Further Analysis of the Zipf's Law: Does the Rank-Size Rule Really Exist?

Fungisai Nota
Department of Business Administration and Economics, Wartburg College, Waverly, IA 50677, U. S. A.; Tel: 1-319-693-9084; Email: fungisai.nota@wartburg.edu

Shunfeng Song
Department of Economics, University of Nevada, Reno, NV 89557, School of Economics, Zhejiang University, Hangzhou, Zhejiang, China; Tel: 1-775-785-6860; Email: song@unr.edu

ABSTRACT. Zipf's law has two striking regularities: excellent fit and an exponent close to 1.0. When the exponent equals 1.0, Zipf's law collapses into the rank-size rule. This paper alters the sample size, the truncation point, and the mix of cities in the sample to analyze the Zipf exponent. Our results demonstrate that the exponent is close to 1.0 only for a number of selected sub-samples. Small samples of large cities provide higher values, while samples of small cities produce lower values. Using the estimated values of the exponent derived from the rolling sample method revealed elasticity in the exponent with regard to sample size. Our results also suggest that the rank-size rule should be interpreted with caution. Although it is well-known and commonly used, the rank-size rule may be more of a statistical phenomenon than an economic regularity.

KEYWORDS. ***Zipf's law, rank-size rule, rolling sample method***

1. INTRODUCTION

Zipf's law describes an empirical regularity observed in both the natural and social sciences (e.g., Zipf, 1949; Shiode and Batty, 2000; Sinclair, 2001; Li and Yang, 2002; Tachimori and Tahara, 2002). It states that the rank associated with a given size S is inversely proportional to S at a given power. If this power is equal to one, Zipf's law collapses into what is commonly called the rank-size rule. This implies that in the case of cities, the second largest city is one-half the size of the first and the third largest city is one-third the size of the largest and so on. In cases where the power is greater than one, Zipf's law suggests that the second largest city is more than half as large as the largest city and the third largest city is more than a third as large as the largest city, and so on. Conversely, an exponent of less than one would suggest that the second largest city is less than half the size of the largest city, and so on. Linearizing the relationship between rank and size using a log transformation can facilitate the estimation of the negative exponent.

One of the striking characteristics of Zipf's law is its excellent fit. Numerous empirical studies have shown that a linear regression of log-rank on log-size generates an excellent fit (very high R2-value). Using data from 44 countries, Rosen and Resnick (1980) found that R2-values were above 0.95 for 36 countries and only Thailand had an R2-value below 0.9 (0.83). Mills and Hamilton (1994) obtained an R2-value of 0.99 using 1990 data on 366 urbanized areas in the U.S. Song and Zhang (2002) obtained an R2-value of 0.91 for 665 Chinese cities in 1998. This astonishing regularity led Krugman (1995, p.44) to claim that the rank-size rule is "a major embarrassment for economic theory: one of the strongest statistical phenomenon we know, lacking any clear basis in theory." Fujita et al. (1999, p. 219) stated "the regularity of the urban size distribution poses a real puzzle, one that neither our approach nor the most plausible alternative approach to city sizes seems to answer."

The second notable observation is related to Zipf's coefficient. In studies related to urban development, the coefficient is very close to 1.0, thus the rank-size rule holds. Gabaix (1999a, b) argued that the rank-size rule is theoretically a natural result of urban growth independent of the initial size of the city. Fujita et al. (1999) suggested that the rank-size rule does indeed approximate the long-run spatial distribution of a mature spatial system. Among the 44 countries empirically studied by Rosen and Resnick (1980), the estimated coefficient ranged from 0.809 for cities in Morocco to 1.963 for cities in Australia. Nitsche (2005) analyzed 515 estimates from 29 studies related to the rank-size relationship and found that two-thirds of the estimated coefficients were between 0.80 and 1.20, with a median estimate of 1.09. This implies that city-size distributions tend to be more even than what is suggested by the rank-size rule.

Some have used economic theory to explain why Zipf's law holds, others cited Gibrat's law, and still others counted it as a purely statistical phenomenon. The economic explanation relies