ZIPF'S LAW AND CITY SIZE DISTRIBUTION: THE CASE OF CITIES AROUND THE WORLD

By

AFAQ AHMAD CHUGHTAI

PIDE2019FMPHILEAF24

Supervisor

Dr. Ahsan Ul Haq Satti

PIDE School of Economics

Pakistan Institute of Development Economics,

Islamabad

2022

Pakistan Institute of Development Economics, Islamabad

PIDE School of Economics

CERTIFICATE

This is to certify that this thesis entitled: **"Zipf's Law and City Size Distribution: The Case of Cities Around the World"** submitted by **Mr. Afaq Ahmad Chughtai** is accepted in its present form by the School of Economics, Pakistan Institute of Development Economics (PIDE), Islamabad as satisfying the requirements for partial fulfillment of the degree in Master of Philosophy in Economics and Finance.

Supervisor: Dr. Ahsan ul Haq Signature: ____________

External Examiner: Dr. Usman Shakoor Signature: ____________

Head,
PIDE School of Economics: Dr. Shujaat Farooq Signature: ____________

Author's Declaration

I Afaq Ahmad Chughtai hereby state that my MPhil thesis titled "Zipf's Law, and City Size Distribution: The Case of Cities Around the World" is my work and has not been submitted previously by me for taking any degree from Pakistan Institute of Development Economics or anywhere else in the country/world.

At any time if my statement is found to be incorrect even after my Graduation theuniversity has the right to withdraw my MPhil degree.

Date: 21/01/2022

Signature: A. Chughtai

Afaq Ahmad Chughtai

Dedications

To my beloved parents, Brothers, Friends Thank you for your moral support.

ACKNOWLEDGEMENTS

All praises to ALLAH, the compassionate, the omnipotent, whose blessing and exaltation flourished my thoughts and thrive my ambitions, provided me with a rich environment of learning and cooperative teachers, helping friends and those who honored me among those who contribute to the scared wealth of humanity. Initially after thanking Allah Almighty, I would like to thank and my dedication goes to my family, my Father Sardar Muhammad Munir Chughtai and my lovely, beautiful, and best Mother (Abida Parveen), my brothers Abrar Munir Chughtai & Afzaal Ahmad Chughtai, my two lovely and caring Sisters who motivate and helps me in completing this milestone.

From the formative stage of this thesis to the first draft, I owe an immense debt of gratitude to my supervisor DR AHSAN UL SATTI for his sound advice and careful guidance through the process. I can never repay you for all the help you have provided me and the precious time you spent making sure my thesis is always on track. Thank you so much!

Finally, I would be remiss without mentioning my friends who help me a lot in this stage of my life, Umar Daraz Khan, M. Babar Basharat Abbasi (BBA) and Uzair Ali who was there in every problem which arises in my life and their extreme generosity will be remembered always.

To each of the above, I extend my deepest appreciation.

ABSTRACT

The question of whether and then why Zipf's law applies to city sizes is at the heart of the research. This study is focusing on whether Zipf's rule remains true in a global context, encompassing all cities on the planet. Previous studies, including this one, depend mostly on traditional official statistics just as population and enumeration department forced city designations. Here we investigate the law further at the level of country discover that Zipf's law is disregarded and varies in time and from country to country. Our research investigates whether the rule of Zipf applies to all situations globally. Zipf law applied to city numbers the outcome is the cities numbers for the first biggest country is double that of the second biggest country, Three times higher than that of the third-largest country, and so on.

An urban system may be described by a Pareto exponent with a parameter (β) value of 1. The empirical validity of Zipf's Law is assessed in this promoting research fresh data from 235 countries and two distinct estimate methods – conventional OLS, Rank-half rule, and the Wald test. Using this Methodology and data set we divided the data into three groups; Developed countries (HIC), Developing (MIC), and underdeveloped countries (LIC). 4 out of 65 developed countries obtained the β- values significantly, while approx. 25% got a higher value than unity. Eight MICs got values near to unity and approx. 18% got a value higher than our hypothesis of (β=1). Likewise, Underdeveloped nation two countries Yemen and DR Congo got statistically sig. a value equal to 1.

Zipf law or Power Law is rejected for most of the countries, most the nations have value less than which implies that overall, there is more concentration, uneven dist. of population, more hierarchies are there in these maximum number countries. These countries have large cities which are more concentrated as compared to small and medium-sized cities.

Keywords: Zipf's law, Rank-size distribution, Pareto distribution, Human Development Index, Concentration, Hierarchies.

TABLE OF CONTENTS

List of Figures:

List of Tables:

List of Abbreviations:

β	Beta as Zif's Coefficient/ Pareto Exponent
Dist.	Distribution.
ETS	Department of Econometrics and Statistics
HIC	High-Income Countries.
LIC	Low-Income Countries.
MIC	Medium Income Countries.
PIDE	Pakistan Institute of Development Economics
PE	Pareto Exponent.
PL	Power law
RSD	Rank- size distributions
ZL	Zipf's law

CHAPTER 1

INTRODUCTION

The chapter mentioned in detail the basic idea of the research the background of the Zipf's law, defining urban system and city size distribution: A case study from all cities in countries around the World. Specifically, this chapter discusses that the introduction of Zipf's law, and the urban systems in different countries. At first, there is a brief introduction to the research work. Then there is a discussion on some questions like; does this law apply to all cities in each country globally? and does this law apply to all cities of the world or just large cities? After discussing the research gap, we will become able to discuss the objectives of the research. The section about the significance of the study examined Zipf's law an emerging tool for urbanization and its possible effect on urban development and sustainable urban growth in these countries.

1.1 Background

An openly observed regularity for cities worldwide is the size of the city that is inversely related to rank. More precisely and accurately ordering cities is a global habitation of a country. It can be observed that the city ranked /ordered one is two times bigger than the city ranked at second position and three times than third positioned city and so forth. This regularity is known as Zipf's law. The name of this law was after the linguist (Zipf, 1949) and was discovered initially by a German physicist (Auerbach, 1913). In our study, we will stick to the exponent of Zipf's Law equal to (1). As mentioned by (Zipf, 1949) for measurement of population concentration the Pareto exponent ought to be utilized between various cities in the entire world. For endorsement of the rank-size rule, it is conditioned/ necessary that the exponent must equal 1. Since, it then shows that on the value of average, these values consist (Jiang & Jia, 2011). This study will tell us more about the estimation of the Pareto Exponent for each census year of all countries to deviation in size distribution.

The power law is a speculative regularity that comes in being when the value of the event or quantity has a negative relation to the power of the value of that event. Zipf's law and Pareto laws are the mathematical/ numerical patterns of regularity that are further assigned to a power law. It has been noticed that Zipf's law is being used in various fields of Economics i.e., world income, gravity model of international trade and the number of employees in firms. Moreover, this well knows the regularity of Zipf's Law (ZL) have prestigious value for the urbanization system which states that the size distribution of cities must be according to the Rank Size Distribution. RSD can easily be observed through the association between the size of the population and then the rank of cities by the population in different states.

Surprisingly, for at least 100 years world is familiar with this law. Despite few investigators are inclined to be ambiguous about this law for all cities even (Gabaix, 1999), who seem clear about it, accepts its cogency for the cities huge in size on a different point of times. Does this uncertainty move around some fundamental inquiry: (1) Either this rule implemented in all urban areas and provinces in all the countries in the world? (2) does this rule applies worldwide to all the urban areas or just large urban areas within the countries? While observing literature and previous studies we counter a contradictory situation that portrays an opposite image of Zipf law (ZL). Though it's been professed as universal, however, an academic perspective above mentioned questions are always not valid, specifically in case when data is big. Zipf's law shows a statistical regularity, and it requires an adequately huge sample. For that reason, Zipf's law would not be implemented in the case of a very small sized number of urban areas. This is quite comprehensible as entire cities are usually interconnected or interdependent because of the global village. However, this interconnectedness is not always there because primate cities (Jefferson, 1989) go beyond their country borders. Through these given, a valid question would come up that does Law of Zipf is applicable worldwide for all the urban areas (or specifically for the large-sized cities).

The motivation for this study comes from many previous studies, (Krugman, 1996), (Gabaix, 1999), (Cordoba, 2001), (Axtell & Florida, 2000), (Reed, 2002) which provides the explanation theoretical explanations for the "empirical fact" that the rank-size-rule for cities holds in general across countries. . The evidence they present for the existence of this fact comes in the form of requests to past work such (Rosen & Resnick, 1980), or some regressions on a small sample of countries which focuses mainly the US. One limitation of such appeals to the Rosen and Resnick result that their study was 20-year-old and the data they have taken was from 1970. Thus, one insistent need is for newer evidence on whether the rank-size-rule remains to hold for a large sample of countries.

The present study sets out to do four things, firstly, to test Zipf's Law, and to examine the Rank Size Distribution of cities considering city-wise data of each country in the world. Using a new dataset that includes a larger sample of countries. The second is to examine the urban and total population size distributions are adequately described by a pure Power Law in all countries. Third, it analyses the distribution of the Pareto exponent to give an indication of its shape and to yield additional insights and to examine the urban and total population size distributions of each country regarding Developed, Developing, and Underdeveloped countries. Finally, this study sets out to explore the relationship between inequality in the sizes of cities as measured by the Pareto exponent and Power law and to estimate the Pareto Exponent for each census year of all countries to find deviation in size distribution and their overall trend by performing the analysis using the OLS estimation suggested by many studies like (Gabaix et al., 2004).

This question of the present study is new. Our primary debate states that the Zipf Law is universal, so data will be extracted on cities, from all countries in the world including developed, developing, and under-developed countries from the whole world. We will take population data from the different census of the country like, China is the most populated country in the world and the population census was taken at a different point in times like (1953, 1964, 1982, 1990, 2000, and 2010). Our investigation is to know whether the Zipf law holds or not. If it holds then for which countries does this law hold? Is it held for large cities only or is it held worldwide and what are the reasons behind this? For estimation of power-law detection for all cities at countries levels, we will use the traditional least square method.

The innovation of our study was possibly observed through different aspects. We will investigate the universality of Zipf Law (ZL) in a cosmopolitan setting including all the urban areas worldwide. We have used the data from different census bureau-imposed cities of the world. By and large, the present study will give a fresh point of view about the issue of Zipf Law (ZL) and analyse the law concerning census data.

1.2 Research Objectives

Specifically, there are four main objectives of our research: -

1. To examine the Rank Size Distribution of cities considering city-wise data of each country in the world.
2. To examine the urban and total population size distributions are adequately described by a pure Power Law in all countries.
3. To examine the urban and total population size distributions of each country regarding Developed, Developing, and Underdeveloped countries.
4. To estimate the Pareto Exponent for each census year of all countries to find deviation in size distribution and their overall trend.

1.3 Research Questions

The research will investigate four main issues:

1. Does the Zipf Law / Rank Size Distribution hold in the most populous countries?

2. How do the urban or total population size distributions deviate in Human Development Index groups?
3. How are the urban and total population size distributions adequately described by a pure Power Law?
4. Finding the value of the Pareto Exponent for each country to know about the deviation of rank-size distribution?

1.4 Significance of Study

Significant and persistent economic regional differences exist between urban regions vary from country to country. This study will provide the guidelines to policymakers and researchers to increase the scope of regional study and understand the benefits of rank-size distribution according to this tool, when there is a case of huge population and not well utilized the fertilized areas for urbanization for better development of the new cities in countries. This study will provide the theoretical and practical concepts for improving the effect of Zipf's law and policies on the urban system and city size distribution by the comparison of different countries in the world.

From the policy perspective, the analysis offers the understanding and treating issues of international concern related to the unequal distribution of urban areas, such as the intense core-periphery disparities and the unplanned expansion of metropolitan systems, outside their legally defined boundaries. Additionally, it addresses questions about whether the distribution of urban areas is adequately described by a pure power law in different countries and estimate the Pareto exponent for each census year of all countries in the world to find deviation in size distribution.

The study of the association between urbanization and Rank Size Distribution (RSD) of the cities from the perspective of countries in the world is new and mixed evidence is being observed. This work is related to nature which means that in different countries the urbanization system is not fully explored and examined Zipf's Law is an emerging tool for urbanization and its possible effect on urban development and sustainable urban growth in these countries.

1.5 Organization of Study

The second chapter of the study contains Literature Review. A brief analysis about Zipf Law, Power Law, Pareto Exponent, City distribution follows the Zipf Law, and City distribution does not conform to Zipf's Law. Third chapter consists of Data & variables and methodology which is to be used for finding out the distribution that conforms the Zipf's law. In first Subsection of chapter 3, contains the Theoretical & Empirical Framework, econometric methodology and defines the data. Fourth chapter of the study consists of results and discussions in this chapter we will be discussing the result in detail. Also, it consists of the sub-heading of the Dynamics of the cities. The results chapter will start with descriptive statistics of data and then results. Fifth chapter explains the conclusion and policy implication.

CHAPTER 2

LITERATURE REVIEW

George Zipf, a linguist, found something strange about how frequently people use terms in each language. He discovered that just a tiny number of terms are used frequently, while the great majority are only used infrequently. Some notable structure exists when he ordered the words in order of popularity. The top-ranking term was always used two times every time the word second-ranking was used, and three times as often as the word third ranking was used. He termed it as the rank vs. frequency rule and discovered that it will also be used to explain income distributions in any region, with the wealthiest person making twice as much as the second richest, etc.

In a general study of languages (Moskowitz, 1959) stated that this law holds for many languages but still, it is unknown why this happens. With regards to urban community sizes, Zipf's law likewise holds. The population of the populous region in any country is usually double that of the next-largest, and so on. From the perspective of cities, Zipf's law (ZL) has remarkably, remained faithful to every region in the globe during the last century. A similar link may be found in many other rankings that aren't connected to languages, such as population rankings in various nations, company sizes, income rankings, and many more. Felix Auerbach in 1913 noted the dispersion in city rankings by population for the first time (Auerbach, 1913). Zipf's law is seen in normal just as in sociology said by (Zipf, 1949), (Shiode & Batty, 2000a), (Sinclair et al., 2010), (Li & Yang, 2002), and (Tachimori & Tahara, 2002) their review depicted that the position with size (S) is adversely identified with S at a given force and this force is equivalent to 1.

The size of cities and their ranking are shown to be inversely connected to cities all over the world. In decreasing order, rank all cities in the globe by population; the size of the first nation will be twice more than the size of the second, and the third country would be three times the size of the first. (Nota & Song, 2012) examined that there are two magnificent consistencies for Zipf's Law that are best applicable and the value of Pareto Exponent close to 1, then Zipf's law (ZL) will collapse into the Rank Size Distribution. A small sample of big cities have higher value and small cities provide us with small values so they suggested that Rank Size Distribution (RSD) must be clarified by taking considering the caution. Though the rank-size rule is more economical than the statistical phenomenon. It is just not easy to have a deeper knowledge of the Zipf exponent. The rank-size rule's validity is dependent on the value of Pareto Exponent (which is expected to β near to 1). Whether this exponent's magnitude differs considerably from one, the rule may be compromised, supporting (Gan et al., 2006) claims that Zipf's Law is a measurable pattern instead of a financial routineness.

An exponent under 1 would propose that a position Second city is not exactly a large portion of the size of the position of number one city. Then again, assuming the exponent power is multiple, Zipf's Law proposes that the Second biggest city is the greater part as rank One city, and the third biggest will be more than a third as extensive as rank one city. In this review they have picked China and the US and reasoned that Zipf's Law doesn't hold dependent on the hypothesis of financial aspects despite the two nations have an enormous number of urban communities with an alternate affordable framework which makes the Zipf's coefficient touchy to monetary variables (Nota & Song, 2012).

One more interesting point to consider is Zipf's coefficient. The coefficient in metropolitan advancement research is approximately near to Pareo Exponent (PE) value, consequently, the rule of Rank Size Distribution holds in that scenario. The rank-size rule, as per (Gabaix, 1999), is hypothetically a characteristic impact of metropolitan development, regardless of the city's starting size. By and large, rich per capita. In contrast, the measure of infrastructure scales subs linearly and indicates economies of agglomeration in bigger cities so the resulting aggregate sum of riches or infrastructure in a system thus depends in the end on the distribution of sizes of the various cities in the system (Cottineau, 2017).

The anticipated coefficient went from below the value of Pareto Exponent for cities in Morocco to above the value of Pareto Exponent for cities in Australia among the 44 countries assessed by (Rosen & Resnick, 1980). In another study, (Nitsch, 2005) took a gander at 515 values from 29 research and tracked down that estimated values for the coefficients lies in between the near value of Pareto Exponent where the middle value of the coefficients probably lies very slightly above the Pareto Exponent value which states that the distribution of the size of the urban areas more equally to the suggested rule of Rank Size Distribution.

According to (Soo, 2004) the estimation strategy and estimation methods selected lead to the working of Zipf's law. Further, he said variety in the worth of the Pareto Exponent (PE) is described by the politically based variables than by economic-based variables. In another study carried out by (Krugman, 1996) about the goodness of fit for Zipf's law which is perhaps quite possibly the most observable feature. For the log size, a linear regression of log-rank produces the amazing fit, as per several experimental research. The rule of Rank Size Distribution is "a significant discomfort for the perspective of economic hypothesis: as we know that one of the strongest statistical wonders, with no persuading hypothetical basis". Furthermore, he said about Zipf's law, which describes a surprisingly steady consistency in the physical layout of market economies is one of the most remarkable empirical laws in economics.

Another study offered the general equilibrium approach to rank-size distribution or Zipf law which mainly concerns the size distribution of cities. This strategy permits the powers of agglomeration and spreading to have been in since a long time ago runs balance, improving our insight into the truth of a city rank-size distribution (Brakman et al., 2001). Later studies observes the distribution of cities (Krugman, 1996) and (Fujita et al., 1999) whereas the distribution of size throughout cannot be explained by most deterministic urban models. Furthermore, cities typically do not increase at the same rate due to parameter variability. Findings by (Dobkins & Ioannides, 2000), the discovery of a significant quadratic term in a log-rank regression, continue to cast serious doubt about Zipf's law's validity as an explanation of the overall distribution of city size in the United States.

The authors' key finding is that Gibrat's rule, or proportionate development, can prompt Pareto distributions. (Gabaix, 1999) shows that the rule of Gibrat's may prompt Zipf's Law distribution in case the quantity of urban areas remains constant, whereas when the number of urban areas arises in that case of Rank Size Distribution only the top tail attains the Zipf's Law. In past some researcher discovers that a summed-up process of Gibrat's law (Córdoba, 2008) which enables the difference however not only the average of the development of the urban area continues to shift with the size of urban areas, may clarify for Pareto Exponent (PE) other than one regardless of whether the quantity of cities remains constant. He also explores it under gentle conditions, this summed up Gibrat's technique is needed to create a Pareto dist. of urban area sizes. We are interested in the broader situation of a discretionary exponent. To represent this proof, many probabilistic and models have been created. (Champernowne, 1953), and (Gabaix, 1999), and (Córdoba, 2008) use the probabilistic model which is the most notable model.

The metropolitan system is described by Zipf's law which had no constraint on size, area, and inside relocation (Kolomak, 2014). Movement patterns have a lot of significance over the Rank Size Distribution of cities of the country. (Simon, 1955) contended the Rank Size Distribution (PE) of an area would possibly adhere to Zipf's Law in case the movement of urban areas is corresponding to the size of that urban area, that is, the expansion or contraction of the population of individual urban areas inside the district was relative to the size of the city. The assertion that the Rank Size Distribution of cities is very much approximated according to the distribution of

Pareto Exponent is upheld by an enormous assemblage of experimental proof physicists (Auerbach, 1913) and (Zipf, 1949).

(Jefferson, 1989) studied that the information for Zipf's law was large because the cities are currently interconnected or reliant internationally however this interconnectedness is not always because primate cities are past their boundaries so there is an issue for discovering the reason for holding Zip's law in these cities. The calculated efforts of present-day geographers to scholarly analysis were studied to answer the question Why Geography? The primate city is an illustration of an idea that has been so extensively taken on an interdisciplinary application that its origins in current geography are ignored, if not neglected. Imprint Jefferson, a geographer, at first presented the idea in a short piece named "The Law of the Primate City," which showed up in the Geographical Review in April 1989.

This idea's use in the analysis of urbanism exemplifies just a single logical impact of present-day geography. The American Geographical Society is pleased to republish this piece for a new generation of readers to mark its fiftieth anniversary. Except for minor formatting changes, the content is verbatim, however, the population statistics on which Jefferson based his estimates is out of date. Aside from a detailed examination of the origins of one of the most significant notions to emerge from contemporary geography, readers will also come across a model of inductive reasoning and evidence of changing styles in geographical presentations. (Dimitrova & Ausloos, 2015).

The study related to quick urbanization in many non-industrial nations, (Henderson, 2002) throughout the last 50 years seems to have been joined by excessively undeniable levels of concentration of the metropolitan population in exceptionally enormous cities. Some level of metropolitan concentration might be desirable at first to diminish between and intraregional infrastructure expenditures. Be that as it may, in an adult system of cities, economic movement is more spread out. Standardized assembling creation tends to be concentrated into smaller and medium-sized metropolitan areas, whereas creation in huge metropolitan areas focuses on services, research and advancement, and non-standardized assembling. Easing excessively high metropolitan concentration requires investments in interregional transport and telecommunications to work with industry least. It also requires less fiscal concentration, so that inside cities can raise the monetary resources and offer the types of assistance expected to contend with primate cities for industry and population (Ades & Glaeser, 1995).

Before we begin evaluating the literature on cities, we'd like to draw your attention to one significant data element. Specifically, it makes a difference whether one is dealing with urban agglomerations (i.e., metropolitan areas) or with city-specific statistics. Conceptually, the proper entity is the urban mass as an urban economy, however international data frequently only supply the city proper data. (Rosen & Resnick, 1980) expressed a new point about the Pareto Exponent where they state the Pareto Exponent must be more noteworthy for the cities-based data than for metropolitan agglomeration information since metropolitan agglomerations are not constrained by lawful definitions of cities-legitimate thus neither have a more extended upper tail. Later, (Brakman et al., 1999) investigated this point of Pareto Exponent. The last provides comparisons based on global information. Despite these distinctions, and unless otherwise specified, the terms metropolitan and metropolitan are used reciprocally.

Rosen and Resnick investigate the size distributions of cities in 44 countries in 1970. Zipf's exponent is 1.13 on average, with a standard deviation of 0.19, and practically all countries fall between 0.8 and 1.5. Roehner studies various nations and in these exercises, the Pareto exponent for the United States is generally close to one, but in the case of other nations, it is not the same. Whereas the theoretical work contributes more prominently to the US-based situation, by considering the Pareto Exponent of one as an assumption.

The case for Brazilian cities of more than 30,000 population has been researched by (Moura Jr & Ribeiro, 2006). They demonstrated that the distribution of Pareto did not apply to smaller towns. The cumulative distribution function in the city size did not follow power-law conduct in these circumstances. The values of the coefficient were determined using three methods: maximum probability estimator, minimum quadrature fit, and average parameter estimator, and the maximum probability estimator was provided more precise findings. (Soo, 2004) revises these findings without changing the fundamental findings. The Zipf exponent has a significant estimated dispersion. Some regard this as contradictory evidence for Zipf's law. Taking a gander at the average of exponential estimates, we can observe that if the average value of the exponent is not precisely corresponding to $\beta = 1$. We infer that power law, with a Zipf exponent normally around one, precisely explains exact consistency.

According to (Brakman et al., 2001; Brakman et al., 1999), city-proper data are related with greater Zipf exponents. $(\text{Mean} = 1.13, \text{S. D.} = 0.19, \text{N} = 42)$ than urban agglomeration data $(\text{Mean} = 1.05, \text{S. D} = 0.21, \text{N} = 22)$. Brakman et al. studied in the case of Netherlands. In the case of two other countries namely Japan and France such instances were investigated by (Eaton & Eckstein, 1997). Numerous country studies and comparative international facts support Zipf's law. The most comprehensive empirical worldwide comparison studies are (Rosen & Resnick, 1980), (Brakman et al., 2001), and (Soo, 2004).

Furthermore, forecasting a value in a wide spectrum such as $[0.8, 1.2]$ may be the listing of the benchmarks applied by (Dobkins & Ioannides, 2000). Further, it provides OLS estimates of the exponent, through recurrent cross-sections data of the US Census for metro regions. Their estimates for power-law distributions decrease over time and they offer optimum predicted values.

As indicated by (Fujita et al., 1999), the estimation of Pareto Exponent (PE) decreases from 1900 to 1990, where the sample number for metro regions consist of 56, whereas, the metro areas consist 0.167 for 1900 (Gabaix, 1999) gives a worth of 1.005 based on the 135 largest metro communities in the United States as recorded in the Statistical Abstract of the United States during 1991. Despite the momentous matches established for Zipf's rule using city size information from the United States, the problem does not change.

Table 1 Literature review of rank-size distribution

Writer (Year)	**Data limit**	**Sample**	**Finding**
(Jadwiga, 1992)	1961-1981	Census-based towns population.	The RSD didn't attain any evidence to ZL, whereas the PE highly move away from coefficient one over the years.
(D'Costa, 1994)	1901–1981	Urban centres-based population.	The RSD of urban centres indicates towards high violation of ZL as the dist. value deconcentrated lower than one all over the years excluding 1961 and 1974 because for those years the coefficient value was more than one.
(Soo, 2005)	1991	Urban areas population and agglomerations.	The RSD of urban areas population does not violate the ZL rules for cities confirmed but the dist. of agglomerations does where the value of coefficient lies below the one.
(Nishiyama et al., 2008)	1991	Agglomerations of urban areas	The RSD of agglomeration of urban areas was not distributed evenly, where the value of PE coefficient lies above the one.
(Jiang et al., 2015)	1992–2010	Natural cities derived based on nightlight data with areas greater than 10 sq. km	The rank-size distribution of natural cities showed an uneven distribution, with a Pareto figure greater than 1 for all three years.

When we look at two recent studies that are extremely significant. (Duranton, 2002), for example, compares an intriguing model's findings, which employs that the empirical dist. of the quality, range violated the ZL rule for the France and US. A system created for the citizens is suggested by (Rossi-Hansberg & Wright, 2003). Firstly, the stimulus findings of the fascinating novel model (Duranton, 2002) for example, that employs quality hierarchies with practical distributions for the United States and France and elaborate deviation from at both sides of the allocation. Further, a system of cities inspired model is developed by (Rossi-Hansberg & Wright, 2003) which hinted at the Zipf’s law (ZL) in certain instances and describes deviations from this law in the case of both ends of the distribution.

Secondly a study conducted by (Black & Henderson, 2003) utilizing twentieth-century city size distributions data of the US to evaluate the performance of Zipf's law. Based on considering the regression which states about the metro area logarithms for the rank of cities vs the size of the cities, they dispute against the Zipf's law (ZL). When entire cities were utilized and just the top one-third of the size distribution was considered the Zipf coefficient decreased in 1900 and climbed in 1990. However, on the flip side work using panel data, Gibrat's law was firstly used and it examine the reliability of Zipf's law by (Ioannides & Overman, 2003) and also represent the initial effort to test the validity of Zipf's law using Gibrat's law.

The notion that big cities expand more rapidly than smaller cities in different times of periods (Bairoch et al., 1988), although smaller cities expand quicker in others. It shows that Gibrat's law would be implacable only in the long term. (Giesen et al.,

2010) reviewed the numerous methodologies employed and found persistent images. Even though the lognormal distribution performs decent work of modelling the verifiable city size distribution over all communities in a country, but "double Pareto lognormal" distribution performs a superior job, even when additional parameters must be estimated. In beginning, they propose that the imprecise Gibrat's process is established may better represent urban expansion across all cities. Even while this data is indirect because we are comparing theoretical steady-state distributions rather than growth processes, it is reliable with some fresh work that likewise shows that the functioning of pure Gibrat's rule is not up to the mark when a wide range of settlements are incorporated. Second, their discoveries might assist with settling the new contention over city size circulations between (Eeckhout, 2004) and others.

In this regard deterministic metropolitan model equipped for showing a consistent state (Black & Henderson, 1999) have notable exceptions. All urban communities grow at a similar rate. The prerequisites for the outcome, on the other hand, are undesirable. As a result, uncommon practical structures for inclinations and advancements are required. (Henderson, 2002) and (Dobkins & Ioannides, 2000), work on panel data yet there were challenges with the meaning of regions in the US that changes over the long haul and made it extremely difficult to depend on panel data.

It would be fascinating to investigate (Reed, 2002) the idea is more still more statistical. which economic forces can give rise to the random city building mechanism? One may, for instance, attempt to extend the Eeckhout model to consider an endogenous number of destinations by coordinating. Some recent publications have begun to investigate comparable concerns, albeit in a somewhat different context.

Regardless of their broad observational discoveries, Rosen and Resnick's review closed with a supplication. They guaranteed that the experimental examinations required one key part: a thorough hypothetical model clarifying city size dispersion. Consistent with their monetary foundation, they implied an instinctively engaging causal story that lays on a proverbial establishment when they said "theory." Carroll's literature evaluation also focused on the issue of theoretical explanation. Following a review of the empirical literature to determine whether any clear theoretical perspectives were suggested. According to him, at this point, we don't need new models, but rather some basis for ruling out several existing ones (Carroll, 1982).

An alternate model for city size information (Sarabia & Prieto, 2009) portrayed the Pareto-positive stable conveyance. This conveyance is portraying city size information in the nation and gives an adaptable model to the best attack of a whole scope of the informational index where the informational collection could have zero and the old-style Pareto and Zipf dispersions are additionally included. There exist articulations for the shape, minutes, and other enlightening probabilistic measurements. Assessment approaches are investigated, and an essential graphical technique for deciding the information's reasonableness for demonstrating is given. In this review, they utilized the information of Spain for various years. Proposed clear model for best fitting the entire scope of an informational collection which is known as Pareto-positive stable (PPS).

The advancement of city size appropriations has drawn in supported by the interest of specialists over an extensive stretch. The presence of extremely huge urban communities, the exceptionally wide scattering in city estimates, the momentous soundness of the progressive system between urban communities over many years or even hundreds of years, and the job of urbanization in monetary advancement are generally especially intriguing subjective components of metropolitan design around the world. Another astonishing consistency, Zipf's law for urban communities, has itself drawn in extensive interest from scientists. In this manner, to consider the metropolitan development of various economies through the determination of specific, this phenomenon is enticing for examples in case of the size of the city's conveyance around the world. It's of exceptional preference for a hypothesis to foresee Zipf's law (ZL) and other experimentally significant elements.

The size conveyance of urban communities in the United States (Krugman, 1996) is startlingly very much depicted by an easier force law: the quantity of urban areas whose populace surpasses S is relative to 1/S. This straightforward consistency is bewildering; considerably more perplexing is the way that it has stayed valid for basically the previous century. Standard models of metropolitan frameworks don't clarify the force law. An arbitrary development model proposed by Herbert Simon 40 years prior is the best attempt to date, yet while it can clarify a force law, it can't recreate one with the right type. Now, we are in the baffling situation of having a hitting experimental consistency with no decent hypothesis to represent it.

The creator gives the feeling that in some decade enormous urban communities (Eaton & Eckstein, 1997) develop than little in some decade, small urban areas develop quicker than the huge one. The creator recommends that the law of Gibrat's for normal exist in since quite a while ago works yet till now nobody has utilized that information which merits more consideration. This review is a correlation to the powerful advancement of urban communities in France and Japan. These two nations have kept public boundaries and have a metropolitan framework with a few urban communities staying consistent. It is additionally seen that the dispersion over the long haul alludes to approach development in various years and they affirmed it by utilizing a few procedures, for example, Lorenz bends, Zipf's regression (log of Rank with log of size). Besides, the issue of city size conveyance is at the core of (Eaton & Eckstein, 1997) study, which utilized exact change lattices to explore a similar inquiry: how frequently have city size appropriations developed over significant stretches of history? Utilizing nonparametric information on France from 1876 to 1990 and Japan from 1925 to 1985, they found proof of "parallel" extension, which implies that city size conveyance in those countries remained almost steady through time.

In a study researcher gather information about the city size dispersion and the elements of metropolitan development (Ioannides & Overman, 2003). They have analysed Zipf's law and the significance of metropolitan power. They initially centred around pragmatic proof on the upper tail of urban areas dispersion then they examine the speculations that were progressed to clarify the inexact dependability of the dissemination among monetary and social frameworks improving the uncovered bone and creating financial hypotheses. They also discussed growth, economic explanations of city size distributions, consequences of shocks in US urban evolution. The evolution of city size distribution was sustained over a significant stretch and the presence of enormous urban communities, the wide scattering in city size. The hierarchy among the cities over decades and the urbanization role in development are the features of urban structure globally. This is especially the interest in theory to foresee Zipf's law.

Researcher inspected information for the time of 1950-1990 from the biggest 142 Chinese urban areas (Luckstead & Devadoss, 2014) provides findings of their study concluded that the size of the city's conveyance was lognormal at that period, yet that it's shifting towards the rule of Zipf Law (however not by and large ZL) somewhere within the range that lies between 2000-2010. Furthermore, from 1950 to 2010, Luckstead and Devadoss inspected the size dissemination of Indian urban communities. The findings of their research conclude that for the period that lies between 1950-1980, the conveyance was lognormal, however, circulation was Pareto Exponent (PE) in the evaluation period of 1990-2010. Whereas, after some time, it was experienced by India that by industrialization there was quick financial development, which is the reason for the inescapable relocation of labours between the country and metropolitan regions.

According to a study which examine the position size rule's importance (Ezzahid & ElHamdani, 2015) in the Moroccan metropolitan framework. The point is to exactly examine the chain of command and describe the conveyance examples of urban communities dependent on their size. The data came from the censuses of1982, 1994, and 2004. To limit the data, three thresholds are considered: 5000, 50000, and 100000 residents. The example is assessed utilizing the ordinary least squares (OLS) approach. For cities with more than 100,000 inhabitants, using the OLS method without the Gabaix-Ibragimov correction (GIC) provides evidence of Zipf's law's validity. When the GIC is utilized, it appears that the Zipfian distribution (exponent = -1) is also applicable for cities with populations greater than 50,000. It appears that intermediate cities developed faster than other cities between 1982 and 2004. This could bring about a more adjusted dispersion of Morocco's metropolitan framework.

Zipf's law and tried position size law on US urban communities (Kosmopoulou et al., 2007) reconsider over different periods and city limits. We exhibit the existence of Zipf’s Law (ZL) for more intently metropolitan regions in 1900 and, all the more as of late, for metropolitan regions in 1990 and 2000. With the development of the contemporary city, changes in the foundation, and the expense of movement that have added to a metropolitan spread, the metropolitan region might prefer to address the present networks over metropolitan spots completed 100 years prior. In that sense, the position size rule stays substantial given the right redefinition of a city (Cordoba 2000).

In the case of Pakistan (Arshad et al., 2018) researched at country level and provincial level to explore the existence of Zipf’s law. The process of collection of the data was used from different population censuses of 1951, 1961, 1972, 1981, and 1998 on cities where the settlement has a local administrative government. Like (Gabaix & Ibragimov, 2011) where they use Monte Carlo simulation and estimate OLS regression on dependent variable Rank and independent variable Population. OLS has a problem that standard error would be underestimated which has maximum possibilities to reject Zipf law based on t-test. At the national level, city size distribution does not hold but it holds only at the provincial level if the urban system relay on consistent property. The system of urbanization is coherently based on languages and their culture at the level of the province.

Further, a few examinations have taken a gander at Zipf's law as far as the complete metropolitan populaces of areas or states inside a solitary nation, just as the general populace of the world's nations. A couple of late investigations have investigated Zipf's law for city size conveyance inside an area, state, or some other sub-local grouping at the sub-public level. (Giesen & Südekum, 2011) for instance, led a critical report wherein they broke down the size circulation of all German urban communities just as the size dispersion of urban communities inside German districts.

The current examination on the exact approval of Zipf’s Law (ZL) for the size of cities dispersion uncovers clashing and questioned results. One way of thinking declined that the size conveyance of urban communities complies with Zipf's standard and such Zipf’s Law regarding the size of the city’s appropriation is widespread. (Josic & Bašić, 2018) in their investigation they discover that Zipf’s law (ZL) holds for greater part settlements, however, it doesn't hold for minuscule and amazingly enormous settlements. As indicated by them, Zipf's law holds just for metropolitan agglomeration, and it is disregarded starting with one country then onto the next country and now and again. (Cheshire, 1999) and (Soo, 2005) contended that useful metropolitan districts can all the more precisely portray a metropolitan framework, contrasted with the city/metropolitan limits, and, subsequently, a more legitimate trial of the Zipf’s Law, they can offer. (Fujita et al., 1999) expressed "the consistency of the metropolitan size dissemination represents a genuine riddle, one that neither our methodology nor the most conceivable elective way to deal with city sizes appears to reply." The rank-size rule, as per (Fujita et al., 1999) in any case, does resemble the since a long time ago runs spatial patterns of a developed spatial system. A classical experimental paper wherein cross-section information of countries was used. They track down that the Pareto Exponent contrast among different countries, going from decreasing to increasing way.

Likewise a study concentrated on worldwide urban areas like New York, London, and Tokyo (Sassen, 1991) present their findings, these urban communities go past their nation's boundaries, so it became hard to decide Zipf's law for such sorts of urban areas. (Yang, 2011) evaluated that, there is too much literature in the field of remote sensing for extracting the cities or equally populated settlements from satellite imagery. Furthermore, (Weber, 2003) also studied remote detecting symbolism gives an incredible way to characterize urban areas as far as degrees and areas, yet there is no assurance for removing urban areas by symbolism. There is also a problem in the definition of a city that the cities are the product of census and their literal meaning states about the estimation of the population for the purpose of taxation. Cities are referred to as urbanized areas, or metropolitan-based on some population said by US Census Bureau. (Jiang & Liu, 2012) recommended that the "Regular urban areas" as an option in contrast to the traditional meaning of urban communities dependent on the population, but still, something is missing in that imagery method so that we will fill the gap of this problem and addresses the problem of this law holding in larger cities

On a discussion on city size dispersion for US metropolitan districts and metro regions (Berry & Okulicz-Kozaryn, 2012) stated about the hypothetical support of the rank-size rule. They have concentrated on the metropolitan development that whether metropolitan development complies with Gibrat's law and finding that whether the Pareto example and log-typical conveyance catches Zipf's law or not. In their study, they defined the Economic areas same as the US Department of Commerce under the Bureau of Economic Analysis as units of observation and mainly focus on the size circulation is Pareto in the upper tail which means that these areas must be metropolitan in scale or largest urban areas. Furthermore, when urban areas are poorly defined then city size distribution has conflicts and urban growth does not obey the law of Gibrat's and size dist. of cities is not firmly rank size of Zipf's Law.

The law of Zipf for being universal condemns by the study of (Arshad et al., 2018) which conducts a comprehensive evaluation of the available research and application for Zipf's Law to size dist. of cities. Some practical proves from their study reveal that for a region having upper-tail, Zipf's law isn't constantly taken note of. However, the process of the selection of samples is not unbiased, the framework of methodology shortcomings and limitations over the data may cause conflicts. Zipf's Law speculation may be bound and be dismissed in the complete size of cities dispersion, whereas different appropriations have been presented in this instance. On the contrary, in case of better techniques of the empirical process applying where the urban areas are identified properly, where hypothesis seems to be more likely to be approved. The discussion is still ongoing. The discussion is still a long way from being settled. We additionally recognize four new spaces of exploration for Zipf's Law and size of cities appropriation, considering circulation of the size of the tail of lower urban areas, dispersion's size for urban areas in sub-public locales, elective types of Zipf's Law, and the connection among the Zipf's Law and the metropolitan framework's lucidness property.

Some issues regarding Zipf's Law on the size of cities (Jiang et al., 2015) studied to see whether and why the law holds. They studied more about whether the law holds instead of why this law holds in the global setting, they have adopted the natural cities instead of relying on the census data. Their primary contention is Zipf's law is general, and they applied this law to city numbers taken through Google night-time imagery to collect data and see that the number of metropolitan regions in the position 1 country is twice more than the second-greatest country, triple simply that various in the third greatest country, and so on (Jiang et al., 2015). The sample size of each country will choose the virtue of rank size. Furthermore, each sample uses a defined sample size or fixed number of cities or threshold level of the population (Rosen & Resnick, 1980).

Urban development and growth are complex in the case of developing countries with constraints impose on financial resources. (Knudsen, 2001) studied that for most of the countries in the world the size dispersion of urban communities should fit 'power law': the quantity of populace more noteworthy than (S) is contrarily corresponding to (1/S). according to the findings, the Danish case does not refute Zipf's law, and 14,541 firms were taken for the purpose of checking the holding of the Law, and the law holds which states that there is clear rank-size distribution with Zipf's Law. This study concluded that given the success of this law in the accounting of cities growth it is not possible but also credible that growth may follow the updated version of it.

Some researcher argue that the greater part of the contention on the legitimacy of Zipf's Law (Malevergne et al., 2011) studied that the legitimacy is lost and propose that "a portion of the discussion on Zipf's law ought to be projected as far as how well, or inadequately, it fits, as opposed to if it tends to be dismissed." (Ioannides & Overman, 2003) planned and executed a methodology for working out neighborhood Zipf types for the US city size dispersion and testing the legitimacy of Zipf's Law for urban areas. There are two significant disclosures. To start with, Gibrat's Law applies to city development measures overall. Second, Zipf's Law generally holds for a wide assortment of the size of cities. Notwithstanding, our discoveries infer that the Zipf example's nearby qualities can shift essentially relying upon the size of the city. These progressions in the nearby Zipf example can be clarified, as indicated by Gabaix, by taking a gander at the mean development rate and changes in development rates dependent on city size. Moreover, our assessments of neighbourhood Zipf examples help us in appreciating some notable marvels. Moreover, their appraisals of neighbourhood Zipf types help us in fathoming a few all-around reported characteristics of the United States city size circulation.

At the point when the likelihood of estimating a specific worth of some amount differs contrarily as a force of that worth, the amount is said to adhere to a Power Law, additionally referred to differently as Zipf's Law or the Pareto appropriation. This law is utilized in the field of Physics, Economics, Finance, demography, and sociologies. For instance, the distribution of the size of cities, earthquakes, forests, and people's fortunes are also seen by Power law. Here we review some of the empirical evidence for power law and theories to explain. (Newman, 2005) said that measuring or identifying the power law has different ways like by plotting on logarithms scale the straight line appears in Histogram that straight line in the form of the power law. However, the plot is not a very good way to see power law. Another way to measure is to see the data which is so large causes a noisy curve than there is a simple solution to put the data into the tail of the curve but if that data is useful information where sometimes the Power law holds only in the tail then we will be in danger. Alternatively, we can shift the width of the canisters in the Histogram. The best choice is to design containers so that every bin is multiple more widely than the previous one. (i.e., bin sizes will be 0.1, 0.2, 0.4, and so forth). This will reduce the statistical errors in the tail effectively.

As indicated by (Gabaix, 1999), the size circulation of urban areas in many countries adheres to a force law." the extent of urban communities with populaces bigger than S is relative to 1/S." He exhibited that urban communities with an upper tail follow an observationally demonstrated development design, making the dissemination unite to Zipf's law. He accumulated information from the metropolitan regions of America which was recorded in the time of 1991. We notice a straight line, which is uncommon (there is no redundancy making the information produce consequently a straight line). Moreover, we find that its slant is (-1). We might run the relapse, Zipf's law, into a significantly more justifiable example, Gibrat's law. Then again, the strength of these laws gives an exhortation to city development scholars: city development models should supply Gibrat's standard in the upper tail.

As revealed by (Cristelli et al., 2012) they blended and disagreeable experimental information is inferable from the fundamental meaning of the things shaping the framework about the application of Zipf's Law. However, by their research, they argued saying that the city framework doesn't display genuine Power-law conduct since May 2012 it's deficient or conflicting within those conditions with whom the PL are relied upon to embrace. Force rules must apply for the gathering of urban areas that shared principles and are financially associated and have co-advanced over the long run. The gathering of urban communities, which has generally advanced as one, has met on a natural monetary unit. Thus, the size dispersion of urban areas for the gathering turns out to be inside reliable and submits to measurable components of force laws. In such a manner, some new examinations have zeroed in on financial factors, contending that Zipf's standard is the result of a powerful interaction between monetary movement and the development cycle of urban areas. (Córdoba, 2008), (Duranton, 2006, 2007), (Hsu, 2012), (Lee & Li, 2013).

There is a requirement on permissible models of nearby development (Gabaix, 1999) inspects which expresses that Power law is a shocking size appropriation of urban communities for most of the nations. The quantity of urban areas with a populace of urban areas is more noteworthy than S is relative to 1/S, in the upper tail, all urbanized areas follow the corresponding development cycle and prompt the dispersion towards Zipf's law. (Feenberg & Poterba, 1993) the distinction between the Pareto law of pay dissemination and Zipf's law for urban areas is that the type for money conveyance appears to change from one year to another and across.

Furthermore, according to the base literature, the exponent of power-law was between 0 and 2 while (Jiang & Yin, 2014) took the exponent value is 1. On the other side, (Gabaix, 1999), (Li, 2002), (Mitzenmacher, 2004), (Newman, 2005) described that Zipf's Law held approximately 100 years. Some researchers have doubtful thoughts regarding this like (Soo, 2005) and (Gabaix, 1999) those who were clear about Zipf's Law sometimes accept that it is valid in the case of larger urban areas. Literature also provides a contradictory picture about Zipf's Law. Another study examined that there are more industries in larger cities which allows them to diversify the shocks and smaller cities have small variances. Furthermore, the existence of power-law is because of "scale variance" as all scales have equal growth process, but the final distribution process should be scale-invariant to follow a PL (Gabaix, 1999).

Table. 2 provides the previous studies about the Zipf's Law in which we have gathered the literature on where the Zipf's Law (ZL) holds or not. A few impressions don't keep Power-law like exceptionally slanted circulation and assessment that took on somewhere else (Moura Jr & Ribeiro, 2006) concentrated on Brazilian urban communities with more than 30,000 occupants. They showed that Pareto circulation was not legitimate for more modest urban communities. For these cases, the city size combined dissemination work didn't adhere to a force law conduct. The coefficient ˜α values were determined with three strategies: greatest probability assessor, least-squares fitting, and normal boundary assessor, and where more precise outcomes are given by the most extreme probability assessor.

Table 2 Previous Literature regarding Zipf's Law

Major Outcomes	Researchers	Country	Sample
Size dist. of cities exactly confirms to ZL.	(Giesen & Südekum, 2011)	Germany	Cities with more than 100,000 inhabitants.
	(Gligor & Gligor, 2008)	Romania	265 large- medium cities-based adjustment menus.
	(Ezzahid & ElHamdani, 2015)	Morocco	Cities with more than 50,000 inhabitants.
	(Rastvortseva & Manaeva, 2016)	Russia	
Size dist. of cities does not confirm to ZL.	(Lalanne, 2014)	Canada	152 largest urban areas.
	(Dubé & Polèse, 2016)		135 largest urban areas.

	(Lanaspa et al., 2003)	Spain	
	(Le Gallo & Chasco, 2008)		722 Spanish municipalities.
Size dist. of cities approaches to ZL as countries experience urbanization	(Gangopadhyay & Basu, 2009)	India	Different samples minimum threshold is the cities above 10000 inhabitants in the census year 2001.
	(Gangopadhyay & Basu, 2013)		Cities above 212523 inhabitants in the census year 2011.
	(Luckstead & Devadoss, 2014)		58 largest cities from 1950 to 2010.
	(Moura Jr & Ribeiro, 2006)	Brazil	Cities with 30000 inhabitants or more.
	(Matlaba et al., 2013)		185 largest functionally defined urban areas.
	(Ignazzi, 2015)		Census years data from 1871 to 2010.
	(Luckstead & Devadoss, 2014)		58 largest cities from 1950 to 2010.
	(Moura Jr & Ribeiro, 2006)		Cities with 30000 inhabitants and more.
	(Matlaba et al., 2013)		185 largest functionally defined urban areas.
Size dist. of cities may evolve diverge from ZL over time.	(Soo, 2007)	Malaysia	Cities with more than 10000 inhabitants.
	(Pérez-Campuzano et al., 2015)	Mexico	Cities with more than 15000 inhabitants.
	(Duran & Özkan, 2015)	Turkey	Cities with more than 37522 inhabitants in the year 2012.

Table. 2 provides the previous studies about the Zipf's Law in which we have gathered the literature on where the Zipf's Law (ZL) holds or not. A few impressions don't keep Power-law like exceptionally slanted circulation and assessment that took on somewhere else (Moura Jr & Ribeiro, 2006) concentrated on Brazilian urban communities with more than 30,000 occupants. They showed that Pareto circulation was not legitimate for more modest urban communities. For these cases, the city size combined dissemination work didn't adhere to a force law conduct. The coefficient ˜α values were determined with three strategies: greatest probability assessor, least-squares fitting, and normal boundary assessor, and where more precise outcomes are given by the most extreme probability assessor.

CHAPTER 3

RESEARCH METHODOLOGY

In this chapter, we discuss in detail the data, the abstracted variables for the data, the collection sources of the data, the applied techniques for checking the nature of the data and the methodology that we are supposed to apply to move forward to know more about Zipf's Law.

3.1 Data

Zipf's Law and size distribution of cities have both focused on the definition of a city. This research relies on that definition where cities are defined in way of administratively, as the census of the population is also defined according to that manner, which states that adjustment is deemed an urbanized area if there is a local administrative administration. Recent works have demonstrated that the Zipf's Law (ZL) application is quite sensitive by the definition of cities [Eeckhout (2004), (Berry & Okulicz-Kozaryn, 2012), (Nitsch, 2005), (Cheshire, 1999), (Jiang & Jia, 2011), (Jiang & Liu, 2012), (Jiang et al., 2015), (Rosen & Resnick, 1980), (Rozenfeld et al., 2011), (Schmidheiny & Suedekum, 2015), (Veneri, 2016)]. In general, cities definitions consist of three forms which are discussed in previous studies: (Eeckhout, 2004) takes the definition of cities according to administrative way, [(Berry & Okulicz-Kozaryn, 2012), (Rozenfeld et al., 2011), (Schmidheiny & Suedekum, 2015), (Veneri, 2016)], functionally defined cities and natural cities [(Jiang & Jia, 2011), (Jiang & Liu, 2012), (Jiang et al., 2015)].

However, in case the applied definition of the cities is valid, still, official statistics prefer the limitations of the city definition assisted by the authorities of statistical. Such definitions can be or can't correspond to an economic-based significant definition of "city" (see (Rosen & Resnick, 1980) or (Cheshire, 1999)). Data set for the agglomerations may be closer to a functional definition because it often contains the adjacent suburbs where city working people live. For collecting the best set of data, we cross-checked that set within the official numbers provided by statistical agencies of many countries, to relieve doubts about the dependability of internet data. The data in each case corresponded to one or more of these sources.

In our Research the data set is updated and obtained from the following website: *http://www.citypopulation.de*. This site includes data of all cities in each country of the world. The total number of countries is 235 listing all cities with their respective census. The five most populous countries are China, India, followed by the United States, the island nation of Indonesia, and Pakistan. Among the smallest countries in the world in terms of population are the island nations in the Caribbean and the Southern Pacific Ocean (Oceania). The five countries in the world with the smallest population are: Vatican City, an enclave in the city of Rome in Italy, Tuvalu, an island country in the Polynesian part of the Pacific Ocean, Nauru, a tiny island country in Micronesia in the southwestern Pacific Ocean, Palau, a Micronesian group of islands in the western Pacific Ocean, and San Marino, a small country that forms an enclave in Italy. We have considered all countries to check the validity of Zipf's law. For each country, this site provides us with the data for two to four sometimes five, but it depends on census years. The earliest is in the 1980s and the latest is in the 2010s and some of the countries arranged their census in 2019.

3.2 Variables

Every country's data is available for administratively defined cities. We are using panel data comprised of population data with their ranks being the largest in the population. The dependent variable is the natural logarithm (ln) of the Rank minus half $(\text{Rank} - 0.5)$ and the independent variable is the natural log of Population (Pop) of every single city. Regression was applied on different sizes of the independent variable on the Population (Pop). The coefficient of the population as shown in Eq. 5 shows the validity of Zipf's law (ZL) and about the system of cities and the dispersion of population in different cities. We obtained the β values for each census of every country. For our ease, we have taken the average of the coefficient to make it clear and easy to interpret.

Rank is a sequential variable the size is measured by the population of cities. The data set will be compiled in a country-by-country style by the grouping on the base of Human Development Index (HDI) in three categories i.e., Developed as High-Income Countries (HIC), Developing as Medium Income Countries (MIC) and Under-develop as Low-Income Countries (LIC). We will comprise "Panel data" from cities populations throughout various census periods as variables. We have gathered the population of each city inside each country, as well as their ranking because of their population (Largest population gets Rank 1). In the long run, the results will aid in the development of national planning frameworks to achieve the goal of sustainable and equitable economic growth for all cities in all countries throughout the world. The standards for the lowered population about the urban area which is supposed to be taken in the sample varies by country—on average, in the case of larger sized countries, there are higher thresholds, however with the larger number of cities taken for the sample. All the countries picked have minimum criteria of at least 1,000 people.

3.3 Methodology

The size distribution of cities was initially suggested by (Auerbach, 1913). This law was named on linguist George Kingsley Zipf (1949). He stated that "The rank of the city is negatively related to the size of that city with the power value, however that value of power stated as 1. For the 140 large-sized cities of the US in 1990, (Nota & Song, 2012) compare Log of (Size) vs. Log of (Rank). The y-axis has the log values of Rank, whereas the log values of the population are plotted on the x-axis. In their investigation, they evaluated that the exponent is equivalent to 1, stating that the law of Zipf will collapse into the rule of rank-size. According to him, the Pareto exponent is utilized as a proportion of populace consideration among urban communities of different sizes. If Pareto Exponent (PE) got the value equal to one the Rank Size Distribution rule is valid, and because of that, this value of Pareto Exponent becomes an average value to get achieved according to the Rank Size Distribution rule (Jiang & Jia 2011). The PE is the most significant component of Pareto's law, and it is mentioned in virtually every research (Fonseca & Tartar, 1989) and (Cheshire, 1999). When the exponent has a power of one, Pareto law applies. Pareto's law is associated with the position rule of size dist. and Zipf's Law (Rozenfeld et al., 2008). City frameworks that stick to the law of Pareto may likewise observe Zipf's Law appropriation and the position size rule (Batten, 2001); (Chlebus & Ohri, 2005).

Empirically Pareto Exponent distribution is examined through the following equation:

$$\ln(y) = \ln(A) - \beta \ln(x) \qquad \textbf{\textit{… (2)}}$$

Where:

Y = rank of the city (k).

$\varkappa$ = size of the city (k), which most of the time is measured through the number of the population of that city.

A = Intercept which may be known as Pareto Exponent (PE).

β = slope of the curve which indicates the size of the rank.

Zipf's Law is taken solely to indicate about β = 1. If β lies between the range of $0 - 1$, then this demonstrates that the Rank Size Distribution for the cities is not even, so in that situation Zipf's Law violated. If $\beta = 1$ then the Rank Size Distribution for the cities is more even (Reed, 2002).

Pareto Exponent nonlinearity reveals departures from the Pareto distribution. (Soo, 2005) asserts that Zipf's Law is strongly rejected in some nations' agglomerations (Rosen & Resnick, 1980), (Cheshire, 1999), in favour of the alternative that agglomerations are more unequal in size than Zipf's Law would anticipate. Previously, Larger city expansion has primarily taken the shape of suburbanization, and therefore this increase is represented not just in administratively designated cities as it is in growing population concentration in bigger cities when urban agglomerations are utilized.

Unless an exponent amount is less than one, it implies that a rank 2 for that area is much below as compared to the rank 1 area, implying that it is undersized. (Shiode & Batty, 2000b), (Li, 2002) and others investigated the influence of growing suburbanization on the expansion of big cities. By using the estimator techniques of Ordinary Least Square (OLS), if the average Pareto Exponent value is lesser in the case of clustered areas than that for other normal areas because Pareto Exponent is an indicator of how equally spread the population is, and urban agglomerations seem to be larger compared to the city centre for the largest cities than that for smaller cities, and the small sample doesn't create this outcome primarily significantly.

If the exponent value is more than the average, Zipf’s Law implies that second biggest urban area is higher than half. One city and third biggest will be more than a third the size of rank One city and will be enormous. The greater the value of the coefficient, the more evenly spaced the cities are, indicating that the concentration of settlements is large and evenly dispersed. According to (Rosen & Resnick, 1980), an exponent greater than just one implies that people in most areas are more spread equally than would be anticipated by the rank-size criterion.

The number of cities with a population higher than S is proportionate to 1/S, and even in the upper tail, all cities obey the equal growth process, resulting in a distribution that is closer to Zipf's law. It considered following dist. for Pareto:

$$Y = \frac{A}{X^{\beta}}$$

or

$$\mathrm{Y} = \mathrm{A} \div X^{\beta} \qquad \ldots (1)$$

Or

$$Y = \mathrm{A}\ X^{-\beta}$$

X = Rank of Country with population P or more,

Y = Population of city,

Q = Constant

β = Pareto Exponent.

Since (Zipf, 1949), it is already a common practice to test that law by simply plotting the natural logarithm of the rank against both the log of the size and hoping to find a straight line with a negative slope to minus one. In more technical terms, a ‘test' is frequently built by estimating the regression using OLS to see if the estimate is ‘close' to (-1). Because of the right-skewed distribution of city sizes, another way to write this formula is to utilize natural logarithms, which offers a good perception for the sampling spanning tiny urban areas (Gabaix, 1999) (Jiang et al., 2015).

By taking (ln) of equation (1) we will get:

$$\ln(y) = \ln(A) - \beta \ln(x) \qquad \text{.... (2)}$$

Where ($\varkappa$) is the population size and (y) is the number of cities with a population of more than a million people ($\varkappa$). (Zipf, 1949) did contribute that the sizes of population dist. couldn't be characterized as dist. of Pareto, however, it can be described as a new form of that distribution with equal to 1, which means that the (β)corresponds to the largest city's size.

The econometric model of this equation is:

$$\ln(Y) = A - \beta \ln X + \mu \qquad \text{... (3)}$$

When we use the OLS estimator to determine the Pareto coefficient in an economic model, we get downward biased results. The fact that Zipf regression is biased in small samples is a major problem. Utilizing Monte Carlo simulations, (Gabaix &

Ioannides, 2004) represent that perhaps the estimation regression OLS coefficient is skewed decreasing way considering sizes of the samples in the range that is often evaluated for size dist. of cities. Furthermore, OLS standard errors are greatly overestimated, resulting in Zipf's Law rejections.

In past study the (Gabaix & Ibragimov, 2011) provide us with a simple way to improve the OLS estimation of tail exponents by using “Rank – ½” even there are more classy methods that exists. Usually, OLS regression is used to estimate the Pareto exponent (α) but this method is strongly biased in a small sample despite being a simple and robust method. The easy practical solution for this problem of biasness is that if Pareto’s exponent is found by this method, then we will be using Rank $- \frac{1}{2}$(as shown in Eq. 4 below).

The corrected formula is presented below:

$$\ln(Y - \frac{1}{2}) = A - \beta \ln X + \mu \qquad \dots (4)$$

The term $(-\frac{1}{2})$ is sample correction bias to correct the error of downward bias after applying OLS regression. In the current study, we will rely on census data from all cities in each country of the whole world and the purpose is to check Zipf’s law globally. And it also helps to make it optimal and reduces the biasness. This approach over the standard OLS estimation procedures and indicate that it performs well under dependent heavy-tailed processes exhibiting deviations from power laws. Therefore, to correct this issue of biasedness an effective solution is by using or multiplying the $(-\frac{1}{2})$ from the dependent variable which is known as "Sample bias correction".

The present study sets out to do some sort of things: firstly, for checking the Zipf's law (ZL), using a fresh set of data which consist of large sample-sized countries. Secondly, for the analysis purpose run Wald's coefficient restriction test where we will restrict the second coefficient equal to 1. The reason behind this test is to check the value of the coefficient of $\ln$ (Population) to be equal to 1. Inferential statistics are used to estimate the hypothesis. Zipf's Law is estimated through the model OLS on *Eq. (5)* with the sample of two to three census data of each country.

After estimation with Ordinary Least Square method, we used the restriction on coefficient (β) which is the Pareto exponent which tells us about whether Zipf law holds or not. The coefficient will lead us to know about the spread of population in different cities of each country.

$$\ln(\text{Rank} - 0.5) = A - \beta\,(\ln \text{Pop}) \qquad \ldots (5)$$

This research has a design that involves the use of a set of Small-bias correction (SBC), and OLS regression to test whether and to what extent the distribution of urban and total population departs from the Zipf's law, using different sample sizes of the population of each city in a country. The population vary from country to country. We have divided these countries into three main groups is High-Income Countries (HIC's), Medium Income Countries (MIC's) and Lower Income Countries (LIC's) according to the Human Development Index.

In this study, we will compare the results of all cities of each country to check if they follow Zipf's law or not? and about their distribution or dispersion. The coefficient (β) also known as the Pareto exponent used for the estimation of the concentration between cities of different sizes. The Zipf's law or rank-size rule holds when the value of coefficient $\beta = 1$ the reason of this is that the value of β move around the average and if the value is near to 1 or equal to 1 then it will be statistically significant. The Independent variable will significantly explain the dependent variable (DV). Smaller the value of ($\beta < 1$) Pareto exponent (β) shows more inequality among cities and with more hierarchy and depict a highly concentrated urban system. The government focused on a specific city, Migration happens and people moving in large numbers to a specific city will cause a decrease in their Pareto exponent. Higher Pareto exponent (β) suggest that City-size distribution (CSD) became more equal or even, and the difference among the size of small and large cities has decreased.

Alternatively, (Gan et al., 2006) described that there will be less hierarchy and more equality when the value of the Zipf coefficient is higher than one. The relation of the size of rank for the cities is linear in logarithm, where the value of R2 is high approximately near to One. The value of R2 shows the goodness of fit of the data. If the value of R2 is (0.93) then it implies that the Independent Variable is explaining 93% of the Dependent Variable.

CHAPTER 4

RESULTS AND DISCUSSIONS

This chapter consists of a detailed discussion of all the applied result and their findings. Specifically, initially, there is the discussion about the Dynamics of countries which were divided into three parts, then descriptive statistics and correlation matrix results which help us to understand the nature of the data and the relationship among the variables. Then we discuss the result of those estimation techniques and tests that are carried out to attain the fulfilment of the objectives of the study, also relation with previous finding if any of the previous study and investigate the Zipf's Law.

4.1 Results & Discussion

This section demonstrates the findings of the study. We used Pareto Exponent (β) to check whether Zipf law holds or not for all countries. Our results of the study are too big, we could not display the whole picture in this study due to space constraints. We discuss only the results for the significant (β) for the countries that lies in the three main groups of the Human Development Index (HDI), from the regressions of Zipf's Law and the OLS estimator. This is to reduce the size of the tables. Full details are available from the author upon request.

Here, we explain the average of the different census year's corresponding to those countries, for the estimation techniques for Zipf's Laws OLS and the Wald Coefficient restriction test. This procedure is utilized for size reduction of the tables and result section instead of discussing and including all data in table form. The whole data include four to five census years for each country in the world. It is also difficult to show all 235 countries graphically here. For ease we have divided this section into three categories which include the graphs and discussion on the average of all β results for a country, each country had three or more β values. Countries with less estimated Pareto Exponent values are more hierarchical and more concentrated, lesser the value will show more of the hierarchies and concentration and there is more equality in dist. of population. Even dist. describes that there is more attention by population in small or medium-sized cities. Moreover, the values of data have less spread or less dispersion around the mean value of (β). On the other hand, countries with a high β value or greater than 1 will show less hierarchy and attention of population and their values are more dispersed from the mean value. A higher value than β suggests there is a large portion of people in big cities as compared to small and medium-sized cities.

Countries are divided based on Human Development Index (HDI) which includes developed categories with very high-income including 65 countries in this group, developing countries which includes the countries having income near to very high comprising of 90 countries in this group and those which are near low income and lastly the underdeveloped countries with low income have 33 countries in their list. A list of countries without estimation is 47, the reason behind this is because, either they comprise of one or two census data, or they comprise of less than 20 cities. While collecting data of these countries we have taken all available data. Statistical regularity of Zipf's law is that it is only be applicable on huge data but not on small data and in Appendix Table VI include all those countries having less data.

4.1.1 Dynamics of the Cities

To present the worldwide application of Zipf's law we have separated the countries into three main categories namely Developed countries, Developing countries and under-developing countries. The purpose of categorizing the worldwide cities in this manner is to investigate the objective of checking out how much effectively Zipf's hold in these categorized countries by comparing the outcomes of the study. There were 235 countries, and we have the OLS estimation for all of them, this topic consists, of three categories each one of them has one graph for the country which obtained the value greater or maximum, two of them shows the value near to one or statistically significant and other one represents the minimum or lower value. This is because we have too big data and results, and it is difficult to add all of them**.**

4.1.1.1 Higher β Value in the HIC group.

Graph for that country obtained the value greater or maximum value of β in estimation by OLS estimate.

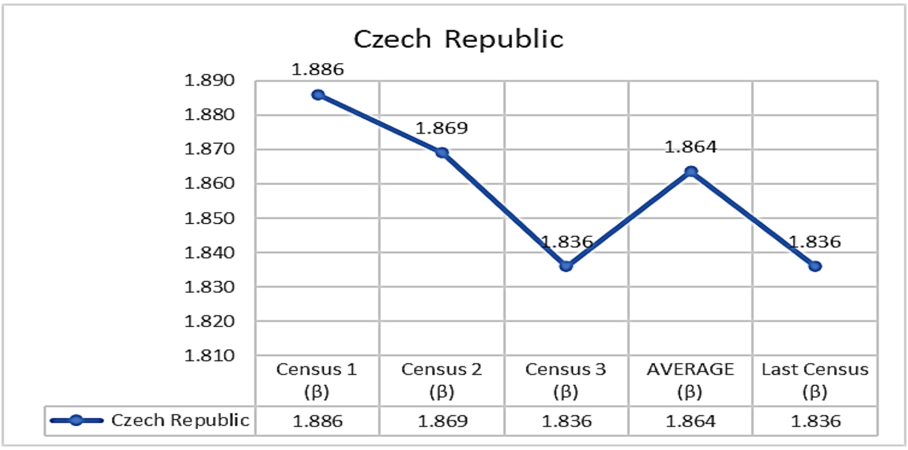


Figure 1 *Maximum Value of β in HIC group*

In the case of Czech Republic, (Figure 1). from High Income Countries (HIC) where the all censuses value of Pareto Exponent (β) value is maximum or greater than 1 as compared to other countries in this group. We have found that in case of the Czech Republic, three censuses took place. According to data of censuses depicted in *Figure 1*, we found the decreasing trend of the variations of the population. The values of β are 1.886, 1.869, and 1.836 for censes 1 to 3 respectively with the average of 1.864. The estimated value of Pareto exponent decreased from its highest value (1.886) in census 1st to its lowest value (1.836) in 3rd census. Data shows that the Pareto exponent's (β) has a decreasing trend over time shows that City Size Distribution (CSD) is more unevenness overtime. So, our result demonstrated that Czech Republic's hypothesis show that Pareto exponent is much greater than 1, which is rejected for all census years shows more even city size. Together, these results show that Zipf's law does not hold for Czech Republic.

4.1.1.2 Significant β Values in the HIC group.

We took the example of two countries i.e., Italy and France from the higher income group where the value to Pareto exponent is near or equal to 1 or significant value of β in estimation by OLS estimate. The below *Table 3*. states the summary statistics for the developed countries where the Pareto Exponent (β) value is near to 1 by indicating the slight evenness in the population dispersion. Also depicted the estimation results of the two countries namely Italy and France which are treated as a sample to represent the overall results of Zipf's law holding in developed countries where the exponent value is near to 1. In the case of Italy, there are four censuses of the population which took place by the gap of years as 1991, 2001, 2011 and in 2020.

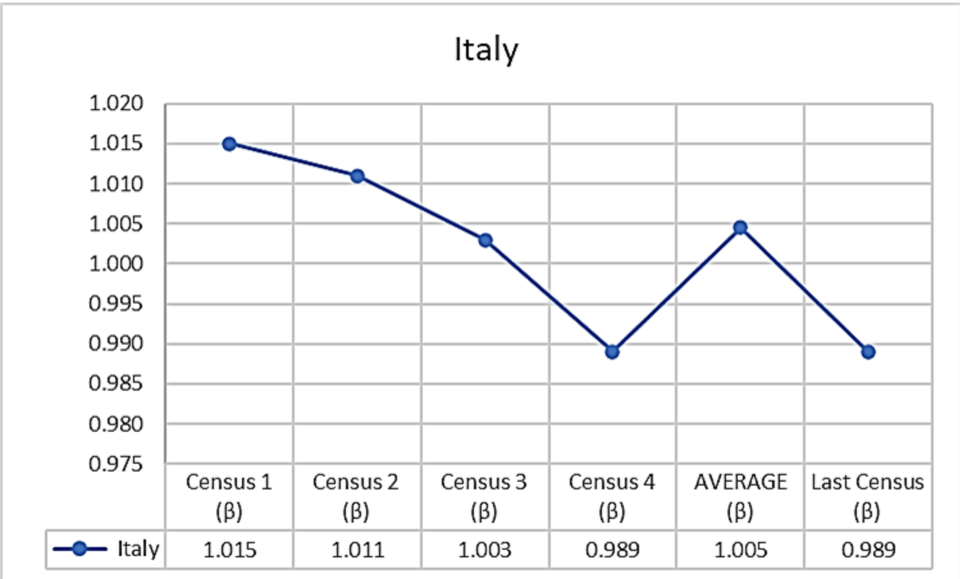


Figure 2 Significant Value of β in HIC group

The above graph *Figure. 2* presented the case of developed countries of the world. We took the Italy as a sample from HIC where the value of PE (β) value is near or equal to 1. We have found that in case of the Italy four censuses took place. According to data of censuses depicted in *Figure 2*, we found the mixed trend of the variations of the population. The values of β are 1.015, 1.011, 1.003 and 0.989 for censes 1 to 4 respectively. there is more evenness of the population of cities by the average of 1.005 and the overall trend of this evenness of the population is decreasing in all the census. So, Our result demonstrated that Zipf's law hold appropriately in case Italy as we found the mixed trend of the variations of the population near to the 1.

On average, the value of Italy, the R-Squares indicates about the 94% variations in the population due to cities Rank Size Distribution (RSD). Where the value of exponent moved slightly closer to 1 in all four censuses as in first three census lies slightly above the 1 but in case of last recent of 2020, it lies slightly below the Pareto Exponent (β) value.

According to probability value, the null hypothesis of the research was rejected in all four censuses as the P-value is less than 0.05 for the coefficient of population. So it is concluded that in Italy there are no proper perceptions about the appropriate holding of Zipf's Law for Rank Size Distribution of cities.

On the other hand, In the case of France, *(Figure 3)*. from HIC where the all censuses value of Pareto Exponent (β) value is near or equal to 1. We have found that in case of the France four censuses took place.

Table 3 Value of HIC's significant to 1

COUNTRY (DEVELOPED)	YEAR	**ln (Rank - 0.5) = A - α (ln Pop)** Wald test (t, F, Chi-square): P<0.05
Italy	1991	$\ln(R - \frac{1}{2}) = 7.460 - 1.014 \ln(Pop)$ (0.127) (0.022) $R^2 = 0.942$
Italy	**2001**	$\ln(R - \frac{1}{2}) = 7.443 - 1.013 \ln(Pop)$ (0.127) (0.022) $R^2 = 0.942$
Italy	**2011**	$\ln(R - \frac{1}{2}) = 7.412 - 1.003 \ln(Pop)$ (0.129) (0.022) $R^2 = 0.940$
Italy	**2020**	$\ln(R - \frac{1}{2}) = 7.329 - 0.989 \ln(Pop)$ (0.129) (0.022) $R^2 = 0.938$
France	**1990**	$\ln(R - \frac{1}{2}) = 6.404 - 0.933 \ln(Pop)$ (0.007) (0.001) $R^2 = 0.990$
France	**1999**	$\ln(R - \frac{1}{2}) = 6.501 - 0.953 \ln(Pop)$ (0.006) (0.001) $R^2 = 0.993$
France	**2007**	$\ln(R - \frac{1}{2}) = 6.623 - 0.978 \ln(Pop)$ (0.005) (0.001) $R^2 = 0.994$
France	**2018**	$\ln(R - \frac{1}{2}) = 6.707 - 0.993 \ln(Pop)$ (0.004) (0.001) $R^2 = 0.995$

According to data of censuses depicted in Figure 3, we found the mixed trend of the variations of the population. The values of β are 0.934, 0.954, 0.978 and 0.994 for censes 1 to 4 respectively. there is evenness of the population or cities size distribution by the average of 0.965 and the overall trend of this evenness of the population is increasing. So, our result demonstrated that Zipf's law hold appropriately in case France as we found the increasing trend of the variations of the population near to the 1.

On average, the value of R-Squares of France indicates about the 93% variations in the population due to cities Rank Size Distribution (RSD). Where the value of exponent moved slightly closer to 1 in all four censuses as in first three census lies slightly above the 1 but in case of last recent of 2020, it lies slightly below the Pareto Exponent (β) value. According to probability value, the null hypothesis of the research was rejected in all four censuses as the P-value is less than 0.05 for the coefficient of population. So, Our result demonstrated that the France hypothesis show that Pareto exponent equals to 1 is rejected for all census years. Together, these results show that Zipf's law holds for France.

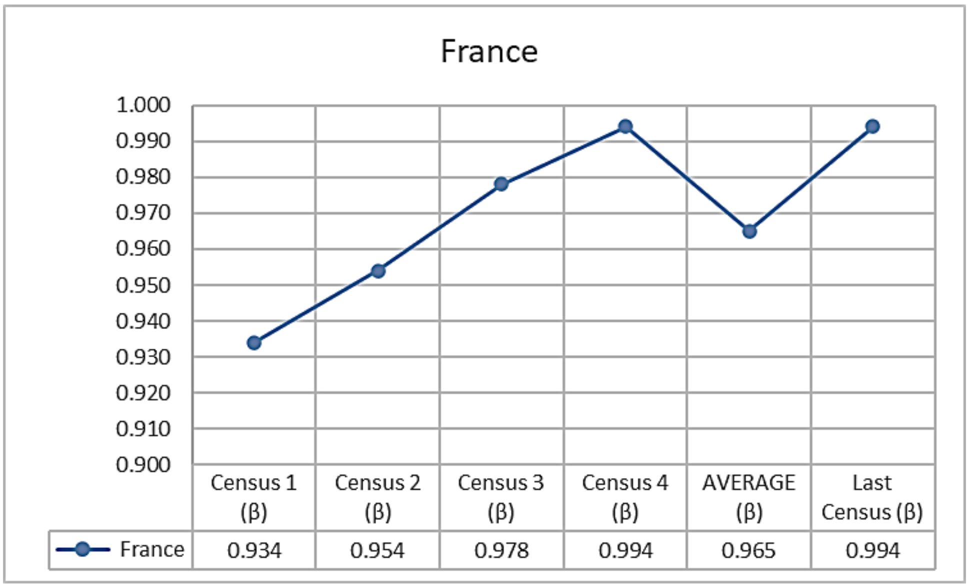


Figure 3 Significant value of β in HIC group

4.1.1.3 Lower β Value in the HIC group.

Graph for that country obtained the minimum value of β in estimation by OLS

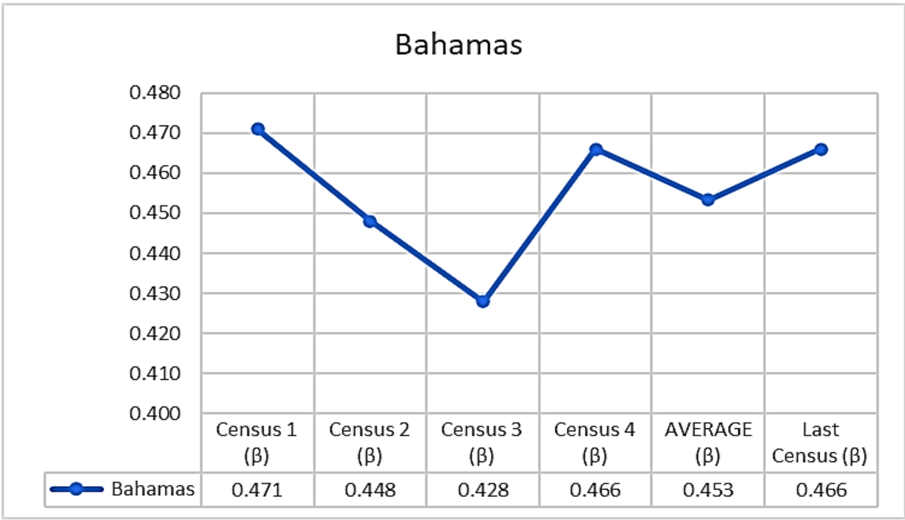


Figure 4 Lower Value of β in HIC group.

In the case of Bahamas, (Figure 4). from High Income Countries (HIC) where the all censuses value of Pareto Exponent (β) value is less or lower than 1 as compared to other countries in this group We have found that in case of the Bahamas, four censuses took place. According to data of censuses depicted in Figure 4, we found values of β as 0.374, 0.732, 0.601 and 0.423 for censes 1 to 4 respectively wih the average of 0.533. The estimated value of Pareto exponent increased from its lowest value (0.374) in census 1st to its highest value (0.732) in Census 2nd , and then it decreased to value (0.610) in census 3rd and got much less upto (0.423) in Census 4th. Data shows that the Pareto exponent's (β) has a mix trend over time shows that City

Size Distribution (CSD) is more evenness overtime with increasing trend and more unevenness when it has decreasing trend. So, our result demonstrated that hypothesis for Bahamas show that Pareto exponent is much greater than 1, It also depicts that there is Primatial or Macrocephalous distribution which means that one big city dominates the whole urban system. Hence the hypothesis is rejected for all census years shows more uneven city size distribution. Together, these results show that Zipf's law does not hold for Bahamas.

4.1.1.4 Higher β Value in MIC group.

Country that obtained the value greater or maximum value of β in estimation by OLS estimate.

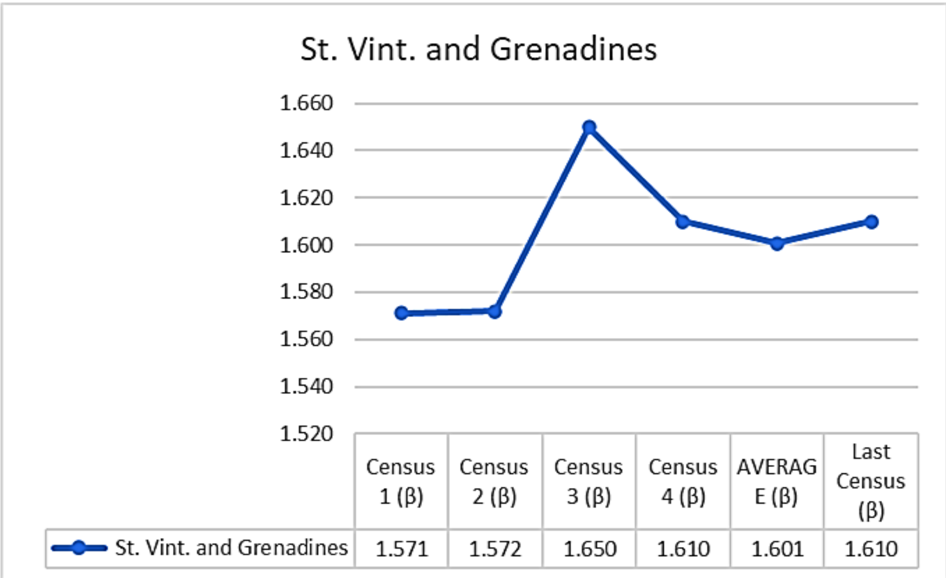


Figure 5 Maximum Value of β in MIC group.

In the case of St. Vint. and Grenadines, (Figure 5). from Medium Income Countries (MIC) where the all censuses value of Pareto Exponent (β) value is maximum or greater than 1 as compared to other countries in developing group. We have found that in case of the St. Vint. and Grenadines, four censuses took place. According to data of censuses depicted in Figure 5, we found the decreasing trend of the variations of the population. The values of β are 1.571, 1.572, 1.650, and 1.610 for censes 1^{st} to 4^{th} respectively wih the average of 1.601. The estimated value of Pareto exponent increased from its value (1.571) in census 1^{st} to value (1.572) in Census 2^{nd} , and further increased to value (1.650) in census 3^{rd} than decreased to (1.610) in Census 4^{th}. Data shows that the Pareto exponent's (β) has a mix trend over time shows that City Size Distribution (CSD) is more evenness overtime with increasing trend and more unevenness when it has decreasing trend. So, our result demonstrated that Pareto exponent of St. Vint. and Grenadines is much greater than 1, which is rejected for all census years shows more even city size distribution. The size of large and small cities approximately equal. Together, these results show that Zipf's law does not hold for St. Vint. and Grenadines**.**

4.1.1.5 Significant β Values in MIC group.

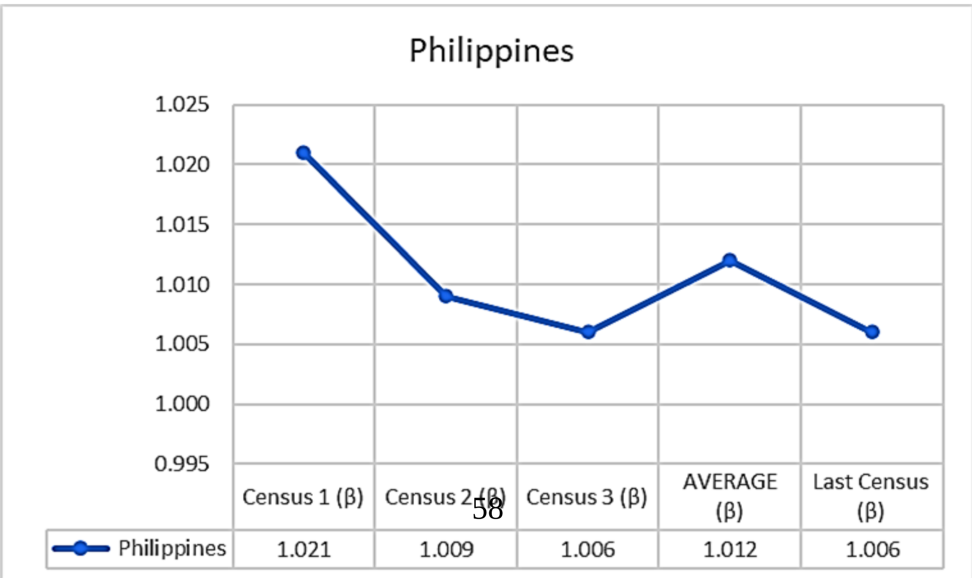


Figure 6 Significant Value of β in MIC group.

We took the example of two countries i.e., Philippines and Lebanon from the Medium Income Countries group where the value to pareto exponent is near or equal to 1 or significant value of β in estimation by OLS estimate. The below *Table 4.* states the summary statistics for the developing countries where the Pareto Exponent (β) value is near to 1. Also depicted the estimation results of the two countries namely Philippines and Lebanon which are treated as a sample to represent the overall results of Zipf's law holding in developing countries where the exponent value is near to 1. In the case of Philippines, there are three censuses of the population which took place by the gap of years as 2000, 2010 and 2015.

The above graph Figure. (6) presented the case of developing countries of the world. We took the Philippines as a sample from Medium Income Countries (MIC) where the value of PE (β) value is near or equal to 1. In case of the Philippines three censuses took place. According to data of censuses depicted in *Figure 6*, we found the mixed trend of the variations of the population. The values of β are 1.021, 1.009 and 1.006 for censes 1 to 3 respectively with the average of 1.005 and the overall decreasing trend which shows that there is slightly unevenness of the population or city size distribution. So, our result demonstrated that Zipf's law hold appropriately in case Philippines as we found the mixed trend of the variations of the population near to the 1.

On average, the value of Philippines, the R-Squares indicates about the 98% variations in the population due to cities Rank Size Distribution (RSD). Where the value of exponent moved slightly closer to 1 in all three censuses as in first three census lies slightly above but very close to 1. According to probability value, the null hypothesis of the research was rejected in all four censuses as the P-value is less than 0.05 for the coefficient of population. So it is concluded that in Philippines has appropriate holding of Zipf's Law for Rank Size Distribution of cities.

On the other hand, In the case of Lebanon, (Figure 7). from Medium Income Countries (MIC) where all censuses value of Pareto Exponent (β) value is near or equal to 1. We have found that in case of the Lebanon four censuses took place. According to data of censuses depicted in Figure 7. The estimated values of β are 0.792, 1.069, 1.092 and 1.031 for censes 1 to 4 respectively, with an average of 0.996 and the data shows that the Pareto exponent's (β) has a mix trend overtime which shows that City Size Distribution (CSD) is more evenness overtime with increasing trend and more unevenness when it has decreasing trend. So, our result demonstrated that hypothesis for Lebanon show that Pareto exponent significantly equal to 1, Hence the hypothesis does not rejected for all census years. Together, these results show that Zipf's law does hold for Lebanon.

Table 4 Value of MIC's Significant to 1

COUNTRY (DEVELOPING)	**YEAR**	**ln (Rank - 0.5) = A - α (ln Pop)** **Wald test (t, F, Chi-square): P<0.05**
Philippines	2000	$\ln(R - \frac{1}{2}) = 6.695155 - 1.021899 \ln(Pop)$ (0.065935) (0.013191) $R^2 = 0.982485$
Philippines	2010	$\ln(R - \frac{1}{2}) = 6.731973 - 1.009365 \ln(Pop)$ (0. 067996) (0.013340) $R^2 = 0.981653$
Philippines	2015	$\ln(R - \frac{1}{2}) = 6.758833 - 1.006950 \ln(Pop)$ (0.067409) (0.013125) $R^2 = 0.982145$
LEBANON	1996	$\ln(R - \frac{1}{2}) = 4.948 - 0.792 \ln(Pop)$ (0.310) (0.062) $R^2 = 0.867$
LEBANON	2007	$\ln(R - \frac{1}{2}) = 6.603 - 1.069 \ln(Pop)$ (0.408) (0.078) $R^2 = 0.883$
LEBANON	2011	$\ln(R - \frac{1}{2}) = 6.767 - 1.092 \ln(Pop)$ (0.428) (0.081) $R^2 = 0.879$

LEBANON	2017	$\ln(R - \frac{1}{2}) = 6.558 - 1.031 \ln(Pop)$ (0.484) (0.089) $R^2 = 0.841$

On average, the value of Lebanon, the R-Squares indicates about the 86% variations in the population due to cities Rank Size Distribution (RSD). Where the value of exponent moved slightly closer to 1 in all four censuses as in first three census lies slightly above but very close to 1. According to probability value, the null hypothesis of the research was rejected in all four censuses as the P-value is less than 0.05 for the coefficient of population. So, it is concluded that in Philippines has appropriate holding of Zipf's Law for Rank Size Distribution of cities.

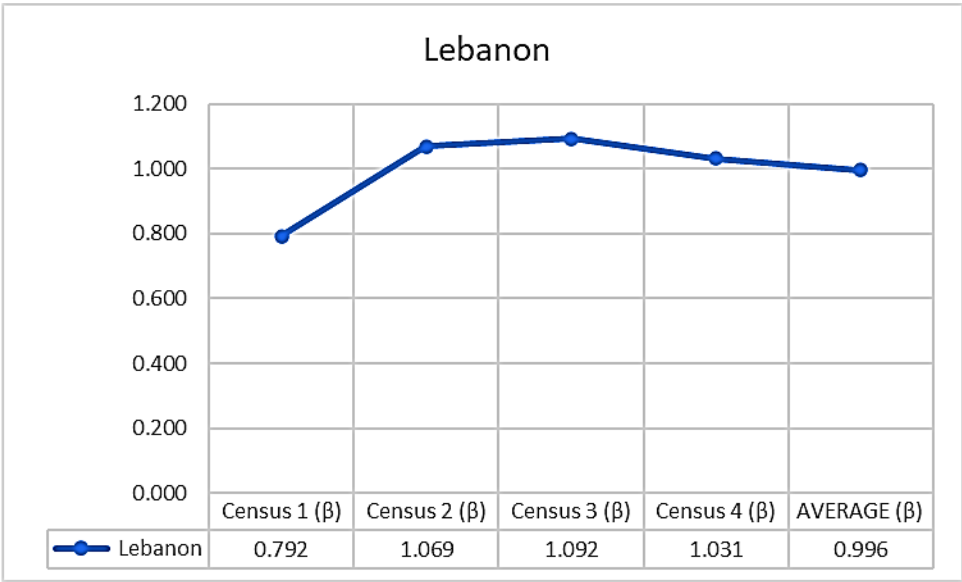


Figure 7 Significant Value of β in MIC group

4.1.1.6 Lower β value in the MIC group.

Graph for that country obtained the minimum value of β in estimation by OLS estimate.

In the case of Zimbabwe, (Figure 8). from Medium Income Countries (MIC) where the all censuses value of Pareto Exponent (β) value is minimum or lesser than 1 as compared to other countries in developing group. Four censuses took place in case of the Zimbabwe. According to data of censuses depicted in Figure 8, we found the decreasing trend of the variations of the population. The values of β are 0.442, 0.459, 0.485 and 0.438 for censes 1 to 4 respectively with the average of 0.456. The estimated value of Pareto exponent increased from its value (0.442) in census 1st to value (0.459) in census 2nd , and further increased to value (0.485) in census 3rd than decreased to (0.438) in 4th census. Data shows that the Pareto exponent's (β) has a mix trend over time shows that City Size Distribution (CSD) is more evenness overtime with increasing trend and more unevenness when it has decreasing trend. So, our result demonstrated that Pareto exponent of Zimbabwe is much lower than 1, which is rejected for all census years shows more uneven city size distribution. The size of large and small cities are much bigger, means that one or small number of countries dominates the urban system. Together, these results show that Zipf's law does not hold for Zimbabwe**.**

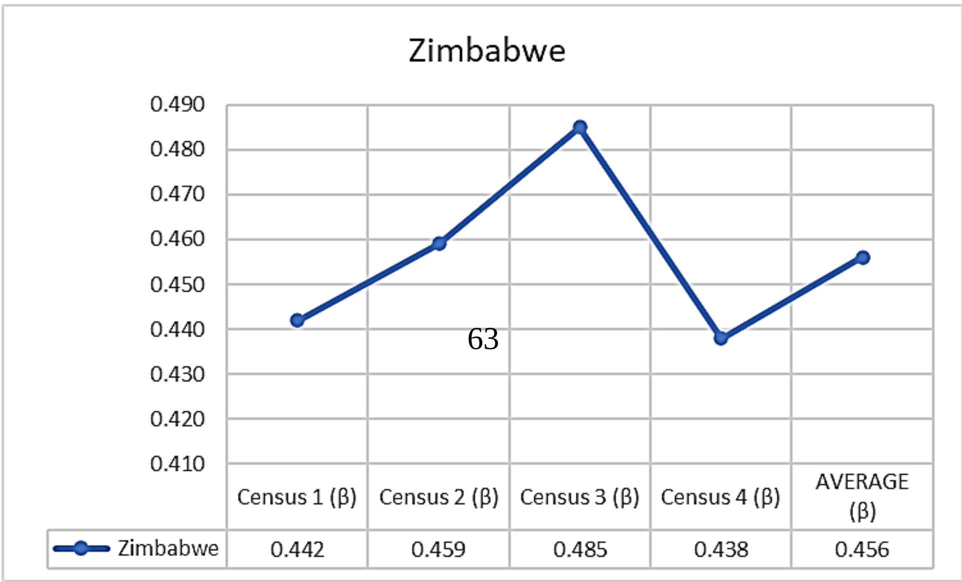


Figure 8 *Lower β Values in MIC group*

4.1.1.7 Higher β Value in the LIC group.

Graph for that country obtained the minimum value of β in estimation by OLS estimate.

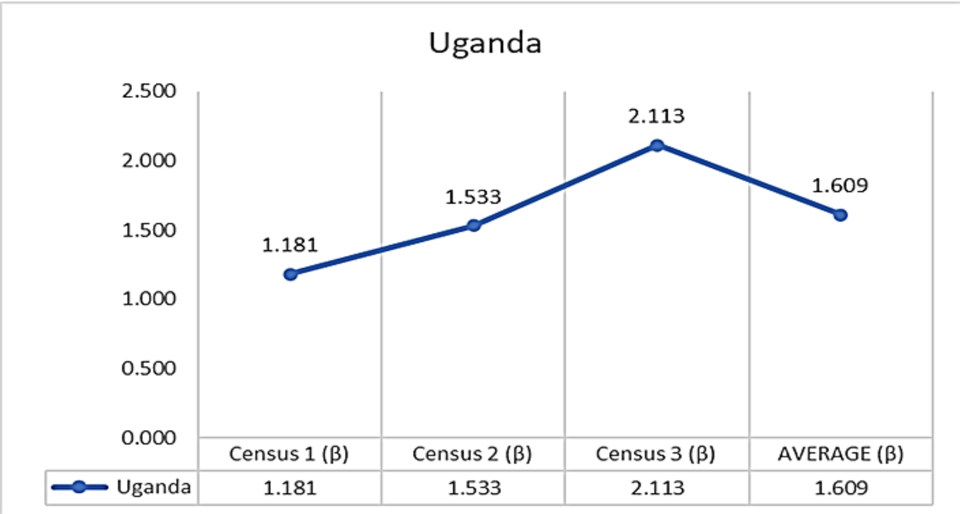


Figure 9 *Higher Value of β in LIC group.*

In the case of Uganda, (Figure 9). from Lower Income Countries (LIC) where the all censuses value of Pareto Exponent (β) value is maximum or greater than 1 as compared to other countries in under-developing group. Three censuses took place in case of the Uganda. According to data of censuses depicted in Figure 9. The estimated values of β are 1.181, 1.533 and 2.113 for censes 1st to 3rd respectively wih the average of 1.609. The estimated value of Pareto exponent increased from its value 1.181 in 1st census to value (1.533) in 2nd census, and further increased to value (2.113) in 3rd census. Estimation results shows that the Pareto exponent's (β) has a increasing trend over time shows that City Size Distribution (CSD) is more evenness overtime with increasing trend. So, our result demonstrated that Pareto exponent of Zimbabwe is much higher than 1, which is rejected for all census years shows more even city size distribution. The size of large and small cities are very little or approximately same. Together, these results show that Zipf's law does not hold for Uganda**.**

4.1.1.8 Significant β values in the LIC group.

We took the example of two under-developed countries i.e., Yemen and DR. Congo from the Lower Income group where the value to pareto exponent is near or equal to 1 or significant value of β in estimation by OLS estimate. The below *Table 5.* states the summary statistics for the under-developed countries where the Pareto Exponent (β) value is near to 1 by indicating the slight Rank Size Rule in the population dispersion. Also depicted the estimation results of the two countries namely Yemen and DR. Congo which are treated as a sample to represent the overall results of Zipf's law holding in under-developed countries where the exponent value is near to 1. In the case of Yemen, there are four censuses of the population which took place by the gap of years as $1994, 2004, 2009, and, 2012$.

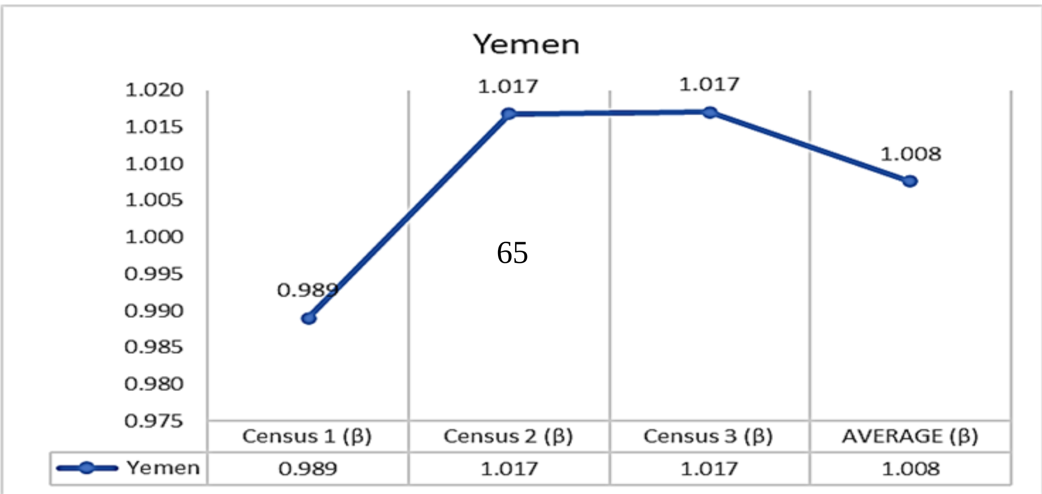


Figure 10 *Significant Value of β in LIC group.*

The above graph Figure. (10) presented the case of under-developed country of the world. We took the Yemen as a sample from Lower Income Countries (LIC) where the value of PE (β) value is near or equal to 1. In case of the Yemen four censuses took place. According to data of censuses depicted in *Figure 10*, we found the mixed trend of the variations of the population. The values of β are 0.989 , 1.017, 1,017 and 1.018 for censes 1 to 4 respectively with an average of 1.008. Estimation results shows that the Pareto exponent's (β) has a little increasing trend over time shows that City Size Distribution (CSD) is evenness overtime with increasing trend. So, Our result demonstrated that Zipf's law hold appropriately in case Yemen as we found the Pareto exponent near to the 1.

On average, the value of Yemen, the R-Squares indicates about the 72% variations in the population due to cities Rank Size Distribution (RSD). Where the value of exponent moved slightly closer to 1 in all four censuses. According to probability value, the null hypothesis of the research was rejected in all four censuses as the P-value is less than 0.05 for the coefficient of population. So, it is concluded that in Yemen there are evidence about the appropriate holding of Zipf's Law for Rank Size Distribution of cities.

On the other hand, In the case of DR Congo, (Figure 11). from Lower Income Countries (LIC) where all censuses value of Pareto Exponent (β) value is near or equal to 1. In case of the DR Congo three censuses took place.

Table 5 Value of LIC's significance to 1.

COUNTRY (UNDER-DEVELOPED)	**YEAR**	**ln (Rank - 0.5) = A - α (ln Pop)** **Wald test (t, F, Chi-square): P<0.05**
YEMEN	1994	$\ln(R - \frac{1}{2}) = 6.56 - 0.989 \ln(Pop)$ (0. 0.82) (0.143) $R^2 = 0.716$
YEMEN	2004	$\ln(R - \frac{1}{2}) = 6.85 - 1.017 \ln(Pop)$ (0. 827) (0.141) $R^2 = 0.732$
YEMEN	2009	$\ln(R - \frac{1}{2}) = 6.911 - 1,017 \ln(Pop)$ (0.835) (0.141) $R^2 = 0.733$
YEMEN	2012	$\ln(R - \frac{1}{2}) = 6.951 - 1.018 \ln(Pop)$ (0. 840) (0.141) $R^2 = 0.733$
DR CONGO	1984	$\ln(R - \frac{1}{2}) = 6.21 - 1.021 \ln(Pop)$ (0.033) (0.007) $R^2 = 0.995$

DR CONGO	2010	$\ln(R-\frac{1}{2}) = 6.017 - 0.924 \ln(Pop)$ (0.036) (0.008) $R^2 = 0.993$

According to data of censuses depicted in Figure 11, we found the mixed trend of the variations of the population. The values of β are 1.021, and 0.924 for censes 1st and 2nd respectively with an average of 0.973. Estimation results shows that the Pareto exponent's (β) has a little decreasing trend over time shows that City Size Distribution (CSD) is unevenness overtime with increasing trend. So, Our result demonstrated that Zipf's law hold appropriately in case DR Congo Pareto exponent (β) near to the 1.

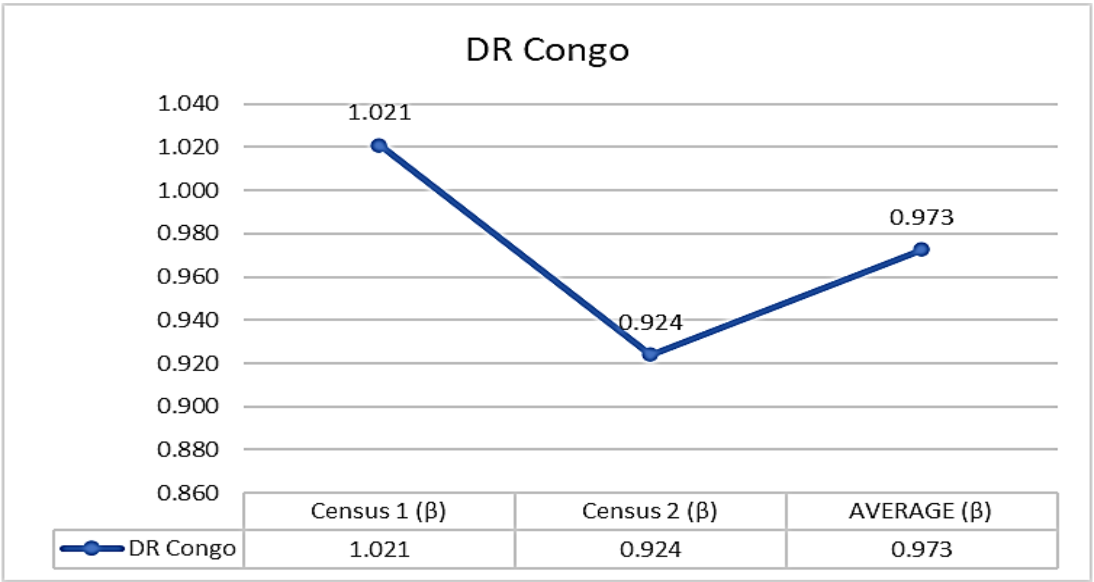


Figure 11 Significant Value of β in LIC group.

On average, the value of R-Squares of DR Congo indicates about the 99% variations in the population due to cities Rank Size Distribution (RSD). According to probability value, the null hypothesis of the research was rejected in all two censuses as the P-value is less than 0.05 for the coefficient of population. So, Our result demonstrated that the France hypothesis show that Pareto exponent equals to 1, and we do not reject the null hypothesis. Together, these results show that Zipf's law holdsfor DR Congo.

4.1.1.9 Lowest β values in the LIC group.

Graph for that country obtained the minimum value of β in estimation by OLS estimate.

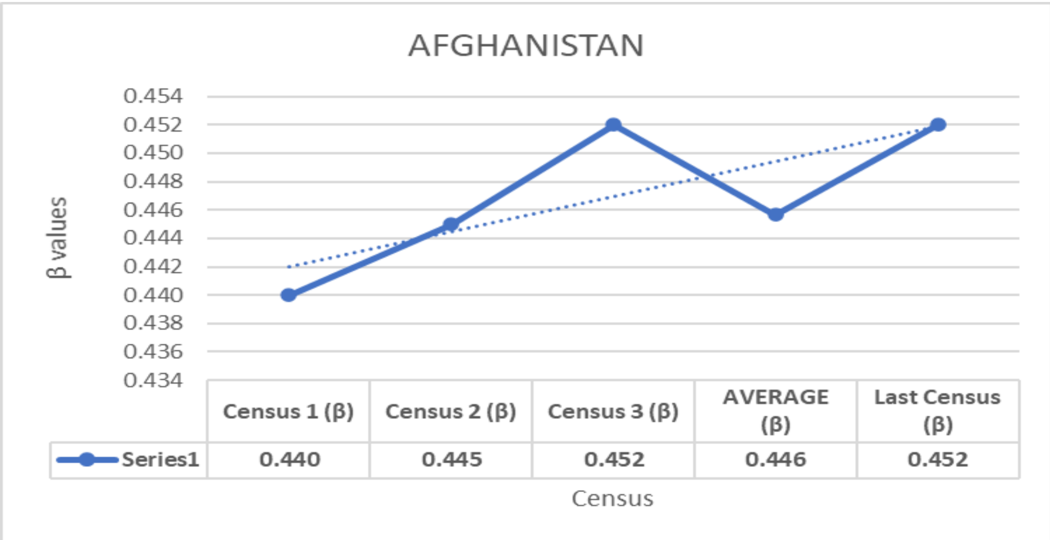


Figure 12 Lowest Value of β in LIC group.

In the case of Afghanistan, (Figure 12). from Lower Income Countries (LIC) where all censuses value of Pareto Exponent (β) value is lesser or lower than 1 as compared to other countries in this group. We have found that in case of Afghanistan, three censuses took place. According to data of censuses depicted in Figure 12, we found values of β as 0.440, 0.445, and 0.452 for censes 1st to 3rd respectively with the average of (0.466). The estimated value of Pareto exponent increased from its lowest value (0.440) in census 1st to its highest value (0.445) in Census 2nd , and further increased to value (0.452) in census 3rd. Data shows that the Pareto exponent's (β) has a increasing trend over time shows that City Size Distribution (CSD) is more even overtime. So, our result demonstrated that hypothesis for Afghanistan shows that Pareto exponent is lower than 1, It also depicts that there is Primatial or Macrocephalous distribution which means that one or small number of big city dominates the whole urban system. Hence the hypothesis is rejected for all census years shows more uneven city size distribution. Together, these results show that Zipf's law does not hold for Afghanistan.

4.1.2 Zipf Law in Human Development Index groups

Countries with more or high estimated Pareto Exponent value are less hierarchical and less concentrated, higher the value will show the less of the hierarchies and less concentration and there is more equality in dist. of population. Moreover, the values of data have less spread or less dispersion around the mean value of β. Countries having an upward tendency in their estimated value of β or Zipf's coefficient, and their value is greater or higher than unity indicates evenness in dist. of cities in these countries. More the higher value than unity described as less concentration, more equally distributed in the population and less hierarchy was seen in these countries.

Countries with less estimated PE value are more hierarchical and more concentrated, lesser the value will show more of the hierarchies and more concentration and there is inequality in dist. of population Moreover, the values of data have less spread or less dispersion around the mean value of β and for higher value the interpretation became opposite. More the lower value than unity described as more concentration, lesser equally distributed in the population and more hierarchy was seen in these countries.

Division of these countries are on the base of Developed countries (HIC's), Developing countries (MIC's), Underdeveloped countries (LIC's). Pareto exponent (PE) for all countries using updated available observations by using Ordinary Least Square and Wald coefficient restriction test including the 95%. A significant value equal to 1 means Zipf's law holds, and the outcome is the cities numbers for the first biggest country is double that of the second biggest country, Three times higher than that of the third-largest country, and so on. The distribution of city sizes or rank-size distribution (RSD) reflects hierarchical mechanisms involved in urban systems.

4.1.2.1 Zipf's Law for Developed Countries (HIC's).

Many studies are there in the literature about countries, some of them are on individual country and some of them are on cross country studies. Appendix Table I, include details of all countries lies in HIC's group. Appendix Table II, shows higher, significant, and lower values of the countries

Table 6 shows the average value of the developed countries as 0.867. Observations are 65, we found the β values of 16 out of 65 countries (24.6 %) are significantly greater than 1 and for the rest of the 49 countries their estimated value is significantly less than unity (approx. 75.4 %). and the maximum value is of the Czech Republic obtained is 1.864 and the minimum value is of Bahamas. The kurtosis shows the distribution is leptokurtic because the value of kurtosis is positive five. The total sum of the value of the coefficient is 56.4. Our data is 28.9% scattered from the average value, or the data is 28.9% above or below the mean value.

Table 6 Descriptive Statistics of Developed Countries

Mean	0.867
Median	0.778
Maximum	1.864
Minimum	0.453
Std. Dev.	0.289
Skewness	1.425
Kurtosis	5.040
Jarque-Bera	33.26
Probability	0.000
Sum	56.38
Sum Sq. Dev.	5.356
Observations	65

The lesser the value of the Pareto exponent than unity it implies that the people from around the world travels and get immigration to developed countries. There are many reasons why in developed countries, there is less hierarchy and Zipf's law does not hold when we see the probability of all countries was less than 0.05 so we do not accept the hypothesis [Ho: $\mathrm{C}\ (2) = 1$]. hence Zipf's Law does not hold.

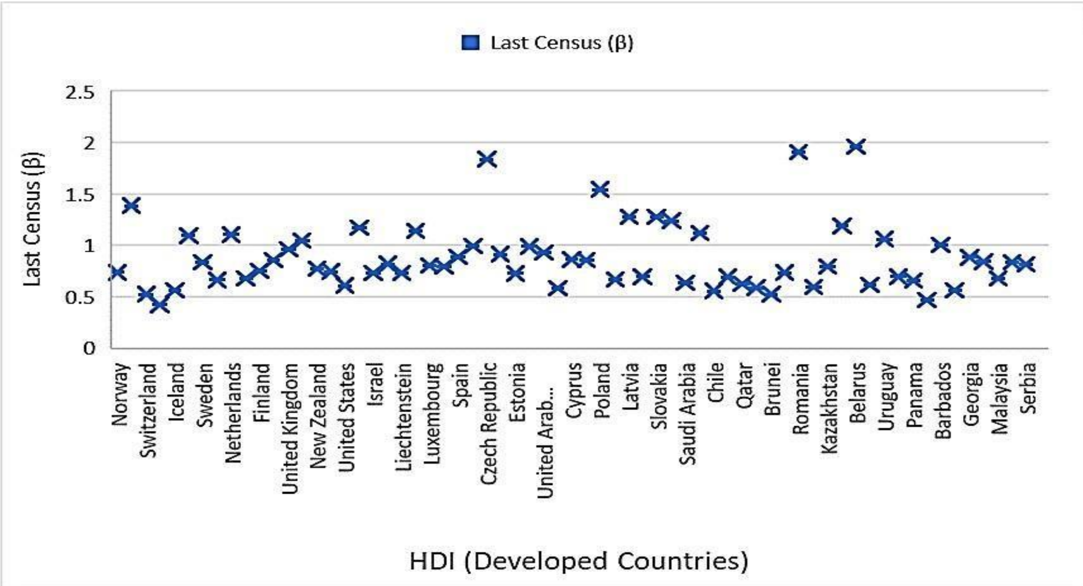
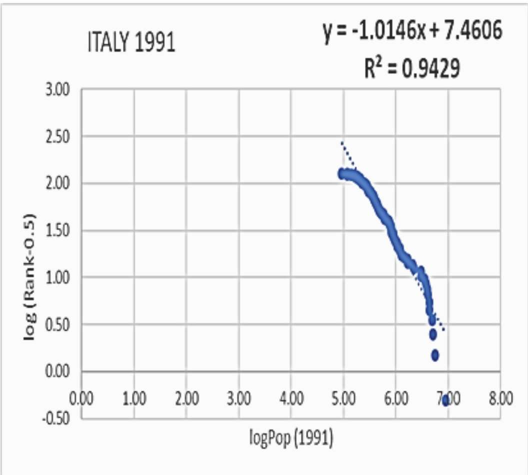
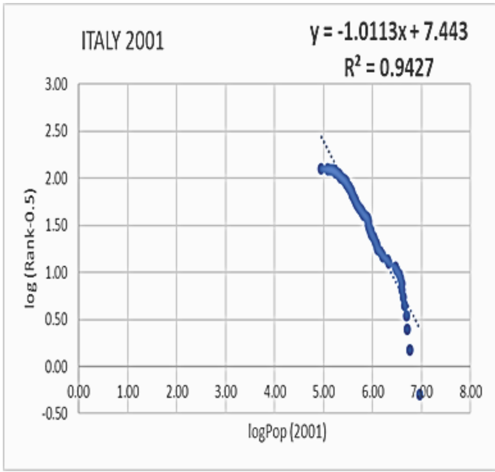


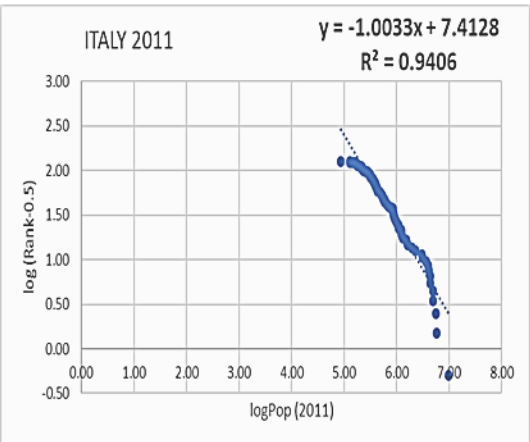
Figure 13 Last Census coefficients obtained from HIC's

Figure (13) is the estimation of Pareto exponent (PE) for all countries that lies under the developed nations. By taking an average of all census coefficients for each country we have constructed the graph showing the average β values corresponding to the countries named on the vertical axis. This graph is roughly showing us that most of the values are below unity (1) Graph summarizes what table tells. More of the (β) value lies below 1 which is the benchmark to know about holding of Zipf's law (ZL). This figure is the relation between the HICs and the last census values of PE. Pareto exponent's (PE) of these countries tilt downwards which demonstrates that these countries jointly obtained lesser value than unity.

Appendix Table 1 highlights the complete findings of the OLS regression of equation (5) where we calculated the coefficient of Zipf's law or Pareto exponent of developed countries. The results are very large in number, so we had to use the average values of the (β) coefficient. Results show that the largest value of the Pareto exponent of 1.864 is obtained by the Czech Republic and its overall trend of β values are decreasing in nature, whereas the lowest value of 0.453 and overall decreasing trend for the country named the Bahamas, Expectedly, these countries are associated with many small cities and no primate city, Switzerland comprises 27 cities whereas Hong Kong has 18 cities.







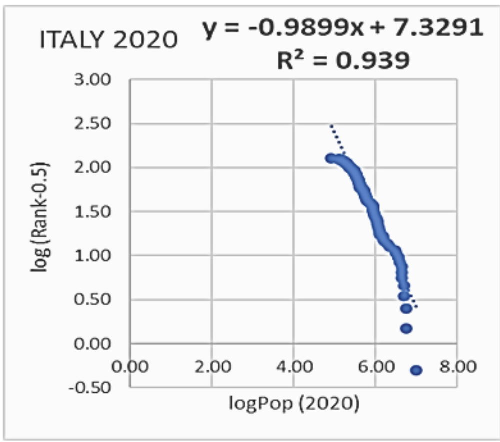


Figure 14 Log (R − 0.5) vs. log (Pop) plots for significant HICs

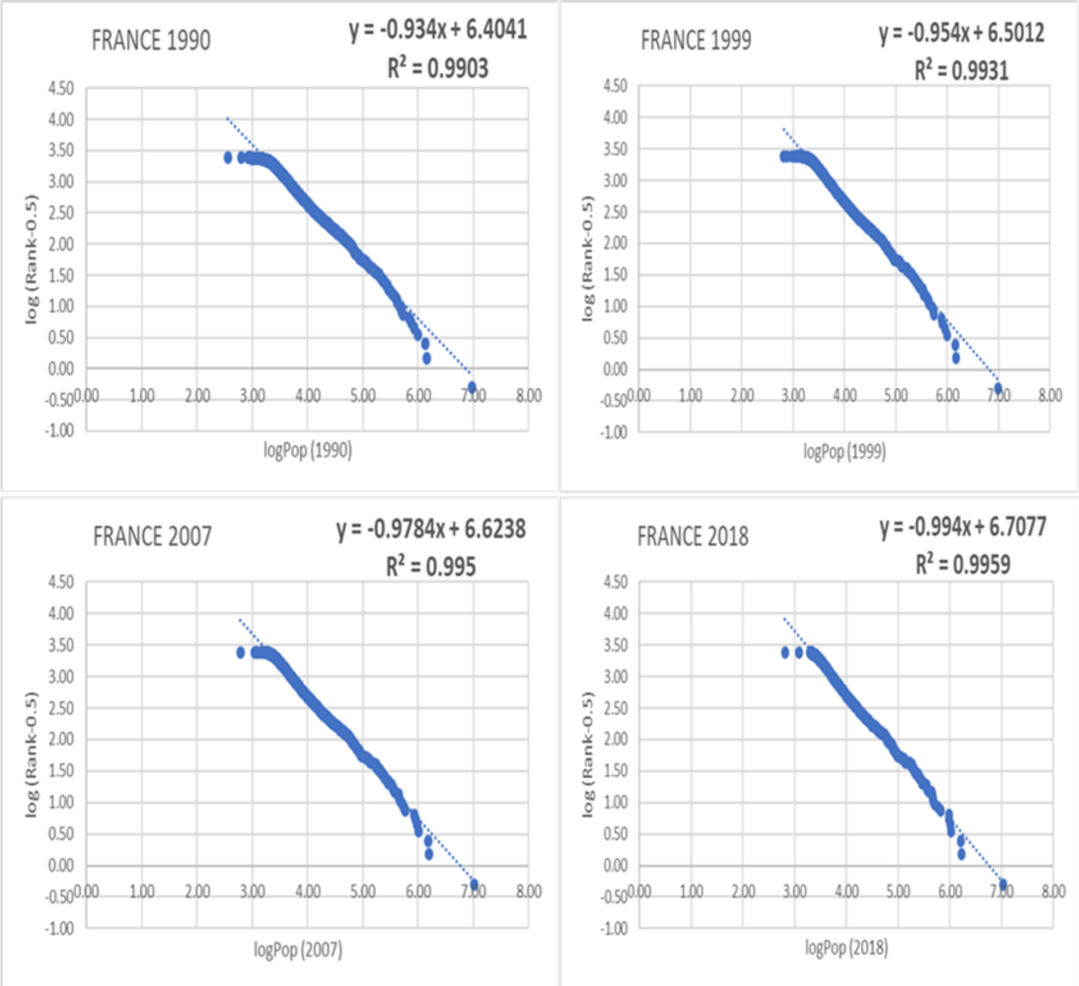


Figure 15 Log (R − 0.5) vs. log (Pop) plots for significant HICs

Figure 14 and Figure 15 depicts Italy has an average value of 1.005, its exponent of Pareto distribution decrease overtime from 1.015, 1.011, 1.003 to 0.989 means that the β-value is declining overtime which represents that city in Italy are more uneven distributed and have more hierarchy. France got an average value of 0.965 and its exponential trend is upward which means that there is more equal dispersion of population and less hierarchy. United Kingdom’s mean value for the coefficient is 0.958, overtime the exponent value increases so with time the cities in the United

Kingdoms are more evenly distributed. Researchers have found that the (β) coefficient of the UK was not significantly higher than unity in previous studies (Rosen & Resnik 1980) and (Cheshire 1999).

We inspect the city size distribution (CSD) at the country level for Poland having an average value of β is 1.572 and which is greater than unity, Ireland 1.392, Belarus 1.343, Slovakia 1.331 and for Hungary, we obtain 1.277, Austria takes 1.191, Netherlands obtained the 1.114 and generally the β values for these countries is decreasing over time, Lativa 1.251, Russia 1.141 and their overtime value of β is increasing in nature and these values are far higher than our hypothesis value of 1. From the different census, the estimated value of the Pareto exponent (PT) for Germany is 1.097 and Uruguay estimated value is 1.074 which is higher than the unity and have less uneven dispersion of population than the above countries.

Lesser the value of β shows higher concentration among the population, more inequality in distribution and high hierarchy. These countries obtained far less amount of coefficient value Barbados 0.698, Australia 0.689, Denmark 0.683, Croatia 0.680, Palau 0.673, Turkey 0.642, and United States 0.640. Same with the case of Saudi Arabia 0.635, Oman 0.611 and others like Greece and Chile obtained 0.575 and 0.565. These countries have more outsiders to visit their country or take immigration to move to these countries for their better future.

4.1.2.2 Zipf's Law for Developing Countries (MIC's).

There are many studies of countries in the literature, some of which are individual country studies and others which are cross-country studies. Appendix Table III shows all nations which are often known as developing countries. Appendix Table (IV) displays the estimation result after doing an OLS regression

Table 7 Descriptive Statistics of Developing Countries

Mean	0.778
Median	0.716
Maximum	1.601
Minimum	0.39
Std. Dev.	0.238
Skewness	0.44
Kurtosis	4.414
Jarque-Bera	10.521
Probability	0.005
Sum	70.87
Sum Sq. Dev.	5.109
Observations	91

Table (7) shows the average value of the developing countries as 0.778 which is smaller than the average of developed nations. Lesser the value of the Pareto exponent means that there is more hierarchy and uneven distribution of cities than developed nations. Observations are 90, we found the β values of 16 out of 90 countries (17.78%) are significantly greater than 1 and for the rest of the 74 countries their estimated value is significantly less than unity (approx. 82.2%). The total sum of these (β) values is 70.87 and the maximum value is of Saint Vincent and Grenadines obtained is 1.601 and the minimum value is of Brazil 0.390. The kurtosis shows the distribution is leptokurtic because the value of kurtosis is positive four. The probability of this group of medium-income countries known as developing nations is 0.005 which is less than 0.05 and the Standard Deviation of the distribution is 0.238 demonstrating the information about the data. 23.8% is the dispersion among the value from average or this amount shows that our data is this much above or beneath the mean value. hence Zipf's law (ZL) does not holds.

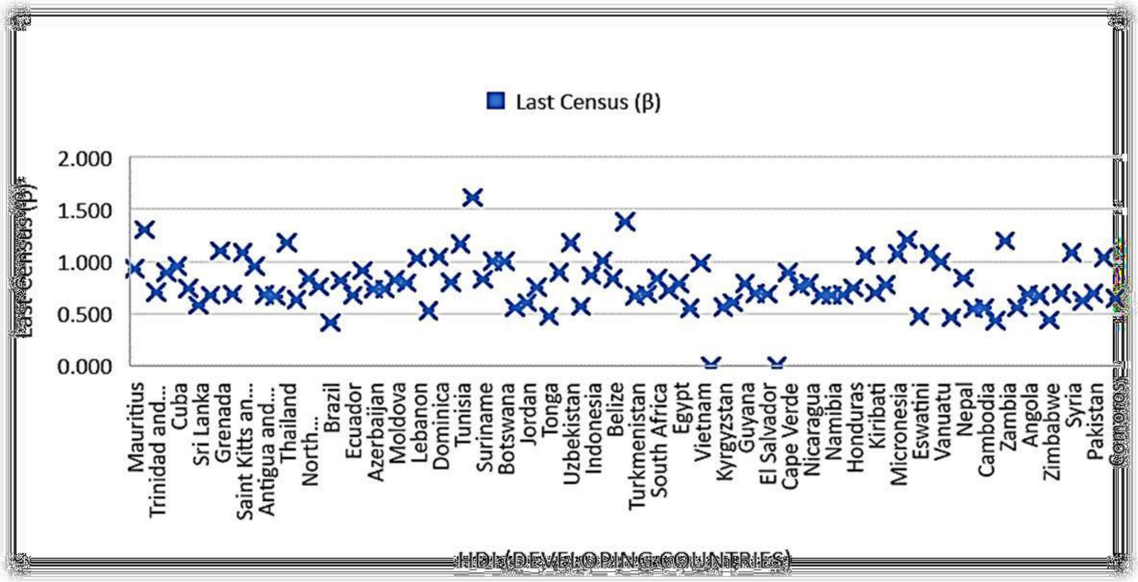


Figure 16 MIC's Pareto Exponents of the last census

Figure 16, is the estimation of (PE) Pareto exponent for all countries lies under the medium-income nations with (MIC's), using updated available observations and applying (OLS) Ordinary Least Square and Wald coefficient restriction test including the 5% (SL) level of significance. Graph summarizes what the table states. As shown in Figure 16, more of the (β) value lies below 1 which is the benchmark to check about holding of (ZL) Zipf's law. This figure reveals the connection between the MIC and the last census value of PEs. Pareto exponent's (PE) of these countries tilt downwards which demonstrates that these countries jointly obtained lesser value but somehow near to one as compared to very high-income nations.

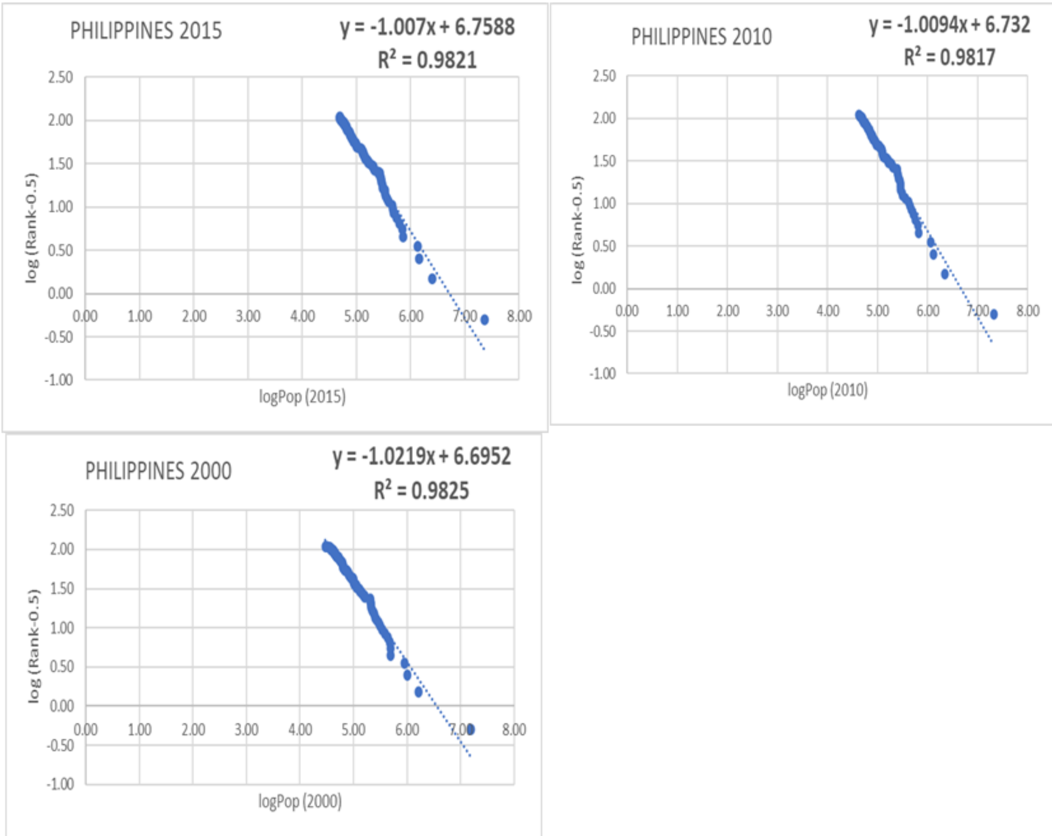


Figure 17 Log (R-0.5) vs. log (Pop) plots for significant MICs.

Figure 17 and Figure 18 shows calculated PE (β) for those emerging countries that are significant, as a result, we obtained Philippines 1.01, Dominica 1.017 which are slightly higher but significant, while Lebanon 0.996 and Micronesia 0.992 obtained the values slightly below the unity which are also statistically significant

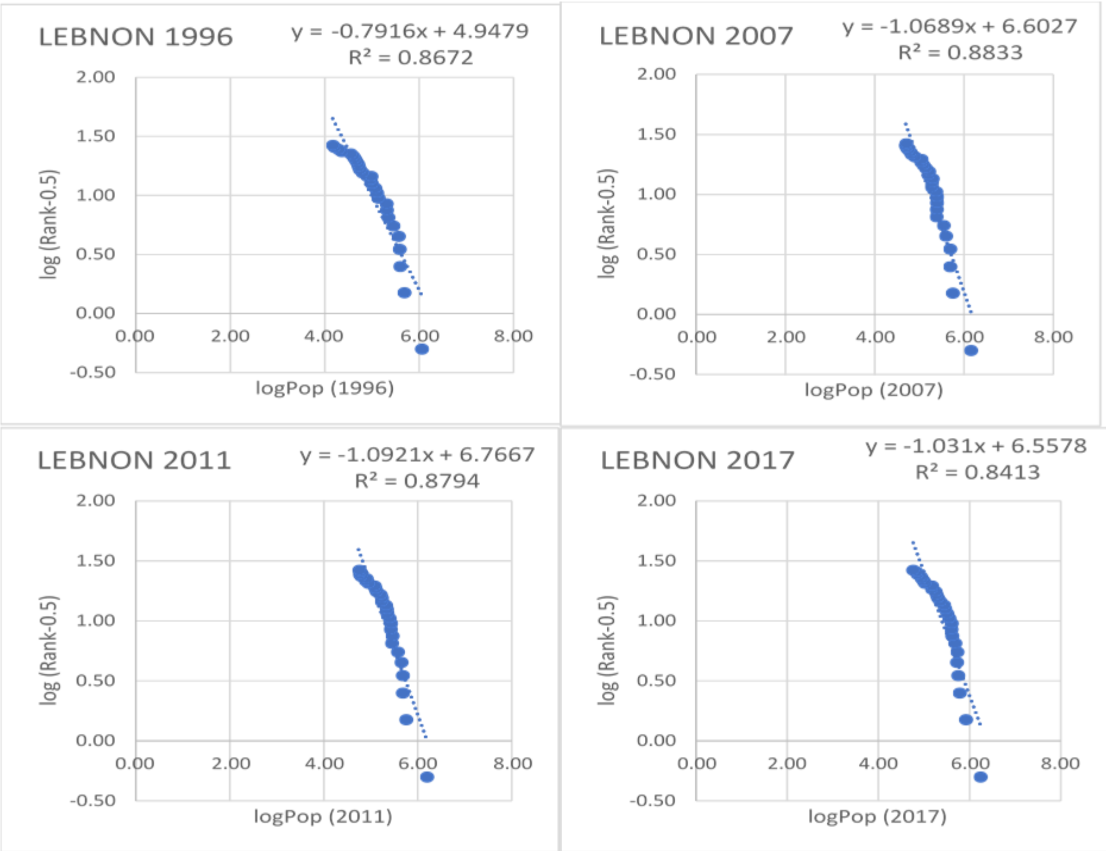


Figure 18 Log (R-0.5) vs. log (Pop) plots for significant MICs

Countries with less estimated PE value is more hierarchical and more concentrated, lesser the value will show more of the hierarchies and concentration and there is more equality in dist. of population. Moreover, the values of data have less spread or less dispersion around the mean value of β. We will be discussing this in this para. about the countries having a downward tendency in their estimated value of β but obtained higher value than unity and their value is greater or higher than unity indicates unevenness in dist. of cities in these countries, OLS reg. indicates the Zipf's coefficient of Tunisia 1.203, Mongolia 1.152, Grenada 1.119, Vanuatu 1.022, and Dominica got 1.017 the more the higher value than unity described as less concentration, more equally distributed in pop and less hierarchy was seen in these countries.

Result to conform Zipf's law (ZL) about the countries having a downward tendency starting from Ukraine and ended up Botswana got the value ranging from 0.971 to 0.902. The estimated values from Suriname to Belize ranged the values from 0.881 to 0.805 which are even lesser in amount than the previous group of countries. Moreover, Maldives value for Zipf's coefficient is 0.796 and goes to Kiribati 0.700. Likewise, the list continued but the reason for breaking down the data is differentiated which group have more hierarchies.

We examine the city size distribution (CSD) at the country level for Samoa having an average value of β is (1.357) and which is greater than unity Zambia (1.222), and for Seychelles we obtain (1.211), Thailand got (1.206), Tunisia takes (1.203) Mongolia (1.152), Laos takes (1.139) Grenada took (1.119). From the different census, the estimated value of the Pareto exponent (PT) for Papua New Guinea obtained the (1.048) and Saint, Kitts & Nevis estimated value is (1.053) which is higher than the unity and have less uneven dispersion of population than the above countries.

4.1.2.3 Zipf's Law for Under-Developed Countries (LIC's).

Low-income countries are generally known as third world countries. Underdeveloped nations have many problems because of their less development and minimum resources. Underdeveloped nations obtained the mean value of β greater than MIC's but more in the case with HIC's developed nations. for Under-Developed Countries. Appendix Table (V), include details of all countries lies in LIC's group. Appendix (Table VI), shows higher, significant, and lower values of the countries in LIC's.

Table 8 Descriptive statistics of Underdeveloped countries

Mean	0.827
Median	0.863
Maximum	1.609
Minimum	0.446
Std. Dev.	0.293
Skewness	0.51
Kurtosis	2.761
Jarque-Bera	1.509
Probability	0.47
Sum	27.28
Sum Sq. Dev.	2.738
Observations	33

Table (8) shows the descriptive analysis of given countries in LIC group is 33, we found the β values of 10 out of 31 countries (32.2 %) are significantly greater than 1 and for rest of the 23 countries have estimated value which is significantly less than unity (approx. 74.2 %). The total sum of these (β) values is 27.28 and the maximum value is of the Uganda obtained is 1.601, with an average value of 0.827 which is smaller than the average of (HIC) but larger as compared to (MIC) average value. The minimum value is of Brazil 0.390. The kurtosis shows the distribution is Meso-Kurtic because the value of kurtosis is 2.76. The probability of the group of low-income countries known as the under-developed nation is 0.005 which is less than 0.05 and the Std. Dev of the distribution is 0.293 greater as compared to MIC's demonstrates the information about the data. 29.3% is because the probability of all countries was significantly lesser than 0.05 and except two countries, we reject the hypothesis of the coefficient being equal to unity, hence Zipf's law (ZL) does not holds.

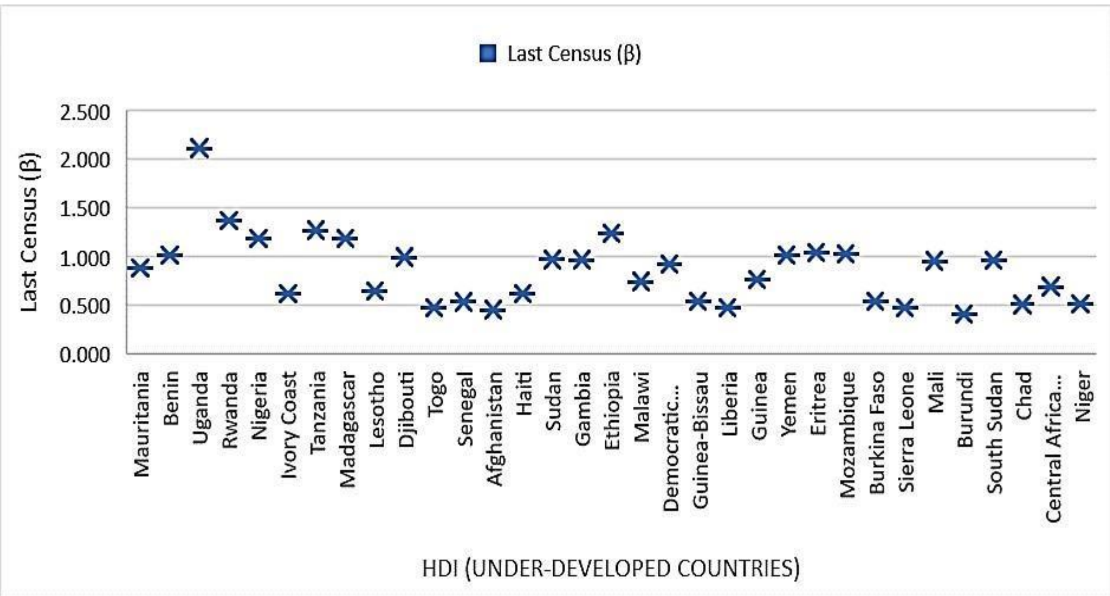


Figure 19 LIC's Pareto Exponents of the Last census

Figure. 19, is the estimation of last census Pareto exponent for all countries lies under the group's underdeveloped nations with (LIC's) low-income countries using updated available observations by using (OLS) Ordinary Least Square and Wald coefficient restriction test including the 95% (CI). Graph summarizes what table tells. More of the (β) value lies below 1 which is the benchmark to know about holding of (ZL) Zipf's law. The result exhibits that 10 out of 33 countries in the group of LICs grabs the value greater than 1, rest of the 23 countries have a value less than 1, which depicts that most of them lies below the unity and give us a clear picture that there is a vastly unequal spread of the population among LICs. Underdeveloped nations collectively have more hierarchies as compared to developed but not for developing countries.

(Figure 20) above depicts the Zipf curves of those countries which hold the law, it shows the significance of the β coefficient which are statistically near or significant to one. Furthermore, Yemen's averagely value became 1.008 which high than 1 but in a smaller amount and it became statistically significant which conforms the Zipf's law. It implies that distribution in Yemen is according to Zipf, Yemen having an increasing trend but there is a slight change in β values. The Democratic Republic of Congo has an average value near to 1 (0.973) obtained from two census data with a decreasing trend. The value is slightly smaller than 1 but it is statically substantial. Table IV. shows calculated PE (β) for those emerging countries that are significant, as a result, we obtained Philippines 1.01, Dominica 1.017 which are slightly higher but significant, while Lebanon 0.996 and Micronesia 0.992 obtained the values slightly below the unity which are also significant statistically.

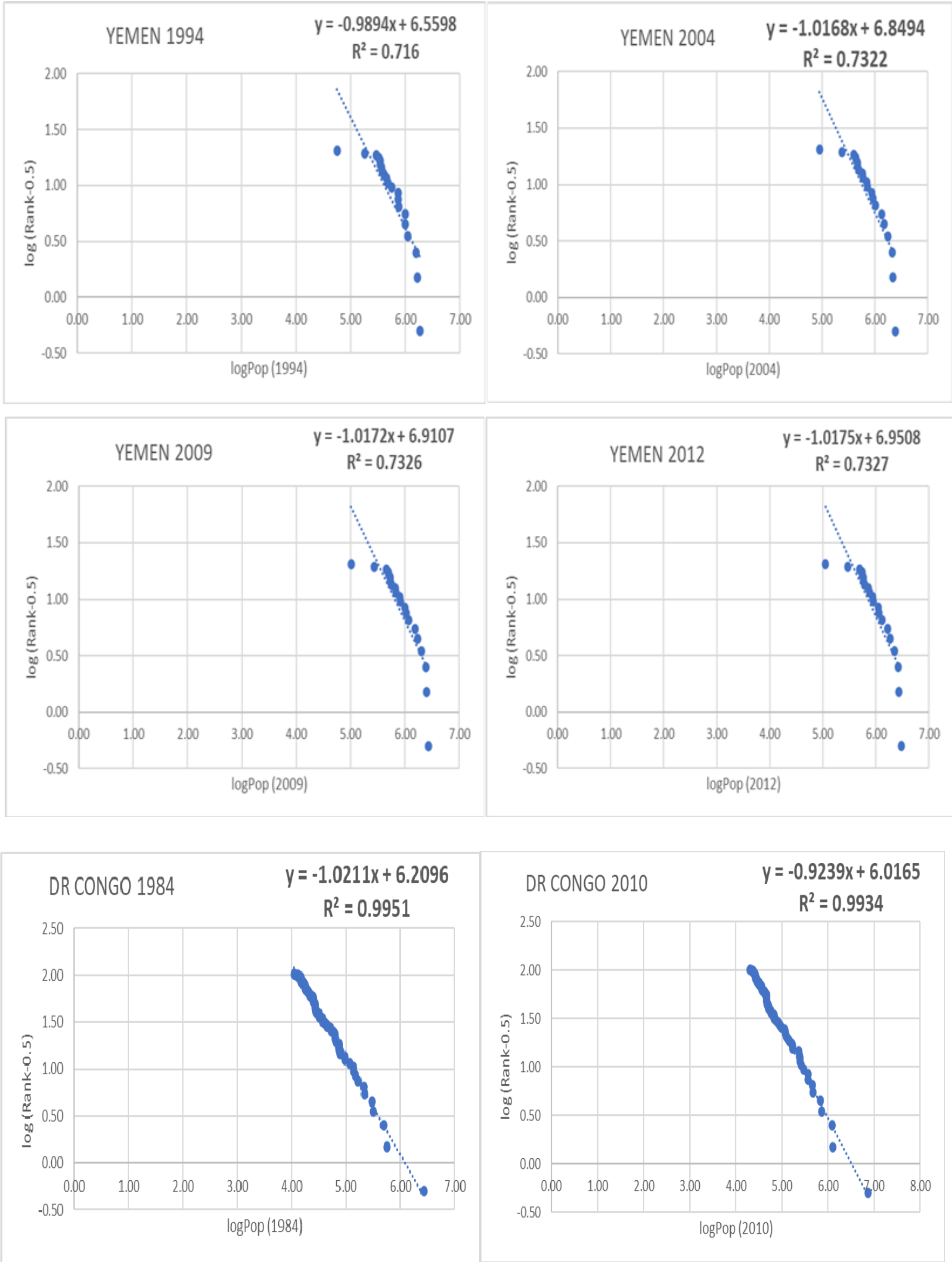


Figure 20 Log (R − 0.5) vs. log (Pop) plots for significant LICs

We will be discussing in this paragraph the countries having a value greater or higher than unity indicates unevenness in dist. of cities in these countries, Ordinary Least Square regression indicates the Zipf's coefficient of Ethiopia has a value of 1.238, Tanzania 1.090, Eritrea 1.061, Benin 1.034, Mozambique 1.023, and the countries having a value greater than 1. Furthermore, Countries having average values greater than unity and with an upward trend are Madagascar 1.153, Nigeria 1.129.

Those countries having lesser value than unity are Gambia 0.863, Djibouti 0.85, Malawi 0.742, Guinea Bissau 0.729, Central African Republic 0.671, Guinea 0.558, Lesotho 0.557, Haiti 0.520, Niger 0.514, and Senegal 0.509. and like South Sudan 0.962, Mali 0.959, Mauritania 0.910, Sudan 0.907, Ivory Coast 0.610, Burkina Faso 0.536, Liberia & Togo have average values 0.505 and 0.488. Table VI, shows the lowest value of 0.446 is obtained for Afghanistan, followed by Sierre Leone (0.471) and Chad (0.508).

4.2 Empirical Results

Previous ''empirical studies'' for the countries in past along with the result of the current finding. Some studies found City size distribution conform Zipf's law in the United States**,** much research has been done on Urban Areas and their hierarchies by (Krugman, 1996) using *130* Metropolitan Areas, (Gabaix, 1999) has used 135 largest metropolitan areas for the year 1990. (Ioannides & Overman, 2003) used the data of 112 cities in the census year 1900 to 334 cities in 1990 and obtained the results that city size distribution holds. (Berry & Okulicz-Kozaryn, 2012) also, study the American case and their results depict that Zipf Law holds for America. Some researchers also find that the city size distribution does not obey Zipf law in the US like (Black & Henderson, 2003) when they applied on 282 larger metro cities, (Eeckhout, 2004) with 25,359 places from 2000 US census with population 1 to 8 million. Same as this study (Bee, 2013) used 28,916 cities or 17,569 clusters and found that Zipf law does not hold.

Our finding for the US is that it took the Pareto exponent lower than 1. United States obtained the β values as $0.673, 0.652, 0.627,$ and 0.608 from 1st to 4th census. Which depicts that US has value statistically less than 1 indicates the city size distribution is more uneven where one or a few cities dominate the whole urban structure. This is also referred as primatial or macrocephalous distribution.

(Giesen & Südekum, 2011) concluded that Zipfian power law holds for Germany with sample of those cities having more than 100,000 inhabitants. Our finding for demonstrate the value $1.100, 1.099, and\ 1.093$ with average of 1.097 show that exponent got a higher value than 1 which says that a value statistically higher than 1 indicates a more even city size distribution where the size difference between larger and smaller cities is little and Zipf law does not hold which is opposite finding than the previous study.

In case of Russia, (Rastvortseva & Manaeva, 2016) studies that Zipf law holds exactly in the cities, but our finding shows that in four census the Russia took the exponent values as $1.081, 1.120, 1.172, and\ 1.189$ with an average of 1.141 which is statistically higher than 1 indicates a more even city size distribution where the size difference between larger and smaller cities is little and Zipf law does not hold which is opposite finding than the previous study.

Other researcher studies about different nations like in case of Canada the researcher named (Lanaspa et al., 2003) took 152 large urban cities, (Dubé & Polèse, 2016) took 135 largest urban areas, and his finding is that city size distribution does not conform to Zipf's law. Current study estimated the pareto values as the $0.785,\ 0.767,\ 0.758\ \ and\ \ 0.746$ with an average of 0.764. which is statistically lower than 1 indicates a more uneven city size distribution where the size difference between larger and smaller cities is much bigger and Rank size distribution does not conform for holding Zipf's law.

In case of China studies that follows Zipf's law is carried out by (Gangopadhyay & Basu 2009) by using different sample compositions, minimum threshold in cities above 50,000 inhabitants in census year 2000. Secondly Ziqin found Zipf's law to hold in China by using 655 largest cities in year 2010. On the other hand, the studies whichstates that Zipf's law does not hold in China were done by Song & Zhang by using 665 largest cities in 1998 census. Additionally, Anderson & Ge used cities more than 100,000 inhabitants; Luckstead & Devadoss have taken 142 largest cities from thedata 1950 to 2010. Furthermore, a study on 657 largest cities in 2010 the Li, Wei & Ning demonstrate that Zipf law does not hold.

Literature includes the studies based on countries in different time periods, Indian case was studied in (Gangopadhyay & Basu, 2013) indicated results that city having an upper tail follows Zipf's Law. In the current study we obtained the average value of Pareto exponent as 0.689 which is less than 1, shows that Zipf's law rejected for case of India. In case of Malaysia, (Soo, 2005) using cities having 10,000 inhabitants and rejected City size distribution and in our study we got an average value of 0.670 which is less than 1 So Zipf's law does not hold in this case.

In case of Brazil studies that Size distribution of cities approaches to Zipf's Law as countries experience urbanization is carried out by (Moura Jr & Ribeiro, 2006) on Cities with 30000 inhabitants or more. (Matlaba et al., 2013), used the data of 185 largest functionally defined urban areas , (Ignazzi, 2015), took different data set from 1871-2010, and other study on Brazil the (Luckstead & Devadoss, 2014), used 58 largest cities from 1950 to 2010. Another study (Moura Jr & Ribeiro, 2006), used cities with 30,000 inhabitants and more. (Matlaba et al., 2013) used 185 largest functionally defined urban areas. Our finding for Brazil demonstrates that the average of 0.390 which far low than 1, and we reject the hypothesis and Zipf's law does not follow which is opposite to the previous studies. Morrocco's trend was reversed (Ezzahid & ElHamdani, 2015) and there was more uneven distribution in its cities with having more than 50,000 inhabitants. Our finding shows that the average value is 0.566 which is less than 1, and city size distribution does not conform Zipf's law.

Zipf's law (ZL) examination was also carried out on large scale by using 44 countries by (Rosen & Resnick, 1980), but the results do not confirm the Zipf's law (ZL) hold because the mean coefficient for these countries was 1.136 that reject the law. Our finding regarding developed countries the average is 0.867. Developing countries got average value 0.778 which also depicts the zipf law is rejected. For underdeveloped country the mean value is 0.827 which is also less than and depicts that zipf law does not hold in underdeveloped countries.

In case study of Pakistan carried out by (Arshad et al., 2018) examined that the Zipf's law does not hold at national level but City size distribution is more likely to follow at four province level explained the coherence property of urban system over time at different provinces. Sindh Province got value non-significant because there is uneven distribution due to Karachi which is mega city. Punjab's value was significant and equal to 1 in three census and urban system is becoming more uneven overtime. Khyber Pakhtunkhwa follows Zipf's law because urban population decreased due to departure of Hindus from this province in 1947. The estimated value Pareto exponent for Baluchistan is statistically significant in first three census, higher value in next two census. Landscape was not changed much during partition in 1947 but after that urban system got more even because of employment opportunities and political instability.

Our study examined that Pakistan has obtained non-significant value in four censuses when estimated on ranking wise city population, 137 administrative cities were taken as a sample of Pakistan and in last census 2017, FATA was not merged in KP. Pakistan has very less value than the unity demonstrates More concentration in urban System. More Inequality in distribution of cities population which means that small number of big or primate cities dominates the urban system and Pakistan has more hierarchy in urban system.

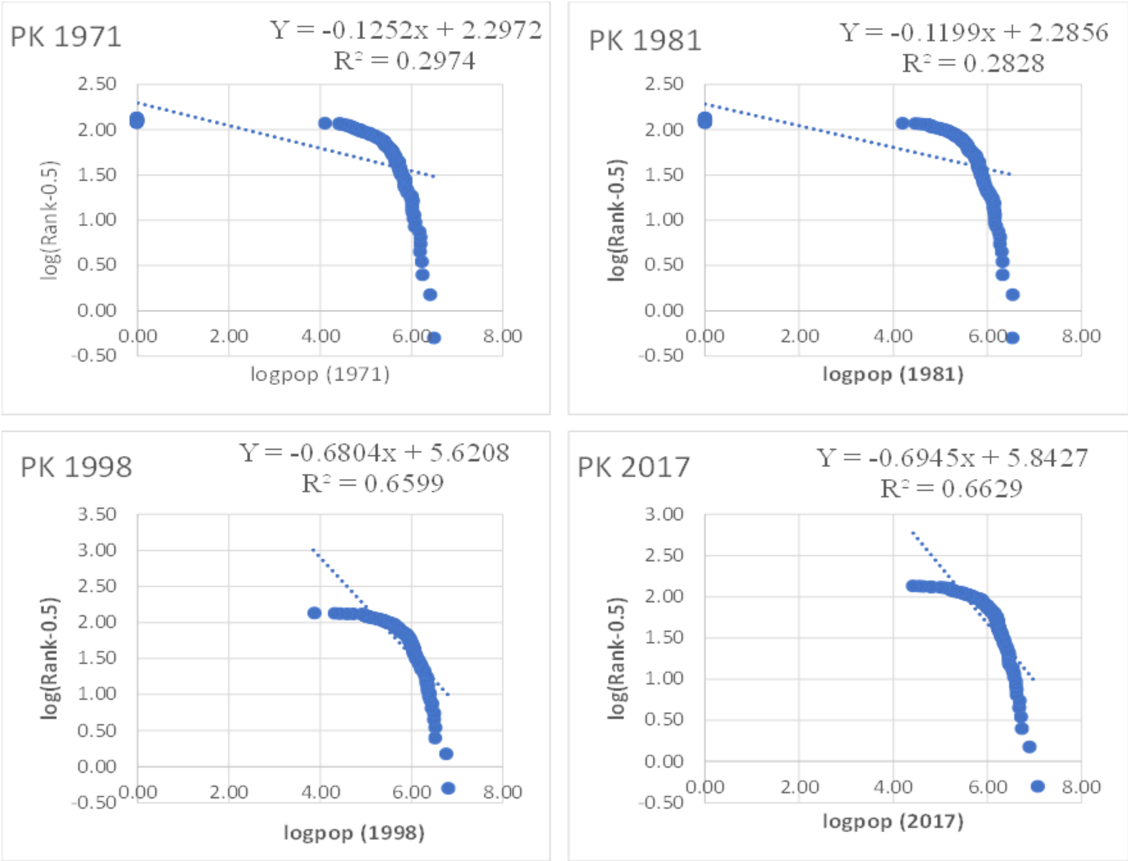


Figure 21 Decade wise Case of Pakistan

Qualitative work in this regard has been done by interviewing one of the officials of CDA that their policies are now diverted to High-rise building and projects like Ravi River, while the cities are more populated with more concentration towards big cities. So, they are scheduling to move big city by expanding the main city. New cities will be formed near the big old cities so that the construction cost will be less in developing a new city near to old one rather than developing a new city away from old city.

CHAPTER 5

CONCLUSION AND RECOMMENDATIONS

In the last chapter of the research, we concluded the overall findings of our research with justification and by comparison of the previous studies. After the conclusion about the overall findings of the thesis, we became able to provide some policy recommendations. In the end, we also indicate to the readers about the direction of untouched areas for future research that they can also put forth an effort in literature by exploring them.

5.1 Conclusion

Globally scholars believe from several disciplines of social sciences is that the so-called rank-size rule is a truthful representation of city size distribution (CSD). Many researchers construct the literature for providing empirical evidence to support this distribution. Our study investigated the validation of the RSD rule according to ZL. For testing the ZL for urban areas in all countries of the world, we have taken the new dataset of all countries comprising the city population of more than two census years. The indication from this dataset includes 235 countries with the data of their census year available, roughly speaking the census year of most of the countries are in $1980, 1990, 2000, 2010, \textit{and}\ 2018$. We have used the method of ordinary least square (OLS), Rank-minus-half rule, and Wald coefficient restriction test for estimating the β (PE). Every country has data for three to four census years; therefore, we have calculated the values for each country. We examined the comparison based on Human

Development Index. countries are divided into three categories. Firstly, High-Income Countries which are known as Developed nations which are having High-Income (HIC), Secondly medium-income countries are there which usually known as Developing nations having income near to very high and near too low. Lastly, the Low-income countries are known as Underdeveloped nations.

Countries with less estimated Pareto Exponent values are more hierarchical and more concentrated, lesser the value will show more of the hierarchies and concentration and there is more equality in dist. of population. Even dist. describes that there is more attention by population in small or medium-sized cities. Moreover, the values of data have less spread or less dispersion around the mean value of β. On the other hand, countries with a high β value or greater than 1 will show less hierarchy and attention of population and their values are more dispersed from the mean value. A higher value than β suggests there is a large portion of people in big cities as compared to small and medium-sized cities.

For most countries, In High Income Countries (HIC) the first category includes 65 nations. We have estimated that 16 out of 65 nations (approx. 24.6 %) of the developed countries have greater value than unity and for 49 nations the value estimated is less than unity. Developing nations (MICs) includes 90 countries out of which 16 countries have a value greater than 1, meanwhile 49 countries have a value lower than unity which is approximately 75.4% of the total countries. LIC's (β) values have a ratio of 32.2% for the higher side in which 10 out of 33 countries are included and the rest of 23 countries have a value less than 1 having a ratio of 74.2%.

Under developing nations (32.2%) have more ratio for having a higher value of β than Developed nations (24.6%), and HIC or developed countries are having more percentage of higher (β) value countries than medium-income countries or developing countries (17.78%). More the ratio is higher meaning more countries with higher exponent value have less hierarchy, less concentration with non-divergent cities and most importantly their large cities are more concentrated as compared to small-medium urban areas. Additionally, when taking the ratio of those countries with lesser value than 1, the MICs having more (82.2%) percentage meaning more countries having a value less than unity than HICs (75.4%) and then LICs (74.2%).

Zipf's Law (ZL) examination of (β) as Pareto exponent must be equal to 1 for conforming to the law which says that the first city or region population will be twice greater than the second one. Furthermore, there is an equal concentrated system of urban areas, population dist. ratio, and population dispersion ratio. Summarizing results, the estimated value of these countries is very near to unity which includes $Belgium$ 1.039, $Italy$ 1.005, France, and United Kingdom from the first group of developed nations, exponents values which are statistically near in developing nations are Bangladesh, Dominica, Philippines, Lebanon, Micronesia, Ukraine, and Vietnam. In under developing nations Benin, Mozambique, Yemen, DR Congo, and South Sudan are having exponent value near to unity.

Urban system development is more when your cities are evenly distributed in magnitude and many new cities should be formed. Development arises when new cities emerged but not when one big city is getting bigger than others. The development system of urban areas is still mainly under the pattern of the planned economy. However, most of the towns are governmentally progressed to be into smaller sized cities, while such development steps are quite similar in different urban areas. By the enlargement of the economical improvements, an economic model of market-leading economic progressively takes over the planned economic model. These abundant urban areas frequently developed however remaining many other urban areas failed to absorb the benefit of such chances of development and fell last in the average rate of growth at the national level.

5.2 Policy Recommendations

During this time, urban growth is still mainly governed by the economical planned economic model. Whereas, numerous areas are formally enhanced to be the minor urban areas, so the development rate in each city is comparable. An economical model which is market-oriented, progressively superseded the planned economic model as economic reforms were broadened and deepened. As a result, several cities grew fast, while many communities failed to capitalize on development possibilities and lagged the national average growth rate.

This research recommended that as seen in literature that we can use the 20-80 rule for Zipf's law to know what the basic problem with the countries is for not conforming to the law. In 1906, Italian economist Vilfredo Pareto determined that just 20% of the population in Italy controlled 80 per cent of the land. He expanded his investigation and discovered that unequal wealth distribution was consistent throughout Europe. The 80-20 rule was technically described as follows: the richest 20% of a country's population accounts for an estimated 80% of the country's wealth or total income. The advantage of using this law is to identify and determine the root cause of the problem, if the problem is identified then the organization and government can solve them on an urgent basis. It also shows us the cumulative impact of the problem.

Primate cities are those areas that are bigger than the normal size of the city. In contrast, economic richness and substantial sized population of urban areas might act in inverse way direction, which firmly helps in narrowing disparity within the size of the population of primate city: as the settlements of the system of the country and its wealthiness helps the many other urban areas to succeed within the large-sized city. Whereas it's not unexpected, then, that the largest divergence from the ZL, which is evaluated by this research through the size difference between the 1st and 2nd urban areas, which are found in ‘limited resource' countries such as Liberia, Barbados, and Eritrea. To sum-up, big by the area of land, very badly developed, and sparsely inhabited such countries where the first urban area serves as the capital at the national level are more likely to highly attain disparities in [2nd sized urban area/ 1st sized urban area] ratios. On the other hand, in densely population sized and largely urbanized huge-income countries, first and second urban areas are typical of comparable size. Countries' geographic location also plays a role: B-ratios generally adhering to Zipf's rule are most likely to be found in Europe, whilst less developed and sparsely urbanized countries elsewhere are more likely to exhibit more 'extreme' B-ratios. In the case of Pakistan, they should increase the newly constructed cities as compared to relay on big old cities or primate cities, so that the growth of all cities tends to increase which will overall maximize the growth of the country. After 1947, the refugees highly adjusted in small urban areas of Punjab, but policies of the government were on other big urban areas like Rawalpindi, Faisalabad, Lahore, has increased their size will cause the lower value of the exponent.

REFERENCES

Ades, A. F., & Glaeser, E. L. (1995). Trade and circuses: explaining urban giants. *The Quarterly journal of economics*, *110*(1), 195-227.

Arshad, S., Hu, S., & Ashraf, B. N. (2018). Zipf's law and city size distribution: A survey of the literature and future research agenda. *Physica A: Statistical Mechanics and its Applications*, *492*, 75-92.

Auerbach, F. (1913). Das gesetz der bevölkerungskonzentration. *Petermanns Geographische Mitteilungen*, *59,* 74-76.

Axtell, R., & Florida, R. (2000). Zipf. s Law of City Sizes: a Microeconomic Explanation far From Equilibrium. *Center on Social and Economic Dynamics The Brookings Institution and Johns Hopkins University & Carnegie Mellon University*.

Bairoch, P., Batou, J., & Pierre, C. (1988). *Population des villes européennes de 800 à 1850: banque de données et analyse sommaire des résultats (la)*. Librairie Droz.

Batten, D. F. (2001). Complex landscapes of spatial interaction. *The Annals of Regional Science*, *35*(1), 81-111.

Berry, B. J., & Okulicz-Kozaryn, A. (2012). The city size distribution debate: Resolution for US urban regions and megalopolitan areas. *Cities*, *29*, S17-S23.

Black, D., & Henderson, V. (1999). A theory of urban growth. *Journal of political economy*, *107*(2), 252-284.

Black, D., & Henderson, V. (2003). Urban evolution in the USA. *Journal of economic geography*, *3*(4), 343-372.

Brakman, S., Garretsen, H., & Van Marrewijk, C. (2001). *An introduction to geographical economics: Trade, location and growth*. Cambridge university press.

Brakman, S., Garretsen, H., Van Marrewijk, C., & Van Den Berg, M. (1999). The return of Zipf: Towards a further understanding of the rank-size distribution. *Journal of Regional Science*, *39*(1), 183-213.

Carroll, G. R. (1982). National city-size distributions: what do we know after 67 years of research? *Progress in Human Geography*, *6*(1), 1-43.

Champernowne, D. G. (1953). A model of income distribution. *The Economic Journal*, *63*(250), 318-351.

Cheshire, P. (1999). Trends in sizes and structures of urban areas. *Handbook of regional and urban economics*, *3*, 1339-1373.

Chlebus, E., & Ohri, R. (2005). Estimating parameters of the pareto distribution by means of zipf's law: application to internet research. GLOBECOM'05. IEEE Global Telecommunications Conference, 2005.,

Córdoba, J.-C. (2008). On the distribution of city sizes. *Journal of urban Economics*, *63*(1), 177-197.

Cordoba, J. (2001). Zipf's law: a case against scale economies. *Processed, University of Rochester*.

Cottineau, C. (2017). MetaZipf. A dynamic meta-analysis of city size distributions. *PloS one*, *12*(8), e0183919.

Cristelli, M., Batty, M., & Pietronero, L. (2012). There is more than a power law in Zipf. *Scientific reports*, *2*(1), 1-7.

D'Costa, J. (1994). Changes in urban structure in Bangladesh. *Urban Geography*, *15*(8), 698-719.

Dimitrova, Z. I., & Ausloos, M. (2015). Primacy analysis in the system of Bulgarian cities. *Open Physics*, *13*(1).

Dobkins, L. H., & Ioannides, Y. M. (2000). Dynamic Evolution of the Size Distribution. *Economics of Cities: Theoretical Perspectives*, 217.

Dubé, J., & Polèse, M. (2016). À propos du rôle de la taille dans la croissance urbaine: Une analyse pour 135 agglomérations canadiennes entre 1971 et 2011. *The Canadian Geographer/Le Géographe canadien*, *60*(4), 541-555.

Duran, H. E., & Özkan, S. P. (2015). Trade openness, urban concentration and city-size growth in Turkey. *Regional Science Inquiry*.

Duranton, G. (2002). City size distributions as a consequence of the growth process. *Available at SSRN 348580*.

Duranton, G. (2006). Some foundations for Zipf's law: Product proliferation and local spillovers. *Regional science and urban Economics*, *36*(4), 542-563.

Duranton, G. (2007). Urban evolutions: The fast, the slow, and the still. *American Economic Review*, *97*(1), 197-221.

Eaton, J., & Eckstein, Z. (1997). Cities and growth: Theory and evidence from France and Japan. *Regional science and urban Economics*, *27*(4-5), 443-474.

Eeckhout, J. (2004). Gibrat's law for (all) cities. *American Economic Review*, *94*(5), 1429-1451.

Ezzahid, E., & ElHamdani, O. (2015). Zipf'S Law In The Case Of Moroccan Cities. *Review of Urban & Regional Development Studies*, *27*(2), 118-133.

Feenberg, D. R., & Poterba, J. M. (1993). Income inequality and the incomes of very high-income taxpayers: evidence from tax returns. *Tax policy and the economy*, *7*, 145-177.

Fonseca, I., & Tartar, L. (1989). The gradient theory of phase transitions for systems with two potential wells. *Proceedings of the Royal Society of Edinburgh Section A: Mathematics*, *111*(1-2), 89-102.

Fujita, M., Krugman, P. R., & Venables, A. (1999). *The spatial economy: Cities, regions, and international trade*. MIT press.

Gabaix, X. (1999). Zipf's law for cities: an explanation. *The Quarterly journal of economics*, *114*(3), 739-767.

Gabaix, X., & Ibragimov, R. (2011). Rank− 1/2: a simple way to improve the OLS estimation of tail exponents. *Journal of Business & Economic Statistics*, *29*(1), 24-39.

Gabaix, X., Ioannides, Y., Henderson, J., & Thisse, J. (2004). Handbook of regional and urban economics. *North-Holland, Amsterdam*, *2341*.

Gabaix, X., & Ioannides, Y. M. (2004). The evolution of city size distributions. In *Handbook of regional and urban economics* (Vol. 4, pp. 2341-2378). Elsevier.

Gan, L., Li, D., & Song, S. (2006). Is the Zipf law spurious in explaining city-size distributions? *Economics Letters*, *92*(2), 256-262.

Gangopadhyay, K., & Basu, B. (2009). City size distributions for India and China. *Physica A: Statistical Mechanics and its Applications*, *388*(13), 2682-2688.

Gangopadhyay, K., & Basu, B. (2013). Evolution of Zipf's law for Indian urban agglomerations vis-à-Vis Chinese urban agglomerations. In *Econophysics of Systemic Risk and Network Dynamics* (pp. 119-129). Springer.

Giesen, K., & Südekum, J. (2011). Zipf's law for cities in the regions and the country. *Journal of economic geography*, *11*(4), 667-686.

Giesen, K., Zimmermann, A., & Suedekum, J. (2010). The size distribution across all cities–double Pareto lognormal strikes. *Journal of urban Economics*, *68*(2), 129-137.

Gligor, L., & Gligor, M. (2008). The fractal city theory revisited: new empirical evidence from the distribution of Romanian cities and towns. *Nonlinear dynamics, psychology, and life sciences*, *12*(1), 15-28.

Henderson, V. (2002). Urbanization in developing countries. *The World Bank Research Observer*, *17*(1), 89-112.

Hsu, W. T. (2012). Central place theory and city size distribution. *The Economic Journal*, *122*(563), 903-932.

Ignazzi, C. A. (2015). The Brazilian Urban System: the trajectories of Brazilian cities between general dynamics and specific peculiarities. *Cybergeo: European Journal of Geography*.

Ioannides, Y. M., & Overman, H. G. (2003). Zipf's law for cities: an empirical examination. *Regional science and urban Economics*, *33*(2), 127-137.

Jadwiga, W. (1992). Bangladesh Cities according to the" Rank-Size Rule". *Miscellanea Geographica. Regional Studies on Development*, *5*(1), 195-200.

Jefferson, M. (1989). Why geography? The law of the primate city. *Geographical review*, *79*(2), 226-232.

Jiang, B., & Jia, T. (2011). Zipf's law for all the natural cities in the United States: a geospatial perspective. *International Journal of Geographical Information Science*, *25*(8), 1269-1281.

Jiang, B., & Liu, X. (2012). Scaling of geographic space from the perspective of city and field blocks and using volunteered geographic information. *International Journal of Geographical Information Science, 26*(2), 215-229.

Jiang, B., & Yin, J. (2014). Ht-index for quantifying the fractal or scaling structure of geographic features. *Annals of the Association of American Geographers, 104*(3), 530-540.

Jiang, B., Yin, J., & Liu, Q. (2015). Zipf's law for all the natural cities around the world. *International Journal of Geographical Information Science, 29*(3), 498-522.

Josic, H., & Bašić, M. (2018). Reconsidering Zipf's law for regional development: The case of settlements and cities in Croatia. *Miscellanea Geographica. Regional Studies on Development*, *22*(1), 22-30.

Knudsen, T. (2001). Zipf's law for cities and beyond: The case of Denmark. *American Journal of Economics and Sociology, 60*(1), 123-146.

Kolomak, E. (2014). Development of Russian urban system: tendencies and determinants. *Voprosy Economiki*, *10*.

Kosmopoulou, G., Buttry, N., Johnson, J., & Kallsnick, A. (2007). Suburbanization and the rank-size rule. *Applied Economics Letters, 14*(1), 1-4.

Krugman, P. (1996). Confronting the mystery of urban hierarchy. *Journal of the Japanese and International economies, 10*(4), 399-418.

Lalanne, A. (2014). Zipf's law and Canadian urban growth. *Urban Studies*, *51*(8), 1725-1740.

Lanaspa, L., Pueyo, F., & Sanz, F. (2003). The evolution of Spanish urban structure during the twentieth century. *Urban Studies*, *40*(3), 567-580.

Le Gallo, J., & Chasco, C. (2008). Spatial analysis of urban growth in Spain, 1900–2001. *Empirical economics, 34*(1), 59-80.

Lee, S., & Li, Q. (2013). Uneven landscapes and city size distributions. *Journal of urban Economics*, *78*, 19-29.

Li, W. (2002). Zipf's Law everywhere. *Glottometrics, 5*, 14-21.

Li, W., & Yang, Y. (2002). Zipf's law in importance of genes for cancer classification using microarray data. *Journal of Theoretical Biology*, *219*(4), 539-551.

Luckstead, J., & Devadoss, S. (2014). A comparison of city size distributions for China and India from 1950 to 2010. *Economics Letters*, *124*(2), 290-295.

Malevergne, Y., Pisarenko, V., & Sornette, D. (2011). Testing the Pareto against the lognormal distributions with the uniformly most powerful unbiased test applied to the distribution of cities. *Physical Review E*, *83*(3), 036111.

Matlaba, V. J., Holmes, M. J., McCann, P., & Poot, J. (2013). A Century of the Evolution of the Urban System in Brazil. Review of Urban & Regional Development Studies: Journal of the Applied Regional Science Conference,

Mitzenmacher, M. (2004). A brief history of generative models for power law and lognormal distributions. *Internet mathematics*, *1*(2), 226-251.

Moskowitz, B. (1959). *Psychological aspects of energetics* The University of Oklahoma.].

Moura Jr, N. J., & Ribeiro, M. B. (2006). Zipf law for Brazilian cities. *Physica A: Statistical Mechanics and its Applications*, *367*, 441-448.

Newman, M. E. (2005). Power laws, Pareto distributions and Zipf's law. *Contemporary physics*, *46*(5), 323-351.

Nishiyama, Y., Osada, S., & Sato, Y. (2008). OLS estimation and the t test revisited in rank-size rule regression. *Journal of Regional Science*, *48*(4), 691-716.

Nitsch, V. (2005). Zipf zipped. *Journal of urban Economics*, *57*(1), 86-100.

Nota, F., & Song, S. (2012). Further analysis of the Zipf's law: Does the rank-size rule really exist? *Journal of Urban Management*, *1*(2), 19-31.

Pérez-Campuzano, E., Guzmán-Vargas, L., & Angulo-Brown, F. (2015). Distributions of city sizes in Mexico during the 20th century. *Chaos, Solitons & Fractals*, *73*, 64-70.

Rastvortseva, S., & Manaeva, I. (2016). Zipf's law appearance in the Russian cities. *Regional Science Inquiry*, *1*, 51-59.

Reed, W. J. (2002). On the rank-size distribution for human settlements. *Journal of Regional Science*, *42*(1), 1-17.

Rosen, K. T., & Resnick, M. (1980). The size distribution of cities: an examination of the Pareto law and primacy. *Journal of urban Economics*, *8*(2), 165-186.

Rossi-Hansberg, E., & Wright, M. L. (2003). Urban structure and growth. Processed. In: Stanford University.

Rozenfeld, H. D., Rybski, D., Andrade, J. S., Batty, M., Stanley, H. E., & Makse, H. A. (2008). Laws of population growth. *Proceedings of the National Academy of Sciences*, *105*(48), 18702-18707.

Rozenfeld, H. D., Rybski, D., Gabaix, X., & Makse, H. A. (2011). The area and population of cities: New insights from a different perspective on cities. *American Economic Review*, *101*(5), 2205-2225.

Sarabia, J. M., & Prieto, F. (2009). The Pareto-positive stable distribution: A new descriptive model for city size data. *Physica A: Statistical Mechanics and its Applications*, *388*(19), 4179-4191.

Sassen, S. (1991). Jlie Global City: New York, London, Tokio. In: New Jersey. Princeton University press.

Schmidheiny, K., & Suedekum, J. (2015). The pan-European population distribution across consistently defined functional urban areas. *Economics Letters*, *133*, 10-13.

Shiode, N., & Batty, M. (2000a). Power law distribution in real and virtual worlds.

Shiode, N., & Batty, M. (2000b). Power law distributions in real and virtual worlds.

Simon, H. A. (1955). On a class of skew distribution functions. *Biometrika*, *42*(3/4), 425-440.

Sinclair, S. J., Blais, M. A., Gansler, D. A., Sandberg, E., Bistis, K., & LoCicero, A. (2010). Psychometric properties of the Rosenberg Self-Esteem Scale: Overall and across demographic groups living within the United States. *Evaluation & the health professions*, *33*(1), 56-80.

Soo, K. T. (2004). *Zipf's law for cities: a cross country investigation*. Centre for Economic Performance, London School of Economics and Political ….

Soo, K. T. (2005). Zipf's Law for cities: a cross-country investigation. *Regional science and urban Economics*, *35*(3), 239-263.

Soo, K. T. (2007). Zipf's Law and urban growth in Malaysia. *Urban Studies*, *44*(1), 1-14.

Tachimori, Y., & Tahara, T. (2002). Clinical diagnoses following ZIPF's law. *Fractals*, *10*(03), 341-351.

Veneri, P. (2016). City size distribution across the OECD: Does the definition of cities matter? *Computers, Environment and Urban Systems*, *59*, 86-94.

Weber, C. (2003). Urban agglomeration delimitation using remote sensing data. *Remote sensing and urban analysis: GISDATA*, *9*, 131-146.

Yang, X. (2011). *Urban remote sensing: monitoring, synthesis and modeling in the urban environment*. John Wiley & Sons.

Zipf, G. K. (1949). Human behaviour and the principle of least-effort. Cambridge MA edn. *Reading: Addison-Wesley*.

Appendix

TABLE I: COEFFICIENT VALUES FOR DEVELOPED COUNTRIES

HDI (Developed)	1st Census (β)	2nd Census (β)	3rd Census (β)	4th Census (β)	AVERAGE (β)	Last Census (β)
Norway	0.798	0.778	0.756	0.736	0.767	0.736
Ireland	1.426	1.380	1.377	1.384	1.392	1.384
Switzerland	0.478	0.519	0.526	-	0.508	0.526
Hong Kong	0.374	0.732	0.601	0.423	0.533	0.423
Iceland	0.584	0.585	0.889	0.563	0.655	0.563
Germany	**1.100**	**1.099**	**1.093**	-	**1.097**	**1.093**
Sweden	0.898	0.893	0.863	0.837	0.865	0.837
Australia	0.741	0.679	0.671	0.664	0.689	0.664
Netherlands	1.135	1.111	1.108	1.103	1.114	1.103
Denmark	0.687	0.684	0.677	-	0.683	0.677
Finland	0.799	0.789	0.774	0.751	0.778	0.751
Singapore	0.955	0.676	0.812	0.856	0.825	0.856
United Kingdom	0.947	0.963	0.963	-	0.958	0.963
Belgium	1.024	1.037	1.048	1.047	1.039	1.047
New Zealand	0.717	0.726	0.754	0.774	0.743	0.774
Canada	**0.785**	**0.767**	**0.758**	**0.746**	**0.764**	**0.746**
United States	**0.673**	**0.652**	**0.627**	**0.608**	**0.640**	**0.608**
Austria	1.220	1.197	1.179	1.169	1.191	1.169
Israel	0.652	0.706	0.726	0.730	0.704	0.730
Japan	0.914	0.879	0.837	0.818	0.862	0.818
Liechtenstein	0.692	0.710	0.714	0.732	0.712	0.732
Slovenia	1.043	1.075	1.136	1.141	1.099	1.141
Luxembourg	1.265	0.784	0.801	0.802	0.913	0.802
South Korea	0.901	0.824	0.807	0.796	0.832	0.796
Spain	0.889	0.891	0.889	-	0.890	0.889
France	0.934	0.954	0.978	0.994	0.965	0.994
Czech Republic	1.886	1.869	1.836	-	1.864	1.836
Malta	0.751	0.845	0.915	0.914	0.856	0.914
Estonia	0.689	0.717	0.744	0.727	0.719	0.727
Italy	1.015	1.011	1.003	0.989	1.005	0.989
UAE	0.568	0.814	0.932	-	0.771	0.932
Greece	0.567	0.573	0.584	-	0.575	0.584
Cyprus	0.691	0.759	0.803	0.863	0.779	0.863
Lithuania	0.858	0.861	0.854	-	0.858	0.854
Poland	1.592	1.582	1.543	-	1.572	1.543

Andorra	0.482	0.583	0.622	0.669	0.589	0.669
Latvia	1.201	1.276	1.276	-	1.251	1.276
Portugal	0.738	0.713	0.696	-	0.716	0.696
Slovakia	1.338	1.325	1.307	1.275	1.311	1.275
Hungary	1.314	1.279	1.277	1.237	1.277	1.237
Saudi Arabia	0.636	0.634	-	-	0.635	0.634
Bahrain	0.974	0.182	1.120	-	0.759	1.120
Chile	0.575	0.567	0.562	0.557	0.565	0.557
Croatia	0.657	0.683	0.687	0.692	0.680	0.692
Qatar	0.811	0.781	0.711	0.626	0.732	0.626
Argentina	0.597	0.593	0.588	-	0.593	0.588
Brunei	0.574	0.557	0.507	0.524	0.541	0.524
Montenegro	0.966	0.885	0.800	0.734	0.846	0.734
Romania	1.200	1.954	1.965	1.903	1.755	1.903
Palau	0.838	0.646	0.611	0.596	0.673	0.596
Kazakhstan	0.936	0.856	0.796	-	0.863	0.796
Russia	**1.081**	**1.120**	**1.172**	**1.189**	**1.141**	**1.189**
Belarus	1.260	1.144	1.011	1.956	1.343	1.956
Turkey	0.664	0.644	0.617	-	0.642	0.617
Uruguay	1.099	1.075	1.061	1.059	1.074	1.059
Bulgaria	0.729	0.702	0.696	-	0.709	0.696
Panama	0.671	0.663	0.657	-	0.664	0.657
Bahamas	0.471	0.448	0.428	0.466	0.453	0.466
Barbados	0.613	0.119	1.057	1.002	0.698	1.002
Oman	0.629	0.639	0.565	-	0.611	0.565
Georgia	1.011	0.958	0.893	0.884	0.937	0.884
Costa Rica	0.863	0.847	0.843	-	0.851	0.843
Malaysia	**0.662**	**0.678**	**-**	**-**	**0.670**	**0.678**
Kuwait	0.610	0.631	0.828	0.834	0.726	0.834
Serbia	0.905	0.889	0.847	0.816	0.864	0.816

TABLE II: HIC'S COUNTRIES B VALUE SUMMARY.

HDI (Developed Countries)	β	Census 1 (β)	Census 2 (β)	Census 3 (β)	Census 4 (β)	Mean (β)	Last Census (β)
Czech Republic	Higher	1.886	1.869	1.836	-	1.864	1.836
Romania	Higher	1.200	1.954	1.965	1.903	1.755	1.903
Poland	Higher	1.592	1.582	1.543	-	1.572	1.543
Ireland	Higher	1.426	1.380	1.377	1.384	1.392	1.384
Belarus	Higher	1.260	1.144	1.011	1.956	1.343	1.956
Belgium	**Near to 1**	**1.024**	**1.037**	**1.048**	**1.047**	**1.039**	**1.047**
Italy	**Near to 1**	**1.015**	**1.011**	**1.003**	**0.989**	**1.005**	**0.989**
France	**Near to 1**	**0.934**	**0.954**	**0.978**	**0.994**	**0.965**	**0.994**
UK	**Near to 1**	**0.947**	**0.963**	**0.963**	**-**	**0.958**	**0.963**
Greece	Lower	0.567	0.573	0.584	-	0.575	0.584
Chile	Lower	0.575	0.567	0.562	0.557	0.565	0.557
Brunei	Lower	0.574	0.557	0.507	0.524	0.541	0.524
Hong Kong	Lower	0.374	0.732	0.601	0.423	0.533	0.423
Bahamas	Lower	0.471	0.448	0.428	0.466	0.453	0.466

Table III: COEFFICIENT VALUES FOR DEVELOPING COUNTRIES.

HDI (DEVELOPING COUNTRIES)	1st Census (β)	2nd Census (β)	3rd Census (β)	4th Census (β)	AVERAGE (β)	Last Census (β)
ST & Grenadines	1.571	1.572	1.650	1.610	1.601	1.610
Samoa	1.315	1.347	1.387	1.378	1.357	1.378
Zambia	1.238	1.231	1.196		1.222	1.196
Seychelles	1.073	1.236	1.228	1.305	1.211	1.305
Thailand	1.231	1.180	-	-	1.206	1.180
Tunisia	1.251	1.194	1.165	-	1.203	1.165
Mongolia	1.389	1.178	1.038	1.002	1.152	1.002
Laos	1.056	1.084	1.209	1.205	1.139	1.205
Grenada	1.139	1.134	1.101	1.102	1.119	1.102
Saint, Kitts & Nevis	1.010	1.042	1.077	1.083	1.053	1.083
Papua New Guinea	1.059	1.036	-	-	1.048	1.036
Uzbekistan	0.972	0.975	1.175	-	1.041	1.175
Bangladesh	1.024	1.018	1.051	-	1.031	1.051
Vanuatu	1.050	1.020	0.995		1.022	0.995
Dominica	0.978	1.031	1.041	-	1.017	1.041
Philippines	1.021	1.009	1.006	-	1.012	1.006
Lebanon	0.792	1.069	1.092	1.031	0.996	1.031
Micronesia	0.798	1.028	1.072	1.071	0.992	1.071
Ukraine	0.984	0.977	0.951	-	0.971	0.951
Vietnam	0.955	0.982	-	-	0.969	0.982

Cuba	0.963	0.964	0.955	-	0.961	0.955
Syria	0.890	0.886	1.082	-	0.953	1.082
Albania	0.948	0.864	0.981	0.891	0.921	0.891
Saint Lucia	0.918	0.914	-	-	0.916	0.914
Libya	0.922	0.893	-	-	0.908	0.893
Mauritius	0.877	0.884	0.925	0.928	0.904	0.928
Botswana	0.817	0.891	0.998	-	0.902	0.998
Suriname	0.975	0.836	0.831	-	0.881	0.831
North Macedonia	0.936	0.861	0.847	0.832	0.869	0.832
Nepal	0.898	0.853	0.875	0.839	0.866	0.839
Solomon Islands	1.003	0.693	-	-	0.848	0.693
Cape Verde	0.745	0.818	0.928	0.891	0.846	0.891
South Africa	0.847	0.842	0.844	0.837	0.843	0.837
Indonesia	0.844	0.812	0.828	0.865	0.837	0.865
Moldova	0.836	0.842	0.821	-	0.833	0.821
Belize	0.781	0.815	0.792	0.833	0.805	0.833
Maldives	0.736	0.814	0.835	0.799	0.796	0.799
Nicaragua	0.770	0.804	0.786	-	0.787	0.786
China	**0.736**	**0.757**	**0.818**	**-**	**0.770**	**0.818**
Colombia	0.767	0.777	0.773	0.755	0.768	0.755
Paraguay	0.777	0.769	0.759	0.750	0.764	0.750
Honduras	0.770	0.755	0.748	-	0.758	0.748
Guyana	0.677	0.763	0.786	-	0.742	0.786
Palestine	0.706	0.729	0.723	0.723	0.720	0.723
Pakistan	**0.717**	**0.774**	**0.680**	**0.694**	**0.716**	**0.694**

Antigua and Barbuda	0.677	0.788	0.682	-	0.716	0.682
Algeria	0.668	0.641	0.750	0.793	0.713	0.793
Iran	0.684	0.704	0.738	-	0.709	0.738
El Salvador	0.705	0.718	0.685	-	0.703	0.685
Peru	0.732	0.709	0.666	-	0.702	0.666
Kiribati	0.594	0.781	0.728	0.698	0.700	0.698
Ghana	0.516	0.502	1.073	-	0.697	1.073
Mexico	0.694	0.696	0.689	-	0.693	0.689
Iraq	0.693	0.693	-	-	0.693	0.693
Dominican Republic	0.707	0.632	0.694	0.734	0.692	0.734
Azerbaijan	0.695	0.649	0.682	0.732	0.690	0.732
India	0.708	0.679	0.680	-	0.689	0.680
Angola	0.689	0.689	-	-	0.689	0.689
Trinid. & Tobago	0.672	0.703	-	-	0.688	0.703
Bosnia & Her.	0.697	0.678	0.673	-	0.683	0.673
Venezuela	0.679	0.682	0.684	-	0.682	0.684
Guatemala	0.570	0.717	0.756	-	0.681	0.756
Ecuador	0.683	0.678	0.673	-	0.678	0.673
Bhutan	0.627	0.696	0.624	0.674	0.655	0.674
Comoros	0.653	0.643	-	-	0.648	0.643
Turkmenistan	0.629	0.623	0.667	-	0.640	0.667
Armenia	0.639	0.640	0.630	-	0.636	0.630
Jordan	0.641	0.657	0.603	-	0.634	0.603
Egypt	0.489	0.561	0.786	-	0.612	0.786

Cameroon	0.564	0.619	0.622	-	0.602	0.622
Marshall Islands	0.769	0.564	0.519	0.549	0.600	0.549
East Timor	0.614	0.716	0.459	-	0.596	0.459
Namibia	0.440	0.629	0.624	0.674	0.592	0.674
Congo	0.617	0.489	0.668	-	0.591	0.668
Kyrgyzstan	0.601	0.603	0.578	0.564	0.587	0.564
S.T and Príncipe	0.397	0.772	-	-	0.585	0.772
Jamaica	0.577	0.569	0.562	0.558	0.567	0.558
Morocco	**0.517**	**0.579**	**0.603**	**-**	**0.566**	**0.603**
Cambodia	0.559	0.563	0.554	-	0.559	0.554
Fiji	0.561	0.527	-	-	0.544	0.527
Sri Lanka	0.535	0.504	0.579	-	0.539	0.579
Myanmar	0.513	0.556	-	-	0.535	0.556
Equatorial Guinea	0.447	0.462	0.763	0.434	0.527	0.434
Bolivia	0.478	0.522	0.536	0.570	0.527	0.570
Kenya	0.486	0.455	0.493	0.544	0.495	0.544
Tonga	0.500	0.475	-	-	0.488	0.475
Gabon	0.768	0.665	-	-	0.478	0.000
Eswatini	0.453	0.465	0.471	-	0.463	0.471
Zimbabwe	0.442	0.459	0.485	0.438	0.456	0.438
Brazil	**0.364**	**0.415**	**-**	-	**0.390**	**0.415**

TABLE IV: MIC’S COUNTRIES Β VALUE SUMMARY.

HDI (Developing Countries)	**β**	**Census 1 (β)**	**Census 2 (β)**	**Census 3 (β)**	**Census 4 (β)**	**AVERAGE (β)**	**Last Census (β)**
St. Vint. and Grenadines	Higher	1.571	1.572	1.650	1.610	1.601	1.610
Samoa	Higher	1.315	1.347	1.387	1.378	1.357	1.378
Zambia	Higher	1.238	1.231	1.196	-	1.222	1.196
Seychelles	Higher	1.073	1.236	1.228	1.305	1.211	1.305
Thailand	Higher	1.231	1.180	-	-	1.206	1.180
Bangladesh	Near to 1	1.024	1.018	1.051	-	1.031	1.051
Vanuatu	Near to 1	1.050	1.020	0.995	-	1.022	0.995
Philippines	Near to 1	1.021	1.009	1.006	-	1.012	1.006
Lebanon	Near to 1	0.792	1.069	1.092	1.031	0.996	1.031
Micronesia	Near to 1	0.798	1.028	1.072	1.071	0.992	1.071
Ukraine	Near to 1	0.984	0.977	0.951	-	0.971	0.951
Vietnam	Near to 1	0.955	0.982	-	-	0.969	0.982
Kenya	Lower	0.486	0.455	0.493	0.544	0.495	0.544
Tonga	Lower	0.500	0.475	-	-	0.488	0.475
Gabon	Lower	0.768	0.665	-	0.000	0.478	0.000
Eswatini	Lower	0.453	0.465	0.471	-	0.463	0.471
Zimbabwe	Lower	0.442	0.459	0.485	0.438	0.456	0.438

Table V: COEFFICIENT VALUES FOR UNDER-DEVELOPED COUNTRIES

HDI (UNDER-DEVELOPED)	1st Census (β)	2nd Census (β)	3rd Census (β)	AVERAGE (β)	Last Census (β)
Uganda	1.181	1.533	2.113	1.609	2.113
Rwanda	1.370	1.368		1.369	1.368
Madagascar	1.084	1.189	1.187	1.153	1.187
Nigeria	1.008	1.191	1.188	1.129	1.188
Ethiopia	1.082	1.047	1.238	1.122	1.238
Tanzania	1.049	0.952	1.268	1.090	1.268
Eritrea	1.080	1.042		1.061	1.042
Benin	1.053	1.033	1.017	1.034	1.017
Mozambique	1.022	1.021	1.026	1.023	1.026
Yemen	0.989	1.017	1.017	1.008	1.017
DR Congo	1.021	0.924		0.973	0.924
South Sudan	0.966	0.958		0.962	0.958
Mali	0.964	0.954		0.959	0.954
Burundi	2.044	0.349	0.410	0.934	0.410
Mauritania	0.935	0.885		0.910	0.885
Sudan	0.879	0.872	0.970	0.907	0.970
Gambia	0.761	0.965		0.863	0.965
Djibouti	0.705	0.994		0.850	0.994
Malawi	0.737	0.749	0.741	0.742	0.741
Guinea-Bissau	0.578	1.069	0.541	0.729	0.541
Central African Republic	0.652	0.690		0.671	0.690
Ivory Coast	0.621	0.590	0.620	0.610	0.620
Guinea	0.448	0.462	0.763	0.558	0.763
Lesotho	0.473	0.555	0.643	0.557	0.643
Burkina Faso	0.538	0.532	0.537	0.536	0.537
Haiti	0.446	0.497	0.616	0.520	0.616
Niger	0.512	0.513	0.516	0.514	0.516
Senegal	0.489	0.502	0.536	0.509	0.536
Liberia	0.535	0.474		0.505	0.474
Togo	0.500	0.475		0.488	0.475
Chad	0.453	0.508		0.481	0.508
Sierra Leone	0.463	0.466	0.471	0.467	0.471
Afghanistan	0.440	0.445	0.452	0.446	0.452

TABLE VI: LIC'S COUNTRIES B VALUE SUMMARY.

HDI (Developing Countries)	β	Census 1 (β)	Census 2 (β)	Census 3 (β)	AVERAGE (β)	Last Census (β)
Uganda	>1	1.181	1.533	2.113	1.609	2.113
Rwanda	>1	1.370	1.368		1.369	1.368
Madagascar	>1	1.084	1.189	1.187	1.153	1.187
Benin	Near to 1	1.053	1.033	1.017	1.034	1.017
Mozambique	Near to 1	1.022	1.021	1.026	1.023	1.026
Yemen	Near to 1	0.989	1.017	1.017	1.008	1.017
DR Congo	Near to 1	1.021	0.924		0.973	0.924
South Sudan	Near to 1	0.966	0.958		0.962	0.958
Chad	<1	0.453	0.508		0.481	0.508
Sierra Leone	<1	0.463	0.466	0.471	0.467	0.471
Afghanistan	<1	0.440	0.445	0.452	0.446	0.452

Table IX: Remaining Countries because of Statistical regulatory of Zipf's law

Sr.	Country	Sr.	Country
1	Azerbaijan	26	Sint Maarten
2	Tajikistan	27	Monaco
3	Denmark	28	Turks and Caicos
4	Puerto Rico	29	Saint Martin
5	Réunion	30	San Marino
6	Macao	31	Gibraltar
7	Western Sahara	32	British Virgin Islands
8	Guadeloupe	33	Caribbean Netherlands
9	Martinique	34	Cook Islands
10	French Guiana	35	Anguilla
11	New Caledonia	36	Tuvalu
12	French Polynesia	37	Wallis & Futuna
13	Mayotte	38	Nauru
14	Sao Tome & Principe	39	Saint Barthelemy
15	Guam	40	Saint Helena
16	Curaçao	41	Saint Pierre & Miquelon
17	Aruba	42	Montserrat
18	U.S. Virgin Islands	43	Falkland Islands
19	Isle of Man	44	Niue
20	Cayman Islands	45	Tokelau
21	Bermuda	46	Holy See
22	Northern Mariana Islands	47	Greenland
23	Greenland	48	Sint Maarten
24	American Samoa	49	Monaco
25	Faeroe Islands		

Note: These Countries were having very small data with little number of census, therefore Zipf's law was applicable because it has a statistical regularity that Zipf's law cannot be applicable on small data as we have mentioned in Introduction section under the heading of background.