Firm Size, Age and Growth in South Africa

Masters Dissertation

Caitlin Shannon Allen
ALLCAI001

A dissertation presented for the degree of
Master of Commerce specialising in Applied Economics

Department of Economics
University of Cape Town

Supervisor:
Prof. Martin WITTENBERG
martin.wittenberg@uct.ac.za

Co-Supervisor:
Dr. Andrew KERR
andrew.kerr@uct.ac.za

Firm Size, Age and Growth in South Africa

Caitlin Shannon Allen

September 20, 2018

Abstract

The relationship between a firm's size, age and proportional growth rate is examined using multiple samples of South African firm-level data from the early to mid-2000s. The foundation of this study is Gibrat's Law of Proportionate Effect (Gibrat, 1931), which states that a firm's proportional growth rate is independent of its absolute size at the start of a given period. It is assumed that firm growth follows a random walk and, therefore, should not be affected by firm size. An implication of Gibrat's Law of Proportionate Effect is that the firm size distribution is lognormal. However, based on both empirical and theoretical literature, this theory of firm growth has fallen out of favour and been replaced by the proposal that there is an inverse relationship between a firm's proportional growth rate and both its size and age. Two questions are evaluated in this research using the samples of South African firms. The first is whether the firm size distribution is lognormal. If this is not the case then Gibrat's Law of Proportionate Effect can be rejected. However, this approach cannot confirm that Gibrat's theory is valid and will, therefore, be referred to in this paper as a partial test. It was shown that the log firm size distribution was not normal, but rather right-skewed with a Pareto distribution characterising the upper tail. Consequently, Gibrat's Law of Proportionate Effect was rejected for the datasets of South African firms. This evidence is largely observational and does not explicitly assess the relationship between proportional growth rates and firm size. Therefore, the second question is whether Gibrat's Law of Proportionate Effect holds. This was investigated by testing conditions derived from Gibrat's Law of Proportionate Effect, the results of which can lead to either the rejection or acceptance of this proposition. This study extends Gibrat's research in order to determine the relationship between firm age and proportional growth. Statistical methods, such as Ordinary Least Squares regressions, considering only firms that survived the period under consideration, were used. The results revealed that Gibrat's Law of Proportionate Effect was invalid and there was a systematic tendency for the smaller, younger South African

firms in the datasets to grow proportionally faster than the larger, older firms. This finding supports the view that firm growth is not entirely random.

Contents

1 Introduction

In South Africa, the firm is at the heart of the Government's strategy to alleviate poverty, inequality and unemployment (Herrington et al., 2015). With the National Development Plan (National Planning Commission, 2012), hereinafter referred to as the NDP, stating that 90% of new jobs are to be created by small, medium and micro enterprises by 2030, it is clear that the South African State views the small business sector as a key contributor to its development strategy. Specifically, the relationship between firm size and firm growth is of central importance when determining the capacity of these firms to be instrumental in economic development.

This dissertation utilises multiple datasets containing South African firm-level data to investigate the interaction between a firm's proportional growth rate and both its size and age. This research is relevant as understanding the nature of firm growth in South Africa can help the Government to ascertain whether the allocation of resources to programmes that support small business activities will result in economic growth.

The foundation of this paper is Gibrat's Law of Proportionate Effect (Gibrat, 1931), which states that a firm's proportional rate of growth is not dependent on its absolute size at the start of a specified period. An implication of Gibrat's Law of Proportionate Effect is that the firm size distribution is lognormal. However, the validity of Gibrat's Law of Proportionate Effect is a contentious issue, which forms the basis of an ever-growing body of literature on the subject of firm growth. A proposed alternative is that there is an inverse relationship between a firm's proportional growth rate and its size. Similarly, some studies find that a firm's age is negatively related to its proportional growth rate. In sum, there are grounds to suggest that smaller, younger firms are growing proportionally faster than larger, older firms. This theory of firm growth provides support for fiscal policies aimed at cultivating small business (Hart and Oulton, 1999).

This research focuses on two key questions using the samples of South African firms. The first is whether the firm size distribution is lognormal. If lognormality is not observed then Gibrat's Law of Proportionate Effect can be rejected. However, while necessary, the finding of a lognormal firm size distribution is not sufficient to lead to the acceptance of Gibrat's Law of Proportionate Effect. Therefore, the method of examining the shape of the firm size distribution is referred to in this paper as a partial test of Gibrat's Law of Proportionate Effect. The evidence gathered using this approach is largely observational and the relationship between proportional growth and firm size is not examined directly. Therefore, the second question

is whether Gibrat's Law of Proportionate Effect holds, in other words whether proportional growth is independent of firm size. The relationship between firm age and proportional growth is included. An evaluation of this question involves testing the conditions underpinning Gibrat's Law of Proportionate Effect and, unlike the partial test, these findings can lead to either the rejection or acceptance of this proposition.

Criticism by Davis et al. (1996) has been levelled at the quality of the data used to study small firms and job generation as much of the existing evidence is based on publicly listed firms, with a strong focus on those in the manufacturing sector. In South Africa, researchers encounter a particular challenge when trying to find firm-level data that is both comprehensive and publicly available. Some data, while appropriate for the performance of thorough investigation, is not in the public domain and other sources are often limited in sample size and regional coverage (Edwards et al., 2008; Kerr et al., 2014). This study contributes to the literature by making use of sizeable datasets from four provinces, which have been either seldom or never used before. These samples include both publicly listed and unlisted firms covering a broad range of industries.

The structure of this dissertation is as follows. Section 2 will begin with a definition of Gibrat's Law of Proportionate Effect. Thereafter, empirical studies testing the lognormality of the firm size distribution and Gibrat's Law of Proportionate Effect will be presented. Section 3 will outline the datasets used and their advantage over those employed in the literature reviewed in Section 2. Section 4 will detail the method used to answer the two key questions outlined above. Section 5 will provide the results of these analyses. Thereafter, Section 6 will consist of a discussion of these findings. Finally, Section 7 will contain a brief summary of the paper as well as the main conclusions and recommendations for future research.

2 Literature Review

Gibrat's seminal study on firm size distribution uses predominantly French data to provide a point of departure for further analysis of the actual patterns of firm growth (Gibrat, 1931). A key finding is Gibrat's Law of Proportionate Effect, hereinafter referred to as the LPE, which states that a firm's proportional growth rate is not dependent on its absolute size at the beginning of a given period. An implication of the LPE is that the firm size distribution is lognormal. Section (2.1) will begin with a definition of the LPE. Section (2.2) will consolidate the research that has explored the lognormality of the firm size distribution. Thereafter, Section (2.3) will summarise the studies that have tested whether the LPE holds.

This is a large, complicated body of literature and only the most important papers will be discussed in detail. A more comprehensive list of materials is organised in Appendix (A.1) and Appendix (A.2).

2.1 Definition of Gibrat's Law of Proportionate Effect

The LPE proposes that probability of a proportionate change in firm size for a stipulated time frame is the same for firms of all sizes (Mansfield, 1962). This is based on the assumption that a large portion of the evolution of the firm size distribution is the result of some underlying random process. The LPE gives rise to a lognormal firm size distribution.

The formulation of the LPE contained in The New Palgrave Dictionary of Economics (Mata, 2008) was used in this paper. The size of firm i at the beginning of the period is $S_{it-\delta}$, with firm size at the end of the period being denoted as S_{it}. Let the proportional growth between time t and time $t-\delta$ be represented by ϵ_{it}. If δ is equal to 1 then the following relationship holds:

$$(S_{it} - S_{it-\delta})/S_{it-\delta} = \epsilon_{it} \tag{2.1}$$

$$S_{it} = S_{it-\delta}(1+\epsilon_{it}) = S_0(1+\epsilon_{i1})(1+\epsilon_{i2})\dots(1+\epsilon_{it}). \tag{2.2}$$

Taking the natural logarithms and making the assumption that $ln(1+\epsilon_{it}) \approx \epsilon_{it}$ results in:

$$lnS_{it} = lnS_0 + \epsilon_{i1} + \epsilon_{i2} + \dots + \epsilon_{it}. \tag{2.3}$$

As $t \to \infty$ the term lnS_0 becomes small when compared to lnS_{it} and, if $\epsilon_{it} \overset{iid.}{\sim} N(\mu, \sigma_\epsilon^2)$, the approximate distribution of the log firm size is $lnS_{it} \sim N(\mu t, \sigma_\epsilon^2 t)$.

If $lnS_{it} = s_{it}$, then the resultant formulation of the LPE process is as follows:

$$s_{it} = s_{i0} + \sum \epsilon_{it}, \tag{2.4}$$

and the proportional growth rate between period t and period $t - \delta$ is:

$$g_{it} = (s_{it} - s_{it-\delta})/\delta. \tag{2.5}$$

Gibrat's (1931) book has formed the basis of an ever-expanding body of literature dedicated to firm growth. Developments in the field have relied heavily on empirical evidence, most of which is based on two main questions. Firstly, the studies that will be summarised in Section (2.2) have focused on examining whether the observed firm size distribution is lognormal and, therefore, characteristic of the LPE. This approach has dominated the earlier research and will be referred to in this study as a partial test of the LPE. This is because as the finding of a non-lognormal firm size distribution can lead to the rejection of the LPE, but no further conclusions about its validity can be drawn.

Secondly, the more recent research discussed in Section (2.3) has analysed whether or not a firm's proportional growth rate depends on its absolute size at the start of the examined time frame. More specifically, whether the LPE holds. The reason for this departure from the original method is that the partial test has low power since the relationship between firm size and proportional growth is not evaluated explicitly. In contrast, the newer method of analysing the LPE is based on the following conditions:

1. All firms have the same average proportional growth rate.
2. Proportional growth rates for all firms have the same dispersion about this common average.
3. There is no significant relationship between a firm's size and its proportional growth rate.
4. There is no serial correlation in proportional growth rates.

2.2 The lognormality of the firm size distribution

A consequence of the LPE is that the firm size distribution is lognormal.[1] As discussed previously, the finding of a non-lognormal firm size distribution can lead to the rejection of the LPE, however, this is not sufficient to confirm that the LPE

[1] Equivalently, the distribution of the log firm size, s_{it}, is normal.

is valid. Therefore, determining whether the firm size distribution is lognormal is a partial test of the LPE.

2.2.1 Firm size and lognormality

Initial evidence on the shape of the firm size distribution used data on net assets from publicly listed firms, thereby focusing on the largest companies in the economy. Hart and Prais (1956) used a dataset of all publicly listed firms from the United Kingdom, hereinafter referred to as the UK, involved in mining, manufacturing or distribution for the period 1885 to 1950 and found that the log firm size distribution was slightly right-skewed. However, they proposed that this deviation from lognormality was sufficiently small enough to be explained by sampling errors. Therefore, the lognormal firm size distribution expected to arise from the LPE was observable. They employed the following method to evaluate whether the firm size distribution was lognormal. Firstly, they presented transition matricies, which served as a summary of the movement of firms between size classes. Secondly, they generated histograms showing the firm size distribution, which were overlaid with the parametric lognormal distribution for each year. They noted that a shortcoming of this procedure was that some very large values could not be incorporated in the graph. Therefore, they suggested that observing the normality of the log firm size distribution may be more apropriate. Thirdly, they found it necessary to furnish the reader with a table of objective measures of goodness of fit, namely the kurtosis and skewness of the log firm size distribution. In their concluding remarks Hart and Prais (1956) stated that there may exist some theoretical distribution which better explains the firm size distribution, however, the simplicity of the lognormal distribution, and its relation to the normal curve, made it a useful econometric tool. When using a similar dataset for the period 1950 to 1955, Hart (1962) confirmed this finding of a lognormal firm size distribution. In both cases, the firm size distribution was consistent with the result theorised in the LPE.

While the firm size distribution was shown to be lognormal for the UK data in the early 1950s, the impact of firm entry and exit had not yet been accounted for. Simon and Bonini (1958) presented a modified version of Gibrat's model, which addressed firm entry by assuming a random walk with the constant introduction of new firms, rather than focusing only on the firms contained in the system at the start of the period. Their research was based on net asset data from 500 of the largest industrial corporations from the United States of America, hereinafter referred to as the US, for the period 1954 to 1956. They assumed that firms were born into the smallest size category at a rate that is consistent. The result was a firm size distribution with a Pareto distribution characterising the upper tail. When

drawing a comparison between Simon and Bonini's (1958) findings and the expectation of lognormality, Hart (1962) agreed that a model that incorporates firm births and deaths was superior. However, Hart (1962) argued that such an analysis could not be undertaken when considering only data from publicly listed firms as the listing or delisting by a firm was not synonymous with its birth or death. Therefore, while Simon and Bonini (1958) provided an appealing theoretical framework, their empirical results can be called into question on the grounds of too narrow a definition of firm births and deaths.

In more recent work, Stanley et al. (1995) explored the firm size distribution using a sample of 4701 publicly listed US manufacturing firms that existed in 1993. They found that, while the lognormal distribution was a good fit for the data, the upper tail was better approximated by a Pareto distribution. Therefore, for firms of all sizes, Simon and Bonini (1958) and Stanley et al. (1995) found results in contrast with the lognormal firm size distribution that characterises the LPE. Therefore, the LPE was rejected.

The caveat to be placed on the findings thus far is that they have been limited to the large firm context and based on net assets. Cabral and Mata (2003) contributed to the literature by performing a nonparametric analysis using density estimates for a sample of large, publicly listed firms, as well as a more comprehensive survey. In addition, they used number of employees as a measure of firm size. Cabral and Mata (2003) justified this methodology by stating that sampling in the data on publicly listed firms was not necessarily random, more specifically large firms were sampled with increasing probability. They began their analysis with a sample of 587 large, publicly listed Portuguese firms that existed in 1991 and, in keeping with the empirical evidence from the early 1950s, found the firm size distribution to be lognormal. Thereafter, they made use of more comprehensive survey containing the details of 33678 firms employing paid labour in the country and found that, contrary to the large firm case, the log firm size distribution was slightly right-skewed. In sum, lognormality was observed for large, publicly listed firms, but this was not the case when considering a more representative sample. Therefore, when employing the partial test of the LPE, the sampling procedure has a sizeable impact on the results.

A further consideration to be made when taking into account Cabral and Mata's (2003) finding was that, unlike the procedure followed in previous studies, number of employees was used as a proxy for firm size. This raised the question of whether the proxy for firm size used was material. Based on their results, Simon and Bonini (1958) stated that turnover, net assets, number of employees, value

added and profits were all possible candidates for a measure of firm size due to their similarly lognormal distributions. In addition, Hart and Oulton (1996) found the distributions of firm size, as measured by either turnover, net assets or employment, had similar lognormal distributions. However, later research by Segarra and Teruel (2012) based on employment and turnover data for a sample of 50000 Spanish firms showed notable differences between these variables. When considering number of employees, they found that the log firm size distribution was right-skewed, however, when using turnover the log firm size distribution approached normality. Looking at the upper tail for both proxies of firm size they found that it was best described by a Pareto distribution. Similarly, Cefis et al. (2009) considered a sample of over 50000 firms from the Netherlands and found the upper tail of the firm size distribution to be Pareto distributed for number of employees. Therefore, the proxy for firm size used has been shown to have a considerable effect on the shape of the firm size distribution.

2.2.2 Firm age and lognormality

As with entry and exit, the LPE does not include any guidance on the evolution of the firm size distribution. An additional finding of interest from Cabral and Mata's (2003) paper was that, when plotting the log firm size distribution for each age group, these distributions shifted rightward with each movement to an older age class. They used two approaches to conduct this research. Firstly, a cross-section of firms was divided into age groups and density estimates of the log firm size distribution were produced for each grouping. It was shown that these density estimates followed a rightward trend with each move to an older age category. A limitation of this method was that it made the assumption that growth conditions were constant over time. Therefore, a second longitudinal approach identified cohorts of firms as they aged with time. Cabral and Mata (2003) found the same result for this panel data. Similarly, later work by Angelini and Generale (2008) used a dataset of Italian firms covering 1992 to 2001 and found that the log firm size distribution of small firms was very right-skewed at birth and this skewness decreased as firms aged. The result that the firm size distribution approached lognormality with the addition of older firms was confirmed by Segarra and Teruel (2012).

2.2.3 Summary

When looking at net assets, the empirical evidence based on large, publicly listed firms in the early 1950s confirmed the lognormality implied by the LPE. Thereafter, later work by Simon and Bonini (1958) and Stanley et al. (1995) found the

Pareto distribution to be a better fit for the upper tail. When considering number of employees as a measure of firm size Cabral and Mata (2003), Angelini and Generale (2008) and Segarra and Teruel (2012) found that the log firm size distribution was right-skewed and a Pareto upper tail was also observed. However, they found that this right-skewed log firm size distribution approached normality as firms aged. Therefore, post-1950s, analyses based on the lognormality of the firm size distribution rejected the LPE with the exception of firms in the oldest age groups.

2.3 Gibrat's Law of Proportionate Effect

This section will discuss the papers that have tested whether the LPE holds, or rather whether a firm's size is independent of its proportional growth rate. In addition, the LPE will be extended to allow for the examination of the effect of a firm's age on its proportional growth rate. Empirical evidence has been gathered on each of the four conditions stated in Section (2.1). Condition (1) requires all firms to have the same average proportional growth rate, while Condition (2) is valid if the proportional growth rates of all firms have the same dispersion about this common average. Condition (3) is met if the firm's proportional growth rate is independent of its size. Finally, Condition (4) holds if there is no serial correlation in proportional growth rates.

2.3.1 Firm size and proportional growth

Hart and Prais (1956) introduced a formal evaluation of the validity of Conditions (1), (2) and (3) of the LPE by estimating the following equations by Ordinary Least Squares, hereinafter referred to as OLS, as follows:

$$s_{it} - E[s_{it}] = \beta(s_{it-\delta} - E[s_{it-\delta}]) + \nu_{it}, \tag{2.6}$$

where Equation (2.4) is redefined in deviation from the mean form with the assumption that the error term $\nu_{it} \overset{iid.}{\sim} N(0, \sigma_\nu^2)$. The conditional variance of the firm size distribution is given by:

$$V(s_{it}) = \beta_1^2 V(s_{it-\delta}) + \sigma_\nu^2. \tag{2.7}$$

When β_1 is equal to 1 Conditions (1) and (2) are met as firms have equi-proportionate growth rates. Furthermore, Condition (3) is upheld, meaning that firm size is not significantly linked to proportional growth. If $\beta_1 < 1$ then small firms grow proportionally faster than their larger counterparts and there is no reason for a perpetually increasing variance (Hart and Prais, 1956).

Using a sample of those firms that survived the period, Hart and Prais (1956) estimated Equation (2.7) by OLS, and found that the estimated β_1 coefficents for four of the five waves were approximately 1. Therefore, Conditions (1), (2) and (3) were met. In a later study, Hart (1962) estimated Equation (2.7) by OLS and calculated that the estimated $\beta_1 = 0.99$ and was not significantly different from 1. Therefore, Hart and Prais (1956) and Hart (1962) showed that the LPE holds and firm size was independent of proportional growth.

As discussed previously, Simon and Bonini (1958) performed an analysis using the US equivalent of large, publicly listed data for surviving firms from the early 1950s and showed a deviation of the firm size distribution from lognormality. Included in their research was a graphical test of the LPE, whereby a scatter plot of firm sizes for the start and end of a specific time frame was generated on a log scale. Simon and Bonini (1958) stated that, if the regression line has a slope of 45 degrees, then the LPE holds. They confirmed the findings of Hart and Prais (1956) and Hart (1962), namely that the estimated $\beta_1 \approx 1$ and the proportional growth rate of a firm was independent of its size. Simon and Bonini (1958) proposed that this result was justifiable on the basis of the minimum efficient scale hypothesis, which states that there is a critical minimum size for a firm above which there are constant returns to scale. As their sample excluded small firms that would have operated below this minimum threshold, all firms in the dataset faced constant average costs. Therefore, firms had the same chance of growing or downsizing in proportion to their initial size. In sum, as Simon and Bonini's (1958) study focused on large firms, the nature of firm growth for small firms was not accounted for by this theory.

All of the studies discussed so far have confirmed the validity of the LPE. In contrast, a new wave of research emerged, which provided evidence of a breakdown of the LPE. Initially, Mansfield (1962) casted doubt using a sample of practically all firms in the steel, petroleum and rubber tyre industries in the US between 1916 and 1954. In order to allow for comparison with prior research, he tested the hypothesis that the LPE holds for all firms, other than those that leave the industry. Mansfield (1962) estimated the following equation, which is based on Equation (2.4), by OLS:

$$s_{it} = \beta_0 + \beta_1 s_{it-\delta} + \nu_{it}. \tag{2.8}$$

It should be noted that $s_{it-\delta}$ in Equation (2.8) is an example of a stochastic regressor as it is not fixed in repeated samples. It is important to highlight that the

OLS regression holds, despite the fact that the regressor is stochastic, if bivariate lognormailty is assumed so that the joint distribution of s_{it} and $s_{it-\delta}$ is bivariate normal (Maddala and Lahiri, 1992). This regression tests Conditions (1), (2) and (3), and the interpretation of the estimated β_1 value is the same as discussed for Equation (2.7). Mansfield (1962) found that the β_1 estimates were not equal to 1 for four of the ten cases and in half of these cases the β_1 estimates were significantly different from 1.[2] Mansfield (1962) concluded that the LPE was not consistent across the sample and, therefore, could not be the de facto rule for measuring the size-growth relationship of firms.

Samuels (1965) contributed to the literature when he tested Conditions (1) and (2) to see if there were significant variations in the mean proportional growth rates, along with the corresponding variances, across predetermined size groups. He compared the means and variances of the proportional growth rates of the net assets of 400 publicly listed UK firms sampled in both 1951 and 1960. Samuels (1965) found that there were significant differences between the means for the different size classes, although the variances were not significantly dissimilar. Thereafter, he estimated Equation (2.7) by OLS and found an estimated $\beta_1 = 1.08$, which was significantly different from 1. This was a violation of Conditions (1), (2) and (3) as a $\beta_1 > 1$ indicated that larger firms grew proportionally faster than smaller firms, therefore, Samuels (1965) concluded that the LPE did not hold. He provided three possible reasons for this finding, the first of which was that large firms often revalue their assets more frequently than small firms and this provides the illusion of faster proportional growth. Secondly, financing is more readily available to large, publicly listed companies, thereby allowing for economies of scale in finance. The third explanation offered was that large firms experience economies of scale in marketing and production.

Using a dataset of approximately 2000 surviving, publicly listed UK firms from 21 industries in the 1950s Singh and Whittington (1975) estimated Equation (2.8) by OLS and found an estimated β_1 that was significantly greater than 1. Therefore, they determined that larger firms grew proportionally faster than their smaller counterparts and the LPE was rejected. Singh and Whittington (1975) contributed to the literature by proposing that the presence of growth persistence from one period to another was likely to result in serial correlation, which would violate Condition (4). As explained in a later paper by Dunne and Hughes (1994), if this serial correlation was positive then the estimated β_1 value, although consistent, would be biased upward in small samples. Singh and Whittington (1975) investigated

[2] Significance refers to statistical significance, which is at least at the 5% level.

the persistence of growth by calculating the mean proportional growth rate of net assets for the period 1954 to 1960, $g_{i,1960-1954}$, and regressing this on the same measure for the period 1948 to 1954, $g_{i,1954-1948}$ by OLS as follows:

$$g_{i,1960-1954} = \lambda_0 + \lambda_1 g_{i,1954-1948} + \gamma_{i,1960-1954}, \tag{2.9}$$

with the assumption that the error term $\gamma_{i,1960-1954} \overset{iid.}{\sim} N(0, \sigma_\gamma^2)$. Singh and Whittington (1975) obtained an estimate of λ_1, which was both positive and significantly different from 1. Singh and Whittington (1975) proposed that a considerable amount of the positive relationship between a firm's size and its proportional growth rate was as a result of positive serial correlation in proportional growth rates and Condition (4) was violated. They did not, however, retract their criticism of the LPE, but rather attributed much of the result to the existence of serial correlation of proportional growth rates across the examined periods.

Research undertaken in the 1970s, using asset data for surviving UK firms from the 1960s, contributed to the body of evidence that rejected the LPE. As the methods and datasets used were varied it has been difficult to make direct comparisons of these findings, but it has been established that the conclusions were mixed. Utton (1971) found a positive β_1 estimate, and attributed this to the effects of mergers and acquisitions. A complementary study of large, publicly listed firms by Meeks and Whittington (1975) showed that there was a positive relationship between a firm's size and its proportional growth rate. Conversely, Aaronovitch and Sawyer (1975) found that larger firms had no systematic tendency to grow proportionally faster than smaller firms. Deviating from the focus on UK firms, Keating (1974) employed a sample of 101 large, publicly listed Australian firms in the finance sector for the 1960s and found that higher proportional growth rates were found for larger firms.

In the mid-1980s Kumar (1985) conducted a study using number of employees as a measure of firm size and covered the 16 year period from 1960 to 1976, which was broken down into three sub-periods of surviving firms. In order to evaluate Conditions (1), (2) and (3), he estimated Equation (2.8) by OLS and found an estimated β_1 significantly below 1. Therefore, smaller firms grew proportionally faster than their larger counterparts and the LPE was rejected. Kumar (1985) supported his finding by linking it to Jovanovic's (1982) theoretical learning model, which relied on the fact that firms developed an understanding of the effects of random shocks on efficiency and either adjust and grow, or decline and potentially exit the industry. An estimated $\beta_1 < 1$ was evidence of Jovanovic's (1982) conclusion that smaller firms had higher proportional growth rates when compared to

larger firms. However, Jovanovic (1982) stated that often these rates were more variable resulting in a higher failure rate of smaller business endeavours.

2.3.2 Firm age and proportional growth

As mentioned in Section (2.2), a major shortcoming of the LPE is its failure to deal with the evolution of firms over their lifecycle. Using employment data for a sample of 17399 surviving US firms in the manufacturing industry for the period 1976 to 1982, Evans (1987) made an important contribution to the field by looking at the age-growth relationship in the context of the LPE. The firms included in the sample purchased goods on credit for the period and, therefore, small entities were uderrepresented. Evans (1987) estimated the following model by OLS [3]:

$$\begin{aligned} g_{it} =& \alpha_0 + \alpha_1 s_{it-\delta} + \alpha_2 lnA_{it} + \alpha_3 (s_{it-\delta})^2 + \alpha_4 (lnA_{it})^2 + \\ & \alpha_5 (s_{it-\delta})(lnA_{it}) + \upsilon_{it}, \end{aligned} \tag{2.10}$$

where the logarithim of age and its interaction with firm size are included as regressors, and the error term is assumed to be $\upsilon_{it} \overset{iid.}{\sim} N(0, \sigma_\upsilon^2)$. He found the average proportional growth rate across firms to be inversely related to both firm size and age. This meant that smaller, younger firms had proportionally higher growth rates than larger, more established firms and the LPE was rejected. Once again, this was in keeping with the learning model posed by Jovanovic (1982). Evans's (1987) research addressed three econometric issues. Firstly, he showed that the firm size-growth relationship was non-linear and varied across the firm size distribution. Secondly, he reported that the results passed White's specification test, therefore, the inferences were not impaired by either heteroscedasticity or an array of other specification problems.

The third issue identified by Evans (1987) was that his findings may suffer from sample selection bias. In this case, the estimated average proportional growth rates of smaller firms were potentially biased upwards due to endogeneity. The theoretical basis for this hypothesis was that slow-growing larger firms may shift toward the left-hand side of the firm size distribution for a substantial length of time before exiting, whereas, slow-growing smaller firms may exit more rapidly. As much of the investigation with regard to the relationship between proportional growth and both firm age and size was based on a balanced panel of surviving firms, the outcome that smaller firms grew proportionally faster than larger firms may be a result of slow-growing small firms exiting the sample. While this is an important aspect when assessing the validity of the LPE, tests of sample selection

[3] As mentioned previously firm size, s_{it} was included in the log form.

bias are beyond the scope of this paper.

While Evans's (1987) study of the proportional growth rate of firms over their life-cycle was conducted using a comprehensive dataset of surviving firms for specific industries, Hall (1987) and Variyam and Kraybill (1992) contributed to the literature by making use of more specific samples. The findings of Hall's (1987) study of a given cohort of 1778 publicly listed US manufacturing firms between 1972 and 1983 confirmed the inverse relationship finding of Evans (1987). While Variyam and Kraybill's (1992) study also concerned US firms, the sample covered a range of industries from a survey that included small businesses. Variyam and Kraybill (1992) used a modified version of Equation (2.10), which included dummy variables for industry and independence of the enterprise. Variyam and Kraybill's (1992) results were in keeping with those of Evans (1987) and Hall (1987), namely that the LPE was rejected as smaller, younger firms grew proportionally faster than larger, older firms.

Before conducting research based on firm age, to allow for comparison with earlier studies, Dunne and Hughes (1994) carried out three tests of the LPE based on a dataset of surviving UK companies, both publicly listed and unlisted, across the size distribution from the EXSTAT database. Firstly, they tested Conditions (1) and (2) and found that there was a threshold firm size below which average proportional growth rates were significantly different across size groups, while there was an inverse relationship between the variances of proportional growth rates and firm size. Secondly, they estimated Equation (2.8) by OLS for the period 1975 and 1980 and found β_1 estimates, which were less than 1 both overall and within each size class. This implied that smaller firms grew proportionally faster than their larger counterparts and Condition (3) of the LPE was not met. This negative relationship was significant overall and within the smallest size category for the period 1980 to 1985. Thirdly, they evaluated Condition (4), namely whether growth persistence from one period to another would result in serial correlation. Dunne and Hughes (1994) used the methodology followed by Singh and Whittington (1975) and estimated Equation (2.9) by OLS. In contrast with the findings of Singh and Whittington (1975), Dunne and Hughes (1994) found that prior growth was not a significant determinant of the potential for future growth.

Dunne and Hughes (1994) postulated that the result of an negative relationship between the variance of proportional growth rates and firm size may give way to heteroscedastic residuals in the OLS regressions. They argued that the root cause of this heteroscedasticity may be the greater stability, which characterises older firms, when compared to the more volatile proportional growth rates of younger

entities. This informed the proposition that age, rather than size, created the higher variance in proportional growth rates for the smallest cohort of enterprises. Dunne and Hughes (1994) proposed that the log of firm age should be included as a regressor in Equation (2.8), to help reduce any heteroscedasticity, as follows:

$$s_{it} = \beta_0 + \beta_1 s_{it-\delta} + \beta_2 lnA_{it} + \text{controls} + \nu_{it}. \tag{2.11}$$

They estimated Equation (2.11) by OLS for the period 1980 to 1985, and found age to be significantly and inversely related to proportional growth both overall and within each size class. This result implied that the observation that smaller, younger firms grew proportionally faster than their larger, more established counterparts was robust to issues of heteroscedasticity and the LPE did not hold. They stated that the age result was congruent with Jovanovic's (1982) learning model, while the size-growth relationship brought in the threshold effects that are consistent with neo-classical theory. These models are based on a hypothesised minimum efficient scale for firms, whereby smaller enterprises experience rapid proportional growth and declining average costs, up until a point at which the average cost curve tends to flatten and there is no longer motivation to grow.

More recent research by Hart and Oulton (2001) extended Equation (2.11) as follows:

$$s_{it} = \beta_0 + \beta_1 s_{it-\delta} + \beta_2 lnA_{it} + \beta_3 (lnA_{it})(s_{it-\delta}) + \text{controls} + \nu_{it}. \tag{2.12}$$

Hart and Oulton (2001) used a sample of 8103 large UK companies from the OneSource database, which survived the period 1986 to 1995, to estimate the regression. They found that the estimated β_1 values were less than 1 for all periods, but increased over time. In addition, the estimated coefficients on log age were both negative and significantly different from 0, which might have been due to the fact that younger firms had more growth opportunities, but that this decreased over time. The estimated coefficient on the interaction term, β_3, was small and not significantly different from 0. Overall, these results suggested that smaller, younger firms grew proportionally faster than larger, older firms, but that this tendency dissipated as firms became older. In saying this, the β_1 value in Hart and Oulton's (2001) study remained below 1 and, therefore, the LPE was rejected.

While the focus so far has been on the US and the UK, a more diverse body of evidence for developed countries emerged in the early 2000s. In order to present the most relevant results, the following studies have made use of methods consistent with those discussed above. When testing the LPE for multinational European

firms, Falk (2008) found that a firm's size had a significant, negative effect on its proportional growth rate, which was in conflict with the LPE. More country-specific data from Italian manufacturing firms (Lotti et al., 2001) and entities in the retail and manufacturing sector in Canada (Petrunia, 2008), showed that the LPE was invalid and smaller firms grew proportionally faster than larger firms. Later studies of micro and small firms in Sweden (Heshmati, 2001), publicly listed firms in Germany (Elston, 2002) and manufacturers in Spain (Calvo, 2006; Teruel-Carrizosa, 2010) also rejected the LPE.

The applicability of the LPE in the developing country case is of great importance because many established theories of firm growth fail to take into account the institutional dynamics of these markets (Sleuwaegen and Goedhuys, 2002). Empirical evidence from Swaziland, Zimbabwe, Lesotho, Botswana and two townships in South Africa was gathered by McPherson (McPherson, 1996) in order to test the proportional growth of micro and small firms in Southern Africa. It was found that the relationship between a firm's age and its proportional growth rate adhered to the principles of Jovanovic's (1982) learning model. Similarly, the size-growth relationship was also negative, rendering the LPE invalid. Liu et al. (1999) provided evidence from a newly industrialised economy in Taiwan, focused on the electronics industry, to support this finding. Studies based on manufacturing firms in both Ethiopia (Bigsten and Gebreeyesus, 2007; Gunning and Mengistae, 2001) and Cote d'Ivoire (Sleuwaegen and Goedhuys, 2002) confirmed that the LPE was invalid.

2.3.3 Summary

The study of the LPE has been pursued by various researchers. Initially, samples of large, publicly listed firms were used to show that firm size was independent of proportional growth and the LPE was accepted (Hart, 1962; Hart and Prais, 1956; Simon and Bonini, 1958). However, Mansfield (1962) reported conflicting estimates of β_1, which resulted in a host of empirical evidence rejecting the LPE. Evidence using UK data from the early 1970s showed an estimated $\beta_1 > 1$ (Meeks and Whittington, 1975; Utton, 1971), indicating the larger firms grew proportionally faster than smaller entities. Thereafter, the findings strongly suggested that $\beta_1 < 1$, meaning that smaller firms grew proportionally faster than larger firms and the LPE was violated.[4] The introduction of age as an explanatory variable for a firm's proportional growth was the result of a renewed interest in the firm life cycle. Similarly,

[4] See for example Aaronovitch and Sawyer (1975), Kumar (1985), Dunne and Hughes (1994), Hart and Oulton (1996), Hart and Oulton (2001), Lotti et al. (2001), Calvo (2006), Petrunia (2008) and Teruel-Carrizosa (2010).

an inverse age-growth relationship was found.[5] In sum, the overwhelming majority of the results from recent studies suggested that smaller, younger firms grew proportionally faster than larger, more mature entities and the LPE was violated.

3 Data

In order to test the lognormality of the firm size distribution and whether the LPE holds, firm-level data is required in panel format. In South Africa, researchers wanting to examine the economy at the firm level are hindered by a lack of data that is both comprehensive and publicly available. While some samples are appropriate for gathering empirical evidence on firm behaviour, they are not often accessible in the public domain (Edwards et al., 2008; Kerr et al., 2014). In contrast, those sources that are publicly available are, for the most part, limited in sample size and regional coverage (Edwards et al., 2008; Kerr et al., 2014). This study addresses the gap in the literature created by this data-scarcity by using multiple data sources that have either been seldom, or never, used before. The datasets employed in this research came from different regions throughout South Africa and cover most industries. When taken as a whole, the samples provided a representative picture of the economy at the firm level.

The datasets included both publicly listed and unlisted firms. While Jansen (2004) stated that publicly listed firms in South Africa provide an accurate representation of the corporate sector, Cabral and Mata (2003) found these kinds of firms to have a substantially different distribution when compared to a more comprehensive sample of all firms in the economy. In addition, each of the datasets used in this paper included smaller firms and sole proprietors. Although these firms are often excluded in firm-level datasets, the impact of this limitation on the lower tail of the firm size distribution cannot be overlooked (Hart and Oulton, 1996).

A limitation of the following datasets is that there is no complete information with regard to mergers and acquisitions. Therefore, it was not possible to disentangle the effects of internal growth from expansion as a result of a merger. Evans (1987) faced a similar problem when describing his sample, but argued that this was a reasonable sacrifice for a large quantity of data. In this case the same reasoning prevails.

[5] See for example Evans (1987), Hall (1987), Variyam and Kraybill (1992), Dunne and Hughes (1994), McPherson (1996), Liu et al. (1999), Heshmati (2001), Hart and Oulton (2001), Gunning and Mengistae (2001), Elston (2002), Sleuwaegen and Goedhuys (2002), Calvo (2006), Bigsten and Gebreeyesus (2007) and Teruel-Carrizosa (2010).

A description of the sources of the datasets that were used as well as the limitations and advantages of each will be discussed below.[6]

3.1 Regional Services Council data

Data employed in this section were based on a register of Regional Services Council's levy-paying firms and the basic information was gathered from the annual administrative records of the Regional Services Council, hereinafter referred to as the RSC. Up until 2006, when the levies were no longer payable, the compilation of these administrative records was conducted by the metropolitan and district councils in South Africa, known at the time as the RSCs. The firms included were entities employing at least one worker, or with an annual revenue in excess of R10,000 according to Section 12(1)(a) of the *Regional Services Councils Act* (South Africa, 1985).

The RSC data for KwaZulu-Natal (KwaZulu-Natal, 2017) and Pretoria (Pretoria, 2017), hereinafter referred to as the RSC KZN and RSC Pretoria data respectively, was unpublished, raw data and has not yet been used in any research at the time of writing. The firm-level aspect of the RSC data from the City of Cape Town (City of Cape Town, 2017), hereinafter referred to as the RSC Cape Town data, has only been used by Kerr (2015) at this time.

3.1.1 RSC KZN data

The RSC KZN sample was based on the administrative database of RSC levy-paying firms for 2004 and number of employees was used to represent firm size (KwaZulu-Natal, 2017).

3.1.2 RSC Pretoria data

The RSC Pretoria database was comprised of quarterly data for a sample of firms between 2004Q1 and 2004Q3 (Pretoria, 2017). Turnover was used as a measure of firm size in the RSC Pretoria dataset.

3.1.3 RSC Cape Town data

The RSC Cape Town sample was based on firm-level panel data collected by the City of Cape Town RSC between 2001 and 2006 (City of Cape Town, 2017). Included in this administrative database was annual turnover, which was used as a

[6] In all cases data was inflation adjusted to fixed December 2016 prices according to the Consumer Price Index (Statistics South Africa, September 2017).

proxy for firm size.

It should be noted that the City of Cape Town was reliant on VAT and PAYE registration reports to identify firms that were required to pay the RSC levy. Therefore, those firms that had taxable incomes below the thresholds for these levies, often informal businesses, were largely excluded from the RSC tax rolls (Bahl and Solomon, 2001). However, the 2006 Labour Force Survey (Statistics South Africa, 2006) showed that only 8% of individuals surveyed in the City of Cape Town were employed in the informal sector. Therefore, the exclusion of informal firms from the Cape Town RSC data does not pose a great threat to the validity of the results.

The RSC Cape Town data included entities, such as trusts and non-profit organisations, that are not typically classified as firms. In order to deal with this, entities without a positive wage bill were dropped from the sample as this is a good indicator of whether or not an organisation can be classified as a firm.

3.2 Combined Enterprise Survey, PICS and 2007 Enterprise Survey data

The Productivity and Investment Climate Survey (DataFirst, 2004), hereinafter referred to as the PICS, and 2007 Enterprise Survey (Enterprise Analysis Unit - World Bank Group, 2007) data sources were curated by the World Bank and used separately in prior research. It will be shown that these two samples can be combined and analysed together. This is, to the author's knowledge, the first time that this has been done. In each case firm size was measured by annual turnover. While referred to as a panel, unlike the RSC Cape Town data which was collected annually, the World Bank data has been transformed into a panel based on information collected retrospectively.

The first source of World Bank data used in this paper was the PICS, which consisted of randomly selected South African firms in the manufacturing sector that were registered with the Department of Trade and Industry in 2003 (DataFirst, 2004). The survey was carried out in Gauteng, the Western Cape, KwaZulu-Natal and the Eastern Cape and covered the period 2000 to 2002. Although the PICS focused solely on manufacturing, it included unlisted firms. A further advantage of this dataset is the vast number of variables available at the firm level such as firm age, number of employees and wage bill. However, this sample is limited as the PICS was conducted retrospectively in 2003, therefore, firms that exited the market were not included.

The second repository of World Bank data used was based on the 2007 Enterprise Survey (Enterprise Analysis Unit - World Bank Group, 2007). The survey targeted firms in Johannesburg, Cape Town, Port Elizabeth and Durban. The sample used for the 2007 Enterprise Survey was made up of two parts. Firstly, a random selection of firms was drawn from a list obtained from the Department of Trade and Industry Companies and Intellectual Property Registration Office. Retrospective data for 2003 and 2006 were collected in 2007. Secondly, an attempt was made to follow up on the sample of firms surveyed in the PICS and to collect backdated data for 2003 and 2006.

Therefore, there was an overlap of firms surveyed in both the PICS and 2007 Enterprise Survey samples. For this reason the Combined Enterprise Survey panel was created, which combined both of these datasets, in order to take advantage of the additional two years of data collected for the PICS sample. The 2007 Enterprise Survey and PICS data did not use the same identification numbers, therefore, the companies contained in both datasets could not be directly linked. However, what appeared to be the old PICS identification numbers were contained in the dataset. The validity of these identification numbers was confirmed by comparing other characteristics of the matched firms, such as location, legal status and specific type of business activity.

It is important to note that, unlike the RSC datasets, the sampling procedure used in the collection of the World Bank data means that it was not fully representative of the firm size distribution in South Africa. Specifically, these datasets included a disproportionate number of large, manufacturing firms.

4 Method

This section will be subdivided into three parts. Section (4.1) will outline the choice of variables, preparation of the data and building of the panels. Thereafter, Section (4.2) will be dedicated to the methods for determining the lognormality of the firm size distribution. Finally, Section (4.3) will focus on tests of whether the LPE holds.[7] The null hypotheses are as follows:

1. *H1*: The firm size distribution is lognormal.
2. *H2*: The proportional growth rate of each surviving firm is independent of its size for a given period and the Conditions of the LPE set out in Section (2.1)

[7] All analysis was done using Stata 14.

are met.

3. *H3*: The proportional growth rate of each surviving firm is independent of its age for a given period and the Conditions of the LPE set out in Section (2.1) are met.

4.1 Data preparation

This research made use of five complementary datasets and, while each required a different degree of preparation, the methods employed were analogous. This paper used a comprehensive combination of samples, resulting in findings that were not simply an artefact of a single sampling criteria.

Six variables were prepared for the RSC KZN, RSC Pretoria, RSC Cape Town, Combined Enterprise Survey, PICS and 2007 Enterprise Survey samples, data permitting. Firstly, a measure of firm size was determined. Turnover was used as a proxy for firm size for all, except for the RSC KZN datasets, the advantage being that this variable has a more continuous distribution when compared to number of employees. Number of employees was used for the RSC KZN data, but this indicator of firm size was limited as it was largely discrete, especially when considering smaller firms. In the case of the RSC KZN data, number of employees was recorded as starting from 0. As no employment data was missing for this sample it was assumed that firms with 0 employees were sole proprietors and, therefore, 1 was added to the number of employees across the sample.

Secondly, in order to present the data in a concise manner, each of the size scales were allocated to three size groups. Hart and Prais (1956) identified two problems when considering cut-off points for size categories. Firstly, they stated that the impact of the changes in price level during the period on the size measure should be considered. In this case, turnover data was inflation adjusted to fixed December 2016 prices according to the Consumer Price Index (Statistics South Africa, September 2017). The second issue highlighted by Hart and Prais (1956) was the determination of the end-points for each size class, which was particularly difficult when dealing with multiple datasets. They used constant intervals with absolute values. In this research, Hart and Prais's (1956) method was followed initially by using the official size category intervals, which were determined as per Section (1) of the *National Small Business Act* (South Africa, 1996).[8] The official size categories for number of employees are as follows: very small firms ($<$ 20), small firms ($>=$ 20 and $<$50), medium firms ($>=$50 and $<$200)

[8] The cut-off points are as at March 2017.

and large firms (>=200). When considering turnover, the official size categories are the following: very small firms (<R62,500,000), small firms (>=R62,500,000 and <R162,500,000), medium firms (>=R162,500,000 and <R637,500,000) and large firms (>=R637,000,000). Using these predetermined cut-off points was an effective method of allocating firms to size classes for the RSC KZN panel as number of employees was used. However, when using turnover as a measure of firm size, it was found that there were some size groups some size groups containing a disproportionately small amount of firms in them. Owing to this limitation, the approach taken in this paper for firm size, as measured by turnover, was to split each of the samples by tercile as follows: small firms (bottom tercile), medium firms (middle tercile) and large firms (top tercile). Therefore, each dataset had different cut-off points that best represented the corresponding firm size distribution.

Firm age was the third variable and was measured according to the number of years between the date that the firm commenced operations and the year or wave under consideration.[9] A fourth variable was used to allocate each firm to a specific age group which, as was the case for the categorical size variable, provided a useful simplification when comparing the large samples of firms. The fifth variable was a four digit Standard Industrial Classification code, hereinafter referred to as a SIC code, which was used to test for inter-industry effects. Where industry was categorised according to a different system, the appropriate SIC codes were generated. The sixth and final variable that was created was the proportional growth rate as per the definition in Equation (2.5).

For each sample, every firm had a corresponding response pattern, which was a string variable that was produced using a modified version of Stata's `spell` package (StataCorp, 2015). Each character of the string represented a period of the sample in order. If the firm reported positive turnover for a period then the corresponding character in the string was recorded as 1, with the absence of turnover information being indicated by 0. These response patterns were used to determine which firms to include in each panel as explained below.

For each dataset, the variables listed above were transformed into a full panel, which contained all firms that had an observation for firm size for at least one year, or wave. In other words, firms with response patterns that had at least one 1 in the string were included in the full panel.[10] For the purpose of testing Hypothesis (1)

[9] This was not the date of inception or the date of public listing, but rather that date that the firm started conducting business.

[10] The RSC KZN data was only available for 2004 and could not be transformed into a panel, however it was referred to as a full panel.

the full panel data was used as the cross-sectional nature of the firm size distribution analysis did not require continuity along the period of study.

However, once the concept of the proportional growth rate was introduced, as was the case for Hypotheses (2) and (3), it was appropriate to consider excluding firms that did not survive throughout the period. Therefore, only firms with response patterns that contained all 1s were included in the balanced panels. A reason for the use of balanced panels was that the treatment of firm deaths in the datasets was not straightforward. A firm's exclusion from a sample for a given period may be owing to a number of factors, other than a permanent exit from the market (Mata, 1994). A firm may not be present in the dataset because operations were temporarily suspended or because of the amalgamation of two or more entities (Hart and Prais, 1956). Alternatively, a firm could have failed to pay its RSC levy, or not have been followed up on in the Combined Enterprise Survey for a particular year. An additional concern is the lack of explanatory power of the LPE to account for macro-factors, which may impact the distribution of proportional growth rates (Hart, 1962). The best candidate for the construction of a balanced panel was the RSC Cape Town dataset as it covered six years, which allowed for insight into firm growth over a substantial period. In addition, balanced panels were created for the PICS and 2007 Enterprise Survey samples separately as, although the time periods were shorter, the sample size was larger when not limited to the firms that existed in both datasets.

4.2 The lognormality of the firm size distribution

This section will address the question of whether the firm size distribution is lognormal. A finding to the contrary can only lead to the rejection of the LPE and, therefore, this approach will be referred to in this paper as a partial test of the LPE. Although the LPE implies a lognormal firm size distribution, this paper will address the issue by looking at whether the log firm size distribution is normal.

Using the full panels, the following procedures were carried out to assess the lognormality of the firm size distribution for the samples of South African firms. This section will detail the advantages, limitations and justifications for each of the methods used, based on both the assumptions behind the techniques employed and the literature from Section 2.

For each dataset histograms were generated for the log firm size distributions for each period of study, and overlaid with the corresponding parametric normal distributions. This visual aid was used to illustrate whether the log firm size distribution

was best approximated by a normal distribution.

For each sample two nonparametric procedures were performed to avoid imposing structure on the data and limiting the number of assumptions required, as was the procedure followed by Cabral and Mata (2003). Firstly, a set of density estimates for the log firm size distribution was produced for each period, by making use of the Epanechnikov kernel density estimator. Secondly, a set of density estimates was generated by age group for a given year, or wave. While not formal tests of Hypothesis (1), taken together, the graphs from these two nonparametric procedures provided insight into the log firm size distribution in total, as well as for different age classes.

In support of these graphical analyses, two statistical measures of shape were used to test for the normality of the log firm size distribution. The first null hypothesis tested was based on a measure of skewness, which is the third moment of the normal distribution. For a normal distribution, the skewness is approximately equal to 0. If a distribution is right-skewed then the right tail is longer than the left tail, whereas left skewness indicates a longer left tail when compared to the right tail, both of which are a violation of the null hypothesis of a symmetrical normal distribution. The second null hypothesis for the shape of a normal distribution is a kurtosis, or fourth moment of the distribution, of approximately 3. The kurtosis is a measure of how 'peaked' the distribution is, with kurtosis values of higher than 3 indicating thin peaks and, conversely, measures of kurtosis below 3 indicating flat peaks. Two analyses of the measures of shape described above were conducted. Firstly, the skewness and kurtosis of the log firm size distribution for each dataset were calculated and compared to the null hypotheses for normality. Secondly, Stata's skewness and kurtosis test for normality, the `sktest` (StataCorp, 2015), was used to test the null hypothesis of a normal log firm size distribution.[11] Unlike the graphical evidence provided by the histograms and density estimates, the `sktest` allowed for a statistical test of Hypothesis (1).

As discussed in Section (2.2) the existence of a Pareto distributed upper tail was confirmed by Simon and Bonini (1958), Stanley et al. (1995) and Cefis and Marsili (2005). This paper adds to the literature by deviating from entirely parametric tests of the distribution of the upper tail. Instead, a nonparametric approach was taken to test the finding that the upper tail of the firm size distribution followed a Pareto distribution. This procedure necessitates a choice of the number of upper order statistics to be used in estimating the upper tail. This was done using a Hill Plot

[11] This approach implements the test described by D'agostino et al. (1990) including the empirical correction formulated by Royston (1992) (StataCorp, 2015).

(Hill, 1975).

Firstly, the order statistics for firm size, S_{it}, were reversed such that:

$$X_1 \geq X_2 \cdots \geq X_n, \tag{4.1}$$

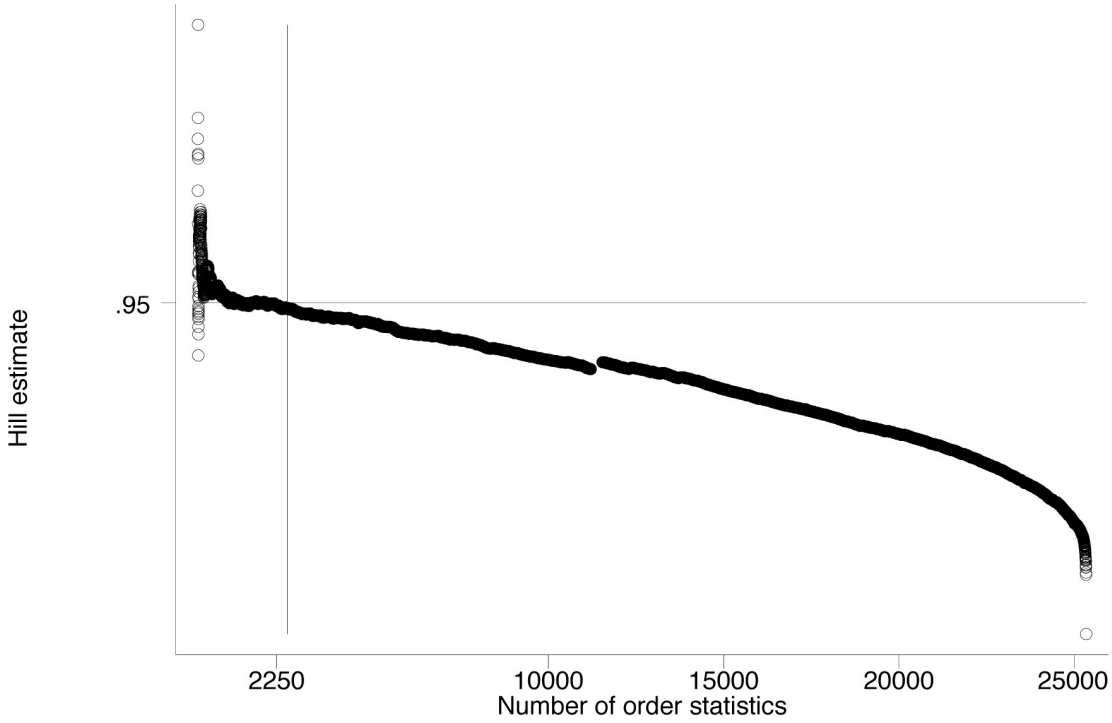
denote the order statistics of the firm size variable.

Secondly, the Hill estimator (Hill, 1975) was estimated for a subset of the highest values of the firm size variable as follows:

$$H_{k,n} = \frac{1}{\kappa} \sum_{i=1}^{\kappa} \frac{lnX_i}{lnX_{\kappa+1}} \qquad where\ \kappa \in \{1, 2, \ldots, n-1\}, \tag{4.2}$$

where κ indicates the number of upper order statistics included in the estimation. Thirdly, the subset was extended to include the next highest set of values and κ was increased. This subset was then continually expanded to incorporate all values of firm size that may have a power-like distribution. Fourthly, these Hill estimates and κ values were graphed on a Hill Plot. An example of this Hill Plot is shown in Graph (4.1) for the RSC Pretoria sample. As the order statistics for S_{it} were reversed in the first step the largest firm has order statistic 1. The κ value used to determine the cut-off value is where the graph is horizontal, just before it becomes downward sloping. In this example $\kappa = 2250$, which means that the largest 2250 firms made up the upper tail. Due to the reordering in step one the cut-off point for the upper tail of the firm size distribution was $S_{n-\kappa}$.

Figure 4.1: RSC Pretoria
Hill Plot



Source: own calculations using Pretoria (2017)

Once the cut-off point for the upper tail, $S_{n-\kappa}$, was determined a non-parametric analysis was undertaken by graphing the relationship between the upper tail and the non-logged firm size for a given period. The upper tail was calculated as follows:

$$P(X > x) = 1 - \left(\frac{S_{n-\kappa}}{S_{it}}\right)^{\theta}, \tag{4.3}$$

in other words, 1 less the cumulative distribution function. The "parametric" line seen in Graphs (C.3), (C.7), (C.11) and (C.16) used the θ parameter obtained from the actual relationship itself as shown by:

$$\ln P = \theta \ln S_{n-\kappa} - \theta \ln S_{it}. \tag{4.4}$$

In order to compare the results of the nonparametric analysis to a parametric approach the maximum likelihood estimate of θ was calculated as follows:

$$\theta_{ML} = \frac{1}{\overline{lnS_{it}} - lnS_{n-\kappa}}, \tag{4.5}$$

which has the same value of the Hill estimator.

4.3 Gibrat's Law of Proportionate Effect

The LPE stipulates that firm proportional growth is independent of absolute firm size at the start of a given period. This implies that all firms should have the same proportional rates of growth. As put forward in Section (2.1), there are four distinct Conditions used to test the LPE, all of which need to be met in order for the LPE to hold. A violation of any one of the following Conditions is sufficient to reject the LPE:

1. All firms have the same average proportional growth rate.
2. Proportional growth rates for all firms have the same dispersion about this common average.
3. There is no significant relationship between a firm's size and its proportional growth rate.
4. There is no serial correlation in proportional growth rates.

If all of these Conditions are met then Hypothesis (2) is accepted and the proportional growth rate of each surviving firm is independent of its size for a predetermined time frame. Furthermore, if age is included in the LPE framework, then meeting all of these Conditions results in an acceptance of Hypothesis (3).

An initial test of Conditions (1) and (2) was performed by calculating the geometric means and the variances of the proportional growth rates for the small, medium and large firms in each dataset. Thereafter, the significance of these differences in the means and variances across size classes was tested. As there were three size categories, the significance of any differences between the mean proportional growth rates were tested using an analysis-of-variance model, with the null hypothesis being that the means for each of the three size groups were equal. Stata's `sdtest` (StataCorp, 2015) for the significance of any inconsistencies in the variances of the proportional growth rates over the size categories was conducted, with the equivalent null hypothesis of equal variances. The advantage of the `sdtest` is that the test statistic used is robust under non-normality.[12] If it was found that variances across the size groups differed significantly then the `ttest` with a Welch correction[13] (Welch, 1947) was used as an appropriate procedure for

[12] This test is based on Brown and Forsythe's (1974) reformulation of Levene's (1960) test statistic for equality of variance (StataCorp, 2015). Brown and Forsythe's (1974) alternative statistic uses more robust measures of central tendency and is, therefore, more appropriate than the F-test when working with skewed distributions (StataCorp, 2015).

[13] This correction specifies that the approximate degrees of freedom for the t-test be determined using Welch's formula (Welch, 1947).

comparing the mean values of distributions with different variances. Conditions (1) and (2) are also applicable when considering the means and variances of the proportional growth rates across firms in different age categories. These Conditions are required to be met in order for Hypotheses (2) and (3) to be accepted, namely that the proportional growth rate of each surviving firm is not related to either its size or age and the LPE holds.

Another way to test whether the Conditions of the LPE were met was to analyse the relationship between the log of a firm's size at the start and end of a given time frame. Equation (2.8) was estimated by OLS to test whether there was a significant relationship between firm size and proportional growth, as per Condition (3), as follows:

$$s_{it} = \beta_0 + \beta_1 s_{it-\delta} + \nu_{it}.$$

This regression is linear in the logarithms owing to the fact that the joint distribution of the sample of firms at each point in time was assumed to be bivariate lognormal (Maddala and Lahiri, 1992). Therefore, the standard results from the classical OLS regression are valid, despite the presence of a stochastic regressor (Maddala and Lahiri, 1992).

The null hypothesis for this test is that the estimated β_1 is equal to 1, resulting in firms having equi-proportional growth rates as required by the LPE. Consequently, if $\beta_1 = 1$, then Conditions (1), (2) and (3) are met. Therefore, all firms have the equal average proportional growth rates, with the same variance. Furthermore, firm size is independent of proportional growth. A test of the significance of $\beta_1 = 1$ was performed using the Wald statistic. An issue raised by, among others, Evans (1987) and Dunne and Hughes (1994), was that of heteroscedasticity. Evans (1987) made use of White's specification test (White, 1980) to identify the presence of heteroscedasticity in the OLS regressions. If the specification passed the test then it was accepted that there was no heteroscedasticity present and no corrections were made (Evans, 1987). A shortcoming of this method is that the failure of an OLS regression to pass White's specification test can be due to a variety of potential issues. Therefore, it is not a given that the specific null hypothesis of homoscedastisity is being tested. Instead, a robust variance estimator for every OLS regression estimated in this paper was computed using Stata's `robust` command (StataCorp, 2015). The result of this process was heteroscedasticity-robust Wald test statistics.

The age-growth relationship was introduced to literature, and this development

has provided the basis for two modifications of Equation (2.8). Firstly, Equation (2.11) was estimated by OLS, whereby age was included as a log regressor as follows[14]:

$$s_{it} = \beta_0 + \beta_1 s_{it-\delta} + \beta_2 lnA_{it} + \nu_{it}.$$

If $\beta_1 = 1$ then Conditions (1), (2) and (3) are met, which is necessary for Hypothesis (2) to be accepted. Similarly, the coefficients on the age covariates need to be insignificant in order for null Hypothesis (3) to hold.

In addition to the above regressions the following, more intuitive, growth specifications were estimated by OLS in order to test Conditions (1), (2) and (3). As the relationship between firm proportional growth and both firm size and age can assume a non-linear form, squared terms and interaction variables were introduced as follows:

$$g_{it} = \alpha_0 + \alpha_1 s_{it-\delta} + \alpha_2 lnA_{it} + \alpha_3 (s_{it-\delta})(lnA_{it}) + \upsilon_{it}, \tag{4.6}$$

$$\begin{aligned} g_{it} =& \alpha_0 + \alpha_1 s_{it-\delta} + \alpha_2 lnA_{it} + \alpha_3 (s_{it-\delta})(lnA_{it}) + \alpha_4 s_{it-\delta}^2 + \\ & \alpha_5 lnA_{it}^2 + \upsilon_{it}. \end{aligned} \tag{4.7}$$

Finally, Condition (4), namely the potential for persistence in a firm's growth pattern, was tested as per the procedure followed by Singh and Whittington (1975). Estimating Equation (2.9) by OLS for the relevant periods, was done to evaluate whether firms having low (high) proportional growth rates in year $t - \delta$ also tended to have low (high) proportional growth rates in year t.[15]

5 Results

This section will present the results of the analysis detailed in Section (4) for each of the samples. Section (5.1) will provide a brief summary of the descriptive statistics by dataset. Thereafter, Section (5.2) will analyse the findings with regard to Hypothesis (1), namely that the firm size distribution is lognormal. This will be followed by Section (5.3), which will present the results obtained from the tests of null Hypotheses (2) and (3). Taken together these Hypotheses state that the proportional growth rate of each surviving firm is independent of both its size and age for a given period, and the Conditions of the LPE set out in Section (2.1) are

[14] OLS regressions including age were estimated using the non-response probability weights calculated by Kerr (2015).

[15] This procedure was only carried out for the RSC Cape Town sample as it could be broken down into two sub-periods.

met.

5.1 Descriptive statistics

Table (B.1) reports some basic descriptive statistics for the firm size distribution by dataset, thereby revealing prominent features of the data. At first glance there were marked differences between the means for firm size across each of the panels despite the fact that most of the data was from the early 2000s.[16] A likely reason for this variation is the composition of these datasets in terms of the distribution of firms across size categories, age groups and SIC sectors.

5.1.1 RSC KZN

Table (B.2) contains the descriptive statistics for all firms in the RSC KZN dataset with positive employee values, including sole proprietors. This sample comprised 55593 firms, of which 50% were sole traders.

The overall mean and median number of employees were 11 and 1 respectively. When excluding sole proprietors, Table (B.3) shows an increase in the mean employees to 21, with a higher median of 4. Kerr (2016) used Pay As You Earn, hereinafter referred to as PAYE, data from the South African Revenue Service and found a median of 6 employees across the firm size distribution between 2011 and 2014. This value was greater than the median for the RSC KZN dataset and this could be due to many small, informal firms not being registered for PAYE. For the RSC KZN sample, excluding sole proprietors, large firms had a mean of 1660 staff members, which was approximately 237 times greater than that of small firms. When looking at the total number of employees for all firms, these larger entities employed 58% of the work force.

Considering the descriptive statistics by industry, removing sole proprietors from the sample, 22% of firms were in the retail and wholesale industry, while 58% operated in the services sector. In addition, 8% of firms were involved in manufacturing activities. These proportions were suggestive of a sample dominated by smaller firms. Interestingly, the 46 public administration entities had a mean of 3552 employees and accounted for 27% of the total number of workers in the sample. This is consistent with the general understanding of a public entity operating on a large scale and generating many jobs.

[16] The arithmetic mean was used as the number of firms is additive in nature.

5.1.2 RSC Pretoria

Table (B.5) summarises the descriptive statistics for the RSC Pretoria full panel for the first three quarters of 2004. In total, 36715 firms were contained in the original RSC Pretoria dataset, however, this paper restricted this panel to the 27906 firms that recorded positive turnover data for at least one quarter. As this sample did not cover more than one year a balanced panel was not constructed. There were fewer firms in 2004Q2 than in the other two quarters, but the reason could not be identified due to a lack of supporting documentation.

The overall mean turnover was R4,946,469 and the corresponding median was R478,172. When categorised by size group, the mean for the smallest firms was R114,486 and the median was R111,889. In contrast, for the largest size category, the mean turnover was R15,914,495 and the median was considerably smaller at R2,630,213. This was indicative of a small number of very large firms. Following the definition of firm size categories by turnover in the *National Small Business Act* (South Africa, 1996), 99% of the firms in the sample were in the very small size group.

There was an approximately even spread of firms across the youngest three age groups, namely 0-4 years, 5-9 years and 10-19 years. In the RSC Pretoria panel there were no firms in the oldest age categories, with the oldest firm being 17 years old. The mean turnover increased monotonically by age group, with the youngest age category having a mean of R2,422,728, while the oldest firms had a mean of R11,799,614.

In 2004Q3, 20% of firms operated in the retail and wholesale sector and 68% of firms were in the services industry. Furthermore, 4% of firms were manufacturers. The mean turnover for the 19 firms in the public administration sector was 286% higher than that of the firms in the largest size category.

5.1.3 RSC Cape Town

Table (B.6) shows the descriptive statistics for the RSC Cape Town full panel between 2001 and 2006. Table (B.7) shows the equivalent values for the balanced panel of 11603 firms.[17] In terms of the full panel, 41757 firms reported positive turnover for at least one of the six years. While not monotonic, the number of firms in the RSC Cape Town full panel increased over time from 21418 in 2001 to 27597

[17] The RSC Cape Town response patterns were not included in Appendix B as there were too many possible response combinations to combine into one meaningful summary table.

in 2006, with 2162 firms entering the sample in 2003. An explanation for this influx of firms was not clear based on the data itself. Furthermore, using information gathered from the City of Cape Town, Kerr (2015) stated that an answer to this question has yet to be finalised.

Considering firm size for the full panel overall, the mean and median turnover were R23,374,647 and R2,617,396 respectively. For the balanced panel the mean and median were 29% and 67% higher than the corresponding values for the full panel respectively. Therefore, the largest firms were in the sample of survivors.

With regard to the full panel, the mean turnover for the smallest size group was R508,782, with a median of R478,905. In the largest size category, the mean of R54,782,622 was considerably higher than the corresponding median of R16,496,172. A similar trend appeared for the balanced panel. Therefore, due to the sizeable gap between the mean and median turnover for the largest size group, there were a few very large firms in the sample. According to the more stringent definition in the *National Small Business Act* (South Africa, 1996), approximately 90% of the firms in the RSC Cape Town dataset were in the very small size category.

There were marked differences in the number of firms in each age group. For both the balanced and unbalanced panels there were at most 2% of firms in the 0-4 years category. However, when combining the five age categories to form three groups the distribution was more proportionate. Considering the full panel, the youngest grouping, consisting of firms up to nine years old, made up 29% of the sample with a mean of R25,926,898. The middle age group included 40% of firms and the mean was R21,286,284. Firms over 40 years old had a mean turnover of R40,372,548. Therefore, a few older firms contributed to the high mean turnover observed for the oldest age grouping.

In terms of distribution by industry, 49% of firms were in the services sector and 25% of firms were involved in finance activities.

5.1.4 Combined Enterprise Survey, PICS and 2007 Enterprise Survey

The number of firms in each year of the Combined Enterprise Survey, PICS and 2007 Enterprise Survey panels was determined on the basis of the response patterns shown in Table (B.8). The first three years of the Combined Enterprise Survey panel, namely 2000, 2001 and 2002, were based on the PICS, whereas the data from 2003 and 2006 were based on the 2007 Enterprise Survey. A total of 1413 firms formed the Combined Enterprise Survey full panel, as each reported

positive turnover for at least one of the years covered. Furthermore, 576 firms were in the PICS full panel and 1018 firms made up the 2007 Enterprise Survey full panel.

This paper made use of the Combined Enterprise Survey full panel to investigate the lognormality of the firm size distribution. In order to disentangle the effects of the sampling procedures used in the PICS and the 2007 Enterprise Survey, the full panel descriptive statistics will be discussed for each of these datasets separately. Furthermore, to test the validity of the LPE, the two balanced panels that were created for the PICS and 2007 Enterprise Survey samples will also be summarised individually.

Table (B.10) shows the descriptive statistics for the PICS full panel. The mean turnover for the period was R718,111,990 and the median was R65,570,352. The inflated mean turnover in 2001 was due to one firm with a sizeable turnover. When looking at number of employees for the full panel, the mean was 350 and the median was 85.

For the PICS full panel the mean and median turnover for the smallest size group were R15,555,721 and R15,860,566 respectively, while the largest firms had a mean of R1,926,274,061 and a median of R470,803,920. A similar pattern occurred for the PICS balanced panel shown in (B.11), whereby the mean turnover for the largest firms of R2,071,321,878 was approximately 336% higher than the median of R474,727,664. According to the definition of a large enterprise in the *National Small Business Act* (South Africa, 1996), 15% of firms fell into the this size category and the mean was just over R4 billion.

The youngest two age classes covered 27% of the firms in the dataset, while 51% of firms were over 20 years old. The descriptive statistics by age group revealed that larger firms existed in the middle age group and firm size did not increase monotonically from the youngest to the oldest age category.

With reference to the 2007 Enterprise Survey, Table (B.12) contains the descriptive statistics for the full panel. The average mean and median were R104,593,728 and R9,360,902 respectively. Furthermore, the mean number of employees was 20, with a median of 87. When examining the 2007 Enterprise Survey balanced panel, Table (B.13) shows that the mean turnover for the period was R117,350,987, with a corresponding median of R11,959,467.

Comparing the firms across size groups, for both the full and balanced panels

the medians were less than half of the means for the largest size group, whereas the small and medium firms had median values close to their respective means. Therefore, there were a few very large firms in the dataset. In addition, based on the definition of a large enterprise in the *National Small Business Act* (South Africa, 1996), 15% of firms fell into this category with a mean of approximately R1,8 billion, while 80% of firms were considered to be very small.

Based on the full panel, as of 2006, it was found that 51% of firms were less than ten years old, with only 86 firms in the oldest age category. On average, firm size increased monotonically by age group, with the exception of the 227 firms in the middle age class. It was found that the oldest firms in the sample coincided with the firms that had the highest mean turnover.

For the full panel, 68% of firms in the 2007 Enterprise Survey data were in the manufacturing sector, while 22% of firms were in the retail and wholesale industry.

5.1.5 Summary

Table (B.1) summaries the descriptive statistics discussed above by dataset, and reveals that there was much variation across the samples. When considering number of employees as a measure of firm size, the RSC KZN sample was made up of mostly sole proprietors and firms with fewer than 4 employees. For this reason the mean and median number of employees were considerably lower than that for the manufacturing-based firms in the PICS sample. The fact that 22% of firms in the 2007 Enterprise Survey dataset operated in the retail and wholesale sector could explain why the mean number of employees of 20 was more in line with the mean for the RSC KZN panel.

Due to the design of the mechanism by which firms were allocated to specific size groups, firms were consistently evenly distributed across size categories within each of the datasets. However, as these size classes were not in absolute terms, it was challenging to gauge the distribution of firms across size groups when comparing the samples. Using the full panels and the official cut-off points for firm size by turnover in the *National Small Business Act* (South Africa, 1996), over 90% of firms in the RSC Pretoria and RSC Cape Town datasets were in the very small size category. Furthermore, 80% of firms were in the very small size category when looking at the 2007 Enterprise Survey dataset. In contrast, 49% of firms in the PICS panel were in the very small size class, with 15% being in the largest size group. Therefore, the descriptive statistics for the comprehensive RSC datasets indicated that the firm size distribution in South Africa is right-skewed due

to the presence of predominately very small firms. The PICS and 2007 Enterprise surveys comprised mostly of manufacturing firms, which explains the presence of more very large enterprises.

A major difference across all five datasets was the distribution of firms over age groups. Firms in the RSC Pretoria dataset were the youngest of all of the panels, with the oldest being only 17 years old. Conversely, the RSC Cape Town data included a few very old firms, which were predominantly in the highest turnover bracket. For the PICS sample most firms were in the oldest age group and with age came higher average turnover levels. However, half of the firms in the 2007 Enterprise Survey dataset were less than 10 years old.

Looking at the samples by industry, the RSC datasets had the broadest coverage. This was in keeping with the comprehensive nature of the administrative databases upon which these samples were based. The PICS panel was made up only of manufacturers, whereas, the 2007 Enterprise Survey sample found a middle ground between the manufacturing-focused PICS sample and the RSC Cape Town dataset. This was consistent with the sampling procedure, whereby a portion of the firms surveyed in the PICS were revisited in the 2007 Enterprise Survey.

5.2 The lognormality of the firm size distribution

This section will provide the results of the analysis of null Hypothesis (1), namely that the firm size distribution is lognormal. In order to do so the normality of the log firm size distribution will be evaluated. In addition, the upper tail of the firm size distribution will be compared to the Pareto distribution. When firm size was measured by number of employees, as was the case for the RSC KZN sample, the log firm size distribution was right-skewed with a Paretian upper tail (Angelini and Generale, 2008; Cabral and Mata, 2003; Cefis et al., 2009; Segarra and Teruel, 2012). However, based on the findings of Segarra and Teruel (2012), the log firm size distribution was normally distributed when using turnover as a proxy for firm size in a comprehensive dataset. This applies to the RSC Pretoria, RSC Cape Town, PICS and 2007 Enterprise Survey datasets. Furthermore, a normal log firm size distribution was shown to be more common for samples of large, publicly listed firms (Cabral and Mata, 2003; Hart, 1962; Hart and Prais, 1956; Stanley et al., 1995).

Initially, the normality of the log firm size distributions for the datasets was examined by generating histograms of these log firm size distributions, which were overlaid with the corresponding normal distributions. For the RSC KZN sample the first

panel of Graph (C.1) shows that the overall log firm size distribution, as measured by number of employees, was highly right-skewed. Furthermore, the remaining panels of Graph (C.1) illustrate that this right-skewness was not only due to the smallest firms in the dataset, but that the log firm size distribution was systematically right-skewed. Similarly, for the RSC Pretoria dataset, Graph (C.4) shows right-skewness of the log firm size distribution for each quarter. Considering the RSC Cape Town sample Graph (C.8) demonstrates that, for 2001, 2002 and 2005, the log firm size distributions were skewed to the right. Based on the Combined Enterprise Survey dataset Graph (C.12) depicts considerable right-skewness for the log firm size distributions for all years of the sample, other than 2006. It was unclear whether the log firm size distribution for 2006 was symmetrical, or slightly left-skewed.

Due to the discrete nature of firm size as measured by number of employees, a frequency plot was used in place of density estimates for the RSC KZN sample. Graph (C.2) confirmed the evidence so far, namely that the RSC KZN dataset had a highly right-skewed log firm size distribution.

In order to limit the assumptions made about the data, density estimates of the log firm size distribution were generated for each period using a kernel density estimator. For the RSC Pretoria sample Graph (C.5) shows density estimates for each quarter and these diagrams reinforced the finding of a right-skewed log firm size distribution. Conversely, Graph (C.9) for the RSC Cape Town dataset shows that the skewness of the density estimates was unclear. Graph (C.13) plots the density estimates for the PICS log firm size distribution and the right-skewness of the log firm size distribution is confirmed. Graph (C.14) provides the same analysis for the 2007 Enterprise Survey and, in keeping with the histograms, the 2003 density estimate was right skewed while symmetry was observed for 2006.

The graphical representations of the log firm size distributions discussed above provided mixed results for each period. However, right-skewness was observed in the majority of cases, which was not consistent with the lognormality implied by the LPE.

Unlike the observational evidence discussed above, the following method allowed for a statistical test of Hypothesis (1). Table (C.1) contains a summary of the descriptive statistics of the log firm size distributions for each of the datasets. Considering the RSC KZN sample, the statistics for skewness and kurtosis, of 1.74 and 7.69 respectively, confirmed the finding of a right-skewed log firm size distribution. Similarly, the `sktest` showed a departure from normality. For the RSC Pretoria

dataset, Table (C.1) includes the average descriptive statistics for the log firm size distribution of the RSC Pretoria dataset over all years. The measures of kurtosis and skewness were 4.45 and 0.20 respectively. Consequently, the `sktest` rejected Hypothesis (1) and the log firm size distribution was skewed to the right.

Similarly, for the RSC Cape Town sample on average, the `sktest` showed a divergence from normality and the log firm size distribution was right-skewed. As the RSC Cape Town panel covered multiple years a more detailed analysis of the log firm size distribution was conducted based on the descriptive statistics contained in Table (C.4). Right-skewness was found for 2001, 2002 and 2005 and, on the basis of the `sktest`, the null hypothesis of a symmetrical distribution was rejected. For 2003 and 2006 the `sktest` showed that the third moment was not significantly different from 0, whereas left-skewness was present in 2004. For each year the kurtosis was greater then 3. Therefore, despite the symmetry observed for two of the periods, for the panel overall Hypothesis (1) of a normal log firm size distribution was rejected.

Finally, using this approach, there was a deviation form normality for the log firm size distribution for the Combined Enterprise Survey dataset on average. As was the case for the RSC Cape Town sample, a detailed analysis was done by year as shown in Table (C.5). For the first three years covered by the PICS, based on the `sktest`, the null hypothesis of a normal log firm size distribution was rejected, instead the distribution was peaked and right-skewed. However, when considering the portion of the Combined Enterprise Survey that was covered by the 2007 Enterprise Survey, it was found that the kurtosis measure was approximately 3 for both years. In addition, while the log firm size distribution for 2003 was right-skewed, symmetry was evident in 2006. Therefore, Hypothesis (1) of a normal log firm size distribution was rejected for four of the five years of the Combined Enterprise Survey.

In sum, Hypothesis (1) of a normal log firm size distribution was rejected for each of the datasets on average. Therefore, the LPE was rejected, meaning that there was a significant relationship between firm size and proportional growth. Instead, as was found by the visual examination conducted previously, right skewness was observed.

A comparable analysis was conducted by evaluating each dataset by age group. Included in Table (C.3) for the RSC Pretoria dataset are summary statistics for the log firm size distribution by age category as in 2004Q1. Contrary to the findings of Cabral and Mata (2003), the log firm size distributions became increasingly right-

skewed over time, with the firms in the youngest age category having the most symmetrical distribution. However, on the basis of the `sktest`, Hypothesis (1) of a normal log firm size distribution for each of the age groups was rejected. These findings were confirmed by the density estimates for the RSC Pretoria panel by age group shown in Graph (C.6).

Table (C.4) summarises the descriptive statistics for the log firm size distribution by age class for the RSC Cape Town sample, as in 2001.[18] Based on the `sktest`, firms over the age of nine had symmetrical distributions. The kurtosis also decreased as with each movement to an older age group, with this shape parameter being approximately 3 for the oldest age group. Therefore, there was a convergence of the log firm size distribution toward normality as firms aged, unlike in the RSC Pretoria dataset and in keeping with Cabral and Mata's (2003) findings.

Lastly, for the Combined Enterprise Survey, Table (C.5) also breaks down the summary statistics for the log firm size distribution by age group, as in 2006. Using the `sktest` it was found that firms in the youngest age class had approximately normal firm size distributions. Thereafter, there was a deviation from normality for the two middle age groups, with a convergence on normality for the oldest age category. Therefore, it cannot be said that the log firm size distribution becomes systematically more normally distributed with each shift to an older age group. However, this result may be due to the fact that this sample is not representative of the economy as a whole, which is in contrast with the dataset used by Cabral and Mata (2003).

For the RSC Pretoria, RSC Cape Town and Combined Enterprise Survey samples it was found that the log firm size distribution shifted to the right when moving toward an older age group and this was confirmation of the rightward trend found by Cabral and Mata (2003). However, with the exception of the RSC Cape Town dataset, these log firm size distributions did not converge on normality.

Due to the unsatisfactory fit of a lognormal distribution to the firm size data, the potential for a Pareto firm size distribution in the upper tail was investigated. Both panels of Graph (C.3) for the RSC KZN sample plot the relationship between the upper tail and the total firm size, which is analogous to 1 less the CDF and was shown by the line $P(X < x)$. The "parametric" line shown in the left-hand panel used the θ parameter obtained from the relationship itself, while the maximum likelihood estimate of θ_{ML} was used in the right-hand panel. Overall, the Pareto distribution appeared to be a good fit for the upper tail of the firm size distribu-

[18] The year 2001 was used as the reference category as it was the period for which the most data on age was available.

tion for all of the samples. This is illustrated by Graph (C.7) for the RSC Pretoria, Graph (C.11) for the RSC Cape Town and Graph (C.16) for the Combined Enterprise Survey samples. Therefore, the lognormal firm size distribution implied by the LPE was not the best fit for the data.

5.3 Gibrat's Law of Proportionate Effect

The LPE states that a firm's proportional growth rate is independent of its absolute size at the start of the period of study. Section (5.2) showed that the lognormality of the firm size distribution implied by the LPE was not observed. However, as this method is a partial analysis of the LPE that is based largely on observational evidence, it follows that an examination of whether the LPE holds is warranted to supplement these findings. The results of these tests of the following Hypotheses will be given in this section. Firstly, Hypothesis (2) states that the proportional growth rate of each surviving firm is independent of its size for a given period and the Conditions of the LPE set out in Section (2.1) are met. Secondly, Hypothesis (3) states that this proportional growth rate is also independent of firm age for a specified time frame and includes the requirement that the Conditions of the LPE given in Section (2.1) are fulfilled.

Condition (1) states that all firms should have the same average proportional growth rate, while Condition (2) is met if the proportional growth rates all have the same dispersion about this common mean. Condition (3) requires a firm's proportional growth rate not to be dependent on its size. Finally, Condition (4) stipulates that the proportional growth rates should not be serially correlated.

Panels of surviving firms were most appropriate in order to conduct tests of whether the LPE held, therefore, only the RSC Cape Town, PICS and 2007 Enterprise Survey datasets were used. For the RSC Cape Town sample the balanced panel was broken down into three year groups as follows: 2001 to 2006 (2006/2001), 2001 to 2003 (2003/2001) and 2004 to 2006 (2006/2004). This was done to make the results more comparable with those from the PICS and 2007 Enterprise Survey panels.

With regard to the RSC Cape Town dataset, Table (D.4) provides the proportional growth rate descriptive statistics for the 2006/2001 year group, whereas Tables (D.5) and (D.6) present these values for the 2003/2001 and 2006/2004 year groups respectively. For the 2006/2001 year group, comparing the proportional growth rates by size category, the standard deviation decreased monotonically from 1.09 to 0.80 when moving from the smallest to the largest size class. Furthermore,

looking at the proportional growth rates by size group for the PICS sample, Table (D.9) shows that the standard deviation for the smallest size group was 1.20, while the standard deviation for the largest size category was substantially lower. For both the RSC Cape Town and PICS datasets, based on an `sdtest`, this negative relationship between firm size and proportional growth rate variability was significant. Considering the 2007 Enterprise Survey sample, Table (D.12) shows no clear pattern for the standard deviations of firms in the different size classes, however, there was a significant difference in the variances across these groups. Condition (2), which requires firms of all sizes to have equal variances, was not upheld for any of the datasets and this was sufficient to reject Hypothesis (2) and, therefore, the LPE.

Comparing the mean proportional growth rates for the RSC Cape Town dataset across size groups for the 2006/2001 year group, Table (D.4) shows that these values decreased from 0.33 for the smallest to 0.09 for the largest firms. Table (D.9) reveals a similar result for the PICS panel as, on average, the turnover for smallest firms increased by 30% for the period, while the increase for the largest firms was less than half of that. Consistent with both of these findings, was the monotonic decrease in the mean proportional growth rate over the period for the 2007 Enterprise Survey sample as shown in Table (D.12). In order to test whether these differences in mean proportional growth rates were significant, the Welch-corrected t-test was used as this procedure does not assume equal variances between size groups. It was found that there was a significant difference in mean proportional growth rates across size classes for all three datasets. Therefore, smaller firms were growing proportionally faster than larger firms, and the significance of this size-growth relationship was a violation of Condition (3). In addition, Condition (1) was not upheld as, by necessity, all firms did not have equal average proportional growth rates. As two of the Conditions were not met, Hypothesis (2) was violated and, consequently, the LPE was rejected. In other words, the proportional growth rates for these samples of surviving firms were dependent on firm size.

For the RSC Cape Town dataset Table (D.4) shows that both average proportional growth rates and their standard deviations were lowest for the oldest age group when compared to the younger firms.[19] With regard to the PICS sample, Table (D.9) indicates that there was also a decrease in the mean proportional growth rate when moving to older age classes, although this relationship was not monotonic. The same result was found for the 2007 Enterprise Survey dataset as given

[19] There was a limitation when testing for significance as there was variation in the standard deviations by age group, therefore an ANOVA was inappropriate. In addition, there were more than two size categories so a Welch-corrected t-test was not an available option.

in Table (D.12). As was found for the firm size-growth relationship, Conditions (1) and (2) were not met and this was sufficient to reject Hypothesis (3) that the proportional growth rate of a firm is independent of its age. On the whole, the LPE was rejected as this statistical analysis suggested that younger firms grew proportionally faster than their older counterparts on average.

Regression estimation was also employed to test the relationship between the log of a firm's size at the start, as well as the end of the period. The overall results of each of the regressions estimated below are summarised by dataset in Table (D.1). More detailed breakdowns of the results for the RSC Cape Town, PICS and 2007 Enterprise Survey samples are presented in Tables (D.7), (D.10) and (D.13) respectively.

Equation (2.8) was estimated by OLS with robust standard errors both overall, and within each size category as follows:

$$s_{it} = \beta_0 + \beta_1 s_{it-\delta} + \nu_{it}.$$

Using the RSC Cape Town sample, the estimated β_1 overall for the 2006/2001 year group was 0.924 and significantly different from 1. Across the size groups, the β_1 estimates were also significantly different from 1, with the smallest coefficient being in the medium size class. These findings held across the other year groups, with 2006/2004 having a higher β_1 estimate of 0.976. Considering the PICS dataset the overall finding was an estimated $\beta_1 = 0.893$ and this coefficient was also significantly different from 1. Similarly, when looking at the 2007 Enterprise Survey sample overall, it was found that the estimated $\beta_1 = 0.923$ and was significantly different from 1. Across all three samples, based on Hart and Prais's (1956) proposition, a β_1 below 1 indicated that, on average, smaller firms grew proportionally faster than larger firms. This was a violation of Condition (3) and Hypothesis (2) as the proportional growth rate was not independent of firm size and this relationship was significant. Consequently, the LPE was rejected.

In order to include age as a logarithmic regressor in this analysis, Equation (2.11) was estimated by OLS with robust standard errors as follows:

$$s_{it} = \beta_0 + \beta_1 s_{it-\delta} + \beta_2 lnA_{it} + \nu_{it}.$$

For the RSC Cape Town sample it was found that overall, for the year group 2006/2001, the estimated $\beta_1 = 0.953$, and this coefficient was significantly different from 1. The estimated β_2 for the 2006/2001 year group as a whole, was signifi-

cantly different from 0 and equal to -0.063. When considering the PICS and 2007 Enterprise Survey panels the results were not as explicit. The estimated overall value of $\beta_1 = 0.960$ for the PICS dataset, but this coefficient was not significantly different from 1. However, within each of the size classes, the estimated β_1 values were significantly different from 1. The estimated β_2 value was negative, but not significantly different from 0. With regard to the 2007 Enterprise Survey panel it was found that the estimated $\beta_1 = 0.915$ and remained significantly different from 1, however the coefficient on log age was both positive and insignificant. Taking all of these findings into consideration it was shown that, for the surviving firms in each sample, smaller firms grew proportionally faster than larger firms on average, and this relationship was significant. This was confirmation of the size-growth relationship found above. However, the effect of age on proportional growth was unclear.

At this point the initial evaluation of the standard deviations and means of the proportional growth rates for each of the datasets showed a significant and inverse relationship between proportional growth and both firm size and age. However, the regression analysis had not provided consistent results for the age-growth relationship. In an attempt to resolve this conflict growth equations in the form $g_{it} = g(s_{it-\delta}, lnA_{it})$ were estimated by OLS. The results of this approach are summarised in Table (D.2). More insight into the findings for each dataset are shown for the RSC Cape Town, PICS and 2007 Enterprise Survey samples in Tables (D.8), (D.11) and (D.14) respectively.

In addition to log initial size and log age, Equation (4.6) included a covariate to test for interaction and was estimated by OLS with robust standard errors as follows:

$$g_{it} = \alpha_0 + \alpha_1 s_{it-\delta} + \alpha_2 lnA_{it} + \alpha_3 (s_{it-\delta})(lnA_{it}) + \upsilon_{it}.$$

With regard to the RSC Cape Town panel for the 2006/2001 year group overall, the estimated $\alpha_1 = -0.028$ was significantly different from 0. This indicated that log initial firm size was weakly, inversely related to proportional growth. For the PICS and 2007 Enterprise Survey datasets the estimated α_1 values of -0.216 and -0.100 respectively were also significantly different from 0. When considering the α_2 estimate on log age for the RSC Cape Town sample overall for the 2006/2001 year group, the value was significantly different from 0 and equal to -0.124. Similarly, for the PICS and 2007 Enterprise Survey datasets there was a significant, inverse relationship between the proportional growth rates and log firm age, with overall estimated α_2 values of -1.307 and -0.534 respectively. For all three samples

the small, significant estimated α_3 coefficients on the interaction terms showed that the impact of firm size on proportional growth weakly depended on firm age. Therefore, based on the estimation of the Equation (4.6) by OLS, there was a significant and negative relationship between proportional growth and both firm size and age. Consequently, Condition (3) and Hypotheses (2) and (3) were rejected. Furthermore, it was found that smaller, younger firms grew proportionally faster than larger, older firms and these inverse relationships were significant. In keeping with the empirical evidence presented up until this point, the LPE was rejected.

Equation (4.7) was then estimated by OLS with robust standard errors, and included squared terms to take into account the potential non-linearity of the growth equation. These results are shown in Table (D.3). For the RSC Cape Town dataset this addition resulted in a larger estimated α_1 across all year groups, however, the magnitude of the estimated coefficient on log age was smaller and less significant. With regard to the PICS and 2007 Enterprise Survey samples the impact of this modification on the previous findings for the sample overall was a decrease in the strength of the age-growth relationship, and an α_1 value, which was negative though no longer significant. Therefore, the significance of the negative age-growth relationship was dependent on model specification.

As the RSC Cape Town balanced panel was broken down into two sub-periods it was possible to test whether firms with higher (lower) proportional growth rates in period t also tended to have higher (lower) proportional growth rates in period $t - \delta$. Equation (2.9) was estimated by OLS and it was found that the estimated coefficient on the previous period was $\lambda_1 = -0.051$. This weak, inverse relationship refuted the proposition that there was a tendency for the proportional growth rates of firms to persist over time and Condition (4) of the LPE was met.

In sum, smaller, younger firms were found to grow proportionally faster than larger, older firms and only weak serial correlation of proportional growth rates was observed. As three of the four Conditions were violated, there was strong evidence to confirm that the LPE did not hold for the three panels of South African firms.

The disaggregation of these growth effects by industry is of particular relevance when comparing the manufacturing and service sectors. Daunfeldt and Elert (2013) provided an example of these industry dynamics, with the finding that the LPE held in most manufacturing sectors but was invalid in the case of services. While the importance of this aspect of firm growth is acknowledged, it is beyond the scope of this paper.

Furthermore, part of the heterogeneity across the samples may be due to regional factors. This type of analysis is challenging as the specific sampling procedures are unknown for the RSC datasets owing to a lack of metadata. An example is the prevalence of sole proprietors in the RSC KZN data not being reflected for the RSC Pretoria and RSC Cape Town datasets. In addition, the RSC KZN dataset is not limited to one large city, but multiple areas in the province, which makes the firms in this dataset fundamentally different from the RSC Pretoria, RSC Cape Town, 2007 Enterprise Survey and PICS cases. Within the 2007 Enterprise Survey and PICS, the distribution across provinces was not even, with 68% of firms in the 2007 Enterprise Survey being situated in Johannesburg. The existence of varied selection criteria in terms of geographical location across the datasets does not allow for a detailed analysis of the the firm characteristics of each sample on a provincial level. These relationships are complex and the detailed analysis required to disaggregate these effects is an area for future research.

6 Discussion

Based on the given samples, this research can be summarised into three key findings about firms in South Africa. Firstly, the firm size distribution was not lognormal. Secondly, based on the panels of surviving firms, smaller firms grew proportionally faster than their larger counterparts and this relationship was significant on average. Thirdly, there was both an inverse and significant age-growth relationship.

The first insight from this research is that the lognormal distribution was not a good fit for the firm size distribution for any of the samples. Conversely, the log firm size distributions for these datasets of South African firms were right-skewed, with the Pareto distribution best characterising the upper tail. This can be partly explained by the nature of the data. As the RSC datasets were based on administrative databases, which were not biased by a non-random sampling procedure, it can be said that they were comprehensive and closely emulated the firm size distribution in South Africa. Consequently, while the literature found the lognormal distribution to be a good description for large, publicly listed firms, this was not the case when a more representative set of samples was taken into consideration. The LPE implies a lognormal firm size distribution, but lognormality was not evident in the South African samples evaluated in this paper. Therefore, on the basis of this partial test, the LPE was rejected.

Secondly, this research shows that smaller firms grew proportionally faster than larger firms for the samples of surviving South African enterprises. In part, this

finding may be accounted for by the theory of minimum efficient scale. An alternative explanation is that the proportionally faster growth of smaller firms may be linked to systematic factors such as favourable tax incentives and small business support interventions. Furthermore, the results indicated that there was a significant, negative relationship between the dispersion of proportional growth rates and firm size. Hart (1962) and Singh and Whittington (1975) proposed that the reason for this finding could be the expectation that large firms, with more diversification between products, should have a smaller dispersion of changes in turnover because they have the ability to offset product lines that are not successful against more profitable offerings.

Finally, the third point of note is that younger firms grew proportionally faster than older firms and this relationship was significant for the samples of surviving firms in South Africa. This finding is in keeping with Jovanovic's (1982) learning model, whereby firms use a Bayesian learning process to discover their actual efficiencies as they grow older. In addition, younger firms are said to be presented with more growth opportunities (Jovanovic, 1982).

These three results were consistent across datasets and, therefore, the rejection of the LPE was not an artefact of a specific sampling procedure.

7 Conclusions

According to the Gibrat model a firm's proportional growth rate is independent of its size and smaller firms do not tend to grow proportionally faster than larger firms. Furthermore, the LPE implies a lognormal firm size distribution. Based on the South African data used in this paper, a close look at the shape parameters of skewness and kurtosis of the log transformed firm size distribution, as measured by turnover, revealed right-skewness and significant deviation from normality. However, this right-skewness dissipated as firms moved toward older age groups, though in most cases there was no convergence on a normal log firm size distribution. When considering the firm size distribution for the upper tail, the Pareto distribution was a good fit for all of the samples. Therefore, the lognormal distribution expected as a result of the LPE process was not evident in any of the datasets and the LPE was rejected.

These deviations may be explained by a process whereby surviving, smaller firms grew proportionally faster than larger firms. This hypothesis was tested by making use of balanced panels of surviving firms. Firstly, a comparison across age

and size groups showed that smaller, younger firms had higher mean proportional growth rates than older, larger firms. Secondly, regression analysis of log firm size both at the start and end of the period revealed a slope coefficient less than 1. This supported the empirical literature, and showed that smaller firms grew proportionally faster than larger firms, and the LPE was rejected. By including log age in this specification it was shown that this covariate had a negative effect on a firm's proportional growth, although this relationship was not statistically significant. When the growth equation was estimated by OLS, it was shown that both firm age and firm size had a significant, negative effect on proportional growth. There was little evidence of serial correlation and, therefore, growth did not persist. Contrary to the independence of firm size and proportional growth proposed by the LPE, the systematic tendency for smaller, younger firms to grow proportionally faster than larger, older firms is a factor when explaining why firm growth is not completely random.

This research is relevant as it provides insight into the role that smaller, younger firms have to play in economic development. As stated in the NDP (National Planning Commission, 2012), the South African Government views the small business sector as an essential contributor to its economic development strategy. The finding that smaller, younger firms grew proportionally faster than their larger, older counterparts and have the capacity to contribute positively to economic growth provides some justification for interventions that provide support to these entities.

However, before using these findings to assist with policy-making with regard to small business development in South Africa, there are five areas that need further research. These gaps should be addressed using a comprehensive sample of South African firms, with broad national coverage, such as SARS data or the QES as was used by Kerr et al. (2014) and Kerr (2016). Firstly, it is necessary to identify why the smallest, youngest firms experienced high levels of proportional growth. Secondly, analysis of the impact of sample selection due to firm births and deaths is needed. Thirdly, the effects of mergers and acquisitions on the firm size distribution need to be measured as surviving firms may appear to be experiencing growth when in fact they are absorbing other entities. The fourth aspect that this paper has not explored is the applicability of the LPE across different industries. Finally, aggregation of the data by province needs to be discussed in order to take into account the diversity of economic factors in each region.

References

Aaronovitch, S. and Sawyer, M. C. 1975. Mergers, growth, and concentration. *Oxford Economic Papers*, 27(1):136–155.

Angelini, P. and Generale, A. 2008. On the evolution of firm size distributions. *The American Economic Review*, 98(1):426–438.

Bahl, R. and Solomon, D. 2001. *Furthering local government fiscal reform in South Africa: Issues, choices and challenges*. South Africa local government financial reform project, RTI International.

Bigsten, A. and Gebreeyesus, M. 2007. The small, the young, and the productive: Determinants of manufacturing firm growth in Ethiopia. *Economic Development and Cultural Change*, 55(4):813–840.

Brown, M. B. and Forsythe, A. B. 1974. Robust tests for the equality of variances. *Journal of the American Statistical Association*, 69(346):364–367.

Cabral, L. and Mata, J. 2003. On the evolution of the firm size distribution: Facts and theory. *The American Economic Review*, 93(4):1075–1090.

Calvo, J. L. 2006. Testing Gibrat's Law for small, young and innovating firms. *Small Business Economics*, 26(2):117–123.

Cefis, E. and Marsili, O. 2005. A matter of life and death: Innovation and firm survival. *Industrial and Corporate change*, 14(6):1167–1192.

Cefis, E., Marsili, O., and Schenk, H. 2009. The effects of mergers and acquisitions on the firm size distribution. *Journal of Evolutionary Economics*, 19(1):1–20.

City of Cape Town. 2017. RSC data: City of Cape Town [dataset]. Unpublished raw data.

D'agostino, R. B., Belanger, A., and D'Agostino Jr, R. B. 1990. A suggestion for using powerful and informative tests of normality. *The American Statistician*, 44 (4):316–321.

DataFirst. 2004. Productivity and Investment Climate Survey, South Africa 2004 [dataset]. DataFirst, University of Cape Town. Version 1.

Daunfeldt, S.-O. and Elert, N. 2013. When is Gibrat's law a law? *Small Business Economics*, pages 1–15.

Davis, S., Haltiwanger, J., and Schuh, S. 1996. *Job Creation and Destruction*. MA: MIT Press.

Dunne, P. and Hughes, A. 1994. Age, size, growth and survival: UK companies in the 1980s. *The Journal of Industrial Economics*, 42(2):115–140.

Edwards, L., Rankin, N., Schöer, V., et al. 2008. South African exporting firms: What do we know and what should we know? *Journal of Development Perspectives*, 4(1):67–92.

Elston, J. A. 2002. *An examination of the relationship between firm size, growth and liquidity in the Neuer Markt*. Economic Research Centre of the Deutsche Bundesbank Discussion Papers, No. 15/02, Economic Research Centre of the Deutsche Bundesbank.

Enterprise Analysis Unit - World Bank Group. 2007. Enterprise Survey, South Africa 2007 [dataset]. World Bank Group.

Evans, D. S. 1987. Tests of alternative theories of firm growth. *Journal of Political Economy*, 95(4):657–674.

Falk, M. 2008. *Testing Gibrat's Law for European Multinational Enterprises*. FIW Research Reports, No. 014, FIW - Research Centre International Economics.

Gibrat, R. 1931. *Les inégalités économiques*. Recueil Sirey.

Gunning, J. W. and Mengistae, T. 2001. Determinants of African manufacturing investment: The microeconomic evidence. *Journal of African Economies*, 10(2): 48–80.

Hall, B. H. 1987. The relationship between firm size and firm growth in the US manufacturing sector. *Journal of Industrial Economics*, 35(4):583–606.

Hart, P. E. 1962. The size and growth of firms. *Economica*, 29(113):29–39.

Hart, P. E. and Oulton, N. 1996. Growth and size of firms. *The Economic Journal*, 106(September):1242–1252.

Hart, P. E. and Oulton, N. 1999. Gibrat, Galton and job generation. *International Journal of the Economics of Business*, 6(2):149–164.

Hart, P. E. and Oulton, N. 2001. Galtonian regression, company age and job generation 1986–95. *Scottish Journal of Political Economy*, 48(1):82–98.

Hart, P. E. and Prais, S. J. 1956. The analysis of business concentration: A statistical approach. *Journal of the Royal Statistical Society. Series A (General)*, 119(2):150–191.

Herrington, M., Kew, J., and Kew, P. 2015. GEM South Africa report: South Africa: The crossroads–a goldmine or a time bomb. *Global Entrepreneurship Monitor*.

Heshmati, A. 2001. On the growth of micro and small firms: Evidence from Sweden. *Small Business Economics*, 17(3):213–228.

Hill, B. M. 1975. A simple general approach to inference about the tail of a distribution. *The Annuals of Statistics*, 3(5):1163–1174.

Jansen, Z. 2004. *Note on the Funding Structure of Non-Financial Companies, 1990-2003*. South African Reserve Bank Quarterly Bulletin, June, South African Reserve Bank.

Jovanovic, B. 1982. Selection and the evolution of industry. *Econometrica: Journal of the Econometric Society*, 50(3):649–670.

Keating, G. 1974. Gibrat's Law and the growth of firms. *Australian Economic Papers*, 13(23):281–286.

Kerr, A. 2015. *The life cycle of South African manufacturing firms*. A DataFirst Technical Paper, No. 33, DataFirst, University of Cape Town.

Kerr, A. 2016. *Job flows, worker flows, and churning in South Africa*. UNU-WIDER Working Paper, No. 2016/37, United Nations University World Institute for Development Economics Research.

Kerr, A., Wittenberg, M., and Arrow, J. 2014. Job creation and destruction in South Africa. *South African Journal of Economics*, 82(1):1–18.

Kumar, M. S. 1985. Growth, acquisition activity and firm size: Evidence from the United Kingdom. *The Journal of Industrial Economics*, 33(3):327–338.

KwaZulu-Natal. 2017. RSC data: KwaZulu-Natal [dataset]. Unpublished raw data.

Levene, H. 1960. Robust tests for equality of variances. *Contributions to probability and statistics*, 1:278–292.

Liu, J.-T., Tsou, M.-W., and Hammitt, J. K. 1999. Do small plants grow faster? Evidence from the Taiwan electronics industry. *Economics Letters*, 65(1):121–129.

Lotti, F., Santarelli, E., and Vivarelli IV, M. 2001. The relationship between size and growth: The case of Italian newborn firms. *Applied Economics Letters*, 8 (7):451–454.

Maddala, G. S. and Lahiri, K. 1992. *Introduction to econometrics*, volume 2. Macmillan New York.

Mansfield, E. 1962. Entry, Gibrat's Law, innovation, and the growth of firms. *The American Economic Review*, 52(5):1023–1051.

Mata, J. *The New Palgrave Dictionary of Economics*, chapter Gibrat's Law. Palgrave Macmillan, 2nd edition, 2008.

Mata, J. 1994. Firm growth during infancy. *Small Business Economics*, 6(1): 27–39.

McPherson, M. A. 1996. Growth of micro and small enterprises in Southern Africa. *Journal of Development Economics*, 48(2):253–277.

Meeks, G. and Whittington, G. 1975. Giant companies in the United Kingdom 1948-69. *The Economic Journal*, 85(340):824–843.

National Planning Commission. 2012. *National Development Plan 2030: Our future – make it work*. Presidency of South Africa, Pretoria.

Petrunia, R. 2008. Does Gibrat's Law hold? Evidence from Canadian retail and manufacturing firms. *Small Business Economics*, 30(2):201–214.

Pretoria. 2017. RSC data: Pretoria [dataset]. Unpublished raw data.

Royston, P. 1992. Tests for departure from normality. *Stata Technical Bulletin*, 1 (2).

Samuels, J. M. 1965. Size and the growth of firms. *The Review of Economic Studies*, 32(2):105–112.

Segarra, A. and Teruel, M. 2012. An appraisal of firm size distribution: Does sample size matter? *Journal of Economic Behavior & Organization*, 82(1):314–328.

Shanmugam, K. and Bhaduri, S. N. 2002. Size, age and firm growth in the Indian manufacturing sector. *Applied Economics Letters*, 9(9):607–613.

Simon, H. A. and Bonini, C. P. 1958. The size distribution of business firms. *The American Economic Review*, 48(4):607–617.

Singh, A. and Whittington, G. 1975. The size and growth of firms. *The Review of Economic Studies*, 42(1):15–26.

Sleuwaegen, L. and Goedhuys, M. 2002. Growth of firms in developing countries, evidence from Cote d'Ivoire. *Journal of Development Economics*, 68(1):117–135.

South Africa. 1985. *Regional Services Council Act, No 109 of 1985*. Pretoria: Government Printer.

South Africa. 1996. *National Small Business Act, No 102 of 1996*. Pretoria: Government Printer.

Stanley, M. H., Buldyrev, S. V., Havlin, S., Mantegna, R. N., Salinger, M. A., and Stanley, H. E. 1995. Zipf plots and the size distribution of firms. *Economics Letters*, 49(4):453–457.

StataCorp. 2015. *Stata glossary and index: Release 14*. StataCorp LLC, Texas.

Statistics South Africa. 2006. Labour Force Survey, March: 2006. Statistics South Africa.

Statistics South Africa. September 2017. Statistical Release: Consumer Price Index. Statistics South Africa.

Teruel-Carrizosa, M. 2010. Gibrat's Law and the learning process. *Small Business Economics*, 34(4):355–373.

Utton, M. A. 1971. The effect of mergers on concentration: UK manufacturing industry, 1954-65. *The Journal of Industrial Economics*, 20(1):42–58.

Variyam, J. N. and Kraybill, D. S. 1992. Empirical evidence on determinants of firm growth. *Economics Letters*, 38(1):31–36.

Welch, B. L. 1947. The generalization of student's' problem when several different population variances are involved. *Biometrika*, 34(1/2):28–35.

White, H. 1980. A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity. *Econometrica: Journal of the Econometric Society*, 48(4):817–838.

A Appendix: Summary of the literature

Table A.1: Section (2.2) The lognormality of the firm size distribution

Study	Country (Period)	Number of firms	Sample of firms	Measure of firm size	Firm size distribution (FSD)
Hart and Prais (1956)	**UK (1885-1950)**	Vary by period	Large, listed	Net assets	Lognormal FSD
Hart (1962)	**UK (1950-1955)**	1981	Large, listed	Net assets	Lognormal FSD
Simon and Bonini (1958)	**US (1954-1956)**	500	Large	Net assets	Yule FSD, with Pareto in upper tail
Stanley et al. (1995)	**US (1993)**	4701	Large, listed	Net assets	Lognormal FSD, with Pareto in upper tail
Cabral and Mata (2003)	**Portugal (1991)**	587	Large, listed	Employees	Lognormal FSD
Cabral and Mata (2003)	**Portugal (1991)**	33678	Comprehensive	Employees	Slightly right-skewed log FSD
Angelini and Generale (2008)	**Italy (1992-2201)**	Vary by period	Comprehensive	Employees	Very right-skewed log FSD
Cefis et al. (2009)	**Netherlands (1996-2003)**	50000+	Comprehensive	Employees	Pareto in upper tail
Segarra and Teruel (2012)	**Spain (2001-2006)**	50000	Comprehensive	Employees	Very right-skewed log FSD, with Pareto in upper tail
Segarra and Teruel (2012)	**Spain (2001-2006)**	50000	Comprehensive	Turnover	Lognormal FSD

Table A.2: Section (2.3) Gibrat's Law of Proportionate Effect

Study	Country (Period)	Number of firms	Sample of firms	LPE test results	Methodology
Hart and Prais (1956)	**UK (1885-1950)**	Vary by period	Large, listed	LPE holds: $\beta \approx 1$	Equation (2.7)
Hart (1962)	**UK (1950-1955)**	1981	Large, listed	LPE holds: $\beta \approx 1$	Equation (2.7)
Simon and Bonini (1958)	**US (1954-1956)**	500	Large	LPE holds	Graphical analysis
Mansfield (1962)	**US (vary by industry)**	Vary by industry	Specific industries	LPE violated in 40% of cases	Equation (2.8)
Samuels (1965)	**UK (1951-1960)**	400	Listed	LPE violated: $\beta > 1$ & large grow faster	Equation (2.7)
Singh and Whittington (1975)	**UK (1948-1960)**	approx. 2000	Listed	LPE violated: $\beta > 1$ & large grow faster	Equation (2.8)
Utton (1971)	**UK (1954-1965)**	1527	Specific industries	LPE violated: $\beta > 1$ & large grow faster	Equation (2.7)
Keating (1974)	**Australia (1961-1969)**	101	Large, listed	LPE violated: large grow faster	Distributional moments
Meeks and Whittington (1975)	**UK (1948-1969)**	Vary by period	Large, listed	LPE violated: large grow faster	Distributional moments
Aaronovitch and Sawyer (1975)	**UK (1958-1967)**	233	Large	LPE holds: $\beta \approx 1$	Equation (2.8)
Kumar (1985)	**UK (1960-1976)**	Vary by period	Listed	LPE violated: $\beta < 1$ & small grow faster	Equation (2.8)
Evans (1987)	**US (1976-1982)**	17399	Specific industries	LPE violated: small, young grow faster	Equation (2.10)
Hall (1987)	**US (1972-1983)**	1778	Listed	LPE violated: small, young grow faster	Equation (2.10)
Variyam and Kraybill (1992)	**US (1986-1991)**	422	Small	LPE violated: small, young grow faster	Equation (2.10) + controls
Dunne and Hughes (1994)	**UK (1975-1985)**	Vary by period	Comprehensive	LPE violated: $\beta < 1$ & small, young grow faster	Equation (2.11)
Hart and Oulton (1996)	**UK (1989-1993)**	87109	Comprehensive	LPE violated: $\beta < 1$ & small grow faster	Equation (2.8)
Hart and Oulton (2001)	**UK (1986-1995)**	8103	Large	LPE violated: $\beta < 1$ & small, young grow faster	Equation (2.12)
Heshmati (2001)	**Sweden (1993-1998)**	Vary by period	Micro- and small	LPE violated: small, young grow faster	Equation (2.10) + controls
Lotti et al. (2001)	**Italy (1987-1993)**	1570	Specific industries	LPE violated: $\beta < 1$ & small grow faster	Equation (2.8)
Elston (2002)	**Germany (1997-2000)**	287	Listed	LPE violated: small, young grow faster	Simplified Equation (2.10)
Calvo (2006)	**Spain (1990-2000)**	967	Specific industries	LPE violated: $\beta < 1$ & small, young grow faster	Equation (2.12)
Petrunia (2008)	**Canada (1984-1996)**	17656	Specific industries	LPE violated: $\beta < 1$ & small grow faster	Equation (2.8)
Falk (2008)	**15 OECD-countries (2000-2004)**	20000	Multinational	LPE violated: small grow faster	Simplified Equation (2.10)
Teruel-Carrizosa (2010)	**Spain (1994-2002)**	139992	Small	LPE violated: $\beta < 1$ & small, young grow faster	Equation (2.11) + quadratic age
McPherson (1996)	**5 Southern African countries**	Vary by country	Micro- and small	LPE violated: small, young grow faster	Equation (2.10) + controls
Liu et al. (1999)	**Taiwan (1990-1994)**	915	Specific industries	LPE violated: small, young grow faster	Equation (2.10) + controls
Gunning and Mengistae (2001)	**Ethiopia (1983-1993)**	220	Specific industries	LPE violated: small, young grow faster	Equation (2.11)
Shanmugam and Bhaduri (2002)	**India (1989-1993)**	392	Specific industries	LPE violated: small, old grow faster	Equation (2.10)
Sleuwaegen and Goedhuys (2002)	**Cote d'Ivoire (1989-1994)**	185	Specific industries	LPE violated: small, young grow faster	Equation (2.10)
Bigsten and Gebreeyesus (2007)	**Ethiopia (1996-2003)**	5542	Specific industries	LPE violated: small, young grow faster	Equation (2.10)

B Appendix: Data description

B.1 Summary

Table B.1: Summary of firm size
Descriptive statistics

Dataset	Period	Measure of firm size	Number of firms	Mean	Median
RSC KZN	**2004**	Employees	55593	11	1
RSC KZN (excluding sole)	**2004**	Employees	27777	21	4
Combined Enterprise Survey	**2000-2006**	Employees	1018	209	48
PICS	**2000-2002**	Employees	571	350	85
2007 Enterprise Survey	**2003-2006**	Employees	1018	20	87
RSC Pretoria	**2004**	Turnover	27906	4,946,469	478,172
RSC Cape Town	**2001-2006**	Turnover	41757	23,374,647	2,617,396
Combined Enterprise Survey	**2000-2006**	Turnover	1413	391,315,148	25,090,047
PICS	**2000-2002**	Turnover	576	718,111,990	65,570,352
2007 Enterprise Survey	**2003-2006**	Turnover	1018	104,593,728	9,360,902

Source : own calculations using KwaZulu-Natal (2017), Pretoria (2017), City of Cape Town (2017), Enterprise Analysis Unit - World Bank Group (2007) and DataFirst (2004).

Notes :

1. Number of firms with firm size data for any period.
2. Arithmetic mean used for this section as number of firms is additive in nature.

B.2 RSC KZN

Table B.2: Employees descriptive statistics - RSC KZN 2004

Category	Number of firms	% of firms	Number of employees	Mean	Median
Sole traders					
Sole traders	27816	50.04	27816	1	1
Not sole traders	27777	49.96	588955	21	4
Total	55593	100.00	616771	11	1
Size category					
Small firm	54647	98.30	204882	4	1
Medium firm	739	1.33	68358	93	82
Large firm	207	0.37	343531	1660	366
Total	55593	100.00	616771	11	1
SIC Code					
Agriculture, Forestry and Fishing	199	0.36	2504	13	2
Construction	2063	3.71	17104	8	2
Manufacturing	3856	6.94	79989	21	2
Transport, Comms and Electricity	2201	3.96	18577	8	2
Retail and Wholesale	12340	22.20	73510	6	1
Finance	4623	8.32	60339	13	1
Services	30224	54.37	201306	7	2
Public Admin	87	0.16	163442	1879	3
Total	55593	100.00	616771	11	1

Source: own calculations using KwaZulu-Natal (2017)

Table B.3: Employees descriptive statistics (excluding sole traders) - RSC KZN 2004

Category	Number of firms	% of firms	Number of employees	Mean	Median
Size category					
Small firm	26831	96.60	177066	7	4
Medium firm	739	2.66	68358	93	82
Large firm	207	0.74	343531	1660	366
Total	27777	100.00	588955	21	4
SIC Code					
Agriculture, Forestry and Fishing	103	0.37	2408	23	9
Construction	1036	3.73	16077	16	5
Manufacturing	2072	7.46	78205	38	9
Transport, Comms and Electricity	1155	4.16	17531	15	4
Retail and Wholesale	6075	21.87	67245	11	4
Finance	1224	4.41	56940	47	3
Services	16066	57.84	187148	12	4
Public Admin	46	0.17	163401	3552	10
Total	27777	100.00	588955	21	4

Source: own calculations using KwaZulu-Natal (2017)

B.3 RSC Pretoria

Table B.4: Response pattern for turnover, by firm - RSC Pretoria

Pattern	Number of firms
000	8809
001	1370
010	340
011	856
100	2531
101	490
110	223
111	22096
Total	36715

Source: own calculations using Pretoria (2017)

Notes :

1. Each character of the string represents a period of the sample in order. If the firm reported positive turnover for a period then the corresponding character in the string was recorded as 1, with the absence of turnover information being indicated by a 0.

Table B.5: Turnover descriptive statistics - RSC Pretoria
Full panel

Category	Number of firms	% of firms	Mean	Median
Wave				
2004Q1	25340	90.80	5,096,388	532,221
2004Q2	23515	84.27	3,664,616	374,788
2004Q3	24812	88.91	6,008,207	560,330
Total	27906	100.00	4,946,469	478,172
Size category 2004Q3				
Small firm	7528	30.34	114,486	111,889
Medium firm	8241	33.21	521,661	498,909
Large firm	9043	36.45	15,914,495	2,630,213
Total	24812	100.00	6,008,207	560,330
Official size category 2004Q3				
Very small firm	24564	99.00	2,218,221	548,988
Small firm	151	0.61	99,745,667	88,276,237
Medium firm	80	0.32	301,347,517	280,415,794
Large firm	17	0.07	3,259,874,363	1,334,532,914
Total	24812	100.00	6,008,207	560,330
Age group 2004Q3				
0-4 years	8965	36.13	2,422,728	446,300
5-9 years	7156	28.84	3,466,383	595,629
10-19 years	8691	35.03	11,799,614	688,208
Total	24812	100.00	6,008,207	560,330
SIC code 2004Q3				
Agriculture, Forestry and Fishing	173	0.70	1,329,704	410,628
Mining	30	0.12	4,721,577	1,008,171
Construction	924	3.72	4,441,573	604,261
Manufacturing	890	3.59	25,529,118	1,477,262
Transport, Comms and Electricity	482	1.94	33,873,442	619,131
Retail and Wholesale	4883	19.68	6,807,327	900,603
Finance	369	1.49	71,334,798	572,210
Services	16785	67.65	2,597,703	461,666
Public Admin	19	0.08	61,465,739	1,477,518
Other	257	1.04	4,743,215	649,076
Total	24812	100.00	6,008,207	560,330

Source: own calculations using Pretoria (2017)

Notes :

1. Total number of firms for the wave section is the number of firms with turnover information for any period.
2. Total number of firms for the rest of the sections is the number of firms with turnover information for 2004Q3.

B.4 RSC Cape Town

Table B.6: Turnover descriptive statistics - RSC Cape Town
Full panel

Category	Number of firms	% of firms	Mean	Median
Year				
2001	21418	51.29	25,598,589	2,355,302
2002	20592	49.31	22,913,697	2,175,192
2003	23021	55.13	25,714,810	2,743,565
2004	24026	57.54	20,338,133	2,709,933
2005	27894	66.80	25,587,847	2,884,026
2006	27597	66.09	20,447,044	2,702,124
Total	41757	100.00	23,374,647	2,617,396
Percentile size category 2006				
Small firm	8556	31.00	508,782	478,905
Medium firm	9281	33.63	2,720,132	2,453,881
Large firm	9760	35.37	54,782,622	16,496,172
Total	27597	100.00	20,447,044	2,702,124
Official size category 2006				
Very small firm	25965	94.09	6,883,700	2,376,450
Small firm	1015	3.68	98,724,916	92,495,151
Medium firm	501	1.82	288,286,357	247,933,548
Large firm	116	0.41	1,214,694,496	1,019,687,885
Total	27597	100.00	20,447,044	2,702,124
Age group 2006				
0-4 years	179	1.59	11,198,196	3,735,344
5-9 years	3081	27.45	27,017,913	3,937,029
10-19 years	4497	40.07	21,286,284	3,995,228
20-39 years	2546	22.69	30,149,324	4,303,716
40+ years	919	8.20	69,764,316	8,313,457
Total	11222	100.00	28,637,392	4,214,296
SIC code 2006				
Agriculture, Forestry and Fishing	161	0.58	28,312,937	4,264,565
Mining	296	1.07	23,650,572	2,782,338
Manufacturing	2800	10.15	42,161,890	7,606,061
Transport, Comms and Electricity	13	0.05	106,989,979	4,158,247
Retail and Wholesale	1168	4.23	18,619,637	3,518,385
Finance	6953	25.19	28,930,193	4,873,953
Services	13419	48.62	13,377,913	1,921,563
Public Admin	2089	7.57	12,798,159	1,789,013
Other	698	2.53	5,904,049	935,423
Total	27597	100.00	20,447,044	2,702,124

Source: own calculations using City of Cape Town (2017)

Notes :

1. Total number of firms for the year section is the number of firms with turnover information for any period.
2. Total number of firms for the rest of the sections is the number of firms with turnover information for 2006.

Table B.7: Turnover descriptive statistics - RSC Cape Town
Balanced panel

Category	Number of firms	% of firms	Mean	Median
Year				
2001	11603	100.00	27,727,407	3,794,563
2002	11603	100.00	25,003,746	3,641,714
2003	11603	100.00	30,232,279	4,614,255
2004	11603	100.00	30,075,976	4,636,120
2005	11603	100.00	34,685,909	5,031,180
2006	11603	100.00	33,568,760	4,743,399
Total	11603	100.00	30,215,679	4,367,429
Percentile size category 2006				
Small	3729	32.14	916,480	845,497
Medium firm	3780	32.58	4,920,846	4,387,215
Large firm	4094	35.28	89,760,619	29,935,311
Total	11603	100.00	33,568,760	4,743,399
Official size category 2006				
Very small firm	10444	90.01	9,259,464	3,731,809
Small firm	697	6.01	100,505,977	95,470,846
Medium firm	368	3.17	291,972,864	246,531,400
Large firm	94	0.81	1,226,529,822	1,020,276,661
Total	11603	100.00	33,568,760	4,743,399
Age group 2006				
0-4 years	11	0.14	8,775,319	5,179,239
5-9 years	2046	25.43	30,787,506	4,520,583
10-19 years	3302	41.03	23,981,303	4,840,888
20-39 years	1968	24.46	35,685,872	5,567,735
40+ years	720	8.95	77,606,394	10,010,637
Total	8047	100.00	33,351,613	5,195,535
SIC code 2006				
Agriculture, Forestry and Fishing	106	0.91	36,148,243	5,356,588
Mining	163	1.40	35,415,218	3,298,286
Manufacturing	1848	15.93	55,968,357	10,395,182
Transport, Comms and Electricity	4	0.03	20,286,759	6,391,589
Retail and Wholesale	561	4.84	28,618,283	5,626,691
Finance	4134	35.63	37,672,372	6,505,996
Services	3431	29.57	23,774,496	3,438,836
Public Admin	1116	9.62	18,940,276	2,125,270
Other	240	2.07	7,846,566	920,696
Total	11603	100.00	33,568,760	4,743,399

Source: own calculations using City of Cape Town (2017)

Notes :

1. Total number of firms for the year section is the number of firms with turnover information for all periods.
2. Total number of firms for the rest of the sections is the number of firms with turnover information for 2006.

B.5 Combined Enterprise Survey

Table B.8: Response pattern for turnover, by firm - Combined Enterprise Survey

Pattern	Number of firms
00000	17
00001	235
00011	602
00100	39
00111	13
01100	29
01111	18
10111	1
11000	1
11011	1
11100	326
11101	7
11111	141
Total	1430

Source: own calculations using DataFirst (2004) and Enterprise Analysis Unit - World Bank Group (2007)

Notes :

1. Each character of the string represents a period of the sample in order. If the firm reported positive turnover for a period then the corresponding character in the string was recorded as 1, with the absence of turnover information being indicated by a 0.

Table B.9: Turnover descriptive statistics - Combined Enterprise Survey
Full panel

Category	Number of firms	% of firms	Mean	Median
Year				
2000	477	33.76	683,383,066	62,814,072
2001	523	37.01	790,495,819	67,619,048
2002	574	40.62	681,019,574	66,289,760
2003	776	54.92	111,566,812	10,309,278
2006	1018	72.05	99,278,293	8,458,647
Total	1413	100.00	391,315,148	25,090,047
Percentile size category 2006				
Small firm	372	36.54	1,530,314	1,409,774
Medium firm	323	31.73	10,630,984	9,398,496
Large firm	323	31.73	300,502,221	93,984,962
Total	1018	100.00	99,278,293	8,458,647
Official size category 2006				
Very small firm	818	80.35	10,545,496	5,075,188
Small firm	88	8.64	104,084,608	96,633,607
Medium firm	79	7.76	302,055,034	261,278,195
Large firm	33	3.24	1,800,524,037	1,127,819,549
Total	1018	100.00	99,278,293	8,458,647
Age group 2006				
0-4 years	248	24.55	13,119,238	2,631,579
5-9 years	269	26.63	66,445,008	4,699,248
10-19 years	227	22.48	53,431,945	11,278,195
20-39 years	180	17.82	104,821,228	31,484,962
40+ years	86	8.51	559,875,251	145,676,692
Total	1010	100.00	99,280,616	8,270,677
SIC code 2006				
Construction	8	0.79	18,819,481	1,766,176
Manufacturing	694	68.17	138,378,719	15,037,594
Transport, Comms and Electricity	6	0.59	2,863,565	2,481,673
Retail and Wholesale	223	21.91	16,120,776	2,255,639
Services	87	8.55	14,572,425	3,383,459
Total	1018	100.00	99,278,293	8,458,647

Source: own calculations using DataFirst (2004) and Enterprise Analysis Unit - World Bank Group (2007)

Notes :

1. Total number of firms for the year section is the number of firms with turnover information for any period.
2. Total number of firms for the rest of the sections is the number of firms with turnover information for 2006.

B.5.1 PICS

Table B.10: Turnover descriptive statistics - PICS
Full panel

Category	Number of firms	% of firms	Mean	Median
Year				
2000	477	82.81	683,383,066	62,814,072
2001	523	90.80	790,495,819	67,619,048
2002	574	99.65	681,019,574	66,289,760
Total	576	100.00	718,111,990	65,570,352
Percentile size category 2002				
Small firm	189	32.93	15,555,721	15,860,566
Medium firm	191	33.28	74,701,761	65,359,476
Large firm	194	33.80	1,926,274,061	470,803,920
Total	574	100.00	681,019,574	66,289,760
Official size category 2002				
Very small firm	279	48.61	24,203,748	21,095,860
Small firm	98	17.07	101,627,246	95,992,372
Medium firm	112	19.51	300,070,920	270,113,288
Large firm	85	14.81	4,006,882,077	1,209,150,336
Total	574	100.00	681,019,574	66,289,760
Age group 2002				
0-4 years	33	12.99	135,567,859	34,625,272
5-9 years	35	13.78	161,954,500	28,520,698
10-19 years	57	22.44	1,156,521,678	31,819,172
20-39 years	74	29.13	656,233,275	88,422,380
40+ years	55	21.65	1,055,516,690	365,440,064
Total	254	100.00	719,206,940	67,359,478
SIC code 2002				
Manufacturing	574	100.00	681,019,574	66,289,760
Total	574	100.00	681,019,574	66,289,760

Source: own calculations using DataFirst (2004)

Notes :

1. Total number of firms for the year section is the number of firms with turnover information for any period.
2. Total number of firms for the rest of the sections is the number of firms with turnover information for 2002.

Table B.11: Turnover descriptive statistics - PICS
Balanced panel

Category	Number of firms	% of firms	Mean	Median
Year				
2000	474	100.00	687,205,497	60,469,850
2001	474	100.00	820,355,419	72,915,476
2002	474	100.00	757,822,985	76,928,104
Total	474	100.00	755,127,967	68,826,192
Percentile size category 2002				
Small firm	147	31.01	17,136,089	17,429,194
Medium firm	161	33.97	79,811,539	68,598,928
Large firm	166	35.02	2,071,321,878	474,727,664
Total	474	100.00	757,822,985	76,928,104
Official size category 2002				
Very small firm	216	45.57	25,681,752	22,028,323
Small firm	85	17.93	101,241,857	93,681,920
Medium firm	99	20.89	304,376,191	281,263,616
Large firm	74	15.61	4,255,703,181	1,255,496,704
Total	474	100.00	757,822,985	76,928,104
Age group 2002				
0-4 years	17	7.83	89,548,636	68,657,952
5-9 years	32	14.75	170,163,674	26,471,678
10-19 years	50	23.04	1,316,183,503	34,752,724
20-39 years	70	32.26	678,068,246	88,913,944
40+ years	48	22.12	1,135,389,426	371,052,272
Total	217	100.00	805,254,420	77,132,896
SIC code 2002				
Manufacturing	474	100.00	757,822,985	76,928,104
Total	474	100.00	757,822,985	76,928,104

Source: own calculations using DataFirst (2004)

Notes :

1. Total number of firms for the year section is the number of firms with turnover information all periods.
2. Total number of firms for the rest of the sections is the number of firms with turnover information for 2002.

B.5.2 2007 Enterprise Survey

Table B.12: Turnover descriptive statistics - 2007 Enterprise Survey
Full panel

Category	Number of firms	% of firms	Mean	Median
Year				
2003	776	76.23	111,566,812	10,309,278
2006	1018	100.00	99,278,293	8,458,647
Total	1018	100.00	104,593,728	9,360,902
Percentile size category 2006				
Small firm	372	36.54	1,530,314	1,409,774
Medium firm	323	31.73	10,630,984	9,398,496
Large firm	323	31.73	300,502,221	93,984,962
Total	1018	100.00	99,278,293	8,458,647
Official size category 2006				
Very small firm	818	80.35	10,545,496	5,075,188
Small firm	88	8.64	104,084,608	96,633,607
Medium firm	79	7.76	302,055,034	261,278,195
Large firm	33	3.24	1,800,524,037	1,127,819,549
Total	1018	100.00	99,278,293	8,458,647
Age group 2006				
0-4 years	248	24.55	13,119,238	2,631,579
5-9 years	269	26.63	66,445,008	4,699,248
10-19 years	227	22.48	53,431,945	11,278,195
20-39 years	180	17.82	104,821,228	31,484,962
40+ years	86	8.51	559,875,251	145,676,692
Total	1010	100.00	99,280,616	8,270,677
SIC code 2006				
Construction	8	0.79	18,819,481	1,766,176
Manufacturing	694	68.17	138,378,719	15,037,594
Transport, Comms and Electricity	6	0.59	2,863,565	2,481,673
Retail and Wholesale	223	21.91	16,120,776	2,255,639
Services	87	8.55	14,572,425	3,383,459
Total	1018	100.00	99,278,293	8,458,647

Source: own calculations using Enterprise Analysis Unit - World Bank Group (2007)

Notes :

1. Total number of firms for the year section is the number of firms with turnover information for any period.
2. Total number of firms for the rest of the sections is the number of firms with turnover information for 2006.

Table B.13: Turnover descriptive statistics - 2007 Enterprise Survey
Balanced panel

Category	Number of firms	% of firms	Mean	Median
Year				
2003	776	100.00	111,566,812	10,309,278
2006	776	100.00	123,135,162	13,157,895
Total	776	100.00	117,350,987	11,959,497
Percentile size category 2006				
Small firm	253	32.60	2,721,747	2,631,579
Medium firm	251	32.35	14,397,360	13,139,098
Large firm	272	35.05	335,479,949	118,421,053
Total	776	100.00	123,135,162	13,157,895
Official size category 2006				
Very small firm	582	75.00	13,167,717	7,518,797
Small firm	84	10.82	104,778,128	98,684,211
Medium firm	78	10.05	302,312,729	257,518,797
Large firm	32	4.12	1,734,609,962	1,127,819,549
Total	776	100.00	123,135,162	13,157,895
Age group 2006				
0-4 years	64	8.32	33,908,210	5,592,105
5-9 years	233	30.30	76,571,846	6,578,947
10-19 years	216	28.09	55,810,993	12,218,045
20-39 years	173	22.50	84,785,555	32,894,737
40+ years	83	10.79	578,639,653	157,894,737
Total	769	100.00	123,226,960	13,157,895
SIC code 2006				
Construction	6	0.77	24,806,234	2,631,579
Manufacturing	590	76.03	154,138,489	18,796,992
Transport, Comms and Electricity	2	0.26	2,162,124	2,162,124
Retail and Wholesale	116	14.95	28,253,221	6,578,947
Services	62	7.99	19,042,607	3,729,323
Total	776	100.00	123,135,162	13,157,895

Source: own calculations using Enterprise Analysis Unit - World Bank Group (2007)

Notes :

1. Total number of firms for the year section is the number of firms with turnover information for all periods.
2. Total number of firms for the rest of the sections is the number of firms with turnover information for 2006.

C Appendix: The lognormality of the firm size distribution

C.1 Summary

Table C.1: Summary of log firm size
Descriptive statistics

Dataset	Mean	Median	SD	Kurtosis	Skewness	`sktest`
RSC KZN	1.64	1.39	0.97	7.69	1.74	Reject
RSC Pretoria	13.23	13.18	1.72	4.45	0.20	Reject
RSC Cape Town	14.81	14.67	1.89	3.94	0.18	Reject
Combined Enterprise Survey	18.17	17.96	1.85	4.75	0.16	Reject
PICS	18.17	17.96	1.85	4.75	0.16	Reject
2007 Enterprise Survey	16.37	16.15	2.04	3.04	0.22	Fail to reject

Source : own calculations using KwaZulu-Natal (2017), Pretoria (2017), City of Cape Town (2017), Enterprise Analysis Unit - World Bank Group (2007) and DataFirst (2004).

Notes :

1. Descriptive statistics are for the first year of the period that is covered by each of the datasets.

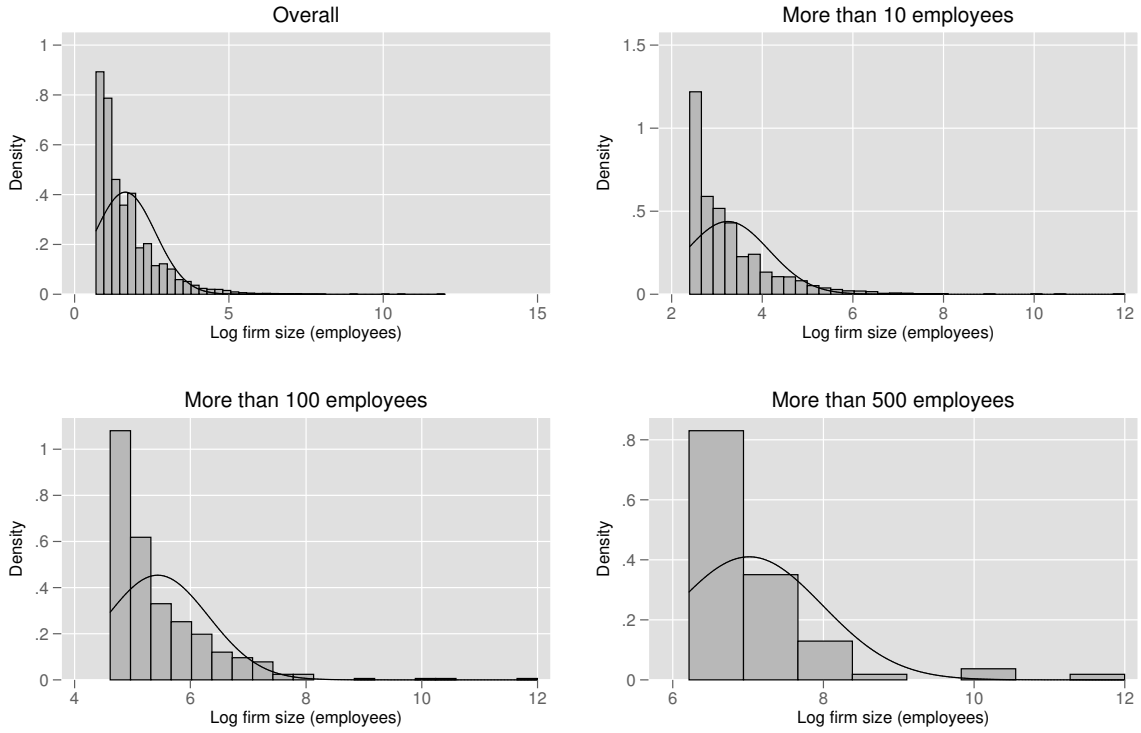
C.2 RSC KZN

Table C.2: Log firm size descriptive statistics - RSC KZN 2004

N	Mean	Median	SD	Kurtosis	Skewness
27777	1.64	1.39	0.97	7.69	1.74

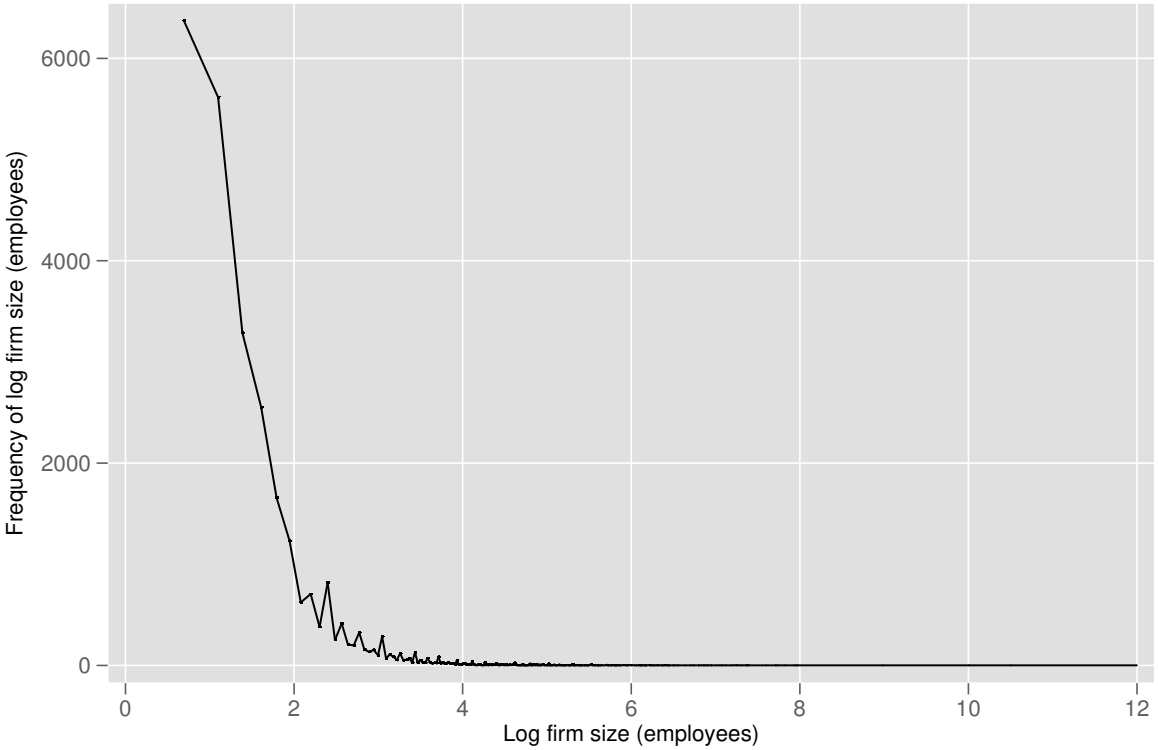
Source: own calculations using KwaZulu-Natal (2017)

Figure C.1: RSC KZN 2004
Histograms



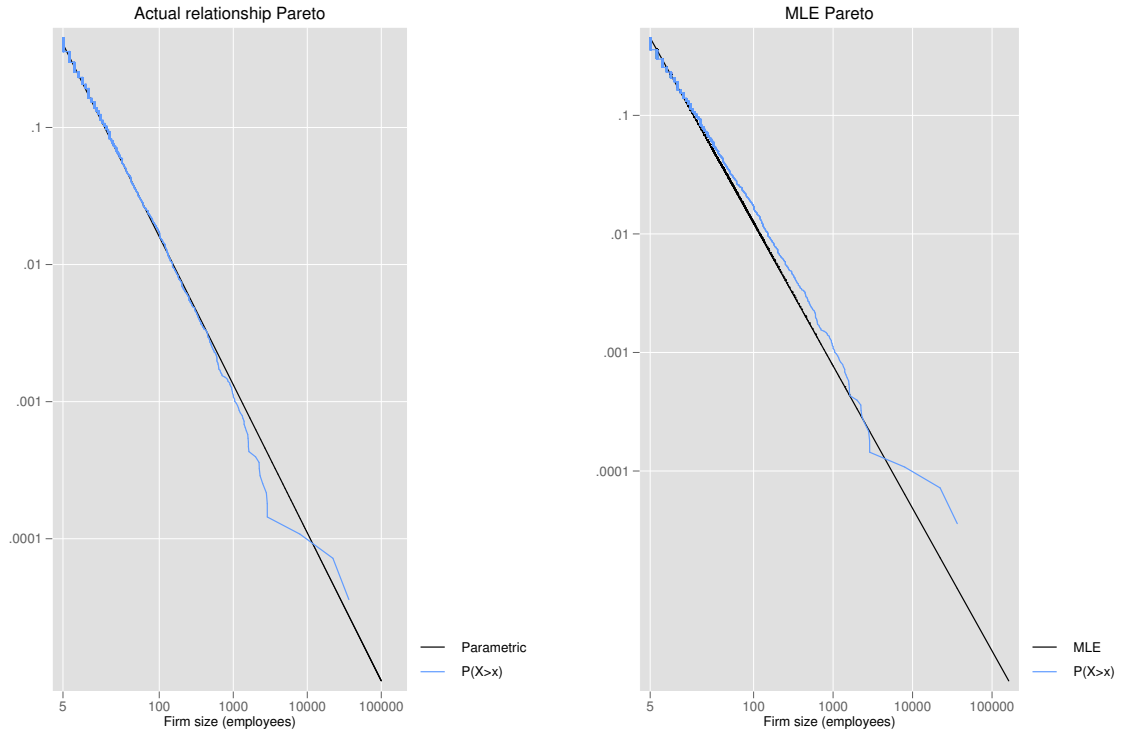
Source: own calculations using KwaZulu-Natal (2017)

Figure C.2: RSC KZN 2004
Frequency plot



Source: own calculations using KwaZulu-Natal (2017)

Figure C.3: RSC KZN 2004
Pareto distribution of firm size in the upper tail



Source: own calculations using KwaZulu-Natal (2017)

C.3 RSC Pretoria

Table C.3: Log firm size descriptive statistics - RSC Pretoria

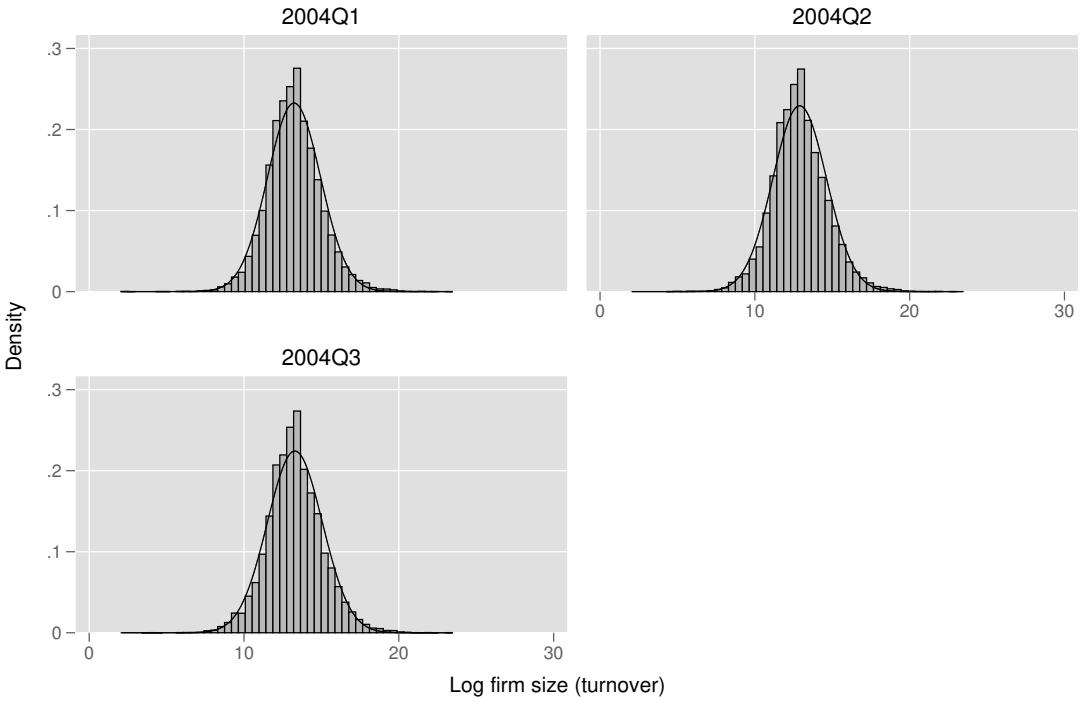
Category	N	Mean	Median	SD	Kurtosis	Skewness
Wave						
2004Q1	25340	13.23	13.18	1.72	4.45	0.20
2004Q2	23515	12.89	12.83	1.74	4.16	0.24
2004Q3	24812	13.28	13.24	1.78	3.97	0.20
Age group 2004Q1						
0-4 years	8965	12.90	13.01	1.81	3.47	0.00
5-9 years	7156	13.38	13.30	1.65	3.75	0.24
10-19 years	8691	13.60	13.44	1.77	4.39	0.46

Source: own calculations using Pretoria (2017)

Notes :

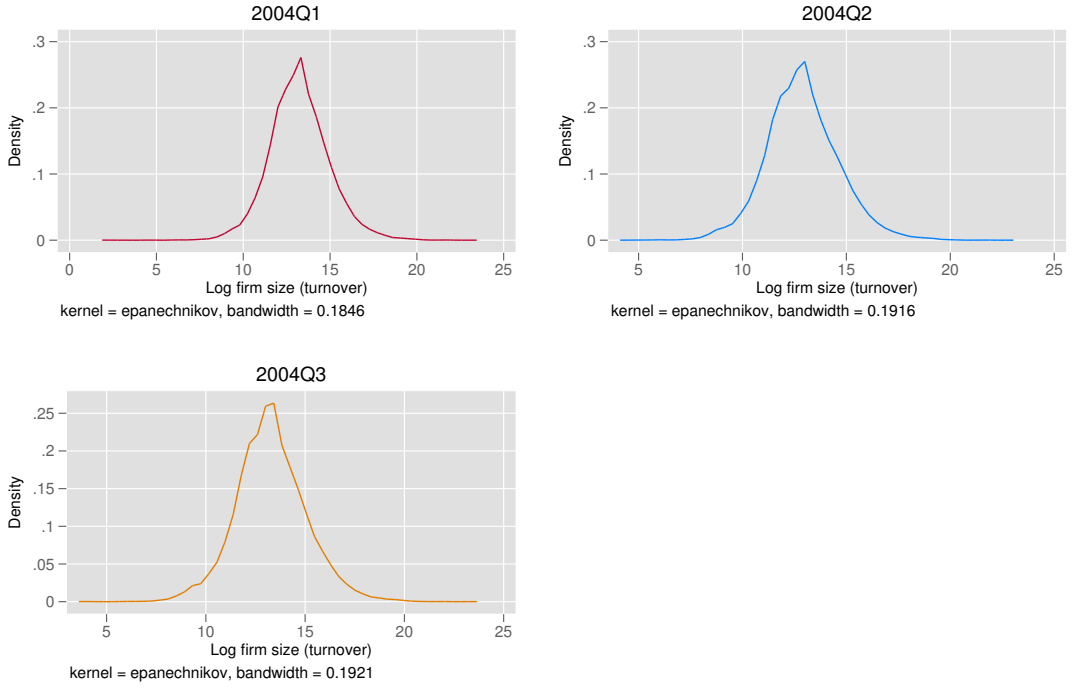
1. Total number of firms is the number of firms with turnover information for any period.

Figure C.4: RSC Pretoria
Histograms



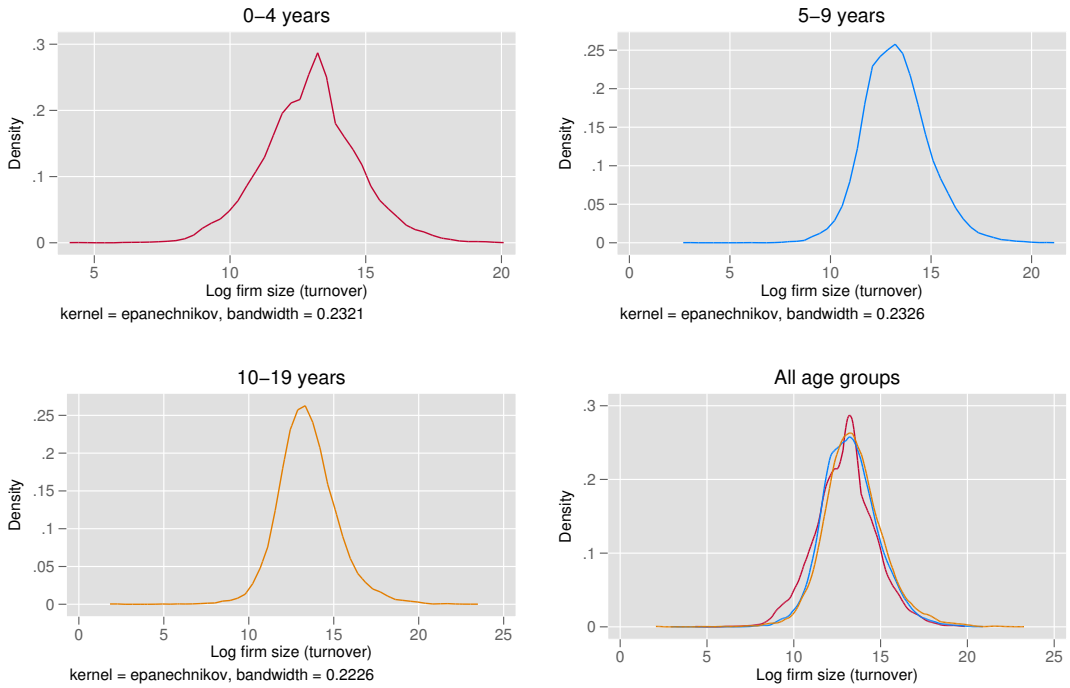
Source: own calculations using Pretoria (2017)

Figure C.5: RSC Pretoria
Densities of log firm size, by year



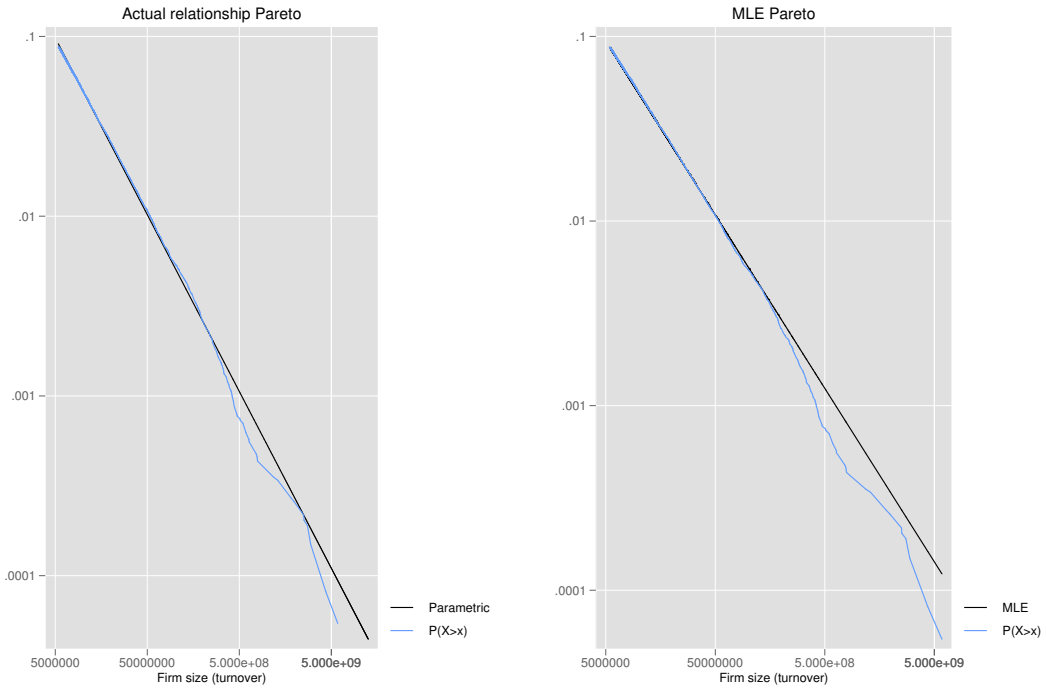
Source: own calculations using Pretoria (2017)

Figure C.6: RSC Pretoria 2004Q1
Densities of log firm size, by age group



Source: own calculations using Pretoria (2017)

Figure C.7: RSC Pretoria 2004Q1
Pareto distribution of firm size in the upper tail



Source: own calculations using Pretoria (2017)

C.4 RSC Cape Town

Table C.4: Log firm size descriptive statistics - RSC Cape Town

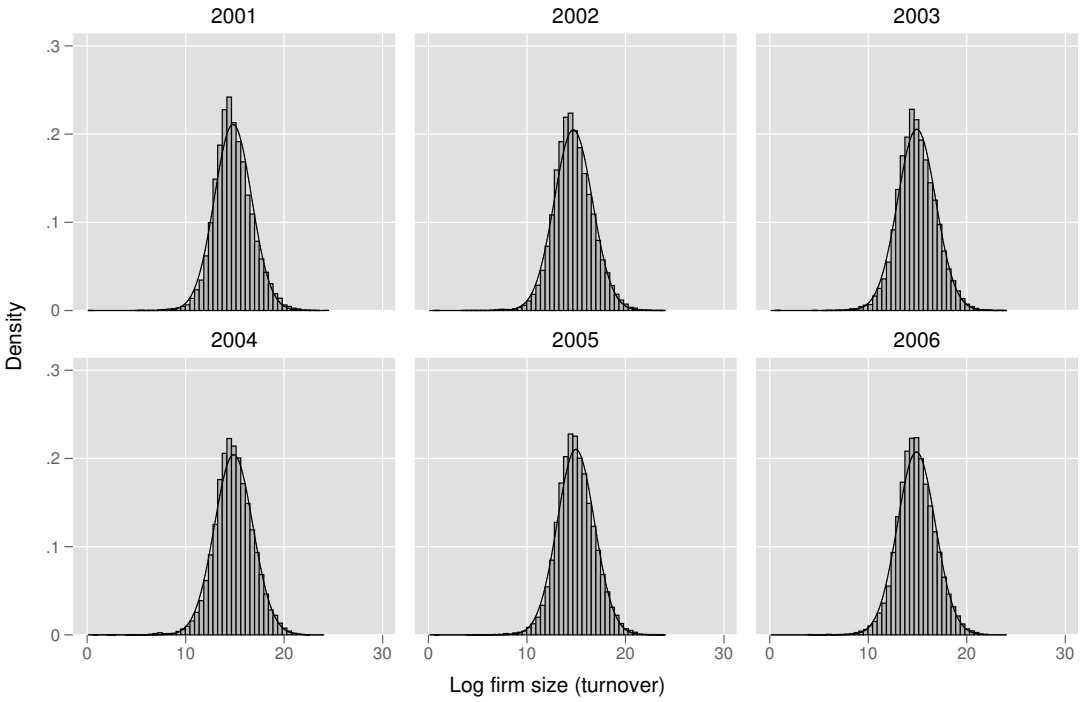
Category	N	Mean	Median	SD	Kurtosis	Skewness
Year						
2001	21418	14.81	14.67	1.89	3.94	0.18
2002	20592	14.71	14.59	1.95	3.86	0.09
2003	23021	14.91	14.82	1.94	3.94	0.00
2004	24026	14.87	14.81	1.95	4.00	-0.11
2005	27894	14.97	14.87	1.90	3.83	0.04
2006	27597	14.88	14.81	1.92	3.82	-0.03
Age group 2001						
0-4 years	3853	14.76	14.65	1.80	3.99	0.15
5-9 years	3413	14.94	14.83	1.79	3.69	0.20
10-19 years	3345	14.98	14.88	1.75	3.69	0.00
20-39 years	2152	15.21	15.05	1.95	3.59	0.04
40+ years	854	15.75	15.70	2.26	3.36	0.08

Source: own calculations using City of Cape Town (2017)

Notes :

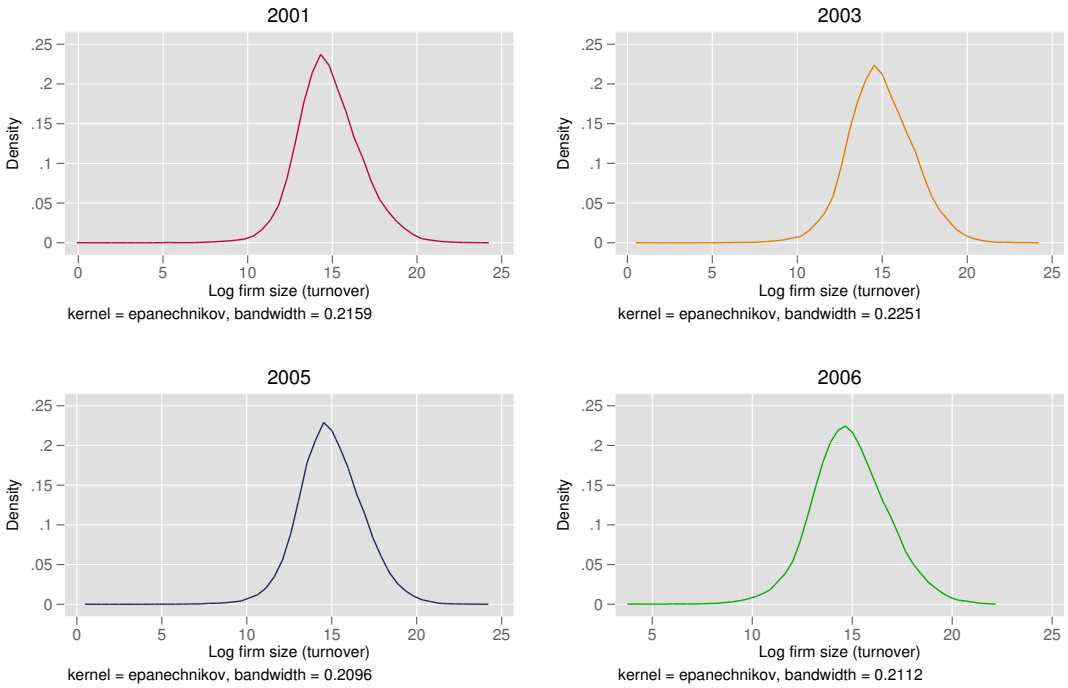
1. Total number of firms is the number of firms with turnover information for any period.

Figure C.8: RSC Cape Town
Histograms



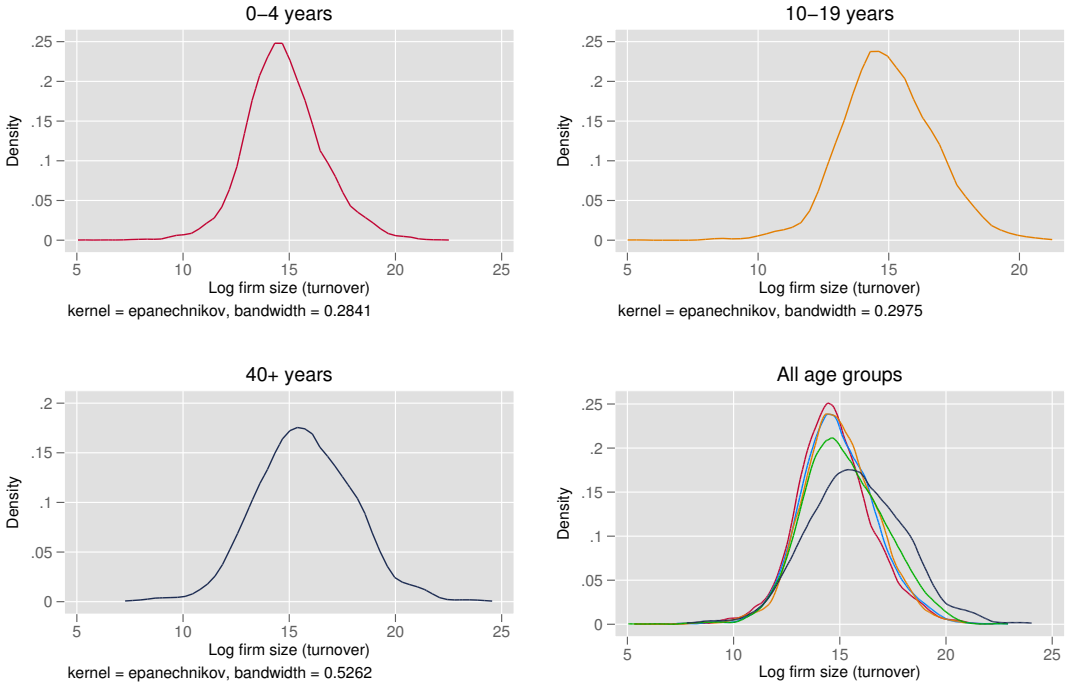
Source: own calculations using City of Cape Town (2017)

Figure C.9: RSC Cape Town
Densities of log firm size, by year



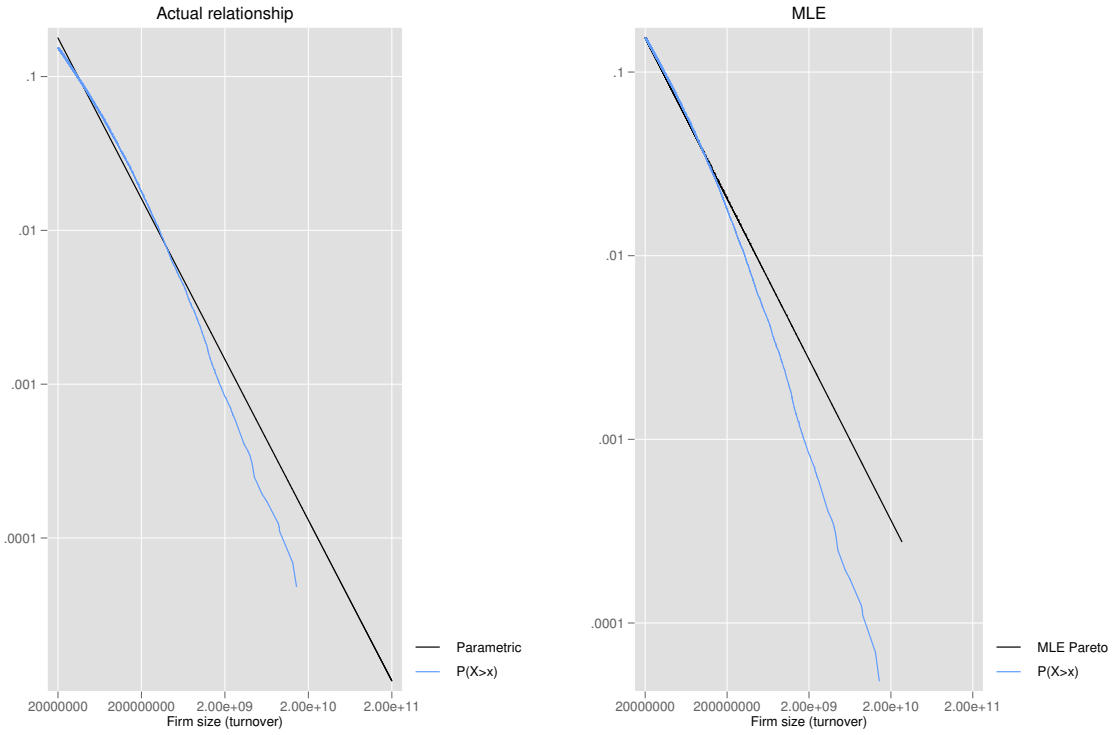
Source: own calculations using City of Cape Town (2017)

Figure C.10: RSC Cape Town 2001
Densities of log firm size, by age group



Source: own calculations using City of Cape Town (2017)

Figure C.11: RSC Cape Town 2001
Pareto distribution of firm size in the upper tail



Source: own calculations using City of Cape Town (2017)

C.5 Combined Enterprise Survey, PICS and 2007 Enterprise Survey

Table C.5: Log firm size descriptive statistics - Combined Enterprise Survey

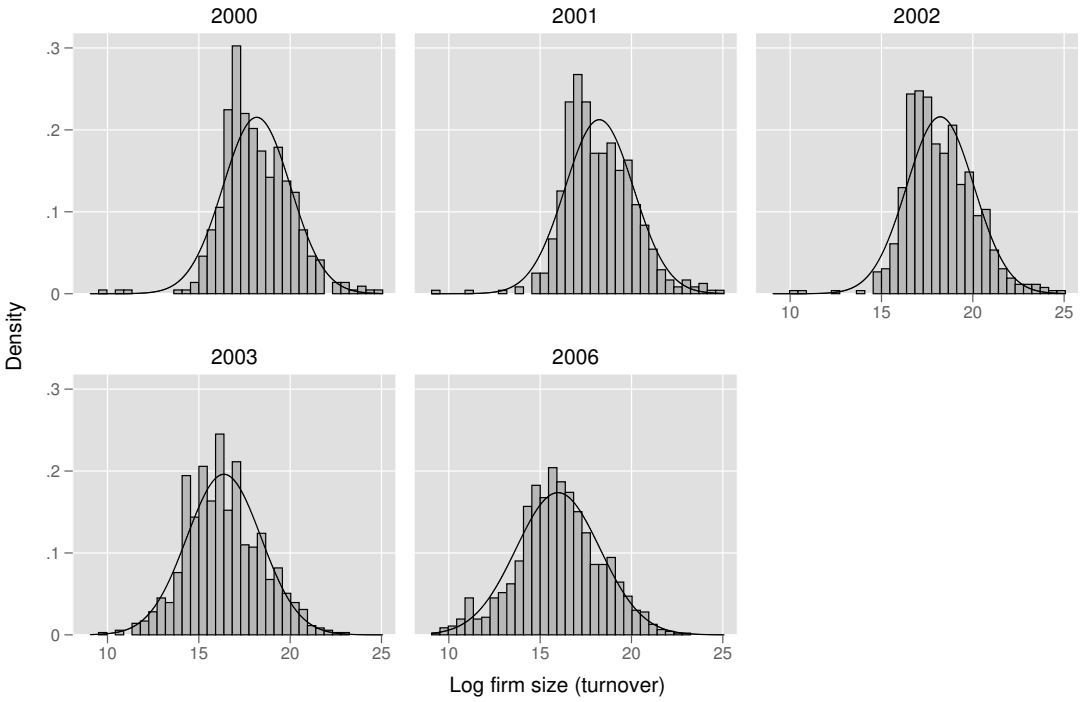
Category	N	Mean	Median	SD	Kurtosis	Skewness
Year						
2000	477	18.17	17.96	1.85	4.75	0.16
2001	523	18.24	18.03	1.88	4.57	0.25
2002	574	18.22	18.01	1.85	4.20	0.28
2003	776	16.37	16.15	2.04	3.04	0.22
2006	1018	15.98	15.95	2.30	3.08	-0.06
Age group 2000						
0-4 years	81	17.68	17.25	1.57	2.44	0.47
5-9 years	61	17.21	17.04	1.22	4.07	0.71
10-19 years	126	17.73	17.46	1.81	7.71	0.25
20-39 years	135	18.10	17.79	1.78	5.46	0.30
40+ years	117	19.58	19.62	1.51	3.40	0.30

Source: own calculations using DataFirst (2004) and Enterprise Analysis Unit - World Bank Group (2007)

Notes :

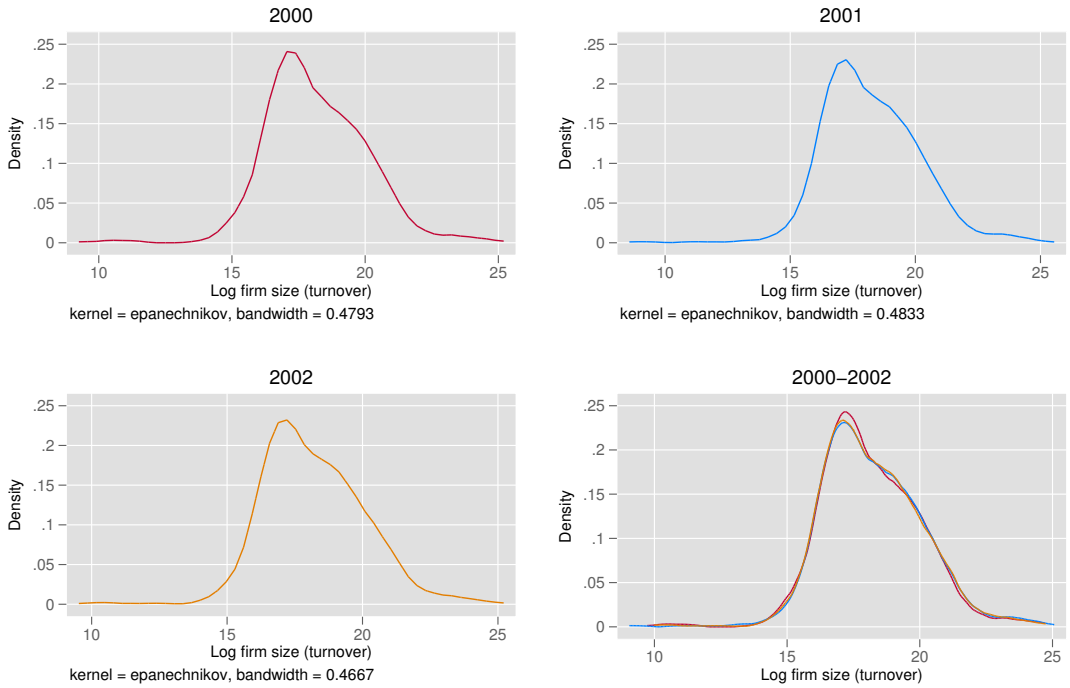
1. Total number of firms is the number of firms with turnover information for any period.

Figure C.12: Combined Enterprise Survey
Histograms



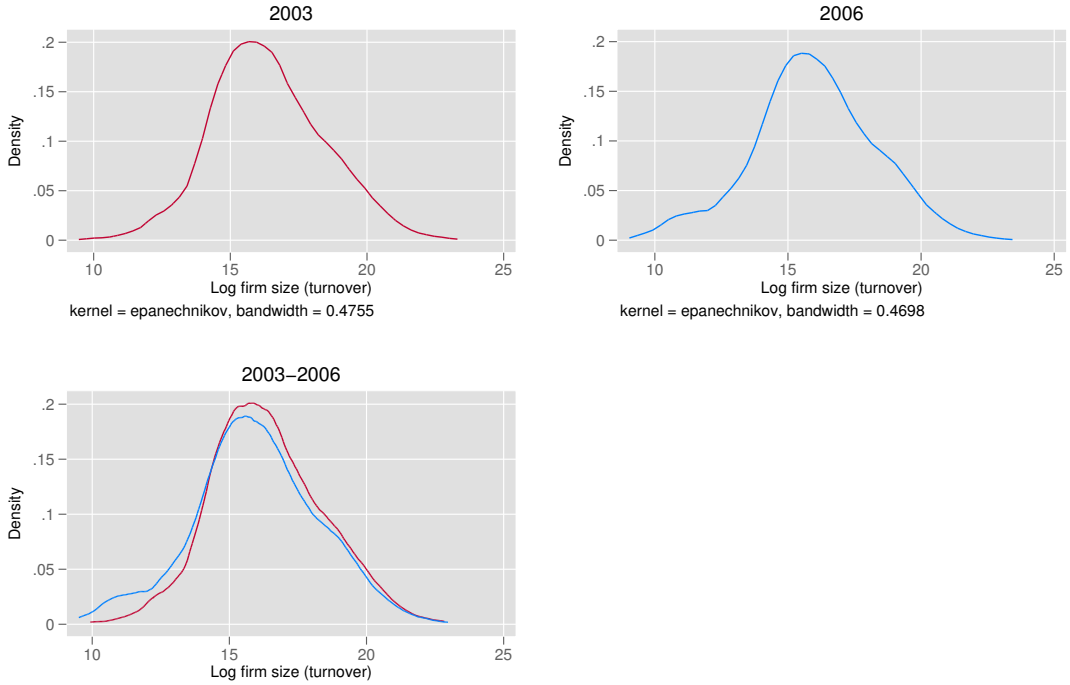
Source: own calculations using DataFirst (2004) and Enterprise Analysis Unit - World Bank Group (2007)

Figure C.13: PICS
Densities of log firm size, by year



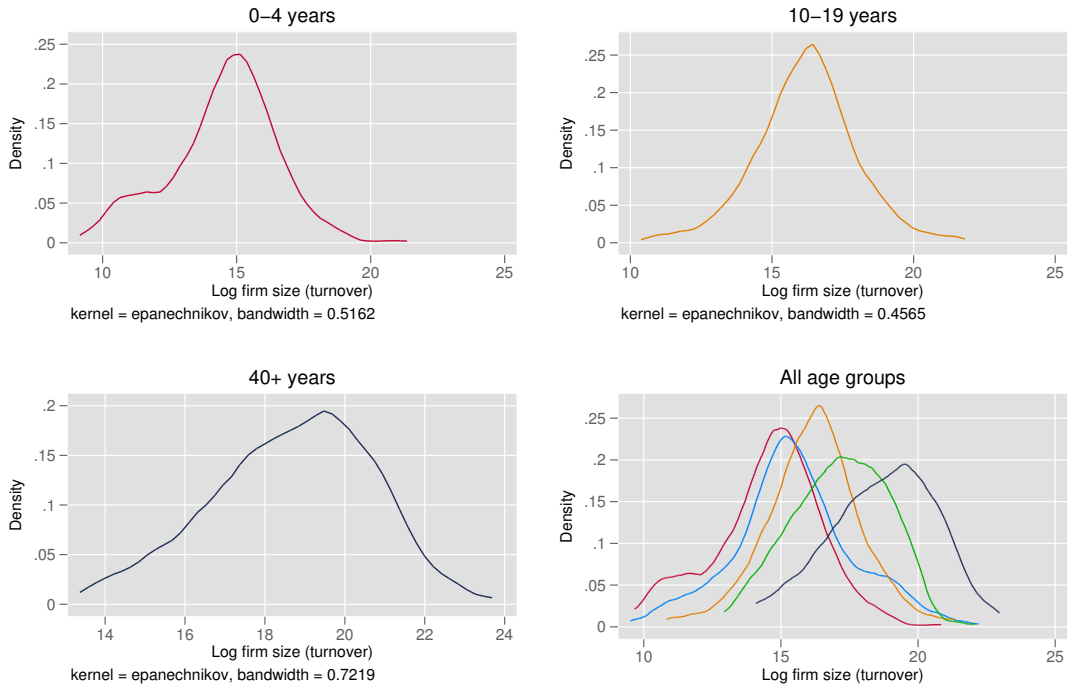
Source: own calculations using DataFirst (2004)

Figure C.14: 2007 Enterprise Survey
Densities of log firm size, by year



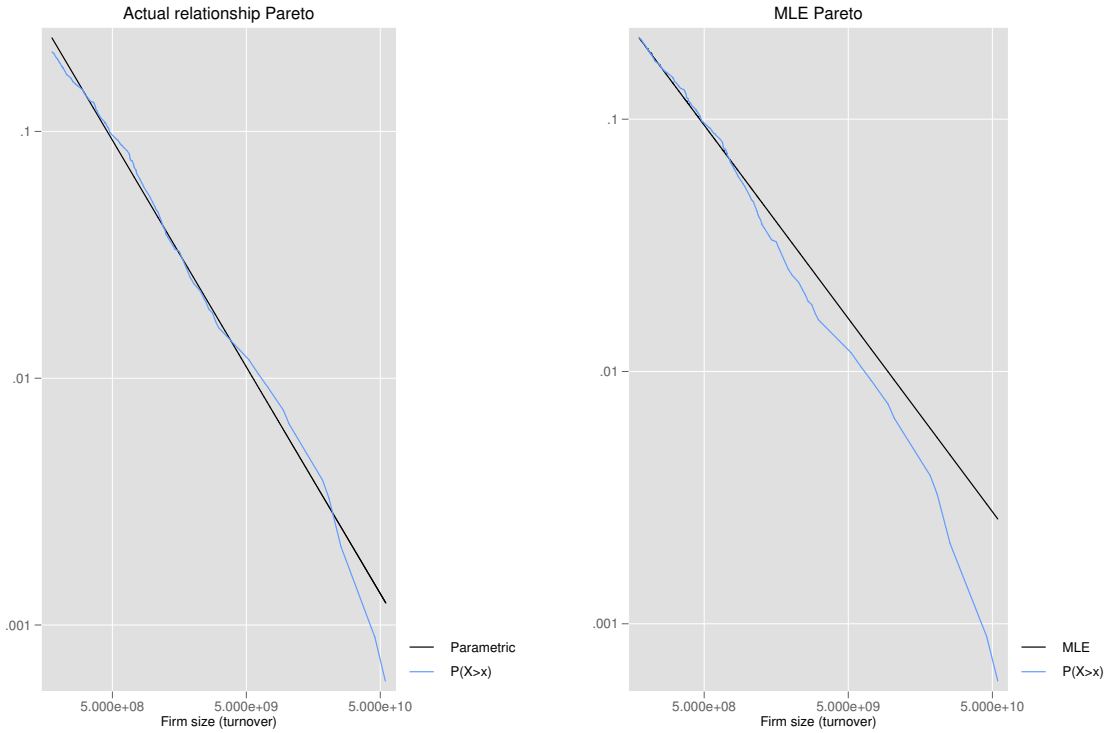
Source: own calculations using Enterprise Analysis Unit - World Bank Group (2007)

Figure C.15: Combined Enterprise Survey 2006
Densities of log firm size, by age group



Source: own calculations using DataFirst (2004) and Enterprise Analysis Unit - World Bank Group (2007)

Figure C.16: Combined Enterprise Survey 2000
Pareto distribution of firm size in the upper tail



Source: own calculations using DataFirst (2004) and Enterprise Analysis Unit - World Bank Group (2007)

D Appendix: Gibrat's Law of Proportionate Effect

Table D.1: Summary OLS estimations

	$s_{it} = \beta_0 + \beta_1 s_{it-\delta} + \nu_{it}$				$s_{it} = \beta_0 + \beta_1 s_{it-\delta} + \beta_2 lnA_{it} + \nu_{it}$				
	N	R^2	β_1	$Wald_{\beta_1=1}$	N	R^2	β_1	β_2	$Wald_{\beta_1=1}$
RSC Cape Town									
2006/2001	11603	0.757	0.924***	157.108***	7635	0.782	0.953***	-0.063***	29.262***
2003/2001	11603	0.845	0.925***	213.594***	7635	0.873	0.957***	-0.026*	38.203***
2006/2004	11603	0.860	0.976***	31.372***	7635	0.866	0.976***	-0.019	15.352***
PICS									
2002/2000	474	0.824	0.893***	4.666*	205	0.819	0.960***	-0.111	0.349
2007 Enterprise Survey									
2006/2003	776	0.943	0.923***	34.305***	769	0.944	0.915***	0.047	28.815***

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Source: own calculations using City of Cape Town (2017), DataFirst (2004) and Enterprise Analysis Unit - World Bank Group (2007)

Table D.2: Summary OLS estimations (continued)

$$g_{it} = \alpha_0 + \alpha_1 s_{it-\delta} + \alpha_2 lnA_{it} + \alpha_3 (s_{it-\delta})(lnA_{it}) + v_{it}$$

	N	R^2	α_1	α_2	α_3
RSC Cape Town					
2006/2001	7635	0.015	-0.028***	-0.124***	0.007***
2003/2001	7635	0.020	-0.040***	-0.165***	0.010***
2006/2004	7635	0.005	-0.019	-0.064	0.004
PICS					
2002/2000	205	0.664	-0.216***	-1.307***	0.070***
2007 Enterprise Survey					
2003/2006	769	0.414	-0.100***	-0.534***	0.032***

$^{*}\ p < 0.05$, $^{**}\ p < 0.01$, $^{***}\ p < 0.001$

Source: own calculations using City of Cape Town (2017), DataFirst (2004) and Enterprise Analysis Unit - World Bank Group (2007)

Table D.3: Summary OLS estimations (continued)

$$g_{it} = \alpha_0 + \alpha_1 s_{it-\delta} + \alpha_2 lnA_{it} + \alpha_3 (s_{it-\delta})(lnA_{it}) + \alpha_4 s^2_{it-\delta} + \alpha_5 lnA^2_{it} + \upsilon_{it}$$

	N	R^2	α_1	α_2	α_3	α_4	α_5
RSC Cape Town							
2006/2001	7635	0.026	-0.128***	-0.071*	0.005**	0.003***	-0.004
2003/2001	7635	0.030	-0.178**	-0.101**	0.009***	0.005**	-0.008*
2006/2004	7635	0.019	-0.193***	-0.008	0.001	0.006***	-0.003
PICS							
2002/2000	205	0.666	-0.149	-1.356***	0.071***	-0.002	0.004
2007 Enterprise Survey							
2003/2006	769	0.483	-0.046	-0.472***	0.042***	-0.002	-0.045***

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Source: own calculations using City of Cape Town (2017), DataFirst (2004) and Enterprise Analysis Unit - World Bank Group (2007)

D.1 RSC Cape Town

Table D.4: Growth rate descriptive statistics - RSC Cape Town 2006/2001

Category	N	Mean	Standard deviation
Size category			
Small firm	4151	0.33	1.09
Medium firm	3945	0.13	0.88
Large firm	3507	0.09	0.80
Total	11603	0.19	0.93
Age group			
0-4 years	1907	0.23	0.99
5-9 years	1971	0.18	0.85
10-19 years	2091	0.11	0.84
20-39 years	1402	0.06	0.83
40+ years	562	0.02	0.84
Total	7933	0.14	0.88

Source: own calculations using City of Cape Town (2017)

Notes :

1. The geometric mean proportionate growth rate is used.
2. The proportionate growth rates are for the six year period 2001-2006.

Table D.5: Growth rate descriptive statistics - RSC Cape Town 2003/2001

Category	N	Mean	Standard deviation
Size category			
Small firm	4151	0.33	0.89
Medium firm	3945	0.14	0.61
Large firm	3507	0.08	0.56
Total	11603	0.19	0.70
Age group			
0-4 years	1907	0.20	0.76
5-9 years	1971	0.15	0.63
10-19 years	2091	0.13	0.60
20-39 years	1402	0.07	0.59
40+ years	562	0.07	0.55
Total	7635	0.14	0.67

Source: own calculations using City of Cape Town (2017)

Notes :

1. The geometric mean proportionate growth rate is used.
2. The proportionate growth rates are for the three year period 2001-2003.

Table D.6: Growth rate descriptive statistics - RSC Cape Town 2006/2004

Category	N	Mean	Standard deviation
Size category			
Small firm	3720	0.06	0.82
Medium firm	3902	-0.02	0.69
Large firm	3981	-0.01	0.60
Total	11603	0.01	0.70
Age group			
0-4 years	457	-0.01	0.80
5-9 years	2605	0.04	0.70
10-19 years	2642	0.01	0.67
20-39 years	1698	0.02	0.64
40+ years	645	-0.03	0.70
Total	8047	0.02	0.68

Source: own calculations using City of Cape Town (2017)

Notes :

1. The geometric mean proportionate growth rate is used.
2. The proportionate growth rates are for the three year period 2004-2006.

Table D.7: OLS regressions - RSC Cape Town

	$s_{it} = \beta_0 + \beta_1 s_{it-\delta} + \nu_{it}$				$s_{it} = \beta_0 + \beta_1 s_{it-\delta} + \beta_2 lnA_{it} + \nu_{it}$				
	N	R^2	β_1	$Wald_{\beta_1=1}$	N	R^2	β_1	β_2	$Wald_{\beta_1=1}$
2006/2001									
Overall	11603	0.757	0.924***	157.108***	7635	0.782	0.953***	-0.063***	29.262***
Small	3729	0.215	0.419***	838.411***	2300	0.251	0.462***	-0.058	288.030***
Medium	3780	0.188	0.220***	6470.035***	2526	0.235	0.266***	-0.008	3467.725***
Large	4094	0.644	0.643***	822.816***	2809	0.667	0.683***	-0.059***	302.883***
2003/2001									
Overall	11603	0.845	0.925***	213.594***	7635	0.873	0.957***	-0.026*	38.203***
Small	3709	0.408	0.567***	352.223***	2271	0.491	0.644***	-0.040	107.097***
Medium	3902	0.264	0.298***	3052.842***	2610	0.331	0.375***	0.000	1175.041***
Large	3992	0.725	0.701***	517.589***	2754	0.749	0.736***	-0.014	190.101***
2006/2004									
Overall	11603	0.860	0.976***	31.372***	7635	0.866	0.976***	-0.019	15.352***
Small	3729	0.356	0.556***	533.625***	2300	0.393	0.586***	-0.033	263.036***
Medium	3780	0.352	0.374***	2731.898***	2526	0.373	0.388***	-0.002	1625.740***
Large	4094	0.796	0.813***	343.994***	2809	0.800	0.817***	-0.027*	198.155***

$^{*}\ p < 0.05$, $^{**}\ p < 0.01$, $^{***}\ p < 0.001$

Source: own calculations using City of Cape Town (2017)

Table D.8: OLS regressions - RSC Cape Town (continued)

$$g_{it} = \alpha_0 + \alpha_1 s_{it-\delta} + \alpha_2 lnA_{it} + \alpha_3 (s_{it-\delta})(lnA_{it}) + v_{it}$$

	N	R^2	α_1	α_2	α_3
2006/2001					
Overall	7635	0.015	-0.028***	-0.124***	0.007***
Small	2300	0.317	-0.134***	-0.236	0.017
Medium	2526	0.702	-0.118***	-0.022	0.002
Large	2809	0.338	-0.109***	-0.354***	0.020***
2003/2001					
Overall	7635	0.020	-0.040***	-0.165***	0.010***
Small	2271	0.261	-0.253***	-0.773***	0.056***
Medium	2610	0.578	-0.195***	0.079	-0.005
Large	2754	0.302	-0.158***	-0.470***	0.027***
2006/2004					
Overall	7635	0.005	-0.019	-0.064	0.004
Small	2300	0.244	-0.159**	-0.114	0.008
Medium	2526	0.597	-0.226***	-0.125	0.008
Large	2809	0.179	-0.108***	-0.292***	0.016***

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Source: own calculations using City of Cape Town (2017)

D.2 PICS

Table D.9: Growth rate descriptive statistics - PICS 2002/2000

Year	N	Mean	Standard deviation
Size category			
Small firm	147	0.30	1.20
Medium firm	161	0.19	0.36
Large firm	166	0.13	0.42
Total	474	0.21	0.79
Age group			
0-4 years	21	0.22	0.63
5-9 years	32	0.01	1.21
10-19 years	36	0.39	1.05
20-39 years	60	-0.01	1.09
40+ years	48	0.13	0.29
Total	210	0.15	0.90

Source: own calculations using DataFirst (2004)

Notes :

1. The geometric mean proportionate growth rate is used.
2. The proportionate growth rates are for the three year period 2000-2002.

Table D.10: OLS regressions - PICS

	$s_{it} = \beta_0 + \beta_1 s_{it-\delta} + \nu_{it}$				$s_{it} = \beta_0 + \beta_1 s_{it-\delta} + \beta_2 lnA_{it} + \nu_{it}$				
	N	R^2	β_1	$Wald_{\beta_1=1}$	N	R^2	β_1	β_2	$Wald_{\beta_1=1}$
2002/2000									
Overall	474	0.824	0.893***	4.666*	205	0.819	0.960***	-0.111	0.349
Small	147	0.115	0.320*	17.971***	68	0.030	0.188	-0.116	19.750***
Medium	161	0.275	0.302*	35.707***	73	0.541	0.605***	0.045	8.558**
Large	166	0.871	0.859***	14.229***	64	0.860	0.850***	-0.203*	6.332*

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$
Source: own calculations using DataFirst (2004)

Table D.11: OLS regressions - PICS (continued)

$g_{it} = \alpha_0 + \alpha_1 s_{it-\delta} + \alpha_2 lnA_{it} + \alpha_3 (s_{it-\delta})(lnA_{it}) + \upsilon_{it}$

	N	R^2	α_1	α_2	α_3
2002/2000					
Overall	205	0.664	-0.216***	-1.307***	0.700***
Small	68	0.974	-0.322***	-1.591***	0.096***
Medium	73	0.883	-0.312***	-1.791***	0.099***
Large	64	0.638	-0.240***	-1.364***	0.065***

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Source: own calculations using DataFirst (2004)

D.3 2007 Enterprise Survey

Table D.12: Growth rate descriptive statistics - 2007 Enterprise Survey 2006/2003

Year	N	Mean	Standard deviation
Size category			
Small firm	253	0.45	0.45
Medium firm	251	0.28	0.34
Large firm	272	0.14	0.61
Total	776	0.30	0.49
Age group			
0-4 years	64	0.39	0.47
5-9 years	233	0.23	0.62
10-19 years	216	0.30	0.47
20-39 years	173	0.23	0.39
40+ years	83	0.23	0.37
Total	769	0.30	0.49

Source: own calculations using Enterprise Analysis Unit - World Bank Group (2007)

Notes :

1. The geometric mean proportionate growth rate is used.
2. The proportionate growth rates are for the three year period 2003-2006.

Table D.13: OLS regressions - 2007 Enterprise Survey

	$s_{it} = \beta_0 + \beta_1 s_{it-\delta} + \nu_{it}$				$s_{it} = \beta_0 + \beta_1 s_{it-\delta} + \beta_2 lnA_{it} + \nu_{it}$				
	N	R^2	β_1	$Wald_{\beta_1=1}$	N	R^2	β_1	β_2	$Wald_{\beta_1=1}$
2006/2003									
Overall	776	0.943	0.923***	34.305***	769	0.944	0.915***	0.047	28.815***
Small	253	0.634	0.587***	18.427***	253	0.634	0.587***	-0.018	18.368***
Medium	251	0.738	0.735***	27.939***	249	0.745	0.727***	0.053	27.767***
Large	272	0.867	0.850***	21.652***	267	0.870	0.856***	-0.012	17.448***

$^{*}\ p < 0.05$, $^{**}\ p < 0.01$, $^{***}\ p < 0.001$

Source: own calculations using Enterprise Analysis Unit - World Bank Group (2007)

Table D.14: OLS regressions - 2007 Enterprise Survey (continued)

$$g_{it} = \alpha_0 + \alpha_1 s_{it-\delta} + \alpha_2 lnA_{it} + \alpha_3 (s_{it-\delta})(lnA_{it}) + \upsilon_{it}$$

	N	R^2	α_1	α_2	α_3
2006/2003					
Overall	769	0.414	-0.100***	-0.534***	0.032***
Small	253	0.896	-0.226***	-1.397***	0.095***
Medium	249	0.852	-0.222***	-1.377***	0.082***
Large	267	0.681	-0.167***	-0.907***	0.048***

$^{*}\ p < 0.05$, $^{**}\ p < 0.01$, $^{***}\ p < 0.001$

Source: own calculations using Enterprise Analysis Unit - World Bank Group (2007)