

# International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 12, December 2015

## Empirical Study on the Relationship between Zipf's law and Classical Pareto Distribution

R. Selvam

Assistant Professor, Department of Statistics, Government Arts College (Autonomous), Salem, Tamil Nadu, India

**ABSTRACT:** In this article, the relationship between zipf's law and classical pareto distribution is discussed. This study uses 1991, 2001 and 2011 census data from the primary census abstract of Tamil Nadu state. The population of all districts in Tamil Nadu state was considered for the empirical analysis. As per this study, the actual distributions of district size of three periods are inconsistent with those of the rank size rule. The value of 'a' in 1991 is 1.108 which is the smallest in three periods. Where as that of 2011 is the largest, 1.418. This means that the distribution of the rank size of districts in 1991 is the closest to the distribution of the rank size rule, while the distribution of the districts in 2011 is the farthest away from the distribution of the rank size rule. The values of 'a' in 1991, 2001 and 2011 are respectively obtained as 1.108, 1.365 and 1.418 which are larger than the value of 'a' in the rank size rule, 1.0. Therefore, all slopes of three periods are steeper than the slope of the rank size rule. This result denotes that the expected district sizes based on the rank size rule are larger than the real size of the districts in Tamil Nadu. The values of slope in actual distribution of districts have been larger and larger over time. Therefore the adoption of rank size rule to the distribution of districts size of Tamil Nadu seems to be less adequate over time.

**KEYWORDS:** District size distribution, Pareto distribution, Regression equation, Rank-size distribution, Zipf's law.

### I. INTRODUCTION

Cities come in different sizes, one enduring line of research has been in describing the size distribution of cities. Population growth is closely related to the size distribution of cities. The zipf's law, is used to describe the size distribution of cities, which states that the size distribution of cities follows a simple Pareto distribution with shape parameter equal to one. The idea that the size distribution of cities can be approximated by a Pareto distribution has fascinated social scientists Auerbach (1913) first proposed it. Over the years, Auerbach's basic proposition has been refined by many others. Most studies have concentrated on rank size distributions, where cities are ranked by population size and the inverse of the rank explains a city's population relative to the population of the largest city. Zipf (1949) popularized a specific distribution in which a population of city is equal to the inverse of its rank multiplied by the population of the largest city. Hence, the term "Zipf's Law" is frequently used to refer to the idea that city sizes follow a Pareto distribution. Zipf's Law states that not only does the size distribution of cities follow a Pareto distribution, but that the distribution has a shape parameter equal to 1. Thus the rank of city sizes is proportional to the inverse of their size.

In empirical studies of the city size distribution, when cities are ordered by population size, regressing the logarithmic of their rank on the logarithm of their population has resulted in a slope coefficient close to minus one in so many instances that the phenomena has acquired the status of the eponymous zipf's law. Generally, the Pareto exponents generated from the rank-size regression are not necessarily equal to unity, and this so called "rank-size rule" is believed to be applicable to almost all countries around world. An attempt has been made to study the district size of population between the census year 1991 and 2011 of Tamil Nadu and find whether the zipf's law relates to Pareto distribution.

# International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 12, December 2015

## II. SOURCE OF DATA

In this present article data on districts of Tamil Nadu State between 1991 and 2011 census have been used for the empirical study.

## III. SCOPE OF THE STUDY

The scope of this article is to study whether the Zipf's law relates to Pareto distribution using the census data of all districts in Tamil Nadu between 1991 and 2011.

## IV. METHODOLOGY

In order to investigate interrelationship between the Pareto distribution and Zipf's law for the districts of Tamil Nadu, regression equations for three periods (1991, 2001 and 2011 census) will be derived and compare the slopes of rank-size lines between the actual distribution of district size and the theoretical rank-size rule or Zipf's law. That is the slopes of the regression equations of districts of Tamil Nadu will be compared with the slope of the rank size rule (-1.0). The relationship between population size and population rank is inherently negative. This is obvious and accounts for much of the strong fit, but alone doesn't account for the multiplicative and convex (to the origin) relationship.

### 4.1 PARETO MODEL

The Pareto distribution was first proposed as a model for the distribution of income. It is also used as a model for the distribution of city population in the given area. The Pareto distribution has been described as follows

The district size ( $x$ ) is assumed to follow Pareto distribution  $f(x; a)$ , where  $f(x; a)$  is described as,

$f(x; a) = \frac{ak^a}{x^{a+1}}$ ,  $a > 0$ ;  $x \geq k$ , Where,  $k$  is the threshold district size. Parameters 'a' and 'k' are estimated by using the method of Maximum likelihood (Rao C.R. 1973) and obtained as

$$\hat{a} = \left[ \frac{\sum \log x_i}{n} - \log k \right]^{-1}$$

$$\hat{k} = \min_{i \leq n} (x_i)$$

Since the district size distribution for the population is skewed in nature, skew model is used to present the district size distribution. The Pareto distribution is one of the skew models. It has been proposed to measure the co-efficient 'a' in rank size rule using the District size distribution given in table (2), table (3), and table (4).

## V. EMPIRICAL RESULTS

The populations of all districts of Tamil Nadu for the census year between 1991 and 2011 were used to find the regression equations and for estimating the parameter of Pareto distribution in order to find their interrelationship. The regression equations for the three periods are obtained and presented in table (1)

# International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

**Vol. 4, Issue 12, December 2015**

**Table.1 Zipf’s Regression Equations**

Census Year	Regression equations	R <sup>2</sup>	Correlation coefficient, r	Zipf’s Coefficient
1991	$\log y = 7.913 - 1.108 \log x$	0.633	-0.796	1.108
2001	$\log y = 9.630 - 1.365 \log x$	0.748	-0.865	1.365
2011	$\log y = 10.039 - 1.418 \log x$	0.755	-0.869	1.418

The population of all distribution of all districts of Tamil Nadu state for the census year between 1991 and 2011 were used for estimating the parameter ‘a’ of the Pareto distribution.

The district size are classified into six classes as per 1991 census classes are I:7-14; II:14-21; III: 21-28; IV:28-35; V: 35-42; VI:42-49. The district size distribution of Tamil Nadu has been formulated and presented in table (2)

**Table. 2 District size distribution –Census 1991**

District size interval(in lakes)	No. of district
7-14	2
14-21	2
21-28	2
28-35	1
35-42	4
42-49	3

The estimate of the parameter is obtained from Table (2) based on Pareto distribution as  $\hat{a}=1.6889$

As per census 2001 the districts are classified into seven classes are I:4-10; II:10-16; III:16-22; IV:22-28; V: 28-34; VI:34-40;VII:40-46. The district size distribution of Tamil Nadu state as per census 2001 has been formulated and presented in table (3)

**Table.3 District size distribution -Census 2001**

District size interval (in lacks)	No. of district
4-10	4
10-16	8
16-22	4
22-28	7
28-34	4
34-40	1
40-46	2

# International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

**Vol. 4, Issue 12, December 2015**

The estimate of the parameter is obtained from Table (3) based on Pareto distribution as  $\hat{a} = 1.504$

As per 2011 census, The districts are classified into eight classes are I: 5-10; II: 10-15; III:15-20; IV:20-25; V:25-30; VI: 30-35; VII: 35-40; VIII: 40-45.The district size distribution of Tamil Nadu as per census 2011 has been formulated and presented in table (4)

**Table. 4 District size distribution – Census 2011**

District size interval (in lakhs)	No. of District
5-10	3
10-15	5
15-20	8
20-25	5
25-30	2
30-35	5
35-40	3
40-45	0
45-50	1

The estimate of the parameter is obtained from Table (4) based on Pareto distribution as  $\hat{a} = 1.6592$

## VI.DISCUSSION

The estimated value of the parameter of Pareto distribution obtained by using the data of the census year 1991,2001 and 2011 is presented in Table (5)

**Table.5 Estimates from Pareto model**

Census Year	Pareto Coefficient
1991	1.6889
2001	1.5040
2011	1.6592

**Table. 6 Zipf's Regression equations, Pareto Coefficient and Zipf's Coefficient**

Census Year	Regression equations	Pareto Coefficient	Zipf's Coefficient
1991	$\log y = 7.913 - 1.108 \log x$	1.6889	1.108
2001	$\log y = 9.630 - 1.365 \log x$	1.5040	1.365
2011	$\log y = 10.039 - 1.418 \log x$	1.6592	1.418

# International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 12, December 2015

There are inverse relationship between populations and ranks. The table (6) shows that the actual distributions of district size of three periods are inconsistent with those of the rank size rule. The value of 'a' in 1991 is 1.108 which is the smallest in three periods. Where as that of 2011 is the largest, 1.418. This means that the distribution of the rank size of districts in 1991 is the closest to the distribution of the rank size rule, while the distribution of the districts in 2011 is the farthest away from the distribution of the rank size rule. The values of 'a' in 1991, 2001 and 2011 are respectively obtained as 1.108, 1.365 and 1.418 which are larger than the value of 'a' in the rank size rule, 1.0. Therefore, all slopes of three periods are steeper than the slope of the rank size rule. This result denotes that the expected district sizes based on the rank size rule are larger than the real size of the districts in Tamil Nadu. The values of slope in actual distribution of districts have been larger and larger over time. Therefore the adoption of rank size rule to the distribution of districts size of Tamil Nadu seems to be less adequate over time.

## VII.CONCLUSION

The empirical studies of the change in rank size system of districts were examined. It provided the ways for an examination of intercept and slope changes over time. First, intercept value could change because of a uniform rate of increases or decreases in the population of all districts. In this case, all districts the attracting migrants in proportion to their population and were experiencing equal rates of natural growth. Second, the slope of the districts rank size distribution changed because smaller settlements were not growing at as rapid a rate as larger ones. Third, slope of the rank size distribution changed because the population of the largest districts had increased to force the slope of the distribution to steepen to meet the increased intercept. This case may occur when the most rapid population growth takes place in the larger districts in the state but without an actual decline in small place population. Finally, the pareto coefficients from both empirical and theoretical distribution are larger than the value of 'a' in the rank size rule, 1.0. Thus, all slopes of three periods are steeper than the slope of the rank size rule over time. Therefore adoption of rank size rule to the distribution of district-size of Tamil Nadu seems to be less adequate over time.

## ACKNOWLEDGEMENT

I gratitude my thanks to Dr. G. Venkatesan M.Sc., Ph.D., Associate Professor and Head, Department of Statistics, Government Arts College (Autonomous), Salem – 7, Tamil Nadu, India, for his immense help and continuous support to complete this work. The author thanks the referees and editor for constructive suggestions that have improved the content and presentation of manuscript to publish in this journal.

## REFERENCES

- [1] Alperovich, G.A, 'An Explanatory Model of city-size distribution: Evidence from cross country data', Urban studies 30, pp. 1591-1595, 1993.
- [2] Black, D. And J.V.Henderson, 'Urban Evolution in the USA', Mimeo, Brown University, 2000
- [3] Brakeman's, H.Garretson, G.V.Marrewijk and M.Vanden Breg, 'The Relation of zipf. Towards a further understanding of the rank-size distribution', Journal of regional science, 79, pp. 183-213, 1999.
- [4] Guerin-pace, F, 'Rank-size distribution and the process of Urban growth', Urban studies 32, pp. 551-562, 1995.
- [5] Hsing, Y, 'A note on function forms and the Urban size distribution', Journal of Urban economics 27, pp. 73-79, 1990.
- [6] Kamecke, U, 'Testing the Rank size rule hypothesis with an efficient Estimator', Journal of Urban economics, 27, pp. 222-231, 1990.
- [7] Normal. Johnson and Samuel Kotz and Balakrishnan, N, Continuous Univariate Distributions, Vol-1 second edition, John Wiley & Sons, INC, New York, pp. 498-584, 1996.
- [8] Parr, J.B, 'A Note on the size distribution of cities over time', Journal of Urban Economics, 18, pp. 199-212, 1985.
- [9] Rao, C. R, Linear Statistical inference and its applications, Wiley eastern, New Delhi, pp.353-374, 1973.
- [10] Rosen, K.T. and M.Resnick, 'The size distribution of cities: An examination of the Pareto law and Primacis', Journal of Urban Economics, 8, pp. 165-186, 1980.
- [11] Simon, H, 'On a class of distribution function', Biometrika, 42, pp. 425-440, 1955.

## International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 12, December 2015

- [12] Census of India 1999: Primary Census Abstract, Director of census operations, Chennai, Tamil Nadu [http:// www.CensusIndia.gov.in](http://www.CensusIndia.gov.in), [www.Census.tn.nic.in](http://www.Census.tn.nic.in)
- [13] Census of India 2001: Primary Census Abstract, Director of census operations, Chennai, Tamil Nadu [http:// www.CensusIndia.gov.in](http://www.CensusIndia.gov.in), [www.Census.tn.nic.in](http://www.Census.tn.nic.in)
- [14] Census of India 2011: Primary Census Abstract, Director of census operations Tamil Nadu series, 34, pp. 51-54.
- [15] Carlos M.Uizua, 'A Simple and efficient test for Zipf's law', Economics Letters, 66, pp 257 – 260, 2000.
- [16] B.Basu and S.Bandyopadhyay ' Zipf's law and distribution of population in Indian cities', Indian J.Phys.83(11), pp. 1575 – 1582, 2009.
- [17] M.Nadi, 'Concentration indices and Zipf's law', Economics Letters, 78, pp. 329 – 334, 2003.
- [18] Xinyue Ye, Yichun Xie, 'Re-examination of Zipf's law and urban dynamic in China: a regional approach', Ann Reg sci, 49, pp. 135 – 156, 2012.
- [19] William J.Reed, 'The Pareto, Zipf and other power laws', Economics Letters 74, pp.15 – 19, 2001.

### BIOGRAPHY



Mrs. R. Selvam, M.Sc.,M.Phil., is currently an Assistant Professor, Department of Statistics, Government Arts College (Autonomous), Salem-636007. Prior to her recent appointment at Govt. Arts College (Autonomous), Salem-7, she was a Lecturer in Statistics at various colleges affiliated to Periyar University, Salem, Tamil Nadu, India. She received her M.Sc and M.Phil from Annamalai University, Chidambaram, Tamil Nadu, India and her Ph.D programme is going under process. In 1993, she has been awarded 'The Sir Norman Strathic Prize' by Annamalai University for her excellence in academic. She has teaching experience in Statistics of more than 15 years. She has presented research papers in National and International conference, National level Seminar and National Workshop. She also published research papers in International Journal. Her current research focus is in the area of Mathematical Demography.