Employment Growth of Establishments in the Brazilian Economy: Results by Age and Size Groups[♦]

Danilo Santa Cruz Coelho (IPEA)
Carlos Henrique Corseuil (IPEA)
Miguel Nathan Foguel (IPEA)

Abstract

Using microdata of Brazilian firms from a longitudinal matched employee-employer data set (RAIS) in the period from 1993 to 2013, this paper confirms that the middle part of the size distribution is "missing" in Brazil and apparently this feature is more intense than in other countries for which there are available results. Our analysis of the dynamics of employment over the life cycle of establishments provides some clues on why there is a missing middle in the size distribution. Considering a representative establishment, the results show that it is born small (perhaps too small) and that the pattern of the growth rate over its life cycle imposes a long time span to surpass the threshold of a mid-sized plant. Our results also indicate that most of the life cycle pattern of establishments can be attributed to age effects. This result is obtained from a application of Deaton's decomposition method that has shown limited scope for the business cycle (year effects) or for the generation to which establishments belong (birth-cohort effects).

Keywords: Firm growth; Size Distribution; Deaton Decomposition.

Resumo

Usando dados longitudinais da RAIS no período de 1993 a 2013, este artigo confirma que a parte do meio da distribuição de tamanho das firmas é "ausente" no Brasil, e aparentemente esta característica é mais intensa do que em outros países para os quais existem resultados disponíveis. Nossa análise da dinâmica do emprego ao longo do ciclo de vida dos estabelecimentos fornece algumas pistas sobre porque há um vazio no meio da distribuição de tamanho. Considerando um estabelecimento representativo, os resultados mostram que ele nasce pequeno (talvez demasiado pequeno) e que o padrão da taxa de crescimento ao longo de seu ciclo de vida impõe um intervalo de tempo considerável para ultrapassar o limiar de uma planta de médio porte. A aplicação do método de decomposição de Deaton aponta que o efeito da idade é de fato responsável por este padrão de ciclo de vida. Nossos resultados apontam um escopo limitado para o ciclo de negócios (efeitos de ano) ou para a geração a que pertencem os estabelecimentos (efeitos de coorte).

Palavras-chave: Crescimento das firmas, Distribuição de Tamanho; Decomposição de Deaton.

Área 6 - Crescimento, Desenvolvimento Econômico e Instituições.

JEL Classification: O14; O12; L25; L26

[♦] We are indebted with the OECD members of Dynemp team for their useful comments in a preliminary version of this paper. Luciana Costa deserves our gratitude due to her efficient research assistance.

1. Introduction

Keeping high employment growth rates over long time periods is considered a desirable characteristic of the development process not only for the direct effect on aggregate employment growth but also because of the connections with other performance indicators such as wage and productivity growth. The pattern of employment growth is thus a key process to be monitored in any economy, in particular those of developing countries like Brazil.

Typically, the monitoring of employment growth in a country is implemented using household surveys. Though rich in information on the workers' traits, this type of survey rarely contains information on the characteristics of the establishments such as their size and age. The process of employment growth is however highly connected with the characteristics and performance of establishments over their lifecycle. For instance, the entry and exit processes of establishments as well as their capacity to grow are important components behind the employment dynamic in an economy. Using establishment-level data is thus a highly valuable source of information if one wants to have a richer understanding of aggregate employment growth.

One of the main theoretical arguments for why aggregate employment and other performance indicators are linked to establishments connects their lifecycle to a learning process through which the plant (decision maker) gradually adjusts to the (new) environment since the beginning of its operations (Nelson and Winter, 1982). This may be motivated by a learning process not only about the evolving environment but also about its own capabilities (Jovanovich, 1982). According to this view, an important indicator to be monitored is the incidence of closing plants by establishment age, which represents an interruption of the learning process.

Given this background, the general objective of this paper is to analyse the statistical patterns of employment dynamics of establishments in the Brazilian economy. In particular, we study the lifecycle evolution of establishments so that we can assess how the age as well as the entry and exit components are related to the employment growth process in the country. To accomplish that we benefit from a large scale longitudinal plant level data to which we have access.

Our focus will be on the performance of small establishments at birth. Studying the pattern of growth of this segment is important for at least two reasons. First, because any small improvement in the rate of employment growth of small establishments tends to have a considerable impact on job creation due to their large share in the establishment and employment size distributions of countries, in particular the developing ones. Second, it has potential effects on the degree of competition in various industries, which in turn affects price adjustments and the innovation impetus in the economy.

The present paper bridges two branches of the empirical literature on either firm or plant size dynamics. The first one encompasses papers exploring large scale, longitudinal firm- or plant-level datasets to unveil basic facts on employment dynamics along the life cycle of the relevant unit of analysis. This literature has two waves of studies, one in the 90s, which includes the important volume edited by Audrescht and Mata (1995) and the survey by Caves (1998) and is centered in analyzing data from European countries, while the more recent one concentrates on data for the US as done by Haltiwanger et al. (2013) and Decker et al. (2014). One of main the stylized facts unveiled by these studies is that the younger plants or firms exhibit higher employment growth rates but also higher death rates. As age and size are strongly related, these findings also hold true for small production units.

A first contribution of this study is that we confirm these stylized facts for Brazil. We offer an additional contribution by innovating on the tools used to obtain a comprehensive picture of establishment employment growth. There are at least two methodological challenges for identifying how employment evolves as plants age. One is a composition effect due to higher probability for small plants to shut down. This shifts the (conditional on age) distribution of plants (across sizes) towards bigger establishments at any comparison of average employment level from one age to the next one. The second challenge is to disentangle a pure age effect from confounding effects of time that are related to the occurrence of economic shocks that hit the establishments as they age. For instance plants tend to experience higher growth rates if their existence coincides with an expansionary phase of the economy. Also, a plant's life cycle pattern may be affected by the prevailing conditions at the time it started operating (for instance, the availability of credit, the costs for registration, the incumbents' market power). Hence one should try to isolate pure age effects from period specific shocks and birth cohort idiosyncratic characteristics. We deal with both issues using a decomposition method put forward by Deaton and Paxson (1994) that separates age, year, and cohort effects.

We show that raw growth rates across ages are indeed influenced by a composition effect that is related to the already mentioned high mortality rate of small and young plants. We also unveil that cohort and macro effects have limited effect on employment dynamics, which is essentially driven by pure age effects.

The second and most recent branch of the literature argues whether small establishments underperform in developing countries creating a "missing middle" in the plant size distribution. This distorted pattern can be the result of a large set of factors such as entry costs, the tax system, the level of development of financial markets, the regulatory environment, and the scale and composition of market demand. This issue is still in debate following important contributions by Tybout (2000) and by Hsieh and Olken (2014). We will follow Tybout (2014) who proposes a method to compare the observable plant size distribution with a Pareto one with estimated parameter.

Our results show that performance of small establishments in Brazil is relatively poor. In particular, we find that though this segment is able to exhibit elevated growth rates in the beginning of their lives, they tend to die early and do not grow enough to increase their scale to that of mid-sized establishments. Connected to these findings, our results indicate that the middle part of the size distribution is "missing" in Brazil. This is robust to different partitions of the size distribution and is valid for the whole (formal) economy as well as for the manufacturing sector alone. Comparing the results for the manufacturing sector to those of other developing countries for which there is available evidence, apparently the missing middle problem is more intense in Brazil than in those countries.

The paper follows with a description of our data in the second section. The following section contains the results on the overall pattern of employment growth, the decomposition results for the age, cohort, and year effects, as well as evidence on the composition effects stemming from the process of death of plants. The fourth section describes the method to identify a missing middle as well as the results from its application to the Brazilian plant size distribution. The last section presents the main conclusions.

2. **Data**

In this paper we use a very large restricted-access administrative file collected by the Brazilian Ministry of Employment and Labor (*Ministério do Trabalho e Emprego*), the *Relação Anual de Informações Sociais* (RAIS). RAIS is a longitudinal matched employee-employer dataset covering by law the universe of formally employed workers in Brazil. Each observation in the dataset consists of a contract-worker-establishment triplet in a given year.

All tax-registered establishments have to report the basic characteristics of the labour contract for every worker formally employed at some point during the previous calendar year.[1] Apart from tax/social security compliance the data has no coverage limitation, as opposed to other similar databases that are limited by geographical region, plant/firm size, or industry.

Apart from information on industry classification, legal form, and location at the municipality level RAIS provides a unique identification number (*CNPJ*) given by the federal tax authority for each establishment. This is a key variable for this study since we use it to: i) aggregate the number of workers within plants at a particular time period, ii) follow this quantity over time, and iii) define establishment age in a particular year.

All the analyses in this study are based upon information on private non-farm establishments. These filters require harmonized information throughout time the on legal form and the industry classification, which are available since 1995. Hence we restrict our sample to establishments born from 1995 to 2013. This would enable us to show the plants' life cycle pattern up to their nineteenth year of existence in the formal sector. We will restrict our results to the twelve first years of plants life in the Brazilian formal sector as some of the results are based on a methodology for which such restriction is desired, as will be properly pointed later on.

The main variables for our analysis are establishment size and age per year. Attrition is a potential source of measurement error for both variables. Some odd pattern follows, possibly due to occasional non-reporting by complier establishments, as some establishments "disappear" from RAIS in a particular year and eventually return in subsequent years.

Our age variable is not affected as it is based on the year of the first appearance since 1992. As so, if occasional non-reporting occurs in the interior of the 1995-2013 interval, we keep the age variable rising until the establishment is back to the sample. We also won't have problem when 1995 is a non-reporting year as long the establishment had reported information in any of the years from 1992 to 1995.

As for establishment size, we measure it as the average number of workers employed by the establishments across the months of the respective year. As in many plant/firm data there is available information on the number of employees at a particular point in time. In RAIS this is readily available for the last day of each year. However there is a significant number of establishments that employ workers along the year but reach the last day without any employee, even when this is not their last year in RAIS. Hence we build the average size across the monthly stock of employees, which is based on information on dates of hiring and separation for each worker. For episodes of non-reporting as mentioned before, the establishment size is not computed in the non-reported year(s)

Over the period from 1995 to 2013 RAIS contains an average of 2.1 million establishment records per year. This number has increased over this period, and can be related to the process of increasing formalization that took place in Brazil. Such trend in formalization encompasses two margins: i) the extensive one, with increasing number of new formal plants; and ii) the intensive margin, with increasing number of formal jobs within a set of formal plants. The results to be presented in the remaining of this paper should be interpreted taking in consideration that the first margin may be driven by informal plants switching status to formal plants. Under this scenario our age variable does not coincide with years of plant existence, but indicates the life cycle under the formal sector environment. Further considerations on this issue will be addressed when discussing specific results in what follows.

---

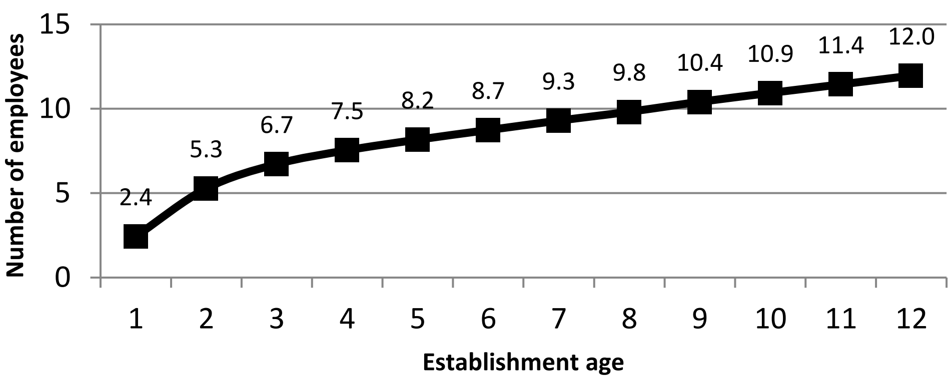
[1] There are incentives for truthful reporting since the main purpose of RAIS is to administer a federal wage supplement (*Abono Salarial*) to formal workers.

3. The plant employment dynamics over their life cycle

3.1. Aggregate life cycle and decomposition results

The aim of this subsection is to illustrate basic facts on employment dynamics over the lifecycle of establishments. We start plotting data on the average number of employees per establishment as shown in Figure 1. The first thing to notice is that the average number of employees of plants in their first year is 2.4, which implies that Brazilian formal establishments are typically born small. One can also see that the average number of employees grows almost fivefold in the first twelve years of life (from 2.4 to 12), which corresponds to an average annual growth rate of 15.5%.

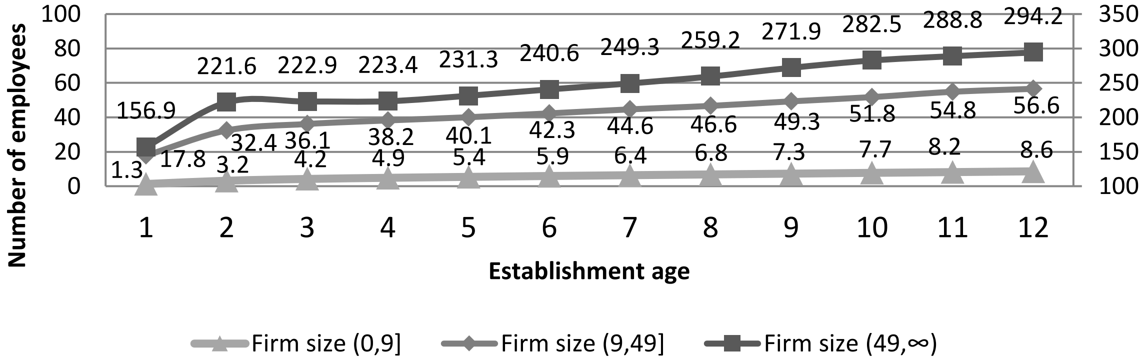
Figure 1: Average employment level by establishment age



Source: Authors´ estimations based on microdata from RAIS.

Figure 1 also shows a great deal of heterogeneity across ages: in the second year the growth rate is very high (116%), then it decreases gradually reaching 4.5% in the twelfth year of age. For future reference, it is worth pointing out that it takes about seven years for the typical establishment born in the Brazilian formal sector (i.e., a plant that starts off in business with 2.4 employees) to reach the lower bound of the range in size associated to middle sized plants (usually taken to be from 9 to 49 employees in the literature as will be detailed later in this paper). Such reference of the "typical" life cycle pattern of an establishment deserves two cautionary notes. First, as attested by Figure 1.A in the appendix, a great deal of heterogeneity can also be found across plants within the same age. Second, if the sample is partitioned in groups of plants according to their size at birth, as done in Figure 2, one can see that, on average, the group of smaller plants at birth does not reach the size of 9 employees even at the twelfth year in the formal sector.

Figure 2: Average employment level by establishment age and establishment size at birth

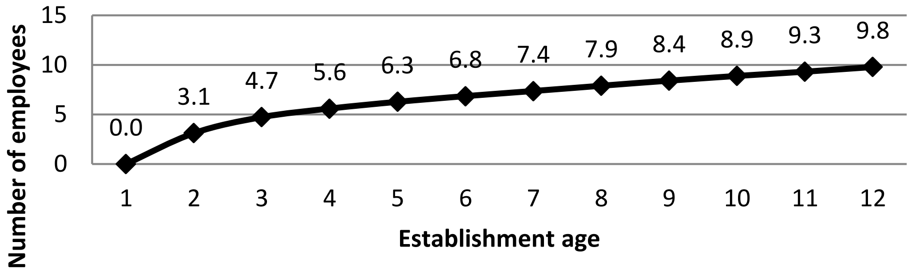


Source: Authors´ estimations based on microdata from RAIS.

As an attempt to isolate the effect of age from other determinants of plant growth such as the macro environment or cohort specific conditions, we perform Deaton and Paxson (1994)'s decomposition by age, cohort, and year (or macro) effects. The implementation is based on a regression model that uses dummies for ages, birth cohorts, and year of the observation to explain the evolution of the establishments' employment levels.[2] We leave the details of the method for the appendix. In principle, we could estimate cohort effects for every entry year of the establishments in our period of analysis. However, for the decomposition exercise of this section, we opted for restricting the sample only to establishments that entered until 2002 (inclusive). The advantage of doing so is that we guarantee that all plants in the sample can potentially reach at least 12 years of age between 1995 and 2013.

Results from the decomposition are reported in Figures 3 and 4. Figure 3 shows that the age effect is remarkably similar to what we have shown from raw data (Figure 1). After getting rid of macro shocks and cohort specific components, establishment size is an increasing function of age displaying high growth rates in the initial years of life and a lower rate as establishments get older.

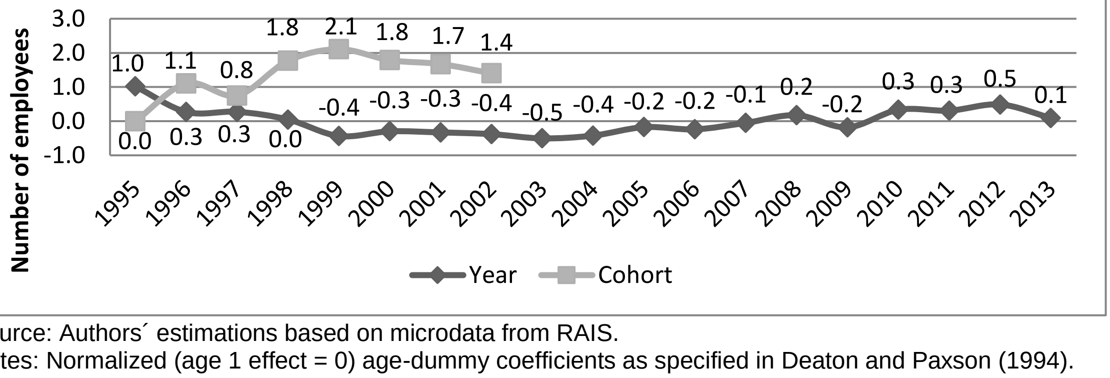
Figure 3: Age effect on plant size from decomposition method



Source: Authors´ estimations based on microdata from RAIS.
Notes: Normalized (age 1 effect = 0) age-dummy coefficients as specified in Deaton and Paxson (1994).

As for the other two components, Figure 4 shows that their magnitudes are much smaller than for the age dimension. The year effects are similar to the pattern of economic growth in the period, while cohort effects depict an inverted U-shape peaking for the cohort born in 1999.[3]

Figure 4: Year and cohort effects on plant size from decomposition method



Source: Authors´ estimations based on microdata from RAIS.
Notes: Normalized (age 1 effect = 0) age-dummy coefficients as specified in Deaton and Paxson (1994).

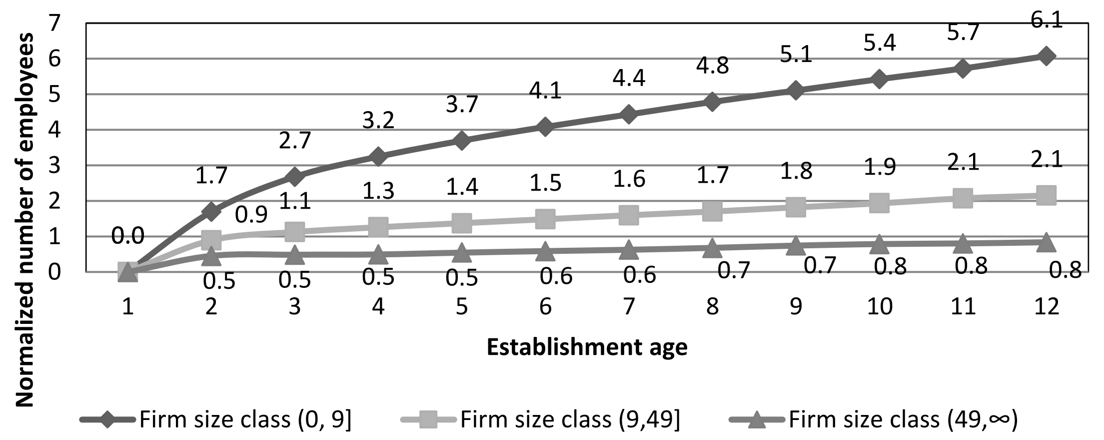
We also estimated the decomposition model for three different groups of the establishments' size at birth. The first group consists of establishments with less than 9 employees (inclusive), the

[2] This was originally proposed for the analysis of wages or consumption but can be applied to any other variable affected by these three dimensions.

[3] As previously mentioned, we excluded all the establishments of cohorts after 2002.

second by establishments with between 9 and 49 employees (inclusive) and, the third by establishments with more than 49 employees. In order to facilitate the comparison across groups, the regression coefficients are divided by the average number of employees in each group when the establishments were born (1.3, 17.8 and 156.9, respectively). Our estimates, reported in Figure 5, reveal that the age effects are higher for the smallest group size, despite an increasing trend for all three groups. For instance, at age 12 the first group's age effect alone would make the establishment grows 607%, while for the second and third groups, this number would be 215% and 84%, respectively. Despite the much higher age effect for the first group, the average plant born in this group (i.e., a plant born with 1.3 employees) does not surpass 8 employees at the twelfth year of existence. If we predict the average size at that point of their life cycle; as a product of its initial size and the predicted growth rate for the first twelve years shown in figure 4; we get to the predicted size of 7.89 (1.3*607%=7.89). In other words, the pure age effect is not strong enough to transform a typical small plant into a middle sized establishment.

Figure 5: Age effect on plant size according to decomposition method - results by establishment birth size



Source: Authors´ estimations based on microdata from RAIS.
Notes: Normalized (age 1 effect = 0) age-dummy coefficients as specified in Deaton and Paxson (1994).

3.2 The composition effect due to establishment deaths

Apart from macro and cohort effects, the pattern of employment growth reported in Figure 1 may be affected by the process of plants' death. The observed pattern is reproduced by the line with triangular markers in Figure 6 and the other two lines represent the average number of employees by age for two partitions of the sample. For each age we split the sample in establishments that survive at least one more year (upper line) and establishments that appear for the last time in our data at that age (lower line). These last two lines clearly show that the overall pattern is influenced by the death of plants. Indeed, there is a striking contrast between the average number of employees in the two partitions of the sample, and the difference increases with establishment age. In the first year the survival plants are three times bigger than non-survival counterparts (closing plants), while in the twelfth year the average size of the two groups differs by a factor of nine.

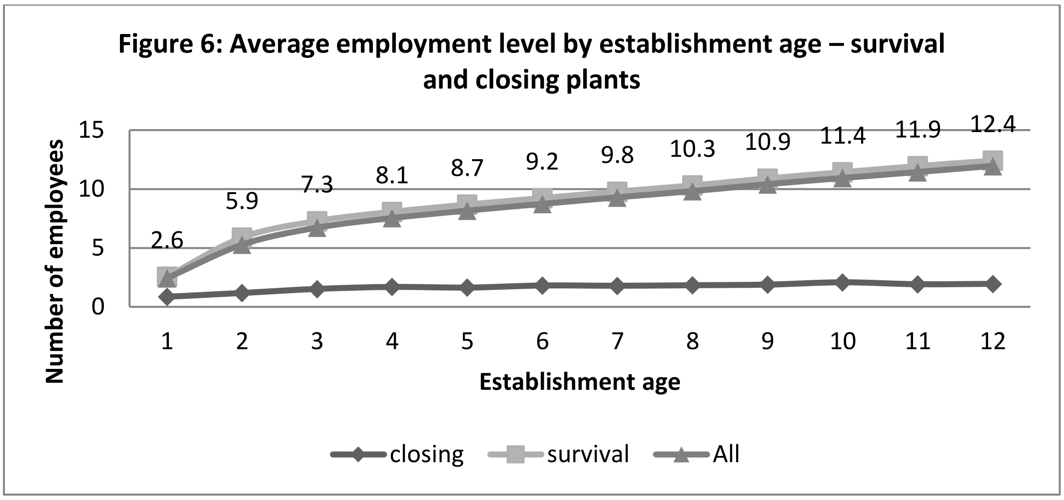


Figure 6: Average employment level by establishment age – survival and closing plants

Source: Authors´ estimations based on microdata from RAIS.

It is worth noticing that the average size of the closing plants stays around one employee for all considered ages, indicating that a typical closing plant is very small at the moment of its death. Figure 7 reinforce this result by comparing the average size of closing plants at the moment of their death with the size of the same group of plants at birth. The results point to a lower average size at death than at birth for the same plants. This may partially explain why closing plants tend to be smaller than surviving ones, as shown in figure 6.

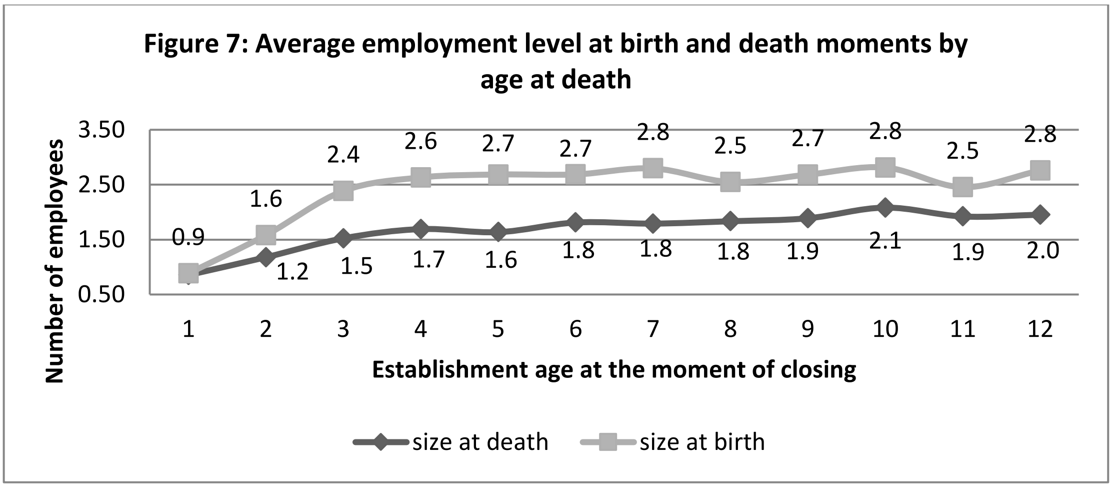


Figure 7: Average employment level at birth and death moments by age at death

Source: Authors´ estimations based on microdata from RAIS.

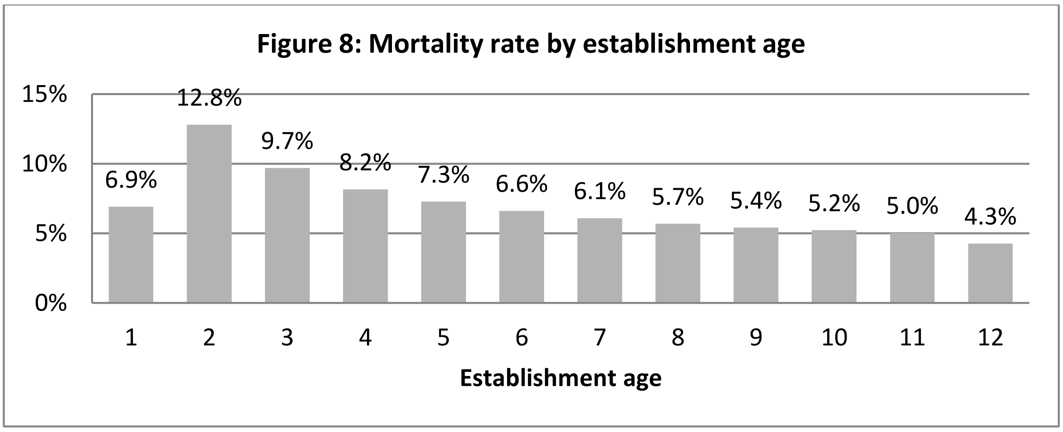


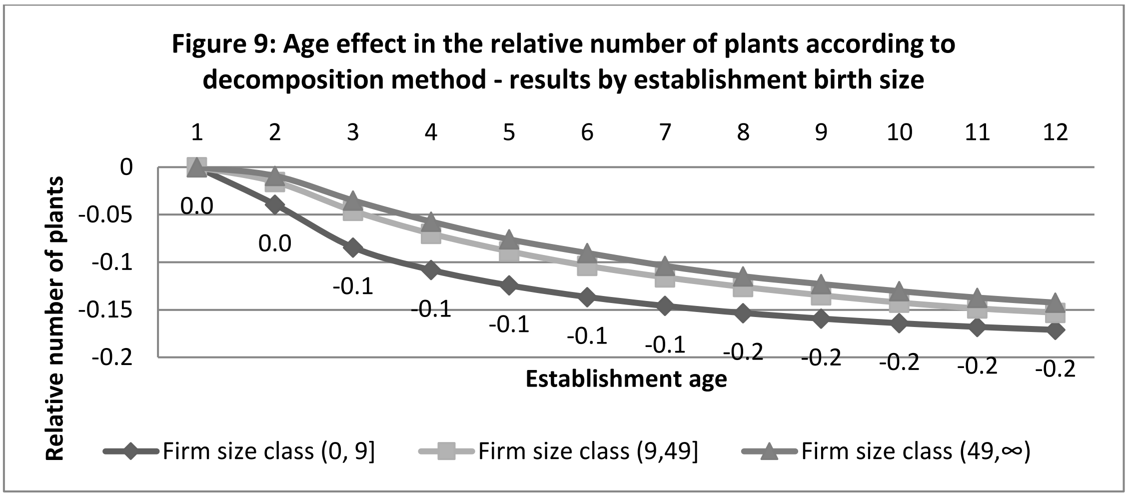
Figure 8: Mortality rate by establishment age

Source: Authors´ estimations based on microdata from RAIS.

The fact that closing plants tend to be small at birth and even smaller when they die generates a composition effect on the evolution of the overall average size. The (conditional on age) distribution of plants across sizes shifts towards bigger establishments when smaller plants close and leave the sample. This composition effect will be higher the larger the share of closing plants in the sample. Figure 8 shows the plant mortality rate by age (the proportion of establishments that die at a given age in relation to the total number of establishments with that age). One can see that this is a critical issue especially in the second (13%) and third year (10%) of plant life, a finding that is in line with evidence for other countries (see Calvino et alli 2015). After the second year mortality rate trend becomes negative, reaching 4% at age 12.

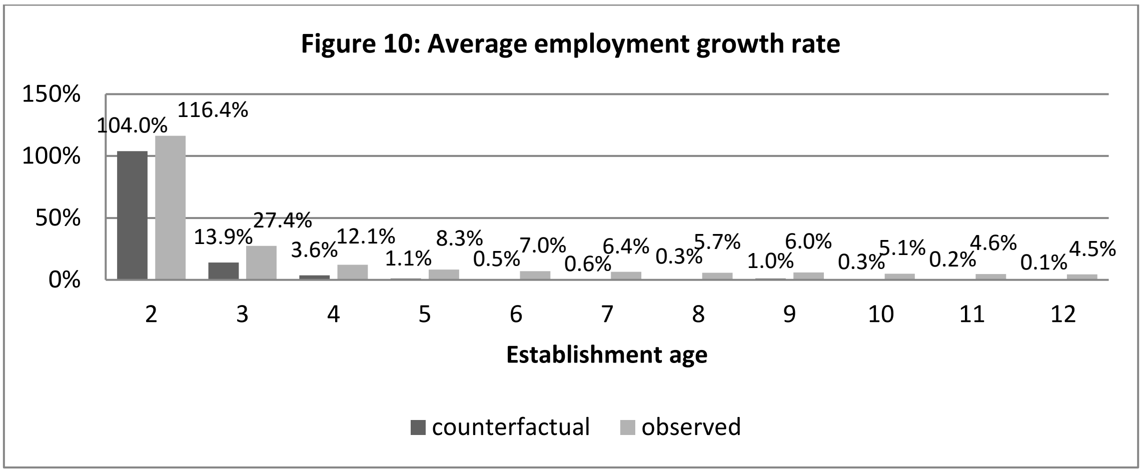
The pattern of mortality rate above can also be influenced by macro or cohort specific shocks. Therefore we also decompose the evolution of the number of establishments in cohort, time, and age effects using the previous decomposition method. This exercise is done for the same three size groups we previously used (i.e., establishments born with up to 9 employees, above 9 but less than 49 employees, and over 49 employees). Figure 9 shows the age effects, which are now normalized with respect to the number of establishments in each size class at age one.

Our results point to negative (normalized) age effects, meaning that the number of establishments diminishes with age, and that their absolute values are higher for the smallest group size. For instance, at age 3 around 8% of plants that were born small (i.e., below 9 employees) would be death from a pure age effect, whereas this figure for the middle and upper groups is around 4%. Though the age effect of the smallest class size gets closer to those of the others at later ages, it is still higher (in module) at age 12 (17% against 15%).



Source: Authors´ estimations based on microdata from RAIS. Notes: Normalized (age 1 effect = 0) age-dummy coefficients as specified in Deaton and Paxson (1994).

We made an attempt to measure the importance of the composition effect. For this purpose we computed two average employment growth rates for each two consecutive ages. One is the observed growth rate from Figure 1 and the other is a counterfactual growth rate that maintains in the sample also those plants that died in the age interval under consideration. Clearly, it is impossible to actually know the employment level of the death plants had they survived one more year. It is likely that the number of employees would diminish for a significant fraction of them but it could also be that many would recover and even increase the sizes of their labor force. We opted to construct the counterfactual imposing zero growth to death plants, that is, for those plants we imputed the same employment level of the previous age. Figure 10 reports the two employment growth rates by plant age.



Source: Authors´ estimations based on microdata from RAIS.

Figure 10 evinces that the composition effect is quite important, especially after the second year of life. For instance, in the second year, the counterfactual growth rate corresponds to a third of the observed one and from the third year on the counterfactual rate is virtually zero. This comparison suggests that the observed growth rates for these ages were substantially influenced by the "cleansing" effect of plants' death. One implication of this counterfactual exercise is that, without the death process, employment growth would be much weaker and the typical establishment born in the Brazilian formal sector would not reach the threshold of middle sized establishments.

In sum, these results allow us to draw a picture of the employment dynamics in the Brazilian formal sector with the following characteristics. A typical establishment is born small, grows relatively fast in the first years but experiences at lower growth rates thereafter. Pure age effects have a much higher impact on the dynamics of employment growth than the macro and cohort effects and they display the same pattern across ages as the one observed for a typical establishment. The results also show that the closing plants tend to be small and that a large fraction of small-born establishments are death before reaching the age of three. We also learnt that growth pattern of employment is affected by the process of death of plants, producing a cleansing effect that arguably inflates the magnitudes of the growth rates across ages.

Our results clearly evince that the employment dynamics of small establishments in Brazil is quite different from that of larger plants. One may then speculate that if a large fraction of small establishments die early and the surviving ones do not grow very much, the establishment size distribution in Brazil is likely to display a lower concentration of establishments of middle size. In the next section we turn into this question.

4. "Missing Middle" and the Establishment Size Distribution in Brazil

Tybout (2000) shows evidence of a much higher concentration of employment in small establishments in low income countries than in industrialized nations. The evidence collected by Tybout (2000) also shows that the employment share of the middle part of the distribution was considerably lower in the former than in the latter group of countries.[4] These results have been interpreted as evidence that small establishments have more difficulties to reach the scale of mid-sized establishments in developing countries. The thinner middle part of the size distribution for this group of countries has become a broadly accepted result in the field of development economics and was baptised the "missing middle" phenomenon.

---

[4] The evidence presented in Tybout (2000, Table 1) was gathered from different studies, in particular from Liedholm and Mead (1987). Most of the figures were only available for three bins of the employment share distribution, namely: 1-9, 10-49, and 50+ employees.

Many explanations have been proposed for the higher concentration of small establishments and the missing middle phenomenon in the developing world. A first group of explanations is based on institutional factors such as the regulatory and tax systems of developing countries. Since larger establishments have to cope with more intricate regulations, face higher labor costs (including the minimum wage and payroll taxes), and become more exposed to the enforcement of the law, many entrepreneurs prefer to be informal and remain small. As a response to some of these constraints, tax subsidy programs have been introduced in many developing countries to stimulate small establishments to formalize their operation and grow in size. But even these initiatives have been criticized because they establish employment or revenue thresholds for tax exemptions that may end up desincentivizing small establishments to grow. The high licensing costs to open a establishment due to public sector inefficiencies (including corruption) can also hinder some talented but credit constrained entrepreneurs to start up their businesses.

A second group of explanations is related to the insufficient development of financial markets and the high supply of unskilled labour that characterise low- and mid-income countries. Since poorly developed financial markets lack instruments to provide long-term finance, potential entrepreneurs and established small establishments find it difficult to obtain credit to invest in fixed capital or even to cope with cash flow problems. As a result, many small establishments are not born, do not increase their scale or even die due to the low development of the financial sector in developing countries. The relative trade closeness of those countries further hinders the access of small establishments to modern machinery, equipment, and technologies. Compounded with these obstacles to invest in modern fixed capital, the abundant supply of low skilled labour pushes small establishments to start up and keep operating with labour intensive, low productive technologies. Lack of supply of training for the basic skills to manage small business is another factor that diminishes the chances of survival and growth of small enterprises in the underdeveloped world.

Another line of arguments has to do with the composition of demand. Most developing countries have a large fraction of low-income families whose consumption expenditures are concentrated in food items and basic goods that can be efficiently produced with small-scale, low-tech plants. This also tends to create a production structure with low diversification, so many product and service markets are lowly developed or missing. The low availability of wide, good-quality transportation networks is another factor that hampers the growth of small establishments. Lack of this type of public good increases the costs and even deters investments to increase the scale of plants.

Despite the arguments for the existence of a missing middle, there are some concerns in the literature on the whether it is in fact a characteristic of the establishment size distribution of developing countries. These concerns have been recently expressed by Hsieh and Olken (2014), who present a set of results using data from India, Mexico, and Indonesia that are not entirely compatible with the missing middle phenomenon. First, the authors show that the (average) productivity of establishments is positively related to their size, a finding that calls into question the common view that small establishments are the ones with high potential returns but do not to grow because they are somehow constrained (say, because of credit constraints). Second, they do not find evidence of kinks in the establishment size or establishment revenue distributions around the thresholds established in the tax policy in those three countries to benefit small establishments. This implies that this type of initiative does not seem to refrain small establishments to grow in size. Finally, and more important for this study, they show that the histograms of establishment size distribution display a monotonic decay, a result that discards a bimodal pattern of high concentration of small and large establishments that would attest the presence of a missing middle.

To explain why the previous empirical literature had improperly established the existence of the missing middle (in terms of a presence of bimodality in the size distribution) in developing countries, Hsieh and Olken (2014) point to the combination of two issues associated with the use of the available data. The first is the incorrect use of the distribution of *employment shares* across plant sizes instead of the distribution of *plant sizes*. They argue that the theory is focused on the establishments' decisions on whether or not to grow, so the relevant distribution for testing the existence of a missing middle is the latter, not the former. The second issue is that the results obtained in the literature are based on an arbitrary number of size bins (as well as the widths of these bins). Hsieh and Olken (2014) show that when the size of bins used in the literature (1-9, 10-49, and 50+ employees) is imposed to the establishment size distribution of India, Indonesia, and Mexico, the bimodal pattern that appears for the employment share distribution vanishes. Hsieh and Olken (2014) conclude that "(…) the existing facts about the missing middle seem to come the combination of these two transformation to the data: the transformation from the distribution of firms to the aggregate employment share, and the arbitrary binning of the employment share distribution" (p. 106).

To circumvent these criticisms, our results are computed for establishment size distribution and, for comparability with the results in Tybout (2014), also for the employment share distribution. Since the international evidence is only available for the manufacturing sector, the analysis is conducted separately for the entire formal sector and for the manufacturing sector alone. To deal with the problem of choosing an arbitrary number of size bins and their widths, we also vary both dimensions in the analysis.

We do not think that bimodality is the only criterion to evaluate the presence of a missing middle in the (establishment) size distribution. Indeed, the empirical distribution of a developing country can be unimodal and yet its middle part could be thinner than the corresponding middle part of the distribution of a developed country (e.g., the US distribution). The hypothetical distributions displayed in Figure 11 – which is based on Tybout (2014, p.2) – shows this case, where the dashed line represents the density of establishments in the developing country and the solid line the density of establishments in the developed country. As the comparison of the two lines shows, though both distributions are unimodal, the share of mid-sized establishments in the developing country is smaller than the corresponding share for the developed one.

The main objective of this section is to analyse whether the establishment size distribution of the Brazilian formal sector displays a missing middle. To conduct this analysis we follow Tybout (2014) who proposes a method that contrasts the observed shares of the establishment size distribution with the corresponding predicted shares of what is considered the best description of establishment size distributions in the literature: the Pareto distribution.[5] The idea is that substantial differences between the theoretical and observed shares constitute evidence of under or over representation of the size groups. If the middle part (however defined) of the observed distribution is relatively underrepresented, we take this as evidence that there is a missing middle.

As in the previous section, our data only covers the formal sector in Brazil. Though the share of informal sector employment decreased markedly in the last decade, informal employment still represents around 20% of the labor force in the country. As informal establishments tend be small, our results will likely understate their weight in the left tail of the establishment size distribution. As before, our universe is the total set of establishments that belong to the private, non-agricultural sector in the Brazilian formal sector. To be compatible with the analysis of other parts of this study, our sample is formed by all establishments that were born from 1995 on. If we computed our results only for establishments in their first years of existence, the size distribution would be too much influenced by the size profile of young establishments. Hence, the results are

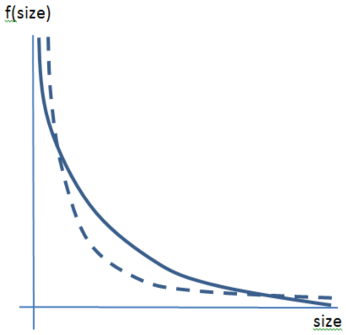
[5] See Axtell (2001) and Luttmer (2007).

obtained using the last year available in our data, 2013, for which the size distribution is more stable, as it becomes also influenced by the presence of older establishments. The main conclusions do not change when the exercise is implemented using the whole sample of establishments, not only those observed in 2013.[6]

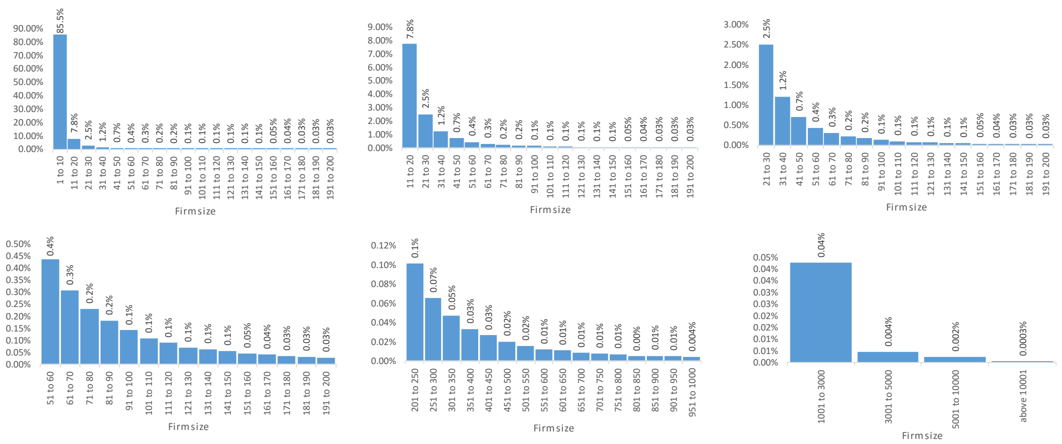
We begin by showing histograms of the size distributions. Panel 1 presents the distribution of plant size while Panel 2 the distribution of employment shares by establishment size. Following Hsieh and Olken (2014), we use bins of ten workers up to size 200 and vary the lower cutoff of the range to make the shares more easily visible. We also show a graph for sizes between 201 and 1000 using bins of 50 workers and a graph for sizes over 1001 employees with bins of different widths.

Panel 1 reveals that the plant size distribution is highly right skewed displaying a very high concentration of small plants and a monotonic decline in the shares of larger establishments. This shape is similar to that presented in Hsieh and Olken (2014) for India, Indonesia, and Mexico. Though we do not think that bimodality is the only criterion to check whether a distribution displays a missing middle, we also do not find any evidence of bimodality in the establishment size distribution for Brazil. Panel 2 shows that the employment share distribution is also right skewed, though the decline in shares is naturally smoother than for the plant size distribution. No bimodality is evinced in Panel 2 either. Though not shown, these findings are also valid for the manufacturing sector alone.

Figura 11: Unimodality and missing middle



Panel 1: Distribution of Establishment Size by Number of Employees – Year: 2013

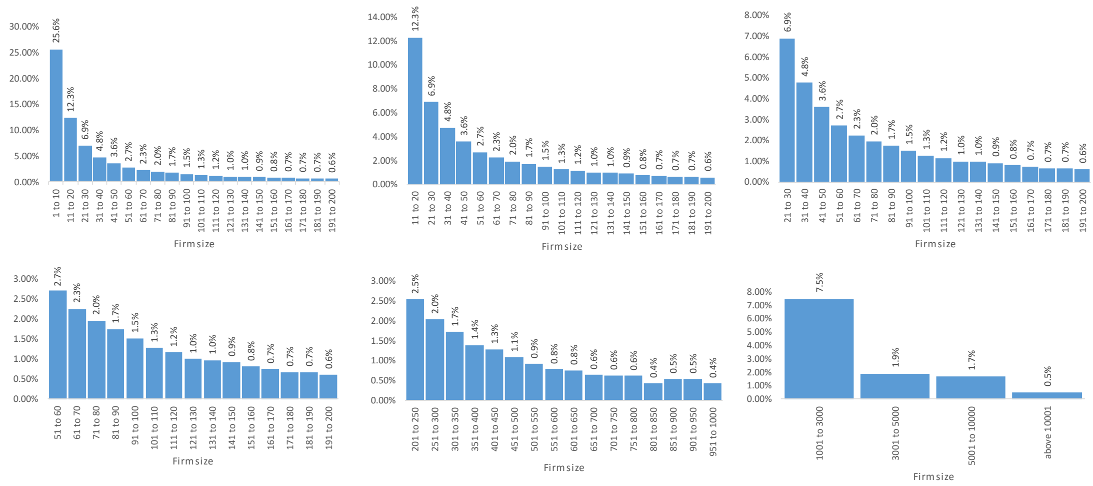


Source: Authors´ estimations based on microdata from RAIS. Notes: The sample is formed by all private, non-agricultural establishments that were born since 1995. The figures refer to the year 2013.

---

[6] The results are available from the authors upon request.

Panel 2: Distribution of Employment Share by Establishment Size – Year: 2013



Source: Authors´ estimations based on microdata from RAIS. Notes: The sample is formed by all private, non-agricultural establishments that were born since 1995. The figures refer to the year 2013.

The Pareto distribution has been established as the best characterization of establishment size distributions for developed countries.[7] Following the method proposed by Tybout (2014), we contrast the empirical size distribution for Brazil with its closest theoretical Pareto distribution. The idea behind this procedure is that observed deviations between the shares of the predicted and the actual Pareto distributions indicate which parts of the empirical distribution are under or over represented.

The upper tail of the Pareto establishment size distribution can be written as:

$$\bar{F}(S) = Pr(S \geq s) = \left(\frac{s_0}{s}\right)^{\alpha}, s \geq s_0, \alpha > 0, \qquad (1)$$

where S denotes establishment size, α is the shape parameter, and s_0 is the scale parameter that represents the minimum possible value assumed by S. We assume that the smallest plant employs one single worker, so $s_0 = 1$. Larger values of α imply higher concentration of smaller establishments in the size distribution. From (1), the share of establishments (h^f) in the size range $s_i \leq s \leq s_j$ is given by:

$$h^f\left(s_i, s_j|\alpha\right) = F\left(s_i \leq s \leq s_j|\alpha\right) = s_i^{-\alpha} - s_j^{-\alpha}. \qquad (2)$$

To calibrate the value of α we use the main approach suggested by Tybout (2014). The method searches for the value of α that minimizes the Euclidian distance between the log of vector $\boldsymbol{s}$ of actual shares given by the Pareto distribution and the log of vector $\hat{\boldsymbol{s}}$ of predicted shares given by the empirical distribution.[8]

Since the bulk of the literature is based on the employment share distribution across plant sizes, we also implement the method for this distribution. Given that the employment share in each bin can be obtained by multiplying the number of establishments in each bin with the average

---

[7] Specifically, the Zipf distribution, a special case of the Pareto, is considered the best description of the firm size distribution. See Axtell (2001).

[8] An alternative method suggested by Tybout (2014) chooses α so that the share of the smallest size category of the actual and predicted distributions matches exactly. We also implemented this method and, though not shown, the results were qualitatively similar.

employment size of establishments in the bin, expression (2) can be written in terms of employment shares (h^e) as:[9]

$$h^e(s_i, s_j|\alpha) = s_i^{1-\alpha} - s_j^{1-\alpha}. \quad (3)$$

We obtain results for different number of size categories and for different values of the cutoff points that define these categories. We initially define three bins (called lower, middle, and upper) and vary the cutoff points to verify the sensitivity of the results to different widths of the bins. We employ the usual three bins that appear in the literature: 1 to 9, 10 to 49, and over 50 employees but as shown in the tables below we use various distinct cutoffs. Results are obtained for the plant size distribution (Table 1 for the entire formal sector and Table 2 only for the manufacturing sector) as well as for the employment share distribution (Table 3 for the entire formal sector and Table 4 only for the manufacturing sector). The second column in these tables reports the calibrated value of α that is retrieved by the method. Columns 3, 4, and 5 display the difference between the actual and predicted shares of the lower, middle, and upper categories respectively. The international evidence is only available for the employment share distribution in the manufacturing sector, so only Table 4 contains results that are comparable to those available in the literature. In Table 5 we increase the number of bins to six to see whether the results are sensitive to finer partitions of both the establishment size and the employment share distributions.

Table 1 presents the results for the plant size distribution for the whole formal sector. It reveals that there is excess concentration of small sized establishments relatively to the benchmark Pareto distribution. In contrast, the middle size category is less populated than would be predicted by the Pareto distribution for all lower and upper bounds used to define the widths of the bins. As for the upper category, column 3 shows that its share is very close to that predicted by the Pareto. These pieces of evidence thus suggest that the missing middle phenomenon is observed for the Brazilian formal sector. Interestingly, the value of α is very close to one, suggesting that the upper tail of the distribution of establishment size in the formal sector in Brazil is almost exactly inversely related to the size of the establishments.

Table 1: Actual Minus Predicted Shares of the Establishment Size Distribution for the Whole Formal Sector - Year: 2013

Width of Bins	α	Lower (L)	Middle (M)	Upper (U)
L=1-5, M=6-49, U=50+	1.010	0.086	-0.116	-0.004
L=1-5, M=6-100, U=101+	1.010	0.086	-0.119	0.000
L=1-5, M=6-200, U=201+	1.010	0.086	-0.120	0.000
L=1-6, M=7-49, U=50+	1.010	0.078	-0.097	-0.004
L=1-6, M=7-100, U=101+	1.010	0.078	-0.101	0.000
L=1-6, M=7-200, U=201+	1.010	0.078	-0.102	0.000
L=1-7, M=8-49, U=50+	1.010	0.069	-0.083	-0.004
L=1-7, M=8-100, U=101+	1.010	0.069	-0.086	0.000
L=1-7, M=8-200, U=201+	1.010	0.069	-0.087	0.000
L=1-8, M=9-49, U=50+	1.010	0.061	-0.070	-0.004
L=1-8, M=9-100, U=101+	1.010	0.061	-0.074	0.000
L=1-8, M=9-200, U=201+	1.010	0.061	-0.075	0.000
L=1-9, M=10-49, U=50+	1.010	0.053	-0.060	-0.004

[9] Letting F be the total numbers establishments and using the density of the establishment size distribution $f(S) = \frac{dF(S)}{dS} = \alpha S^{-\alpha-1}$, total employment in the economy is given by: $E(1,\infty|\alpha,F) = F\int_1^\infty sf(s)ds = \alpha F\int_1^\infty s^{-\alpha}ds = \frac{\alpha F}{\alpha-1}$. Total employment in the size range $s_i \le s \le s_j$ is: $E(s_i,s_j|\alpha,F) = F\int_{s_i}^{s_j} sf(s)ds = \alpha F\int_{s_i}^{s_j} s^{-\alpha}ds = \frac{\alpha F}{\alpha-1}\left[s_i^{1-\alpha} - s_j^{1-\alpha}\right]$. Thus, $h^e(s_i,s_j|\alpha) = \frac{E(s_i,s_j|\alpha,F)}{E(1,\infty|\alpha,F)} = s_i^{1-\alpha} - s_j^{1-\alpha}$.

L=1-9, M=10-100, U=101+	1.010	0.053	-0.064	0.000
L=1-9, M=10-200, U=201+	1.010	0.053	-0.065	0.000

Source: Authors´ estimations based on microdata from RAIS.
Notes: The sample is formed by all private, non-agricultural establishments that were born since 1995. The figures refer to the year 2013. See text for the methods used to construct the table.

Table 2 reports the results for the manufacturing sector alone. Compared to the whole formal sector, the overrepresentation of small establishments is much higher and the underrepresentation of mid-sized establishments is much deeper. Larger establishments in this sector also appear to be underrepresented but much less than the middle category. These figures also evince the presence of a missing middle in the establishment size distribution of the manufacturing sector in Brazil. As in the aggregate formal sector, the value of α is also very close to unity in the manufacturing sector.

Table 2: Actual Minus Predicted Shares of the Establishment Size Distribution - Manufacturing Sector

Width of Bins	α	Lower (L)	Middle (M)	Upper (U)
L=1-5, M=6-49, U=50+	1.010	0.224	-0.225	-0.033
L=1-5, M=6-100, U=101+	1.010	0.224	-0.244	-0.014
L=1-5, M=6-200, U=201+	1.010	0.224	-0.251	-0.006
L=1-6, M=7-49, U=50+	1.010	0.212	-0.203	-0.033
L=1-6, M=7-100, U=101+	1.010	0.212	-0.222	-0.014
L=1-6, M=7-200, U=201+	1.010	0.212	-0.230	-0.006
L=1-7, M=8-49, U=50+	1.010	0.198	-0.183	-0.033
L=1-7, M=8-100, U=101+	1.010	0.198	-0.202	-0.014
L=1-7, M=8-200, U=201+	1.010	0.198	-0.209	-0.006
L=1-8, M=9-49, U=50+	1.010	0.184	-0.165	-0.033
L=1-8, M=9-100, U=101+	1.010	0.184	-0.184	-0.014
L=1-8, M=9-200, U=201+	1.010	0.184	-0.191	-0.006
L=1-9, M=10-49, U=50+	1.010	0.171	-0.149	-0.033
L=1-9, M=10-100, U=101+	1.010	0.171	-0.168	-0.014
L=1-9, M=10-200, U=201+	1.010	0.171	-0.175	-0.006

Source: Authors´ estimations based on microdata from RAIS.
Notes: The sample is formed by all private, non-agricultural establishments that were born since 1995. The figures refer to the year 2013. See text for the methods used to construct the table.

The results for the employment share distribution across plant sizes for the formal sector are reported in Table 3. Differently from what was observed in Table 1, here there is overrepresentation of both the smallest and largest categories. Naturally, the middle of the distribution is thinner than would be expected by the Pareto distribution, so the missing middle phenomenon is also apparent when the employment share distribution is used.

Table 3: Actual Minus Predicted Employment Shares for the Whole Formal Sector – Year: 2013

Width of Bins	α	Lower (L)	Middle (M)	Upper (U)
L=1-5, M=6-49, U=50+	1.160	0.082	-0.169	0.064
L=1-5, M=6-100, U=101+	1.170	0.094	-0.208	0.090
L=1-5, M=6-200, U=201+	1.185	0.113	-0.234	0.096
L=1-6, M=7-49, U=50+	1.170	0.091	-0.155	0.043
L=1-6, M=7-100, U=101+	1.175	0.098	-0.197	0.079
L=1-6, M=7-200, U=201+	**1.190**	**0.117**	**-0.225**	**0.087**
L=1-7, M=8-49, U=50+	1.175	0.093	-0.144	0.033
L=1-7, M=8-100, U=101+	1.185	0.107	-0.184	0.059
L=1-7, M=8-200, U=201+	1.195	0.120	-0.215	0.077
L=1-8, M=9-49, U=50+	1.180	0.095	-0.134	0.023
L=1-8, M=9-100, U=101+	1.190	0.109	-0.174	0.050
L=1-8, M=9-200, U=201+	1.200	0.123	-0.206	0.068
L=1-9, M=10-49, U=50+	1.185	0.096	-0.124	0.014

L=1-9, M=10-100, U=101+	1.190	0.103	-0.167	0.050
L=1-9, M=10-200, U=201+	1.205	0.125	-0.197	0.059

Source: Authors´ estimations based on microdata from RAIS.
Notes: The sample is formed by all private, non-agricultural establishments that were born since 1995. The figures refer to the year 2013. See text for the methods used to construct the table.

Table 4 contains the results for the employment share distribution only for establishments in the manufacturing sector. Similar to Table 3, there is overconcentration of employment in the lower and upper categories and a thinner middle part than predicted by the benchmark Pareto distribution. Table 1 of Tybout (2014), which is based on the 1-9, 10-49, and 50+ partition, reports that the negative gap for the middle category in the manufacturing sector is -0.084 for India (2011), -0.085 for Indonesia (2006), and -0,030 for Mexico (2006). The corresponding figure for Brazil is -0.139, so taking the employment share distribution as the reference for caparison purposes, the missing middle phenomenon seems to be slightly stronger in Brazil than in the Asian countries and considerably more severe than in Mexico.

Table 4: Actual Minus Predicted Employment Shares for the Manufacturing Sector – Year: 2013

Width of Bins	α	Lower (L)	Middle (M)	Upper (U)
L=1-5, M=6-49, U=50+	1.090	0.069	-0.174	0.090
L=1-5, M=6-100, U=101+	1.095	0.076	-0.240	0.149
L=1-5, M=6-200, U=201+	1.100	0.083	-0.289	0.191
L=1-6, M=7-49, U=50+	1.095	0.076	-0.166	0.077
L=1-6, M=7-100, U=101+	1.100	0.083	-0.231	0.134
L=1-6, M=7-200, U=201+	1.105	0.091	-0.280	0.176
L=1-7, M=8-49, U=50+	1.100	0.081	-0.157	0.063
L=1-7, M=8-100, U=101+	1.105	0.089	-0.221	0.120
L=1-7, M=8-200, U=201+	1.110	0.097	-0.270	0.161
L=1-8, M=9-49, U=50+	1.105	0.087	-0.148	0.050
L=1-8, M=9-100, U=101+	1.110	0.095	-0.212	0.106
L=1-8, M=9-200, U=201+	1.120	0.111	-0.254	0.132
L=1-9, M=10-49, U=50+	1.110	0.092	-0.139	0.037
L=1-9, M=10-100, U=101+	1.115	0.100	-0.203	0.092
L=1-9, M=10-200, U=201+	1.125	0.117	-0.245	0.118

Source: Authors´ estimations based on microdata from RAIS.
Notes: The sample is formed by all private, non-agricultural establishments that were born since 1995. The figures refer to the year 2013. See text for the methods used to construct the table.

In Table 5 the number of bins is doubled to six in order to verify whether the patterns of the previous tables are sensitive to way the distribution is partitioned. When using the establishment size distribution (Panel A), we also see an underrepresentation of the middle part, particularly in the second bin, whose upper cutoff is 29 workers (instead of the 49 threshold used before). This is valid for the entire formal sector as well as for the manufacturing sector alone. A similar pattern emerges for the employment share distribution (Panel B) but with the underrepresentation of the middle part being more spread across the central categories.

Table 5: Actual Minus Predicted Shares of Establishment Size and Employment Shares Distributions for the Whole Formal Sector and the Manufacturing Sector

Width of Bins	α	First (1-9)	Second (10-29)	Third (30-49)	Fourth (50-99)	Fifth (100-200)	Sixth (201+)
A - Establishment Size Distribution							
A1 - Whole Formal Sector	1.010	0.053	-0.054	-0.007	-0.004	-0.001	0.000
A2 - Manufacturing Sector	1.010	0.171	-0.124	-0.026	-0.019	-0.007	-0.006
B - Employment Share Distribution							
B1 - Whole Formal Sector	1.220	0.145	-0.081	-0.036	-0.045	-0.037	0.033
B2 - Manufacturing Sector	1.175	0.196	-0.062	-0.043	-0.060	-0.048	-0.002

Source: Authors´ estimations based on microdata from RAIS.
Notes: The sample is formed by all private, non-agricultural establishments that were born since 1995. The figures refer to the year 2013. See text for the methods used to construct the table.

In sum, the method applied in this section shows an overrepresentation of small-sized establishments in the formal sector in Brazil. It also evinces that the middle part of the establishment size distribution is underrepresented, with the upper part's share corresponding to what would be expected by the Pareto distribution. Similar results apply when the employment share distribution is used. The result of a relatively thinner middle part provides support for the assertion that the missing middle phenomenon is also observed in Brazil.

5. Conclusions

Our results confirm that the middle part of the size distribution is "missing" in Brazil and apparently this feature is more intense than in other countries for which there are available results. Our analysis of the dynamics of employment over the life cycle of establishments provides some clues on why there is a missing middle in the size distribution. Considering a representative establishment, the results show that it is born small (perhaps too small) and that the pattern of the growth rate over its life cycle imposes a long time span to surpass the threshold of a mid-sized plant.

Our results also point that most of this lifecycle pattern can actually be attributed to age effects, as the application of a novel decomposition method revealed a limited scope for potential confounding determinants such as the conditions prevailing at the time the plant was born (cohort effects) or the phase of the business cycle through which the plant existed (year effects).

Focusing on establishments that are actually born small, they start very small and, though the age effects are positive and high in their first years of life, they are not strong enough to transform them into plants of middle size. In addition, their mortality rate is quite high, especially within the first three year of their lives.

As in many other countries, the segment of micro and small establishments has received a great deal of attention from public policy in Brazil. Indeed, a myriad of programs specifically targeted to this segment have been implemented by the national, state, and municipal spheres of government over the last decades. The number of initiatives and their specificities are too extensive to fit in this paper but amongst the most important ones are an ample set of national and regional programs that provide credit at low interest rates and credit guarantees to micro and small establishments, a wide program that concedes tax subsidies for establishments whose revenues are below a defined threshold, a large program of government procurement targeted to micro and small establishments, and a broad supply chain of training courses and technical assistance dedicated to help potential entrepreneurs and already established small businesses to improve their operation. As claimed by Nogueira (2014) Brazil is certainly the Latin American country with the largest and more diversified institutional framework to support this type of establishments.

Unfortunately, the effectiveness of these interventions has not been assessed, so it is difficult to say whether and to what extent they have actually affected the performance of the micro and small establishments in the country. Though our results indicate that the performance of small establishments in the country is relatively poor, it is possible that the situation would be even worse had these programs not been in place. Nevertheless, they clearly need to be redesigned, in particular towards increasing the articulation of the various initiatives within and between the three spheres of government.

References:

AUDRESCHT, D. and MATA, J. (1995) "The post-entry performance of firms: Introduction" International Journal of Industrial Organization Vol. 13, Issue 4, (Dec. 1995), pp. 413–419

Axtell, R.L. (2001): Zipft Distribution of U.S. Firm Sizes, *Science*, 293: 1818-20.

Calvino, F.; Criscuolo, C.; and Menon, C. (2015) "Cross-country evidence on start-up dynamics", *OECD Science, Technology and Industry Working Papers*, 2015/06, OECD Publishing, Paris.

Caves, R. (1998) "Industrial Organization and New Findings on the Turnover and Mobility of Firms" Journal of Economic Literature, Vol. 36, No. 4 (Dec., 1998), pp. 1947-1982

Deaton, A. and Paxson, C. (1994) "Saving, Growth and Aging in Taiwan", in WISE, D. (ed.) *Studies in the economics of aging*, University of Chicago Press.

Decker, R., Haltiwanger, J. Jarmin, R. e Miranda, J. (2014): "The Role of Entrepreneurship in US Job Creation and Economic Dynamism", *Journal of Economic Perspectives*, 28: 3-24.

Haltiwanger, J., Jarmin, R. e Miranda, J. (2013): "Who Creates Jobs? Small versus Large versus Young", *The Review of Economics and Statistics*, 95: 347-361.

Hsieh, C-T. and Olken, B.A. (2014): The Missing "Missing Middle", Journal of Economic Perspectives, 28: 89-108.

Jovanovic, B. (1982) "Selection and the Evolution of Industry". Econometrica, Vol. 50, Issue 3, p. 649-670.

Leidholm, C. and Mead, D. (1987): Small-Scale Industries in Developing Countries: Empirical Evidence and Policy Implications, International Development Paper 9, Agricultural Econ. Dept., Michigan State U.

Luttmer, E.G.T. (2007): Selection, Growth, and the Size Distribution of Firms, Quarterly Journal of Economics, 122: 1103-1114.

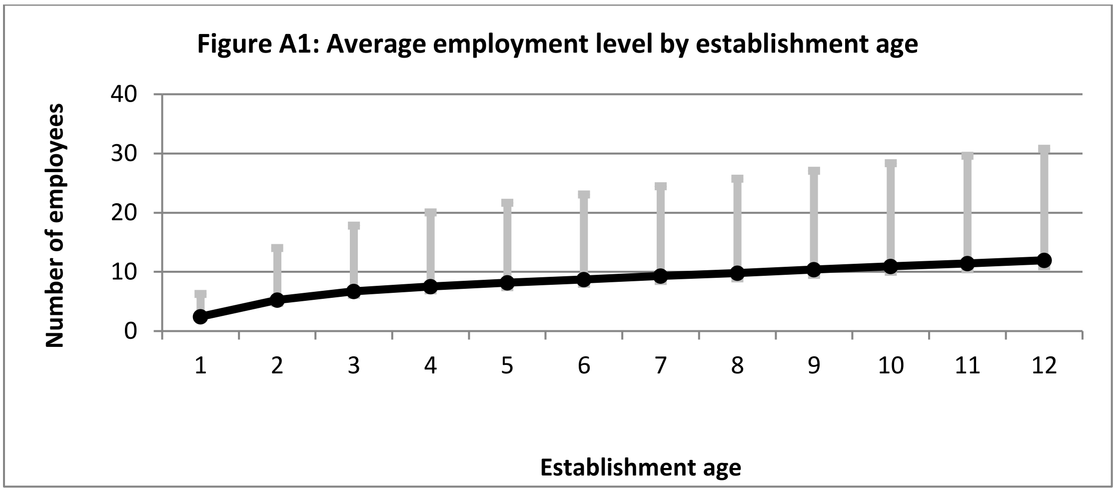
Nelson, R. and Winter, S. (1982) *An Evolutionary Theory of Economic Change*, Cambridge MA: Harvard University Press.

Nogueira, M. O. (2016) "O panorama das políticas públicas federais brasileiras voltadas para as empresas de pequeno porte" IPEA, mimeographed

Tybout, J.R. (2000): Manufacturing Firms in Developing Countries: How Well They Do, and Why?, *Journal of Economic Literature*, 38: 11-44.

Tybout, J.R. (2014): The Missing Middle, Revisited, mimeo.

***Appendix**: Complementary Data*



Source: Authors´ estimations based on microdata from RAIS.

Appendix: *Methodological details*

The model we estimate is the following :

$$averagesize = \beta + \alpha_1 age_1 + \cdots + \alpha_{10} age_{19} + \gamma_{1995} cohort_{1995} + \cdots + \gamma_{2002} cohort_{2002} + \varphi_{1995} year_{1995} + \cdots + \varphi_{2013} year_{2013} + u \qquad (1)$$

where $averagesize$ is a column vector of the average number of employees of establishments with $\boldsymbol{m}$ rows (where $\boldsymbol{m}$ is equal to the product of the number of cohorts, ages, and years), $(age_1, \ldots, age_{19})$ are column vectors of age dummies, $(cohort_{1995}, \ldots, cohort_{2002})$ are column vectors of cohort dummies and $(year_{1995}, \ldots, year_{2013})$ are column vectors of year dummies.

The model (1) cannot be estimated due perfect collinearity since the cohort is a linear combination of year and age:

$$cohort = year - age + 1 \qquad (2)$$

In order to estimate model (1), Deaton proposes a normalization that is based on the assumption that year effects capture cyclical fluctuations that have mean zero in the long run. This assumption makes the year effects orthogonal to a time-trend:

$$s_y{'}\varphi = 0 \qquad (3)$$

where s_y is an arithmetic sequence {0,1,2,3;..}.

Under assumption (3), Deaton and Paxson (1994) suggests estimating the model (1) by regressing the response vector on cohorts dummies (omitting the first cohort), age dummies (omitting the first age) and a set of T-2 years dummies, defined as follows:

$year_t^* = year_t - [(t-1995)year_{1996} - (t-1996)year_{1995}]$ for every t=1997,...,2013 (4)

The coefficients of the year$_t^*$ dummies give the coefficients of $year_{1997}, \ldots, year_{2013}$, the coefficients of $year_{1995}$ *and* $year_{1996}$ can be recovered by conditions (4) and from the fact that all year effects add to zero.