# A survey on Stylized Facts for Brazilian Manufacturing: 1996-2013

Elias Haddad*
Ivette Luna†

This paper investigates the statistical properties and patterns of Brazilian manufacturing firms related to size, productivity, growth, and their seemingly ubiquitous heterogeneity. Using microdata from the Brazilian Industrial Survey from 1996 to 2013, we conducted exercises using panel and cross-sectional data at different levels of aggregation. Our primary objective is to see which of the most common stylized facts explored in the evolutionary economics literature are empirically supported for Brazil. We find that, despite significant differences among individual sectors, there is a core set of regularities that seems to hold for all of them, such as the lognormal-Pareto distribution of firm size and the Laplacian distribution of firm growth and productivity rates. As such, the evidence for Brazil corroborates the results found for developed countries. These stylized facts may then describe ubiquitous processes driving market organization in Economics.
**JEL: L11, D22, L60, L80.**
**Keywords:** developing countries, industrial dynamics, heterogeneity, firm size distributions.
**ANPEC AREA:** 09.

Este trabalho investiga propriedades estatísticas e padrões da manufatura brasileira relacionados a tamanho, produtividade, crescimento e sua aparente ubíqua heterogeneidade. Através de microdados da pesquisa industrial (PIA) de 1996 a 2013, nós conduzimos exercícios usando recortes em painel e *cross-section* a diferentes níveis de agregação. Nosso objetivo principal é verificar quais dos fatos estilizados mais comuns explorados na literatura de economia evolucionária tem corroboração empírica para o caso brasileiro. Nossos resultados mostram que, apesar de diferenças significativas entre diferentes setores, existe um núcleo de regularidades que aparentemente ocorrem para todos, como o formato lognormal-Pareto das distribuições de tamanho das firmas, e o formato Laplaciano das distribuições das taxas de crescimento e variação da produtividade. Dessa forma, as evidências para o Brasil se alinham com os resultados encontrados para os países desenvolvidos, e reforçam a ideia de que esses fatos estilizados descrevem processos universais que impulsionam a organização dos mercados em Economia.
**JEL: L11, D22, L60, L80.**
**Palavras-chave:** países em desenvolvimento, dinâmica industrial, heterogeneidade, distribuição de tamanho das firmas.
**ÁREA ANPEC:** 09.

*School of Electrical and Computer Engineering (FEEC), University of Campinas; Av. Albert Einstein, 400, Barão Geraldo, Campinas (Brazil).
†Institute of Economics (IE), University of Campinas; Rua Pitágoras, 353 Barão Geraldo - Campinas (Brazil).

# 1 Introduction

There is a rich strand of literature concerning patterns in Industrial Organization that begins with the works of Pareto (1896) and Gibrat (1931), all the way down to Simon (1955). However, only recently has there been a systematic effort to compile evidence for several countries outside the US, developed or not, such as China, India, France, and Italy (Bottazzi and Secchi, 2003, 2005; Bottazzi et al., 2007, 2011; Mathew, 2017; Yu et al., 2015b). Our work aims to cover the gap concerning the Brazilian manufacturing sector.

Specifically, we will ascertain if 1) Brazilian firms are characterized by the same large skewness and wide dispersion in most economic metrics as found in other countries; 2) if there is favorable evidence for the Pareto or lognormal distribution of firm size; and 3) if there is favorable evidence for the Laplacian distribution of size and productivity growth rates. By doing this, we find a compelling case against some common hypotheses in economics.

First, there is a wide heterogeneity in firm behavior for all proxies used, robust under any level of disaggregation and persistent over time. Such an outcome cast doubts about the existence of an optimal firm size or the usefulness of a representative agent as a tool to model the Brazilian economy. Particularly, our work disavows its usage from an empirical point of view, complementing other works such as Shaikh (1974), Fisher (2005) and Felipe and McCombie (2013), which took a theoretical approach for their critique. Since this widespread heterogeneity was found for several countries[1], the theoretical economist is on very unstable grounds when using such tools to model anything minimally resembling the real world.

Second, the general occurrence of a very fat left tail on productivity distributions shows that even if market selection were an effective promoter of virtuous change on the economy, and most empirical studies up to this point show that it is not (Bottazzi et al., 2010; Dosi et al., 2015; Yu et al., 2015a), its impact is not enough to avoid the entry and survival of new low-productivity firms, which makes any convergence for an efficiency frontier a non-occurrence in practical terms. In other words, market selection does not seem enough to promote productivity growth, and it appears that firms that are already in activity very seldom radically change their productive structure.

The lack of empirical validation for competition as broadly understood has consequences on several grounds, since most of the evolutionary tradition uses selection mechanisms as the basis for their models. While the very high entry and exit rates indeed point to an underlying choice mechanism by the customer, it is not clear whether this mechanism is directly related to firm efficiency. Whereas better overall productivity could be translated into better prices, price by itself, as the innovation literature indicates, may not be the main factor of a purchase decision, and other factors related to social status, such as brand power and design, have to be incorporated. These results would diminish the importance of productivity in overall economic modeling and points to the necessity of searching for proxies of cultural selection.

Third, Brazilian growth distributions have very extreme events, which are closely approximated by a Laplacian distribution, a curve characterized by its fat tails. As a consequence, growth is not random. While predicting growth until now has been an impossible task, specially when trying to differentiate high growth firms from generic firms (Coad et al., 2014), the overall distribution of firm growth rates does not follow a normal curve, showing that growth itself is not a series of independent events, but has some underlying correlation instead, even if very complex to be easily described.

Finally, while the idea of technological trajectories may be useful to understand how firms acquire capabilities and how new technological paradigms can create instability, the incremental nature of progress that such theory predicts is not apparent in enterprise surveys. Instead, productivity growth rates for Brazil have even fatter tails than size growth rates. The occurrence of such short-term extreme events highlights that productivity, when measured by monetary terms, is very volatile. Particularly, since there is an abundance of negative events, the data suggests that monetary productivity may be mostly dominated by demand, in a sort of micro Kaldor-Verdoorn law.

[1] See Bartelsman and Doms (2000) for a review.

Since this kind of variation occurs even when we measure it by proxies that seem to be only internal to the firm, such as value added, we hypothesize that it must somehow reflect changes in capacity utilization[2]. The problem is that the lack of correlation between growth and productivity change as evidenced in the literature shows that this process is not of a simple nature, and is probably related to price and inventory policies.

In order to show our investigation, our work is divided into three main lines. The first gives some contextual information about the Brazilian Manufacturing and its evolution through the period 1996-2013.

The second line explores the market concentration of Brazilian Firms. We perform aggregated and disaggregated estimations for size distributions in cross-sectional and annual views. Since Gibrat (1931) stated the Law of Proportionate Effect, i.e., that firms' growth appears to be uncorrelated with their size, patterns in distributions were found regarding market structure and organization for a broad range of countries and metrics. There is evidence of skewed distributions for firm size, closely approximated by a Pareto or lognormal distribution, at least in the aggregated level (Hart and Prais, 1956; Ijiri and Simon, 1977; Stanley et al., 1995; Axtell, 2001; Cabral and Mata, 2003), which usually extends over a broad support, implying the coexistence of firms with significant differences in size, spanning several orders of magnitude.

The third line of investigation deals with dynamics and performance. We perform estimations for distributions of growth and productivity accompanied by a parametric fit. This exercise is related to a more recent strand of research regarding the investigation of patterns for distribution of firm growth, productivity and productivity variation. Starting with Stanley et al. (1996), followed by Bottazzi and Secchi (2003, 2005); Bottazzi et al. (2007), the Laplacian distribution, a curve characterized by its fat tails, seems to be well suited to describe the distribution of these metrics for countries as dissimilar as China, India, the United States and Italy (Yu et al., 2015b; Mathew, 2017). The exponential decay of the Laplacian curve associated with its fat tails predicts far more frequent extreme events than they would be if short-term events were completely uncorrelated, such as in a normal distribution.

The rest of this work is divided as follows. The next section presents the data description and some context on the Brazilian manufacturing sector. The second section overviews the methodology used in this study. The third section presents the results and a discussion, and the fourth one ends the paper with some highlights and a conclusion.

# 2 Data Description

Our analysis is based on the Brazilian Industrial Survey (PIA), which contains yearly census information for firms with more than 30 employees and in sectors with CNAE (National Classification of Economic Activities) codes between 5-33[3]. Our total sample comprehends 467,695 observations from 1996 to 2013 and the monetary values were deflated using two-digit sectoral prices constructed with the GDP Implicit Deflator from the National Accounts.[4] While the number of employees, total revenue and value added are used as proxies for firm size, labor productivity is used as a proxy for efficiency, given by value added per employee.

We opted to use labor productivity for several reasons. The first is that the data on firm capital is unrepresentative of the whole sample due to a large amount of missing data. Labor productivity also

[2]Unless someone supposes that firms intentionally destroy their capital, which would somewhat produce a smaller output given the same input.

[3]The split in sectors agrees with the ISIC Rev. 4 structure at the two-Digit level, with minimal differences. Most expressively, alcohol production, which enters ISIC as a chemical product (Sector 20), is classified by the Brazilian Institute of Geography and Statistics' CNAE 2.0 as a biofuel (Sector 19), due to its extreme importance both as sole fuel and as a mixture with gasoline.

[4]The access to the data is restricted, and due to privacy reasons we have to exclude any sector with less than three firms in any particular exercise. These exclusions make some sectors, such as petroleum extraction, an activity that was a State monopoly until recently, to appear only in certain views. To avoid errors and fill-in mistakes, we also exclude firms with negative value added, negative total revenue, with less than 30 employees or that were registered as inactive.

does not require any intuition about the relationship of the productive structure, nor does it require strong hypotheses about the substitution between capital and labor.[5] Finally, it guarantees comparability between our study and those of several other scholars (Dosi et al., 2012; Yu et al., 2015b; Mathew, 2017).

Since we are dealing with census information for firms only above the threshold of 30 employees, most of the Brazilian manufacturing firms are not captured by our subsample. In 1996, firms with up to 29 employees represented 76% of the number of firms in manufacturing and mining, or about 82,940 firms. In 2013, this number increased to 86% of the total, or 296,154 firms.[6] However, according to IBGE (2013), they have a low share in the number of employees (17% in 1996, 22% in 2013) and value added (6% in 1996, 8% in 2013). Thus, despite the importance that small firms have on the Brazilian economy and which our subsample ignores, our database is still responsible, on average, for about 80% of the employment and 90% of the value added in manufacturing and mining (SEBRAE, 2014).

Table 1 shows a summary with statistics for the metrics used in this study considering the full data sample. The sectors that command the highest productivity are activities related to basic commodities, such as petroleum extraction and refining (ISICs 6 and 19) and mining (ISICs 7, 8 and 9). These sectors, along with motor vehicles (ISIC 29), were also the ones that experienced the highest growth in this period. The worst performer is the tobacco industry (ISIC 12), which suffered from restrictive domestic policies regarding product design, marketing, and places allowed for consumption. The ranking is followed by leather (ISIC 15) and textiles (ISIC 13), which were subject to heavy Chinese competition (Soares and Castilho, 2016).

Note that for the variation metrics we lost almost a hundred thousand observations due to firms that were not present in any of the years of the survey. Growth has a positive average rate for all but a few sectors, and productivity change shows mostly an insignificant or negative result. Only one sector shows a negative median growth, while productivity change presents 14 sectors with a weak or negative result. Both metrics present very fat tails for the majority of sectors.

For the metrics of size, most sectors have indeed significant positive skewness and kurtosis, which for unimodal distributions means that they are fatter on the left side with long tails on the right side.[7] We know that this dispersion comes mostly from large enterprises, and this becomes apparent by the distance between average and median in several industries. Particularly in metal ores (ISIC 7), the average is almost eight times the median for the number of workers, 34 times for total revenue and 27 times for value added.

The data shows a significant intersectoral heterogeneity, with some sectors having great productivity but most being much less prolific, which renders a poor overall result for total manufacturing. Our results are in accordance with other recent studies regarding the structural heterogeneity hypothesis for Brazilian manufacturing (Catela et al., 2015). These facts provide sound evidence for the ECLAC (Economic Comission for Latin America) tradition of centre-periphery (Prebisch, 1981; Cimoli and Porcile, 2013). The structural heterogeneity approach advocated by ECLAC assumes that underdeveloped countries, particularly those in Latin America, have a hard-cut division between sectors that are well-integrated in international trade and those that are only competitive in the national market, against a soft decay found for most developed countries.

Figure 1 shows the changes in shares of value added from 1996 to 2013. Most sectors reduced their participation to refined petroleum and metal ores. In the same period, these two industries, together with soybeans, produced the main products exported from Brazil. This decline in the complexity of manufactured and exported goods from Brazil has been appointed as a cause of low economic dynamism (Hausmann and Hidalgo, 2014), which is undoubtedly observed due to the reduced total manufacturing growth experienced in the period, and as a possible symptom of Dutch disease (Gala et al., 2017).

[5]Issues related to the empirical estimation of these metrics and their relationship with account identities are discussed by Felipe and McCombie (2013)

[6]Part of this increase is due to the IBGE starting to consider firms with less than five employees in the total.

[7]Not reported due to space limitations.

Table 1: Statistics on Brazilian manufacturing - cross-section - 1996-2013

| ISIC | Industry | Total Obs. Lvl. | Workers | | Total Revenue | | Value Added | | Productivity | | Total Obs. Diff. | Δ% Tot. Rev. | | Δ% Prod. | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | Avg. | Median | Avg. | Median | Avg. | Median | Avg. | Median | | Avg. | Median | Avg. | Median |
| 5 | Coal and lignite | 240 | 354 | 326 | 87 | 67 | 44 | 35 | 135 | 104 | 206 | −1 | −2 | −5 | −4 |
| 6 | Crude petroleum | 54 | 172 | 99 | 1,052 | 287 | 417 | 122 | 2,678 | 771 | 39 | 10 | 11 | 23 | 6 |
| 7 | Metal ores | 928 | 827 | 128 | 1,850 | 54 | 808 | 29 | 449 | 217 | 744 | 7 | 4 | 2 | 0 |
| 8 | Other mining | 8,269 | 82 | 53 | 15 | 7 | 8 | 4 | 84 | 59 | 6,734 | 3 | 3 | 1 | 2 |
| 9 | Mining support | 614 | 436 | 236 | 184 | 72 | 117 | 49 | 294 | 208 | 492 | 7 | 2 | −1 | −4 |
| 10 | Food | 52,966 | 337 | 74 | 136 | 15 | 42 | 4 | 86 | 44 | 43,208 | 4 | 4 | 0 | 1 |
| 11 | Beverages | 6,239 | 312 | 83 | 173 | 12 | 74 | 4 | 142 | 45 | 5,208 | 4 | 3 | 0 | 1 |
| 12 | Tobacco | 559 | 609 | 138 | 760 | 51 | 333 | 18 | 236 | 107 | 485 | −4 | 0 | −3 | −3 |
| 13 | Textiles | 20,244 | 203 | 78 | 35 | 8 | 13 | 3 | 51 | 36 | 16,792 | 2 | 2 | 0 | 1 |
| 14 | Wearing | 53,903 | 102 | 54 | 9 | 2 | 4 | 1 | 27 | 16 | 41,119 | 3 | 1 | 2 | 1 |
| 15 | Leather | 26,843 | 195 | 68 | 24 | 3 | 10 | 2 | 36 | 25 | 20,874 | 1 | 0 | 1 | 0 |
| 16 | Wood Manufacturing | 20,122 | 109 | 57 | 15 | 3 | 6 | 2 | 41 | 27 | 15,599 | 0 | 0 | −1 | 0 |
| 17 | Paper | 12,544 | 191 | 75 | 85 | 10 | 36 | 3 | 80 | 48 | 10,490 | 3 | 3 | 1 | 1 |
| 18 | Printing | 5,458 | 101 | 51 | 25 | 5 | 14 | 3 | 91 | 53 | 4,188 | 3 | 2 | 2 | 1 |
| 19 | Refined petroleum | 3,178 | 735 | 217 | 1,157 | 76 | 617 | 24 | 177 | 103 | 2,777 | 6 | 5 | 0 | 0 |
| 20 | Chemicals | 20,382 | 184 | 75 | 184 | 24 | 53 | 8 | 185 | 98 | 17,234 | 3 | 3 | −1 | −1 |
| 21 | Pharmaceutical | 4,544 | 300 | 118 | 142 | 24 | 70 | 12 | 153 | 104 | 3,991 | 5 | 4 | 0 | 1 |
| 22 | Rubber and plastic | 33,526 | 133 | 66 | 40 | 10 | 15 | 3 | 75 | 52 | 27,502 | 2 | 2 | −1 | −1 |
| 23 | Other non-metallic | 34,731 | 115 | 55 | 34 | 3 | 15 | 1 | 58 | 26 | 28,330 | 2 | 2 | 0 | 0 |
| 24 | Basic metals | 10,546 | 310 | 85 | 264 | 20 | 92 | 6 | 123 | 70 | 8,883 | 3 | 3 | 0 | 0 |
| 25 | Fabricated metal | 37,951 | 116 | 59 | 25 | 6 | 10 | 3 | 63 | 45 | 30,410 | 4 | 3 | 1 | 1 |
| 26 | Computer and electronic | 10,128 | 242 | 83 | 125 | 14 | 38 | 6 | 102 | 66 | 8,372 | 4 | 4 | −1 | 1 |
| 27 | Electrical equipment | 12,729 | 235 | 76 | 88 | 12 | 31 | 5 | 86 | 60 | 10,643 | 3 | 3 | −1 | 0 |
| 28 | Machinery | 30,780 | 154 | 68 | 55 | 12 | 21 | 6 | 105 | 77 | 25,582 | 2 | 2 | −1 | −1 |
| 29 | Motor vehicles | 16,512 | 402 | 87 | 249 | 14 | 77 | 6 | 90 | 63 | 14,025 | 3 | 2 | 0 | −1 |
| 30 | Other transport | 3,402 | 333 | 85 | 131 | 9 | 46 | 4 | 71 | 42 | 2,769 | 7 | 7 | 1 | 3 |
| 31 | Furniture | 22,672 | 107 | 59 | 15 | 4 | 6 | 2 | 39 | 27 | 18,047 | 3 | 3 | 1 | 2 |
| 32 | Other manufacturing | 11,356 | 116 | 59 | 19 | 4 | 9 | 2 | 56 | 35 | 9,117 | 3 | 3 | 2 | 1 |
| 33 | Repair of machinery | 6,275 | 147 | 60 | 26 | 5 | 12 | 3 | 69 | 49 | 4,358 | 5 | 4 | 4 | 2 |
| | Total Manufacturing | 467,695 | 189 | 65 | 81 | 7 | 30 | 3 | 75 | 40 | 378,218 | 3 | 2 | 0 | 0 |

Source: Our elaboration. Revenue values are presented in BRL 1M (millions of reais). Number of workers are in units of headcount. Variations are in percentage points (%). Productivity values are presented in BRL 1K (thousands of reais), and are calculated at the firm level using value added divided by the number of employees.

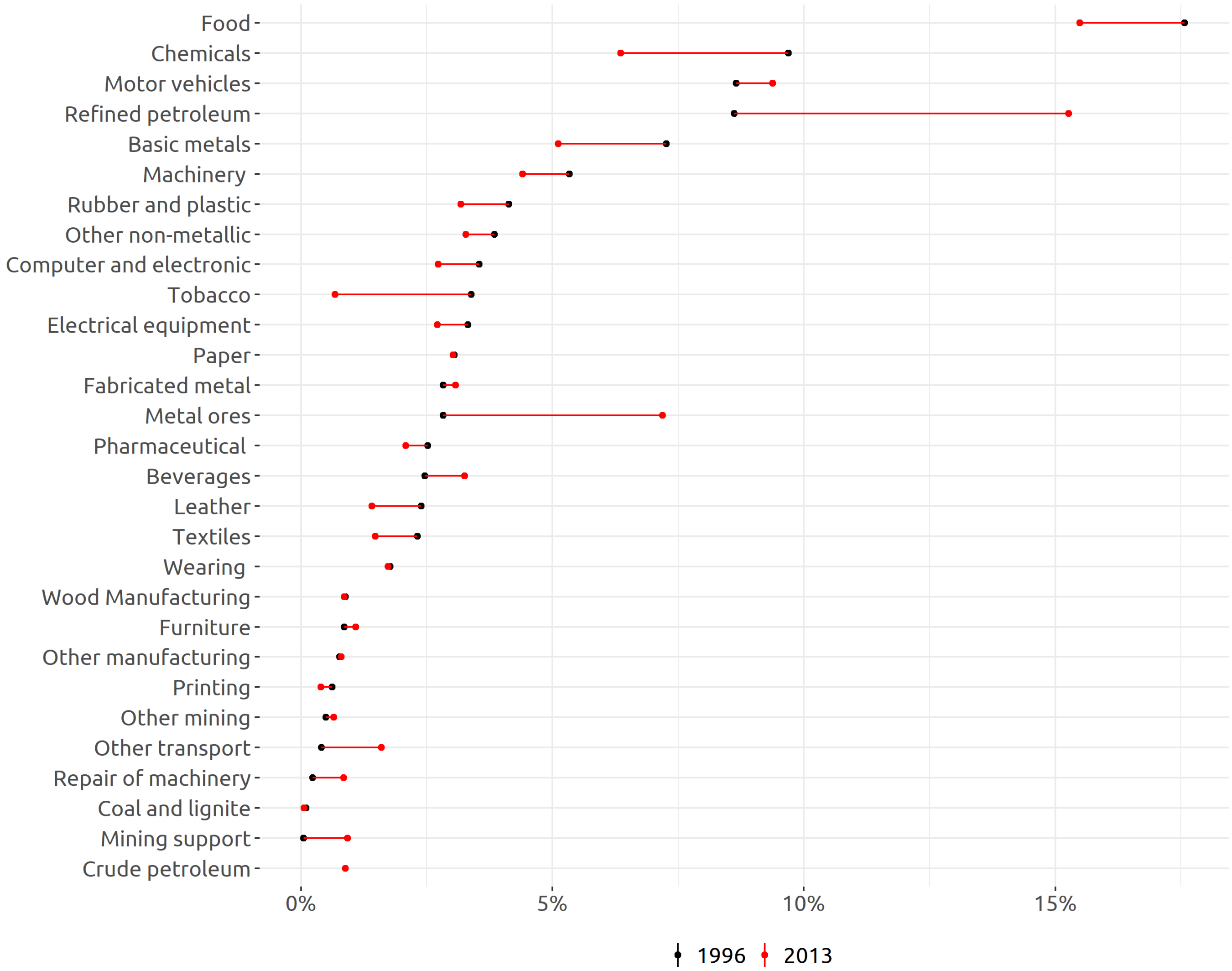


Figure 1: Decreasingly ranked shares in value added for each sector - 1996 and 2013.

In fact, several studies have already indicated the failure of economic policies to improve the capacities of the Brazilian industry (Negri and Cavalcante, 2014) and the low technological intensity demonstrated in most sectors (Negri and Cavalcante, 2015) is a cause of concern due to their consequence in wealth concentration and increased gap of income against developed countries (Hartmann et al., 2017).

# 3 Methodology

Our work presents empirical density distribution estimations with their parametric fitting on proxies for firm size, growth and efficiency. These exercises permits us to give a broad categorization of the Brazilian manufacturing industry and the heterogeneity of its performance metrics, while also allowing us to assess the evidence of more delimited stylized facts, such as the lognormal and Pareto distributions of firm size and the Laplacian distribution of firm productivity, growth and productivity rates.

The analysis is performed in two contexts: a) annual, where all the data sample from each year, regardless of the sector, is pooled; and b) sectoral, where data from all years are pooled by sector.

Due to space limitations, the visualization of the annual context is limited to three periods (1996, 2004 and 2013) and, in the sectoral context, to three two-digit sectors - mining of metal ores (07), manufacture of food products (10) and manufacture of motor vehicles, trailers and semi-trailers (29). These periods and sectors were deemed the most representative of the sample, considering differences in technological intensity, number of firms and share of value added in total manufacturing.

In relation to empirical kernel distributions, we explore the ones associated to the most important proxies for performance, size, and growth for Brazilian manufacturing, using the methodology exposed in works such as Silverman (1986), Tsybakov (2009) and Scott (2015). For all estimations, we use the Epanechnikov Kernel with bandwidth selected according to Silverman's rule of thumb (Silverman, 1986, pp. 48). In our study, we estimated the distributions for any given year or sector that have at least 300 observations. We also established 512 equally spaced bins for each distribution as a standard. Each density plot is accompanied by a normal distribution fit to serve as a benchmark. The normal fittings were made using maximum-likelihood estimation.

## Pareto Distributions

For parametric estimations of size distributions we conduct two exercises, one using the Hill estimator and the other using ordinary least squares (OLS) rank regressions.

Consider that the size distribution follows a cumulative distribution function (CDF) given by:[8]

$$F(x) = \text{Prob}(S \leq x_i) = 1 - \left(\frac{x_{\min}}{x_i}\right)^{\frac{1}{\gamma}} \tag{1}$$

where $\text{Prob}(S \leq x)$ represents the probability of a random value sampled from the distribution to be smaller than $x_i$, $\gamma$ represents the format of the tail, $S$ is a random variable, and $x_{\min}$ represents the smallest observation considered (i.e., the cut-off point from which the right tail of the distribution is modeled). Also, note that:

$$\gamma = \frac{1}{\alpha} \tag{2}$$

Then, making the $x_i$ decreasingly ranked, the Hill estimator (Hill, 1975) is defined as:

$$\hat{\gamma} = \frac{1}{n-1}\sum_{j=1}^{n}\ln(x_j) - \frac{n}{n-1}\ln(x_{\min}) \tag{3}$$

where $n$ represents the number of observations used until the cut-off point $x_{\min}$. Equation (3) includes a correction for small sample bias and constitutes a Maximum Likelihood estimator, being asymptotically normal and efficient with smooth distributions.[9]

Since the Hill estimator can be a poor estimator when the true distribution of the log-CDF is not linear[10] and since the point estimation is very dependent on the chosen cut-off point, we also did a log-rank regression using the whole distribution and a binned equipopulated empirical distribution for each plot.

Alternatively, it is possible to write a complementary cumulative distribution function (CCDF) of Equation (1) as:

$$R(x) = \text{Prob}(S > x) = \left(\frac{x_{\min}}{x_i}\right)^{\alpha} \tag{4}$$

Lower values of the parameter $\alpha$ are associated with more concentrated markets, since it makes the decrease of probability of finding firms up to size $x_i$ more steep. This means that there are more smaller firms but also that the few largest firms have most of the market share. Particularly, when $\alpha = 1$ Equation (4) is reduced to the so-called Zipf Law, a discrete distribution used to describe various physical and social phenomena (Newman, 2005). The $R(x)$ distribution can be estimated for a sample by:

$$\hat{R}(x) = \frac{j}{n} \tag{5}$$

[8]We follow the exposition adopted by Bottazzi et al. (2015).

[9]See (Bottazzi et al., 2015, footnote 6) for a discussion and list of references.

[10]Basically because of misspecification bias due to an incorrect functional form.

where $j$ represents the rank of the firm decreasingly ordered and $n$ represents the sample size. Equation (5) is an empirical survival function, or alternatively, a discrete complementary cumulative distribution function. By taking the log transformation on both sides, we have:

$$\log(\hat{R}(x)) = \hat{\alpha}\log(x_{\min}) - \hat{\alpha}\log(x_j) \tag{6}$$

with $\alpha\log(x_{\min})$ being the scale factor for the probability function to sum up to unity. In practice, we can use the ranking $j$ directly, since the number of observations does not affect the value of $\alpha$, as it is a constant:

$$\log(j) = \alpha\log(x_{\min}) + \log(n) - \alpha log(x) \tag{7}$$

A simple OLS regression can then estimate the parameters in this equation. This procedure is called an OLS rank regression.

Both methods deliver close point estimations for the true value of the coefficient $\alpha$ for the same $x_{\min}$ when $X \geq x_{\min}$ follows a power law (Bottazzi et al., 2015), but the Hill estimator is preferred because of its properties.

The exercises were conducted for the different contexts expressed at the beginning of this section. We were unable to use maximum likelihood methods to establish the optimal cut-off point of the Hill estimator initially, so we set the cut-off point on the 500th observation (Dosi et al., 2008). For the OLS rank regression, we used all the data in the respective context. Also, following Newman (2005) and Clauset et al. (2009), instead of reporting the value of $|\alpha|$ for the CCDF, we report the values of $|\alpha|+1$, as this gives the decay value of the probability density function distribution.

## Subbotin Fit

For parametric estimations of productivity, growth rates and productivity change distributions we use asymmetric exponential power (AEP) densities, a class of distributions introduced by Bottazzi et al. (2011) which belongs to the Subbotin Family of parametric fits (Subbotin, 1923). This distribution is composed by five parameters, which present both Laplacian and Gaussian distributions as special cases. Its functional form is:

$$f_{\mathrm{AEP}}(x; b_l, b_r, a_l, a_r, m) = \frac{1}{C} e^{\left(-\left[\frac{1}{b_l}\left|\frac{x-m}{a_l}\right|^{b_l}\theta(m-x) + \frac{1}{b_r}\left|\frac{x-m}{a_r}\right|^{b_r}\theta(x-m)\right)\right]} \tag{8}$$

with

$$C = \frac{a_l}{b_l}^{\frac{1}{bl}-1}\Gamma(1/b_l) + \frac{a_r}{b_r}^{\frac{1}{br}-1}\Gamma(1/b_r) \tag{9}$$

where $\theta(x)$ and $\Gamma(x)$ are, respectively, the Heaviside theta and the gamma function, $x$ represents the sample of the variable for which we want to estimate the parametric fit, $m$ is the sample average, $a_l$ and $a_r$ are the left and right scale parameters, respectively, and $b_l$ and $b_r$ are the shape parameters.

Specifically, when $b = 1$ the fit identifies a Laplacian distribution, and when $b = 2$ it turns to a normal distribution. The AEP allows each tail to be determined independently, and the lower the $b$, the fatter the tail. The parameters are estimated using maximum likelihood estimation, following Bottazzi et al. (2011).

# 4 Results

## Size Distributions

The objective of this section is to search for characteristic patterns in the market structure using different proxies for size. The existence of a specific shape may suggest a particular mechanism behind market organization. When this commonality is shared across countries, it creates favorable signs for the existence of processes that are of a purely economic nature, trespassing cultural and regional differences.

We begin by exploring the basic shape for the distribution of number of employees, total revenue and value added. These are the most common metrics for firm size and are what we commonly define as the "market" from the supply side. The distributions of these proxies are so extreme that it is not possible to have a meaningful visualization of their shapes as they are. Therefore, our plots will present the values from either the log of the variable or the log-rank version of CCDF.

Figure 2 shows the estimated probability density distributions for the natural logarithm of size proxies. The dotted line in each plot represents the fit of a normal distribution[11] using maximum likelihood estimations. Results are depicted for total manufacturing in three years of our sample. The shapes present a shift to the left in all metrics.

The total revenue shows evidence of the emergence of a bimodality, which we suspect to be caused by the introduction of a new tax regime privileging smaller firms[12]. This is the first time that we, as authors, see such a clear-cut effect of policy in the market structure, a fact that we pretend to investigate further.

These patterns, despite the evidence of bimodality, seem to follow the parametric distributions rather closely, with the worst case occurring when using the number of employees. When we move to a sectoral view, the apparent goodness of fit of these distributions seems to be improved, particularly for the monetary proxies, a result that contrasts with the literature. The European and US results tended to show that the apparent lognormal shape occurred as a consequence of sheer aggregation, exposed for example in Hymer and Pashigian (1962) for the UK and Bottazzi and Secchi (2003) for the US. While demonstrating the same fact for the Italian industry, this was the main argument of Bottazzi et al. (2007) to reduce the importance of the lognormal shape as a stylized fact. Instead, they emphasize a very skewed shape.

It is important to highlight that age was shown to have an essential role in these distributions. Cabral and Mata (2003) demonstrate that the distributions became less skewed when only old firms were considered. So, there is compelling evidence indicating that the entry-exit dynamic is responsible for the highly asymmetrical shape found in size distributions. Unfortunately, no accurate age-related data of the sample used is available at this time, so it is not possible to implement the more recent advances regarding Gibrat's law tests and age-split density probabilities of firm size distributions.

Figure 3 shows the distribution of these metrics in a log-rank plot[13]. In this plot, we compare both the lognormal (blue) and the Pareto (red) fits for each distribution. These plots show the right tail of the distribution on the top-left side, with the body and the left tail of the distribution concentrated in the bottom-right of each graph. The bimodality of the total revenue is not visible anymore.

Both fits seem very close for the data in the annual view. Especially when the number of employees is considered, the Pareto fit seems favored over the lognormal. Monetary values, on the other hand, display a more lognormal appearance, particularly in the body. However, it is also important to evaluate the robustness to disaggregation of these fits. Generally, the quality of the adjustment of sectoral values seems even better than in the aggregate case, with value added presenting an almost perfect fit of a lognormal distribution. As in the case with the annual view, the Pareto fit seems favored only for the number of employees. Different sectors and years also share similar inclinations.

[11] That means that the original distribution is fitted by a lognormal fit.

[12] The *simples*, a special tax regime that was implemented in the Complementary Law No. 123, from December 14th, 2006, introduces the option for firms under a specific constraint of revenue to be taxed by a fixed percentage of their sales. The limits, around R$ 2-4 million for the period, and the date of the law both coincide with the emergence of the bimodality.

[13] In this visualization, we plot the previous adaptation on the complementary cumulative distribution function, i.e., we took the logarithm of the decreasingly ranked firms and plotted it against the logarithm of the proxy used to measure size.

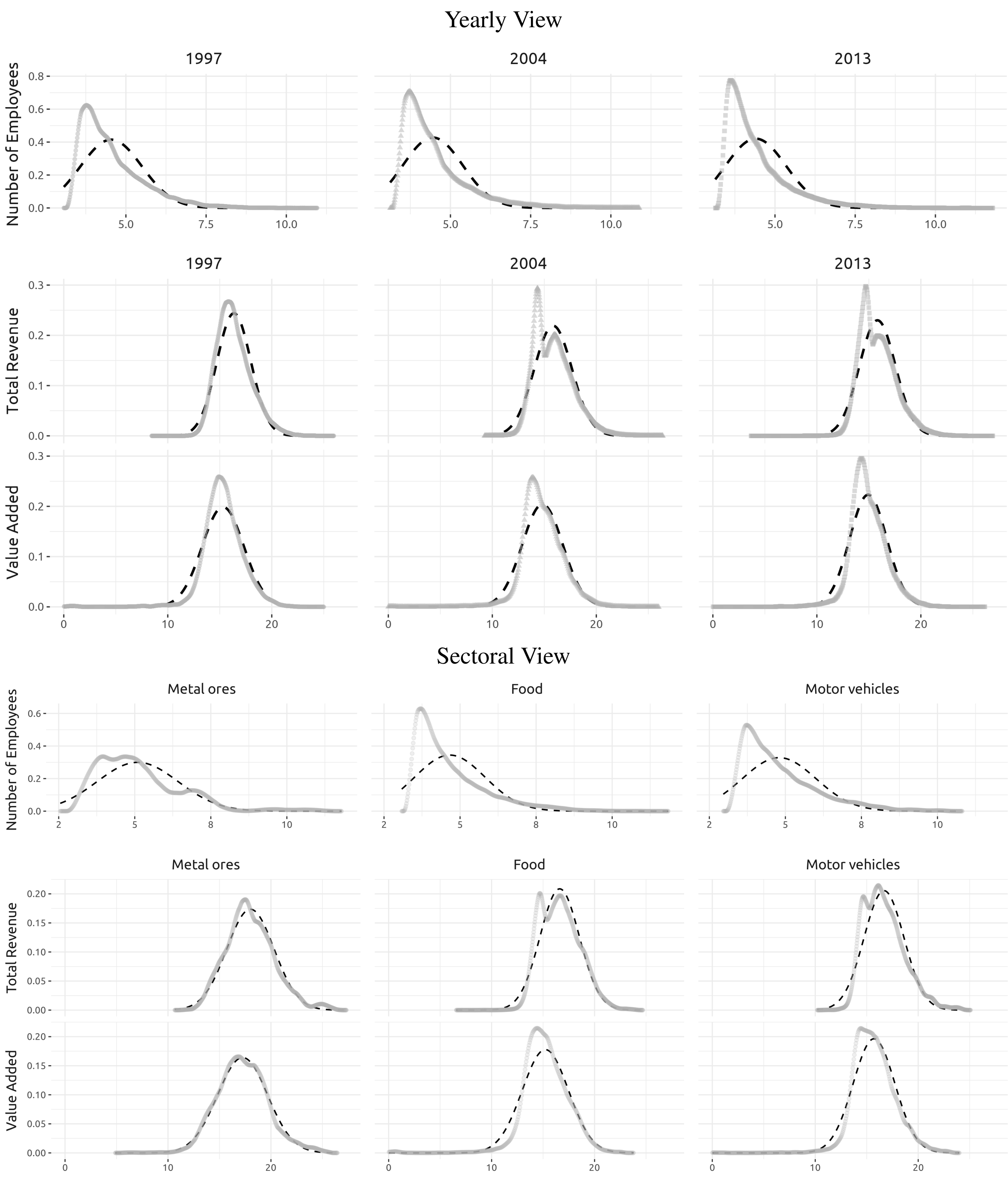


Figure 2: Size - annual and sectoral probability density plots. Variables in log, axes in level. Dashed lines represent a normal fit for each distribution.

In order to formally present this results, we proceed to report OLS rank and Hill estimations of the right tail of size distributions. For the Hill estimation, we considered the five hundred biggest firms in each context, whereas we used the whole distribution for the OLS rank regression. The estimations are presented in Table 2. The OLS rank regression showed great explanatory power of the model, in general over 90%, which we do not report in detail here. This result should be understood as the model being generally a good "fit" for the data rather than suggesting the superiority of any particular distribution.

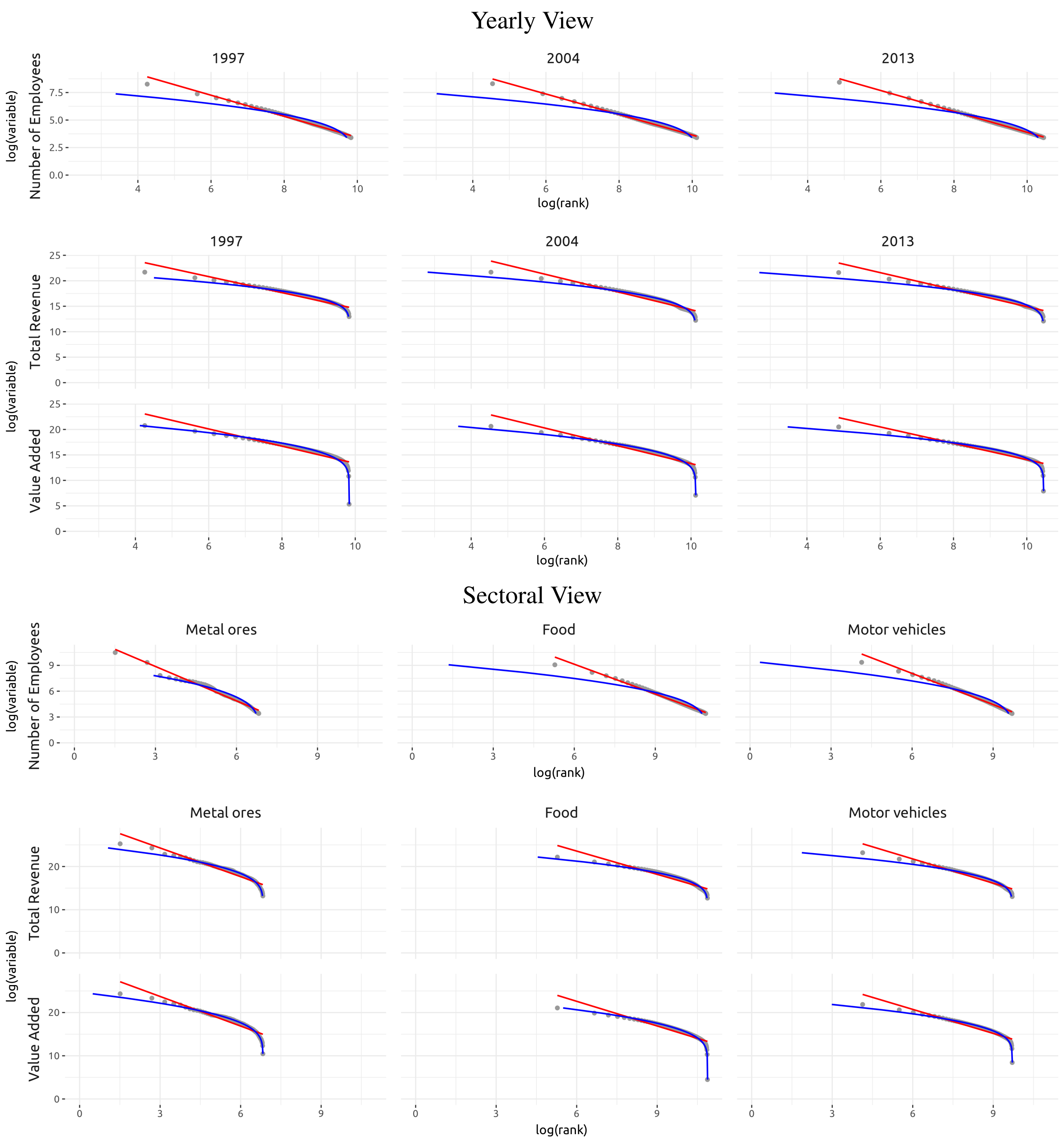


Figure 3: Size - annual and sectoral log-rank plots. The red line represents the Pareto fit from the OLS rank estimation, while the blue line represents the lognormal fit using maximum likelihood estimation.

More interestingly, sectors that present non-smooth formats or bimodalities are still very well represented by the model. Of all proxies, value added is the one with the "poorest" fit, which, as shown visually in the previous graphs, reflects the apparent superiority of the lognormal fit. A similar result was found by Dosi et al. (2008), regarding the evidence of a concavity. Yet, more investigation is still necessary to compare the goodness of fit of different parametric distributions with the use of formal tests (Clauset et al., 2009).

The OLS rank estimation coefficients vary from 1.62 to 2.33 for the number of workers, 2.11 to 3.28 for total revenue and 2.24 to 3.28 for value added. The same metrics using the ML Hill estimator provide coefficients that range from 1.60 to 3.07, 1.47 to 2.89, and 1.33 to 2.86, respectively.

The estimates obtained by the two methods are close for some sectors, but generally do not agree. This should come as no surprise. As they are both very precise, their differences are caused by disparities in the cut-off value, with the Hill method being susceptible to the choice of $x_{\min}$.

In general, our OLS rank estimates for total manufacturing are in accordance with the literature, with a coefficient of 1.94 when using total employees to measure size, whereas Axtell (2001) found a slope of 2.06 for the US. Our results for total revenue, nevertheless, are less agreeable. While Axtell (2001) found a slope of 1.99 and Dosi et al. (2008), using Italian and French firms, a range between 1.8 and 2.05, our results show coefficients for the OLS rank regression and Hill estimates equal to 2.7 and 2.45, respectively.

Size distributions, therefore, present significant right-skewed distributions regardless of the metric used, which closely resemble lognormal distributions for total revenue and value added and Pareto distributions (also known as power-law distributions) for the number of employees. This pattern seems robust to both different disaggregation levels and time frames.

The considerable heterogeneity evidenced by these distributions also corroborates the outcomes of other works for the Brazilian economy (Nogueira et al., 2014; Esteves, 2015; Squeff and Nogueira, 2015; Catela et al., 2015). This shows that even controlling by sector at diverse levels of disagreggation, the researcher still faces a very diverse set of enterprise characteristics.

We hypothesize that the explanation for this last fact may go beyond the usual suspects, such as gains from different scales of operation and access to better prices through suppliers. In addition to them, market niches and brand power could create earning differentials that would not be mitigated even if firms shared the same costs or technology (Sturgeon, 2002).

Also, the widespread heterogeneity in performance metrics found for other countries (Griliches and Regev, 1995; Bartelsman and Doms, 2000) indicates that this evidence is not a Brazilian peculiarity. Neither is it necessarily a problem in itself, although it generally carries along the side effects of wealth concentration, which is a growing cause of concern and instability in modern societies (Atkinson and Piketty, 2007).

From all that we know, heterogeneity may very well be a feature of capitalist societies that is still poorly understood. When the interpretation of the market goes beyond basic price mechanisms (which we explore better on the section of performance), other hypotheses about customer behavior and market formation start to make sense.

Hierarchies can constitute a more natural way to transmit signaling information, helping to organize markets (Krugman, 1996). The idea that markets are intelligent and self-organized goes as back as Hayek (1945), but the atomized information that the market contains is not necessarily optimal under a set with uniform agents. Studies from network theory show that networks following power laws are very robust to random shocks (or, in our case, bankruptcy, for example) since there are few large hubs and many small components. So, if stability is a required feature for a high-functioning economy (at least, it seems to be present in most developed countries), it may as well be a required characteristic of any network that is candidate to describe the connection between customers and enterprises.

That is because customers expect to find goods and services available when they search for them, and if the markets do not have a minimal degree of predictability, they cannot meet their expectations. Market turnover rates would have to be much lower in a uniform network than in a power-law distributed one for bankruptcy not to affect a customer. Given the empirical failure rates found in most countries, some kind of hub-based distribution seems to be our best guess to describe these interactions, since it both prevents customers from having to spend time creating a wide number of relationships with different firms (such as in a uniform distribution) and creating a more predictable environment at the same time; only very rarely would a customer miss a product that would have been on the shelf some time ago.

Therefore, if size is a good proxy for the number of transactions that a firm carries, and then, a proxy for the number of other individuals a firm is connected to, then this heterogeneity may imply some kind of power law or a similarly very skewed robust network of producers and consumers.

Table 2: Pareto coefficients from OLS rank regression and Hill estimations for firm size in Brazilian manufacturing - cross-sectional data

| ISIC | Industry | Number Workers | | | | Total Revenue | | | | Value Added | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | OLS-Rank | Hill | | | OLS-Rank | Hill | | | OLS-Rank | Hill | | |
| | | $\alpha+1$ | $\alpha+1$ | 95% Interval | | $\alpha+1$ | $\alpha+1$ | 95% Interval | | $\alpha+1$ | $\alpha+1$ | 95% Interval | |
| 7 | Metal ores | $2.33^{***}$ | 1.79 | 1.72 | 1.86 | $3.21^{***}$ | 1.47 | 1.43 | 1.52 | $3.28^{***}$ | 1.48 | 1.44 | 1.53 |
| 8 | Other mining | $1.62^{***}$ | 2.73 | 2.58 | 2.89 | $2.11^{***}$ | 2.40 | 2.28 | 2.53 | $2.24^{***}$ | 2.31 | 2.19 | 2.42 |
| 9 | Mining support | $2.06^{***}$ | 1.64 | 1.58 | 1.70 | $2.25^{***}$ | 1.67 | 1.61 | 1.73 | $2.3^{***}$ | 1.57 | 1.52 | 1.62 |
| 10 | Food | $2.15^{***}$ | 2.58 | 2.44 | 2.72 | $2.8^{***}$ | 2.24 | 2.13 | 2.35 | $2.9^{***}$ | 2.17 | 2.07 | 2.28 |
| 11 | Beverages | $2.13^{***}$ | 2.05 | 1.96 | 2.15 | $2.91^{***}$ | 1.94 | 1.86 | 2.02 | $3.14^{***}$ | 1.86 | 1.78 | 1.94 |
| 12 | Tobacco | $2.32^{***}$ | 1.60 | 1.55 | 1.65 | $3.22^{***}$ | 1.30 | 1.28 | 1.33 | $3.24^{***}$ | 1.33 | 1.30 | 1.35 |
| 13 | Textiles | $1.97^{***}$ | 2.71 | 2.56 | 2.86 | $2.53^{***}$ | 2.78 | 2.62 | 2.93 | $2.53^{***}$ | 2.62 | 2.48 | 2.76 |
| 14 | Wearing | $1.7^{***}$ | 2.39 | 2.27 | 2.51 | $2.31^{***}$ | 2.52 | 2.38 | 2.65 | $2.31^{***}$ | 2.43 | 2.30 | 2.56 |
| 15 | Leather | $1.91^{***}$ | 2.18 | 2.08 | 2.29 | $2.52^{***}$ | 2.38 | 2.26 | 2.50 | $2.38^{***}$ | 2.26 | 2.15 | 2.37 |
| 16 | Wood Manufacturing | $1.76^{***}$ | 2.95 | 2.78 | 3.12 | $2.28^{***}$ | 2.21 | 2.10 | 2.31 | $2.36^{***}$ | 2.24 | 2.13 | 2.35 |
| 17 | Paper | $1.95^{***}$ | 2.57 | 2.43 | 2.71 | $2.61^{***}$ | 1.90 | 1.82 | 1.98 | $2.73^{***}$ | 1.81 | 1.74 | 1.88 |
| 18 | Printing | $1.73^{***}$ | 2.44 | 2.31 | 2.56 | $2.45^{***}$ | 2.03 | 1.94 | 2.12 | $2.42^{***}$ | 2.06 | 1.96 | 2.15 |
| 19 | Refined petroleum | $2.16^{***}$ | 2.42 | 2.29 | 2.54 | $2.3^{***}$ | 2.17 | 2.07 | 2.27 | $2.46^{***}$ | 2.14 | 2.04 | 2.24 |
| 20 | Chemicals | $1.93^{***}$ | 2.58 | 2.45 | 2.72 | $2.75^{***}$ | 2.45 | 2.33 | 2.58 | $2.74^{***}$ | 2.44 | 2.32 | 2.57 |
| 21 | Pharmaceutical | $2.07^{***}$ | 3.02 | 2.84 | 3.19 | $2.69^{***}$ | 2.49 | 2.36 | 2.62 | $2.72^{***}$ | 2.46 | 2.33 | 2.58 |
| 22 | Rubber and plastic | $1.81^{***}$ | 2.52 | 2.38 | 2.65 | $2.37^{***}$ | 2.31 | 2.20 | 2.42 | $2.44^{***}$ | 2.26 | 2.15 | 2.37 |
| 23 | Other non-metallic | $1.78^{***}$ | 2.81 | 2.65 | 2.97 | $2.62^{***}$ | 2.47 | 2.35 | 2.60 | $2.58^{***}$ | 2.43 | 2.31 | 2.56 |
| 24 | Basic metals | $2.1^{***}$ | 2.23 | 2.12 | 2.34 | $2.9^{***}$ | 1.90 | 1.82 | 1.98 | $2.88^{***}$ | 1.85 | 1.77 | 1.92 |
| 25 | Fabricated metal | $1.77^{***}$ | 2.89 | 2.73 | 3.06 | $2.34^{***}$ | 2.51 | 2.37 | 2.64 | $2.36^{***}$ | 2.59 | 2.45 | 2.72 |
| 26 | Computer and electronic | $2.05^{***}$ | 2.58 | 2.44 | 2.72 | $2.73^{***}$ | 2.01 | 1.92 | 2.10 | $2.67^{***}$ | 2.11 | 2.01 | 2.21 |
| 27 | Electrical equipment | $2.01^{***}$ | 2.32 | 2.21 | 2.44 | $2.59^{***}$ | 2.26 | 2.15 | 2.37 | $2.59^{***}$ | 2.18 | 2.08 | 2.28 |
| 28 | Machinery | $1.86^{***}$ | 2.63 | 2.49 | 2.77 | $2.38^{***}$ | 2.52 | 2.39 | 2.65 | $2.36^{***}$ | 2.48 | 2.35 | 2.61 |
| 29 | Motor vehicles | $2.21^{***}$ | 2.31 | 2.20 | 2.43 | $2.87^{***}$ | 1.95 | 1.86 | 2.03 | $2.85^{***}$ | 1.96 | 1.88 | 2.04 |
| 30 | Other transport | $2.13^{***}$ | 2.06 | 1.97 | 2.16 | $2.85^{***}$ | 1.85 | 1.77 | 1.92 | $2.89^{***}$ | 1.88 | 1.80 | 1.95 |
| 31 | Furniture | $1.74^{***}$ | 3.07 | 2.89 | 3.25 | $2.3^{***}$ | 2.89 | 2.72 | 3.06 | $2.42^{***}$ | 2.86 | 2.70 | 3.03 |
| 32 | Other manufacturing | $1.78^{***}$ | 2.69 | 2.54 | 2.84 | $2.36^{***}$ | 2.45 | 2.33 | 2.58 | $2.43^{***}$ | 2.40 | 2.28 | 2.53 |
| 33 | Repair of machinery | $1.87^{***}$ | 2.31 | 2.19 | 2.42 | $2.33^{***}$ | 1.91 | 1.83 | 1.99 | $2.28^{***}$ | 2.01 | 1.92 | 2.10 |
| | Total Manufacturing | $1.94^{***}$ | 2.91 | 2.74 | 3.07 | $2.7^{***}$ | 2.45 | 2.33 | 2.58 | $2.68^{***}$ | 2.37 | 2.25 | 2.49 |

Source: Our elaboration. Stars represent significance at the 1% level. $\alpha+1$ refers to the inclination of the PDF, as reported in Newman (2005) and Clauset et al. (2009), while $\alpha$ represents the inclination of the CDF.

Another advantage is that such kind of market organization reduces the distance between agents (Barabasi, 2016). The "small world" effect is much more pronounced in networks that have hubs, which makes the number of connections needed for any two enterprises to interact to be usually in the single-digit stance. So, collaborations, at least from the perspective of small businesses, may be more clear. They know who the key players are, and probably what they are looking for.

At the same time, those networks replicate other aspects of real economies, such as crisis behavior. Hub-based networks are much more fragile against targeted failures, or meltdowns of important players, which is usually understood in economics under the concept of "too big to fail" (Nurisso and Prescott, 2017), popularized in the post-2008 crisis after the bailouts of several financial and industrial firms. If those big firms are more inclined, by necessity, to trade and lend among themselves, a failure of an important node starts a contagion mechanism that spreads quickly through the whole network, and forces government intervention.

## Productivity Distributions

We repeat the previous exercise for productivity, which we consider to be the most critical metric of fitness and performance, and acts as the primary mechanism of survival in evolutionary theories, forming the "replicator dynamics" of models such as Metcalfe (1994).

Table 3: Subbotin (AEP) Coefficients for Productivity in Brazilian Manufacturing - Cross-Sectional Data

| ISIC | Industry | Labor Productivity | | | |
|---|---|---|---|---|---|
| | | $b_l$ | $\sigma(b_l)$ | $b_r$ | $\sigma(b_r)$ |
| 7 | Metal ores | NA | NA | NA | NA |
| 8 | Other mining | 0.87 | ( 0.02) | 2.57 | ( 0.10) |
| 9 | Mining support | NA | NA | NA | NA |
| 10 | Food | 0.55 | ( 0.01) | 3.35 | ( 0.04) |
| 11 | Beverages | 0.69 | ( 0.02) | 2.06 | ( 0.07) |
| 12 | Tobacco | NA | NA | NA | NA |
| 13 | Textiles | 0.62 | ( 0.01) | 3.58 | ( 0.08) |
| 14 | Wearing | 0.56 | ( 0.00) | 2.14 | ( 0.02) |
| 15 | Leather | 0.62 | ( 0.01) | 2.38 | ( 0.04) |
| 16 | Wood Manufacturing | 0.63 | ( 0.01) | 2.56 | ( 0.05) |
| 17 | Paper | 0.71 | ( 0.01) | 2.11 | ( 0.05) |
| 18 | Printing | 0.76 | ( 0.03) | 2.78 | ( 0.11) |
| 19 | Refined petroleum | 0.83 | ( 0.03) | 1.82 | ( 0.10) |
| 20 | Chemicals | 1.06 | ( 0.02) | 2.55 | ( 0.07) |
| 21 | Pharmaceutical | NA | NA | NA | NA |
| 22 | Rubber and plastic | 0.80 | ( 0.01) | 2.18 | ( 0.04) |
| 23 | Other non-metallic | 0.58 | ( 0.01) | 2.35 | ( 0.03) |
| 24 | Basic metals | 0.79 | ( 0.02) | 2.65 | ( 0.08) |
| 25 | Fabricated metal | 0.75 | ( 0.01) | 2.49 | ( 0.04) |
| 26 | Computer and electronic | 0.79 | ( 0.02) | 2.23 | ( 0.07) |
| 27 | Electrical equipment | 0.84 | ( 0.02) | 2.42 | ( 0.07) |
| 28 | Machinery | 0.99 | ( 0.02) | 1.92 | ( 0.04) |
| 29 | Motor vehicles | 0.87 | ( 0.02) | 2.29 | ( 0.06) |
| 30 | Other transport | 0.72 | ( 0.03) | 3.11 | ( 0.17) |
| 31 | Furniture | 0.57 | ( 0.01) | 3.17 | ( 0.06) |
| 32 | Other manufacturing | 0.61 | ( 0.01) | 3.25 | ( 0.09) |
| 33 | Repair of machinery | 0.85 | ( 0.03) | 1.79 | ( 0.07) |
| | Total Manufacturing | 0.63 | ( 0.00) | 3.05 | ( 0.01) |

Source: Our elaboration. $b_l$ and $b_r$ represents the left and right tail, respectively, while $\sigma(b)$ represents the standard deviation of the estimated parameters.

Following Dosi et al. (2012); Yu et al. (2015b); Mathew (2017), we proceed to test the parametric fit of the productivity distributions using the asymmetric exponential power distribution (AEP). The AEP distribution, introduced by Bottazzi et al. (2011), belongs to a family of distributions started by Subbotin (1923), which assumes a normal or Laplacian shape accordingly with the values of the $b$ coefficients used, with values $b = 1$ generating a Laplacian, and values $b = 2$ generating a normal distribution. This distribution estimates the values of $b$ for each tail independently, so $b_l$ represents the coefficient for the left tail, while $b_r$ represents the right one.

We estimate the fit of these parameters for the natural logarithm of productivity for each sector, which in turn produces lognormal and log-Laplacian fits. We used a maximum likelihood method, but we were unable to achieve convergence for all sectors. The results are detailed in Table 3.

Somewhat more intensively than expected, the AEP estimation reveals tails significantly fatter on the left side (notably for the food sector). They are even fatter than what a log-Laplacian distribution would produce, and the estimates are smaller than the ones found for China and Italy (Yu et al., 2015b; Dosi et al., 2012). Following the international results, the right side presents a steeper decline, very close to a lognormal distribution, with few exceptions.

Figure 4 shows the distributions of (log) productivity with the parametric fits of (log) normal and (log) AEP fits. The AEP fits seem rather good and superior to the one produced by a (log) normal. They also seem very robust to different periods and sectors.

The overall picture provides supporting evidence from what Dosi et al. (2012) called an "efficiency frontier". Firms that are at the top of productivity in their sectors face constraints that are technological in their nature, which in turn create barriers for increases in productivity that are similar for all leaders, with far fewer outliers. Firms are more widely dispersed at the "bottom" of productivity, since their survival may be more attached to spatial or contextual advantages. Alternatively, their low productivity may reflect not low physical productivity *per se*, but a low capacity to capture market earnings and their adverse positioning in their production network (Sturgeon, 2002; Gereffi et al., 2005) - particularly if they are producing for intermediate consumption, which may see them subjugated by the monopsonistic power of the leading firms.

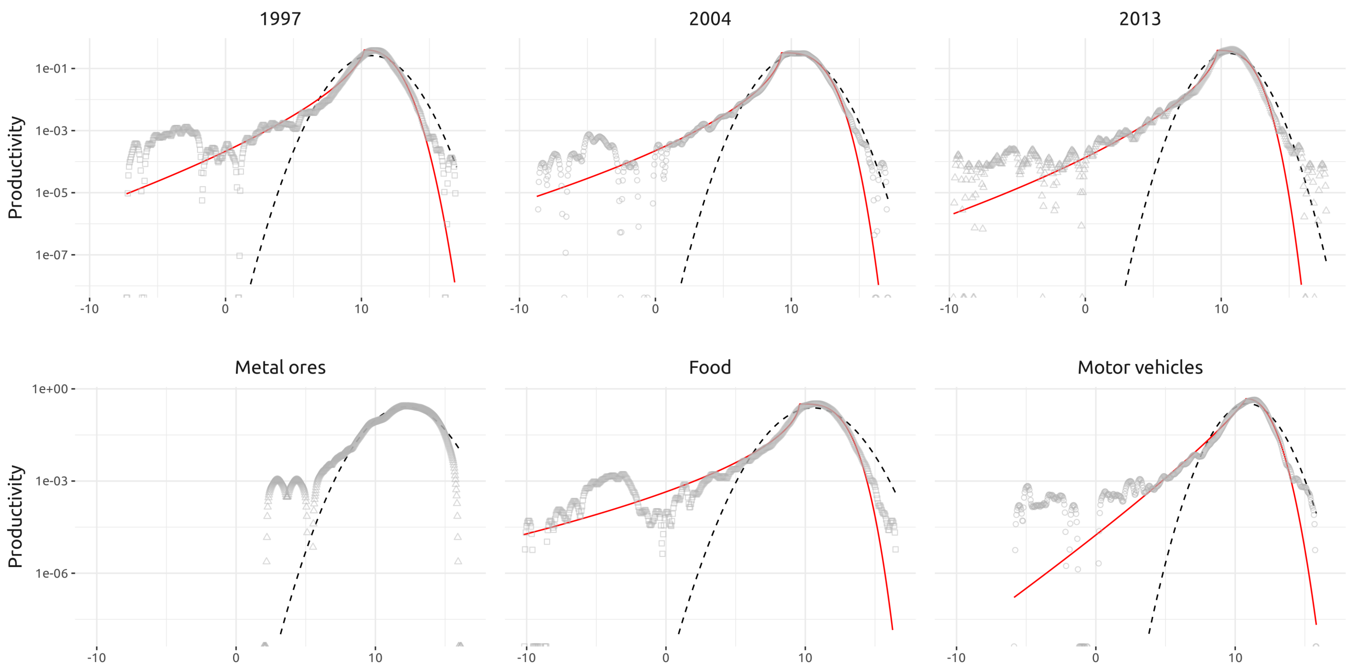


Figure 4: Log labor productivity - probability density plots. Dashed lines represent a normal fit for each distribution, while the red lines represent the AEP fit.

## Rate Distributions

In this section, we analyze the nature of the distributions of firm growth and productivity change rates. These variables are fundamental to understand the economic process as they are the power that shapes the markets. In fact, there is no capitalism without dynamics. To understand them, thus, is to understand how markets and consumers interact to decide who will be chosen to produce and what will keep being produced.

It is remarkably intriguing that this process would ever assume any particular shape. There is no apparent reason why dynamics should have to follow a specific mechanism or be modeled by simple stochastic principles. However, as imaginative as nature is, such patterns do exist and have now been found for several countries.

The results for the AEP estimates are presented in Table 4. They show that distributions on growth rates and productivity change are fairly symmetrical for most sectors, with all the $b$ estimates close to one or below, indicating tails that are at least Laplacian. Growth rates present values that are comparable with the ones from India (Mathew, 2017) and lower than the ones found for Italy and the US (Bottazzi and Secchi, 2003; Bottazzi et al., 2007), while productivity change presents values lower than those found for China (Yu et al., 2015b).

Table 4: Subbotin (AEP) Coefficients for Firm Growth and Productivity Change in Brazilian Manufacturing - Cross-Sectional Data

| ISIC | Industry | Δ% Tot. Rev. | | | | Δ% Prod. | | | |
|---|---|---|---|---|---|---|---|---|---|
| | | $b_l$ | $\sigma(b_l)$ | $b_r$ | $\sigma(b_r)$ | $b_l$ | $\sigma(b_l)$ | $b_r$ | $\sigma(b_r)$ |
| 7 | Metal ores | 0.77 | (0.06) | 0.80 | (0.06) | 0.65 | (0.05) | 0.82 | (0.06) |
| 8 | Other mining | 0.97 | (0.03) | 0.99 | (0.03) | 0.62 | (0.01) | 0.63 | (0.02) |
| 9 | Mining support | 0.92 | (0.12) | 1.21 | (0.14) | 0.92 | (0.10) | 0.84 | (0.09) |
| 10 | Food | 0.72 | (0.01) | 0.78 | (0.01) | 0.51 | (0.00) | 0.57 | (0.01) |
| 11 | Beverages | 0.86 | (0.03) | 0.85 | (0.03) | 0.50 | (0.01) | 0.51 | (0.01) |
| 12 | Tobacco | 0.67 | (0.06) | 0.80 | (0.08) | 0.92 | (0.10) | 0.80 | (0.08) |
| 13 | Textiles | 0.88 | (0.02) | 0.88 | (0.02) | 0.54 | (0.01) | 0.59 | (0.01) |
| 14 | Wearing | 0.78 | (0.01) | 0.91 | (0.01) | 0.54 | (0.00) | 0.61 | (0.01) |
| 15 | Leather | 0.87 | (0.01) | 0.90 | (0.01) | 0.63 | (0.01) | 0.68 | (0.01) |
| 16 | Wood Manufacturing | 0.86 | (0.02) | 0.92 | (0.02) | 0.66 | (0.01) | 0.69 | (0.01) |
| 17 | Paper | 0.81 | (0.02) | 0.82 | (0.02) | 0.56 | (0.01) | 0.52 | (0.01) |
| 18 | Printing | 0.85 | (0.03) | 0.79 | (0.03) | 0.72 | (0.02) | 0.62 | (0.02) |
| 19 | Refined petroleum | 0.95 | (0.04) | 0.77 | (0.03) | 0.64 | (0.02) | 0.66 | (0.03) |
| 20 | Chemicals | 0.82 | (0.01) | 0.85 | (0.01) | 0.60 | (0.01) | 0.66 | (0.01) |
| 21 | Pharmaceutical | 0.84 | (0.03) | 0.83 | (0.03) | 0.58 | (0.02) | 0.70 | (0.02) |
| 22 | Rubber and plastic | 0.81 | (0.01) | 0.85 | (0.01) | 0.55 | (0.01) | 0.55 | (0.01) |
| 23 | Other non-metallic | 0.90 | (0.01) | 0.91 | (0.01) | 0.65 | (0.01) | 0.69 | (0.01) |
| 24 | Basic metals | 0.96 | (0.03) | 0.83 | (0.02) | 0.62 | (0.01) | 0.67 | (0.01) |
| 25 | Fabricated metal | 0.81 | (0.01) | 0.92 | (0.01) | 0.57 | (0.01) | 0.65 | (0.01) |
| 26 | Computer and electronic | 0.89 | (0.02) | 0.82 | (0.02) | 0.61 | (0.01) | 0.60 | (0.01) |
| 27 | Electrical equipment | 0.81 | (0.02) | 0.99 | (0.02) | 0.59 | (0.01) | 0.65 | (0.01) |
| 28 | Machinery | 0.90 | (0.01) | 0.97 | (0.02) | 0.67 | (0.01) | 0.69 | (0.01) |
| 29 | Motor vehicles | 0.85 | (0.02) | 0.97 | (0.02) | 0.59 | (0.01) | 0.58 | (0.01) |
| 30 | Other transport | 0.72 | (0.03) | 0.79 | (0.03) | 0.57 | (0.02) | 0.71 | (0.03) |
| 31 | Furniture | 0.82 | (0.01) | 1.01 | (0.02) | 0.51 | (0.01) | 0.59 | (0.01) |
| 32 | Other manufacturing | 0.86 | (0.02) | 0.88 | (0.02) | 0.58 | (0.01) | 0.66 | (0.01) |
| 33 | Repair of machinery | 0.84 | (0.03) | 0.90 | (0.03) | 0.66 | (0.02) | 0.64 | (0.02) |
| | Total Manufacturing | 0.82 | (0.00) | 0.88 | (0.00) | 0.57 | (0.00) | 0.61 | (0.00) |

Source: Our elaboration. $b_l$ and $b_r$ represents the left and right tail, respectively, while $\sigma(b)$ represents the standard deviation of the estimated parameters.

Figure 5 presents the distribution of growth rates and productivity change for three years (notice the log-transformation in the vertical axis), together with both AEP and normal fits. The graph for each period and proxy shows a very clear, "tent-like" shape. Also, note the poor fit of the normal distribution to describe the tails. The normal fit falls much faster than the empirical rates, which demonstrates that infrequent events of extreme impact are much more “common” than it would be expected under normality. The bottom side shows the same plot for three different sectors. A similar tent-like shape as before is found, proving this pattern to be robust under disaggregation. Specifically, productivity change for the food sector (ISIC 10) presents some symmetrical inflections at both ends of the distribution that deserve more investigation. Overall, the tent-like shape is very solid and characterizes a Laplacian curve.

Bottazzi and Secchi (2006) explored this phenomenon in detail in a model following Simon’s tradition of “islands of opportunity”(Ijiri and Simon, 1977). If we suppose that there is a limited availability of growth episodes available for firms and that the ones that have taken these opportunities in the past have more chance of winning them in the future, thus generating a path-dependent mechanism of competition, then the model can reproduce this Laplacian shape asymptotically.

At the same time, these distributions offer an interesting contrast with some notions from innovation theory. First, the notion of capabilities, which are incremental in the sense that they are hard to obtain and must be accumulated and built upon, thus constituting the core of value generated by firms (Penrose, 1959; Malerba, 1992; Teece et al., 1997; Gereffi et al., 2005), with learning by doing being a primary factor (Arrow, 1962). Second, the idea of technological trajectories, which are mostly subject to periods of incremental improvement with discontinuities following structural breaks due to radical or disruptive innovation (Dosi, 1982; Dosi and Nelson, 2010). These two concepts, together, would make one expect fairly smooth periods of incremental perfecting followed by large leaps of rapid growth due to paradigm changes.

Instead, the shape of growth or productivity change rate distributions is continuously bombarded by a process that generates extreme, symmetrical events. It sounds implausible that in all these cases some disruptive innovation is happening for a few, and not necessarily the same, enterprises all the time, notably for sectors that are already mature or stagnate. Stochastic and simpler models as the ones proposed by Bottazzi and Secchi (2006) seem closer to the empirical data.

This, of course, does not disavow any theory of incremental innovation or continuous improvement, but suggests that there are essential middle steps between what configures learning in the sense of technological advancement and organizational management and what in fact generates financial returns, the latter being somewhat more extreme in its deviation and, at the same time, relatively constant in its nature. Increases in physical productivity do not necessarily translate into increased monetary productivity, and quality change does not imply sales growth. Especially with highly standardized products, a lot of these gains became customer surplus, e.g., transistors and steel production (Dosi and Nelson, 2010).

Thus, while we are obviously not disagreeing with the idea that physical productivity and technology change play an important role in monetary growth and monetary productivity change, these theories must be adapted to faithfully incorporate the kind of short-term competition and the network nature of markets, such as to define who is more probably to take the gains of innovation: consumers, leaders or innovators. (Gereffi et al., 2005).

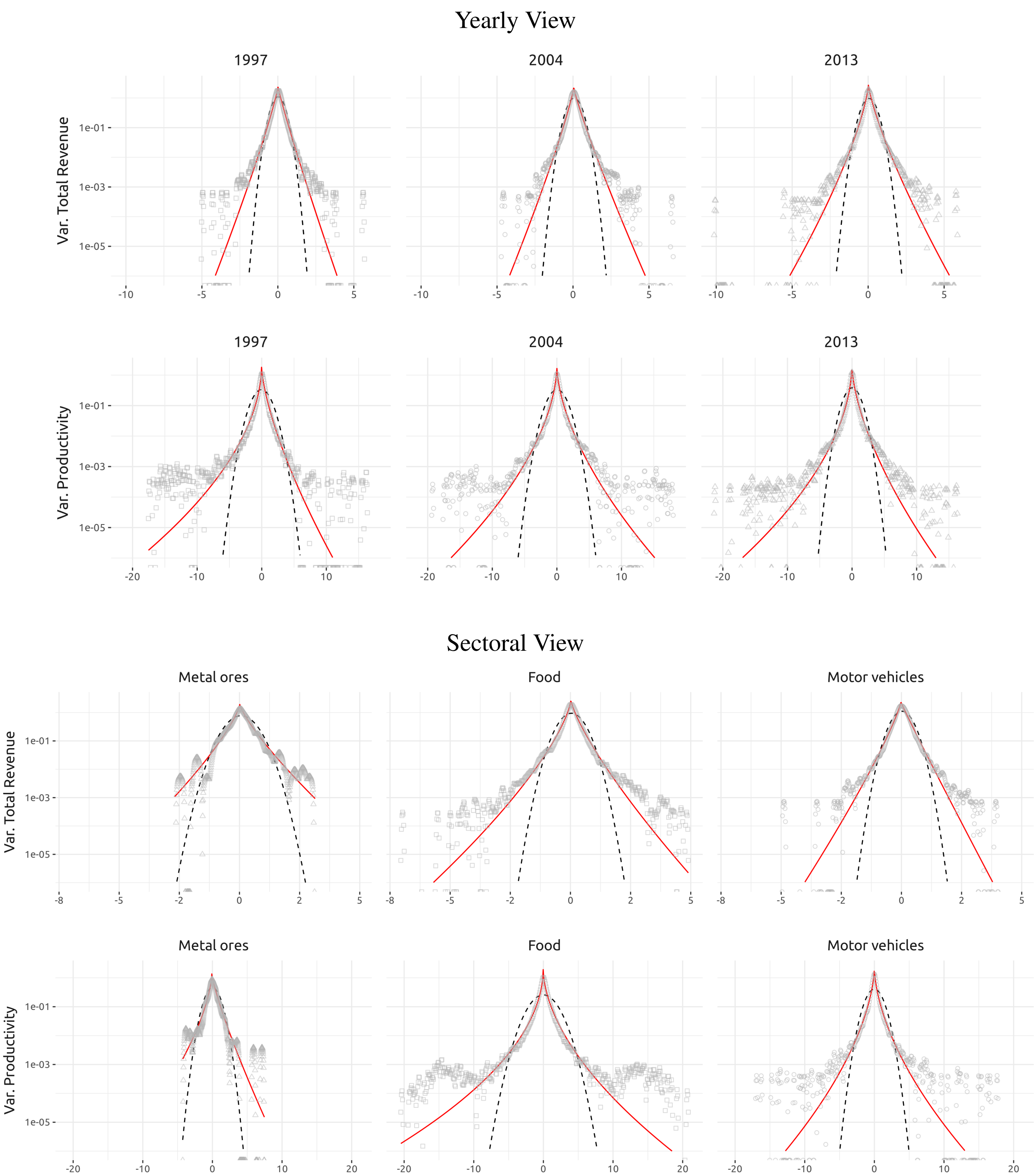


Figure 5: Growth and productivity change - annual and sectoral probability density plots. Note the vertical axis in natural logarithm. Dashed lines represent a normal fit for each distribution, while the red lines represent the AEP fit.

# 5 Conclusion

This article presented a list of statistical measures regarding Brazilian manufacturing. Our results corroborate the international literature and weight favorably to the hypothesis that stylized facts on growth, productivity and size may describe timeless economic phenomena.

Among them, our highlights are 1) the ubiquitous heterogeneity found in the most significant economic proxies for size, performance, and growth; 2) the skewness of firm size distributions, well described by lognormal and Pareto distributions; 3) the efficiency frontier and the roles that hierarchies may play in productivity distributions, and finally 4) the Laplacian shape of firm growth rates and productivity change, implying some type of short-term correlation and competition across business opportunities.

Our interpretation of these results is that they move us to a more complex representation of the markets than what is usually thought. At the same time, the periodicity and robustness of these stylized facts put the theorist in a much firmer ground. We feel that this kind of characterization of empirical results in stylized facts helps more to advance the field than oblivious testing of *a priori* hypotheses. In this sense, we follow the spirit of the words of Gabaix (2009, p. 285), "Estimate, don't test!" and Tukey (1962, p. 13),"it is better to have an approximate answer to the right question, which is often vague, than an exact answer to the wrong question, which can always be made precise". So, what are the consequences of these results for the economics profession?

We think that the current stream of empirical evidence regarding both industrial organization, behavioral economics, labor markets and the most useful tools developed by the great empiricists of the 20th century, such as Wassily Leontief, Colin Clark and Simon Kuznets, known formally as the National Accounts, needs a deep integration with models that can adequately reproduce what is empirically found while having depth in economic thought. The class of models broadly named as "Schumpeter meets Keynes" (Dosi et al., 2010) is a valid effort in this direction, but one that is only in its infancy. The network nature of economics must be recognized, and we need to develop a more realistic representation of the intermediate expenditure, e.g., drawing in the literature of complexity (Hausmann and Hidalgo, 2014; Hartmann et al., 2017), such as input-output tables at the firm level, which will enable to enrich representations as the ones developed by Gereffi et al. (2005).

The overall prognostic is optimistic, and while our knowledge of economics will probably always be only of a statistical nature, the lack of data and computational power that affected the previous generations are no longer a problem, and now we have the opportunity to bring the economic field to more solid grounds.

# References


K. Arrow. The economic implications of learning by doing. *Review of Economic Studies*, 29(3):155–173, 1962. URL https://EconPapers.repec.org/RePEc:oup:restud:v:29:y:1962:i:3:p:155-173.

A. Atkinson and T. Piketty. *Top Incomes over the Twentieth Century: A Contrast between European and English-Speaking Countries*. 2007. ISBN 0199286884,9780199286881,9781429493000.

R. L. Axtell. Zipf distribution of U.S. firm sizes. *Science*, 293(5536):1818–1820, 2001. ISSN 0036-8075. doi: 10.1126/science.1062081. URL http://science.sciencemag.org/content/293/5536/1818.

A.-L. Barabasi. *Network Science*. Cambridge University Press, United Kingdom, 2016. ISBN 978-1107076266.

E. J. Bartelsman and M. Doms. Understanding productivity: lessons from longitudinal microdata. *Journal of Economic Literature*, 38:569–594, 2000.

G. Bottazzi and A. Secchi. Common properties and sectoral specificities in the dynamics of U.S. manufacturing companies. *Review of Industrial Organization*, 23:217–232, 2003.

G. Bottazzi and A. Secchi. Growth and diversification patterns of the worldwide pharmaceutical industry. *Review of Industrial Organization*, 26:195–216, 2005.

G. Bottazzi and A. Secchi. Explaining the distribution of firm growth rates. *The RAND Journal of Economics*, 37(2):235–256, 2006. ISSN 1756-2171. doi: 10.1111/j.1756-2171.2006.tb00014.x. URL http://dx.doi.org/10.1111/j.1756-2171.2006.tb00014.x.

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*, 29:137–159, 2007.

G. Bottazzi, G. Dosi, N. Jacoby, A. Secchiy, and F. Tamagni. Corporate performances and market selection: some comparative evidence. *Industrial and Corporate Change*, 19(6):1953–1996, 2010.

G. Bottazzi, A. Coad, N. Jacoby, and A. Secchi. Corporate growth and industrial dynamics: evidence from french manufacturing. *Applied Economics*, 43(1):103–116, 2011. doi: 10.1080/00036840802400454. URL https://doi.org/10.1080/00036840802400454.

G. Bottazzi, D. Pirino, and F. Tamagni. Zipf law and the firm size distribution: a critical discussion of popular estimators. *Journal of Evolutionary Economics*, 25(3):585–610, Jul 2015. ISSN 1432-1386. doi: 10.1007/s00191-015-0395-7. URL https://doi.org/10.1007/s00191-015-0395-7.

L. M. B. Cabral and J. Mata. On the evolution of the firm size distribution: Facts and theory. *American Economic Review*, 93(4):1075–1090, September 2003. doi: 10.1257/000282803769206205. URL http://www.aeaweb.org/articles?id=10.1257/000282803769206205.

E. Catela, M. Cimoli, and G. Porcile. Productivity and structural heterogeneity in the Brazilian manufacturing sector: Trends and determinants. *Oxford Development Studies*, 43(2):232–252, 2015. doi: 10.1080/13600818.2015.1020939.

M. Cimoli and G. Porcile. Tecnología, heterogeneidad y crescimiento. *ECLAC - Serie Desarrollo Productivo*, 194:1–38, Sep 2013. ISSN 1020-5179. URL http://repositorio.cepal.org/handle/11362/4592.

A. Clauset, C. R. Shalizi, and M. E. J. Newman. Power-law distributions in empirical data. *SIAM Review*, 2009. doi: https://doi.org/10.1137/070710111.

A. Coad, S.-O. Daunfeldt, W. Hölzl, D. Johansson, and P. Nightingale. High-growth firms: introduction to the special section. *Industrial and Corporate Change*, 23(1):91–112, 2014. doi: 10.1093/icc/dtt052. URL http://dx.doi.org/10.1093/icc/dtt052.

G. Dosi. Technological paradigms and technological trajectories: A suggested interpretation of the determinants and directions of technical change. *Research Policy*, 11(3):147–162, 1982. URL https://EconPapers.repec.org/RePEc:eee:respol:v:11:y:1982:i:3:p:147-162.

G. Dosi and R. Nelson. volume 1, chapter Technical Change And Industrial Dynamics As Evolutionary Processes. North Holland., New York, 2010.

G. Dosi, A. Gambardella, M. Grazzi, and L. Orsenigo. Technological revolutions and the evolution of industrial structures: Assessing the impact of new technologies upon the size and boundaries of firms. *Capitalism and Society*, 3(1), 2008.

G. Dosi, G. Fagiolo, and A. Roventini. Schumpeter meeting Keynes: A policy-friendly model of endogenous growth and business cycles. *Journal of Economic Dynamics and Control*, 34:1748–1767, 2010.

G. Dosi, M. Grazzi, C. Tomasi, and A. Zeli. Turbulence underneath the big calm? the micro-evidence behind Italian productivity dynamics. *Small Business Economics*, 39(4):1043–1067, 2012. ISSN 1573-0913. doi: 10.1007/s11187-011-9326-7. URL http://dx.doi.org/10.1007/s11187-011-9326-7.

G. Dosi, D. Moschella, E. Pugliese, and F. Tamagni. Productivity, market selection, and corporate growth:comparative evidence across US and Europe. *Small Business Economics*, 45(3):643–672, 2015.

L. Esteves. chapter Economias de Escala, Economias de Escopo e Eficiência Produtiva na Indústria Brasileira de Transformação. ABD, Brasília, 2015.

J. Felipe and J. McCombie. *The Aggregate Production Function and the Measurement of Technical Change: Not Even Wrong*. Edward Elgar Publishing, 2013. ISBN 9781782549680. URL https://books.google.com.br/books?id=BevDAQAAQBAJ.

F. M. Fisher. Aggregate production functions - a pervasive, but unpersuasive, fairytale. *Eastern Economic Journal*, 31(3): 489–491, 2005. ISSN 00945056, 19394632. URL http://www.jstor.org/stable/40326426.

X. Gabaix. Power laws in economics and finance. *Annual Review of Economics*, 1(1):255–294, 2009. doi: 10.1146/annurev.economics.050708.142940. URL https://doi.org/10.1146/annurev.economics.050708.142940.

P. Gala, J. Camargo, and G. Magacho. The resource curse reloaded: revisiting the Dutch disease with economic complexity analysis. 10 2017. FGV EESP - Textos para Discussão / Working Paper Series - TD 448.

G. Gereffi, J. Humphrey, and T. Sturgeon. The governance of global value chains. *Review of International Political Economy*, 12(1):78–104, 2005. doi: 10.1080/09692290500049805. URL http://dx.doi.org/10.1080/09692290500049805.

R. Gibrat. *Les Inégalités économiques*. Librairie du Recuil Sirey, France, Paris, 1931.

Z. Griliches and H. Regev. Firm productivity in Israeli industry 1979–1988. *Journal of Econometrics*, 65:175–203, 1995.

P. E. Hart and S. J. Prais. The analysis of business concentration: A statistical approach. *Journal of the Royal Statistical Society. Series A (General)*, 119:150–191, 1956.

D. Hartmann, M. R. Guevara, C. Jara-Figueroa, M. Aristarán, and C. A. Hidalgo. Linking economic complexity, institutions, and income inequality. *World Development*, 93:75–93, 2017. ISSN 0305-750X. doi: https://doi.org/10.1016/j.worlddev.2016.12.020. URL http://www.sciencedirect.com/science/article/pii/S0305750X15309876.

R. Hausmann and C. Hidalgo. *The Atlas of Economic Complexity: Mapping Paths to Prosperity*, volume 1. MIT Press Books, The MIT Press, 1 edition, 2014. ISBN 9780262317726.

F. A. Hayek. The use of knowledge in society. *The American Economic Review*, 35(4):519–530, 1945. ISSN 00028282. URL http://www.jstor.org/stable/1809376.

B. Hill. A simple general approach to inference about the tail of a distribution. *The Annals of Statistics*, 3:1163–1174, 1975.

S. Hymer and P. Pashigian. Firm size and rate of growth. *Journal of Political Economy*, 70:556–569, 1962.

IBGE. *Pesquisa Industrial Anual: empresa*, volume 32(1). IBGE, Rio de Janeiro, 2013. Série Relatórios Metodológicos. Coordenação de Indústria.

Y. Ijiri and H. A. Simon. *Skew Distributions and the Sizes of Business Firms*. North-Holland Publishing Company, 1977.

P. R. Krugman. *The Self-Organizing Economy*. Blackwell Publishers., Cambridge, Massachusetts, 1996. ISBN 1557866996.

F. Malerba. Learning by firms and incremental technical change. *The Economic Journal*, 102(413):845–859, 1992. ISSN 00130133, 14680297. URL http://www.jstor.org/stable/2234581.

N. Mathew. Drivers of firm growth: micro-evidence from Indian manufacturing. *Journal of Evolutionary Economics*, pages 1–27, 2017. ISSN 1432-1386. doi: 10.1007/s00191-017-0492-x. URL http://dx.doi.org/10.1007/s00191-017-0492-x.

J. S. Metcalfe. Competition, Fisher's principle and increasing returns in the selection process. *Journal of Evolutionary Economics*, 4(4):327–346, 1994.

F. D. Negri and L. R. Cavalcante, editors. *Produtividade no Brasil: desempenho e determinantes*, volume 1. IPEA, Brasília, 2014. Desempenho.

F. D. Negri and L. R. Cavalcante, editors. *Produtividade no Brasil: desempenho e determinantes*, volume 2. IPEA, Brasília, 2015. Determinantes.

M. Newman. Power laws, Pareto distributions and Zipf's law. *Contemporary Physics*, 46(5):323–351, 2005. doi: 10.1080/00107510500052444. URL https://doi.org/10.1080/00107510500052444.

M. Nogueira, R. Infante, and C. Mussi. chapter Produtividade Do Trabalho E Heterogeneidade Estrutural No Brasil Contemporâneo. ABD, Brasília, 2014.

G. Nurisso and E. Prescott. Origins of too-big-to-fail policy. *Federal Reserve Bank of Cleveland, Working Paper*, (17-10), 2017. doi: https://doi.org/10.26509/frbc-wp-201710.

V. Pareto. *Cours d'Economie Politique*. Droz, Geneva, 1896.

E. Penrose. *The Theory of the Growth of the Firm*. Wiley, New York, 1959.

R. Prebisch. *Capitalismo Periférico: Crisis y Transformación*. Fondo de Cultura Económica., Mexico, 1981.

D. Scott. *Multivariate Density Estimation: Theory, Practice, and Visualization.* John Wiley & Sons, Inc., Hoboken, New Jersey, second edition. edition, 2015.

SEBRAE. *Participação das Micro e Pequenas Empresas na Economia Brasileira.* Serviço Brasileiro de Apoio às Micro e Pequenas Empresas, 2014.

A. Shaikh. Laws of production and laws of algebra. the humbug production function: a comment. *Review of Economics and Statistics*, 56:115–120, 1974.

B. Silverman. *Density Estimation for Statistics and Data Analysis.* Chapman and Hall, London, 1986.

H. Simon. On a class of skew distribution functions. *Biometrika*, 44:425–440, 1955.

E. Soares and M. Castilho. O impacto das importações chinesas na indústria brasileira nos anos 2000. Anais do xlii encontro nacional de economia [proceedings of the 42ndd brazilian economics meeting], ANPEC - Associação Nacional dos Centros de Pósgraduação em Economia [Brazilian Association of Graduate Programs in Economics], 2016. URL https://EconPapers.repec.org/RePEc:anp:en2014:152.

G. Squeff and M. Nogueira. chapter A heterogeneidade estrutural no Brasil de 1950 a 2009. Nacoes Unidas, Santiago, 2015.

M. Stanley, S. Buldyrev, S. Havlin, R. Mantegna, M. Salinger, and H. E. Stanley. Zipf plots and the size distribution of firms. *Economics Letters*, 49(4):453–457, 1995. ISSN 0165-1765. doi: https://doi.org/10.1016/0165-1765(95)00696-D. URL http://www.sciencedirect.com/science/article/pii/016517659500696D.

M. Stanley, L. Amaral, S. Buldyrev, S. Havlin, H. Leschhorn, P. Maass, M. Salinger, and H. Stanley. Scaling behaviour in the growth of companies. *Nature*, 379:804–806, 1996. doi: 10.2307/1883096. URL +http://dx.doi.org/10.1038/379804a0.

T. J. Sturgeon. Modular production networks: a new american model of industrial organization. *Industrial and Corporate Change*, 11(3):451–496, 2002. URL https://EconPapers.repec.org/RePEc:oup:indcch:v:11:y:2002:i:3:p:451-496.

M. Subbotin. On the law of frequency of errors. *Matematicheskii Sbornik*, 31:296–301, 1923.

D. J. Teece, G. Pisano, and A. Shuen. Dynamic capabilities and strategic management. *Strategic Management Journal*, 18(7): 509–533, 1997. ISSN 01432095, 10970266. URL http://www.jstor.org/stable/3088148.

A. Tsybakov. *Introduction to Nonparametric Estimation.* Springer Series in Statistics. Springer-Verlag, New York, 2009. DOI: 10.1007/b13794.

J. W. Tukey. The future of data analysis. *Ann. Math. Statist.*, 33(1):1–67, 03 1962. doi: 10.1214/aoms/1177704711. URL https://doi.org/10.1214/aoms/1177704711.

X. Yu, G. Dosi, M. Grazzi, and J. Lei. Inside the virtuous cycle between productivity, profitability, investment and corporate growth: An anatomy of China industrialization. *Quaderni - Working Paper DSE N. 1006*, 2015a. Available at SSRN: http://ssrn.com/abstract=2605030 or http://dx.doi.org/10.2139/ssrn.2605030.

X. Yu, G. Dosi, J. Lei, and A. Nuvolari. Institutional change and productivity growth in China's manufacturing: the microeconomics of knowledge accumulation and creative restructuring. *Industrial and Corporate Change*, 24(3):565–602, 2015b.