UNIVERSIDAD EAFIT
Abierta al mundo

Escuela de Economía y Finanzas

Documentos de trabajo

Economía y Finanzas

Centro de Investigación Económicas y Financieras

No. 17-01
2017

Measuring Firm Size Distribution With Seminonparametric Densities

Lina Cortés, Andrés Mora-Valencia, Javier Perote

cief
Centro de Investigaciones Económicas y Financieras
Universidad EAFIT

Measuring firm size distribution with semi-nonparametric densities

Lina M. Cortés
Professor Department of Finance, School of Economics and Finance. Universidad EAFIT. Address: Carrera 49 No 7 Sur-50 Medellin, Colombia.
E-mail: lcortesd@eafit.edu.co
Corresponding author. Phone: +5742619500 ext. 9756. Fax: +5743120649. E-mail address: lcortesd@eafit.edu.co.

Andrés Mora-Valencia
Professor School of Management, Universidad de los Andes. Address: Calle 21 No. 1-20 Bogota, Colombia.
E-mail: a.mora262@uniandes.edu.co

Javier Perote
Professor Department of Economics and IME, University of Salamanca. Address: Campus Miguel de Unamuno 37007 Salamanca, Spain.
E-mail: perote@usal.es

Abstract

In this article, we propose a new methodology based on a (log) semi-nonparametric (log-SNP) distribution that nests the lognormal and enables better fits in the upper tail of the distribution through the introduction of new parameters. We test the performance of the lognormal and log-SNP distributions capturing firm size, measured through a sample of US firms in 2004-2015. Taking different levels of aggregation by type of economic activity, our study shows that the log-SNP provides a better fit of the firm size distribution. We also formally introduce the multivariate log-SNP distribution, which encompasses the multivariate lognormal, to analyze the estimation of the joint distribution of the value of the firm's assets and sales. The results suggest that sales are a better firm size measure, as indicated by other studies in the literature.

Keywords: Firms size distribution; Heavy tail distributions; Semi-nonparametric modeling; Bivariate distributions.

JEL Codes: C14, C53, L11.

1. Introduction

Studies of firm size distribution have raised great interest among researchers in the fields of physics and economics [1] [2] [3] [4]. This topic is relevant because knowledge about the shape of the firm size distribution can provide researchers and policymakers with information about the levels of industrial concentration and economic cycles that is useful in implementing competition policies [5] [6] [7].

In a pioneering study on firm size, Gibrat [8] found that firm size could be described by the lognormal distribution. Since then, several studies have supported the use of this distribution [1] [9]. However, different types of distributions have been proposed. Some empirical studies have argued that the size distribution can be adjusted according to a Pareto or power-law distribution [10] [11] or that it can be well estimated based on Zipf's law [12].

A strand of the empirical literature has thus sought to examine the application of lognormal and Pareto or power-law distributions using firm size data as cross-sectional data [13] [14] [15]. However, there is evidence that, on some occasions, a poor approximation of the empirical distributions of the firm size in the upper tail, which typically exhibit greater asymmetry as a small number of large firms exist alongside a large number of smaller firms [1] [12] [16], is obtained.

For example, in their article, Stanley et al. [1] find that the size distribution for a series of firms listed in the US stock market has a good fit with the lognormal distribution, with the exception of the upper tail. In this case, the lognormal distribution overestimates the size of the large firms. On the other hand, Goddard et al. [15] examine firm size among banks and credit unions based on Zipf's law. Their study rejects Zipf's law as a descriptor of the firm size distribution in the upper tail.

The differences obtained in applying these types of firm size distributions have led researchers in this area to discuss the stability of a single firm size probability model over time and across industries and countries [5] [17] [13] [14] [18]. These discrepancies likely occur because the distributions that are traditionally used to model data with very thick tails have the disadvantage of relying on very few parameters for capturing the entire shape of the

firm size distribution, including its right tail [18]. In this regard, Newman [19] and Martínez-Mekler et al. [20] state that few processes in the real world follow the Pareto or power-law distribution across their entire range and, in particular, these types of distributions do not fit the smaller values of the variable being measured.

Meanwhile, the common point of departure under the hypothesis of Zipf's law is to assume that the firm size distribution is well described by a Pareto or power-law distribution above a certain minimum threshold [21] [15] [16] [22]. In this manner, if we seek to study the growth of smaller firms compared to that of larger firms, then we cannot use a Pareto or power-law distribution because the small firms are found in the upper tail, below the threshold value [2] [23].

With the objective of modeling the firm size distribution, we propose to use semi-nonparametric approximations (SNP) based on Edgeworth and Gram-Charlier expansions. These distributions have been applied in very diverse fields in which precision in measuring distribution tails is important for correctly measuring the occurrence of extreme values (for examples of applications in thermodynamics, astronomy, finance and scientometrics, see Kuhs [24], Blinnikov and Moessner [25], Mauleon and Perote [26] and Cortés et al. [27], respectively).

In this article, for the first time, we propose to use these distributions to model the firm size distribution, and in particular, we propose logarithmic transformations of an SNP distribution (log-SNP), which are extensions of a lognormal distribution that enable an approximation of any empirical distribution through the introduction of additional parameters. With this transformation, we seek to maintain the parameter flexibility of the Gram-Charlier distributions while restricting the domain of positive values. We find that in comparison to the lognormal distribution, the log-SNP distribution provides a better fit in modeling the firm size distribution using different levels of industrial aggregation. We also show that the log-SNP distribution allows us to obtain a better fit in the upper quantiles without having to impose a minimum threshold. This aspect is important because understanding the behavior of the largest firms and those with the greatest weight in the market is essential for analyzing the economy as a whole [23]. Additionally, we extend the log-SNP to the multivariate context by providing an expression for the bivariate log-SNP

distribution, whose marginal densities act as univariate distributions of the log-SNP. The advantage of developing a multivariate framework is based on the fact that more efficient estimations are obtained, making it possible to jointly analyze the behavior of variables that are highly correlated and testing differences in marginal specifications through traditional linear restrictions tests – likelihood ratio (LR), Wald or Lagrange multipliers (LM).

This paper is structured as follows. Section 2 provides the definitions and main characteristics of both univariate and bivariate log-SNP distributions. Section 3 compares the performance of these distributions to their nested lognormal counterparts for studying the size of a sample of US firms in different industries and analyzes 'sales' and 'assets' for measuring firm size. The final conclusions are summarized in the last section.

2. Log-SNP distribution

This section defines the log-SNP probability density function (PDF) and provides a straightforward extension of this distribution to the multivariate case. Because this distribution is a logarithmic transformation of the so-called Gram-Charlier (or SNP) distribution, we begin by defining this class of densities and reviewing some of its main properties.

Definition 2.1: *The Gram-Charlier density of a random variable* x_i *is a general class of densities of the type*

$$f(x_i; \boldsymbol{d_i}) = \phi(x_i) \sum_{s=0}^{n} d_{is} H_s(x_i) = \phi(x_i) p_i(x_i),\ x_i \in \mathbb{R}, \tag{1}$$

where $\phi(x_i) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}{x_i}^2}$ *is the standard normal PDF,* $H_s(x) = \frac{(-1)^s}{\phi(x_i)} \frac{d^s \phi(x_i)}{d{x_i}^s}$ *is the* s^{th} *order Hermite polynomial*[1]*,* $\boldsymbol{d_i} = (d_{i1}, \dots, d_{iS})' \in \mathbb{R}^n$ *and n is the (even) order of the expansion.*

[1] The first Hermite polynomials are $H_0(x_i) = 1$, $H_1(x_i) = x_i$, $H_2(x_i) = x_i^2 - 1$, $H_3(x) = x_i^3 - 3x_i$, $H_4(x_i) = x_i^4 - 6x_i^2 + 3$.

The condition $d_{i0} = 1$ is sufficient to guarantee that function (1) integrates one, but non-negativity is not guaranteed for all $\boldsymbol{d_i} \in \mathbb{R}^n$; thus, positivity constrains must be considered to ensure a well-defined family of densities.[2]

This density nests the standard normal (when $\boldsymbol{d_i} = \mathbf{0}$) and presents two main advantages with respect to other PDFs used to fit empirical distributions: *(i)* It incorporates enough degrees of freedom to capture any moment of the distribution with a flexible parametric structure; and *(ii)* the asymptotic expansion (as $n \to \infty$) captures the true distribution – see Jarrow and Rudd [28]. Moreover, despite its apparent complexity, the Gram-Charlier PDF is very tractable due to the orthogonality of the Hermite polynomials, which satisfy, among other properties, the following:

$$\int_{-\infty}^{\infty} H_s(x) H_j(x) \phi(x) dx = \begin{cases} 0, & s \neq j \\ s!, & s = j \end{cases}. \tag{2}$$

For instance, the cumulative distribution function (CDF) and the moment generating function (MGF) can be computed, respectively, as follows:

$$\int_{-\infty}^{a} f(x_i, \boldsymbol{d_i}) dx_i = \int_{-\infty}^{a} \phi(x_i) dx_i - \phi(a) \sum_{s=1}^{n} d_{is} H_{s-1}(a) \tag{3}$$

and

$$\int_{-\infty}^{\infty} e^{t x_i} f(x_i, \boldsymbol{d_i}) dx_i = e^{t^2/2} \sum_{s=0}^{n} d_{is} t^s. \tag{4}$$

Therefore, it can be easily checked that the even (odd) moment of order *k* depends on the *k* first even (odd) parameters. For instance, if $d_1 = 0$, then the density has a zero mean, and d_3 accounts for asymmetry; and if $d_2 = 0$, then the density has unit variance, and d_4 captures excess kurtosis.

Definition 2.2: *Let z_i be log-SNP distributed with location parameter $\mu_i \in \mathbb{R}$, scale $\sigma_i{}^2 \in \mathbb{R}^+$ and shape parameters $\boldsymbol{d_i} = (d_{i1}, \dots, d_{iS})' \in \mathbb{R}^n$. Then, its PDF can be expressed as follows:*

[2] For a description of the positivity region in terms of skewness and kurtosis, see Jondeau and Rockinger [39]. Alternatively, Gram-Charlier can be defined as $f^*(x_i; \boldsymbol{d}) = \phi(x_i) p_i(x_i)^2$, though at the cost of an increasing complexity. However, the positive formulation can be represented in terms of a larger expansion of the type defined in (1) – see León et al. [40] – and maximum likelihood estimation algorithms necessarily converge to values that guarantee a well-defined PDF.

$$h(z_i;\ \mu_i, {\sigma_i}^2, \boldsymbol{d_i}) = \left[1 + \sum_{s=1}^{n} d_{is} H_s\left(\frac{\ln(z_i)-\mu_i}{\sigma_i}\right)\right]\left(\frac{1}{z_i\sigma_i\sqrt{2\pi}} e^{-\frac{(\ln(z_i)-\mu_i)^2}{2{\sigma_i}^2}}\right), \quad z_i \in \mathbb{R}^+. \tag{5}$$

Consequently, the log-SNP is the exponential transformation of a Gram-Charlier distributed variable, i.e., $z_i = \exp(x_i)$, where x_i is distributed according to the PDF (1), which has also been "location-scale" transformed so that the lognormal distribution is a particular case (for $\boldsymbol{d_i} = \mathbf{0}$). The resulting density presents the same parameter flexibility as the Gram-Charlier but is defined only on the positive real axis. The properties of this distribution can be easily obtained from those of the Gram-Charlier – for further details, see Ñíguez et al. [29] and [30]. In particular, central moments can be obtained directly from the MGF of the Gram-Charlier distribution – equation (4) – as follows:

$$E[{z_i}^t] = e^{\mu_i t + \frac{1}{2}t^2{\sigma_i}^2}[1 + \sum_{s=1}^{n} d_{is}(\sigma_i t)^s]. \tag{6}$$

Both the Gram-Charlier and the log-SNP can be extended to the multivariate case in different ways. This paper defines the multivariate log-SNP PDF in terms of the so-called multivariate Edgeworth-Sargan density defined by Perote [31]. In what follows and without loss of generality, we describe the bivariate case, which is applied in next section.

Definition 2.3: *Let* $\boldsymbol{Z} = (z_1, z_2)' \in \mathbb{R}^{2+}$ *be a random vector distributed in the bivariate log-SNP with the mean* $\boldsymbol{\mu} = (\mu_1, \mu_2)' \in \mathbb{R}^2$ *and the (positive definite) variance matrix* $\boldsymbol{\Sigma} = \begin{pmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{pmatrix}$, $\sigma_i > 0$, i=1,2, *and* $|\rho| < 1$. *Thus, the joint density of* $\boldsymbol{Z}$ *is described by*

$$H(\boldsymbol{Z};\ \boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{D}) = F(\boldsymbol{Z}) + \frac{1}{\sigma_1\sigma_2}\phi\left(\frac{\ln(z_1)-\mu_1}{\sigma_1}\right)\phi\left(\frac{\ln(z_2)-\mu_2}{\sigma_2}\right)\left[p_1\left(\frac{\ln(z_1)-\mu_1}{\sigma_1}\right) + p_2\left(\frac{\ln(z_2)-\mu_2}{\sigma_2}\right)\right],$$

$$z_i \in \mathbb{R}^+, i = 1,2,\ \boldsymbol{D} = (\boldsymbol{d_1} \quad \boldsymbol{d_2})' \in \mathbb{R}^{2n}, \tag{7}$$

where $F(\boldsymbol{Z})$ *is a multivariate lognormal distribution – Aitchinson and Brown* [32] –

$$F(\boldsymbol{Z};\ \boldsymbol{\mu}, \boldsymbol{\Sigma}) = \frac{1}{2\pi x_1 x_2 \sigma_1\sigma_2\sqrt{1-\rho}} exp\left(-\frac{q}{2}\right), \tag{8}$$

$$q = \frac{1}{1-\rho^2}\left[\left(\frac{\ln(z_1)-\mu_1}{\sigma_1}\right)^2 + \left(\frac{\ln(z_2)-\mu_2}{\sigma_2}\right)^2 - 2\rho\left(\frac{\ln(z_1)-\mu_1}{\sigma_1}\right)\left(\frac{\ln(z_2)-\mu_2}{\sigma_2}\right)\right], \tag{9}$$

$\phi(x_i)$ *is the standard normal and* $p_i(x_i) = \sum_{s=0}^{n} d_{is} H_s \left(\frac{\ln(z_i)-\mu_i}{\sigma_i}\right)$, i=1,2.

The marginal densities of this multivariate log-SNP are distributed as the univariate log-SNP – equation (5) – and therefore, empirical applications are very tractable because the parameter estimates of the marginals can be used as initial values for the joint estimation maximum likelihood (ML) procedures, and the relationship between (sample) moments and density parameters can be exploited for this purpose.

3. Distribution of firm size

3.1. Data description and statistics

Our study is conducted using a group of firms in the United States during the 2004-2015 period. The analysis is based on financial information from the set of firms available in the Thomson Reuters Datastream, a global database with time series reports on the accounts of firms listed in the stock markets. The following selection criteria were applied to the available information to obtain the final sample utilized for the estimation: Firms that did not have accounting data for the review period were excluded. Only companies that were active during this period (total assets and positive operating results) were considered. Firms with no available Standard Industrial Classification (SIC) code were also excluded. Given these criteria, a total sample of N=2,349 firms per year was used.[3]

Additionally, to analyze the firm size distribution by groups according to economic activity, the firms were divided in the following manner: Manufacturing (SIC codes 20-39), Non-manufacturing (SIC codes 10-14, 15-17, 40-49, 50-51, 52-59 and 70-89), Finance, Insurance and Real Estate (SIC codes 60-67) and Economy-wide, which includes all three aforementioned groups. Finally, the variable used to measure firm size is the value of average sales in dollars (USD).

[3] Note that in this study, we are not controlling for changes in the shape of the distribution of firms by the entrance and exit of companies or by mergers and acquisitions (M&As). In this regard, Cefis et al. [38] determine that M&As do not affect the size of the distribution when considering the entire population of firms. This result could be due to the balancing effect in which the entrance and exit of firms counteract the effect of M&As. However, this issue is presented as a limitation of the present study.

Table 1 contains the descriptive statistics for each of the four groups of industries examined in this study. The table shows the temporal behavior at the moments of distribution, calculated to the fourth order. In particular, the third and fourth central moments provide useful information regarding the shape of the distribution, in addition to the average and standard deviation. In general, the firm size distribution presents positive skewness, with a very large presence of small firms. The positive kurtosis also shows that the upper tail of the distribution is heavier than that observed in a lognormal distribution.

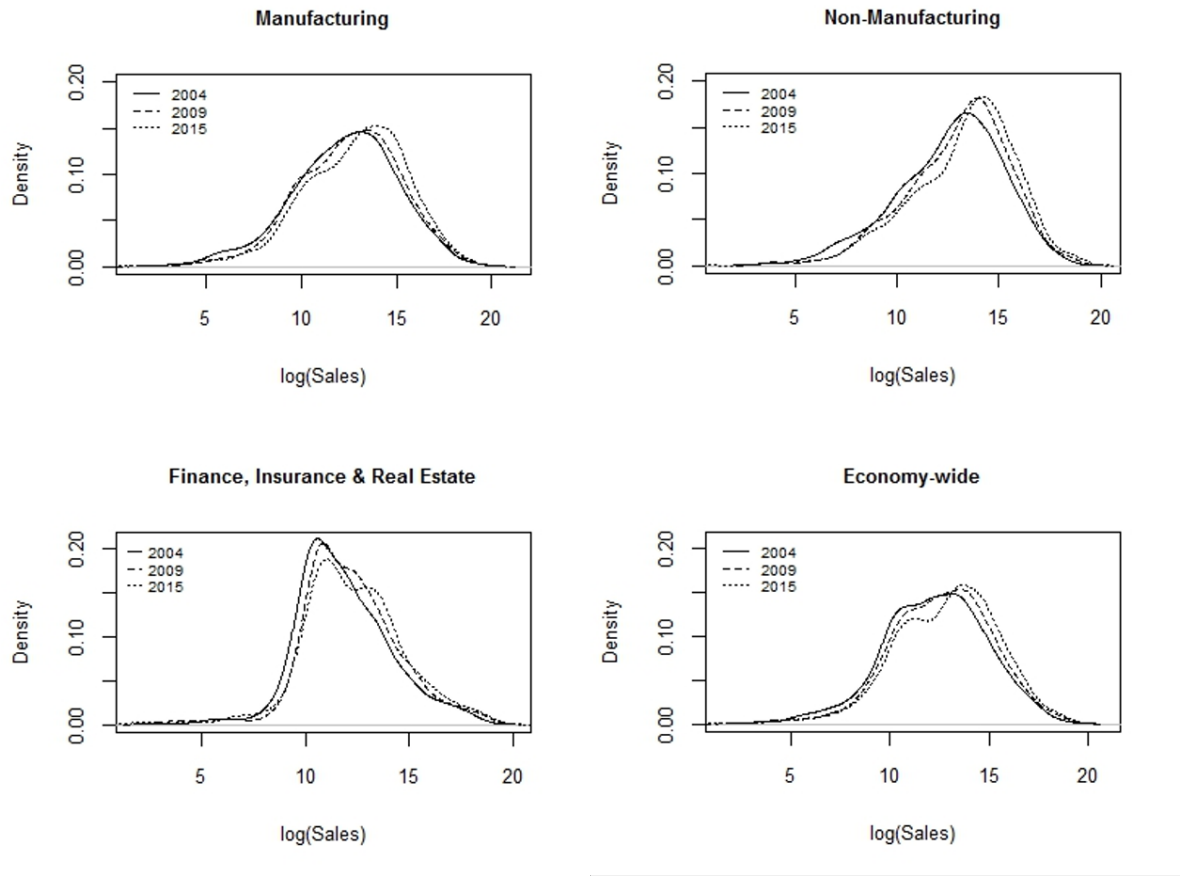
In the long term, for the groups of firms in Non-manufacturing and Finance, Insurance and Real Estate industries, the four moments exhibit a slight increase. However, in the Manufacturing industry, the tendency is reversed by the year 2013. In general, observing the Economy-wide category during the period studied, the firm size distribution became less dispersed around the average, less skewed toward the small firms and less thick at the tails.

Table 1 Descriptive statistics

Industry	Statistic	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015
Manufacturing (N=947)	Min $x10^6$	0.019	0.023	0.007	0.008	0.046	0.008	0.008	0.008	0.002	0.004	0.030	0.003
	Mean $x10^9$	3.38	3.80	4.06	4.35	4.79	3.94	4.44	5.04	4.99	4.96	4.99	4.55
	Max $x10^9$	263.99	328.21	335.09	358.60	425.07	275.56	341.58	433.53	420.71	390.25	364.76	236.81
	St.Dev $x10^{10}$	1.45	1.71	1.76	1.86	2.15	1.55	1.81	2.18	2.06	1.99	1.93	1.61
	Skewness	10.90	11.68	11.39	11.49	12.16	9.85	10.71	11.83	11.93	11.18	10.56	8.92
	Kurtosis	152.15	175.94	169.26	173.18	191.72	130.63	156.48	188.62	198.25	172.76	154.41	103.90
Non-manufacturing (N=784)	Min $x10^6$	0.025	0.035	0.046	0.021	0.034	0.048	0.050	0.007	0.005	0.002	0.009	0.001
	Mean $x10^9$	2.76	3.08	3.37	3.70	4.04	3.73	4.01	4.45	4.69	4.91	5.20	5.12
	Max $x10^9$	96.29	91.13	91.42	118.93	124.03	123.02	124.28	126.72	127.43	128.75	139.37	179.05
	St.Dev $x10^{10}$	0.79	0.84	0.92	1.06	1.14	1.14	1.18	1.28	1.37	1.40	1.49	1.60
	Skewness	6.32	6.00	6.03	6.47	6.27	6.79	6.42	6.06	6.05	5.89	6.00	6.81
	Kurtosis	51.25	45.00	44.41	50.61	47.36	53.89	48.49	43.18	42.36	40.15	41.87	53.91
Finance, Insurance & Real Estate (N=618)	Min $x10^6$	0.026	0.021	0.013	0.011	0.011	0.017	0.008	0.009	0.012	0.003	0.016	0.003
	Mean $x10^9$	2.11	2.44	2.88	3.19	2.91	3.09	3.39	3.41	3.53	3.54	3.61	3.72
	Max $x10^9$	108.28	120.28	146.56	159.23	112.45	143.27	155.70	146.70	160.84	179.54	194.17	209.85
	St.Dev $x10^{10}$	0.90	1.04	1.27	1.41	1.18	1.32	1.50	1.46	1.48	1.47	1.50	1.57
	Skewness	7.54	7.32	7.31	7.29	6.62	7.01	6.98	6.70	6.58	7.03	7.30	7.72
	Kurtosis	68.07	62.53	60.62	59.55	48.53	54.74	53.50	49.57	48.72	58.65	65.28	74.26
Economy-wide (N=2349)	Min $x10^6$	0.019	0.021	0.007	0.008	0.011	0.008	0.008	0.007	0.002	0.002	0.009	0.001
	Mean $x10^9$	2.84	3.20	3.52	3.83	4.04	3.65	4.02	4.41	4.51	4.57	4.70	4.52
	Max $x10^9$	263.99	328.21	335.09	358.60	425.07	275.56	341.58	433.53	420.71	390.25	364.76	236.81
	St.Dev $x10^{10}$	1.13	1.30	1.40	1.51	1.64	1.37	1.54	1.74	1.70	1.68	1.68	1.60
	Skewness	11.16	12.13	11.18	10.89	12.59	8.81	9.62	11.25	10.64	9.87	9.13	7.89
	Kurtosis	182.20	219.63	185.53	175.95	241.23	108.94	137.87	199.19	183.50	152.72	125.97	79.39

Figure 1 plots the density and the logarithm of the variable of sales, resulting from a softening of the corresponding histogram. To obtain a better visualization of the densities, three years (2004, 2009 and 2015), randomly selected, are displayed. The behavior described above can be observed for each of the four groups of industries. Additionally, it is clear that the firm size distributions have a different shape from the lognormal and are also bimodal or even multimodal, as in the studies by Marsili [14] and Bottazzi et al. [33].

Figure 1 Empirical density of the logarithm of sales



3.2. Results and discussion

Tables 2-5 present the ML estimation of the univariate case for each of the four groups of industries and 12 years selected in the sample. Panel A shows the estimated parameters for the lognormal distribution, and Panel B those of the log-SNP distribution. In Panel C, the LR statistic for the comparison of the log-SNP versus the lognormal is displayed.

Table 2 Estimation of the firm size distribution under lognormal and log-SNP, Manufacturing Industry

Year	Panel A lognormal					Panel B log-SNP									Panel C LR
	μ	σ	logL	AIC	KS test	μ	σ	d_1	d_2	d_3	d_4	logL	AIC	KS test	
2004	5.3598	2.7296	-7370.44	14744.87	(0.0231)	3.9696	2.1032	0.6610	0.5607	0.1308	0.0673	-7357.01	14726.02	(0.6906)	26.86
	(<.0001)	(<.0001)			Not rejected	(<.0001)	(<.0001)	(<.0001)	(<.0001)	(0.0030)	(0.0023)			Not rejected*	(<.0001)
2005	5.4834	2.7026	-7478.08	14960.15	(0.1427)	3.9566	2.1056	0.7251	0.5866	0.1631	0.0690	-7466.14	14944.28	(0.4308)	23.87
	(<.0001)	(<.0001)			Not rejected	(<.0001)	(<.0001)	(<.0001)	(<.0001)	(0.0003)	(0.0011)			Not rejected*	(<.0001)
2006	5.6136	2.6747	-7591.56	15187.13	(0.1754)	3.9653	2.1235	0.7763	0.5946	0.1842	0.0633	-7580.17	15172.34	(0.2831)	22.79
	(<.0001)	(<.0001)			Not rejected	(<.0001)	(<.0001)	(<.0001)	(<.0001)	(<.0001)	(0.0022)			Not rejected*	(0.0001)
2007	5.7088	2.6693	-7679.77	15363.55	(0.2584)	3.9904	2.1016	0.8177	0.6409	0.2095	0.0789	-7666.27	15344.55	(0.8677)	27.00
	(<.0001)	(<.0001)			Not rejected	(<.0001)	(<.0001)	(<.0001)	(<.0001)	(<.0001)	(<.0001)			Not rejected*	(<.0001)
2008	5.7688	2.6581	-7732.62	15469.24	(0.3981)	4.1036	2.0983	0.7936	0.6173	0.2024	0.0735	-7721.31	15454.63	(0.8022)	22.61
	(<.0001)	(<.0001)			Not rejected	(<.0001)	(<.0001)	(<.0001)	(<.0001)	(<.0001)	(<.0001)			Not rejected*	(0.0002)
2009	5.6448	2.6342	-7606.57	15217.14	(0.1584)	3.8858	2.1308	0.8255	0.6049	0.2011	0.0713	-7595.21	15202.42	(0.1939)	22.72
	(<.0001)	(<.0001)			Not rejected	(<.0001)	(<.0001)	(<.0001)	(<.0001)	(<.0001)	(<.0001)			Not rejected*	(0.0001)
2010	5.7446	2.6509	-7707.10	15418.19	(0.2138)	3.7281	2.2052	0.9144	0.6406	0.2235	0.0708	-7693.91	15399.83	(0.9740)	26.37
	(<.0001)	(<.0001)			Not rejected	(<.0001)	(<.0001)	(<.0001)	(<.0001)	(<.0001)	(<.0001)			Not rejected*	(<.0001)
2011	5.8257	2.7323	-7812.53	15629.07	(0.3981)	3.6698	2.2339	0.9651	0.7137	0.2389	0.0836	-7790.51	15593.02	(0.8962)	44.05
	(<.0001)	(<.0001)			Not rejected	(<.0001)	(<.0001)	(<.0001)	(<.0001)	(<.0001)	(<.0001)			Not rejected*	(<.0001)
2012	5.8615	2.7386	-7848.61	15701.22	(0.1283)	3.7389	2.2005	0.9646	0.7397	0.2500	0.0922	-7824.27	15660.53	(0.2353)	48.69
	(<.0001)	(<.0001)			Not rejected	(<.0001)	(<.0001)	(<.0001)	(<.0001)	(<.0001)	(<.0001)			Not rejected*	(<.0001)
2013	5.8823	2.7532	-7873.40	15750.80	(0.0010)	3.6045	2.2394	1.0172	0.7731	0.2708	0.0927	-7845.51	15703.02	(0.6133)	55.78
	(<.0001)	(<.0001)			Rejected	(<.0001)	(<.0001)	(<.0001)	(<.0001)	(<.0001)	(<.0001)			Not rejected*	(<.0001)
2014	5.9594	2.6853	-7922.73	15849.45	(0.0131)	3.9886	2.0708	0.9517	0.7937	0.2806	0.1059	-7897.59	15807.17	(0.8962)	50.28
	(<.0001)	(<.0001)			Not rejected	(<.0001)	(<.0001)	(<.0001)	(<.0001)	(<.0001)	(<.0001)			Not rejected*	(<.0001)
2015	5.9591	2.6700	-7917.02	15838.05	(0.0820)	4.0037	2.0577	0.9503	0.7933	0.2729	0.1065	-7888.91	15789.83	(0.9740)	56.22
	(<.0001)	(<.0001)			Not rejected	(<.0001)	(<.0001)	(<.0001)	(<.0001)	(<.0001)	(<.0001)			Not rejected*	(<.0001)

P-values are in parentheses. Not rejected* indicates a better fit in the KS test.

Table 3 Estimation of the firm size distribution under lognormal and log-SNP, Non-manufacturing Industries

Year	Panel A lognormal					Panel B log-SNP									Panel C LR
	μ	σ	logL	AIC	KS test	μ	σ	d_1	d_2	d_3	d_4	logL	AIC	KS test	
2004	5.6920	2.5976	-6323.36	12650.72	(0.2340)	4.1361	1.9310	0.8058	0.7294	0.1884	0.0725	-6297.28	12606.57	(0.9171)	52.16
	(<.0001)	(<.0001)			Not rejected	(<.0001)	(<.0001)	(<.0001)	(<.0001)	(0.0003)	(0.0055)			Not rejected*	(<.0001)
2005	5.8579	2.5606	-6442.18	12888.37	(0.1058)	4.4427	1.8864	0.7502	0.7027	0.1726	0.0616	-6416.19	12844.38	(0.8203)	51.99
	(<.0001)	(<.0001)			Not rejected	(<.0001)	(<.0001)	(<.0001)	(<.0001)	(0.0073)	(0.0202)			Not rejected*	(<.0001)
2006	5.9917	2.5499	-6543.85	13091.70	(0.0317)	4.5243	1.8624	0.7880	0.7478	0.1902	0.0711	-6514.36	13040.71	(0.8886)	58.98
	(<.0001)	(<.0001)			Not rejected	(<.0001)	(<.0001)	(<.0001)	(<.0001)	(0.0016)	(0.0054)			Not rejected*	(0.0001)
2007	6.1032	2.5241	-6623.24	13250.48	(0.0274)	4.3527	1.8995	0.9216	0.8075	0.2545	0.0873	-6593.65	13199.29	(0.9754)	59.19
	(<.0001)	(<.0001)			Not rejected	(<.0001)	(<.0001)	(<.0001)	(<.0001)	(<.0001)	(<.0001)			Not rejected*	(<.0001)
2008	6.2082	2.5122	-6701.89	13407.78	(0.0032)	4.1176	2.0048	1.0428	0.8289	0.2931	0.0881	-6671.73	13355.46	(0.6569)	60.32
	(<.0001)	(<.0001)			Rejected	(<.0001)	(<.0001)	(<.0001)	(<.0001)	(<.0001)	(<.0001)			Not rejected*	(0.0002)
2009	6.1111	2.4789	-6615.24	13234.49	(0.0149)	3.8505	2.0423	1.1069	0.8493	0.3120	0.0858	-6586.47	13184.94	(0.6144)	57.55
	(<.0001)	(<.0001)			Not rejected	(<.0001)	(<.0001)	(<.0001)	(<.0001)	(<.0001)	(<.0001)			Not rejected*	(0.0001)
2010	6.1958	2.4821	-6682.65	13369.30	(0.2106)	3.8191	2.0698	1.1482	0.8782	0.3224	0.0946	-6652.11	13316.22	(0.8561)	61.09
	(<.0001)	(<.0001)			Not rejected	(<.0001)	(<.0001)	(<.0001)	(<.0001)	(<.0001)	(<.0001)			Not rejected*	(<.0001)
2011	6.2690	2.5358	-6756.89	13517.79	(0.0203)	3.9474	2.0741	1.1194	0.8738	0.3087	0.0905	-6724.04	13460.09	(0.7817)	65.70
	(<.0001)	(<.0001)			Not rejected	(<.0001)	(<.0001)	(<.0001)	(<.0001)	(<.0001)	(<.0001)			Not rejected*	(<.0001)
2012	6.3150	2.5448	-6795.72	13595.44	(0.0108)	3.9355	2.0671	1.1511	0.9204	0.3346	0.1047	-6760.12	13532.24	(0.2866)	71.20
	(<.0001)	(<.0001)			Not rejected	(<.0001)	(<.0001)	(<.0001)	(<.0001)	(<.0001)	(<.0001)			Not rejected*	(<.0001)
2013	6.3541	2.5790	-6836.84	13677.67	(0.0027)	4.0593	2.0753	1.1058	0.8835	0.3096	0.0943	-6800.73	13613.45	(0.9859)	72.22
	(<.0001)	(<.0001)			Rejected	(<.0001)	(<.0001)	(<.0001)	(<.0001)	(<.0001)	(<.0001)			Not rejected*	(<.0001)
2014	6.4151	2.5903	-6888.11	13780.23	(0.0824)	3.9697	2.0955	1.1670	0.9449	0.3549	0.1093	-6846.59	13705.18	(0.7412)	83.05
	(<.0001)	(<.0001)			Not rejected	(<.0001)	(<.0001)	(<.0001)	(<.0001)	(<.0001)	(<.0001)			Not rejected*	(<.0001)
2015	6.3842	2.5857	-6862.43	13728.86	(0.0127)	3.9279	2.0970	1.1713	0.9461	0.3578	0.1037	-6820.71	13653.42	(0.2340)	83.44
	(<.0001)	(<.0001)			Not rejected	(<.0001)	(<.0001)	(<.0001)	(<.0001)	(<.0001)	(<.0001)			Not rejected*	(<.0001)

P-values are in parentheses. Not rejected* indicates a better fit in the KS test.

Table 4 Estimation of the firm size distribution under lognormal and log-SNP, Finance, Insurance and Real Estate Industries

Year	Panel A lognormal					Panel B log-SNP									Panel C LR
	μ	σ	logL	AIC	KS test	μ	σ	d_1	d_2	d_3	d_4	logL	AIC	KS test	
2004	4.9348	2.2025	-4414.58	8833.15	(0.0561)	4.1876	1.9570	0.3818	0.2061	0.1077	0.0984	-4381.61	8775.22	(0.4603)	65.94
	(<.0001)	(<.0001)			Not rejected	(<.0001)	(<.0001)	(<.0001)	(<.0001)	(0.0011)	(<.0001)			Not rejected*	(<.0001)
2005	5.0986	2.1826	-4510.20	9024.40	(0.0179)	4.3265	1.9581	0.3943	0.1990	0.0990	0.0988	-4475.77	8963.54	(0.3788)	68.86
	(<.0001)	(<.0001)			Not rejected	(<.0001)	(<.0001)	(<.0001)	(0.0016)	(0.0030)	(<.0001)			Not rejected*	(<.0001)
2006	5.2375	2.2216	-4606.95	9217.91	(0.0872)	4.4807	2.0436	0.3703	0.1594	0.0661	0.0923	-4569.33	9150.66	(0.3788)	75.24
	(<.0001)	(<.0001)			Not rejected	(<.0001)	(<.0001)	(<.0001)	(0.0103)	(0.0457)	(<.0001)			Not rejected*	(0.0001)
2007	5.3224	2.2742	-4673.91	9351.82	(0.0017)	4.4390	2.1017	0.4203	0.1738	0.0480	0.0936	-4631.00	9274.00	(0.4184)	85.82
	(<.0001)	(<.0001)			Rejected	(<.0001)	(<.0001)	(<.0001)	(0.0070)	(0.1664)	(<.0001)			Not rejected*	(<.0001)
2008	5.3157	2.2614	-4666.27	9336.53	(0.0179)	4.4822	2.0080	0.4151	0.2203	0.0704	0.1056	-4623.26	9258.52	(0.3416)	86.01
	(<.0001)	(<.0001)			Not rejected	(<.0001)	(<.0001)	(<.0001)	(<.0001)	(0.0384)	(<.0001)			Not rejected*	(<.0001)
2009	5.2705	2.3032	-4649.67	9303.34	(0.0481)	4.3158	2.1107	0.4523	0.1977	0.0478	0.1013	-4602.54	9217.07	(0.8284)	94.26
	(<.0001)	(<.0001)			Not rejected	(<.0001)	(<.0001)	(<.0001)	(0.0019)	(0.1714)	(<.0001)			Not rejected*	(<.0001)
2010	5.2595	2.3541	-4656.39	9316.77	(0.0481)	4.3900	2.1587	0.4028	0.1757	0.0391	0.0993	-4608.45	9228.90	(0.4603)	95.88
	(<.0001)	(<.0001)			Not rejected	(<.0001)	(<.0001)	(<.0001)	(0.0049)	(0.2504)	(<.0001)			Not rejected*	(<.0001)
2011	5.2829	2.3531	-4670.56	9345.12	(0.1317)	4.2784	2.1454	0.4682	0.2111	0.0443	0.1078	-4617.88	9247.76	(0.1501)	105.36
	(<.0001)	(<.0001)			Not rejected	(<.0001)	(<.0001)	(<.0001)	(0.0011)	(0.2125)	(<.0001)			Not rejected*	(<.0001)
2012	5.2923	2.3979	-4688.05	9380.10	(0.0072)	4.3261	2.1399	0.4515	0.2297	0.0510	0.1062	-4640.82	9293.64	(0.3788)	94.46
	(<.0001)	(<.0001)			Rejected	(<.0001)	(<.0001)	(<.0001)	(<.0001)	(0.1470)	(<.0001)			Not rejected*	(<.0001)
2013	5.3019	2.4297	-4702.08	9408.16	(0.0017)	4.3472	2.1709	0.4398	0.2230	0.0523	0.0989	-4660.23	9332.46	(0.1004)	83.70
	(<.0001)	(<.0001)			Rejected	(<.0001)	(<.0001)	(<.0001)	(0.0011)	(0.1352)	(<.0001)			Not rejected*	(<.0001)
2014	5.3788	2.3615	-4732.01	9468.03	(0.1931)	4.4620	2.0670	0.4435	0.2510	0.0766	0.1003	-4699.90	9411.80	(0.5970)	64.23
	(<.0001)	(<.0001)			Not rejected	(<.0001)	(<.0001)	(<.0001)	(<.0001)	(0.0252)	(<.0001)			Not rejected*	(<.0001)
2015	5.4117	2.4111	-4765.25	9534.49	(0.2180)	4.5211	2.2107	0.4028	0.1759	0.0376	0.0759	-4740.59	9493.18	(0.4603)	49.31
	(<.0001)	(<.0001)			Not rejected	(<.0001)	(<.0001)	(<.0001)	(0.0238)	(0.2618)	(<.0001)			Not rejected*	(<.0001)

P-values are in parentheses. Not rejected* indicates a better fit in the KS test.

Table 5 Estimation of the firm size distribution under lognormal and log-SNP, Economy-wide

Year	Panel A lognormal					Panel B log-SNP									Panel C LR
	μ	σ	logL	AIC	KS test	μ	σ	d_1	d_2	d_3	d_4	logL	AIC	KS test	
2004	5.3589	2.5722	-18140.36	36284.72	(0.5641)	4.0772	2.0039	0.6396	0.5283	0.1438	0.0797	-18113.55	36239.11	(0.9327)	53.62
	(<.0001)	(<.0001)			Not rejected	(<.0001)	(<.0001)	(<.0001)	(<.0001)	(<.0001)	(<.0001)			Not rejected*	(<.0001)
2005	5.5072	2.5439	-18462.70	36929.39	(0.4937)	4.1626	2.0102	0.6689	0.5245	0.1555	0.0738	-18439.89	36891.78	(0.1942)	45.61
	(<.0001)	(<.0001)			Not rejected	(<.0001)	(<.0001)	(<.0001)	(<.0001)	(<.0001)	(<.0001)			Not rejected*	(<.0001)
2006	5.6409	2.5371	-18770.48	37544.97	(0.4491)	4.2123	2.0282	0.7044	0.5305	0.1562	0.0705	-18745.70	37503.40	(0.8044)	49.57
	(<.0001)	(<.0001)			Not rejected	(<.0001)	(<.0001)	(<.0001)	(<.0001)	(<.0001)	(<.0001)			Not rejected*	(<.0001)
2007	5.7388	2.5396	-19002.83	38009.66	(0.0740)	4.1594	2.0343	0.7764	0.5807	0.1797	0.0837	-18967.08	37946.15	(0.5167)	71.51
	(<.0001)	(<.0001)			Not rejected	(<.0001)	(<.0001)	(<.0001)	(<.0001)	(<.0001)	(<.0001)			Not rejected*	(<.0001)
2008	5.7963	2.5334	-19132.03	38268.06	(0.0239)	4.1958	2.0424	0.7836	0.5764	0.1858	0.0830	-19096.15	38204.31	(0.5167)	71.75
	(<.0001)	(<.0001)			Not rejected	(<.0001)	(<.0001)	(<.0001)	(<.0001)	(<.0001)	(<.0001)			Not rejected*	(<.0001)
2009	5.7019	2.5199	-18897.95	37799.91	(0.0686)	3.9814	2.0941	0.8216	0.5616	0.1768	0.0751	-18862.74	37737.47	(0.2629)	70.44
	(<.0001)	(<.0001)			Not rejected	(<.0001)	(<.0001)	(<.0001)	(<.0001)	(<.0001)	(<.0001)			Not rejected*	(<.0001)
2010	5.7676	2.5449	-19075.26	38154.51	(0.2949)	3.8747	2.1719	0.8715	0.5662	0.1761	0.0714	-19034.35	38080.70	(0.4711)	81.81
	(<.0001)	(<.0001)			Not rejected	(<.0001)	(<.0001)	(<.0001)	(<.0001)	(<.0001)	(<.0001)			Not rejected*	(<.0001)
2011	5.8309	2.5992	-19273.51	38551.01	(0.5402)	3.9387	2.1652	0.8739	0.6023	0.1826	0.0797	-19222.03	38456.07	(0.9746)	102.95
	(<.0001)	(<.0001)			Not rejected	(<.0001)	(<.0001)	(<.0001)	(<.0001)	(<.0001)	(<.0001)			Not rejected*	(<.0001)
2012	5.8631	2.6176	-19365.85	38735.70	(0.1142)	3.9475	2.1432	0.8938	0.6453	0.2047	0.0904	-19308.43	38628.86	(0.3669)	114.84
	(<.0001)	(<.0001)			Not rejected	(<.0001)	(<.0001)	(<.0001)	(<.0001)	(<.0001)	(<.0001)			Not rejected*	(<.0001)
2013	5.8871	2.6442	-19445.94	38895.87	(0.0364)	3.8808	2.1976	0.9130	0.6407	0.2026	0.0815	-19388.47	38788.93	(0.5167)	114.94
	(<.0001)	(<.0001)			Not rejected	(<.0001)	(<.0001)	(<.0001)	(<.0001)	(<.0001)	(<.0001)			Not rejected*	(<.0001)
2014	5.9587	2.6022	-19576.70	39157.39	(0.1602)	4.1042	2.0662	0.8976	0.6959	0.2405	0.1001	-19522.27	39056.55	(0.5167)	108.85
	(<.0001)	(<.0001)			Not rejected	(<.0001)	(<.0001)	(<.0001)	(<.0001)	(<.0001)	(<.0001)			Not rejected*	(<.0001)
2015	5.9570	2.6027	-19572.92	39149.84	(0.1823)	4.0172	2.1084	0.9200	0.6852	0.2290	0.0886	-19519.64	39051.28	(0.5882)	106.55
	(<.0001)	(<.0001)			Not rejected	(<.0001)	(<.0001)	(<.0001)	(<.0001)	(<.0001)	(<.0001)			Not rejected*	(<.0001)

P-values are in parentheses. Not rejected* indicates a better fit in the KS test.

The results of the estimation show that all of the models sufficiently capture the mean and standard deviation of each of the groups of industries; these statistics are represented by the location parameters μ and scale parameters σ, respectively. As shown, the p-values indicate that these parameters are highly significant for both distributions. However, as shown in Panel B, for the log-SNP distribution, the d_s parameters are also highly significant for the majority of the years and industries. Analyzing the Akaike Information Criteria (AIC), which penalizes the inclusion of additional parameters for the distributions, it is found that this criterion is consistently lower for the log-SNP, which suggests that modeling based on this distribution is clearly superior.

Meanwhile, the LR statistics included in Panel C conclude that for the majority of the years and industries selected, the incorporation of the d_s parameters improves the LR of the model. These results are consistent with the Kolmogorov-Smirnov (KS) test applied to each of the distributions. Based on a significance level of 0.01, for the majority of the years and across the four groups, the test cannot reject the null hypothesis that the data-generating process comes from a theoretical lognormal or log-SNP distribution. However, for all years and across each of the industries, the log-SNP distribution shows a better fit. Note that despite the differences in the number of firms and economic activity performed by each of the four groups selected, the shape of the firm size distributions is similar.

An example of the quality of fit obtained for each of the industries in the year 2015 can be observed in Figure 2, which shows, on a logarithmic scale, the relationship between the rank and size of sales. As shown through a comparison of empirical values (unfilled dots) and values estimated theoretically under the lognormal specification (discontinuous line) and log-SNP (solid line), the log-SNP distribution more accurately captures not only the values regarding the averages but also the extreme values.[4]

[4] The behavior for the remaining years is similar.

Figure 2 Logarithm of firm rank vs. logarithm of sales by the firm

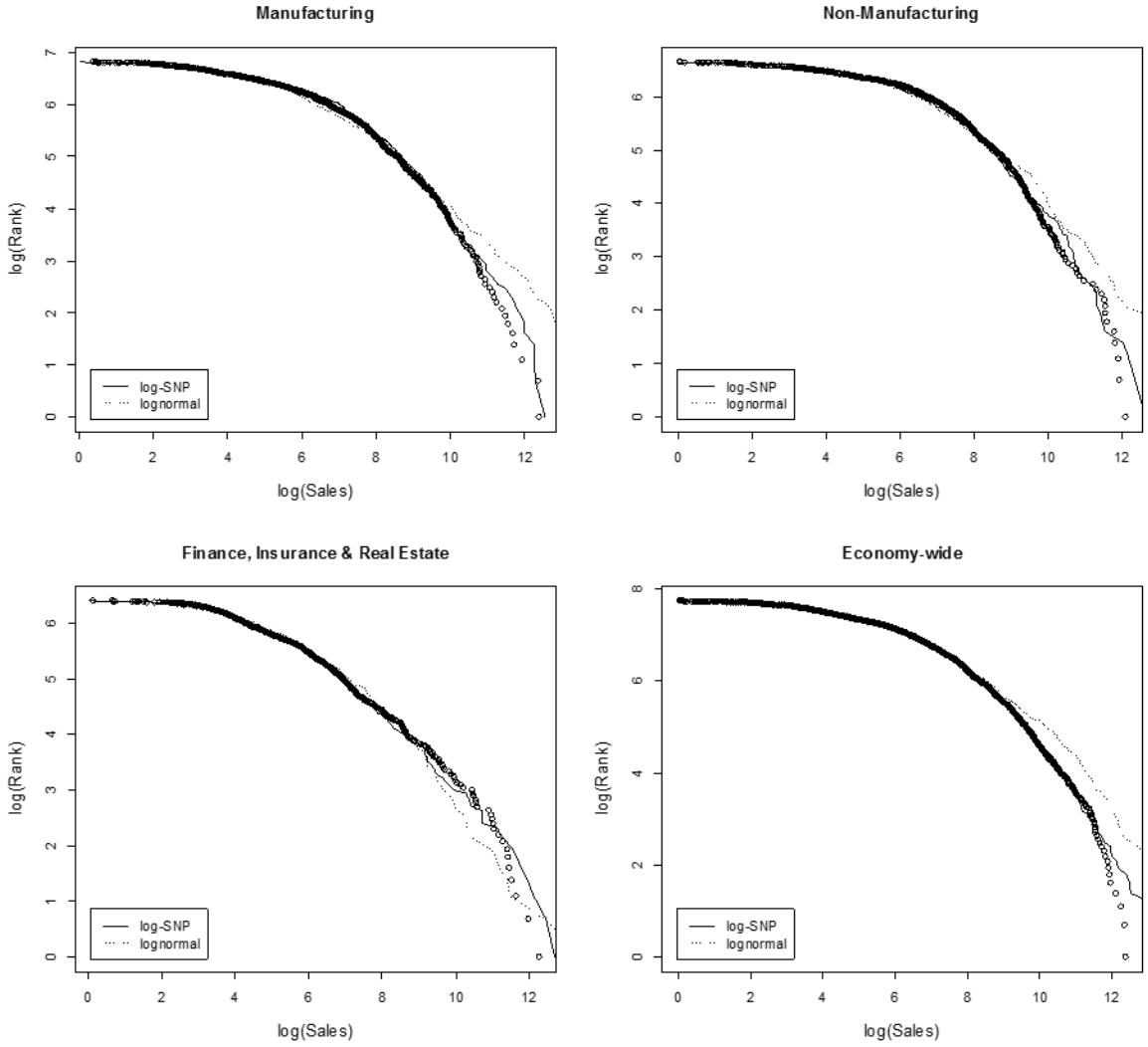


Figure 2 illustrates how the lognormal distribution overestimates the extreme values in the distribution. These results are consistent with those reported in previous research in which the lognormal distribution consistently underestimates or overestimates the theoretically expected values at the upper tail of the firm size distribution [1] [2]. Table 6 corroborates these effects for the Manufacturing industry, calculating the empirical and estimated upper quantiles under the lognormal and log-SNP for confidence levels of 10%,

5% and 1%.[5] We provide the results for a single industry here, conducting the analysis for the remaining industries and obtaining qualitatively similar results.[6]

Table 6 Value of sales empirically observed versus theoretically expected under lognormal and log-SNP

Year	Observed Sales value (millions, US dollars)			Expected Sales value (millions, US dollars)					
				Lognormal			Log-SNP		
	10%	5%	1.0%	10%	5%	1.0%	10%	5%	1.0%
2004	6,134.60	14,356.17	51,974.00	7,030.40	18,952.33	121,772.43	6,014.81	13,237.28	54,958.85
2005	6,603.66	15,043.38	55,868.84	7,684.46	20,512.97	129,392.69	6,598.52	14,322.48	58,345.11
2006	7,433.20	18,030.20	60,788.02	8,446.12	22,319.20	138,139.58	7,251.54	15,638.38	63,459.94
2007	7,750.20	19,133.65	60,310.24	9,224.81	24,328.55	150,016.27	7,887.45	16,788.62	66,459.65
2008	8,076.61	20,111.00	67,024.26	9,655.94	25,362.41	155,204.98	8,394.62	17,921.06	71,209.05
2009	7,218.28	17,030.90	65,292.00	8,271.88	21,539.06	129,678.04	7,250.99	15,625.23	63,327.55
2010	7,856.00	18,692.20	66,107.04	9,338.44	24,464.64	148,982.21	8,011.79	17,432.64	72,667.96
2011	8,369.81	19,970.51	75,190.16	11,240.25	30,330.24	195,230.32	8,991.23	19,557.65	81,846.26
2012	8,673.40	20,145.28	75,039.96	11,744.70	31,764.51	205,347.60	9,282.22	19,935.42	81,395.02
2013	8,772.35	20,860.57	78,253.70	12,218.77	33,222.39	216,919.80	9,454.11	20,367.08	84,142.50
2014	9,512.72	20,191.80	78,041.34	12,097.21	32,089.90	200,043.57	9,406.80	19,227.65	71,621.03
2015	9,330.04	20,170.31	68,975.52	11,859.39	31,285.45	193,014.95	9,175.00	18,691.47	69,169.41

Analyzing the tendency of the values at the upper tail of the distribution for sales during the period studied, it is observed that the flexible parametric structure of the log-SNP distribution allows for a better fit of the expected values. The interpretation of the values from this table highlights the errors induced in the estimation of the firm size distribution by the use of traditional parametric distributions such as the lognormal.

3.3. The log-SNP bivariate distribution: Sales vs. assets

Firm size can be measured by different variables: sales, assets, employees, or benefits, among others [34] [35] [7]. This diversity of measures suggests that there is no single best indicator of size and that the selection primarily depends on the available data [36]. To shed light on the robustness of the results obtained above, we take the size of the value of the total

[5] To obtain the quantiles of the log-SNP distribution, we use the CDF presented in equation (3) and the inverse transform method (ITM).
[6] The results for the other industries are available upon request.

assets of firms in Manufacturing as an additional measure. Table 7 shows the descriptive statistics for this variable during the entire period studied.

Table 7 Descriptive statistics, total assets

Industry	Statistic	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015
Manufacturing (N=947)	Min $x10^6$	0.145	0.174	0.153	0.050	0.050	0.030	0.034	0.023	0.029	0.020	0.005	0.010
	Mean $x10^9$	4.26	4.35	4.77	5.20	5.11	5.35	5.71	6.03	6.32	6.61	6.78	6.94
	Max $x10^9$	750.33	673.34	697.24	795.34	797.77	781.82	751.22	717.24	685.33	656.29	645.81	508.14
	St.Dev $x10^{10}$	2.85	2.64	2.79	3.10	3.03	3.07	3.07	3.07	3.05	3.06	3.07	2.88
	Skewness	20.73	19.10	18.20	19.26	20.12	18.63	17.08	15.62	14.44	13.35	12.92	10.18
	Kurtosis	512.32	449.59	415.24	459.05	499.00	440.85	379.15	323.04	280.09	240.54	225.10	135.90

As shown in the table, the firm size distribution measured based on the variable of assets shows positive asymmetry, with the presence of a very high quantity of small firms and a low number of large firms. The shape of the distribution suggested by these statistics is consistent with that observed using the sales variable. However, the kurtosis also shows that the upper tail of the distribution is even thicker than that exhibited for the sales variable.

Table 8 provides an example of the results of the estimation of the joint distribution of the value of assets and of sales for firms in the years 2008 and 2009. These years are relevant because, as observed in Tables 1 and 7, they mark a break in the tendency of the moments of distribution due to the global financial crisis. Specifically, using ML, we estimate the parameters of the bivariate case of the densities of the lognormal and log-SNP distributions described in the above sections. We implement the estimation in a sequential manner, beginning with the simplest univariate density, the lognormal, and recursively add the parameters whose results served as initial values.

Regarding the results, Panel A of the table gathers the estimated parameters for the lognormal distribution, and Panel B displays the estimated parameters for the log-SNP distribution. As shown, we only estimate the parameters d_{3i} and d_{4i} (i = 1 assets, i = 2 sales), reinforcing the fact that the densities must be expanded to higher polynomials to capture the probabilistic mass at the extreme end of the tails. Note that for both distributions, all parameters are significant.

Table 8 Estimation of the assets-sales bivariate case

	Panel A lognormal		Panel B log-SNP	
	2008	**2009**	**2008**	**2009**
μ_1	5.8310	5.7997	5.2752	5.6858
	(<.0001)	(<.0001)	(<.0001)	(<.0001)
σ_1	2.5085	2.5385	3.3442	2.8759
	(<.0001)	(<.0001)	(<.0001)	(<.0001)
d_{31}			-2.1829	0.4319
			(<.0001)	(<.0001)
d_{41}			1.3327	0.4059
			(<.0001)	(<.0001)
μ_2	5.7685	5.6448	5.3791	5.6201
	(<.0001)	(<.0001)	(<.0001)	(<.0001)
σ_2	2.6589	2.6342	3.1727	2.8036
	(<.0001)	(<.0001)	(<.0001)	(<.0001)
d_{32}			2.3255	0.5729
			(<.0001)	(<.0001)
d_{42}			-0.1443	-0.1569
			(0.0210)	(<.0001)
ρ	0.9539	0.9547	0.9890	0.9871
	(<.0001)	(<.0001)	(<.0001)	(<.0001)
logL	-3343.27	-3339.48	-2832.68	-2983.98
BIC	3360.41	3356.61	2863.52	3014.82

P-values are in parentheses.

Regarding the estimated correlation, ρ is also significant, and although this parameter does not exactly capture the correlation between the two variables, a very high dependency between them is observed. Comparing the Bayesian Information Criterion (BIC), which penalizes the inclusion of additional parameters, for the two distributions, it is found that this criterion is consistently lower for the bivariate log-SNP distribution. This finding suggests that as in the univariate case, the model based on this distribution is clearly superior.

Additionally, we performed the Wald test to analyze the relationship between the estimated parameters in the log-SNP distribution. The results obtained are found in Table 9. Note that, in general, for the two years selected, the null hypothesis of equal values of the counterpart coefficients in both marginals is rejected, which indicates that, although the series are highly correlated, significant differences regarding the behavior of the extreme values can be found. However, for 2009, the difference between the d_{3i} parameters is significantly equal to zero.

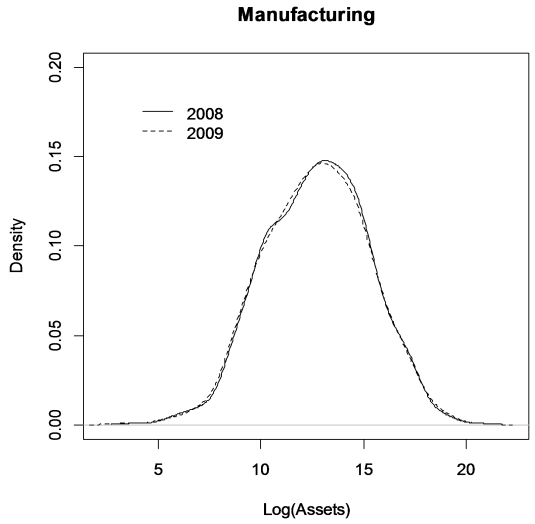
Table 9 Wald test, log-SNP distribution

	2008	2009
μ_1-μ_2	-0.1039 (<.0001)	0.0657 (<.0001)
σ_1-σ_2	0.1714 (<.0001)	0.0723 (<.0001)
d_{31}-d_{32}	-4.5084 (<.0001)	0.1411 (0.2110)
d_{41}-d_{42}	1.4770 (<.0001)	0.5628 (<.0001)
Chi Sq(4)	5049.63 (<.0001)	203.42 (<.0001)

P-values are in parentheses.

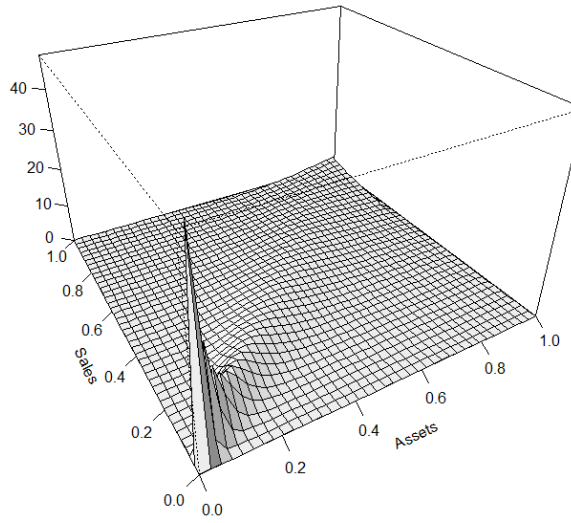
As shown in Figure 3, in 2009, the value of the assets measured in logarithms became more symmetrical with respect to the previous year, which also resulted in a minor difference between the location parameters μ_i and the scale parameters σ_i. According to Pascoal et al. [22], a possible explanation for this behavior is that assets act as a survival mechanism in times of economic instability. Additionally, a reduction in sales in addition to large fixed costs can lead to heavy losses that immediately result in a reduction in assets (which can act as a safety net in adverse economic situations), which can lead to new sales losses. As a result, this behavior can also imply a reduction in the difference between large and small firm sizes. In this manner, the value of sales appears to be a better firm size measure compared to other measures, as indicated in the literature [37] [22].

Figure 3 Empirical density of the logarithm of assets



These effects can also be observed in Figure 4, which shows the histogram of the joint distribution of the values of assets and of sales for firms during the year 2009. Normally, firms with low sales values have low asset values (and vice versa); however, some firms may have high sales values in addition to asset values that, comparatively, are relatively lower.

Figure 4 Histogram of the assets-sales joint distribution



4. Conclusions

We propose a new methodology based on the log-SNP distribution for modeling the firm size distribution. This distribution nests the lognormal, including new parameters that are capable of better capturing the behavior of the upper tail of the firm sizes, which makes it possible to contrast the deficiencies of the lognormal distribution in this direction. In the empirical application we compare the performance of both distributions, adjusting a sample of US firms in the 2004-2015 period.

The results show that the lognormal distribution tends to consistently overestimate the expected values at the upper tail of the distribution. This finding raises the need to propose other flexible distributions that allow for the gathering of more reliable information regarding the level of industrial concentration and economic cycles and, therefore, the implementation of competition policies. Taking different aggregation levels by economic activity, our study

shows that the log-SNP provides a better fit for the distribution of firm sizes. Meanwhile, it is more flexible than the lognormal when the data are very skewed and there are possible jumps in the upper tail due to the extreme observations.

We are also the first to develop an expression for the density of the multivariate log-SNP distribution and to analyze the estimation of the distribution together with the value of the assets and sales of the firms. This distribution nests the multivariate lognormal and has marginal log-SNP densities, which facilitates the estimation procedures. The results suggest that sales are a better firm size measure, as has been determined in previous studies. Despite the high correlation between the value of a firm's assets and sales, in periods of financial crisis, assets can act as a safety net for survival in the face of economic instability. This fact could lead to distorted conclusions in the analysis of the behavior of the distribution of firm sizes based on this variable.

References

[1] M. H. Stanley, S. V. Buldyrev, S. Havlin, R. N. Mantegna, M. A. Salinger and H. E. Stanley, "Zipf plots and the size distribution of firms," *Economics Letters,* vol. 49, no. 4, pp. 453-457, 1995.

[2] P. E. Hart and N. Oulton, "Zipf and the size distribution of firms," *Applied Economics Letters,* vol. 4, no. 4, pp. 205-206, 1997.

[3] H. M. Gupta, J. R. Campanha, D. R. de Aguiar, G. A. Queiroz and C. G. Raheja, "Gradually truncated log-normal in USA publicly traded firm size distribution," *Physica A: Statistical Mechanics and its Applications,* vol. 375, no. 2, pp. 643-650, 2007.

[4] R. Hernández-Pérez, "An analogy of the size distribution of business firms with Bose–Einstein statistics," *Physica A: Statistical Mechanics and its Applications,* vol. 389, no. 18, pp. 3837-3843, 2010.

[5] H. Simon and C. Bonini, "The size distribution of business firms," *The American Economic Review,* vol. 48, no. 4, pp. 607-617, 1958.

[6] G. Barba Navaretti, D. Castellani and F. Pieri, "Age and firm growth: evidence from three European countries," *Small Business Economics,* vol. 43, no. 4, p. 823–837, 2014.

[7] T. Heinrich and S. Dai, "Diversity of firm sizes, complexity, and industry structure in the Chinese economy," *Structural Change and Economic Dynamics,* vol. 37, pp. 90-106, 2016.

[8] R. Gibrat, Les Inégalités Economiques, Paris: Recueil Sirey, 1931.

[9] J. Voit, "The growth dynamics of German business firms," *Advances in Complex Systems,* vol. 4, no. 1, pp. 149-162, 2001.

[10] T. Kaizoji, H. Iyetomi and Y. Ikeda, "Re-examination of the size distribution of firms," *Evolutionary and Institutional Economics Review,* vol. 2, no. 2, p. 183–198, 2006.

[11] A. Coad, "The exponential age distribution and the Pareto firm size distribution," *Journal of Industry, Competition and Trade,* vol. 10, no. 3, p. 389–395, 2010.

[12] R. Axtell, "Zipf distribution of U.S. Firm sizes," *Science,* vol. 293, p. 1818–1820, 2001.

[13] L. M. Cabral and J. Mata, "On the evolution of the firm size distribution: facts and theory," *American Economic Review,* vol. 93, no. 4, pp. 1075-1090, 2003.

[14] O. Marsili, "Stability and turbulence in the size distribution of firms: evidence from Dutch manufacturing," *International Review of Applied Economics,* vol. 20, no. 2, pp. 255-272, 2006.

[15] J. Goddard, H. Liu, M. Donal and J. O. Wilson, "The size distribution of US banks and credit unions," *International Journal of the Economics of Business,* vol. 21, no. 1, pp. 139-156, 2014.

[16] G. Bottazzi, D. Pirino and F. Tamagni, "Zipf law and the firm size distribution: a critical discussion of popular estimators," *Journal of Evolutionary Economics,* vol. 25, no. 3, p. 585–610, 2015.

[17] G. Dosi, O. Marsili, L. Orsenigo and R. Salvatore, "Learning, market selection and the evolution of industrial structures," *Small Business Economics,* vol. 7, no. 6, p. 411–436, 1995.

[18] L. Crosato and P. Ganugi, "Statistical regularity of firm size distribution: the Pareto IV and truncated Yule for Italian SCI manufacturing," *Statistical Methods and Applications,* vol. 16, no. 1, pp. 85-115, 2007.

[19] M. J. Newman, "Power laws, Pareto distributions and Zipf's law," *Contemporary Physics,* vol. 46, no. 5, pp. 323-351, 2005.

[20] G. Martínez-Mekler, R. A. Martínez, M. B. del Río, R. Mansilla, P. Miramontes and G. Cocho, "Universality of rank-ordering distributions in the arts and sciences," *PLoS ONE,* vol. 4, no. 3, p. e4791, 2009.

[21] J. di Giovanni, A. A. Levchenko and R. Rancière, "Power laws in firm size and openness to trade: measurement and implications," *Journal of International Economics,* vol. 85, no. 1, pp. 42-52, 2011.

[22] R. Pascoal, M. Augusto and A. M. Monteiro, "Size distribution of Portuguese firms between 2006 and 2012," *Physica A: Statistical Mechanics and its Applications,* vol. 458, pp. 342-355, 2016.

[23] P. Cirillo and J. Hüsler, "On the upper tail of Italian firms' size distribution," *Physica A: Statistical Mechanics and its Applications,* vol. 388, no. 8, pp. 1546-1554, 2009.

[24] W. F. Kuhs, "The anharmonic temperature factor in crystallographic structure analysis," *Australian Journal of Physics,* vol. 41, no. 3, pp. 369-382, 1988.

[25] S. Blinnikov and R. Moessner, "Expansions for nearly Gaussian distributions," *Astronomy and astrophysics Supplement Series,* vol. 130, no. 1, p. 193–205, 1998.

[26] I. Mauleón and J. Perote, "Testing densities with financial data: an empirical comparison of the Edgeworth-Sargan density to the Student's t," *European Journal of Finance,* vol. 6, no. 2, pp. 225-239, 2000.

[27] L. M. Cortés, A. Mora-Valencia and J. Perote, "The productivity of top researchers: a semi-nonparametric approach," *Scientometrics,* vol. 109, no. 2, pp. 891-915, 2016.

[28] R. Jarrow and A. Rudd, "Approximate option valuation for arbitrary stochastic processes," *Journal of Financial Economics,* vol. 10, no. 3, pp. 347-369, 1982.

[29] T.-M. Ñíguez, I. Paya, D. Peel and J. Perote, "On the stability of the constant relative risk aversion (CRRA) utility under high degrees of uncertainty," *Economics Letters,* vol. 115, no. 2, pp. 244-248, 2012.

[30] T.-M. Ñíguez, I. Paya, D. Peel and J. Perote, "Higher-order moments in the theory of diversification and portfolio composition," *Economics Working Paper Series 2013/003. Lancaster University,* 2013.

[31] J. Perote, "The multivariate Edgeworth-Sargan density.," *Spanish Economic Review,* vol. 6, no. 1, pp. 77-96, 2004.

[32] J. Aitchison and J. Brown, The Lognormal Distribution, Cambridge: Cambridge University Press, 1957.

[33] G. Bottazzi, E. Cefis, G. Dosi and A. Secchi, "Invariances and diversities in the patterns of industrial evolution: some evidence from Italian manufacturing industries," *Small Business Economics,* vol. 29, no. 1, pp. 137-159, 2007.

[34] F. Delmar, P. Davidsson and W. B. Gartner, "Arriving at the high-growth firm," *Journal of Business Venturing,* vol. 18, no. 2, pp. 189-216, 2003.

[35] J. Zhang , Q. Chen and Y. Wang, "Zipf distribution in top Chinese firms and an economic explanation," *Physica A: Statistical Mechanics and its Applications,* vol. 388, no. 10, pp. 2020-2024, 2009.

[36] N. Barbosa and V. Eiriz, "Regional variation of firm size and growth: the Portuguese case," *Growth and Change,* vol. 42, no. 2, pp. 125-158, 2011.

[37] E. Gaffeo, M. Gallegati and A. Palestrini, "On the size distribution of firms: additional evidence from the G7 countries," *Physica A: Statistical Mechanics and its Applications,* vol. 324, no. 1-2, pp. 117-123, 2003.

[38] E. Cefis, O. Marsili and H. Schenk, "The effects of mergers and acquisitions on the firm size distribution," *Journal of Evolutionary Economics,* vol. 19, no. 1, pp. 1-20, 2009.

[39] E. Jondeau and M. Rockinger, "Gram-Charlier densities," *Journal of Economic Dynamics & Control,* vol. 25, no. 10, pp. 1457-1483, 2001.

[40] A. Leon, J. Mencía and E. Sentana, "Parametric properties of semi-nonparametric distributions, with applications to option valuation," *Journal of Business and Economic Statistics,* vol. 27, no. 2, pp. 176-192, 2009.