Small Business Economics (2006) 27: 301–311
DOI 10.1007/s11187-005-8571-z

The Evolution of the Distribution of Plant Size: Evidence from Luxemburg*

Luisito Bertinelli
Olivier Cardi
Teoman Pamukçu
Eric Strobl

ABSTRACT. In a recent paper, [Cabral, L. M. B. and J. Mata, 2003, *American Economic Review*, **93**, 1075–1090] using Portuguese manufacturing data have shown for the first time that the plant size distribution is significantly right-skewed at the early part of the life cycle and then changes to become more log-normal. Using a similar proxy for plant age we compare this stylised fact for both Luxembourg manufacturing and services. Use of plants' true age, however, reveals that the size distribution reverses back towards right-skewness for very old plants. We confirm the finding that the right-skewness is likely driven by financial constraints.

KEY WORDS: financial constraints, plant size distribution

JEL CLASSIFICATION: F23, L11, L60

1. Introduction

Conventional wisdom has it that the plant size distribution is approximately log-normal (see, e.g., Sutton, 1996). In a recent study Cabral and Mata (2003), however, examine the evolution of the plant size distribution in Portuguese manufacturing and, importantly, show that it is initially at birth significantly skewed to the right and only over time evolves to become more log-normal. They conjecture that this pattern arises because plants initially face financial constraints and, therefore, only over time grow to reach their optimal size. Some supporting evidence for this is provided by using the entrepreneur's age and education as a proxy of the financial constraints faced by the plant.

In this paper, we similarly investigate the evolution of the plant size distribution in Luxembourg using exhaustive social security data. Our study adds to the Cabral and Mata (2003) in a number of ways. First, we provide evidence for another country to further establish whether indeed the patterns found by the authors are a stylised fact or just peculiar to Portugal. Second, we also examine the evolution of the plant size distribution in services sectors to determine whether these patterns also hold outside manufacturing. Third, our data provides an exact indicator of plant age. This allows us to investigate whether results found by Cabral and Mata (2003) are driven by their proxy of age, namely workers' maximum tenure. Finally, we also examine whether financial constraints are driving the observed right-skewness by using a different approach than in the Portuguese study.

The remainder of the paper is organised as follows. In the next section, we describe our data set. Section 3 contains our empirical results. Concluding remarks are provided in the final section.

2. Data

Our data set consists of information collected by the Luxembourg Social Security (Inspection Générale de la Sécurité Sociale, IGSS). It contains information, inter alia, on all the persons

Final version accepted on June 9, 2005

Luisito Bertinelli
CREA
Université du Luxembourg
162A av1511, de la Faïencerie, Luxembourg
E-mail: Luisito.Bertinelli@uni.lu

Olivier Cardi
IRES
Université catholique de Louvain

Teoman Pamukçu
STATEC
Ministère de l'Economie
Luxembourg

Eric Strobl
Université de Paris X
Nanterre

employed and affiliated to the social security system in Luxembourg, so that we have at our disposal the whole universe of workers in Luxembourg. Workers employed by the same establishment can be linked via a unique establishment identifier. This allows us to determine and follow establishments and their level of employment over time. Additionally, there is information on the sector of activity (5-digit NACE) and the year of start-up of all establishments, so that one can derive information about the age and industry of each establishment at any point in time. One should also note that if an establishment is taken over it keeps its original establishment number, so that actual start-ups cannot be confused with takeovers. The data is collected on a monthly basis and we have at our disposal information for the month march of each year for a period spanning from 1992 to 2003. For the purposes of the paper here, we exclude the government sector and household employers and split our total sample into manufacturing (NACE5 codes 15000-37990) and services (NACE5 codes 50000-67990) establishments. Our decision to not investigate the size distribution of household employers and the public sector is driven by the realisation that size is likely to be severely restricted for these two sectors and hence their pattern of little interest. The remaining data leaves us with a total number of observations of roughly 194 thousand plant-years.

Our two aspects of interest are plant size and age. As is standard in the literature, including Cabral and Mata (2003), we use total employment as an indicator of the size of a plant at any point in time. In terms of plant age, one should note that Cabral and Mata (2003) do not use the actual plant age in terms of following the distribution of plant size over the life cycle of plants. Rather they implement a proxy, measured as the maximum of workers' tenure within an establishment. As will be discussed in more detail later, this will likely introduce some measurement error in the categorisation of plants by age, in particular for older plants. As we have information on the tenure of workers and the 'true' start-up year of the plant we experiment and compare both measures and their impact on the results.[1]

Figure 1 displays the evolution of employment in manufacturing and services industries since 1992. As can be seen, manufacturing employment, after experiencing an initial dip in the mid 1990s, has remained fairly stable around 40,000. In contrast, employment in the services sector has been consistently rising since the beginning of our sample period, particularly, since the late 1990s. The services sector currently is more than five times that of manufacturing and employs over 200,000 individuals.

A more detailed sectoral breakdown in manufacturing in terms of employment and number of plants is provided in Table I. Accordingly, the food and beverages industry accounts for the largest number of plants, although the metal industry is still the largest manufacturing employer in Luxembourg. In the service industries, shown in Table II, retail trade accounts for the largest number of plants and workers.

We graphed the average size for the two broad sectors in Figure 2. As can be seen, the average size of plants in manufacturing has been on a general decline since the early 1990s – a

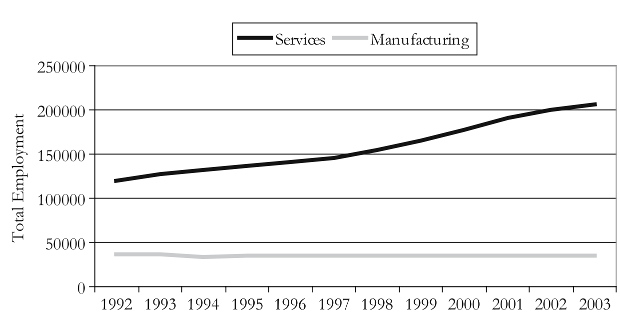

Figure 1. Total employment.

TABLE I
Sectoral breakdown (manufacturing) of employment and plants (2003)

NACE 2-digit manufacturing	Employment		Plants	
	Percentage of total empl.	Percentage of manuf. empl.	Percentage of total no. of plants	Percentage of manuf. plants
Food products and beverages	1.44	11.43	0.83	21.88
Tobacco products	0.18	1.41	0.01	0.27
Textiles	0.16	1.26	0.09	2.42
Wearing apparel	0.01	0.06	0.04	1.07
Wood and of products of wood and cork, except furniture	0.20	1.63	0.08	2.15
Pulp, paper and paper products	0.11	0.86	0.01	0.27
Publishing, printing and reproduction of recorded media	1.00	7.95	0.57	15.17
Chemicals and chemical products	0.52	4.17	0.07	1.74
Rubber and plastic products	1.84	14.62	0.12	3.22
Other non-metallic mineral products	1.13	8.98	0.20	5.23
Basic metals	2.37	18.81	0.07	1.74
Fabricated metal products, except machinery and equipment	1.64	13.00	0.78	20.67
Machinery and equipment n.e.c.	0.84	6.66	0.20	5.37
Electrical machinery and apparatus n.e.c.	0.19	1.53	0.10	2.55
Radio, television and communication equipment and apparatus	0.01	0.09	0.01	0.27
Medical, precision and optical instruments, watches and clocks	0.65	5.21	0.22	5.91
Motor vehicles, trailers and semi-trailers	0.10	0.83	0.05	1.21
Other transport equipment	0.01	0.07	0.01	0.27
Furniture	0.09	0.70	0.26	6.98
Recycling	0.09	0.74	0.06	1.61

pattern that is similarly found for most developed countries. On average a manufacturing plants nowadays employs about 45 individuals. In contrast, services plants have grown very slightly in size, and have on average about 10 employees. Figure 3, depicting the evolution of average plant age, shows that both sectors have experienced a decline in age. It is noteworthy that manufacturing plants tend at any point in time to be somewhat younger than their services counterparts.

Given that this paper focuses on how size evolves over the life cycle, it is of interest to determine how dynamic manufacturing and services sectors in Luxembourg are. Thus, we graph their plant turnover rates, measured as the percentage of plants at any point in time that just started-up or will exit in the following year, in Figure 4. Accordingly, services' plants experience a much greater turnover than those in manufacturing. At any point in time on average about 13% of the plant population are either new or are just about to shut down. For manufacturing this rate is usually 3–5% points lower. It is also noteworthy that plant turnover experiences considerable volatility over our sample period.

3. Empirical analysis

We utilise the data just described to estimate the non-parametric kernel density distribution estimates of the plant size, measured as $x = \log$ (employment), of the manufacturing and services sectors. A kernel density estimator approximates a probability density function $f(.)$ of a random variable X. Assume we have n independent observations $x_1, x_2, \ldots, x_n$ from the random variable X. The kernel density estimator $\hat{f}_k$ for the estimation of the density value $f(.)$ at point x is defined as

TABLE II
Sectoral breakdown (services) of employment and plants (2003)

NACE 2-digit services	Employment		Plants	
	Percentage of total empl.	Percentage of service empl.	Percentage of total no. of plants	Percentage of services plants
Sale, maintenance and repair of motor vehicles and motorcycles; retail sale of automotive fuel	2.43	6.32	2.93	7.64
Wholesale trade and commission trade, except of motor vehicles and motorcycles	4.72	12.27	7.79	20.31
Retail trade, except of motor vehicles and motorcycles; repair of personal and household goods	6.00	15.60	9.29	24.21
Hotels and restaurants	4.25	11.04	8.35	21.77
Land transport; transport via pipelines	4.36	11.33	2.19	5.70
Water transport	0.71	1.83	0.84	2.18
Air transport	1.24	3.22	0.05	0.13
Supporting and auxiliary transport activities; activities of travel agencies	0.99	2.57	0.87	2.27
Post and telecommunications	1.57	4.09	0.28	0.74
Financial intermediation, except insurance and pension funding	9.05	23.50	3.42	8.92
Insurance and pension funding, except compulsory social security	0.95	2.47	0.53	1.37
Activities auxiliary to financial intermediation	2.21	5.73	1.82	4.75

$$\hat{f}_k(x) = \frac{1}{nh}\sum_{i=1}^{n} K\left(\frac{x_i - x}{h}\right)$$

where $K(.)$ denotes the so-called kernel function, n is the number of observations, and h is the bandwidth. A number of possible kernel functions are available (see Haerdle et al., 2003). Here we employ the most widely used kernel function, i.e., the Epanechnikov kernel:

$$K(z) = \begin{cases} \frac{3}{4}\left(1 - \frac{1}{5}z^2\right)/\sqrt{5} & \text{if } |z| < \sqrt{5} \\ 0 & \text{if otherwise} \end{cases}$$

One should note that the kernel density estimator is closely related to a histogram, where histograms are a particular example of a density estimation. The shape of a histogram depends on both the choice of the origin (this is the lower boundary of the first interval) and the widths of

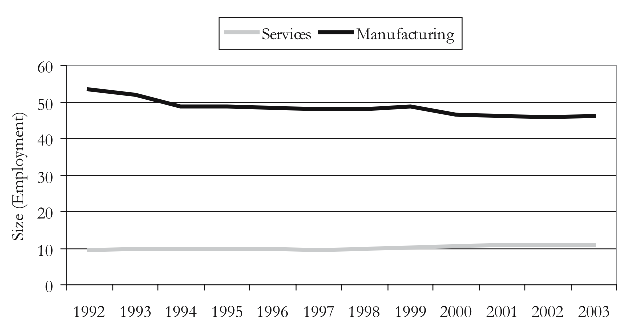

Figure 2. Average size.

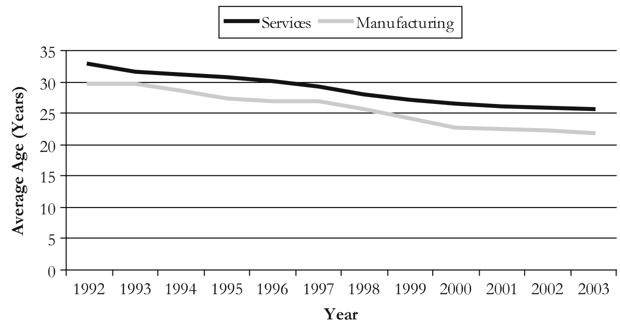

Figure 3. Average age.

the chosen intervals. In contrast, for the kernel density estimates, the range is still divided into intervals, but these intervals are allowed to overlap. One can think of its estimation in terms of a sliding interval, called the window, along the range of the data. Rather than merely counting the number of observations in a window, a weight between 0 and 1 is assigned, based on the distance from the centre of the window, and the weighted values are summed. The function that determines these weights is called the kernel.

The kernel density estimates of the plant size distribution for manufacturing and services plants in Luxembourg are shown in Figure 5.[2] As can be seen, the distribution of the plant size in the services sector is more skewed to the right than that of manufacturing, suggesting a greater incidence of small establishments, as was already suggested by our earlier figure on average size. More importantly, however, neither of these distributions appears to be *log*-normally distributed.

Next we divided our plants into age groups in the spirit of Cabral and Mata (2003) by using the maximum of workers' tenure within an establishment: (a) less than one year, between 2 and 4 years, between 5 and 9 years, between 10 and 19 years, between 20 and 29 years, and at least 30 years of age. The kernel density distributions of these for manufacturing establishments are graphed in Figure 6. As can be seen, at start-up plants are very much skewed to the right. After about 5 years, however, the distribution starts to noticeable shift to the right, moving to what appears to be more and more like a *log*-normal distribution. This process continues across age groups until the very oldest plants, i.e. the plants with the longest tenure. The visual results for Luxembourg are thus qualitatively very similar to those found by Cabral and Mata (2003) for Portugal manufacturing establishments.

The graphical depiction of the plant size distribution of similar age groups of establishments operating in the services sectors in Figure 7 shows that these are characterised by a virtually identical pattern over the plant life cycle. In particular, plant size similarly moves from a

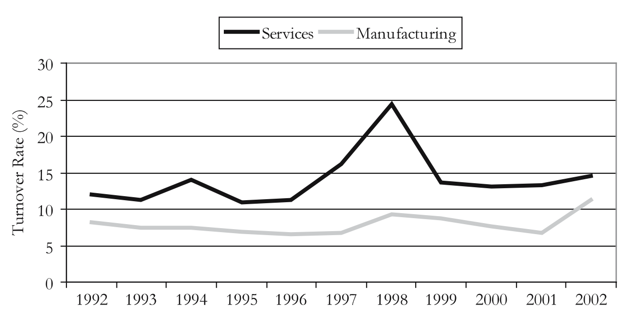

Figure 4. Turnover rate.

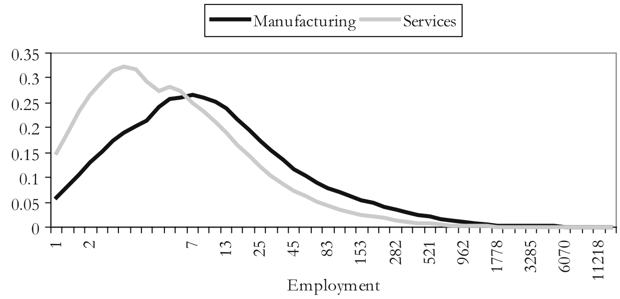

Figure 5. Plant size distribution.

right skewed slowly towards a *log*-normal distribution as plants age. Hence the stylised fact about the evolution of the plant size distribution unearthed by Cabral and Mata (2003) seems to hold for services plants as well.

One possibility may be that our results are due to sample selection bias – i.e., non-surviving plants that exit at an early age drive the right skewness of domestic plants, particularly at earlier ages. To investigate this we follow Cabral and Mata (2003) and depict the size distribution of manufacturing plants that survive at least 10 years in their first and tenth year alongside the first year size distribution of plants that do not survive 10 years in Figure 8. Accordingly, the plant size distribution of survivors and non-survivors in their first year is both skewed to the right. The distribution of survivors after 10 years then clearly moves more towards a *log*-normal distribution. A similar pattern is even more prominent for services sectors, as shown in Figure 9. Thus, sample selection bias due to survival is not driving our overall patterns.

In order to verify our results more formally, we, following Cabral and Mata (2003), employ a parametric test based on the extended gamma distribution, where $w = (\ln s - \mu)/\sigma$ is a function of plant size s with mean μ and standard deviation σ and has the p.d.f.:

$$\begin{cases} \text{if } q \neq 0: |q|(q^{-2})^{q^{-2}} \exp(q^{-2}(qw - \exp(qw)))/\Gamma(q^{-2}); \\ \text{if } q = 0: (2\pi)^{-1/2} \exp(-\frac{1}{2}w^2) \end{cases}$$

such that Γ is the gamma function. Accordingly, if q, the shape parameter is equal to zero, then w, and hence plant size s, follows the standard *log*-normal distribution. On the other hand, if $q < 0$ there is right skewness, while $q > 0$ implies left skewness.

Results for the total sample and the age groups for the manufacturing and services sectors using workers' maximum tenure as a proxy of plant age are provided in the first and second column of Table III, respectively. Accordingly, for both manufacturing and services one can reject the hypothesis of a *log*-normal distribution for the total samples, as evidenced by the

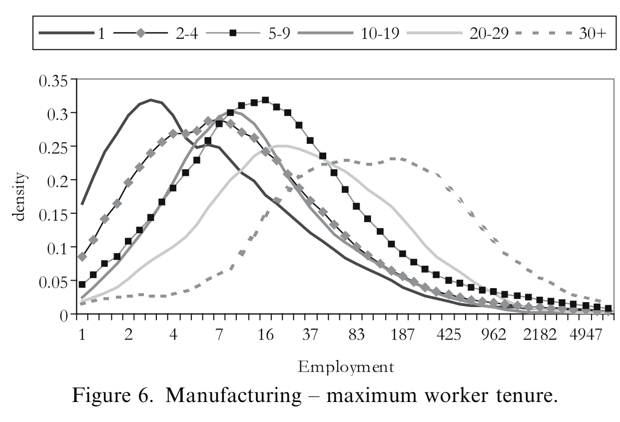

Figure 6. Manufacturing – maximum worker tenure.

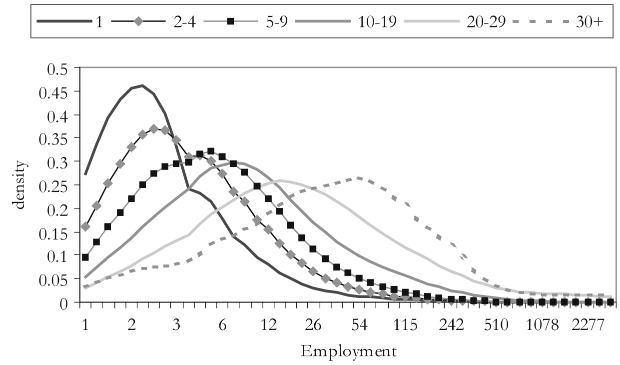

Figure 7. Services – maximum worker tenure.

negative and significant estimate of q. Examining the first age group in manufacturing separately, one discovers that while the estimated value of q continues to be negative and significant, it clearly decreases in absolute size as one moves up the age categories, thus indicating a move towards log-normality. As a matter of fact, by the time establishments reach the age group of 20–29 the null hypothesis of a *log*-normal distribution can no longer be rejected, a result that holds for the oldest age group as well. For services we find a similar story, although the degree of right skewness does not noticeably fall until the age group 5–9. Additionally, log-normality can still be rejected for plants that are aged 20–29 years. Nevertheless, there is evidence that by the age 30 plants have the tendency to follow a *log*-normal distribution in terms of their size.

As suggested earlier, workers' maximum tenure within an establishment may be riddled with considerable measurement error as a proxy for plant age. While one can be confident that it serves as a lower threshold of the establishment's true age, its deviation from the actual age is likely to rise as plants age. This will be influenced by the degree of worker turnover and retirement among those originally employed. To get a preliminary idea of the bias we graph the average maximum worker tenure by establishment age up until 50 years in Figure 10.[3] As one can see, there is even at an early age some bias in the use of the maximum worker tenure as a proxy for the true age of a plant – the further the line departs from a 45 degree line the greater the bias. Nevertheless, this bias seems to be constant for the first 20 years of a plant's existence, as evidenced by the stable slope. More precisely, worker maximum tenure rises proportionally with a plant's age. As the plant, however, reaches its 20th year of existence this trend changes dramatically. From this point onward maximum tenure actually begins to fall with rising age until a plant reaches around 30. A

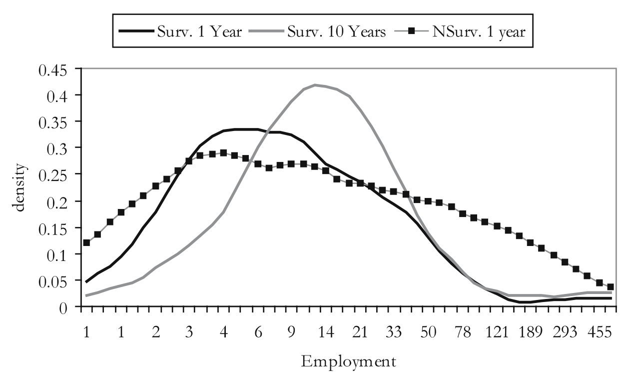

Figure 8. Manufacturing by survival – maximum worker tenure.

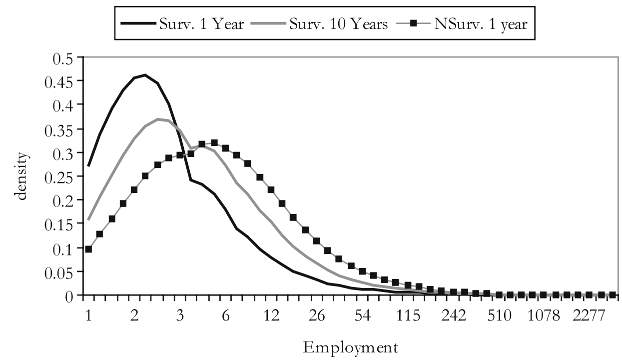

Figure 9. Services by survival – maximum worker tenure.

simple explanation for this is the retirement of the oldest workers in a plant. Thus, while worker maximum tenure rises for establishments above 30 years in existence, clearly many of these will be on average misclassified into much younger age groups.

We depict the plant size distribution by age groups using actual plant age as a measure for manufacturing and services in Figures 11 and 12, respectively. As can be seen for both groups of sectors, while the general patterns evolve similarly to using maximum tenure as a proxy up until the 20–29 age group, the size distribution of the very oldest group seems to revert back to considerable right skewness.[4]

As before, we verify our graphical results using the simple parametric test outlined above. The results using actual plant age are given in the third and fourth columns of Table III for manufacturing and services, respectively. In manufacturing one can reject the null hypothesis of log-normality for the first four age groups. However, when considering the two oldest groups one can, in contrast to using maximum worker tenure as a proxy for age, continue to reject a *log*-normal size distribution. Moreover, it is noteworthy that for the last size group the test statistic is higher than for any of the younger ones. For services, log-normality can under actual plant age be rejected for the oldest group, while results are qualitatively similar to the tenure proxy for earlier ones. One should note that we also verified these results excluding plants above 50 years of age.[5]

TABLE III
Skewness test – Estimates of q

Sector: Age proxy:	Manufacturing Worker tenure	Services Worker tenure	Manufacturing Plant age	Services Plant age
All Plants	−0.840***	−10.630***	−0.840***	−10.630***
	(0.032)	(0.410)	(0.032)	(0.300)
Age ≤ 1	−11.074***	−7.621***	−0.934***	−9.284***
	(0.001)	(0.144)	(0.166)	(0.400)
Age 2–4	−1.091***	−9.418***	−0.693***	−10.413***
	(0.078)	(0.010)	(0.080)	(0.258)
Age 5–9	−0.661***	−0.788***	−0.511*	−10.995***
	(0.052)	(0.023)	(0.065)	(0.210)
Age 10–19	−0.724***	−0.606***	−0.341***	−0.735***
	(0.052)	(0.023)	(0.066)	(0.028)
Age 20–29	−0.148	−0.494***	−0.178*	−14.529***
	(0.095)	(0.058)	(0.104)	(0.928)
Age ≥ 30	0.195	−0.009	−1.751***	−9.576***
	(0.121)	(0.128)	(0.117)	(0.827)

Note: ***, **, and * signify 1, 5, and 10 per cent significance levels, respectively.

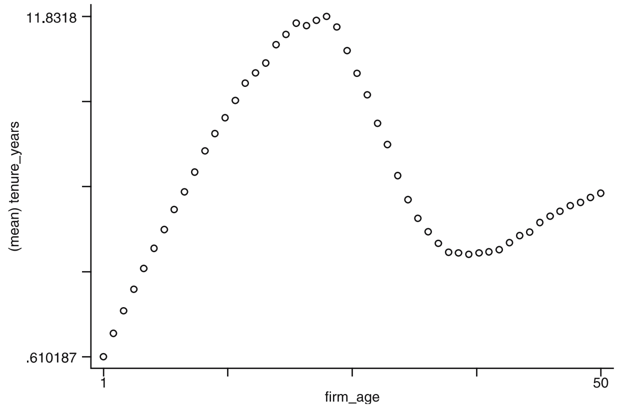

Figure 10. Average of maximum worker tenure by establishment age.

Cabral and Mata (2003) make the argument that the observed patterns in the evolution of the Portuguese manufacturing plant size distribution is due to financial constraints at the early part of the life cycle. They provide indirect support for this by showing that the size of a plant is positively related to the owner's age and level of education, where these two factors are arguably proxies for financial constraints. Given that we do not have such information in our data set we take a slightly different approach to provide supporting evidence of the importance of financial constraints in determining similar distributional patterns found in Luxembourg. Specifically, it is reasonable to argue that financial constraints will play a major role in plants shutting down completely. The actual event of a plant shutdown, however, is unlikely to be an instantaneous event, but rather there will at least be some time period of 'downsizing' prior to the exit from the market. If financial constraints are the driving force behind observed right-skewnesses in size distributions, then one would expect to observe such right-skewness regardless of age just prior to exit.

In order to investigate this proposition we thus graphed the size distribution of plants just prior to exit by our age groups for manufacturing and services in Figures 13 and 14, respectively. As can be seen, for both sets of industries all age groups depict rather similar right-skewness in their distributions. Note that this result cannot exclude Jovanovic (1982)'s efficiency effect, stating that after a series of bad signals regarding its efficiency, a firm gradually reduces output level and eventually exits, even if firms were not subject to financing constraints.

We further complemented our results by employing our parametric test and report the results in Table IV. Accordingly, not only are all values of q significantly negative, just indicating

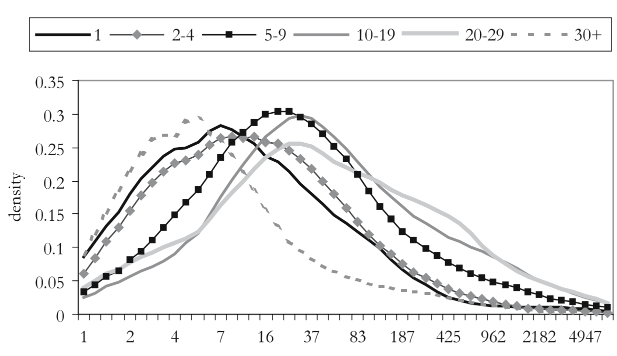

Figure 11. Manufacturing – actual plant age.

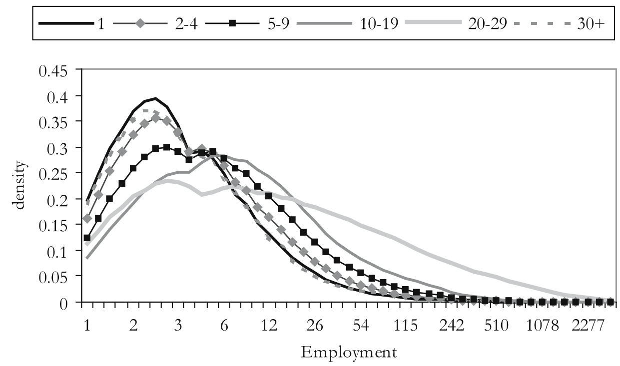

Figure 12. Services – actual plant age.

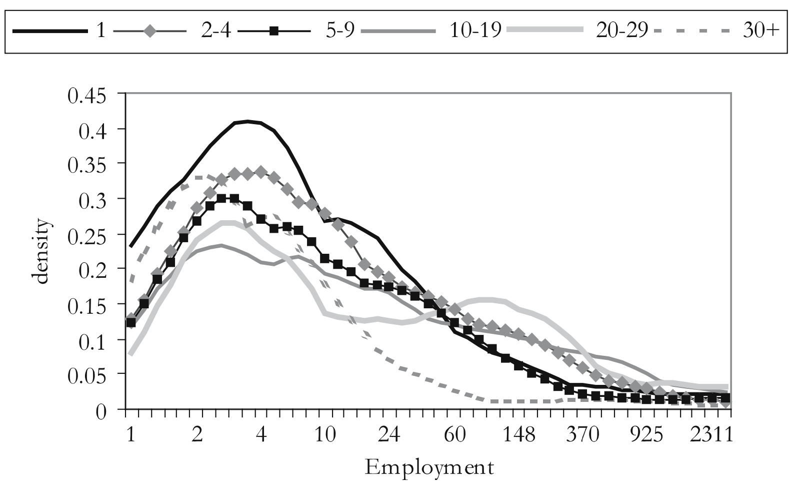

Figure 13. Manufacturing – year prior to exit.

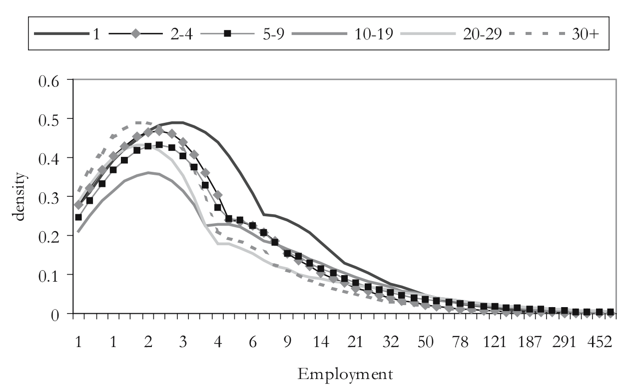

Figure 14. Services – year prior to exit.

TABLE IV
Skewness test – Estimates of q for plant exit

Sector	Manufacturing	Services
All plants	−11.480***	−9.124***
	(1.205)	(0.247)
Age ≤ 1	−11.618***	−10.378***
	(02.831)	(1.168)
Age 2–4	−13.286***	−9.377***
	(0.934)	(1.231)
Age 5–9	−13.324***	−10.121***
	(02.306)	(3.214)
Age 10–19	−15.078***	−11.338**
	(4.431)	(5.127)
Age 20–29	−15.305***	−10.026***
	(4.549)	(0.789)
Age ≥ 30	−10.538***	−8.444***
	(4.207)	(0.166)

Note: ***, **, and * signify 1, 5, and 10 per cent significance levels, respectively.

right-skewness, but their absolute levels do not differ drastically either for manufacturing or services across age groups.

4. Conclusions

In this paper, using data from Luxembourg we corroborate previous findings by Cabral and Mata (2003) in that we find that the plant size distribution is initially right-skewed and evolves over time to a *log*-normal distribution. In contrast to their study, however, we find that this process is reversed for very old plants, in that the distribution of these reverts back towards right-skewness. One likely reason for the difference in our findings is that we used a 'true' measure of plant age rather than a proxy, and results using a similar proxy to the previous authors suggests as much. Similar to the authors, although using a slightly different approach, we also find that financial constraints are likely to play a role in causing right-skewness in the plant size distribution.

Notes

[*] The first author gratefully acknowledges financial support from the National Research Fund of Luxembourg. We are grateful for comments by an anonymous referee.

[1] Importantly, one has to acknowledge for the fact that actually, our data does not allow us to distinguish between the effect of age on the firm size distribution and cohort effects. In order to do so, we would ideally have to follow a cohort of firms of more than X years, but unfortunately, such data is not available to us, nor was it for the Mata and Cabral (2003) study.

[2] We use the optimal smoothing parameter as suggested by Fox (1990).

[3] We restrict this graphical depiction to plants no older than 50 because very few plants greater than half a century exist in the data.

[4] Excluding all plants older than 50 years one gets a similar picture.

[5] Results are available from the authors.

References

Cabral, L. M. B. and J. Mata, 2003, 'On the Evolution of the Plant Size Distribution: Facts and Theory', *American Economic Review* **93**, 1075–1090.

Fox, J., 1990, Describing Univariate Distributions, in J. Fox and J. S. Long (eds.), *Modern Methods of Data Analysis*, Sage.

Girma, S., D. Greenaway and K. Wakelin, 2001, 'Who Benefits from Foreign Direct Investment in the UK?', *Scottish Journal of Political Economy* **48**, 119–133.

Haerdle W., M. Mueller, St. Sperlich and A. Werwatz, 2003, 'Non- and Semiparametric Modelling', Springer.

Ijiri, Y. and H. Simon, 1964, 'Business Plant Growth and Size', *American Economic Review* **54**(2), 77–89.

Jovanovic, B., 1982, 'Selection and Evolution of Industry', *Econometrica* **50**(3), 649–670.

Mansfield, E., 1962, 'Entry, Gibrat's Law, Innovation, and the Growth of Plants', *American Economic Review* **52**(5), 1023–1051.

Sutton, J., 1996, 'Gibrat's Legacy', *Journal of Economic Literature* **35**(1), 40–59.