Power-Law Distributions in Economics: Empirical Evidence for an Emerging Economy, Kazakhstan

by

Chingis Matayev

A thesis submitted in partial fulfillment of the requirements for the degree of

Master of Arts
Economics

at

NAZARBAYEV UNIVERSITY

SCHOOL OF HUMANITIES AND SOCIAL SCIENCE

2018

Acknowledgements

I want to thank Dr. Jozef Konings for his guidance throughout the project. Without his valuable assistance, this work would not have been completed. I am also indebted for Dr. Dana Bazarkulova for the support and valuable comments. I also want to thank Economics Department of Nazarbayev University in the face of every faculty member.

Abstract

Heterogeneity between workers, firms or regions is an aspect in macroeconomics that has been largely overlooked. Typically, a representative agent approach is followed, like the respresentative firm or worker in an economy. However, this approach potentially misses a lot of what may be important drivers and transmission mechanisms in the economy at large. In other words, it misses the underlying granularity of the economy. Rather than taking the average firm or worker or region, it is therefore important to study the entire distribution. This allows a better assessment of how firm-specific idiosyncratic shocks may affect the overall economy. This is particurlarly relevant given that recent work has demonstrated that many economic variables, such as firm size and city size, follow a power law (with respect to the number of employees). This paper first documents power laws in economics and then estimates power laws for firms and cities in Kazakhstan.

Contents

1 Introduction

Scholars have assimilated a number of things by studying statistical distributions of empirical datasets, such as populations of bacteria, lifetimes of a physical particles, stock market prices, the worldwide seismicity activity rate, and so on. Many, if not the most, of such observations are distributed around some value, which turns out to be an average value, or the mean. Alternatively speaking, such distributions put the lesser amount of probability the further away from the mean, and therefore the mean is a representative for majority of observations. For example, an information that the most female Kazakhs are about 160 cm tall would be a valuable statement, since large deviations from this norm are very exceptional.

However, there are many other distributions that can not be simply described by the mean and standard deviation. While such distributions are considered to be puzzling, many scientists find them some of the most challenging and interesting observations. Exactly because these distributions are not easily defined by their mean and variance makes scholars conduct more studies of the topic.

The *power-law*, which belongs to the group of such distributions, with its mathematical properties, is of specific interest for the scientific community. The power law distribution tends to appear in the remarkable number of regularities in both natural and human-influenced phenomena in Finance and Economics. Some examples include U.S. firms size (employment level) and market capitalization (Axtell 2001, Gabaix & Landier 2008), distribution of wealth and income (Atkinson & Piketty 2007), etc. For instance, according to U.S. Census Bureau data for 1997, the average U.S. firm size was 19.0 employees. However, this fact would not tell us much about the distribution of the firm sizes since the significant number of employees work in companies whose sizes differ from the mean by several orders of magnitude.

One importance of this paper comes from the fact that there is a lot of heterogeneity in underlying macroeconomic fluctuations. Even within narrowly defined sectors: there is a difference between firm sizes, a difference within a city sizes, etc. This heterogeneity is an aspect in macroeconomics that has been largely overlooked. Mostly, the scientists concentrate on the representative structure that has typical parameters. However, while looking on the averages, i.e. on "representative" of the data, we miss a lot of what potentially could be going in the economy, the underlying granularity of the economy. That is why it is interesting to start thinking about the implications of this heterogeneity, and look not only on average but also on the entire distribution of firms to see then how shocks may affect individual firms.

Acemoglu (2010) looks at propagation of shocks through networks in the economy. The main result of his paper is that one shock of some large enough firm will affect other firms and cause the "cascades" of effects. This result is in line with the search for the power-law behavior in many of the macro(-micro)economic structures. The reason behind that is the so-called "fat tail" of the power-law distribution where a significant part of the whole data lies. Thus, if the power-law behavior is being observed, then in the case of some economical instability the governing authority would like to concentrate more on the firm in the fat tail of the distribution.

Literature on the the power-law behavior with its applications in Economics and Finance is abundant. Axtell (2001) found the evidence of the Zipf's law (a special case of the power law with exponent $\alpha \approx 1$, to be discussed in section 3) for US firms. City sizes in most of the countries also turn out to exhibit the power-law behavior (Gabaix 2009). However, to the best of my knowledge, there was no previous research on power-law behavior in countries with centralized governance. It may be due to the fact that government has a strong influence on the much of the macroeconomic structures like firms. This paper will identify the power-law behavior in Kazakhstani cities and firms with respect to their sizes (population for the cities and number of employees for firms).

Economic papers on trying to detect a power law use a simple Ordinary Least Squares (OLS) regression. OLS is the method of measuring unknown parameter. However, it requires specific assumptions of the data to hold. The objective of this paper is to use more advanced tools (which will result in more general outcomes) like Kolmogorov-Smirnov test statistic in searching for the power-law behavior in Kazakhstan.

The paper is organized as follows: section two will provide a brief introduction to the power-law theory. The next section will present the estimate of the power-law parameters. Afterwards, section four will detect the existence or absence of the power law behavior in Kazakhstani distributions of city and firm sizes. I find that the city sizes do follow power-law statistical regularity, while only the upper-tail of the fim-size distribution does so. Finally, section five will test for the fraud detection in our obtained data with concluding remarks followed.

2 Power-law properties

The variable of interest X follows a power-law if it is drawn from a distribution with the following probability function

$$p(x) = Cx^{-\alpha},$$

where $\alpha \in \mathbb{Q}$ is a power-law exponent, and $C \in \mathbb{Q}$ is some unostentatious constant.

Power law, like many other distributions, might come in one of two types: continuous distributions governing continuous real numbers, and discrete distributions where the variable of interest takes on only natural values.

Let's assume that we have some empirical dataset $x_i = \{x_1, x_2, \ldots, x_n\}$. Then, a continuous power-law distribution, $p(x)$ has the following form:

$$p(x)\,dx = \Pr(x \leq X \leq x + dx) = Cx^{-\alpha}\,dx, \tag{1}$$

where $X \in x_i$, and C is a normalization constant (which reduces the function to pdf with total probability being equal to one). We could clearly see that this probability function diverges as $x \to 0$, which in turn means that (1) does not work $\forall\, x_i \geq 0$. Thus, there has to be some threshold value $\mathrm{x_{min}}$, starting from which the powel-law starts to behave. Then, given that $\alpha > 1$, and calculating the normalizing constant, we find that

$$p(x) = \frac{\alpha - 1}{x_{\min}} \left(\frac{x}{x_{\min}} \right)^{-\alpha} \tag{2}$$

Coming to a discrete case, the observed values can take only integer values, therefore, probability distribution function will be of the following form:

$$p(x) = Pr(X = x) = Cx^{-\alpha}. \tag{3}$$

As in continuous case, this function diverges at zero. Thus, there has to be some lower bound $x_{\min}$ for the power-law bahavior. Calculating the normalizing constant, we obtain:

$$p(x) = \frac{x^{-\alpha}}{\zeta(\alpha, x_{\min})}, \tag{4}$$

where

$$\zeta(\alpha, x_{\min} = \sum_{n=0}^{\infty}(n + x_{\min})^{-\alpha} \tag{5}$$

is Hurwitz zeta function (for more details see Clauset 2011).

3 Procedure

As we have seen from the previous part, the pdf of power-law distribution depends on two parameters, the power-law exponent α, and the threshold value $x_{\min}$, from which the power-law starts to behave. Thus, we need to find these two parameters in order to see whether the data fits power-law or not.

3.1 Estimating the power-law exponent α

It is relatively easier to find the value of α rather than that of $x_{\min}$. Let's for a second assume that we know $x_{\min}$ (which we will in fact estimate later in this paper). Then for the continuous power-law case, the way we derive the power-law exponent is by using the maximum likelihood estimation (MLE). We begin by stating the pdf of the power-law first:

$$p(x) = \frac{\alpha - 1}{x_{\min}} \left(\frac{x}{x_{\min}}\right)^{-\alpha}.$$

The log-likelihood function is then

$$\mathcal{L} = \ln\Big[\prod_{i=1}^{n} \frac{\alpha - 1}{x_{\min}} \left(\frac{x}{x_{\min}}\right)^{-\alpha}\Big] \tag{6}$$

Simplifying (6),we get

$$\begin{aligned} \ln \mathcal{L} &= \ln\Big[\prod_{i=1}^{n} \frac{\alpha - 1}{x_{\min}} \left(\frac{x}{x_{\min}}\right)^{-\alpha}\Big] \\ &= n \ln\left(\frac{\alpha - 1}{x_{\min}}\right) - \alpha \sum_{i=1}^{n} \ln\left(\frac{x_i}{x_{\min}}\right) \\ &= n\ln(\alpha - 1) - n\ln(x_{\min}) - \alpha \sum_{i=1}^{n} \ln\left(\frac{x_i}{x_{\min}}\right) \end{aligned} \tag{7}$$

Now we set partial derivative of (7) with respect to parameter α equal to zero, and solve for α:

$$\begin{aligned} \frac{\partial \mathcal{L}}{\partial \alpha} &= \frac{n}{\alpha - 1} - \sum_{i=1}^{n} \ln\left(\frac{x_i}{x_{\min}}\right) = 0 \\ \alpha &= 1 + n\Big[\sum_{i=1}^{n} \ln\frac{x_i}{x_{\min}}\Big]^{-1}. \end{aligned}$$

Standard error of the MLE is $\hat{\sigma} = (\hat{\alpha} - 1)/\sqrt{n}$. Since α depends on $x_{\min}$, so does the standard error (Clauset, 2007).

3.2 Estimating $x_{\min}$

Using MLE for estimating $x_{\min}$ will not bring us the desired result, because as we increase the value of $x_{\min}$, we decrease the number of observations in our sample. As we could see from (6), the likelihood function gets smaller as number of observations gets larger, that is why the likelihood will never decrease when we truncate our sample. As a result, the MLE approach will converge to the value of $x_{\min} = \max_i x_i$, and will make the size of our sample n equal to one.

3.2.1 Hill Plot

One of the methods of estimating $x_{\min}$ is the Hill plot, which was often used by scientists in the past (Clauset, 2007). The idea behind this method is to construct a graph of the function $\alpha(x_{\min})$ against the domain $x_{\min} \in \{x_i\}$. The value of x_i, where the region of $\alpha(x_{\min})$ function starts to be more or less flat, would become our $x_{\min}$. Thus, in the past scholars were visually estimating the value of $x_{\min}$ to observe a power-law. The downside of this method is that the estimation of the threshold value occured to be quite subjective. We will refer to Clauset (2011) who provided a Hill plot for $n = 500$ observations, with $\alpha = 2.5$ (see figure 1).

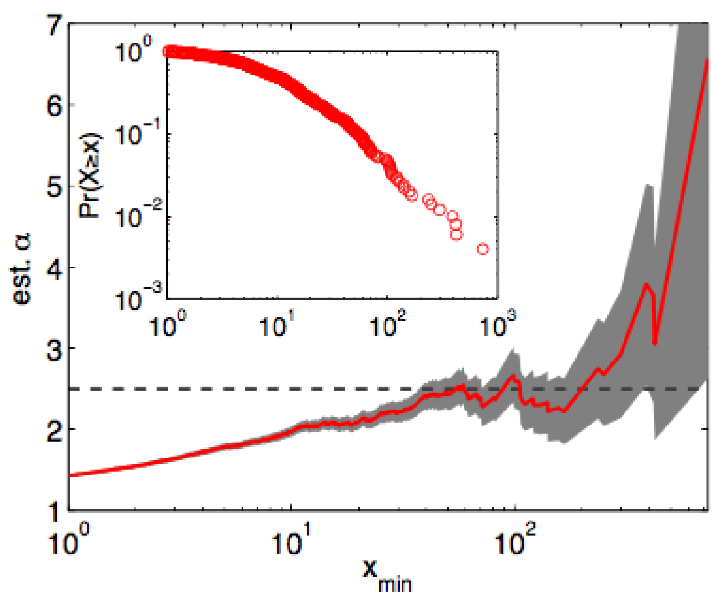

Figure 1: Hill Plot

Retrieved from http://tuvalu.santafe.edu/ aaronc/courses/7000/csci7000-001_2011_L3.pdf

From the figure above we could see the Hill plot for $n = 500$ i.i.d. observations from a shifted power law, with parameters $\alpha = 2.5$, and $C = 15$; inside there is also a plot of counter cumulative distribution function (ccdf) of this sample. The diagram suggests that stability starts approximately from $x_{\min} = 20$, which, using the maximum likelihood approach, gives us $\hat{\alpha} \approx 2$, which makes the right tail of the distribution much heavier that it actually is. For smaller values of $x_{\min}$ the deviations from the true value of α is evoked from fitting a power-law model to non-power-law-data. For larger values of $x_{\min}$, the deviations from the true value of α is caused by the noise in sample (Stoev et al. 2006).

3.2.2 Theoretical Approach

Several theoretical approaches have been proposed in order to estimate the value of $x_{\min}$. Handcock and Jones (2004) presented the general model of the discrete case which was considering the data above and below $x_{\min}$, where the values of $x_i > x_{\min}$ would exhibit a power-law behavior, and the values below $x_{\min}$ would follow some probability distribution p_x. Then they search for the best fit of the model to the sample data, letting $x_{\min}$ vary. Their model, however, has some downsides. The most important one is the fact that it can not be generalized to the continuous power-law type (Clauset 2007).

The method that I will use in this paper, and which works for both discrete and continuous cases, was presented by Clauset et al. (2007). The idea behind the method is choosing the value $x_{\min}$ such that probability distribution of the observed data above $x_{\min}$ and the theoretical power-law distribution is as congruent as possible. Choosing $\hat{x}_{\min}$ lower than the true value of $x_{\min}$ will bring us statistical fluctuations due to differences in the samlpe and the model describing it. On the other hand, choosing $\hat{x}_{\min}$ greater than the actual value of $x_{\min}$ will reduce the efficiency of the data because observations between $x_{\min}$ and $\hat{x}_{\min}$ would have fitted the power-law model, and as a result our dataset became poorer. What is in beetween these two cases is what we are looking for.

I will rely on the most widely used measure of estimating the distance between two distributions - Kolmogorov-Smirnov Statistic (KS) (Press et al. 1992). KS statistics calculates the maximum distances between the cumulative distribution functions of the theoretical model and the data:

$$D = \max_{x \geq x_{\min}} |S(x) - P(x)|, \tag{8}$$

where $P(x)$ is the CDF of the theoretical power-law model with the best fit for all $x_i \geq x_{\min}$, and $S(x)$ is the CDF of our data where observations are at least equal to $x_{\min}$. Now, since we are looking for the minimal distance between two distributions, our aim is to minimize KS statistic D by finding an appropriate value of $\hat{x}_{\min}$. The theoretical CDF is given by the following equation:

$$F(x) = \left(\frac{x}{x_{\min}} \right)^{(-\alpha+1)}. \tag{9}$$

We compare (9) to the empirical CDF, which is simply the vector (1,(n-1)/n,. . . , 1/n). It is so because empirical CDF turns out to be a step function, which shows the part, or fraction of the whole sample less than some x. For each point of our observation data, we find KS statistic by comparing these two CDFs. Then, after getting the vector of these KS statistics, we choose the minimum one. The value of x_j corresponding to this minimum KS statistic point will be our $x_{\min}$.

Clauset et al (2007) describes other test statistics in order to estimate the lower bound $x_{\min}$. They include: Reweighted KS, Kuipers, and Anderson-Darling statistic. However, it the author claims these test statistics do not bring much different results from the regular KS statistic.

3.3 Power-Law Hypothesis Test

In the previous section we showed how to fit a power-law distribution to some sample data by finding appropriate threshold value $x_{\min}$, and the power-law exponent α. However, we do not know how well the data actually fits to a power-law distribution. For instance, if the data is generated from exponential, or lognormal distribution, whose shape even slightly resembles that of the power-law distribution - could always fit to the power-law. Even if the fit is very poor, it is still a fit.

In the literature there are few tests that would identify whether the data fits the hypothesized distribution. However, following Clauset (2007) I will rely on the test based on KS statistics. As described in section 3.2, the KS statistics was giving us back a number which was used as a quantity to test how close the data to the hypothesized distribution is. The method would be to calculate the value of D defined in previous section, and then if the returned number is small enough, the powel-law distribution is a suitable model for our observations sample; otherwise if it is large, then power-law is not a good enough fit. Consequently, the appropriate question would be how large it should be so we can rule-out the hypothesized distribution. Clauset et al (2007) suggest that the best option would be to pick up the significance level equal to ten percent. That is, if p-value is greater than 0.10, then we would accept the null hypothesis that both data sets are coming from the same distribution, otherwise we reject it.

Since we want to detect a power-law behavior, we will compare our data set with a randomly power-law generated data. Clauset et al (2007) suggest to generate 2500 random points in order to get correct p-value within two decimal points, and thus to perform 2500 KS-tests.

3.4 Competing distributions

A high number of p-values larger than our siginficance level does not guarantee that the power-law is the best fit. It just tells us that our data set fits it plausibly well. Thus, we need to test for lognormal and exponential distributions that have approximately the same shape as the one of the power-law. If these two other distributions also would fail to reject the null hypothesis, then we will have to perform likelihood ratio test. This test that would help us to determine which of two distributions is the better fit for our data points. The idea behind this test is to compute the likelihood ratio of two distributions, and then the distribution with a higher likelihood ration would be chosen as a better fit.

4 Power-Laws in Kazakhstan: City and Firm Sizes

This section establishes the existence or absence of the cases of power-law behavior for city and firm sizes in Kazakhstan. To the best of our knowledge, the test for power-law behavior has not been conducted in Kazakhstan. Previous line of research has concentrated on developed economies (Segarra & Terruel, 2010). Therefore, it is of particular interest to detect statistical regularity in the state because Kazakhstan is a young country with emerging economy.

4.1 City Sizes Distribution

4.1.1 Literature Review

Auerbach (1913) was the first one to propose that the city sizes distribution could be approximated by the power-law distribution (Ioannides 2003). In other words, if we rank the cities in terms of their sizes from the most populated (rank=1) to the least populated one (rank=N), then the resulted distribution of city sizes would approximately follow a power law. Thus, if we denote the variable Y to represent the rank of the city, and the variable X to represent its size, then we would have a relationship of the type $Y = kX^{-\alpha}$, or equivalently

$$\log(\text{rank}) = \log k - \alpha \log \text{size}$$

Zipf (1949) made a further extension of the work done by Auerbach, and stated that the power-law exponent α of the city size distributions should be approximately equal to one. Gabaix (2009) analyzed the largest 135 US metropolitan areas for the power-law behavior. He put the cities in descending order in terms of their population (New-York as first, followed by Los-Angeles, etc.). Afterwards, a graph of the log of the rank of the city on the y-axis was put against the log of the population of the repsective city on the x-axis. Surprisingly, the graph revealed a straight line.The linear regression of Gabaix's data set yielded:

$$\ln \text{Rank} = 10.53 - 1.005 \ln \text{Size}$$

The graph below shows the emerging plot for the 135 US metropolitan areas taken from the Statistical Abstract of the United States for the year 1991."

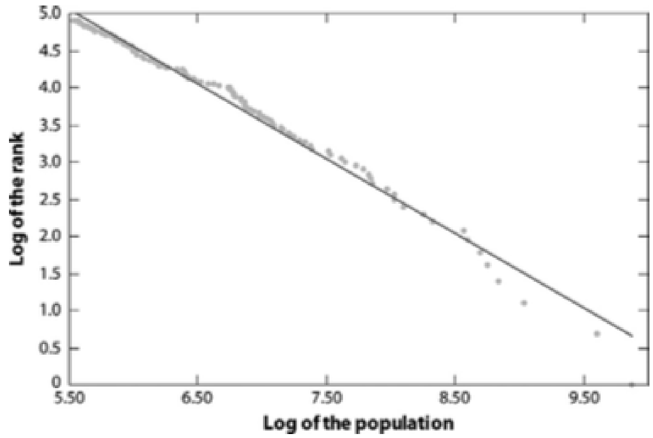

Figure 2: The figure shows the log size vs log rank for
135 US metropolitan areas, 1991. Obtained from Gabaix (1999).

Gabaix, one of the prominent researchers in this series of literature, showed a validity of the Zipf's law for cities using stochastic processes (Gabaix, 1999). Among other things, he shows that under assumption that distinct cities' growth is a random process but with the same variance and with the same expected growth rate (Gibrat's Law), then the city size distribution in its limit converges and tends to follow the special case of the power-law - Zipf's law. The support for Zipf's law has been validated for a number of countries (Rosen and Resnick, 1980).

However, cases where the power-law has been refuted for cities is also not a new for the literature (Brakman, 1999). This kind of deviations from the power-law regularity is a source of inspiration for this paper. Since the line of research concentrated mostly on developed countries, it possesses an interest whether developing country like Kazakhstan is one more exception from such deviation or not.

4.1.2 Case of Kazakhstan

I have obtained the data for Kazakhstani cities from Committee on Statistics of the Ministry of National Economy of the Republic of Kazakhstan (stat.gov.kz) for the year 2014. In total there are eighty seven (87) populated localities that have a status of the city. The mean size of the city was 107942.6, with standard deviation being equal to 211836.5. The largest city inside this data is Almaty with population of the city equal to one and a half million, while the lowest populated city is Zhem city (2033 people living there).

I applied procedures described in sections two and three. Kolmogorov-Smirnov (KS) test returned that 2459 out of 2500 data points, or equivalently saying, 98% of the whole data set fail to reject the null hypothesis (which states that two data sets are coming from the same distribution). Thus, we can claim the existence of the power-law behavior in Kazakhstani city sizes. The overall city size distribution in Kazakhstan fits the power law with parameters $\alpha = 1.53$, and $x_{\min} = 24302$. This result is in line with many other papers detecting the power-law behavior in city sizes (Rosnick & Resnick, 1980). Figure 3 below shows the plot of Kazakhstani city sizes distribution, obtained by regressing the log of the rank (y-axis) on log of population (x-axis), as Gabaix (1999) did in his work. The fitted power-law line is marked green.

As was mentioned above, Gabaix (1999) showed that the the city size distribution should converge to the "Zipf's line" in its limit, given cities' identical expected growth rate and variance. Thus, according to the theory, we might expect continual population reduction in the middle size cities. Consequently, six data points below the theoretical line, referring to the group of the biggest cities in terms of population (Almaty, Astana, Shymkent, Karaganda, Aktobe, and Taraz) are expected to approach the theoretical Zipf's line. Again, the latter is going to happen due to the size reduction of the middle size cities.

Another year I have picked up for the test of the power-law behavior in Kazakhstani cities is 1999. The reason behind looking for this particular year was motivated by the transfer of the state capital from Almaty to Astana in 1997. It posed a specific interest of how that event influenced the city size distribution after a couple of years. Applying the same

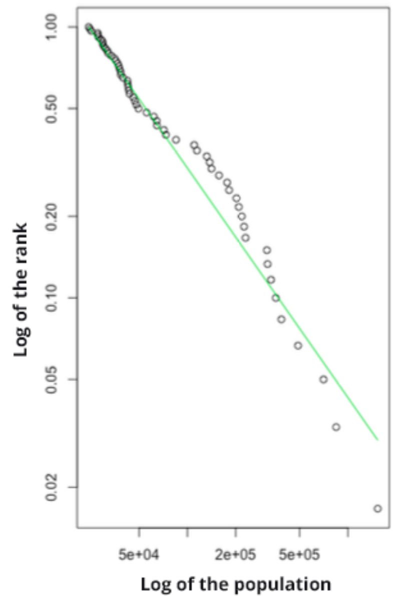

Figure 3: City sizes distribution in Kazakhstan, 2014

techniques, I found that the power-law behavior failed to exist in 1999. KS-test returned that 0% of the whole data set fail to reject the null, impying that the power-law is not a good fit for the city size distribution of Kazakhstan in 1999. City sizes distribution for the year 1999 is given in Figure 4 below.

Such deviation from the statistical regularity might be explained by the resettlement of considerable part of population during the years 1998-1999. For example, the total number of people migrating in years 1998-1999 was close to 450,000. This number is only four times less than the total number of internal migrants in Kazakhstan in the next ten years period (2000-2009).

I also want to find the support for the theory of Gabaix (1999) about the city size distribution convergence to the Zipf's law. For this reason, the data of Kazakhstani cities for the year 2018 was used. Applying the same techniques, I find that KS-test returned that 2418 out of 2500 data points (approximately 97% of the whole data) failed to reject the null (two data sets are of the same distribution). Therefore, there is a one more evidence of the power-law behavior. What is more important, I find that the power-law exponent α is equal

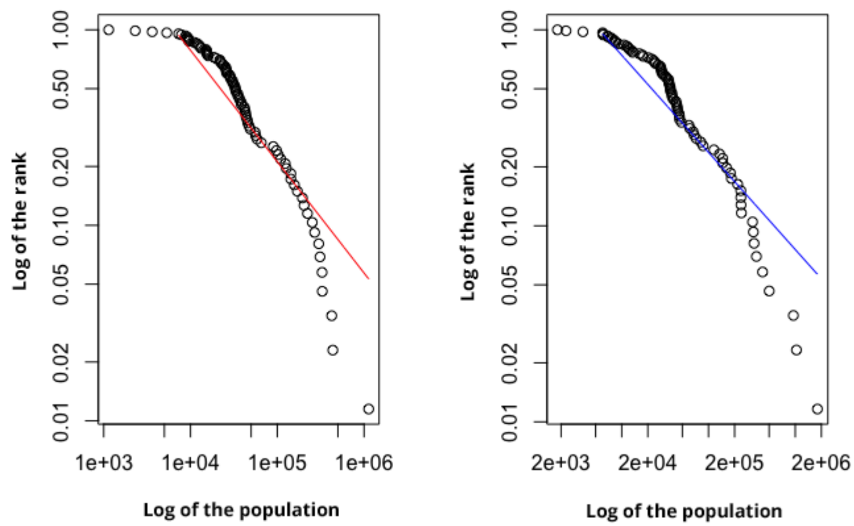

Figure 4: City size distributions for the years 1999 (on the left) and 2018 (on the right)

to 1.49. In a four-year period there was a slight decline in the exponent, from 1.53 in 2014 to 1.49 in 2018.

Having considered years 2014 and 2018, I find a support of the theory proposed by Gabaix (1999), in a way that the limit of the city size distribution will converge in a manner to obey the Zipf's Law. The future extension of the work might have a look on evolution of the city size distribution after some period, and seek for the new evidence to support Gabaix's theory (1999).

The Table 1 below briefly summarizes findings of this section.

Year	**PL behavior**	**p-value**	**PL exponent** α
1999	N	-	-
2014	Y	0.02	1.53
2018	Y	0.03	1.49

Table 1

4.2 Firm Sizes Distribution (FSD)

4.2.1 Literature Review

Power-law behavior in the distribution of city sizes is a classical example in the line of research. As we have seen previously, Kazakhstan is not an exemption from this wide-spread stastical regularity. Having detected that, I now focus on the firm size distribution (FSD) in Kazakhstan.

Gibrat's (1931) was the first one to raise the interest in analyzing FSD. In his PhD thesis, he proposed that the FSD can be approximated by the lognormal distribution. He aimed to provide both theoretical and empirical base for his finding. Today's interpretation of Gibrat's law is an extension of his primary argument. Loti et al. (2013) claims that the main usage of the Gibrat's law presented in a number of related articles is following: growth rate of the firm is independent from its size. It has to be noted that in Gibrat's original work that was only an assumption.

Conventionally, it was believed that the firm size is log-normally distributed (Hall, 1987 and Evans, 1987). However, in the last decades a number of empirical studies reveal that the FSD is better described by Pareto distribution, also known as the power-law distribution (see Axtell, 2001). Crosato and Ganugi (2007) demonstrate that both power-law and lognormal distributions can be derived from multiplicative random growth models à la Gibrat.

Yet, there are examples of FSD in some countries which do not exhibit a power-law behavior. Still, most of those empirical findings estimate that at least the right-tail of the FSD fits power-law well (Okuyama et al. 1999; Steindl, 1965; Ijiri and Simon, 1977). The literature on this topic is abundant, because the distribution of the firm sizes is rather important. For instance, FSD is a critical element in estimating the connection between alteration in employment and output (Hölger et al, 2016).

However, this section's main insipiration for looking for a FSD is coming from the relatively recent research about aggregate fluctuations in the economy. Before it was established that idiosynchratic shocks on the firm level are not able to produce aggregate movements as they “average out” as a part of the whole economy. Nonetheless, fresh line of research shows that this is not necessarily true.

For instance, Gabaix (2011) proposes that idiosyncratic firm-level shocks can result in aggregate fluctuations, if the firm-size is power-law distributed. Therefore, if there is some sign of the shock in the macroeconomy, then the government would want to contribute to the recovery of the firm located in the fat-tail of the distribution.

Also, the importance of studying the size distribution of firms aligns with the theory of ”Cascades of Effects” in the paper “Cascades in Networks and Aggregate Volatility” (Acemoglu, 2010). In his paper Acemoglu studies the consequence of the low productivity of the relatively large firm on the rest of the economy. It turns out that the microeonomic shock will originate a negative aftereffect, and may cause aggregate fluctuations. (Acemoglu, 2012 shows that this assumption strongly depends on the configuration of the networks in the economy).

Factors mentioned in the last paragraph are not the only ones that make the study of the firm size distribution interesting and important. Thus, the analysis of the firm-level distribution is of vast importance.

4.2.2 Case of Kazakhstan

The firm level data being used is the 2008-20014 Bureau van Dijck’s ORBIS commercial database. In total there are approximately 67,000 firms from all sectors of the Kazakhstani economy (manufacturing, agriculture, and services), and from every region in the country. The database consists of the company’s year end accounts. Almost every firm in our data sample has information on employment (firm size), age of the firm, and the region where it is operating.

The year 2014 had the most observations, so I decided to test our hypothesis for this year first. The largest size of the firm for 2014 was 16761, with the mean and standard deviation being equal to 60.61 and 204.4, respectively. I have applied tecnhiques described in section three. For this year, the power-law behavior failed to exist. Kolmogorov-Smirnov (KS) test revealed that out of 2500 points, 837, or approximately 34 % failed to reject the null hypothesis (which states that two data sets are coming from the same distribution). That is why we say that there is no evidence of the power-law behavior in Kazakhstani firms levels. Exponential distribution - 0 out of 2500, 0% of the points fail to reject the null. Figure 5 below shows Kazakhstani FSD for the year 2014. Log of the rank is placed on the y-axis, while log of the size is on x-axis. Theoretical red line is a fitted power-law distribution line.

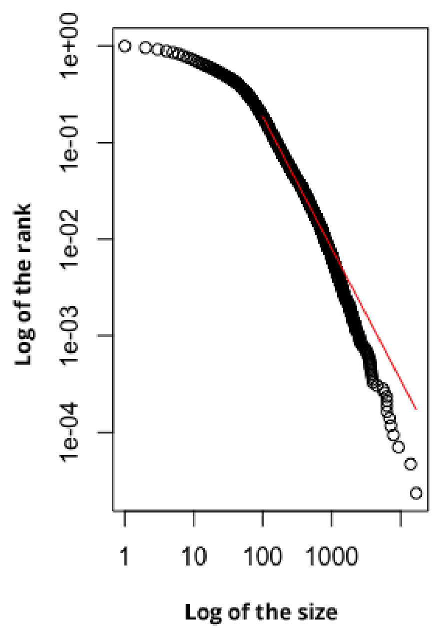

Figure 5: Firm Size Distribution, 2014

Yet, it would be still useful to acknowledge whether the upper-tail of the distribution follows the power-law or not (Okuyama, 1999; Ijiri & Simon, 1977; Steindl, 1965). Applying again procedure described in section three, we find that the upper-tail of the FSD for the year 2014 fits the power-law with $x_{\min}$ equal to 99, and the power-law exponent α equal to 2.36.

One of the possible explanations behind the absence of the overall power-law behavior might be the year that we have picked up for analysis. So the hypothesis was tested on other years in the data set. However, KS-test again did not support the existence of the power-law behavior in Kazakhstani firm levels.

Thus, Kazakhstani firm-level distribution did not exhibit a power-law behavior. Therefore, we are not able to line our results with the works of Gabaix (2011) and Acemoglu (2010, 2012). Equivalently saying, our result about the absence of the power-law behavior in Kazakhstan cannot follow the theory of Gabaix (2011) and Acemoglu (2010, 2012).

Possible explanation may be derived from Sutton (1997). Based on the model proposed by Simon (1977), Sutton presents an alternative way of stating the Gibrat's law: "The probability that the next opportunity is taken up by any particular active firm is proportional to the current size of the firm." This law is not in line with results obtained by Konings (2016).

Konings (2016) showed that Kazakhstani FSD is less skewed than in typical market economies. Therefore, a firm growth barriers like access to global markets and finance, as well as the deficiency of high managerial skills might take place. By running a few firm size growth regressions, Konings (2016) discovers that in Kazakhstan small firms grow faster than large ones.

This incongruence of results obtained by Konings (2016) with Sutton's (1997) version of the Gibrat's law is one of the possible explanations behind the absence of the power-law behavior in Kazakhstani FSD.

5 Statistical Discrepancy or Made-up Data?

Earlier in this paper, we have seen that FSD for the most of the countries does follow a power-law. The absence of such regularity in Kazakhstani firms level might be explained by misleading in the data set collection, or some statistical errors. However, there even might be the case that the data was made up.

Here in this chapter we will try to see whether there is a fraud in our data using one more statistical regularity (law), the Benford's Law. Irmay (1997) presents the evidence that

Zipf's law is closely related to Benford's distribution.

Benford's law has been applied to fraud detection: Nigrini (1996) proposed using Benfords Law to detect the fraudulous tax returns. Also, B. Rauch and Gottsche (2011) analyzed financial data of European Union countries using Benfords law. They found that among these countries, Greece had the largest deviation from Benfords Law.

Since the previous section revealed the absence of the power-law behavior in the distribution of Kazakhstani firm sizes, unlike in the other countries (Axtell, 2001), this section will test the data we have for the fraudulence (using Benford's Law).

5.1 Benford

Benford's Law is named after Frank Benford, an American physicist and electrical engineer at General Electric. In 1938 Benford analyzed 20 data sets, including areas of rivers, population census data, specific heat of chemical substances, atomic weights, house address data, and mathematical sequences. He published his findings in a paper "The Law of Anomalous Numbers" (1938).

Benford found that for nearly all data sets the distribution of the first digits is approximately given by

$$P(d) = \log_{10}\left(1 + \frac{1}{d}\right), \quad \text{for } d = 1, 2, \ldots, 9$$

d	1	2	3	4	5	6	7	8	9
$P(d)$	30.1%	17.6%	12.5%	9.7%	7.9%	6.7%	5.8%	5.1%	4.6%

Table 2: Benford's Distribution

5.2 Mathematical Explanation of Benford's Law

Let d_1 be the first non-zero digit of $n \in \mathbb{R}$. Then

$$d_1 10^k \leq n < (d_1 + 1)10^k \quad \text{for some } k \in \mathbb{Z}.$$

Taking the base 10 logarithm on each side of this inequality, we have

$$k + \log_{10} d_1 \leq \log_{10} n < k + \log_{10}(d_1 + 1).$$

From this we conclude that

$$\log_{10} d_1 \leq \{\log_{10} n\} < \log_{10}(d_1 + 1),$$

where $\{t\}$ is the fractional part of t.

Assuming that the numbers $\{\log_{10} n\}$ for the n's from our data set are distributed uniformly modulo 1, the proportion of n's satisfying

$$\log_{10} d_1 \leq \{\log_{10} n\} < \log_{10}(d_1 + 1),$$

is given by

$$\begin{aligned} P(\text{first digit is } d_1) &= \log_{10}(d_1 + 1) - \log_{10} d_1 \\ &= \log_{10}\left(\frac{d_1 + 1}{d_1}\right) = \log_{10}\left(1 + \frac{1}{d_1}\right). \end{aligned}$$

See Diaconis (1977) for more details.

5.3 Benford's Law in Action: Firm Size Data Test

The firm level data obtained from Konings (2016) was tested against Benford's law. Three tests were used in order to detect statistical irregularity in the dataset: Pearson's Chi-Square, Mantissa Arc, and Mean Absolute Deviation (MAD) tests. For Chi-Square and Mantissa Arc test, we tested the following null hypothesis: our empirical distribution has no association with the Benford's distribution. The p-value for chi-square and Mantissa tests is approximately the same and is low enough to reject the null (p-value $< 2.2e^{-16}$). In MAD test, we are looking for the average distance between the empirical and theoretical data points. It turns out that the MAD test's mean absolute deviation is also quite small (0.008440135), so there is no significant deviation from the theoretical distribution.

Figure 4 below shows actual (firm level data) versus theoretical data points. Vertical axes show the frequency of the first digit appearing on each data point, while horizontal axes show the digits itself. Theoretical distribution is shown by the red dashed line, while blue bars show first-digit distribution of our data set. It could be seen that the firm level data distribution is not far away from Benford's theoretical distribution.

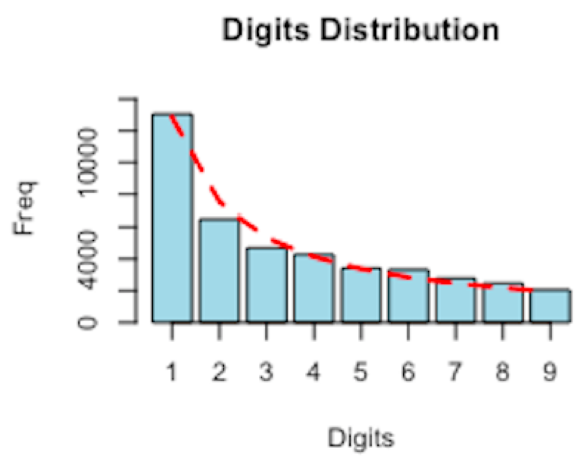

Figure 6: Firm Level Data vs Benford's Law

Thus, we make an inference that the data set we obtained is not a subject for suspicion, and apparently there should be some other specific reason behind the absence of the power-law behavior in Kazakhstani firms.

6 Conclusion & Future Work

This paper tested Kazakhstani city and firm levels for existence of the power-law behavior. The procedures described in sections two and three were used in the process. Thus, Kolmogorov-Smirnov test revealed existence of the power-law behavior for the distribution of city sizes in Kazakhstan for years 2014 and 2018. Moreover, our results are in line with the work by Gabaix (1999), where he showed that the power-law exponent α should converge to one, if specific assumptions hold. Future work may look on evolution of the FSD (particluarly on its power-law exponent) after some period of time.

In regards with the firm level, we failed to find the evidence of the power-law distribution in Kazakhstani firm sizes. The outcome of the analysis was rather unexpected, since it is believed that the firm size distribution typically follows a power-law. However, I found out that the upper-tail of the firm size distribution of our dataset follows a power-law, which is in line with most of the papers (Okuyama, 1999; Ijiri & Simon, 1977; Steindl, 1965; etc). Feasible explanation behind the deviation from the popular statistical regularity is the fact that Gibrat's law does not work well in Kazakhstani market.

One more reason might have been in data fraudulence. The test for Benford's law in detecting the fraudulence in the data was applied. According to theoretical results, the fraudulence in the data was not detected.

The future work is to find a credible explanation of the absence of the power-law behavior in Kazakhstan. One possible explanation might be the fact that Kazakhstani government has a strong interference on the state-owned firm levels.

References

1. D. Acemoglu, A. Ozdaglar, A. Tahbaz-Salehi, *Cascades in Networks and Aggregate Volatility.* NBER Working Paper No. 16516 (2010)

2. D. Acemoglu, V.M. Carvalho, A. Ozdaglar, A. Tahbaz-Salehi, *The network origins of aggregate fluctuations.* Econometrica (2012), 80, 1977-2016.

3. T. C. Anderson, et. al. *Benford's Law for Coefficients of Modular Forms and Partition Functions.* Proceedings of the American Mathematical Society (2011), 139(5), 1533-1541.

4. A.B. Atkinson, T. Piketty, E. Saez, *Top Incomes in the Long Run of History.* Journal of Economic Literature 2011, 49(1), 371.

5. F. Auerbach, *Das Gesetz der Belvo lkerungskoncentration.* Petermanns Geographische Mit- teilungen (1913), 59, 7476.

6. R. L. Axtell, *Zipf Distribution of U.S. Firm Sizes.* Science (2001), 293 (5536), 1818-1820.

7. F. Benford, *The law of anomalous numbers.* Proceeding of the American Philosophical Society (1938), 78(4), 551-572.

8. S. Brakman, H. Garretsen, C. Marrevijk, M. Berg, *The Return of Zipf: Towards a Further Understanding of the Rank-Size Distribution.* Journal of Regional Science (1999), 39(1), 183-213.

9. A. Clauset, C.R. Shalizi, M.E.J. Newman, *Power-law distributions in empirical data.* SIAM Review (2009), 51, 661-703.

10. W.K.T. Cho, B.J. Gaines, *Breaking the (Benford) law: Statistical fraud detection in campaign finance.* American Statistician (2007), 61(3), 218-223.

11. L. Crosato, P. Ganugi, *Statistical regularity of firm size distribution: the Pareto IV and truncated Yule for Italian SCI manufacturing.* Statistical Methods and Applications (2007), 16(1), 85-115.

12. P. Diaconis, *The Distribution of Leading Digits and Uniform Distribution Mod 1.* The Annals of Probability (1977), 5(1), 72-81.

13. P. Duparcq, J. Konings, *An Enterprise Map of Kazakhstan: A Guide for Industrial Policy.* Retrieved from http://shss.nu.edu.kz/shss/academics/departments/economics (2006).

14. X. Gabaix, A. Landier, *Why Has CEO Pay Increased So Much?.* Quarterly Journal of Economics (2007), 123(1), 49-100.

15. X. Gabaix, *Power Laws in Economics and Finance.* Annual Review of Economics (2009), 1, 25593.

16. X. Gabaix, *The Granular Origins of Aggregate Fluctuations.* Econometrica (2011), 79(3), 733-772.

17. R. Gibrat, *Les Inègalites Économiques.* Paris: Librarie du Recuiel Sirey (1931).

18. M.S. Handcock, J. H. Jones, 2004, Theoretical Population Biology 65, 413.

19. G. Hölger, P. Henze, et. al, *Firm Size Distribution and Employment Fluctuations: Theory and Evidence.* Kiel Working Paper (2016).

20. Y.M. Ioannides, H.G. Overman, *Zipfs law for cities: an empirical examination.* Regional Science and Urban Economics (2003), 33, 127-137.

21. S. Irmay, *The Relationship Between Zipf's Law and the Distribution of FIrst Digits.* Journal of Applied Statistics (1997), 24(4), 383-393.

22. M. Nigrini, *A taxpayer compliance application of Benford's law.* Journal of the American Taxation Association (1996), 18(1), 72-91.

23. M. Nigrini, *Forensic Analytics: Methods and Techniques for Forensic Accounting Investigations*. John Wiley & Sons, Hoboken, New Jersey (2011).

24. B. Rauch, M. Gottsche, *Fact and Fiction in EU-Governmental Economic Data*. German Economic Review(2011), 12(3), 243-255.

25. K.T. Rosnick, M. Resnick, *The size distribution of cities: An examination of the Pareto law and primacy*. Journal of Urban Economics (1980), 8(2), 165-186.

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

27. G. K. Zipf, *Human behavior and the principle of least effort.* Cambridge, (Mass.): AddisonWesley, (1949) pp. 573.