ipea
Institute for Applied Economic Research

277

Brasilia, June 2023

EXECUTIVE SUMMARY

FEATURES AND DETERMINANTS OF THE FIRM SIZE DISTRIBUTION: AN EMPIRICAL ANALYSIS WITH BRAZILIAN DATA

Miguel Nathan Foguel
Researcher at the Institute for Applied Economic Research (Ipea).
E-mail: <miguel.foguel@ipea.gov.br>.

Eduardo Pontual Ribeiro
Professor at the Federal University of Rio de Janeiro (UFRJ).
E-mail: <eribeiro@ie.ufrj.br>.

DOI: http://dx.doi.org/10.38116/td277-eng

The firm size distribution in a country can influence an ample set of economic variables. This includes the level and quality of employment, earnings inequality, the degree of competition, the innovation rate, and the productivity of the economy. Investigating the features and the determinants of this distribution is thus key not only to understand some of the underlying factors that affect the functioning of the economy but also to shed some light onto which areas policy makers should focus their attention.

Given its importance, it is not surprising that the firm size distribution has received a lot attention in the academic literature. There is a large set of studies that focus on the statistical properties of this distribution, and there is wide consensus that the tails of the firm size distribution in many countries is consistent with a Power Law (or Pareto distribution). Additionally, the literature has considered a special case of the Pareto distribution, namely the Zipf distribution, for which the shape parameter is equal to unity. The empirical regularity on the Pareto distribution and its special Zipf case implies that the frequency of firms' sizes is inversely proportional to their sizes, at least above a minimum threshold.

There is a large theoretical literature that provides a conceptual reference that can be used to empirically investigate the importance of a set of determinants of the firm size distribution. However, there are only a few studies that pursue this line of investigation, namely Henrekson e Johansson (1999) and Halvarsson (2013), both of which based on data from Sweden. This paper seeks to not only fill part of this gap but also characterize the firm size distribution and its determinants for a developing country such as Brazil. To the best of our knowledge, this is the first study that does that for a developing country.

To accomplish that we employ the two-stage methodology of Halvarsson (2013). In the first stage, the shape parameter of the Pareto distribution is estimated for different 3-digit industries over the years. Given that the Pareto distribution is more appropriate for firms' sizes above a certain (unkown) level, we follow Halvarsson (2013) and employ the procedure proposed by Clauset, Shalizi and Newman (2009) to determine this level. The Clauset, Shalizi and Newman (2009) method is based on the Kolgomorov-Smirnov (KS) test. However, as Goerlich (2013) points outs, this method has some limitations and thus we also employ the Lagrange Multiplier (LM) test proposed by Goerlich (2013) to pinpoint the minimum size level. Using the minimum threshold sizes determined by the aforementioned methods, we estimate the shape parameter using the rank-size log-linear regression proposed by Gabaix and Ibragimov (2011). The second stage

uses the estimates of the shape parameters from the first stage as the dependent variable in pooled and panel regressions that are run against a set of explanatory variables that intends to measure relevant determinants of the shape of tail of the firm size distribution.

The main source of information used in the study is the *Relação Anual de Informações Sociais* (Rais – Annual Roll of Social Information), a very large administrative dataset collected by Ministry of Labor that contains information on every labor contract on the formal sector in Brazil. Apart from data on the number of firms' employees during the year, the available information includes the industry of firms. This allows us to construct industry-level firm size distribution for every year between 2007 and 2019. We also use Rais to construct some of the explanatory variables used in the second stage regression. Industry-level information for the manufacturing sector is also used for the second stage.

Our main explanatory variables are: i) the proportion of workers with at least university degree, which intends to capture industry human capital; ii) the sum of absolute changes in industry payroll, which intends to measure industry instability; iii) the standard deviation of growth in industry payroll, which intends to gauge industry uncertainty; and iv) the average age of firm in the industry, to capture the maturity of the industry.

Our estimates indicate that the Pareto distribution is valid for the tails of the majority of industries in Brazil. The tail itself tend to be high, with the estimated minimum size on the upper quarter centiles of the distributions. The Parameter estimates are mostly over unity, with median estimate of 1.2 under the KS minimum size criteria and 1.3 under the LM minimum size criteria. Interestingly, about half of the industry-year distribution we can accept the hypothesis that the distribution follows the special case of the Zipf's distribution.

Results for the second stage show that human capital is negatively associated with the shape parameter of the Pareto, as expected in the literature. Our estimates partially confirm the predictions on the uncertainty and instability variables, although the results are insignificant when considering short-term variations of these variables, i.e. in models with fixed effects. The results are generally the same when considering only manufacturing or all industries, suggesting that the known differences between, say, services and manufacturing such as size and human capital intensity do not seem to affect the main results.

REFERENCES

CLAUSET, A.; SHALIZI, C. R.; NEWMAN, M. E. J. Power-law distributions in empirical data. **SIAM Review**, v. 51, n. 4, p. 661-703, 2009.

GABAIX, X.; IBRAGIMOV, R. Rank – ½: a simple way to improve the OLS estimation of tail exponents. **Journal of Business and Economics Statistics**, v. 29, n. 1, p. 24-39, 2011.

GOERLICH, F. J. A simple and efficient test for the Pareto law. **Empirical Economics**, v. 45, p. 1367-1381, 2013.

HALVARSSON, D. **Firm dynamics**: the size and growth distribution of firms. 2013. Dissertation (Doctoral Degree) – Department of Industrial Economics and Management, Royal Institute of Technology, Stockholm, 2013.

HENREKSON, M.; JOHANSSON, D. Institutional effects on the evolution of the size distribution of firms. **Small Business Economics**, v. 12, n. 1, p. 11-23, 1999.