The Firm Size Distribution: Evidence from Belgium*

Lionel Artige[†‡]

University of Liège

Sousso Bignandi

University of Liège

June 23, 2022

Abstract

The firm size distribution (FSD) is one of the most well-known economic empirical regularity in market economies. Its functional form, postulated in macroeconomic models with heterogenous firms, can be approximated by a parametric distribution. However, the parametric approximations proposed in the literature have long been contested due to the lack of unbiased databases and robust statistical methods. This paper adresses these shortcomings. First, a robust estimation method is proposed to test the fit of parametric distributions and to determine the one offering the best fit at different truncation thresholds. Then, by applying it to a comprehensive database of Belgian firms for the period 2006-2012, the results show that the lognormal distribution is a better approximation of the empirical FSD than the Pareto distribution. These results hold true

*We thank Laurent Cavenaile, Cédric Heuchenne, Bernard Lejeune, Joseph Tharakan and Maren Ulm for their valuable comments and suggestions. Responsibility for all errors is our own.

[†]The authors acknowledge the financial support of BELSPO.

[‡]Corresponding author: Université de Liège, Place des Orateurs 3, B. 31, 4000 Liège, Belgium. E-mail: lionel.artige@uliege.be

at the aggregate, sectoral and regional levels establishing that the shape of the aggregate distribution is not an aggregation artefact arising from the potential distributional heterogeneity of sectoral or regional subsets. This empirical evidence is essential information for tracing the root causes of a plausible underlying stochastic model explaining firm dynamics.

Keywords: Firm size distribution, Gibrat's law, lognormal, Pareto, power law, Zipf.

JEL Classification: C13; L11; L25.

1 Introduction

The firm size distribution (FSD) appears to be similar in all market economies: many small firms and few large corporations. The shape of this distribution is the culmination of a historical process, initiated during the Industrial Revolution, in which self-employed workers, mostly farmers, became over time employees of private firms of varying sizes, particularly in the industrial sector. Gibrat (1931) proposed a probabilistic model based on proportional random growth (known as Gibrat's law) to replicate and explain the dynamics of firms observed during this historical period. He found that the lognormal distribution, which approximates the asymptotic distribution of his random walk model, fitted well the entire empirical distribution of the number of employees in French establishments in 1896, 1901, and 1921. By introducing a minimum size in Gibrat's model, Champernowne (1953) found that the asymptotic distribution is approximated by a Pareto distribution in its upper tail (i.e., the distribution of the largest firms). Simon (1955) modified Gibrat's model by assuming that the number of firms increases over time at a constant rate. The resulting stochastic process is no longer a random walk but a Yule process. Simon and Bonini (1958) apply the Yule process to firm sizes, restrict the constant rate of firm entries to small firms, and find that the Pareto distribution approximates the asymptotic distribution of the Yule distribution in its upper tail. Since then, the findings of the empirical literature have oscillated between the lognormal and the Pareto distributions.

This paper makes an important contribution to this long-standing debate in a multidisciplinary literature marked by disputed estimation methods and databases with incomplete or truncated information. The strengths of our contribution are threefold. First, we address the flaws of the estimation methods usually found in the literature and propose a simple and robust method. Second, we use a complete database on firm sizes in Belgium for each year from 2006 to 2012, thus ruling out the sampling biases that are frequently suspected in the literature. Third, our database includes sectoral and regional information allowing us to study possible aggregation effects arising from the potential distributional heterogeneity of these subsets. Our results show that the lognormal distribution is a better approximation of the empirical FSD than the Pareto distribution at all levels of truncation and disaggregation.

The first major attempt to fit an entire empirical distribution of firm sizes was made by Axtell (2001) who used a comprehensive database of US firms' employees, though with grouped data, and concluded that the entire empirical US distribution of firm sizes fitted well a Zipf distribution, which is a rank-frequency distribution that Zipf (1949) identified in the frequency of words in written English texts. The Zipf distribution, linear on a log-log plot, is a discrete version of a Pareto distribution with a scale parameter equal to one. Since then, the Pareto FSD distribution has been assumed extensively in heterogeneous firm models for its appealing analytical convenience[1]. Nevertheless, Axtell's paper did not close the FSD debate for three reasons. First, the approximate linearity of an empirical distribution on a log-log plot is a necessary but not sufficient condition to conclude that it is a Pareto distribution. Other parametric distributions are possible and should be tested for comparison.[2] Second, the fitting method used by Axtell has been questioned for its reliability (Kleiber and Kotz, 2003; Goldstein et al., 2004; Perline, 2005; Clauset et al., 2009).

[1]See, for instance: Antras and Helpman (2004); Helpman et al. (2004); Luttmer (2007); Rossi-Hansberg and Wright (2007); Chaney (2008); Gabaix and Landier (2008); Eaton et al. (2011). In the trade literature, see among others: Arkolakis et al. (2008), Helpman et al. (2008), Melitz and Ottaviano (2008) and Melitz and Redding (2015).

[2]A lognormal distribution with a high enough value of its variance relative to its mean can look like a straight line in a log-log plot. For an example involving lognormal, Pareto, and exponential distributions, see Figure 5a in Clauset et al. (2009).

Third, the result obtained by Axtell could be specific to the distribution of US firms. Unfortunately, the accessibility of comprehensive national databases on firm sizes is still too rare to generalize his conclusion. As far as we know, three other papers use complete ungrouped data. Bee et al. (2017) test the fit of the distribution of the total revenues of all French firms in 2003 by the Pareto and lognormal distributions and find that neither of them provides a good fit to the French entire FSD but the lognormal distribution is a better approximation of French firm sizes. The other two papers use the method and algorithm proposed by Clauset et al. (2009) to estimate which of the lognormal and Pareto distributions best fits the empirical distribution. Montebruno et al. (2019) find that the Pareto distribution best fits the employment distribution of the 19th-century firms in England and Wales. Using the same US firm data as Axtell (2001), a contemporary paper by Kondo et al. (2020) eventually finds that the lognormal distribution provides a better fit.

Like these last two papers, our work uses a complete database and the algorithm of Clauset et al. (2009) – more precisely its version for R-software written by Gillespie (2015, 2020) – to estimate the fit of the lognormal and Pareto distributions to the empirical distribution. However, we question the validity of the test that Clauset et al. (2009) propose to conclude on the best fit between the two parametric distributions. Their test allows the two distributions to be tested on different samples, which violates the assumptions of hypothesis testing. We modify their test by imposing the same sample on both tested distributions and apply it to the empirical distribution of ungrouped Belgian firms each year from 2006 to 2012 at the aggregate, sectoral and regional levels. The result is unambiguous at the aggregate, sectoral and regional levels: of the two candidate distributions, the lognormal distribution provides the best fit in all cases. Thus, the sectoral and regional information in our database allows us to conclude that the best fit obtained by the lognormal distribution is not the result of an aggregation artifact.

The rest of the paper is organized as follows. Section 2 describes our exclusive complete database on Belgian firm sizes. Section 3 discusses our estimation approach. We then present the estimation results for the aggregate level (Section 4), the sectoral level (Section 5), and the regional level (Section 6). Section 7 concludes.

2 Data

This paper uses an exclusive complete database of firm sizes in Belgium that was obtained from the Belgian Ministry of Economy (SPF Economie - Direction générale Statistique - Statistics Belgium). The database includes the exact number of employees of all registered firms and establishments of the NACE sectors A to N in Belgium for each year from 2006 to 2012.[3] In Belgium, each enterprise must provide its list of employees to the social security administration every quarter. Our database contains the exact number of employees of the last quarter. In this study, the unit of observation is the firm, which may be a combination of several legal units if they share a common economic activity. This choice is justified by the fact that economic decisions, such as hiring and firing decisions, are made at the firm level. In 2012, for instance, there were 202,480 private firms in Belgium hiring more than 2.28 million employees (Table 1). The average size was 11.2 employees while the median size was 3 employees, emphasizing the right-skewness of the distribution.

[3] For the list and description of the 2008 NACE sectors in the European Union, see Appendix A.

	Firms	Employees	Median	Mean	Std Dev	skewness
2006	201,677	2,080,570	3.00	10.32	55.28	39.43
2007	197,731	2,069,337	3.00	10.47	55.75	38.61
2008	204,563	2,264,683	3.00	11.07	59.34	46.57
2009	203,424	2,230,109	3.00	10.96	57.25	44.12
2010	203,963	2,262,225	3.00	11.09	57.04	42.64
2011	203,733	2,277,888	3.00	11.18	57.87	42.46
2012	202,480	2,280,598	3.00	11.26	57.63	41.36

Table 1: Belgian firms' database: summary statistics.

3 Method

Our complete database contains 202,480 firms from 1 employee to 8,198 employees in 2012. The complementary cumulative frequency distribution of this sample is represented in a log-log plot in Figure 1. This distribution is the realization in 2012 of a discrete random variable X, where X is the size of the firm measured by its number of employees. The probability of a firm to be of size x is

$$p(x) = P(X = x) \tag{1}$$

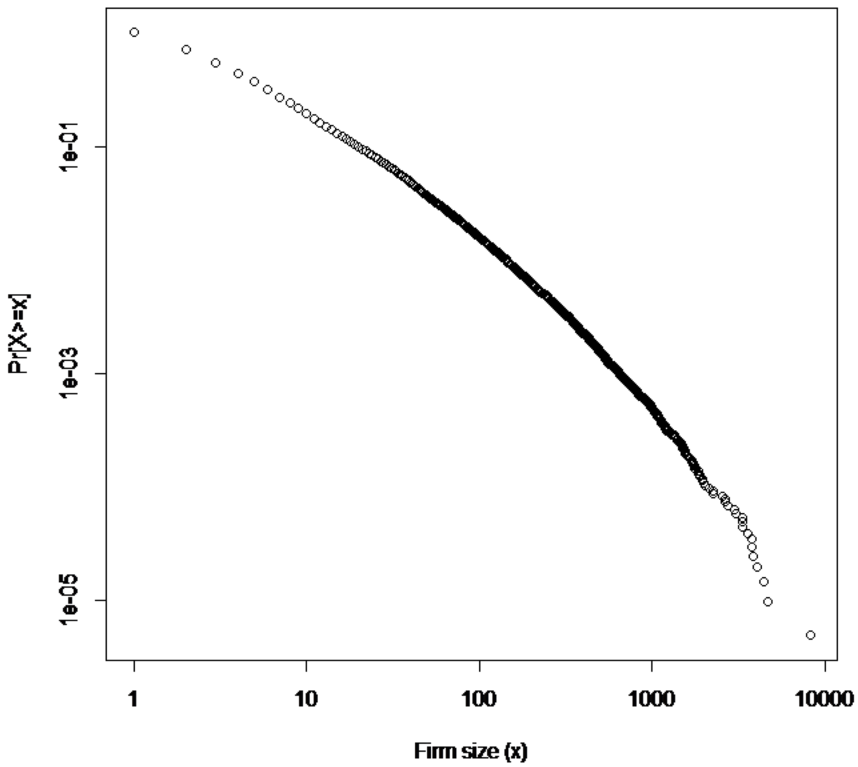


Figure 1: Complementary cumulative distribution of firm sizes in Belgium in 2012.

The scientific literature on firm size distribution has sought to fit the observed frequencies to the probabilities $p(x)$ of a parametric distribution function. Since Gibrat (1931), two objectives motivate this scientific research. The first is to use the parameters of the parametric distribution as an indicator of firm concentration and to measure its variability over time. The second objective is to use the shape of the parametric distribution as the long-run equilibrium of a stochastic model of firm dynamics to be identified. Historically, two parametric distributions have been proposed corresponding to different models of firm dynamics. First, the lognormal distribution, proposed by Gibrat (1931), is the asymptotic distribution of a stochastic model in which the growth rates of firms are independent and identically distributed across firms and over time. The CDF of the lognormal distribution is

$$F^L(x) = P(X \leq x) = \frac{1}{\sqrt{2\pi}} \int_0^x \exp\left[-\frac{1}{2}\left(\frac{\ln s - \mu}{\sigma\sqrt{2}}\right)^2\right] ds \qquad x > 0, \tag{2}$$

where μ is the mean and σ the standard deviation. Gibrat claimed that the lognormal distribution fitted well the entire distribution of French establishments in grouped data for the years 1896, 1901 and 1921. One may want to fit only part of the empirical distribution. In this case, the CDF of the truncated lognormal distribution is

$$F^L_{x_{min}}(x) = P(X \leq x) = \frac{F^L(x) - F^L(x_{min})}{1 - F^L(x_{min})} \qquad x \geq x_{min} > 0, \tag{3}$$

where x_{min} is the cutoff on the domain of the CDF.

The alternative distribution is the Pareto distribution proposed by Champernowne (1953) and Simon (1955) to fit the upper tail of the firm size distribution. Champernowne's model of firm dynamics is essentially identical[4] to Gibrat's model but assumes a mininum size in the range of firm sizes and finds that the upper tail of the resulting asymptotic distribution is Pareto.[5] Simon (1955) modifies the two preceding models by assuming that the number of firms increases over time thanks to a constant flow of entries of new small firms. The firm dynamics now is like a Yule process whose asymptotic upper tail is Pareto.

[4]Gibrat's model is a Markov chain and Champernowne's model is a Markov chain with a reflecting barrier.

[5]Champernowne (1953) proposed such a model to account for the distribution of incomes and, therefore, assumed a minimum income. However, his model can also be applied to other asymmetric empirical distributions such as the firm size distribution to fit their upper tail.

The CDF of the Pareto distribution is

$$F^P_{x_{min}}(x) = P(X \leq x) = 1 - \left(\frac{x_{min}}{x}\right)^{\alpha} \qquad x \geq x_{min} > 0, \tag{4}$$

where x_{min} is the cutoff on the domain of the CDF and $\alpha > 0$ is the scale parameter of the Pareto distribution function. Simon and Bonini (1958) find that data on assets of large US firms, i.e. the upper tail of the US FSD, collected by *Fortune* in 1955 fit well with a Pareto distribution. They also find that the Pareto distribution is a good approximation of the FSD in the US steel industry. More recently, in the first study with comprehensive data on the FSD, Axtell (2001) concludes that the entire empirical distribution of US firm sizes, grouped in logarithmic bins, is well approximated by the Zipf distribution, i.e. the Pareto distribution with α close to 1. Both papers estimate the scale parameter α of the upper tail or the entire distribution by simple linear regression. Simon and Bonini (1958) prefer not to conclude on the goodness-of-fit whereas Axtell (2001) uses the R-squared to evaluate the Zipf fit to the entire US distribution. Given the linearity of the Pareto pmf and CDF in log-log, it is tempting to want to use the linear regression model as a fitting method. Since Pareto, many researchers have adopted this method. Nevertheless, Aigner and Goldberger (1970) show that one must be careful because the sampling errors are heteroskedastic and non-independent especially when using the cumulative frequencies of the CDF. In their paper, they propose different efficient least squares estimators, which turn out to be neither simpler nor more efficient than the maximum likelihood estimator (MLE) for estimating the scale parameter α.

Our empirical strategy is based on the MLE for discrete data since our empirical distribution contains observations on the discrete number of employees per firm. Therefore, it is necessary to discretize the CDFs (3) and (4). This can be done by defining the discrete density of a firm size as

$$p(x) = P(X = x) = F(x+1) - F(x). \tag{5}$$

Aitchison and Brown (1957) and Seal (1952) provide the derivation of the MLE to estimate the parameters of the lognormal and Pareto discrete distributions respectively. Both log-likelihood functions can be solved numerically to obtain the ML estimators. To do so, we use the package for R-software proposed by Gillespie (2015, 2020) based on the MATLAB programme by Clauset et al. (2009). The goodness-of-fit test for each parametric distribution is based on the Kolmogorov–Smirnov (KS) distance with the empirical distribution and the p-value of the test is obtained by bootstrap.

If $x_{min} = 1$, we assess the fit of each parametric distribution to the entire empirical FSD. If we want to fit only the upper tail of the empirical FSD, we need to choose a value for x_{min}. Clauset et al. (2009) propose a way to select the value for x_{min} that provides the best fit for each parametric distribution. Their algorithm computes all possible pairs (parameters, x_{min}) and selects the pair that minimizes the KS for each parametric distribution. We believe that their selection process of the optimal lowest bound is questionable because the KS distances across pairs (parameters, x_{min}) are not statistically comparable since the support of the truncated distributions changes with the values of x_{min}. For this reason, we proceed differently. We choose some values for x_{min}, possibly including the 'optimal lowest bound', estimate the parameters by MLE and perform the goodness-of-fit test for each. Given the large size of our sample, the power of the goodness-of-fit test is very big and, hence, we expect the rejection of any candidate parametric distribution as a statistical significant fit to the empirical distribution. Therefore, there is no reason to repeat the whole process for all possible values of x_{min}. Moreover, we do not perform the test for high values of x_{min} as the number of observations and, hence, the power of the test decreases rapidly as x_{min} increases.

If one parametric distribution is rejected and the other is not, our study can conclude. But, in the most likely case where both parametric distributions are rejected, our investigation must continue and determine which of the two is a better model to summarize our empirical distribution. Following Clauset et al. (2009), we can use the log-likelihood ratio:

$$LLR = \sum_{i=1}^{n} [\ln p^P(x_i) - \ln p^L(x_i)], \tag{6}$$

where $p^P(x_i)$ and $p^L(x_i)$ are the probabilities determined respectively by a Pareto distribution and a lognormal distribution when $x_{min} = 1$. If the sign of (6) is positive, then the log-likelihood of the Pareto fit is higher, which means that the Pareto fit is better than the lognormal fit. If the sign is negative, the lognormal fit is better. To conclude that, under the null hypothesis, one of the two parametric distributions provides a better fit than the other, the LLR must be statistically different from zero. This depends on the sampling variance of the LLR. Vuong (1989) proposes the calculation of a *p*-value of the likelihood ratio test for non-nested models. If this *p*-value is sufficiently small, it can be concluded that the negative or positive value of the LLR is statistically different from zero. Otherwise, the two parametric fits cannot be statistically distinguished.

We can continue our investigation by focusing on the upper tail of the empirical distribution, i.e. when $x_{min} > 1$, and determine which of the parametric distributions offers the best fit. The log-likelihood ratio then becomes

$$LLR_{x_{min}} = \sum_{i: x_i \geq x_{min}} [\ln p^P_{x_{min}}(x_i) - \ln p^L_{x_{min}}(x_i)], \tag{7}$$

where the support of the distributions starts with the value of x_{min}. The statistical test of this $LLR_{x_{min}}$ can be performed as previously for a series of given values for x_{min} to conclude about the best candidate parametric distribution in the upper tail.

4 The Shape of the Belgian Aggregate Firm Size Distribution

This section presents the fitting estimation results of the Pareto and lognormal distributions for the entire size distribution of Belgian firms each year from 2006 to 2012. Following the method we detailed in Section 3, we first calculate the maximum likelihood estimates of the parameters of the entire lognormal and Pareto distributions for each year. We then compute the KS distance and the p-value for each parametric distribution. Table 2 reports, for each year and $x_{min} = 1$, the ML estimates of the parameters, the values of the KS and their p-values for the Pareto and lognormal distributions. As expected due to the power of the test, the fit of each of the two parametric distributions to the Belgian empirical distribution is statiscally rejected for all years. We then repeat the exercise for different values of $x_{min} > 1$ to test the goodness-of-fit of the two parametric distributions in the right tail of the FSD.[6] Unsurprinsingly, the statistical tests give the same results as for $x_{min} = 1$: both parametric distributions are rejected for all tested cases. As an example, Figure 2 shows the log-log plot of the complementary cumulative distribution of firm sizes in Belgium in 2006 and 2012 for $x_{min} = 10$. It can be checked visually that none of them fits perfectly the truncated empirical distribution (black line). However, the figure suggests that the lognormal distribution provides a better fit to the

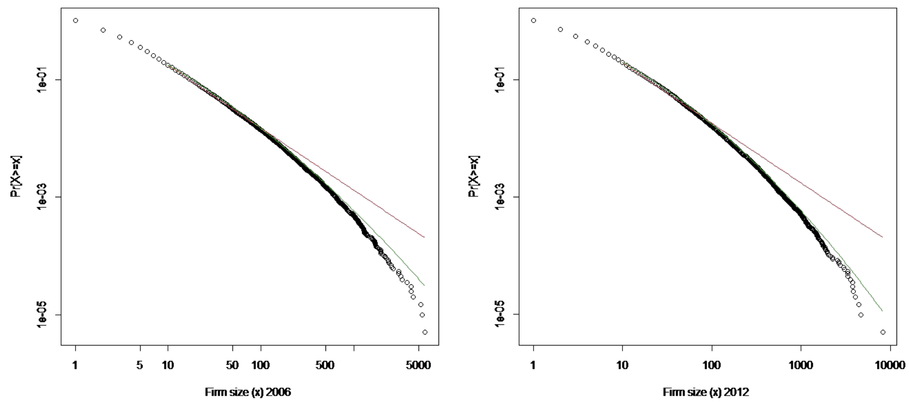
[6] The estimation results are available on request.

Belgian FSD than the Pareto distribution in 2006 and 2012. In order to confirm it, we perform the Vuong test. As explained in Section 3, a negative value for the Vuong statistic indicates that, for a given x_{min}, the fit of the lognormal is better than the fit of the Pareto. Table 3 shows, for 2012, the maximum likelihood estimates for the parameters of the Pareto and lognormal distributions and the Vuong statistic as x_{min} increases. The LLR ratio is negative and statistically significant at 10% every year for all values of x_{min} up to a firm size equal to 200 employees. From a size equal to 300, the LLR remains negative but is no longer statistically significant because the power of the test decreases with the number of firms. We repeated the exercise for all years in the database and found similar results.[7] Therefore, we can conclude that the lognormal distribution provides a better fit of the entire or truncated firm size distribution in Belgium from 2006 to 2012.

	Pareto distribution				Lognormal distribution				
Year	α	KS	p-value	R/F	μ	σ	KS	p-value	R/F
2006	1.58	0.017	0.00	R	0.46	1.69	0.004	0.000	R
2007	1.58	0.018	0.00	R	0.45	1.71	0.004	0.020	R
2008	1.57	0.021	0.00	R	0.47	1.73	0.005	0.000	R
2009	1.57	0.021	0.00	R	0.48	1.72	0.004	0.060	R
2010	1.57	0.022	0.00	R	0.47	1.74	0.004	0.090	R
2011	1.57	0.023	0.00	R	0.48	1.74	0.005	0.000	R
2012	1.57	0.021	0.00	R	0.49	1.75	0.004	0.000	R

Table 2: ML estimates of the parameters and goodness-of-fit test of the Pareto and lognormal distributions to the annual empirical aggregate FSD when $x_{min} = 1$. KS: Kolmogorov-Smirnov statistic; R/F: Reject/Fail to reject.

[7] The results are available on request.



(a) 2006

(b) 2012

Figure 2: Log-log plot of the complementary cumulative distribution of firm sizes in Belgium in 2006 and 2012: fit of the Pareto (red) and lognormal (green) distribution to the empirical distribution (black) for $x_{min} = 10$.

Aggregate FSD								
	Empirical distribution		P	LN		Vuong test		
x_{min}	Firms	Employees	α	μ	σ	LLR	p-value	Winner
1	202 480	2 280 598	1.57	0.48	1.75	-95.22	0.00	LN
5	74 470	2 034 564	2.22	0.31	1.99	-20.36	0.00	LN
15	26 416	1 656 133	2.22	0.31	1.99	-12.93	0.00	LN
25	16 126	1 463 591	2.22	0.31	1.99	-8.67	0.00	LN
50	7 336	1 158 748	2.43	0.29	1.99	-5.91	0.00	LN
100	3 211	874 089	2.43	2.40	1.63	-3.33	0.00	LN
200	1 270	608 851	2.75	2.44	1.62	-2.36	0.02	LN
300	714	474 166	2.75	2.12	1.67	-1.64	0.10	None
400	455	384 660	2.77	4.90	1.17	-1.29	0.19	None
500	315	321 947	2.80	4.90	1.17	-1.06	0.28	None
1000	99	176 873	3.17	5.21	1.12	-0.60	0.54	None
2500	17	62 804	4.71	7.33	0.59	-0.57	0.56	None

Table 3: Aggregate FSD in 2012. Number of firms and employees, ML estimates of the parameters of the Pareto (P) and lognormal (LN) distributions, and Vuong test results for the year 2012 as x_{min} increases.

5 The Shape of the Belgian Firm Size Distribution by Sectors

The objective of this section is to further our investigation at the sectoral level and identify possible distributional heterogeneity at this level of disaggregation. NACE sectors are very diverse, some are much more competitive than others, have higher average firm sizes than others, are more capital intensive than others, or employ more skilled personnel than others. This heterogeneity could suggest that the distributions of firm sizes might have different shapes across sectors.

We propose a simple sectoral disaggregation by distinguishing between manufacturing (NACE sector C) and service activities (NACE sectors G, H, I, J, K, M and N). The share of manufacturing in total employment is declining rapidly, as in many developed countries. The descriptive statistics in Appendix 2 show that it was 25 per cent in 2006 and is only 22 per cent six years later (Table 13). It can also be observed that the average and median sizes are significantly higher than in services (Table 12) and that the exit rate of industrial firms, close to that of services, is nevertheless slightly higher than the entry rate (Table 14), confirming the deindustrialisation trend in Belgium.

As in the aggregate case, we first test the fit of each of the two parametric distributions to the empirical distribution for all available years and for different values of x_{min}. At this level of disaggregation, the power of the test remains very high and all the tests we have performed conclude to reject the fit of the two parametric distributions whatever the value of x_{min}. Therefore, we perform the Vuong test to determine which parametric distribution provides the best fit to the empirical distribution. Tables 4 and 5 present the Vuong test results for manufacturing and services respectively in 2012. As with the aggregate distribution, the LLR is always negative regardless of the truncation of the distribution. It is statistically significant up to $x_{min} = 200$ for both sectors. These results at the sectoral level therefore confirm our results at the aggregate level, namely that the lognormal distribution offers a better fit to the distribution of firm sizes in Belgium from 2006 to 2012 whatever the truncation threshold.

Since services are a large sector, we push the disaggregation a bit further to ensure that the results we just found are not sensitive to the level of disaggregation chosen. To do this, we divide the services into two subsectors. The first sub-sector includes NACE sectors G, H and I, i.e. services that employ mainly low-skilled labor. The second sub-sector is composed of NACE sectors J,K, M and N, i.e. services that employ a lot of highly educated personnel. We apply the same method as above to both subsectors. Tables 6 and 7 provide the results of the Vuong statistical test for each of the two subsectors. These results show that the lognormal distribution provides a better fit than the Pareto distribution at this level of disaggregation. As an example, Figures 3 and 4 show the log-log plots of the complementary cumulative distribution of firm sizes in Belgium for different sectors in 2012 for $x_{min} = 10$.

We performed the Vuong test with different subsectors of the services, provided the number of observations is sufficiently large, for all the availbale years and obtained similar results.[8] We can conclude that the lognormal distribution also offers the best fit at the sectoral level of the Belgian FSD.

The objective of this section was to identify possible distributional heterogeneity at the sectoral level. Our estimations results show that, while parameter values may vary across sectors, the shape of the parametric distribution that best approximates the different sectoral distributions is the lognormal distri-

[8]Results not presented in this paper are available upon request.

bution. Therefore, these results rule out the possibility that the shape of the aggregate FSD is the result of an aggregation artifact.

FSD in manufacturing								
	Empirical distribution		P	LN		Vuong test		
x_{min}	Firms	Employees	α	μ	σ	LLR	p-value	Winner
1	18 139	507 269	1.42	1.37	1.88	-41.61	0.00	LN
5	9 790	489 904	1.69	0.85	2.09	-15.17	0.00	LN
15	4 971	450 327	1.86	1.39	1.93	-8.93	0.00	LN
25	3 472	421 948	1.97	0.65	2.11	-5.57	0.00	LN
50	1 805	363 707	2.04	2.92	1.58	-5.30	0.00	LN
100	969	304 965	2.23	2.91	1.59	-3.02	0.00	LN
200	447	232 302	2.43	2.95	1.58	-1.66	0.10	LN
300	264	187 722	2.55	3.09	1.56	-1.25	0.21	None
400	176	156 968	2.66	1.43	1.85	-0.77	0.44	None
500	128	135 510	2.78	-10.07	3.12	-0.25	0.80	None
1000	38	75 363	2.81	4.40	1.39	-0.46	0.64	None
2500	12	41 487	4.27	8.05	0.25	-0.99	0.32	None

Table 4: FSD in manufacturing in 2012. Number of firms and employees, ML estimates of the parameters of the Pareto (P) and lognormal (LN) distributions, and Vuong test results for the year 2012 as x_{min} increases.

FSD in services								
	Empirical distribution		P	LN		Vuong test		
x_{min}	Firms	Employees	α	μ	σ	LLR	p-value	Winner
1	144 759	1 491 550	1.57	0.62	1.64	-86.40	0.00	LN
5	53 240	1 312 744	1.98	-2.57	2.46	-16.31	0.00	LN
15	17 811	1 034 689	2.10	0.01	1.96	-10.66	0.00	LN
25	10 648	900 675	2.21	0.11	1.94	-7.31	0.00	LN
50	4 762	696 549	2.35	-0.97	2.15	-3.63	0.00	LN
100	1 935	502 652	2.48	-2.23	2.34	-1.82	0.07	LN
200	707	335 982	2.55	2.73	1.55	-1.96	0.05	LN
300	399	261 200	2.69	2.42	1.61	-1.23	0.22	None
400	252	210 626	2.75	3.57	1.40	-1.11	0.27	None
500	172	174 804	2.78	4.82	1.16	-1.11	0.27	None
1000	58	98 182	3.30	5.79	0.9	-0.54	0.59	None

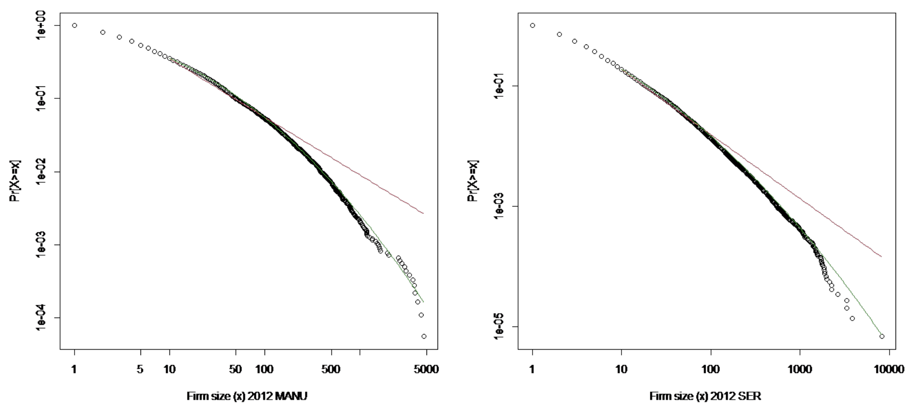
Table 5: FSD in services in 2012. Number of firms and employees, ML estimates of the parameters of the Pareto (P) and lognormal (LN) distributions, and Vuong test results for the year 2012 as x_{min} increases.

FSD in low-skilled services (NACE sectors G, H and I)								
	Empirical distribution		P	LN		Vuong test		
x_{min}	Firms	Employees	α	μ	σ	LLR	p-value	Winner
1	95 748	801 063	1.58	0.80	1.46	-80.89	0.00	LN
5	35 310	680 692	2.08	-1.36	2.05	-14.34	0.00	LN
15	10 552	486 951	2.29	-1.18	2.02	-5.95	0.00	LN
25	5 664	395 670	2.40	-2.19	2.19	-3.44	0.00	LN
50	2 138	274 433	2.51	-6.95	2.83	-1.31	0.19	None
100	756	180 979	2.57	-1.37	2.14	-1.28	0.20	None
200	258	113 636	2.63	3.58	1.31	-1.59	0.11	None
300	145	86 213	2.85	2.17	1.56	-0.81	0.42	None
400	87	66 407	2.89	4.26	1.18	-0.88	0.38	None
500	54	51 608	2.80	6.23	0.69	-1.41	0.16	None
1000	19	27 558	3.91	7.08	0.34	-1.11	0.27	None

Table 6: FSD in low-skilled services (NACE sectors G, H and I) in 2012. Number of firms and employees, ML estimates of the parameters of the Pareto (P) and lognormal (LN) distributions, and Vuong test results for the year 2012 as x_{min} increases.

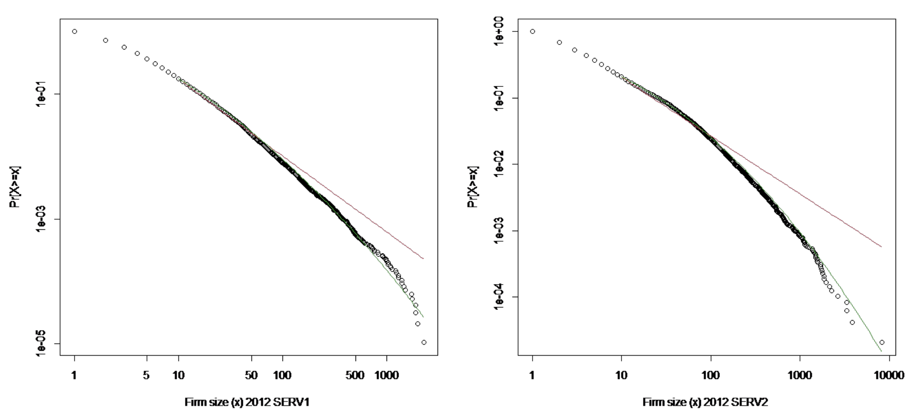
FSD in high-skilled services (NACE sectors J, K, M and N)								
	Empirical distribution		P	LN		Vuong test		
x_{min}	Firms	Employees	α	μ	σ	LLR	p-value	Winner
1	49 011	690 487	1.56	0.11	2.03	-40.91	0.00	LN
5	17 930	632 052	1.82	-1.74	2.56	-12.39	0.00	LN
15	7 259	547 738	1.91	2.03	1.63	-11.67	0.00	LN
25	4 984	505 005	2.05	2.06	1.62	-8.08	0.00	LN
50	2 624	422 116	2.24	1.25	1.81	-3.94	0.00	LN
100	1 179	321 673	2.43	-2.26	2.39	-1.42	0.15	None
200	449	222 346	2.50	2.30	1.67	-1.43	0.15	None
300	254	174 987	2.60	2.85	1.57	-1.07	0.29	None
400	165	144 219	2.68	3.43	1.47	-0.88	0.38	None
500	118	123 196	2.77	3.01	1.55	-0.62	0.54	None
1000	39	70 624	3.09	5.31	1.08	-0.46	0.65	None

Table 7: FSD in high-skilled services (NACE sectors J, K, M and N) in 2012. Number of firms and employees, ML estimates of the parameters of the Pareto (P) and lognormal (LN) distributions, and Vuong test results for the year 2012 as x_{min} increases.



(a) Manufacturing (b) Services

Figure 3: Log-log plot of the complementary cumulative distribution of firm sizes in Belgium in 2012: fit of the Pareto (red) and lognormal (green) distribution to the empirical distribution (black) for $x_{min} = 10$.



(a) Low-skilled services (b) High-skilled services

Figure 4: Log-log plot of the complementary cumulative distribution of firm sizes in Belgium in 2012: fit of the Pareto (red) and lognormal (green) distribution to the empirical distribution (black) for $x_{min} = 10$.

6 The Shape of the Belgian Firm Size Distribution by Regions

Our database includes geographic characteristics for each firm that allows us to extend our investigation at another level of disagregation: the regional level. Among this information, we have the address of each firm. To the best of our knowledge, this is the first FSD regional study that uses a comprehensive database. Belgium has three administrative regions with unequal populations: 58% of the national population lives in Flanders, 32% in Wallonia and 11% in Brussels, which is the capital city and a region. The Brussels region is an urban area while the other two regions include cities and rural areas. Flanders is more densely populated and richer than Wallonia. The descriptive statistics in Appendix 3 show differences and similarities between the regions. The median firm size (3 employees) is the same in all three regions while the average size is around 14 employees in Brussels, 12 employees in Flanders and 9 employees in Wallonia (Table 17). The entry and exit rates of firms are very similar in the three regions (Table 18), suggesting that it is the sector that is the relevant variable to explain the different entry and exit dynamics. Despite the differences in density and income levels between the three regions, the shapes of the regional FSD are similar.

The objective of this section is identical to that of the previous section. Given the specific demographic and economic characteristics of the three Belgian regions, the aim is to detect possible distributional heterogeneity at the regional level. Tables 8, 9 and 10 present the results of the Vuong statistical test for the Brussels region, Flanders, and Wallonia respectively in 2012.[9] Again, the LLR is almost always negative regardless of region and truncation level. It is negative

[9] We repeated the exercise for all the years available in the database and found similar results which are available on request.

and statistically significant up to the threshold of $x_{min} = 100$ and even beyond. Therefore, it can be concluded that the lognormal distribution provides the best approximation to the empirical regional FSDs. As an example, Figures 5 shows the log-log plots of the complementary cumulative distribution of firm sizes in Belgium for the three regions in 2012 for $x_{min} = 10$.

Our results lead to the same conclusion as before: the lognormal distribution offers a better approximation than the Pareto distribution to the empirical FSD of each region. These results offer further evidence that the shape of the aggregated FSD is not an artifact of aggregation.

FSD in Brussels								
	Empirical distribution		P	LN		Vuong test		
x_{min}	Firms	Employees	α	μ	σ	LLR	p-value	Winner
1	24 002	331 760	1.59	-0.23	2.09	-23.36	0.00	LN
5	8 115	302 634	1.86	-4.13	2.99	-5.70	0.00	LN
15	3 118	263 015	1.96	-3.98	2.97	-3.21	0.00	LN
25	1 946	240 915	2.01	-4.46	3.06	-2.32	0.02	LN
50	978	207 856	2.05	2.87	2.31	-2.46	0.01	LN
100	481	173 255	2.11	2.98	1.69	-2.53	0.01	LN
200	247	141 001	2.29	3.62	1.53	-1.78	0.08	LN
300	157	119 137	2.41	3.79	1.49	-1.25	0.21	None
400	109	103 003	2.47	5.22	1.14	-1.44	0.15	None
500	82	90 961	2.55	5.70	1.01	-1.28	0.20	None
1000	34	58 850	3.07	7.08	0.52	-1.43	0.15	None

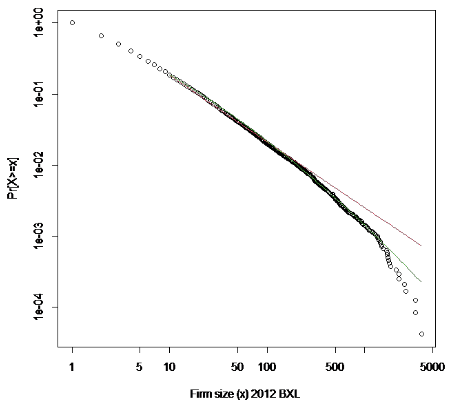
Table 8: FSD in Brussels in 2012. Number of firms and employees, ML estimates of the parameters of the Pareto (P) and lognormal (LN) distributions, and Vuong test results for the year 2012 as x_{min} increases.

FSD in Flanders								
	Empirical distribution		P	LN		Vuong test		
x_{min}	Firms	Employees	α	μ	σ	LLR	p-value	Winner
1	120 760	1 407 225	1.56	0.59	1.73	-78.15	0.00	LN
5	46 280	1 262 917	1.92	-1.85	2.39	-17.68	0.00	LN
15	16 688	1 029 070	2.06	0.37	1.93	-11.45	0.00	LN
25	10 282	909 286	2.17	0.42	1.92	-7.90	0.00	LN
50	4 745	717 080	2.29	0.66	1.89	-4.67	0.00	LN
100	2 034	530 590	2.46	0.32	1.95	-2.41	0.02	LN
200	773	358 658	2.60	2.10	1.64	-1.63	0.10	LN
300	419	272 741	2.71	2.71	1.54	-1.2	0.23	None
400	257	216 186	2.75	3.76	1.36	-1.09	0.27	None
500	183	182 834	2.89	1.25	1.77	-0.56	0.57	None
1000	52	94 340	3.19	7.15	1.03	0.18	0.85	None
2500	9	36 506	3.37	7.58	0.63	-0.48	0.63	None

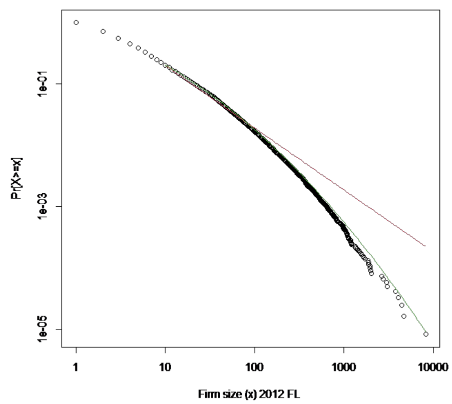
Table 9: FSD in Flanders in 2012. Number of firms and employees, ML estimates of the parameters of the Pareto (P) and lognormal (LN) distributions, and Vuong test results for the year 2012 as x_{min} increases.

FSD in Wallonia								
	Empirical distribution		P	LN		Vuong test		
x_{min}	Firms	Employees	α	μ	σ	LLR	p-value	Winner
1	57 718	541 613	1.59	0.5	1.65	-51.86	0.00	LN
5	20 075	469 013	1.99	-2.69	2.46	-9.84	0.00	LN
15	6 610	364 048	2.14	-0.18	1.97	-6.34	0.00	LN
25	3 898	313 390	2.25	-0.45	2.02	-4.14	0.00	LN
50	1 613	233 812	2.31	2.02	1.57	-3.50	0.00	LN
100	696	170 244	2.53	1.50	1.68	-1.64	0.10	LN
200	250	109 192	2.69	2.97	1.42	-1.04	0.30	None
300	138	82 288	2.91	0.77	1.76	-0.49	0.63	None
400	89	65 471	3.22	2.54	1.68	0.55	0.58	None
500	50	48 152	3.01	3.15	1.53	0.03	0.97	None
1000	13	23 683	3.05	3.74	1.41	-0.28	0.78	None

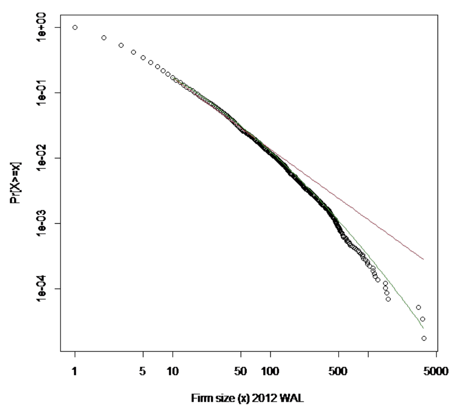
Table 10: FSD in Wallonia in 2012. Number of firms and employees, ML estimates of the parameters of the Pareto (P) and lognormal (LN) distributions, and Vuong test results for the year 2012 as x_{min} increases.



(a) Brussels



(b) Flanders



(c) Wallonia

Figure 5: Log-log plot of the complementary cumulative distribution of firm sizes in Belgium in 2012: fit of the Pareto (red) and lognormal (green) distribution to the empirical distribution (black) for $x_{min} = 10$.

7 Discussion

The statistical method we propose in this paper allows us to classify possible parametric distributions according to their goodness-of-fit to the empirical distribution under study. This method can be used to compare the fits of parametric distributions to many observed economic empirical regularities such as the distribution of city size, income and wealth distributions, the distribution of stock market returns or the firm size distribution as we do in this paper.

The approximation of the empirical distribution of firm size by a parametric distribution allows the researcher:

- to have empirical information on the shape of the FSD, which is essential for the construction of an explanatory model of firm dynamics. Our results show that the shape of the FSD in Belgium is persistent across sectors and regions, even though the degree of firm concentration may vary across sectors and regions. This suggests that the underlying explanatory model of firm dynamics transcends large sectoral or regional particularities.

- to use the parameters of the parametric distribution to evaluate the degree of firm concentration and its evolution over time. It is an alternative or complementary tool to non-parametric indicators.

- to make a statistically informed choice about the functional form for the empirical FSD used in models. For example, a quantitative macroeconomics model with heterogeneous firms must specify the functional form of the distribution of firms (see references in footnote 1). If the model assumes that the functional form is Pareto in the case of Belgium, then the weight of the large firms is greater than that of the empirical distribution, which may substantially bias the model's predictions on aggregate outcomes (income growth, international trade volumes, corporate tax revenues, R&D expenditures, ...) and on the effects of public policies.

It is important to stress that no parametric distribution will ever perfectly fit the empirical distribution. This loss of statistical information, which can be minimized by the choice of the parametric distribution from the method we propose, is the price to pay for obtaining a functional form that is simple to model. This imperfection of the parametric distribution creates an uncertainty on the "true" functional form of the empirical distribution to be retained. Faced with this uncertainty, the approach of Ijiri and Simon (1974) seems reasonable. They consider that the choice of the functional form should be guided by the simplicity and plausibility of the economic assumptions of the underlying firm dynamics model. If the statistical method, which should be robust, shows that there are persistent deviations over time from the empirical distribution, then the model should be able to propose an explanatory mechanism for these deviations. Therefore, the scientific debate should be as much about the fit of the competing parametric distributions as about their underlying explanatory models.

8 Conclusion

In this paper, we propose a reliable statistical method to determine which of two parametric distributions offers the best fit to the empirical distribution. We apply this method to a complete database on the distribution of Belgian firm sizes for the period 2006-2012. Our results show that the lognormal distribution always prevails over the Pareto distribution whatever the level of aggregation.

These unambiguous results argue for considering the lognormal function as the most statistically motivated hypothesis of the functional form of the firm size distribution in a firm dynamics model, at least in Belgium. It remains to be investigated whether this hypothesis is confirmed in other countries.

In order to complete a model of firm dynamics, there is still one distribution to characterise: the conditional distribution of the growth rate of the firm given its size. Indeed, the size of a firm is the outcome of the history of its growth rates. Gibrat (1931) hypothesized that this history is a random walk and that the growth rate of the firm and its size are independent (Gibrat's law). Due to the lack of comprehensive panel data on firm sizes, Gibrat's law still lacks empirical evidence. The increasing availability of such data in the future should make it possible to produce them and thus provide a long overdue robust empirical basis for models of firm dynamics.

References

Aigner, D. J. and A. S. Goldberger (1970). Estimation of Pareto's law from grouped observations. *Journal of the American Statistical Association 65*(330), 712–723.

Aitchison, J. and J. A. Brown (1957). The lognormal distribution with special reference to its uses in economics.

Antras, P. and E. Helpman (2004). Global sourcing. *Journal of Political Economy 112*(3), 552–580.

Arkolakis, C., S. Demidova, P. J. Klenow, and A. Rodríguez-Clare (2008). Endogenous variety and the gains from trade. *American Economic Review 98*(2), 444–50.

Axtell, R. L. (2001). Zipf distribution of U.S. firm sizes. *Science 293*(5536), 1818–1820.

Bee, M., M. Riccaboni, and S. Schiavo (2017). Where Gibrat meets Zipf: Scale and scope of French firms. *Physica A: Statistical Mechanics and its Applications 481*, 265–275.

Champernowne, D. G. (1953). A model of income distribution. *Economic Journal 63*(250), 318–351.

Chaney, T. (2008). Distorted gravity: The intensive and extensive margins of international trade. *American Economic Review 98*(4), 1707–21.

Clauset, A., C. R. Shalizi, and M. E. Newman (2009). Power-law distributions in empirical data. *SIAM Review 51*(4), 661–703.

Eaton, J., S. Kortum, and F. Kramarz (2011). An anatomy of international trade: Evidence from French firms. *Econometrica 79*(5), 1453–1498.

Gabaix, X. and A. Landier (2008). Why has CEO pay increased so much? *Quarterly Journal of Economics 123*(1), 49–100.

Gibrat, R. (1931). *Les inégalités économiques*. Sirey.

Gillespie, C. S. (2015). Fitting heavy tailed distributions: The powerlaw package. *Journal of Statistical Software 64*(2).

Gillespie, C. S. (2020). Powerlaw package. Technical report, P27.

Goldstein, M. L., S. A. Morris, and G. G. Yen (2004). Problems with fitting to the power-law distribution. *The European Physical Journal B-Condensed Matter and Complex Systems 41*(2), 255–258.

Helpman, E., M. Melitz, and Y. Rubinstein (2008). Estimating trade flows: Trading partners and trading volumes. *Quarterly Journal of Economics 123*(2), 441–487.

Helpman, E., M. J. Melitz, and S. R. Yeaple (2004). Export versus FDI with heterogeneous firms. *American Economic Review 94*(1), 300–316.

Ijiri, Y. and H. A. Simon (1974). Interpretations of departures from the Pareto curve firm-size distributions. *Journal of Political Economy 82*(2, Part 1), 315–331.

Kleiber, C. and S. Kotz (2003). *Statistical Size Distributions in Economics and Actuarial Sciences*, Volume 470. John Wiley & Sons.

Kondo, I. O., L. T. Lewis, and A. Stella (2020). Heavy tailed, but not Zipf: Firm and establishment size in the US. *http://ltlewis.net/KLS.pdf*.

Luttmer, E. G. (2007). Selection, growth, and the size distribution of firms. *Quarterly Journal of Economics 122*(3), 1103–1144.

Melitz, M. J. and G. I. Ottaviano (2008). Market size, trade, and productivity. *Review of Economic Studies 75*(1), 295–316.

Melitz, M. J. and S. J. Redding (2015). New trade models, new welfare implications. *American Economic Review 105*(3), 1105–46.

Montebruno, P., R. J. Bennett, C. Van Lieshout, and H. Smith (2019). A tale of two tails: Do power law and lognormal models fit firm-size distributions in the mid-Victorian era? *Physica A: Statistical Mechanics and its Applications 523*, 858–875.

Perline, R. (2005). Strong, weak and false inverse power laws. *Statistical Science*, 68–88.

Rossi-Hansberg, E. and M. L. Wright (2007). Establishment size dynamics in the aggregate economy. *American Economic Review 97*(5), 1639–1666.

Seal, H. L. (1952). The maximum likelihood fitting of the discrete Pareto law. *Journal of the Institute of Actuaries 78*(1), 115–121.

Simon, H. A. (1955). On a class of skew distribution functions. *Biometrika 42*(3/4), 425–440.

Simon, H. A. and C. P. Bonini (1958). The size distribution of business firms. *American Economic Review 48*(4), 607–617.

Vuong, Q. H. (1989). Likelihood ratio tests for model selection and non-nested hypotheses. *Econometrica* (57), 307–333.

Zipf, G. K. (1949). *Human behavior and the principle of least effort: An introduction to human ecology.* Addison-Wesley Press.

A Appendix 1

Code	Economic Area
A	Agriculture, Forestry and Fishing
B	Mining and Quarrying
C	Manufacturing
D	Electricity, Gas, Steam and Air Conditioning Supply
E	Water Supply; Sewerage, Waste Management and Remediation Activities
F	Construction
G	Wholesale and Retail Trade; Repair of Motor Vehicles and Motorcycles
H	Transportation and Storage
I	Accommodation and Food Service Activities
J	Information and Communication
K	Financial and Insurance Activities
L	Real Estate Activities
M	Professional, Scientific and Technical Activities
N	Administrative and Support Service Activities
O	Public Administration and Defence; Compulsory Social Security
P	Education
Q	Human Health and Social Work Activities
R	Arts, Entertainment and Recreation
S	Other Service Activities
T	Activities of Households as Employers; Undifferentiate Goods and Services Producing Activities of Households for Own Use
U	Activities of Extraterritorial Organisations and Bodies

Table 11: Statistical classification of economic activities in the European Community Rev. 2 (2008): Level 1 codes.

B Appendix 2

	Manufacturing				**Services**			
Year	Firms	Employees	Median	Mean	Firms	Employees	Median	Mean
2006	20,094	526,258	5.00	26.19	148,073	1, 312,583	3.00	8.86
2007	19,306	499,347	5.00	25.86	144,623	1,327,698	3.00	9.18
2008	19,887	553,291	5.00	27.82	145,600	1,432,516	3.00	9.84
2009	19,460	523,367	5.00	26.89	144,993	1,430,032	3.00	9.86
2010	19,192	516,654	5.00	26.92	144,878	1,462,325	3.00	10.09
2011	18,822	517,531	5.00	27.50	144,750	1,473,427	3.00	10.18
2012	18,139	507,269	5.00	27.96	144,759	1,491,550	3.00	10.30

Table 12: Summary statistics by sector: Manufacturing (NACE sector C) and Services (NACE sectors G, H, I, J, K, M and N).

	Manufacturing		Services	
Year	Employment (%)	Firms (%)	Employment (%)	Firms (%)
2006	25.3	9.9	63.1	73.4
2007	24.1	9.7	64.2	73.1
2008	24.4	9.7	63.3	71.2
2009	23.5	9.5	64.1	71.3
2010	22.8	9.4	64.6	71
2011	22.7	9.2	64.7	71
2012	22.2	8.9	65.4	71.5

Table 13: Manufacturing and Services: share of employment and firms in national totals.

NACE sector		2007	2008	2009	2010	2011	2012
C	Entry rate	5.04	8.14	3.73	4.40	4.11	3.32
	Exit rate	5.83	7.00	5.16	5.21	5.37	5.92
G H I	Entry rate	6.84	8.04	5.74	6.33	6.39	5.57
	Exit rate	6.61	6.97	6.40	6.35	6.46	6.80
J K M N	Entry rate	6.81	7.65	5.33	6.47	6.21	5.83
	Exit rate	5.33	10.88	5.07	5.23	5.55	5.52
G H I J K M N	Entry rate	6.77	7.72	5.63	6.37	6.31	5.68
	Exit rate	6.11	8.13	5.97	5.96	6.13	6.38
All	Entry rate	6.29	6.78	7.62	5.78	6.61	6.48
	Exit rate	6.34	6.00	6.26	6.19	6.11	6.37

Table 14: Manufacturing and Services: exit and entry rates (%).

C Appendix 3

	Brussels		**Flanders**		**Wallonia**	
Year	Firms	Employees	Firms	Employees	Firms	Employees
2006	24,273	315,183	121,980	1,275,689	55,424	489,698
2007	23,625	315,671	119,111	1,266,477	54,995	487,189
2008	24,241	336,380	123,256	1,400,871	57,066	527,432
2009	23,912	337,252	122,434	1,373,686	57,078	519,171
2010	23,894	337,295	122,202	1,391,019	57,867	533,911
2011	24,110	340,023	121,415	1,395,037	58,208	542,828
2012	24,002	331,760	120,760	1,407,225	57,718	541,613

Table 15: Summary statistics by region.

	Brussels		**Flanders**		**Wallonia**	
Year	Firms	Employees	Firms	Employees	Firms	Employees
2006	12.04%	15.15%	60.48%	61.31%	27.48%	23.54%
2007	11.95%	15.25%	60.24%	61.20%	27.81%	23.54%
2008	11.85%	14.85%	60.25%	61.86%	27.90%	23.29%
2009	11.75%	15.12%	60.19%	61.60%	28.06%	23.28%
2010	11.71%	14.91%	59.91%	61.49%	28.37%	23.60%
2011	11.83%	14.93%	59.60%	61.24%	28.57%	23.83%

Table 16: Firms and employees by regions as a percentage of the national totals.

	Brussels		Flanders		Wallonia	
Year	Median	Mean	Median	Mean	Median	Mean
2006	3.00	12.98	3.00	10.46	3.00	8.84
2007	2.00	13.36	3.00	10.63	3.00	8.86
2008	3.00	13.88	3.00	11.36	3.00	9.24
2009	3.00	14.10	3.00	11.21	3.00	9.09
2010	3.00	14.12	3.00	11.38	3.00	9.22
2011	3.00	14.10	3.00	11.50	3.00	9.32
2012	3.00	13.82	3.00	11.65	3.00	9.38

Table 17: Median and mean firm size by region.

Region		**2007**	**2008**	**2009**	**2010**	**2011**	**2012**
Brussels	Entry rate	7.41	8.20	6.88	7.34	7.51	7.39
	Exit rate	6.92	7.11	7.57	7.20	6.92	7.84
Flanders	Entry rate	6.55	7.49	5.54	6.09	5.82	5.31
	Exit rate	5.90	6.07	6.01	5.94	6.05	6.24
Wallonia	Entry rate	7.14	7.88	6.39	7.55	7.13	6.36
	Exit rate	6.04	6.42	6.48	6.25	6.70	7.40
All	Entry rate	6.29	6.78	7.62	5.78	6.61	6.48
	Exit rate	6.34	6.00	6.26	6.19	6.11	6.37

Table 18: Regions: exit and entry rates (%).

D Appendix 4

Abbreviation	Meaning
CDF	Cumulative distribution function
FSD	Firm size distribution
KS	Kolmogorov-Smirnov
LLR	Log-likelihood ratio
ML(E)	Maximum likelihood (estimator)
NACE	Nomenclature statistique des Activités économiques dans la Communauté Européenne (Statistical Classification of Economic Activites in the European Community)

Table 19: List of abbreviations.