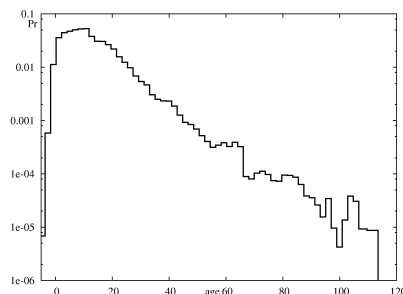


Figure 3: Kernel density of the age distribution of Spanish firms in 2005.
Source: authors elaboration based on the data in Segarra et al. (2008, p.92).

distribution of a census of Italian firms of all sizes in the year 2000, based on the data in Bottazzi et al. (2008). Furthermore, analysis of the age distribution of Irish firms can be found in Kinsella (2009). These distributions appear to be well approximated by a straight line of negative slope over most of the support, covering several orders of magnitude, which on these semi-log axes would signify that the empirical distribution is well approximated by an exponential distribution. Although data on firm age may contain a certain amount of measurement error, the fact that these three diagrams constructed from independent datasets are in conformity with each other is encouraging.

In an attempt to go beyond mere ‘eyeball-testing’ and provide a parametric characterization of the aggregate age distribution, we perform some Subbotin distribution estimation in order to obtain maximum likelihood estimates of the distribution parameters (Bottazzi (2004)). In doing so, we should bear in mind that very young firms are likely to be under-represented in these databases. For the Indian data in Figure 2 the modal age is 6 years, while it is 10 years in the Spanish data in Figure 3. As a consequence, we estimate an asymmetric distribution, with the mode specified as 6 years in the case of the Indian data and 10 years in the case of the Spanish data. The functional form of the asymmetric Subbotin distribution (also known as the exponential power family of densities) depends on five parameters: a positioning parameter m , two scale parameters a_l and a_r , respectively, for the values below or above m , and two shape parameters b_l and b_r .

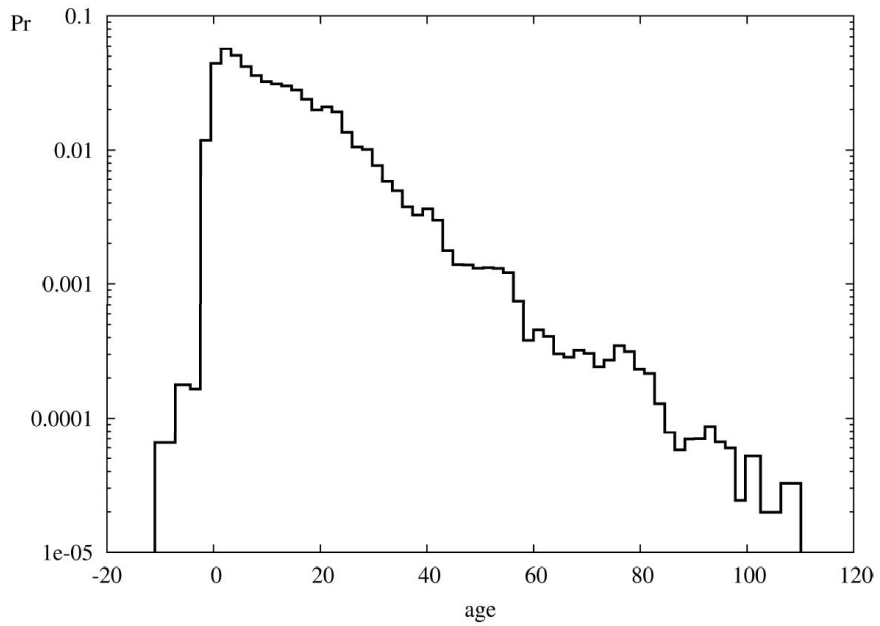


Figure 4: Kernel density of the age distribution of Italian firms in 2000, based on the Unicredit dataset in Bottazzi et al. (2008).

characterizing, respectively, the lower and upper tail of the density. The functional form is:

$$f(x; a, b, m) = \begin{cases} \frac{1}{A} e^{-\frac{1}{b_l} \left| \frac{x-m}{a_l} \right|^{b_l}} & x < m \\ \frac{1}{A} e^{-\frac{1}{b_r} \left| \frac{x-m}{a_r} \right|^{b_r}} & x > m \end{cases} \quad (9)$$

where

$$A = a_l b_l^{1/b_l} \Gamma(1 + 1/b_l) + a_r b_r^{1/b_r} \Gamma(1 + 1/b_r) \quad (10)$$

We focus on the b_r parameter, which characterizes the age distribution above the mode. If $b_r = 2$, we would have a Gaussian right-tail, whereas with $b_r = 1$ we would have an exponential distribution. The estimates for b_r are 1.219 for the Indian data (with a model (negative) log-likelihood of 3.463) and 1.069 for the Spanish data (with a model (negative) log-likelihood of 3.494).¹³

The evidence presented in Figures 2, 3 and 4, as well as asymmetric Subbotin analysis, suggests that the exponential distribution appears to be a reasonable approximation for describing the central part of the distribution, but little attention was given to the age distribution for the youngest firms. The modal age of firms is 6 years in the Indian dataset (displayed in Figure 2), and it is 10 years in the Spanish dataset (displayed in Figure 3), suggesting that very young firms are under-represented in these datasets. These values for the modal age contrast with the modal age implied by an exponential distribution, which would correspond to the very youngest age group. To investigate the age distribution of very young firms, therefore, it will be necessary to use a different database. Furthermore, the age structure for the oldest firms was not well explored. In the rest of the paper, we investigate the age distribution for these two extremes, focusing on the age distribution for very young establishments and also very old firms, taking the exponential distribution as a reference point. We also show that, while the exponential is a good representation of the age distribution at the aggregate level, it is not always valid for individual sectors (such as the international airline sector).

¹³ To my knowledge, standard errors for b_r are not calculated in asymmetric Subbotin estimation in Subbotools 0.9.8.1.

4 The Age Distribution of Young Establishments

In this section we investigate the age distribution of very young establishments, beginning with a survey of the previous literature on the survival rates of new cohorts, and then moving on to an analysis of Census data on new US establishments.

4.1 Previous Literature

In an economic system characterized by constant entry rates, an exponential age distribution in a cross-section of firms implies a constant survival rate for firms. Consider again the exponential age distribution:

$$P(t) = \lambda e^{-\lambda t} \tag{11}$$

The probability of a firm being of age t is $\lambda e^{-\lambda t}$, while the probability of a firm being of age $t + 1$ is equal to $\lambda e^{-\lambda(t+1)} = \lambda e^{-(\lambda t + \lambda)} = \lambda e^{-\lambda t} \cdot e^{-\lambda}$, where $e^{-\lambda} < 1$ is the survival rate. In other words, the probability of a firm surviving to age $t + 1$ is equal to the probability of it surviving to age t , multiplied by the survival rate $e^{-\lambda}$.

While the assumption of constant entry rates is seen to be approximately valid at the aggregate level (we explore this later), the hypothesis of constant survival rates for firms of different ages is rejected for small firms. Figure 5 summarizes results from previous research that shows how the exit hazard decreases over time for new plants and firms. Figure 5 is plotted with a logarithmic y-axis, because a constant survival probability for different years would be represented by a straight line on these axes. The lines do appear to be approximately straight, but a closer inspection suggests that they are slightly ‘droopy’ or convex with respect to the origin. As such, it is worth investigating whether or not survival probabilities are constant or increasing over time in the case of new plants and firms. Our analysis on the BDS dataset on new US establishments in the following section complements these studies by providing stronger evidence that annual survival rates increase in the years immediately following entry.

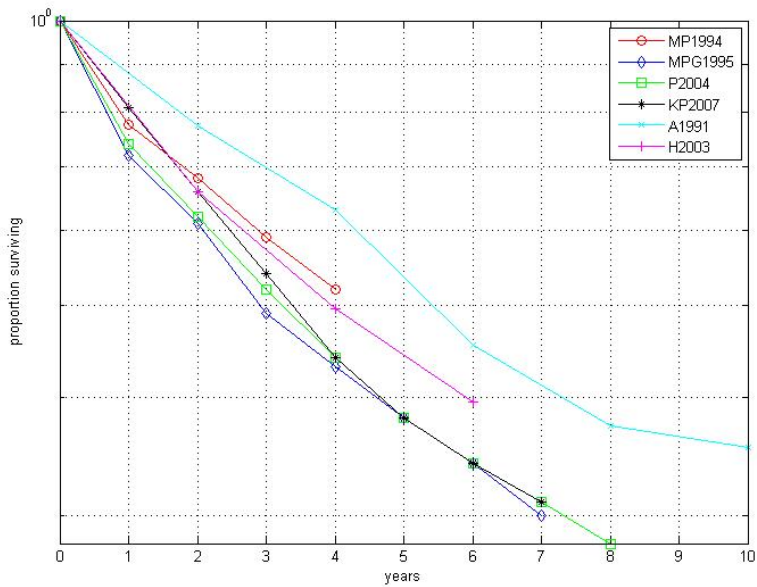


Figure 5: Percentage of surviving plants and firms reported in previous analyses. Note the log scale on the y-axis – if the survival probability is constant across years, the data should be represented as a straight line on these semi-log axes.

Source: the legend refers to the following data respectively: Mata and Portugal (1994, p. 235) on Portuguese firms (1983 cohort), Mata et al. (1995, p. 468) on Portuguese plants (1983 cohort), Persson (2004, p. 428) on Swedish plants (1987 cohort), Audretsch (1991, p. 443) on US firms (1976 cohort), Headd (2003, p. 59) on US firms (1989-1998) and Knaup and Piazza (2007, p. 8) for US plants (1998 cohort).

4.2 Database

The database we analyze is publicly available and can be found online at the following URL: http://www.ces.census.gov/index.php/bds/bds_database_list.¹⁴ The Business Dynamics Statistics (BDS) database is a comprehensive government database on the population of young US establishments (also referred to hereafter as plants), which contains relatively detailed information on the number of young establishments and their ages.¹⁵

Birth year is defined as the year an establishment first reports positive employment in the US LBD database. Establishment age is computed by taking the difference between the current year of operation and the birth year. Given that the LBD series starts in 1976 observed age is by construction left censored at 1975.

In the case of multi-plant firms, establishments are assigned a firm age based upon the age of the parent firm. This reflects the idea that new establishments that are set up by incumbent parent firms can already benefit from their parents market experience, and so are not considered to be entirely new establishments. The age of the parent firm, in turn, is based on the age of the oldest establishment in the firm. The vast majority of new firms are single-unit firms, however.

The vast majority of establishment openings are true greenfield entrants. Similarly, the vast majority of establishment closings are true establishment exits. Note, however, that mergers and acquisitions and divestitures could lead to abrupt changes in firm age purely from establishment composition issues if we defined firm age in each year using age of the oldest establishment owned in that year. Unfortunately there is no way to control for this effect in the database.

We begin by taking the number of plants of age 0, 1, 2, 3, 4, 5, 6-10, 11-15, 16-20, 21-25, and 26+. While the BDS database has detailed information on the ages of very young plants, older plants are grouped together into age classes such as the ‘6-10 years’ age class. We take the midpoint of those classes that span more than one year (e.g. firms in the 6-10 year class are represented by the age 8), and divide the total number of plants in the class by the number of years spanned

¹⁴ The online database was accessed and downloaded by the author on the 21st November 2009.

¹⁵ Since the data series on Business Dynamics Statistics are based on administrative rather than sample data, there are no issues related to sampling error. Nonsampling error, however, still exists. Nonsampling errors can occur for many reasons, such as the employer submitting corrected employment data after the end of the year as well as late filers. Other sources of error include typographical errors made by businesses when providing information. Such errors, however, are likely to be distributed randomly throughout the dataset.

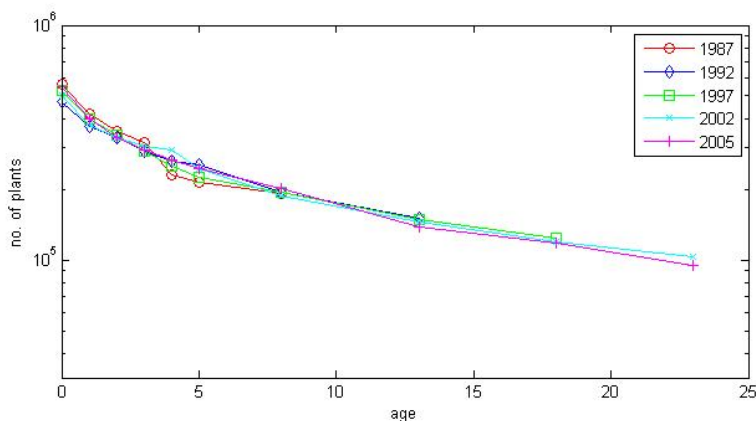


Figure 6: Aggregate age distribution for different years

by the class, to obtain a representative frequency for the midpoint. We ignore the last category (26+) because it is unbounded.

4.3 Analysis

Figure 6 plots the aggregate age distribution for different years. Instead of pooling the years together, we focus on the age distribution for individual years.¹⁶ Nonetheless, we observe that the age distribution changes little over time. We also observe that the age distribution is visibly convex with respect to the origin, whereas an exponential age distribution for this sample of young plants would suggest a straight line.

An aggregate age distribution such as the one observed here can be decomposed into two distinct factors. First, it could arise because the number of entrants in each year is steadily increasing (i.e. approximately exponentially increasing). Second, it could be the artifact of the survival rates within cohorts, such that a roughly constant proportion of plants within any cohort exit each year. The first scenario is investigated in Figure 7. The number of entering plants appears to be

¹⁶ It doesn't make good sense to pool years together, because the age distribution taken from different years is not independent. For instance, if in one year an unusually large number of establishments are observed to enter, this will probably be visible in cross-sections of the age distribution taken in subsequent years.

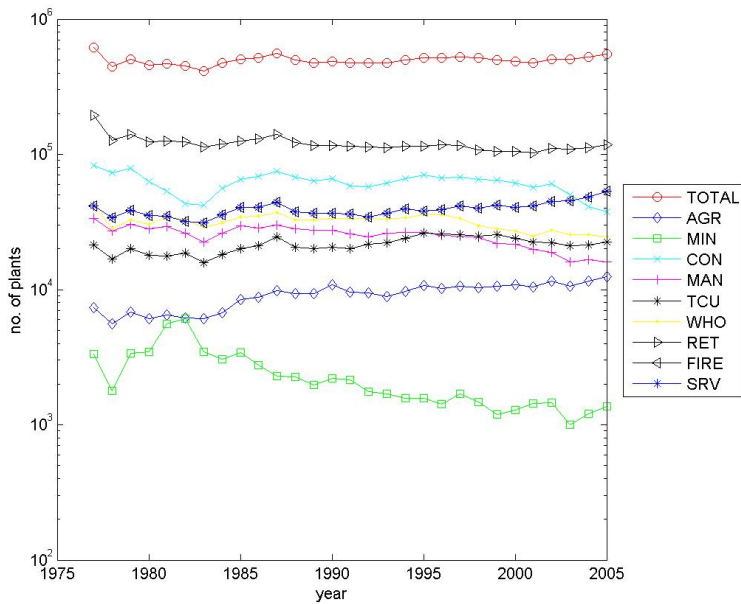


Figure 7: Number of establishments aged zero in each year. By order of appearance in the legend, the sectors are: Agricultural services, forestry and fishing; Mining; Construction; Manufacturing; Transportation and public utilities; Wholesales trade; Retail trade; Finance, insurance and real estate; Services.

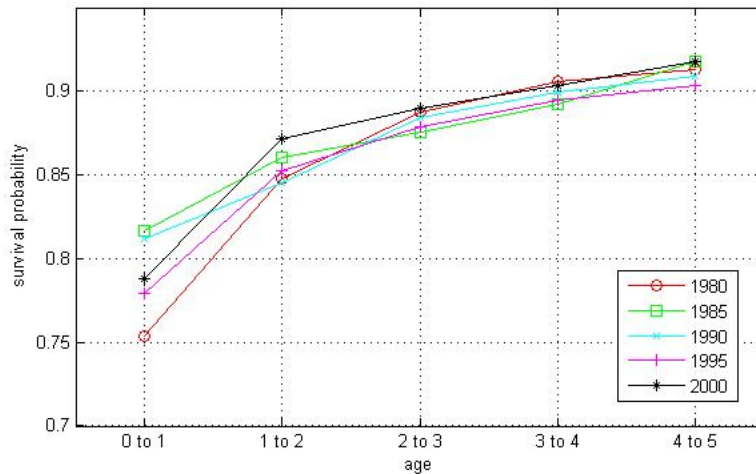


Figure 8: Survival rates of cohorts at different ages

roughly constant over the time period at the aggregate level, although at a sectorally disaggregated level the pattern is much messier. For instance, the number of entrants per year appears to be steadily increasing in the Agricultural services, forestry and fishing sector, for example, while it tends to decrease in the Mining sector (which is a relatively mature industry). The second scenario is investigated in Figure 8, which plots the survival rates for different cohorts. In each year a certain proportion of establishments are observed to exit. This proportion is not constant over time, however. The observation that the survival rate is lower for very young plants is consistent with the unexpectedly high number of very young plants in the aggregate age distribution that was observed in Figure 6.

Figure 8 shows that survival rates are lowest over the period 0-1 year, and that they rise steadily over time. Survival of the first year is hardest, but survival gradually becomes easier over time. While only 79% survive their first year, 91% of establishments survive from their 4th year to their 5th year, on average. Needless to say, these differences in survival probabilities are highly statistically significant (for details see Table 1).

To summarize, the exponential age distribution does not hold in the case of very young plants because, although the number of entrants is roughly constant

across years, the youngest plants are observed to have a higher exit hazard. This stands in contrast to a constant exit hazard over time predicted by the exponential age distribution benchmark.

5 The Age Distribution of the Oldest Firms

In this section we investigate the upper tail of the firm age distribution — the case of ‘Methuselah’ firms. Our dataset on the oldest firms includes a high proportion of family firms,¹⁷ because joint-stock corporations do not face the same survival imperatives as family firms, where firm failure might be interpreted in terms of letting the family down. To give an idea of the kind of ages these firms reach, Table 2 shows the ages of the world’s oldest family firms. Some old firms, such as family firms, go to great lengths to continue their operations (Bertrand and Schoar, 2006). Consider for example the gun-maker Beretta (founded in 1526): the current president is Ugo Gussalli Beretta, who was adopted by his childless uncle in order to inherit the Beretta name and keep the succession within a direct family line.¹⁸ In cases such as this, firms may continue for reasons that are not purely commercial, and as a result we may expect departures from the exponential age distribution benchmark due to the extreme longevity of a small number of firms at the upper tail of the age distribution.

In this section we investigate the age distribution of the world’s oldest firms, basing our analysis on databases available on the internet.¹⁹ Figure 9 shows a

¹⁷ Among the world’s oldest companies, there is a disproportionately high number of family firms. For example, the world’s oldest company is a family company, and the 8 oldest family companies are all among the 44 oldest companies.

¹⁸ See <http://www.familybusinessmagazine.com/worldsoldest.html> (accessed 18th January 2010).

¹⁹ The main dataset comes from http://en.wikipedia.org/wiki/List_of_oldest_companies (accessed on the 14th December 2009), and is based on data from Tokyo Shoko Research, Japan. This dataset of the oldest companies includes brands and companies, excluding associations, educational, government or religious organizations. To be listed, a brand or company name must remain, either in whole or in part, since inception. If the original name has subsequently changed due to acquisition or renaming, this must be verifiable. Age is calculated as 2009 minus year of founding. Another useful data source is http://www.servinghistory.com/topics/List_of_oldest_companies (accessed on the 25th January 2010), which contains a list of the world’s oldest companies as well as describing their line of business and providing external links to the companies’ websites (interestingly enough, all of these firms seem to have their own websites). Some minor discrepancies can be found between the Wikipedia (W hereafter) database, the Serving History (SH) database, and the Economist (E) database on old family firms presented in Table 2. W and SH report the age of Hoshi Ryokan as

Table 1: Survival rates for the years following entry for different cohorts of young establishments. Survival rates for individual cohorts are followed by average survival rates, and pairwise two-sample *t*-tests that reject the hypotheses that the survival rates are constant over time within ageing cohorts.

	0-1	1-2	2-3	3-4	4-5
1977	0.7702	0.8610	0.8780	0.8617	0.9221
1978	0.8974	0.8238	0.8552	0.8905	0.9221
1979	0.7349	0.8589	0.8737	0.9046	0.9153
1980	0.7539	0.8479	0.8876	0.9057	0.9132
1981	0.7969	0.8590	0.8855	0.9047	0.9100
1982	0.7717	0.8538	0.8855	0.8952	0.9062
1983	0.8282	0.8493	0.8751	0.9074	0.9071
1984	0.8727	0.8605	0.8931	0.8993	0.9055
1985	0.8168	0.8602	0.8757	0.8922	0.9180
1986	0.8059	0.8408	0.8708	0.9035	0.9032
1987	0.7571	0.8398	0.8854	0.8910	0.9047
1988	0.7826	0.8627	0.8816	0.8890	0.9114
1989	0.8121	0.8643	0.8719	0.9019	0.9101
1990	0.8121	0.8455	0.8843	0.8991	0.9093
1991	0.7824	0.8596	0.8815	0.9020	0.9142
1992	0.7749	0.8526	0.8831	0.9007	0.9075
1993	0.7765	0.8535	0.8828	0.8997	0.9069
1994	0.7765	0.8563	0.8781	0.8926	0.9054
1995	0.7794	0.8521	0.8787	0.8951	0.9034
1996	0.7758	0.8488	0.8740	0.8927	0.9028
1997	0.7689	0.8485	0.8781	0.8941	0.9116
1998	0.7943	0.8702	0.8976	0.9117	0.9230
1999	0.7874	0.8628	0.8942	0.9064	0.9120
2000	0.7881	0.8718	0.8899	0.9038	0.9175

Table 1 (continued)

Average	0.7924	0.8543	0.8809	0.8977	0.9109
Std Dev	0.0357	0.0105	0.0090	0.0098	0.0062
<i>t</i> -stat	0-1 & 1-2		-8.1541		
<i>p</i> -value			1.75E-10		
DoF			46		
<i>t</i> -stat	1-2 & 2-3		-9.4322		
<i>p</i> -value			2.53E-12		
DoF			46		
<i>t</i> -stat	2-3 & 3-4			-6.1847	
<i>p</i> -value				1.53E-07	
DoF				46	
<i>t</i> -stat	3-4 & 4-5				-5.5873
<i>p</i> -value					1.20E-06
DoF					46

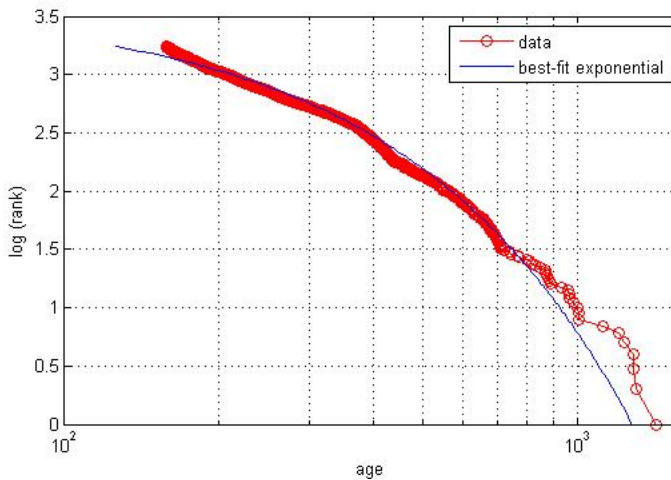


Figure 9: Zipf plot of the age distribution for the world's oldest companies. A best fit exponential is also plotted as a reference.

Table 2: The World's oldest family companies. Dates of founding are approximate in some cases.

	Date of founding	Country
Kongo Gumi	578	Japan
Hoshi Ryokan	718	Japan
Château de Goulaine	1000	France
Fonderia Pontificia Marinelli	1000	Italy
Barone Ricasoli	1141	Italy
Barovier & Toso	1295	Italy
Hotel Pilgrim Haus	1304	Germany
Richard de Bas	1326	France
Torrini Firenze	1369	Italy
Antinori	1385	Italy
Camuffo	1438	Italy
Baronnie de Coussergues	1495	France
Grazia Deruta	1500	Italy
Fabbrica D'Armi Pietro Beretta	1526	Italy
Wiliam Prym	1530	Germany

Source: The Economist (2004), based on data from familybusinessmagazine.com

Zipf plot of the world's oldest companies (on double-log axes). Alongside the plotted data is the best-fit exponential distribution. The empirical distribution is noticeably more right-skewed than the best-fit exponential distribution, indicating that the world's oldest firms are much older than the exponential distribution would predict. On these double-log axes, a straight line would correspond to a Pareto or power law distribution. The concavity of the empirical distribution on these axes therefore indicates that the empirical age distribution is less skewed than the Pareto case.

We consider these extremely old companies as meaningful observations with a plausible economic explanation. We can only remark, however, that these extremely old firms are small in number, and that even though the exponential benchmark for the empirical age distribution is not verified exactly, it remains a useful approximate benchmark in practical terms. Indeed, one might even suppose that the popular empirical methodology of excluding extreme observations as 'outliers' may well overlook this upper-tail phenomenon completely.

6 Sector-level Analysis

Concerning the size distribution, a number of researchers have suggested that the smooth shape observed at the aggregate level is merely a statistical artifact generated through aggregation, and that the disaggregated size distribution observed at the level of individual sectors is much less regular (Dosi et al., 1995). For example, Bottazzi and Secchi (2005) observe significant bimodality in their analysis of the firm size distribution of the worldwide pharmaceutical industry, and relate this to a cleavage between the industry leaders and fringe competitors. Bottazzi et al. (2010) also observe significant bimodality in the French clothing sector.

In this section, we investigate whether the age distribution still resembles the exponential at a more disaggregated level of analysis. Previous work on the

717, whereas it is 718 in E. E describes an Italian firm called Camuffo, founded in 1438; W and SH report that a German firm called Andechs was founded in 1438, but there is no Italian firm called Camuffo. W reports the age of Chivas as 1802 whereas it is recorded as 1801 in SH. Furthermore, SH has 14 'missing' observations when compared to W, corresponding to firms founded in the years 1295, 1398, 1399, 1498, 1499, 1599, 1649, 1699, 1749, 1774, 1799, 1824 (2 firms), and 1851. Thus, SH has only $n=1726$ observations for the years up to 1851, whereas W has $n=1740$. The analysis in Figure 9 is based on W, but it comes as no surprise to discover that when we repeated the analysis on SH data the results were virtually identical.

evolution of specific narrowly-defined industries has suggested that entry and exit processes are not constant over time, and that they may tend to concentrate at certain times such as industry shake-out events. As such, we suspect that the smooth shape of the aggregate age distribution is an aggregate regularity that does not necessarily hold at a disaggregated level.

In this section, we focus on the age distribution of the international airline industry. We focus on this particular industry because we consider it to be a special case that is particularly likely to show a multimodal age distribution. In the early 20th century, when airline technology began to take off, many countries invested heavily in national airlines. International air regulations (in particular, bilaterally-negotiated air traffic rights which were allocated by foreign government departments to specific airline companies by name) provided incentives to governments to subsidize their airlines even if they made losses – and these national airlines did frequently make losses.²⁰ In other words, failure of airlines was often artificially avoided through government intervention. Furthermore, new firms were often denied air traffic rights on key international routes. As a result, we anticipate that the airline industry contains an unnaturally high frequency of old airline companies — a historical characteristic that will presumably be discernable in the empirical age distribution.

Airlines are included in the dataset based on whether they are IATA members. This list of members corresponds to the population of major international airlines. We identified 231 airlines as IATA members on the basis of the member list downloaded from the IATA website.²¹ After scanning the internet for the relevant websites, age data was obtained for all 231 of these companies, and age is measured relative to the year 2010.²²

²⁰ For instance, the Belgian carrier Sabena only reported a positive financial result twice in its 78 year history. These two profitable years were hardly impressive – Sabena made a profit in 1958 only because of the EXPO held in Brussels, and in 1998 due to some financial window-dressing by means of a sales-and-lease-back deal with Flightlease (Swissair) (Knorr and Arndt, 2004)

²¹ Data was downloaded from http://www.iata.org/membership/airline_members_list?All=true on the 18th January 2010. The main advantages of IATA membership concern international transport (as opposed to transport within domestic airspace), being useful for such issues as interline transport agreements, global distribution systems, foreign currency management, and baggage handling. National airlines and low-cost airlines are therefore not likely to be interested in becoming IATA members.

²² For example, since Adria Airways was founded in 1961, its age will be calculated as 2010 - 1961 = 49. In the few cases where year of founding is not recorded as being the same as the year of commencement of operations, we calculate a company's age on the basis of its year of founding. In

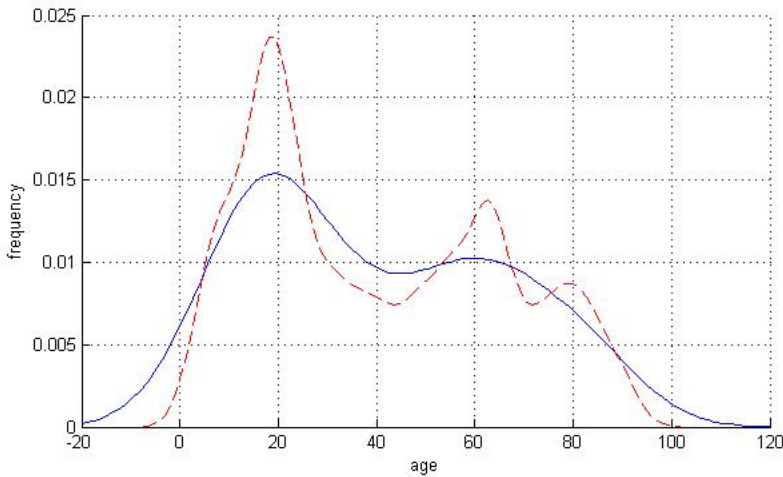


Figure 10: Kernel density of the age distribution for international airlines (IATA members in 2010). Kernel densities obtained using the normal kernel function. The smoother line is the Matlab 7.9.0 default for estimating normal densities ($u=10.5758$). The dotted line is obtained using a kernel bandwidth that is three times smaller than this default value.

The age distribution is presented in Figure 10. The empirical age distribution displays clear multimodality, reflecting the fact that the international airline industry contains many old companies. The exponential distribution is therefore not a useful approximation to the empirical age distribution in this industry, although it provides a useful benchmark which allows us to comment on the unexpectedly large number of old international airline companies.

7 Conclusion

We began the paper by showing some age distribution plots at the aggregate level, observing that the exponential distribution appeared to be a useful approximation over most of the support of the empirical age distribution. In later sections of the paper, however, we focused on situations in which the exponential gave only an

these cases where year of founding and year of commencement of operations are different, year of founding precedes year of commencement of operations by only one year in the majority of cases.

imperfect representation. In contrast to the exponential benchmark, we observed that young establishments seem to be especially numerous, the oldest firms seem to be exceptionally long-lived, and at the disaggregated level of certain specific sectors we can observe a particularly irregular age structure.

In spite of these departures from the exponential distribution, we argue that the exponential is still a useful benchmark for understanding the age structure of industries. Concerning the large number of young establishments, it could be that this excess weight in the age distribution corresponds to over-entry by inefficient establishments who exit shortly afterwards (the case of ‘hopeful monsters’); excess entrepreneurship undertaken by overoptimistic entrepreneurs. (Santarelli and Vivarelli (2007) provide an interesting survey of this phenomenon of over-entry.) To the extent that departures from the exponential benchmark among young establishments represent over-entry, then the exponential age distribution could be used to gauge the magnitude of this phenomenon.

Departures from the exponential benchmark in the case of the oldest firms also have a ready economic explanation, in that certain long-lived firms, and especially family firms, do not pursue economic rationality in the sense of maximization of expected profits, but instead they may seek to maximize their chances of survival (e.g. by pursuing risk-averse strategies).

We also presented evidence that the exponential distribution may not always be a valid heuristic at the disaggregated level of individual sectors. We focused on a particular sector that we suspected of having an irregular age distribution — the international airline industry. In contrast to the smooth shape observed at the aggregate level, the age distribution of this particular sector is much messier and displayed conspicuous multimodality.

To summarize, therefore, there are a number of situations in which the empirical age distribution strays from the exponential benchmark. Nonetheless, we consider the exponential to be a useful approximation. In the words of Herbert Simon, “statistically significant deviations of data from a generalization should not always, or usually, lead us to abandon the generalization” (Simon, 1968, p. 454). We argue that even in those situations where the exponential can be rejected on statistical grounds, it still serves as a useful benchmark against which these distortions can be gauged. For example, we suggest that theoretical models of firm entry, exit, and industry evolution would do well to generate an exponential age distribution as part of their output, even though the empirical data is not exactly exponentially distributed.

This paper contains preliminary investigations on a patchwork of datasets that have detailed information on different parts of the aggregate age distribution. We anticipate that, in future, more detailed databases will become available that provide better coverage of all sections of the aggregate age distribution. Future work should aim to apply more advanced statistical tests to estimate the shape of the aggregate age distribution. While we have suggested that the exponential distribution is a good fit to the empirical density, other candidate distributions such as the Pareto, or intermediate cases (between the exponential and the Pareto) could also be fruitfully investigated.

Acknowledgements: I am grateful to Zakaria Babutsidze, Marco Capasso, Tommaso Ciarli, Christina Guenther, Brian Headd, Geoff Hodgson, André Lorentz and Ulrich Witt and seminar participants at the Max Planck Institute of Economics, Jena, for helpful discussions; to Mercedes Teruel-Carrizosa for help in preparing Figure 3, and to Federico Tamagni for preparing Figure 4. Katja Mehlis provided excellent research assistance. I am also grateful to Thomas Jeitschko (the associate editor), two anonymous referees, and to some anonymous *E-economics* readers for many helpful comments. The usual caveat applies.

References

- Adamic, L. A., and Huberman, B. A. (1999). The Nature of Markets in the World Wide Web. SSRN working paper IEA5. DOI: 10.2139/ssrn.166108.
- Angelini, P., and Generale, A. (2008). On the Evolution of Firm Size Distributions. *American Economic Review*, 98(1): 426–438.
- Audretsch, D. (1991). New-firm survival and the technological regime. *Review of Economics and Statistics*, 73(3): 441–450.
- Axtell, R. L. (2001). Zipf distribution of US firm sizes. *Science*, 293: 1818–1820.
- Barron, D. N., West, E., and Hannan, M. T. (1994). A Time to Growth and a Time to Die: Growth and Mortality of Credit Unions in New York, 1914-1990. *American Journal of Sociology*, 100(2): 381–421.
- Bertrand, M., and Schoar, A. (2006). The role of family in family firms. *Journal of Economic Perspectives*, 20(2): 73–96.
- Bottazzi, G. (2004). Subbotools User’s Manual. LEM Papers Series 2004/14, Laboratory of Economics and Management (LEM), Sant’Anna School of Advanced Studies, Pisa, Italy.
- Bottazzi, G., Coad, A., Jacoby, N., and Secchi, A. (2010). Corporate Growth and Industrial Dynamics: Evidence from French Manufacturing. *Applied Economics*, Forthcoming: DOI: 10.1080/00036840802400454.
- Bottazzi, G., and Secchi, A. (2005). Growth and Diversification Patterns of the Worldwide Pharmaceutical Industry. *Review of Industrial Organization*, 26: 195–216.
- Bottazzi, G., Secchi, A., and Tamagni, F. (2008). Productivity, Profitability and Financial Performance. *Industrial and Corporate Change*, 17(4): 711–751.
- Bruderl, J., and Schussler, R. (1990). Organizational Mortality: The Liabilities of Newness and Adolescence. *Administrative Science Quarterly*, 35(3): 530–547.
- Buenstorf, G., and Klepper, S. (2009). Heritage and agglomeration: The Akron tyre cluster revisited. *Economic Journal*, 119(April): 705–733.

- Cabral, L. M. B., and Mata, J. (2003). On the Evolution of the Firm Size Distribution: Facts and Theory. *American Economic Review*, 93(4): 1075–1090.
- Cirillo, P. (2010). An analysis of the size distribution of Italian firms by age. *Physica A: Statistical Mechanics and its Applications*, 389: 459–466.
- Coad, A. (2010). The exponential age distribution and the Pareto firm size distribution. *Journal of Industry Competition and Trade*, forthcoming, doi 10.1007/s10842-010-0071-4.
- Coad, A., and Tamvada, J. P. (2008). The Growth and Decline of Small firms in Developing Countries. Papers on Economics and Evolution 2008-08, Max Planck Institute of Economics, Evolutionary Economics Group.
- Cook, W., and Ormerod, P. (2003). Power law distribution of the frequency of demises of US firms. *Physica A: Statistical Mechanics and its Applications*, 324(1-2): 207–212.
- de Wit, G. (2005). Firm size distributions An overview of steady-state distributions resulting from firm dynamics models. *International Journal of Industrial Organization*, 23(5-6): 423–450.
- Delacroix, J., and Carroll, G. (1983). Organizational foundings: An ecological study of the newspaper industries of Argentina and Ireland. *Administrative Science Quarterly*, 28(2): 274–291.
- Dosi, G., Marsili, O., Orsenigo, L., and Salvatore, R. (1995). Learning, market selection and the evolution of industrial structures. *Small Business Economics*, 7(6): 411–436.
- Fagiolo, G., and Luzzi, A. (2006). Do Liquidity Constraints Matter in Explaining Firm Size and Growth? Some Evidence from the Italian Manufacturing Industry. *Industrial and Corporate Change*, 15(1): 1–39.
- Fichman, M., and Levinthal, D. (1991). Honeymoons and the liability of adolescence: A new perspective on duration dependence in social and organizational relationships. *Academy of Management Review*, 16(2): 442–468.

- Gaffeo, E., Gallegati, M., and Palestrini, A. (2003). On the size distribution of firms: additional evidence from the G7 countries. *Physica A: Statistical Mechanics and its Applications*, 324(1-2): 117–123.
- Gibrat, R. (1931). *Les inégalités économiques: applications, aux inégalités des richesses, a la concentration des entreprises, aux populations des villes, aux statistiques des familles, etc.: d'une loi nouvelle la loi de l'effet proportionnel*. Recueil Sirey, Paris.
- Growiec, J., Pammolli, F., Riccaboni, M., and Stanley, H. (2008). On the size distribution of business firms. *Economics Letters*, 98(2): 207–212.
- Guenther, C. (2009). Structural change in the German machine tool industry: an empirical investigation of technological, regional and firm-specific developments. Doctoral dissertation, Friedrich-Schiller-Universitaet Jena, Germany. February 4th, 2009.
- Hannan, M., and Freeman, J. (1984). Structural inertia and organizational change. *American Sociological Review*, 49(2): 149–164.
- Headd, B. (2003). Redefining Business Success: Distinguishing Between Closure and Failure. *Small Business Economics*, 21(1): 51–61.
- Headd, B., and Kirchhoff, B. (2009). The Growth, Decline and Survival of Small Businesses: An Exploratory Study of Life Cycles. *Journal of Small Business Management*, 47(4): 531–550.
- Huberman, B. A., and Adamic, L. A. (1999). Growth dynamics of the world-wide web. *Nature*, 401: 131.
- Huergo, E., and Jaumandreu, J. (2004a). Firms' age, process innovation and productivity growth. *International Journal of Industrial Organization*, 22(4): 541–559.
- Huergo, E., and Jaumandreu, J. (2004b). How does probability of innovation change with firm age? *Small Business Economics*, 22(3): 193–207.
- Kinsella, S. (2009). The Age Distribution of Firms in Ireland, 1961–2009. Department of Economics, University of Limerick, Ireland.

- Klepper, S. (2002). Firm survival and the evolution of oligopoly. *RAND Journal of Economics*, 33(1): 37–61.
- Klepper, S., and Thompson, P. (2006). Submarkets and the evolution of market structure. *Rand Journal of Economics*, 37(4): 861–886.
- Knaup, A. E., and Piazza, M. C. (2007). Business Employment Dynamics data: survival and longevity, II. *Monthly Labour Review*, 130: 3–10.
- Knorr, A., and Arndt, A. (2004). Alliance strategy and the fall of Swissair: a comment. *Journal of Air Transport Management*, 10(2): 119–123.
- Marsili, O. (2005). Technology and the Size Distribution of Firms: Evidence from Dutch Manufacturing. *Review of Industrial Organization*, 27(4): 303–328.
- Mata, J., and Portugal, P. (1994). Life duration of new firms. *Journal of Industrial Economics*, 42(3): 227–245.
- Mata, J., Portugal, P., and Guimaraes, P. (1995). The survival of new plants: Start-up conditions and post-entry evolution. *International Journal of Industrial Organization*, 13(4): 459–481.
- Persson, H. (2004). The survival and growth of new establishments in Sweden, 1987-1995. *Small Business Economics*, 23(5): 423–440.
- Reed, W. (2001). The Pareto, Zipf and other power laws. *Economics Letters*, 74(1): 15–19.
- Salter, W. E. G. (1960). *Productivity and Technical Change*. Cambridge University Press, Cambridge UK.
- Santarelli, E., and Vivarelli, M. (2007). Entrepreneurship and the process of firms' entry, survival and growth. *Industrial and Corporate Change*, 16(3): 455–488.
- Segarra, A., Teruel, M., Arauzo, J., and Iranzo, S. (2008). Dinámica empresarial, creación de empleo y productividad en las manufacturas españolas. Project for the Spanish government (Ministerio de Industria, Turismo y Comercio), Madrid.
- Simon, H. (1955). On a class of skew distribution functions. *Biometrika*, 42(3): 425–440.

- Simon, H. A. (1968). On Judging the Plausibility of Theories. In *Logic, methodology and philosophy of science; proceedings of the international congress for Logic, Methodology and Philosophy of Science*, pages 439–459. Amsterdam: North-Holland.
- Stinchcombe, A. (1965). Social structure and organizations. In J. G. March (Ed.), *Handbook of Organizations*, pages 142–193. Rand McNally, Chicago.
- Sutton, J. (1997). Gibrat's Legacy. *Journal of Economic Literature*, XXXV: 40–59.
- The Economist (2004). The world's oldest companies: The business of survival. December 18th, pp122-3.
- Thompson, P. (2005). Selection and firm survival: Evidence from the shipbuilding industry, 1825-1914. *Review of Economics and Statistics*, 87(1): 26–36.
- Winker, P. (1999). Causes and effects of financing constraints at the firm level. *Small Business Economics*, 12(2): 169–181.

Please note:

You are most sincerely encouraged to participate in the open assessment of this article. You can do so by either recommending the article or by posting your comments.

Please go to:

www.economics-ejournal.org/economics/journalarticles/2010-17

The Editor