



# Documentos de trabajo

## **Economía y Finanzas**

**Nº 20-16**

2020

---

### **Firm size and economic concentration: An analysis from lognormal expansion**

Lina M. Cortés, Juan M. Lozada, Javier Perote

# Firm size and economic concentration: An analysis from lognormal expansion<sup>1</sup>

Lina M. Cortés<sup>2</sup>  
Juan M. Lozada<sup>3</sup>  
Javier Perote<sup>4</sup>

June 8, 2020

## Abstract

This paper proposes a semi-nonparametric (SNP) generalization of the lognormal distribution for studying firm size and providing accurate measures for economic concentration and inequality in terms of the Gini index adjusted to flexible functional forms. The empirical application for a sample of 1,772 Colombian companies from 2002 to 2015 shows that the log-SNP provides a better fit to firm size distribution, especially for the extreme quantiles, for which lognormal distribution overestimates economic concentration. In addition, dynamic panel model estimates indicate that firm characteristics, including size, age, and leverage, are determining factors in explaining firm growth, thus rejecting Gibrat's law.

**Keywords:** Firm growth, Firms size, Gini index, Semi-nonparametric approach.

**JEL Classification:** C14, L11, L25.

---

<sup>1</sup> Authors hold sole responsibility for the views expressed in this working paper, which may not necessarily reflect the opinion or policy of the Universidad EAFIT or the Center for Research in Economics and Finance (Cief). This study was funded by Universidad EAFIT [project 828-000019].

<sup>2</sup> 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](mailto:lcortesd@eafit.edu.co). Corresponding author. Phone: +5742619500 ext. 9756. Fax: +5743120649.

<sup>3</sup> Department of Finance, School of Economics and Finance. Universidad EAFIT. Address: Carrera 49 No 7 Sur-50 Medellin, Colombia. E-mail: [jmlozadah@eafit.edu.co](mailto:jmlozadah@eafit.edu.co)

<sup>4</sup> Department of Economics and IME, University of Salamanca. Address: Campus Miguel de Unamuno 37007 Salamanca, Spain. E-mail: [perote@usal.es](mailto:perote@usal.es)

## 1. Introduction

The relationship between firm size (FS) and firm growth (FG) has been extensively studied since the early seminal study of Gibrat (1931). The so-called Gibrat's law postulates that these two variables are not correlated, and the probability density function (PDF) of FS is stable and approximately lognormal. In an economy, many small businesses coexist with a few large companies, and Gibrat's law is used as an explanation for the high bias in FS distribution (Stanley, et al., 1995; Hart & Oulton, 1997; McCloughan, 1995; Lotti, Santarelli, & Vivarelli, 2009). Although this topic has been addressed in several studies, FS distribution is still an open question that arouses increasing interest among researchers and policy makers, since firm distribution is correlated with the degree of economic concentration and, consequently, is a cornerstone of antitrust policy (Hart & Prais, 1956; Simon & Bonini, 1958; Barba Navaretti, Castellani, & Pieri, 2014; Heinrich & Dai, 2016).

However, in the literature, there is no consensus regarding the functional model that should be adopted to analyze FS distribution. Although studies have found evidence that the lognormal distribution accurately fits to FS, favoring Gibrat's law (Voit, 2001; Kaizoji, Iyetomi, & Ikeda, 2006; Gallegati & Palestrini, 2010), other studies feature a poor performance of this distribution, especially in the higher quantiles (Gupta, Campanha, de Aguiar, Queiroz, & Raheja, 2007; Cefis, Marsili, & Schenk, 2009). In this line, some empirical studies have shown that FS distribution can be adjusted using a Pareto or Power-law distribution (Axtell, 2001; Coad, 2010; Simon & Bonini, 1958; Akhundjanov & Toda, 2019), although this latter distribution presents the shortcoming of requiring the selection of a minimum threshold to assume that FS distribution is well defined (di Giovanni, Levchenko, & Ranci re, 2011; Goddard, Liu, Donal, & Wilson, 2014; Bottazzi, Pirino, & Tamagni, 2015; Pascoal, Augusto, & Monteiro, 2016; Hart & Oulton, 1997; Cirillo & H sler, 2009).

There is also a strand of literature that argues that the discrepancies on the data fits may be due to the fact that the distributions traditionally used to accommodate fat tails usually depend on very few parameters to determine the entire shape of the FS distribution, including the right tail of the distribution (Newman, 2005; Crosato & Ganugi, 2007; Mart nez-Mekler, et al., 2009; Cort s, Mora-Valencia, & Perote, 2017). This may result in density misspecification and misleading conclusions on economic policy

recommendations, since FS dynamics is a determining factor of economic growth and stability. Note that, small changes in the way companies are distributed may have a significant macroeconomic effect, e.g., increased employment and income distribution (Segarra & Teruel, 2012; Heinrich & Dai, 2016).

Although several studies have investigated FS distribution, other research sought to understand what determines FG. In this framework, and taken Gibrat (1931) as a reference, studies have mainly focused on analyzing the effects of FS distribution on FG (Coad, 2009). However, the assumptions behind Gibrat's law remain one of the most controversial and explored topics in the studies on industrial organization because the empirical evidence shows that in some industries or economies, FG depends on FS and/or company history (Santarelli, Klomp, & Thurik, 2006; Coad & Hölzl, 2012).

This study sheds some light on this topic with an empirical study based on Colombian firms. The motivation for choosing Colombia was that the empirical literature to date has focused on characterizing FS distribution and its growth determinants in other regions, including the United States, Europe, and Asia (e.g., Stanley, et al., 1995; Hart & Oulton, 1997; Cirillo & Hüsler, 2009; di Giovanni, Levchenko, & Ranci  re, 2011; Heinrich & Dai, 2016; Canarella & Miller, 2018, among others), but research on this area for Latin American emerging economies are very limited.<sup>5</sup> Therefore, a study in a Latin American country represents a particularly relevant contribution to this literature, since companies in Latin America are characterized by a highly concentrated structure and less developed capital markets, even among emerging countries (Chong & L  pez-de-Silanes, 2007; C  spedes, Gonz  lez, & Molina, 2010). These conditions can generate potentially different results between the markets previously studied and the Latin American market.

In line with the above, this study has three primary objectives. The first is comparing the adjustment of FS distribution using the lognormal distribution (Gibrat, 1931) with the more flexible log-semi-nonparametric (log-SNP) distribution (Cort  s, Mora-Valencia, & Perote, 2017). The log-SNP distribution, which generalizes the lognormal, is derived from a logarithmic transformation of SNP distributions, which are based on Edgeworth and Gram-Charlier expansions. This transformation keeps the flexibility of the Gram-Charlier distributions' parametric structure, but constraining the domain to positive

---

<sup>5</sup>Among the studies conducted in Latin America, it is worth mentioning those by Gupta, Campanha, de Aguiar, Queiroz, & Raheja (2007) and Capelleras & Rabetino (2008).

values. The log-SNP distribution has been applied in diverse fields in which the precision in the measurement of distribution tails is crucial for accurately measuring the occurrence of extreme values. The studies by Kuhs (1988), Blinnikov & Moessner (1998), Mauleón & Perote (2000), and Cortés, Mora-Valencia, & Perote (2016) have used this distribution in the areas of thermodynamics, astronomy, finance and scientometrics, respectively.

Second, firm distribution is closely associated with the level of economic concentration (Hart & Prais, 1956; Cefis, Marsili, & Schenk, 2009). Therefore, this study proposes using the log-SNP distribution to analyze the economic concentration in a market according to the Gini index. In this respect, Hart and Prais (1956) pointed out that the Gini index could be interpreted as a measure of average dominance within a group of companies, that is, the difference in the size between two companies can provide a measure of the degree of power that one company can exert over the other. Starting from the definition proposed by Sen (1973), we propose calculating the Gini index using the log-SNP distribution and comparing its performance with the lognormal distribution.

The third objective is to analyze the determinants of FG. The validity of Gibrat's law in an emerging Latin American market was assessed by estimating the relationship between FG, FS, firm age, leverage, and profitability. The dynamic panel methodology proposed by Arellano and Bond (1991) and Blundell and Bond (1998) was used to control for the endogeneity and unobservable heterogeneity associated with this type of models.

Our results evidenced that Gibrat's law was not applicable to the Colombian economy. Compared with the lognormal distribution, the log-SNP distribution provided a better fit when modeling FS distribution. Moreover, the log-SNP distribution allowed a better adjustment in the upper quantiles without having to impose a minimum threshold, which allowed us to obtain a better quantification of the Gini index. This is relevant because knowing the characteristics of larger companies and having a larger share of the market is essential to analyze the entire economy. In addition to variables such as growth rate and the correlation between FG and FS, other characteristics are fundamental determinants of FG.

This paper is structured as follows. Section 2 contains definitions about FS distribution and a description of the log-SNP distribution. Section 3 defines the economic concentration and approaches to its quantification using the log-SNP distribution. Section 4 reviews the relevant literature on the determinants of FG and presents the hypotheses

to be analyzed. Section 5 reports the collected data and descriptive statistics on the evaluated variables. Section 6 describes the results of the comparison of the performance of lognormal and log-SNP distributions and discusses their compliance with Gibrat's law. The last section summarizes the conclusions.

## 2. Firm size distribution

Gibrat (1931) proposed that FS distribution (measured by sales or number of employees) is adequately estimated using a lognormal distribution because FG tends to be multiplicative and independent of its size at a certain point in time. Formally, let  $z_t \in \mathbb{R}^+$  be a random variable (with finite variance) that represents FS at a time  $t$ , and let  $\gamma_t$  denote its corresponding growth rate, i.e.  $z_t = \delta_t z_{t-1}$ , where  $\delta_t = 1 + \gamma_t$ . It follows that

$$z_t = \delta_t z_{t-1} = \delta_t \delta_{t-1} z_{t-2} = \dots = \delta_t \delta_{t-1} \dots \delta_1 z_0, \text{ and in logarithmic form}$$

$$\ln(z_t) = \ln(\delta_t) + \ln(\delta_{t-1}) + \dots + \ln(\delta_1) + \ln(z_0).$$

Assuming that the terms  $\ln(\delta_i)$ , with  $i = 1, \dots, t$  are independent and identically distributed, and applying the central limit theorem, it can be concluded that  $\ln(z_t) \in \mathbb{R}$  approximately follows a normal distribution and thus  $z_t$  is lognormal distributed (Santarelli, Klomp, & Thurik, 2006; Pascoal, Augusto, & Monteiro, 2016).

Therefore, a strand of empirical literature has been devoted to the evaluation of the performance of lognormal distribution using cross-sectional data on FS (e.g., Kaizoji, Iyetomi, & Ikeda (2006); Gupta, Campanha, de Aguiar, Queiroz, & Raheja (2007), Gallegati & Palestini (2010), among others). As a result, the empirical evidence does not support the lognormal hypothesis since this distribution seems to either underestimate or overestimate the theoretically expected values in the upper quantile of FS distribution (Stanley, et al., 1995; Hart & Oulton, 1997; Cortés, Mora-Valencia, & Perote, 2017).

Given the poor performance of lognormal for fitting FS distribution, Cortés, Mora-Valencia, & Perote (2017) proposed modeling this variable using the log-SNP distribution. The latter is an extension of the former that allows the estimation of a general family of density specifications by introducing additional parameters.

Let  $z_i$  be the variable that measures FS at a specific time, then, it is said to be log-SNP distributed if its PDF can be expressed as

$$h(z_i; \mu, \sigma^2, \mathbf{d}) = \left( \frac{1}{z_i \sigma \sqrt{2\pi}} e^{-\frac{(\ln(z_i) - \mu)^2}{2\sigma^2}} \right) \left[ 1 + \sum_{s=1}^n d_s H_s \left( \frac{\ln(z_i) - \mu}{\sigma} \right) \right], z_i \in \mathbb{R}^+, \quad (1)$$

where  $\mu \in \mathbb{R}$  and  $\sigma^2 \in \mathbb{R}^+$  represent the location and scale, respectively, and  $\mathbf{d} = (d_1, \dots, d_n)' \in \mathbb{R}^n$  are shape parameters. Note that the lognormal distribution is a particular case when  $\mathbf{d} = 0$ . Consequently, as well as the lognormal corresponds to an exponential transformation of the normal, the log-SNP is the exponential transformation of a variable with SNP distribution (also known as Gram-Charlier Type A). That is,  $z_i = \exp(x_i)$  if  $x_i$  has an SNP distribution.

The PDF of a SNP random variable  $x_i$  is a general class of densities of the type:

$$f(x_i; \mathbf{d}) = [1 + \sum_{s=1}^n d_s H_s(x_i)] \phi(x_i), x_i \in \mathbb{R}, \quad (2)$$

where  $\phi(x_i) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x_i^2}$  is the standard normal PDF and  $H_s(x_i)$  is the Hermite polynomial (HP) of order  $s$ , which is defined as the  $s$ -th order derivative of  $\phi(x_i)$ ,

$$\frac{d^s \phi(x_i)}{dx_i^s} = (-1)^s H_s(x_i) \phi(x_i), \quad (3)$$

e.g., the first four HPs are  $H_0(x_i) = 1$ ,  $H_1(x_i) = x_i$ ,  $H_2(x_i) = x_i^2 - 1$ ,  $H_3(x_i) = x_i^3 - 3x_i$ , and  $H_4(x_i) = x_i^4 - 6x_i^2 + 3$ .<sup>6</sup>

These polynomials form an orthonormal basis and therefore satisfy the following orthogonality property,

$$\int_{-\infty}^{\infty} H_s(x_i) H_j(x_i) \phi(x_i) dx_i = 0 \quad \forall s \neq j, \quad (4)$$

which is the ground for interesting results as the fact that the expansion integrates to one or that the even (odd)  $k$ -order moment only depends on  $d_s$ , for  $s \leq k$  and  $s$  being even (odd) parameters, e.g.  $d_1$  and  $d_2$  account for mean and variance,  $d_3$  and  $d_4$  incorporate bias and excess kurtosis (provided that  $d_1 = d_2 = 0$ ), respectively, and the remaining parameters represent higher-order moments. It is clear that the parameter flexibility of the SNP density represents a major advantage compared to other traditional densities that depend on a limited number of parameters. Furthermore, by adding more parameters data

---

<sup>6</sup>It is worth noting that given a truncation order, the resulting distribution is purely parametric; however, truncation order may vary to allow a more accurate approximation to a given distribution. Conveniently, it is assumed  $d_0 = 1$ , so as the density integrates to one.

fits may be improved, since the asymptotic approximation theoretically captures any distribution that meets the regularity conditions of the Gram-Charlier Type A series<sup>7</sup>.

### 3. Firm size and economic concentration

Economic concentration is a relevant parameter in the characterization of a market or industry. There is a tendency to establish the structure of a market or industry based on company number and size because firm distribution is closely related to the level of economic concentration (Hart & Prais, 1956; Cefis, Marsili, & Schenk, 2009).

In this respect, the Gini index provides an average measure of dominance within a group of companies, i.e., this measure can be used to compare the evolution in FS distribution with the evolution of economic concentration (Carree & Thurik, 1991; Crosato & Ganugi, 2007; Guo, Xu, Chen, & Wang, 2013). Since the Gini index is based on the Lorenz curve, several models of that curve have been developed in the economic literature (e.g., Sarabia, Castillo, & Slottje, 1999; Ogwang & Rao, 2000; Crosato & Ganugi, 2007; Wang & You, 2016, among others).

The Lorenz curve (Lorenz, 1905)<sup>8</sup> is a function that represents the cumulative proportion  $X$  of ordered individuals (from lowest to highest) in cumulative size distribution  $Y$ . For example,  $X$  can represent income or wealth, citations, votes and city population, among other variables. The standard definition of the Lorenz curve starts by determining a particular quantile by solving the equation:

$$p = F(z_i) = \int_0^{z_i} f(t_i) dt_i, \quad (5)$$

and then

$$L(p) = \frac{1}{\eta} \int_0^{z_i} t_i f(t_i) dt_i, \quad (6)$$

where  $\eta$  corresponds to the mean distribution,

$$\eta = \int_0^{\infty} t_i f(t_i) dt_i. \quad (7)$$

---

<sup>7</sup>It is noteworthy that for finite expansions of Gram-Charlier series non-negativity is not guaranteed for all  $\mathbf{d} \in \mathbb{R}^n$ , which requires implementing positive transformations (Leon, Mencía, & Sentana, 2009) or positivity restrictions (Jondeau & Rockinger, 2001). However, the maximum likelihood estimation algorithms tend to converge to values that guarantee a well-defined PDF. This study considers the original SNP expansion in equation (2), which is more usual in economics and finance applications.

<sup>8</sup>Further details can be found in Aitchison and Brown (1957).



Taking the notation proposed by Gastwirth (1971) and given the transformation  $z_i = F^{-1}(p)$ , the Lorenz curve can be expressed as

$$L(p) = \frac{1}{\eta} \int_0^p F^{-1}(t_i) dt_i, 0 \leq p \leq 1, \quad (8)$$

with  $L(0) = 0$ ,  $L(1) = 1$ ,  $L'(p) \geq 0$ ,  $L''(p) \geq 0$  in  $(0,1)$ .

Furthermore, the Gini index, which ranges from zero to one, where zero corresponds to total equality (all individuals in a population have the same income) and one corresponds to total inequality (one individual accumulates the total income), was obtained with the equation

$$Gini = 1 - 2 \int_0^1 L(p) dp. \quad (9)$$

According to Sen (1973), in an empirical sample  $\{z_1, \dots, z_n\}$ , the Gini index can be estimated using the discrete equation

$$\widehat{Gini} = \frac{1}{n} \left[ n + 1 - 2 \frac{\sum_{i=1}^n (n+1-i) z_{(i)}}{\sum_{i=1}^n z_i^*} \right], \quad (10)$$

where  $z_{(i)}$  is the  $i$ -th order statistic.

According to Gibrat's law, the PDF of an empirical sample can be fitted using the lognormal distribution, which assumes the cumulative distribution function (CDF)

$$F(z_i; \mu, \sigma^2) = \Phi\left(\frac{\ln(z_i) - \mu}{\sigma}\right) = \int_0^{z_i} \frac{1}{\sigma t_i \sqrt{2\pi}} e^{-\frac{(\ln(t_i) - \mu)^2}{2\sigma^2}} dt_i. \quad (11)$$

However, some authors proposed using non-parametric or semi-nonparametric distributions to fit the empirical sample and estimate the Gini index described in equation (10) (Hasegawa & Kozumi, 2003; Cowell & Flachaire, 2007; Cowell & Victoria-Feser, 2008; Zhang, Wu, & Li, 2016). Considering that many factors may affect the degree of economic concentration, it can be difficult to summarize the characterization of FS distribution using a few parameters. For instance, when FS distribution is widely dispersed around the mean, and larger companies are relatively large, it may be more challenging to determine extreme values with traditional parametric distributions (Hart & Prais, 1956).

For this purpose, the log-SNP provides an accurate performance of the empirical distribution, especially in the upper quantiles, which have extreme values (Cortés, Mora-

Valencia, & Perote, 2017). Given the properties of the SNP distribution discussed in Section 2, a closed form of the CDF of the log-SNP distribution can be obtained, which results to be very useful for computing the probabilities and quantiles of this distribution<sup>9</sup>.

The CDF of the log-SNP distribution is defined as:

$$\begin{aligned} F(z_i; \mu, \sigma^2, \mathbf{d}) &= \int_0^{z_i} \left( \frac{1}{\sigma t_i \sqrt{2\pi}} e^{-\frac{(\ln(t_i) - \mu)^2}{2\sigma^2}} \right) \left[ 1 + \sum_{s=1}^n d_s H_s \left( \frac{\ln(z_i) - \mu}{\sigma} \right) \right] dt_i, \\ &= \Phi \left( \frac{\ln(z_i) - \mu}{\sigma} \right) - \phi \left( \frac{\ln(z_i) - \mu}{\sigma} \right) \sum_{s=1}^n d_s H_{s-1} \left( \frac{\ln(z_i) - \mu}{\sigma} \right). \end{aligned} \quad (12)$$

This large number of parameters does not result in higher computational difficulty and can be obtained by maximum likelihood (ML), whose log-likelihood (logL) function is given by:

$$\log L(z_i; \mu, \sigma^2, \mathbf{d}) = -\frac{1}{2} \log(2\pi\sigma^2 z_i^2) - \frac{1}{2} \left( \frac{\log(z_i) - \mu}{\sigma} \right)^2 + \log \left[ 1 + \sum_{s=1}^n d_s H_s \left( \frac{\log(z_i) - \mu}{\sigma} \right) \right]. \quad (13)$$

For computational purposes sequential estimation is recommended, i.e. beginning with the simplest density, the lognormal, and adding  $d_s$  parameters recursively (estimates of the previous step are used as initial values for the next one). As models in every stage are nested, the final expansion can be chosen according to accuracy criteria, e.g. logL or Akaike Information Criteria (AIC), and standard tests, as the likelihood ratio (LR). The quantiles of the log-SNP distribution are directly obtained using the CDF presented in equation (12) and the inverse transform method.<sup>10</sup>

#### 4. Firm size and its determinants

This section reviews the relevant literature on the determinants of FG and discuss some conjectures underlying the role of firm characteristics on explaining FG using Gibrat's law. As a by-product, we establish a model to empirically evaluate Gibrat's law compliance.

##### 4.1. Theoretical background and hypothesis formulation

---

<sup>9</sup> Proofs and discussion on the properties of the SNP distribution and log-SNP distribution are derived in Ñíguez, Paya, Peel, & Perote (2012), Ñíguez, Paya, Peel, & Perote (2013), and Cortés, Mora-Valencia, & Perote (2017).

<sup>10</sup>The R code for performing the maximum likelihood estimation algorithm and quantile computation is available upon request.

On the grounds of Gibrat's (1931) seminal paper, several authors have investigated the relationship between FS and FG (Simon & Bonini, 1958; Mansfield, 1962; Lotti, Santarelli, & Vivarelli, 2009; Tang, 2014). According to Gibrat's law, FG rates do not depend on the FS and/or company history. That is, the distribution of FG rates in an economy is identical for all companies, regardless of their current size and/or previous growth history (Coad, 2009).

However, some studies have questioned the validity of Gibrat's law (Evans, 1987; Machado & Mata, 2000; Santarelli, Klomp, & Thurik, 2006; Coad & Hölzl, 2012; Meisenzahl, 2016; Canarella & Miller, 2018). Among them, there are several opinions on the determinants of FG (Delmar, Davidsson, & Gartner, 2003) because the growth patterns may depend on different factors, which were corroborated in previous theoretical and empirical studies. For instance, in addition to FS, other variables may affect firm dynamics and evolution (Becchetti & Trovato, 2002; Moreno & Casillas, 2007; Angelini & Generale, 2008; Canarella & Miller, 2018).

Gibrat's law can be tested using three different approaches: (i) considering all the companies within an industry or a specific economy and time interval, including the companies that did not survive; (ii) considering only surviving companies; (iii) considering companies large enough to reach the minimum efficiency scale (Mansfield, 1962). However, the available studies have focused mainly on the second approach. In this respect, it is necessary to correct heteroscedasticity and serial correlation when analyzing the determinants of FG in a sample of surviving companies because, if the study is based only on surviving companies, it is very likely that sample selection is strongly correlated with the same variables that may affect FG (Becchetti & Trovato, 2002; Angelini & Generale, 2008; Canarella & Miller, 2018).

In this respect, to confirm the validity of Gibrat's law and the impact of other variables on company growth, several studies have focused on dynamic econometric models (e.g., Dunne & Hughes, 1994; Oliveira & Fortunato, 2006; Angelini & Generale, 2008; Huynh & Petrunia, 2010; Barba Navaretti, Castellani, & Pieri, 2014; Canarella & Miller, 2018, among others). In the present research, in addition to evaluating the relationship between company growth and size, other determinants were considered, including firm age, leverage, and profitability.

When analyzing FG, Gibrat's law assumes the absence of autocorrelation in errors or non-persistence of the growth rate. However, previous studies using dynamic econometric models provided evidence of growth rates persistence. However, the magnitude and direction of this effect are not entirely clear. For instance, some studies found that the growth rate in a specific period was positively correlated with its first lag in growth ( $t - 1$ ) (Bottazzi & Secchi, 2003; Angelini & Generale, 2008; Fotopoulos & Giotopoulos, 2010; Canarella & Miller, 2018; Liñares-Zegarra & Wilson, 2018). Other studies reported that negative persistence values indicated that small companies developed at a rate higher than that of their larger counterparts (Evans, 1987; Oliveira & Fortunato, 2006; Lotti, Santarelli, & Vivarelli, 2009; Huynh & Petrunia, 2010; Liñares-Zegarra & Wilson, 2018). According to the above, in the present study, FG is expected to be persistent, which would be evidence against Gibrat's law (Hypothesis 1).

On the other hand, Gibrat's law postulates the lack of correlation between FG and FS. However, empirical studies point to the opposite result (Hart & Oulton, 1996; Cabral & Mata, 2003; Canarella & Miller, 2018). Firm size can be measured using different parameters, including sales, assets, employees, and benefits, among others (Delmar, Davidsson, & Gartner, 2003; Zhang, Chen, & Wang, 2009; Heinrich & Dai, 2016). The number of employees, assets, and sales are the most frequently used. However, each of these measures has advantages and disadvantages. The number of employees is a discrete variable that may not reflect the increase in employee productivity (Tang, 2014). The level of assets, in contrast to the number of employees and level of sales, can assume negative values (Hart & Oulton, 1996). Therefore, previous studies suggest that the level of sales may better represent FS (Pascoal, Augusto, & Monteiro, 2016; Cortés, Mora-Valencia, & Perote, 2017). Based on the hypothesis that small businesses seek growth to achieve a minimum efficient size (Dunne & Hughes, 1994; Becchetti & Trovato, 2002; Barba Navaretti, Castellani, & Pieri, 2014), a negative relationship between FS and FG is expected (Hypothesis 2).

In addition, empirical studies found that FG might be affected by age (Cabral & Mata, 2003; Angelini & Generale, 2008; Meisenzahl, 2016). In this respect, Evans (1987), Reid, & Xu (2012) and Barba Navaretti, Castellani, & Pieri (2014) found a negative relationship between firm age and growth, that is, young companies developed faster than their older counterparts. In contrast, Das (1995) and Shanmugam & Bhaduri (2002) show a positive relationship between FG and firm age. According to Das (1995), the positive effect may

be because over the years, consumers become more aware of the existence of a product or service, which increases their consumption and thus result in greater growth in the firm. Furthermore, the firm's reputation can improve with age and this can be reflected in a positive impact. Also, some authors have evaluated the presence of non-linear relationships. Thereon, Park, Shin, & Kim (2010) found a concave relationship between FG and firm age, suggesting that FG decreased more rapidly as companies aged. A positive correlation between firm age and growth is expected in this study (Hypothesis 3a); however, the effect of age on growth is expected to decrease as companies age (Hypothesis 3b).

Studies on FG used leverage as a control variable (Huynh & Petrunia, 2010; Barba Navaretti, Castellani, & Pieri, 2014). Theoretically, leverage generates benefits (Modigliani & Miller, 1963) and costs (e.g., financial difficulties and agency costs) (Jensen, 1986) which may have variable effects on growth. The studies by Jang & Park (2011) and Canarella & Miller (2018) found a negative relationship between the level of leverage and FG rate. This result is because companies lose financial flexibility as they become more indebted, which may lead to the rejection of projects with a positive net present value in inefficient markets, and consequently less growth. In contrast, Huynh & Petrunia (2010) and Barba Navaretti, Castellani, & Pieri (2014) found a positive association between the level of leverage and FG. The reason is because debt is a mechanism of control used by shareholders over managers. If a company has debts, the manager should be more efficient and pay debts by avoiding waste and poor investments. In addition, a positive relationship can be explained by companies' desire to avoid raising capital and the consequent loss of control (Céspedes, González, & Molina, 2010). In this respect, a positive relationship between the level of leverage and FG is expected in this study (Hypothesis 4).

Finally, and according to the pecking order theory, companies initially prefer to finance investment projects by reinvesting profits because the asymmetry of market information can make other sources of financing more expensive (Myers, 1984). In this respect, it is expected that companies with higher profitability can make investments with lower costs and therefore, grow more. Jang & Park (2011) and Canarella & Miller (2018) found empirical evidence that supports a positive link between profitability and FG. In contrast, Heshmati (2001) and Liñares-Zegarra & Wilson (2018) found that there was no

significant relationship between profitability and FG. A positive effect of profitability on FG is expected in this study (hypothesis 5).

## 4.2 Econometric modeling

In the context of Gibrat's law, the canonical specification for FG is:

$$Growth_{it} = \alpha_i + \beta Growth_{i,t-1} + \gamma \log(Sales_{t-1}) + \varepsilon_{it}, \quad (14)$$

where  $Growth_{it}$  is FG calculated as the first logarithmic difference of sales,  $Growth_{i,t-1}$  is the first lag of FG and  $\log(Sales_{t-1})$  is a proxy of FS measured as the natural logarithm of sales, all for a specific firm  $i$  and time  $t$ .<sup>11</sup> Furthermore,  $\alpha_i$  and  $\varepsilon_{it}$  correspond to the unobserved fixed effect of the company and the error term (which holds the standard assumptions of panel data models), respectively.

However, Huynh & Petrunia (2010) argue that age affects the differences in financial capital between companies. Therefore, their model suggests that firm age and leverage are relevant when studying the dynamics of FS. Following to the theory presented in Section 4.1, we also considered profitability.

In accordance with the above, the specification proposed in equation (14) is extended as follows:

$$Growth_{it} = \alpha_i + \beta Growth_{i,t-1} + \gamma \log(Sales_{t-1}) + \theta_1 \log Age_{it} + \theta_2 [\log Age_{it}]^2 + \varphi Leverage_{i,t-1} + \omega ROE_{i,t-1} + \varepsilon_{it}, \quad (15)$$

where  $\log Age_{it}$  is the logarithm of the age of the company since its foundation, which is considered in both level and quadratic form;  $Leverage_{i,t-1}$  is the first lag of leverage calculated as the sum of the long-term debt and short-term debt divided by the total assets; and  $ROE_{t-1}$  is the first lag of profits, calculated as the net profit divided by common equity.

Given that equations (14) and (15) correspond to a dynamic panel model, the literature proposes using the generalized method of moments (GMM) estimator developed by Arellano & Bond (1991). This estimator uses the lagged differences in the levels of the same variable as instruments of the endogenous variable and induces first order, but not second-order, correlation in the estimated model. However, the GMM difference

---

<sup>11</sup> Sales is the value of net sales in Colombian peso (COP) deflated by Colombian Consumer Price Index (CPI) using 2008 price base (see <https://totoro.banrep.gov.co/analytics/saw.dll?Portal>).

estimator may produce weak instruments if the parameter of interest is close to one, which results in biased and inconsistent finite sample properties. Blundell & Blond (1998) proposed using the system GMM estimator to address the problem. The system estimator uses the lagged differences in endogenous variables, in addition to the variables used in the original estimator. Consequently, system GMM presents a superior performance in finite samples than the difference estimator.

## **5. Data description and statistics**

This study analyzes a sample of Colombian companies from 2002 to 2015. The primary sources of information were reports of financial statements, annexes, and basic information that companies send annually to the Superintendence of Companies of Colombia.<sup>12</sup> This source reports valuable information at the firm level but also has several limitations. Despite the legal provisions that oblige companies in Colombia to present financial statements annually, data from some companies in the database of the Superintendence are available for specific time periods but not for others, which limits the control for inclusion and exclusion criteria. Since the EMIS Benchmark was used to verify data consistency, the sample restricted to 1,772 surviving companies from all economic sectors.<sup>13</sup> However, there were no restrictions on the minimum level of net sales or total assets. The sample included large and small companies, in contrast to other studies on FG, which focused on either large or small companies.

**[Table 1 here]**

Some descriptive statistics for the entire sample period (2002–2015), particularly the first four moments, for each variable are summarized in Table 1.<sup>14</sup> The third and fourth central moments provided useful information about firm distribution shape, and the means and standard deviations are also provided. The variable sales, which in this case was related to FS, featured positive asymmetry, with a very high number of small businesses. Positive

---

<sup>12</sup>Consult the website [www.supersociedades.gov.co](http://www.supersociedades.gov.co). The datasets analyzed during the current study are available from the corresponding author on reasonable request.

<sup>13</sup>Note that this study does not control for the inclusion and exclusion of companies and mergers and acquisitions (M&As). In this respect, Cefis, Marsili, & Schenk (2009) showed that M&As did not affect firm size distribution when all companies were included, and this result may be because the balance generated by the inclusion and exclusion of companies counteracted the effect of M&As. In contrast, Audretsch, Klomp, Santarelli, & Thurik (2004) discussed this topic in more detail and suggested that the selection bias of the sample was not observed in short periods and, in case it was observed, the bias might be difficult to be econometrically determined because companies could be excluded for many reasons.

<sup>14</sup>The Superintendence of Companies of Colombia reports the accounting information in local currency (peso).

kurtosis also indicated that the upper quantile of the distribution was larger than that of a lognormal distribution. On average, companies were at a mature age. However, there was high variability in growth rate. On average, Colombian companies had a high level of leverage and high variability in profits.

**[Figure 1 here]**

The graph of the density of the logarithm of sales resulting from smoothing of the corresponding histogram is presented in Figure 1. The years 2002, 2009, and 2015 were selected at random to visualize the densities better. The picture shows density dynamics, illustrating its deviations from the lognormal distribution over time. Long-term FS distribution becomes more dispersed near the mean, more biased toward small firms, and larger in the higher quantiles. The empirical evidence from Kernel density estimation indicated that the shape of FS distribution was different from that of the lognormal distribution (Fig. 1a). Furthermore, the tail (Fig. 1b) featured multimodality or jumps, as observed by Marsili (2006), Bottazzi, Cefis, Dosi, & Secchi (2007), and Cortés, Mora-Valencia, & Perote (2017) (Fig. 1b).

## **6. Results and discussion**

This section presents and discusses the results of FS distribution, economic concentration, and the determinants of FG.

### **6.1. Modeling FS distribution**

Table 2 reports the ML estimates obtained from equation (13) for lognormal distribution (Panel A) and log-SNP distribution (Panel B). The results indicate that both models adequately determined the mean and standard deviation of the sample of selected companies. These statistics are represented by the location ( $\mu$ ) and scale ( $\sigma$ ) parameters, respectively. The p-values indicate that these parameters are highly significant for both distributions. However, the parameters  $d_s$  were also highly significant for most of the evaluated years in the log-SNP distribution (Panel B).

**[Table 2 here]**

The analysis of the AIC statistic, which penalizes the inclusion of additional parameters in the two distributions, indicates that this criterion is consistently lower in the log-SNP distribution, suggesting that the model for this distribution provides a better performance.



According to the Kolmogorov-Smirnov (KS) test, neither the lognormal nor the log-SNP can be rejected as the data generating process at a 1% significance level and for most of the years. However, the LR statistic for the difference between the log-SNP and lognormal distribution, shown in panel C, presents strong evidence in favor the log-SNP specification. The results of this test confirm the fact that the incorporation of the parameters  $d_s$  is significant and leads to the log-SNP model outperformance. This means that FS distribution presents significant asymmetries (captured by parameter  $d_3$ ) and non-monotonic thick tails (captured by parameter  $d_4$ ), due to the presence of extreme values, which definitely cannot be represented by the lognormal distribution.

**[Figure 2 here]**

The relationship between rank and sales (in logarithmic scale) for the years 2002, 2009, and 2015 is shown in Figure 2. The comparison of empirical values (hollow points) and those estimated using a lognormal distribution (dashed line) and log-SNP distribution (solid line) reveals that the log-SNP captured more adequately the empirical distribution. The lognormal distribution, however, tended to systematically overestimate the values, as previously reported for other regions (Stanley, et al., 1995; Hart & Oulton, 1997; Cortés, Mora-Valencia, & Perote, 2017).<sup>15</sup>

## **6.2. Analysis of economic concentration**

Under Gibrat's law, the Gini index presented in equation (10) should be calculated using the values predicted theoretically by the CDF of the lognormal distribution described in equation (11). However, there is still controversy regarding the distribution function that best represents the upper quantiles, especially for extreme values (Wang & You, 2016). In this respect, Hart & Prais (1956) reported that, in the case of lognormal distribution, changes in the parameter of the scale  $\sigma$  were positively correlated with changes in the level of economic concentration. However, these changes may be the result of different factors that may affect the degree of competition and, in that case, it may be difficult to summarize the changes using a single parameter.

**[Table 3 here]**

---

<sup>15</sup>The quantiles of the log-SNP distribution were obtained using the CDF presented in equation (12) and the inverse transform method (ITM). The lognormal is a special case where  $\mathbf{d}=\mathbf{0}$ .

In this respect, the present study used the log-SNP distribution to analyze the economic concentration, measured from sales in the sample of the selected companies. Our hypothesis is that, as explained in Section 3, the flexible parametric structure of the log-SNP distribution may allow a better adjustment of the expected values in the presence of extreme values. The sales, in millions of Colombian pesos, obtained empirically for the sample of 1,772 Colombian companies versus the values expected theoretically using a lognormal distribution and log-SNP distribution are shown in Table 3. The analysis of the trend of the upper quantile of the distribution of sales at a confidence level of 10%, 5%, and 1% indicated the errors in the estimation of FS distribution using a lognormal distribution, possibly leading to an inadequate measurement of the level of economic concentration.

**[Figure 3 here]**

Equation (10) was employed to measure the Gini index for the level of sales of each company in the sample. The dynamics of the values of this index measured using empirical data and data adjusted theoretically for both distributions is shown in Figure 3. The lognormal distribution tended to overestimate the level of economic concentration, which is consistent with the results presented in Table 3. Moreover, these results are reinforced by those of the KS test for the empirical Gini index and each distribution. For the lognormal distribution (log-SNP), the p-value was 0.002 (0.987) using the KS test, indicating that this distribution was not adequate (null hypothesis could not be rejected) at the usual confidence levels. Consequently, the log-SNP distribution allowed obtaining a better quantification of the economic concentration.

### **6.3 Determinants of firms' growth and the evidence on Gibrat's law**

The results of the system GMM estimator for three dynamic panel models and the statistical tests for analyzing the estimations provided by the models are shown in Table 4. First, the validity of the instruments was assessed using the Hansen test. This test allowed the detection of the overidentification of the model when the heteroscedastic weight matrix was used in the estimation and, therefore, it was appropriate for analyzing the two-step estimates of the table. In the three estimated models, all explanatory variables were considered endogenous (except for age) and were instrumented. The results supported the validity of the instruments used.

Second, to achieve consistent estimation of the system GMM, which uses lagged differences or levels as instruments, correlation analysis of the residuals is performed by the Arellano and Bond test. A first-order serial correlation was expected in these models because the residuals in the first differences should be correlated by construction. However, the validity of these models was confirmed only in cases in which a second-order serial correlation was not found. This condition was met by adding a second lag of the endogenous dependent variable in the models (Huynh & Petrunia, 2010).

The three estimated models use FG as the dependent variable. Model 1 included the lagged growth and FS as explanatory variables, Model 2 included age and leverage, and Model 3 included profitability. The lagged growth variables were significant confirming the dynamic nature and the persistence of FG, which provided evidence against Gibrat's law and confirmed Hypothesis 1 (Evans, 1987; Oliveira & Fortunato, 2006; Lotti, Santarelli, & Vivarelli, 2009; Huynh & Petrunia, 2010; Liñares-Zegarra & Wilson, 2018). The results show a negative impact of past growth on the contemporary one.

**[Table 4 here]**

Similarly, all three models showed evidence of a correlation between FG and FS. The estimated coefficient was negative and significant, corroborating Hypothesis 2. This is, small businesses seek high growth rates to achieve a minimum efficient size (Dunne & Hughes, 1994; Becchetti & Trovato, 2002; Barba Navaretti, Castellani, & Pieri, 2014).

Models 2 and 3 provided evidence on the effect of firm age on growth. There was a positive and significant ( $p < 0.1$ ) linear relationship between these two variables, confirming Hypothesis 3a. This result is like that reported by Das (1995) and Shanmugam & Bhaduri (2002) for a developing economy. Furthermore, the effect of age in its quadratic form indicates that FG is lower as the surviving companies age, which corroborates Hypothesis 3b.

On the other hand, studies have shown that Latin American companies exhibit a higher-than-expected leverage because economic concentration is significantly higher than that in developed countries (Céspedes, González, & Molina, 2010; Chong & López-de-Silanes, 2007). In this respect, leverage plays an essential role as a determinant of FG. Models 2 and 3 indicated that this variable had a positive and significant coefficient as reported by Huynh & Petrunia (2010) and Barba Navaretti, Castellani, & Pieri (2014), and this result confirms Hypothesis 4.

Furthermore, Model 3 proposes the analysis of profitability as a determinant of growth. As a result, a negative and statistically non-significant coefficient was obtained, with which we cannot provide conclusions about Hypothesis 5. However, this result may provide evidence of non-compliance with the pecking-order theory in Latin American companies, as observed in previous studies in the region (Céspedes, González, & Molina, 2010). This result is in line with Hypothesis 4.

## **7. Conclusions**

This paper sheds light on the compliance of Gibrat's law using a sample of 1,772 Colombian companies collected between 2002 and 2015 and comparing the performance of FS distribution using the lognormal distribution (Gibrat, 1931) and log-SNP distribution (Cortés, Mora-Valencia, & Perote, 2017). The latter distribution nests the lognormal distribution and includes new parameters that can better assess the characteristics of the upper and lower quantiles corresponding to larger and smaller companies. The results indicate that the lognormal distribution tends to systematically overestimate the expected values in the distribution tails but the log-SNP becomes a flexible method to fit them more accurately.

This finding emphasizes the need to propose other methodologies to obtain more reliable information on the level of economic concentration. In this line, we demonstrate analytically that the Gini index has a better result if it is fitted with SNP methods formulated in terms of the log-SNP distribution. In fact, the lognormal distribution tends to overestimate the level of economic concentration. This is because the log-SNP distribution is more flexible than the lognormal distribution when the data are skewed, and there are possible jumps in the tails due to outliers.

Furthermore, to test the validity of Gibrat's law and investigate on the determinants of FG, we estimated the relationship between this variable and FS, as well as other potentially explanatory variables: age, leverage, and profitability. Based on the system GMM estimator proposed by Blundell & Bond (1998), we conclude that Gibrat's law does not apply to the selected sample in Colombia. The FG rates strongly depended on the FS and presents a significant persistence over time. We also find that some company characteristics were fundamental determinants of FG, particularly firm age and leverage had a significant impact on growth. There was no evidence of a positive correlation

between profits and FG, which can be explained by the high level of economic concentration in Latin American firms and by their focus on leverage.

These results represent a valuable contribution, not only for researchers on Industrial Organization, but also for policy makers, since the knowledge about FS distribution and their determinants of growth, help to forecast industrial concentration and its impact on economic cycles and, consequently, implement adequate antitrust and economic policies. However, there are still various unsolved problems that should be considered in future research, e.g. addressing some limitations of the data coming from Latin American institutions and the extension of the analysis at the sectoral level. The degree of heterogeneity of the results for different sectors could provide a richer economic structure that could be hidden by the aggregated analysis. In addition to this, Gibrat's law can be tested considering all the companies within an industry or a specific economy and time interval, including the companies that did not survive.

### **Compliance with Ethical Standards**

**Funding:** This study was funded by the Spanish Ministry of Economics and Competitiveness [grant number ECO2016-75631-P] and Universidad EAFIT [project 828-000019].

**Conflict of interest:** The authors declare that they have no conflict of interest.

## References

- Aitchison, J., & Brown, J. (1957). *The Lognormal Distribution*. Cambridge: Cambridge University Press.
- Akhundjanov, S. B., & Toda, A. A. (2019). Is Gibrat's "Economic Inequality" lognormal? *Empirical Economics*, <https://doi.org/10.1007/s00181-019-01719-z>.
- Angelini, P., & Generale, A. (2008). On the evolution of firm size distributions. *American Economic Review*, *98*(1), 426-438.
- Arellano, M., & Bond, S. (1991). Some tests of specification for panel data: Monte Carlo evidence and an application to employment equations. *The Review of Economic Studies*, *58*(2), 277-297.
- Audretsch, D. B., Klomp, L., Santarelli, E., & Thurik, A. R. (2004). Gibrat's law: are the services different?. *Review of Industrial Organization*, *24*(3), 301-324.
- Axtell, R. (2001). Zipf distribution of U.S. Firm sizes. *Science*, *293*, 1818-1820.
- Barba Navaretti, G., Castellani, D., & Pieri, F. (2014). Age and firm growth: evidence from three European countries. *Small Business Economics*, *43*(4), 823-837.
- Becchetti, L., & Trovato, G. (2002). The determinants of growth for small and medium sized firms. The role of the availability of external finance. *Small Business Economics*, *19*(4), 291-306.
- Blinnikov, S., & Moessner, R. (1998). Expansions for nearly Gaussian distributions. *Astronomy and astrophysics Supplement Series*, *130*(1), 193-205.
- Blundell, R., & Bond, S. (1998). Initial conditions and moment restrictions in dynamic panel data models. *Journal of Econometrics*, *87*(1), 115-143.
- Bottazzi, G., & Secchi, A. (2003). Common properties and sectoral specificities in the dynamics of U.S. manufacturing companies. *Review of Industrial Organization*, *23*(3/4), 217-232.
- Bottazzi, G., Cefis, E., Dosi, G., & Secchi, A. (2007). Invariances and diversities in the patterns of industrial evolution: some evidence from Italian manufacturing industries. *Small Business Economics*, *29*(1), 137-159.
- Bottazzi, G., Pirino, D., & Tamagni, F. (2015). Zipf law and the firm size distribution: a critical discussion of popular estimators. *Journal of Evolutionary Economics*, *25*(3), 585-610.
- Cabral, L. M., & Mata, J. (2003). On the evolution of the firm size distribution: facts and theory. *American Economic Review*, *93*(4), 1075-1090.
- Canarella, G., & Miller, S. M. (2018). The determinants of growth in the U.S. information and communication technology (ICT) industry: A firm-level analysis. *Economic Modelling*, *70*, 259-271.
- Capelleras, J.-L., & Rabetino, R. (2008). Individual, organizational and environmental determinants of new firm employment growth: evidence from Latin America. *International Entrepreneurship and Management Journal*, *4*(1), 79-99.
- Carree, M., & Thurik, R. (1991). Recent developments in the Dutch firm-size distribution. *Small Business Economics*, *3*(4), 261-268.
- Cefis, E., Marsili, O., & Schenk, H. (2009). The effects of mergers and acquisitions on the firm size distribution. *Journal of Evolutionary Economics*, *19*(1), 1-20.
- Céspedes, J., González, M., & Molina, C. A. (2010). Ownership and capital structure in Latin America. *Journal of Business Research*, *63*(3), 248-254.

- Chong, A., & López-de-Silanes, F. (2007). *Investor Protection and Corporate Governance: Firm-level Evidence Across Latin America*. Palo Alto, CA: Stanford University Press and World Bank.
- Cirillo, P., & Hüsler, J. (2009). On the upper tail of Italian firms' size distribution. *Physica A: Statistical Mechanics and its Applications*, 388(8), 1546-1554.
- Coad, A. (2007). Testing the principle of 'growth of the fitter': the relationship between profits and firm growth. *Structural Change and Economic Dynamics*, 18(3), 370-386.
- Coad, A. (2009). *The growth of firms*. Cheltenham, UK: Edward Elgar.
- Coad, A. (2010). The exponential age distribution and the Pareto firm size distribution. *Journal of Industry, Competition and Trade*, 10(3), 389-395.
- Coad, A., & Hözl, W. (2012). Firm Growth: Empirical Analysis. En M. Dietrich, & J. Krafft, *Handbook on the Economics and Theory of the Firm*. Cheltenham, UK: Edward Elgar Publishing.
- Cortés, L. M., Mora-Valencia, A., & Perote, J. (2016). The productivity of top researchers: a semi-nonparametric approach. *Scientometrics*, 109(2), 891-915.
- Cortés, L. M., Mora-Valencia, A., & Perote, J. (2017). Measuring firm size distribution with semi-nonparametric densities. *Physica A: Statistical Mechanics And Its Applications*, 485, 35-47.
- Cowell, F. A., & Flachaire, E. (2007). Income distribution and inequality measurement: The problem of extreme values. *Journal of Econometrics*, 141(2), 1044-1072.
- Cowell, F. A., & Victoria-Feser, M. P. (2008). Modelling Lorenz Curves: Robust and Semi-parametric Issues. In D. Chotikapanich, *Modeling Income Distributions and Lorenz Curves. Economic Studies in Equality, Social Exclusion and Well-Being*, vol 5. New York: Springer.
- Crosato, L., & Ganugi, P. (2007). Statistical regularity of firm size distribution: the Pareto IV and truncated Yule for Italian SCI manufacturing. *Statistical Methods and Applications*, 16(1), 85-115.
- Das, S. (1995). Size, age and firm growth in an infant industry: The computer hardware industry in India. *International Journal of Industrial Organization*, 13(1), 111-126.
- Delmar, F., Davidsson, P., & Gartner, W. B. (2003). Arriving at the high-growth firm. *Journal of Business Venturing* 18(2), 189-216.
- di Giovanni, J., Levchenko, A. A., & Rancière, R. (2011). Power laws in firm size and openness to trade: measurement and implications. *Journal of International Economics*, 85(1), 42-52.
- Dunne, P., & Hughes, A. (1994). Age, size, growth and survival: UK Companies in the 1980s. *The Journal of Industrial Economics*, 42(2), 115-140.
- Evans, D. S. (1987). Tests of alternative theories of firm growth. *Journal of Political Economy* 95(4), 657-674.
- Fotopoulos, G., & Giotopoulos, I. (2010). Gibrat's law and persistence of growth in Greek manufacturing. *Small Business Economics*, 35(2), 191-202.
- Gallegati, M., & Palestrini, A. (2010). The complex behavior of firms' size dynamics. *Journal Of Economic Behavior & Organization*, 75(1), 69-76.
- Gastwirth, J. (1971). A general definition of the Lorenz curve. *Econometrica*, 39(6), 1037-1039.
- Gibrat, R. (1931). *Les Inégalités Economiques*. Paris: Recueil Sirey.

- Goddard, J., Liu, H., Donal, M., & Wilson, J. O. (2014). The size distribution of US banks and credit unions. *International Journal of the Economics of Business*, 21(1), 139-156.
- Guo, J., Xu, Q., Chen, Q., & Wang, Y. (2013). Firm size distribution and mobility of the top 500 firms in China, the United States and the world. *Physica A: Statistical Mechanics and its Applications*, 392(13), 2903-2914.
- Gupta, H. M., Campanha, J. R., de Aguiar, D. R., Queiroz, G. A., & Raheja, C. G. (2007). Gradually truncated log-normal in USA publicly traded firm size distribution. *Physica A: Statistical Mechanics and its Applications*, 375(2), 643-650. doi:<http://dx.doi.org/10.1016/j.physa.2006.09.025>
- Hart, P. E., & Oulton, N. (1996). Growth and Size of Firms. *The Economic Journal*, 106(438), 1242-1252.
- Hart, P. E., & Oulton, N. (1997). Zipf and the size distribution of firms. *Applied Economics Letters*, 4(4), 205-206.
- Hart, P., & Prais, S. (1956). The Analysis of Business Concentration: A Statistical Approach. *Journal Of The Royal Statistical Society. Series A (General)*, 119(2), 150-191.
- Hasegawa, H., & Kozumi, H. (2003). Estimation of Lorenz curves: a Bayesian nonparametric approach. *Journal of Econometrics*, 115(2), 277-291.
- Heinrich, T., & Dai, S. (2016). Diversity of firm sizes, complexity, and industry structure in the Chinese economy. *Structural Change and Economic Dynamics*, 37, 90-106.
- Heshmati, A. (2001). On the growth of micro and small firms: evidence from Sweden. *Small Business Economics*, 17(3), 213-228.
- Huynh, K. P., & Petrunia, R. J. (2010). Age effects, leverage and firm growth. *Journal of Economic Dynamics and Control*, 34(5), 1003-1013.
- Jang, S. S., & Park, K. (2011). Inter-relationship between firm growth and profitability. *International Journal of Hospitality Management*, 30(4), 1027-1035.
- Jensen, M. (1986). Agency costs of free cash flow, corporate finance, and takeovers. *American Economic Review*, 76(2), 323-329.
- Jondeau, E., & Rockinger, M. (2001). Gram-Charlier densities. *Journal of Economic Dynamics & Control*, 25(10), 1457-1483.
- Kaizoji, T., Iyetomi, H., & Ikeda, Y. (2006). Re-examination of the size distribution of firms. *Evolutionary And Institutional Economics Review*, 2(2), 183-198.
- Kuhs, W. F. (1988). The anharmonic temperature factor in crystallographic structure analysis. *Australian Journal of Physics*, 41(3), 369-382.
- Leon, A., Mencía, J., & Sentana, E. (2009). Parametric properties of semi-nonparametric distributions, with applications to option valuation. *Journal of Business and Economic Statistics*, 27(2), 176-192.
- Liñares-Zegarra, J., & Wilson, J. O. (2018). The size and growth of microfinance institutions. *The British Accounting Review*, 50(2), 199-213.
- Lorenz, M. O. (1905). Methods of Measuring the Concentration of Wealth. *Publications of the American Statistical Association*, 9(70), 209-219.
- Lotti, F., Santarelli, E., & Vivarelli, M. (2009). Defending Gibrat's Law as a long-run regularity. *Small Business Economics*, 32(1), 31-44.
- Machado, J., & Mata, J. (2000). Box-Cox quantile regression and the distribution of firm sizes. *Journal of Applied Econometrics* 15(3), 253–274.



- Mansfield, E. (1962). Entry, Gibrat's law, innovation, and the growth of firms. *American Economic Review* 52(5), 1023–1051.
- Marsili, O. (2006). Stability and turbulence in the size distribution of firms: evidence from Dutch manufacturing. *International Review of Applied Economics*, 20(2), 255-272.
- Martínez-Mekler, G., Martínez, R. A., del Río, M. B., Mansilla, R., Miramontes, P., & Cocho, G. (2009). Universality of rank-ordering distributions in the arts and sciences. *PLoS ONE*, 4(3), e4791.
- Mauleón, I., & Perote, J. (2000). Testing densities with financial data: an empirical comparison of the Edgeworth-Sargan density to the Student's t. *European Journal of Finance*, 6(2), 225-239.
- McCloughan, P. (1995). Simulation of concentration development from modified Gibrat growth-entry- exit processes. *The Journal Of Industrial Economics*, 43(4), 405-433.
- Meisenzahl, R. R. (2016). Can financing constraints explain the evolution of the firm size distribution? *Review of Industrial Organization*, 48(2), 123–147.
- Modigliani, F., & Miller, M. H. (1963). Corporate income taxes and the cost of capital: a correction. *The American Economic Review*, 53(3), 433-443.
- Moreno, A. M., & Casillas, J. C. (2007). High-growth SMEs versus non-high-growth SMEs: a discriminant analysis. *Entrepreneurship & Regional Development*, 19(1), 69-88.
- Myers, S. (1984). Presidential address: the capital structure puzzle. *Journal of Finance*, 39(3), 575-592.
- Newman, M. J. (2005). Power laws, Pareto distributions and Zipf's law. *Contemporary Physics*, 46(5), 323-351.
- Ñíguez, T., Paya, I., Peel, D., & Perote, J. (2012). On the stability of the constant relative risk aversion (CRRA) utility under high degrees of uncertainty. *Economics Letters*, 115(2), 244-248.
- Ñíguez, T.-M., Paya, I., Peel, D., & Perote, J. (2013). Higher-order moments in the theory of diversification and portfolio composition. *Economics Working Paper Series 2013/003. Lancaster University*.
- Ogwang, T., & Rao, U. (2000). Hybrid models of the Lorenz curve. *Economics Letters*, 69(1), 39-44.
- Oliveira, B., & Fortunato, A. (2006). Testing Gibrat's law: empirical evidence from a panel of Portuguese manufacturing firms. *International Journal Of The Economics Of Business*, 13(1), 65-81.
- Park, Y., Shin, J., & Kim, T. (2010). Firm size, age, industrial networking, and growth: a case of the Korean manufacturing industry. *Small Business Economics*, 35(2), 153-168.
- Pascoal, R., Augusto, M., & Monteiro, A. M. (2016). Size distribution of Portuguese firms between 2006 and 2012. *Physica A: Statistical Mechanics and its Applications*, 458, 342-355.
- Reid, G. C., & Xu, Z. (2012). Generalising Gibrat: using Chinese evidence founded on fieldwork. *Small Business Economics*, 39(4), 1017-1028.
- Santarelli, E., Klomp, L., & Thurik, A. R. (2006). Gibrat's Law: An overview of the empirical literature. En S. E. (eds), *Entrepreneurship, Growth, and Innovation: The Dynamics of Firms and Industries* (págs. 41-73). New York: Springer.

- Sarabia, J., Castillo, E., & Slottje, D. J. (1999). An ordered family of Lorenz curves. *Journal of Econometrics*, 91(1), 43-60.
- Shanmugam, K. R., & Bhaduri, S. N. (2002). Size, age and firm growth in the Indian manufacturing sector. *Applied Economics Letters*, 9(9), 607-613.
- Segarra, A., & Teruel, M. (2012). An appraisal of firm size distribution: Does sample size matter?. *Journal Of Economic Behavior & Organization*, 82(1), 314-328.
- Sen, A. (1973). *On Economic Inequality*. Oxford: Clarendon Press.
- Simon, H., & Bonini, C. (1958). The size distribution of business firms. *The American Economic Review*, 48(4), 607-617.
- Stanley, M. H., Buldyrev, S. V., Havlin, S., Mantegna, R. N., Salinger, M. A., & Stanley, H. E. (1995). Zipf plots and the size distribution of firm. *Economics Letters*, 49(4), 453-457.
- Tang, A. (2014). Does Gibrat's law hold for Swedish energy firms? *Empirical Economics*, 49(2), 659-674.
- Voit, J. (2001). The growth dynamics of German business firms. *Advances in Complex Systems*, 4(1), 149-162.
- Wang, Y., & You, S. (2016). An alternative method for modeling the size distribution of top wealth. *Physica A: Statistical Mechanics and its Applications*, 457, 443-453.
- Zhang, J., Chen, Q., & Wang, Y. (2009). Zipf distribution in top Chinese firms and an economic explanation. *Physica A: Statistical Mechanics and its Applications*, 388(10), 2020-2024.
- Zhang, Y. Y., Wu, X., & Li, Q. (2016). A Simple Consistent Nonparametric Estimator of the Lorenz Curve. In G. González-Rivera, R. C. Hill, & T.-H. Lee (ed.), *Essays in Honor of Aman Ullah (Advances in Econometrics, Volume 36)*. Emerald Group Publishing Limited.

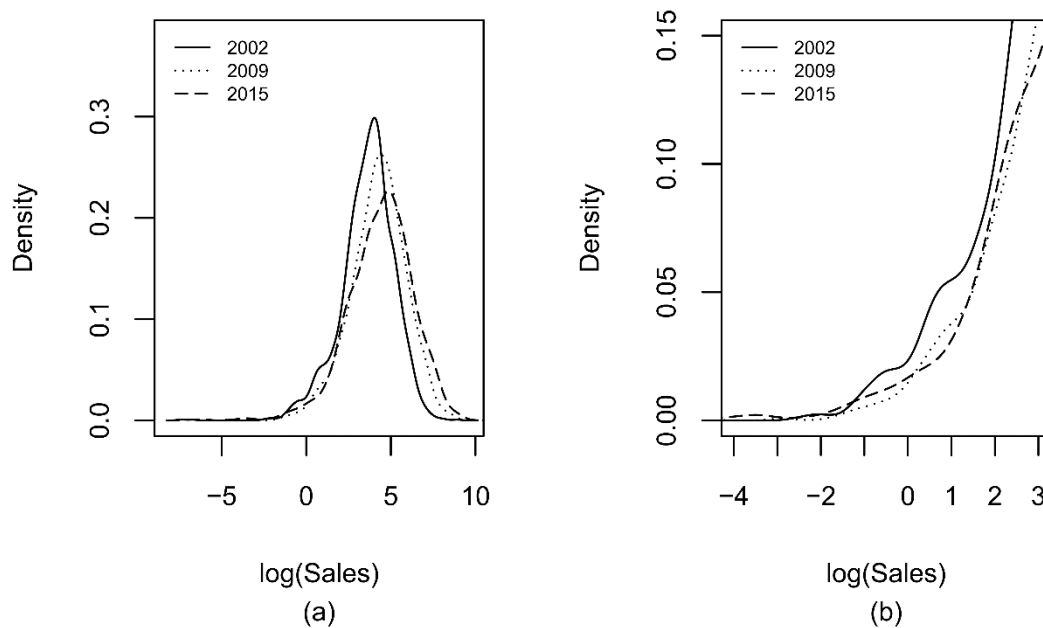
## Figures and Tables

**Table 1.** Descriptive statistics.

Variable	n	Mean	Sd	Skew	Kurtosis
Growth	24,808	1.61%	48.85%	0.78	46.00
Sales	24,808	22,447	68,140	20.08	689.07
Age	24,808	30.04	10.30	0.68	4.21
Leverage	24,808	40.98%	27.67%	2.16	26.01
ROE	24,808	2.52%	654.72%	-116.96	17896.12

Note: The sample is composed of 1,772 Colombian companies in all sectors of the economy. The data is collected over a 14-year period from 2002 to 2015. Growth is the firm's growth, calculated as the first difference of the natural logarithm of sales. Sales is the value of net sales in Colombian peso (COP) deflated by Colombian Consumer Price Index (CPI) using 2008 price base. Age corresponds to the age measured in years. Leverage is calculated as the sum of long-term debt and short-term debt divided by total assets. ROE is computed as net income divided by common equity. n = number of observations, Mean = mean value of the variable, Sd = standard deviation, Skew and Kurtosis correspond to the coefficient of asymmetry and excess kurtosis, respectively.

**Fig. 1.** Empirical density of the logarithm of sales.



Note: The figure shows the density of the logarithm of the sales variable ( $\log(\text{Sales})$ ) resulting from a smoothing of the corresponding histogram. The panel (a) represents the total of the domain and (b) a detail of the left tails, where the smallest firms in terms of sales are located. Sales were divided by the factor  $10^8$ .

**Table 2.** Estimates of firm size distribution using a lognormal and log-semi-nonparametric distribution.

Year	Panel A lognormal					Panel B log-SNP									Panel C LR
	μ	σ	logL	AIC	KS test	μ	σ	d <sub>1</sub>	d <sub>2</sub>	d <sub>3</sub>	d <sub>4</sub>	logL	AIC	KS test	
2002	3.6133	1.5879	-9,736.50	19,476.99	(0.0045)	1.4768	1.5608	1.3689	0.9545	0.3476	0.0690	-9,668.80	19,349.60	(0.9184)	135.39
	( $<.0001$ )	( $<.0001$ )			Rejected	( $<.0001$ )	( $<.0001$ )	( $<.0001$ )	( $<.0001$ )	( $<.0001$ )	( $<.0001$ )			Not rejected*	( $<.0001$ )
2003	3.7475	1.5881	-9,974.50	19,952.99	(0.0253)	1.7825	1.5169	1.2954	0.8870	0.3136	0.0628	-9,922.35	19,856.71	(0.4552)	104.28
	( $<.0001$ )	( $<.0001$ )			Not rejected	( $<.0001$ )	( $<.0001$ )	( $<.0001$ )	( $<.0001$ )	( $<.0001$ )	( $<.0001$ )			Not rejected*	( $<.0001$ )
2004	3.8629	1.5986	-10,190.67	20,385.34	(0.0410)	1.9147	1.5299	1.2734	0.8567	0.2926	0.0561	-10,138.89	20,289.78	(0.8099)	103.57
	( $<.0001$ )	( $<.0001$ )			Not rejected	( $<.0001$ )	( $<.0001$ )	( $<.0001$ )	( $<.0001$ )	( $<.0001$ )	( $<.0001$ )			Not rejected*	( $<.0001$ )
2005	3.9511	1.5829	-10,329.62	20,663.24	(0.0308)	2.3932	1.4447	1.0784	0.6816	0.2075	0.0373	-10,292.80	20,597.61	(0.6455)	73.63
	( $<.0001$ )	( $<.0001$ )			Not rejected	( $<.0001$ )	( $<.0001$ )	( $<.0001$ )	( $<.0001$ )	(0.004)	(0.0401)			Not rejected*	( $<.0001$ )
2006	4.0923	1.6524	-10,655.87	21,315.74	(0.383)	2.0231	1.5973	1.2954	0.8741	0.2963	0.0627	-10,599.79	21,211.59	(0.8099)	112.15
	( $<.0001$ )	( $<.0001$ )			Not rejected	( $<.0001$ )	( $<.0001$ )	( $<.0001$ )	( $<.0001$ )	( $<.0001$ )	( $<.0001$ )			Not rejected*	( $<.0001$ )
2007	4.2018	1.6793	-10,878.49	21,760.98	(0.0169)	2.1939	1.5899	1.2629	0.8553	0.2939	0.0627	-10,826.84	21,665.68	(0.8346)	103.31
	( $<.0001$ )	( $<.0001$ )			Not rejected	( $<.0001$ )	( $<.0001$ )	( $<.0001$ )	( $<.0001$ )	( $<.0001$ )	( $<.0001$ )			Not rejected*	( $<.0001$ )
2008	4.2697	1.6560	-10,974.08	21,952.16	(0.0836)	4.4345	1.7820	-0.0925	-0.0640	-0.0435	0.0235	-10,949.07	21,910.15	(0.6455)	50.01
	( $<.0001$ )	( $<.0001$ )			Not rejected	( $<.0001$ )	( $<.0001$ )	(0.4161)	(0.1352)	(0.0555)	(0.0006)			Not rejected*	( $<.0001$ )
2009	4.2478	1.6638	-10,943.68	21,891.35	(0.2445)	2.0663	1.6910	1.2901	0.8162	0.2760	0.0494	-10,915.71	21,843.42	(0.9001)	55.93
	( $<.0001$ )	( $<.0001$ )			Not rejected	( $<.0001$ )	( $<.0001$ )	( $<.0001$ )	( $<.0001$ )	( $<.0001$ )	( $<.0001$ )			Not rejected*	( $<.0001$ )
2010	4.2745	1.6774	-11,005.39	22,014.78	(0.2797)	3.0176	1.5149	0.8297	0.4573	0.1105	0.0295	-10,987.51	21,987.02	(0.6455)	35.77
	( $<.0001$ )	( $<.0001$ )			Not rejected	( $<.0001$ )	( $<.0001$ )	( $<.0001$ )	( $<.0001$ )	(0.0243)	(0.0546)			Not rejected*	( $<.0001$ )
2011	4.3511	1.7271	-11,192.81	22,389.61	(0.0056)	1.9534	1.8060	1.3277	0.8386	0.2802	0.0533	-11,163.85	22,339.69	(0.8346)	57.92
	( $<.0001$ )	( $<.0001$ )			Rejected	( $<.0001$ )	( $<.0001$ )	( $<.0001$ )	( $<.0001$ )	( $<.0001$ )	( $<.0001$ )			Not rejected*	( $<.0001$ )
2012	4.3619	1.7426	-11,227.72	22,459.44	(0.2445)	2.3906	1.7170	1.1481	0.6740	0.2068	0.0345	-11,205.56	22,423.11	(0.6455)	44.33
	( $<.0001$ )	( $<.0001$ )			Not rejected	( $<.0001$ )	( $<.0001$ )	( $<.0001$ )	( $<.0001$ )	(0.0019)	(0.0145)			Not rejected*	( $<.0001$ )

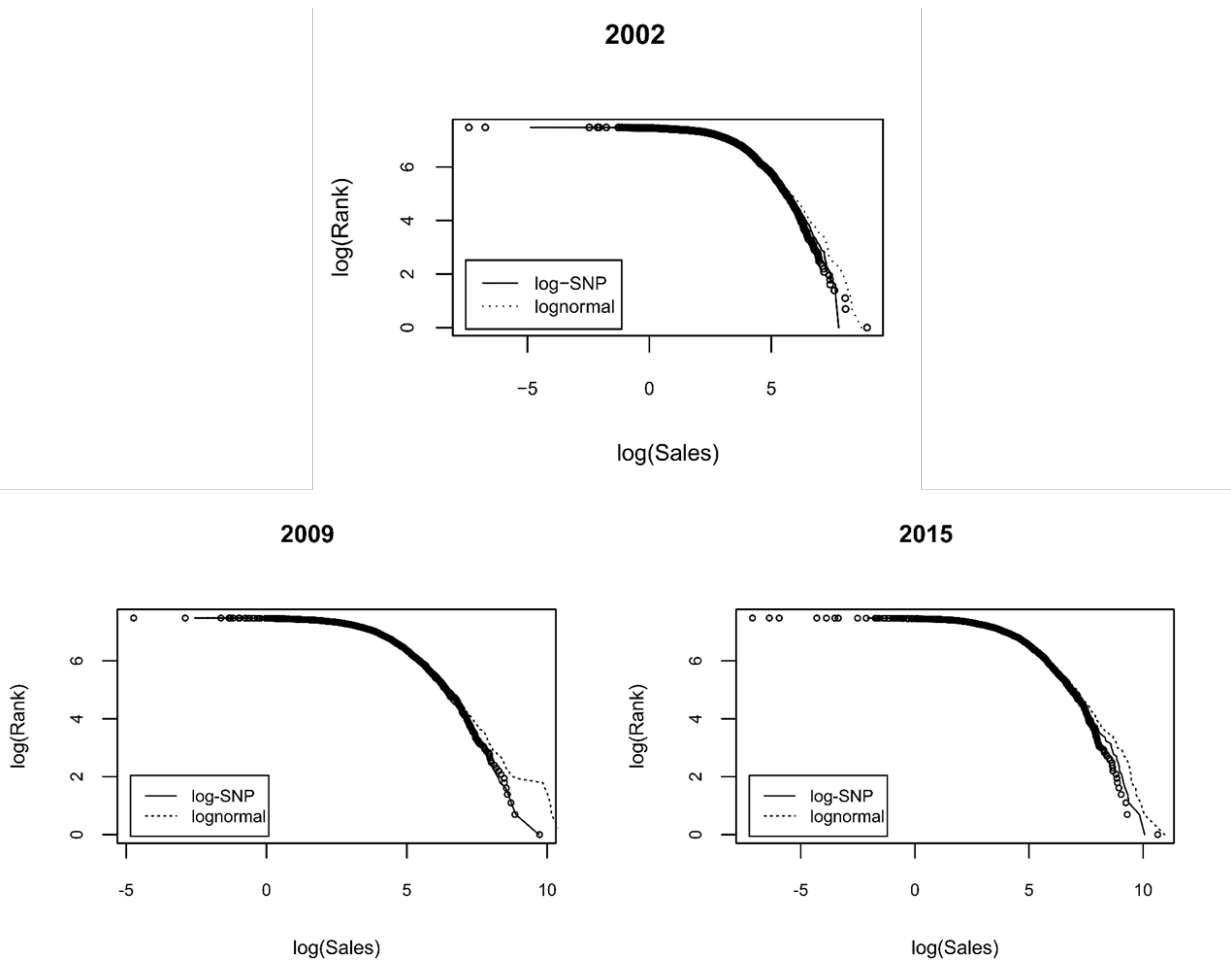
To be continued

Continuation Table 2

Year	Panel A lognormal					Panel B log-SNP									Panel C LR
	μ	σ	logL	AIC	KS test	μ	σ	d <sub>1</sub>	d <sub>2</sub>	d <sub>3</sub>	d <sub>4</sub>	logL	AIC	KS test	
2014	4.3607	1.8548	-11,336.24	22,676.47	(0.383)	1.8202	1.9081	1.3314	0.8588	0.2830	0.0559	-11,293.32	22,598.63	(0.9001)	85.84
	(<.0001)	(<.0001)			Not rejected	(<.0001)	(<.0001)	(<.0001)	(<.0001)	(<.0001)	(<.0001)			Not rejected*	(<.0001)
2015	4.3796	1.9404	-11,449.71	22,903.42	(0.5071)	1.5302	2.0167	1.4129	0.9611	0.3281	0.0704	-11,389.43	22,790.86	(0.7299)	120.55
	(<.0001)	(<.0001)			Not rejected	(<.0001)	(<.0001)	(<.0001)	(<.0001)	(<.0001)	(<.0001)			Not rejected*	(<.0001)

Note: This table reports ML estimates for a sample of 1,772 Colombian firms. Panels A, B and C show the estimated parameters for the lognormal distribution, log-SNP distribution and the likelihood ratio for testing both specifications, respectively.  $\mu$  and  $\sigma$  are location and scale parameters (respectively) and  $d_s$  the shape parameters. logL = log-likelihood, AIC = Akaike Information Criterion, LR = likelihood ratio, KS test = Kolmogorov-Smirnov test. P-values in parentheses. Not rejected\* indicates a better fit in the KS test. Sales were divided by the factor  $10^8$ .

**Fig. 2** Logarithm of firm size vs. logarithm of sales.



Note: The figure compares the empirical values (hollow points) and the estimated values under a lognormal specification (dashed line) and log-SNP (solid line). The axes are in logarithmic scale and correspond to the relationship between Rank and Sales for a sample of 1,772 Colombian firms. Sales were divided by the factor  $10^8$ .

**Table 3.** Sales obtained empirically versus values expected theoretically using a lognormal distribution and log-semi-nonparametric distribution.

Year	Observed sales value (millions, COP pesos)			Expected sales value (millions, COP pesos)					
				Lognormal			Log-SNP		
	10%	5%	1%	10%	5%	1.0%	10%	5%	1%
2002	23,451.86	38,227.88	87,209.13	28,378.75	50,526.95	149,099.32	24,018.42	39,525.33	101,613.90
2003	28,717.54	46,491.08	99,324.82	32,463.40	57,804.70	170,603.97	27,501.74	44,909.17	113,434.40
2004	31,337.34	50,769.05	119,950.90	36,928.78	66,007.68	196,216.30	31,326.16	51,456.53	131,399.60
2005	35,026.29	57,931.51	126,561.90	39,533.83	70,262.34	206,642.20	34,191.24	55,511.25	137,742.60
2006	42,617.09	72,973.43	170,576.80	49,767.04	90,710.48	279,715.40	42,148.36	70,741.10	188,147.10
2007	50,030.84	86,731.06	210,178.50	57,472.96	105,785.12	332,236.20	48,449.59	81,331.37	216,017.80
2008	54,082.71	95,918.84	212,051.90	59,701.06	108,958.68	336,803.50	54,025.03	91,018.41	236,091.90
2009	53,869.23	93,628.22	242,631.30	58,999.08	107,984.97	335,585.60	52,740.64	91,338.40	257,938.90
2010	57,660.52	107,134.72	247,459.70	61,660.73	113,416.07	355,748.80	56,445.67	96,872.13	263,597.20
2011	64,912.65	117,687.80	271,429.20	70,941.61	132,860.90	431,077.00	64,208.30	115,175.00	348,034.50
2012	67,081.24	123,506.30	308,754.10	73,150.65	137,773.90	451,777.60	65,657.94	116,254.40	340,606.30
2013	68,343.02	128,521.50	330,017.00	77,998.47	150,838.30	519,753.80	66,825.25	119,220.40	354,467.90
2014	74,772.49	138,535.60	374,155.20	84,364.18	165,504.10	585,829.90	73,637.94	136,396.90	438,116.60
2015	81,185.61	153,785.60	390,981.30	95,943.81	194,165.70	728,564.00	81,192.43	154,382.10	522,820.90

Note: This table compares the value of sales, in millions of COP pesos, observed empirically in a sample of 1,772 Colombian firms versus the theoretically expected under lognormal and log-SNP distributions. The values 10%, 5% and 1% are percentiles of the distributions.

**Fig. 3** Dynamics of the empirical and theoretical value of the Gini index.



Note: The figure compares the time evolution of the empirical values of the Gini index (solid circle) and the theoretically estimated values under a lognormal (square) and log-SNP (triangle) specification.

**Table 4** Determinants of business growth.

	<b>Model 1</b>	<b>Model 2</b>	<b>Model 3</b>
Growth <sub>i,t-1</sub>	-0.1988*** (0.000)	-0.1795*** (0.000)	-0.1802*** (0.000)
Growth <sub>i,t-2</sub>	-0.0837*** (0.000)	-0.0737*** (0.000)	-0.0740*** (0.000)
log(Sales <sub>i,t-1</sub> )	-0.0640*** (0.002)	-0.1372*** (0.002)	-0.1347*** (0.001)
logAge <sub>it</sub>		32.2789** (0.059)	30.6995** (0.061)
[logAge <sub>it</sub> ] <sup>2</sup>		-4.9241** (0.061)	-4.6812** (0.064)
Leverage <sub>i,t-1</sub>		1.6136*** (0.009)	1.5566*** (0.008)
ROE <sub>i,t-1</sub>			-0.0004 (0.571)
Constant	-1.0254*** (0.001)	50.7655** (0.062)	-48.2463** (0.064)
Observations	24808	24808	24808
Firms	1172	1172	1172
Instruments	7	9	11
Wald Test	148.12	114.54	115.24
(p-value)	(0.000)	(0.000)	(0.000)
Hansen Test	2.27	0.49	0.78
(p-value)	(0.518)	(0.782)	(0.854)
Arellano-Bond Test for AR(1)	-14.88	-6.17	-6.66
(p-value)	(0.000)	(0.000)	(0.000)
Arellano-Bond Test for AR(2)	-1.48	1.28	1.21
(p-value)	(0.138)	(0.199)	(0.226)

Note: Models 1, 2 and 3 correspond to different estimates of firm growth carried out using the System-GMM estimator. The sample is composed of 1,772 Colombian firms in all sectors of the economy. The data is collected over a 14-year period from 2002 to 2015. Growth is the firm's growth, calculated as the first difference of the natural logarithm of sales. Sales is the value of net sales in Colombian peso (COP) deflated by Colombian Consumer Price Index (CPI) using 2008 price base. Age corresponds to the age of the company measured in years. Leverage is calculated as the sum of long-term debt and short-term debt divided by total assets. ROE is calculated as net income divided by equity. \*, \*\*, \*\*\* indicate levels of significance at 10%, 5% and 1%, respectively.