American Economic Review 2009, 99:4, 1676–1683
http://www.aeaweb.org/articles.php?doi=10.1257/aer.99.4.1676

Gibrat's Law for (All) Cities: Reply

By JAN EECKHOUT*

Moshe Levy (2009) raises questions about the fit of the very upper tail of the city size distribution in Eeckhout (2004). More specifically, he is unsure about whether the upper tail fits the lognormal distribution. Basing his findings primarily on a graphical interpretation, he seeks to show that the conclusions of the analysis do not apply to the upper tail of the size distribution. In this reply, I show that the method on which his conclusion is based is unsubstantiated.

The main confusion arises from the peculiarities of the log-log plot. The log-log plot is a commonly used method to visually inspect fat-tail distributions. I show that seemingly similar patterns on log-log plots have radically different interpretations due to the logarithmic scaling and the fact that the tail of the distribution is indeed lognormal. In addition, the methodology used to estimate a truncated subsample of the distribution while testing its significance against a distribution with prespecified parameters is ill-founded. The main conclusion is that Gibrat's law continues to hold: city sizes follow a proportionate growth process, and this gives rise to a lognormal size distribution, with a lognormal upper tail included.

The confusion surrounding the interpretation of the log-log plot is quite natural. My own first reaction after seeing Levy's Figure 1 was that the upper tail does not fit the lognormal. The value of his comment and this debate is therefore to draw attention to the peculiarities of the log-log plot and the methods of analyzing fat-tail distributions. This is especially relevant now, when log-log plots are widely used in different fields and in the popular press.[1]

Finally, this reply also highlights the consistency of a lognormal distribution with Zipf's law. Data generated from a lognormal distribution have tails that closely resemble a Pareto distribution. Because Pareto distributions are much simpler analytically, it may therefore be useful to focus on Pareto distributions when studying the very upper tail of a distribution.

I. The Upper Tail Is Lognormal

The empirical observations of the large city sizes are within the confidence bands of the lognormal estimates. We use the confidence bands generated by the Lilliefors test with a 5 percent significance level, a two-sided goodness-of-fit test suitable when a fully specified null distribution is unknown and its parameters must be estimated. The Lilliefors test statistic is the same as for the Kolmogorov-Smirnov test, and it tests the fit of the data with the normal distribution with sample mean and variance. We choose a confidence interval of 5 percent and show that the tail of the distribution is well within this tight interval. For the upper tail (the top two percentiles), Figure 1 plots the predicted lognormal cumulative distribution function (CDF) with sample mean

* Department of Economics, University of Pennsylvania, McNeil Rm 456, 3718 Locust Walk, Philadelphia, PA 19104, and ICREA-UPF (e-mail: eeckhout@ssc.upenn.edu). I would like to thank numerous colleagues for insightful discussions and comments. Financial support from the European Research Council (grant 208068) is gratefully acknowedged.

[1] There is an extensive academic literature exploiting fat-tail distributions in computer science mapping the connectedness of Web pages and Weblogs, in finance, and in management. In the popular press there is a large body of writing on "The Long Tail": http://en.wikipedia.org/wiki/The_Long_Tail.

This content downloaded from 88.18.104.120 on Sat, 27 Apr 2013 06:08:21 AM
All use subject to JSTOR Terms and Conditions

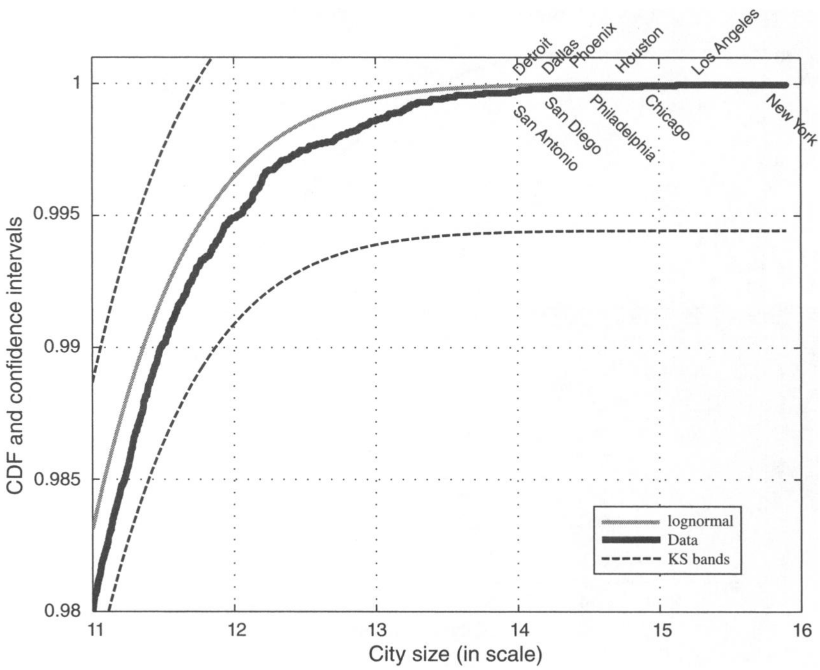

FIGURE 1. CDF ON LINEAR SCALE WITH CONFIDENCE BANDS

and variance, the data on city size, and the confidence bands. Observe that the vertical distance between the predicted distribution and the bands is constant.

How can we reconcile this finding with Levy's (2009) claim that the tail does not fit the lognormal distribution? By considering a logarithmic scale for the rank, the corresponding differences between the data and the estimated rank blow up out of proportion for low-rank observations. The low-rank observations correspond to the larger cities and hence the apparent lack of fit for the upper tail. The way to read the confidence bands is that a value generated by the Lilliefors test of 0.0056 above or below the predicted lognormal CDF $\Phi(\cdot)$ ($\Phi(\cdot) \pm 0.0056$) translates in this case of a sample size of 25,359 into a band of ranks of ± 142 (which is equal to $25{,}359 \times 0.0056$). If your predicted rank is, say, 10,000, then the confidence bands are [9,858; 10,142]. And here the logarithmic scale kicks in. For large enough rank R, the confidence bands do not show since $\ln(R \pm 142) \approx \ln R$. Since for large R the scale "glues" the data to the predicted distribution as well as the bands, Levy (2009) concludes that the "lognormal fits extremely well." However, for small R, the bands fan out since $\ln(\mathrm{R} + 142) \approx \ln 142$. For example, for $R = 1$, the bands are given by $\ln(1 + 142) = 4.96$. This is illustrated in Figure 2. For low rank, the logarithmic scale exacerbates the differences between the predicted distribution and the data, even if the data are well within the bands and close in terms of the linear distance, as in Figure 1. The fact that for low rank both curves fan out therefore cannot be taken as evidence that the lognormal fit breaks down.

To further illustrate that this is an artefact of the logarithmic scale of rank, we now show in Figure 3 that Levy's conclusion should also be reached for any part of the distribution, and not just for the upper tail. To that purpose, we also perform Levy's analysis for the inverse of rank of sizes, i.e., the smallest cities are ranked 1 and NY city is ranked 25,359. While in Figure 2 the confidence bands (and the data) were very close to the predicted distribution in the lower

This content downloaded from 88.18.104.120 on Sat, 27 Apr 2013 06:08:21 AM
All use subject to JSTOR Terms and Conditions

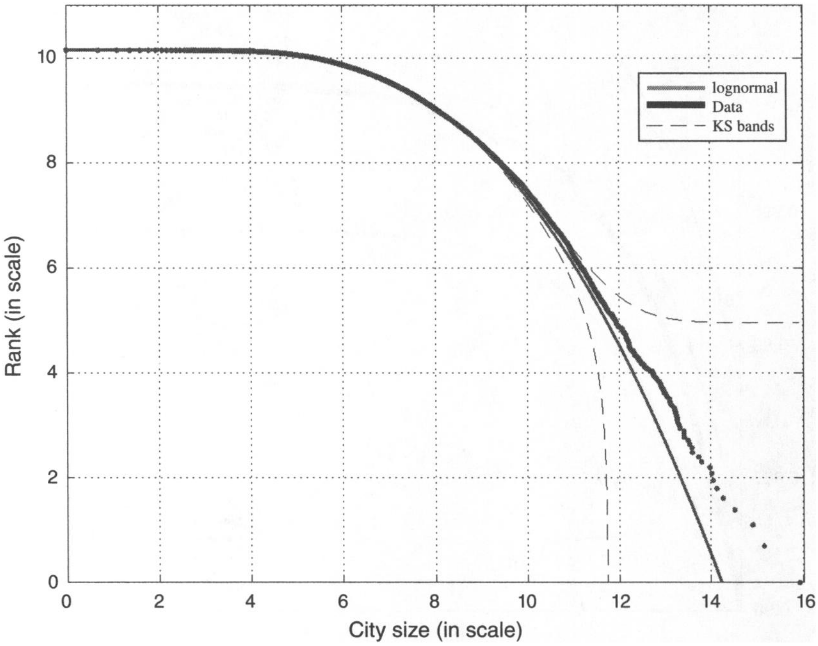

FIGURE 2. LOG-LOG PLOT INCLUDING THE KS BANDS

tail, those confidence bands now fan out because of the impact of the logarithmic scale on the inverse rank.

Given that the log-log plot is misleading at low-rank observations, we should also observe a similar pattern in the middle of the distribution, provided we truncate the distribution. Suppose, for example, we truncate the size distribution and drop all cities above the median observation (12,679 out of 25,359) and we assign rank equal to one to the median city (which happens to have a population of 2,758) and rank 12,679 to the smallest city. Then a log-log plot in Figure 4 of the left half of the data, together with the predicted distribution and confidence bands, should show a similar "fanning" out of the bands near the lower-ranked cities (i.e., at the intersection with the horizontal axis). This fanning out is not apparent, but it is there: the upper bound of the median city is at 4.96 (nearly half the vertical scale) higher, but it happens to coincide with the vertical log-log plot. The reason it is not apparent is due to the fact that there are so many observations in the middle of the distribution and therefore the log-log plot is nearly vertical. As a result, a vertical transformation appears as a perfectly coinciding.

If the lognormal is rejected in the upper tail, it must also be rejected in many other parts of the distribution. In contrast, the author argues that lognormality is accepted everywhere else: "[...] the lognormal distribution indeed provides an excellent fit to the empirical data for 99.4 percent of the size range [...]", Levy (2009, 1). Using the same logic for the entire distribution, either it applies at the upper tail and everywhere else, or it applies nowhere.

Log-log Plot Caveats.—The issues raised by Levy (2009) are instructive because they point out some peculiarities of log-log plots. Log-log plots are widely used in many fields (finance, computer science, engineering), and they are often used to derive conclusions about the fit of an empirical distribution. Four caveats are worth keeping in mind:

This content downloaded from 88.18.104.120 on Sat, 27 Apr 2013 06:08:21 AM
All use subject to JSTOR Terms and Conditions

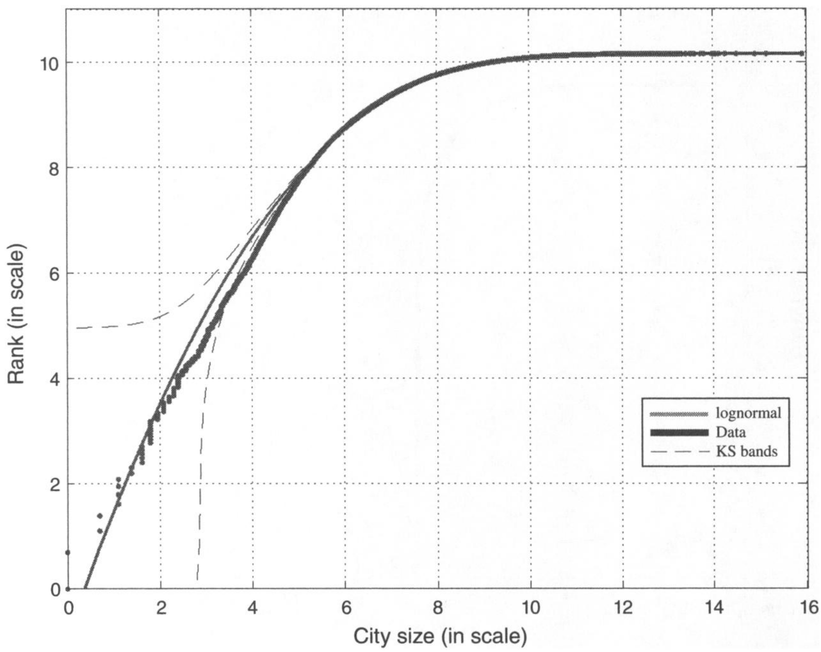

Figure 3. Log-Log Plot with Inverse Rank

1. The log scale heavily *distorts low-rank observations:* the scale blows up bounds and deviations for very few observations at low ranks; the opposite is true for high-rank observations.

2. The data are heavily *concentrated in the middle* of the distribution. The very casual observer may be (*wrongly*) led to believe that nearly half of the observations do not fit the distribution. Figure 2 indicates that all observations between log rank 0 and 5 (on a full domain of 0–10) do not fit the distribution. As it turns out, this is the case for only 150 of 25,359 observations. On the horizontal axis of Figure 2, those 150 observations span log size between 12 and 16. This concentration bias is symmetric. There are also only 150 observations between log size 0 and 4. Therefore, 25,059 observations are concentrated between log size 4 and 12, half the length of the horizontal axis.

3. Log-log plots are very uninformative for parts of the distribution with a high density and a large number of observations. For high-density sizes in large sample distributions, the log-log plot is nearly vertical. Even if the confidence bands are plotted, they are hardly distinguishable from predicted distribution, since they are a vertical transformation.

4. A biased representation similar to the log-log plot occurs with a normal probability plot, where the deviations in the middle of the distribution are not picked up. In the normal probability plot, however, at least the bias is symmetric and it can readily be seen that a similar lack of fit occurs at the bottom of the distribution. Adding the confidence bands to the normal probability plot leads to a similar fanning out in the tails.

This content downloaded from 88.18.104.120 on Sat, 27 Apr 2013 06:08:21 AM
All use subject to JSTOR Terms and Conditions

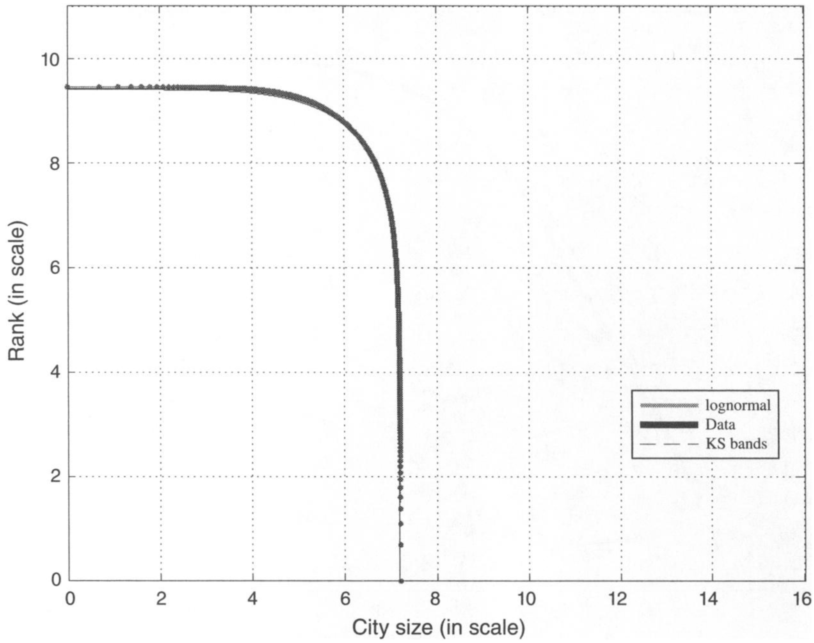

FIGURE 4. LOG-LOG PLOT FOR LEFT HALF OF DISTRIBUTION: MEDIAN CITY HAS RANK = 1

With these caveats in mind, and even though the objective of this Reply is not to propose solutions for estimating Pareto distributions, it is worth using alternative tools beyond the log-log plot for inferring information about Pareto distributions. Using an experiment, Michel Goldstein, Steven Morris, and Gary Yen (2004) point out the inaccuracies of using graphical methods for fitting to a power law distribution. A linear fit on the log-log scale of an ordered sample is biased.[2] They show that using a maximum likelihood estimation (MLE) is far more robust, and develop a Kolmogorov-Smirnov type test in which the power-law exponent is estimated using maximum likelihood estimation.

II. Lognormal Data and Pareto Tails

Even if Gibrat's law holds and the distribution is lognormal, we may be interested in analyzing the tails in isolation. The tails are important, for example, because each observation represents a large population. The fact that the tail fits the lognormal does not mean that it cannot also fit the Pareto at the same time. The lognormal tail in the limit becomes hard to distinguish from a Pareto tail. In other words, there is little power in a test that distinguishes between the Pareto and the tail of a lognormal distribution.

[2] In fact, they point out that the linear OLS estimate based on the data not including the tail observations is more accurate than the full linear regression. This corrects for the sample bias of few tail observations disproportionately affecting the slope of the estimate.

This content downloaded from 88.18.104.120 on Sat, 27 Apr 2013 06:08:21 AM
All use subject to JSTOR Terms and Conditions

To see this, observe that, for a lognormal distribution with a large enough standard deviation (as is the case for the city size data), both the Pareto and the lognormal distribution have a linear logdensity (Michael Mitzenmacher 2003). The Pareto has a logdensity that is given by

$$\ln f(x) = -(\alpha - 1)\ln x + \alpha k + \ln \alpha,$$

whereas the lognormal has a logdensity

$$\ln f(x) = -\ln x - \ln(\sqrt{2\pi}\,\sigma) - \frac{(\ln x - \mu)^2}{2\sigma^2}.$$

The logdensity of the Pareto is linear in ln*x*, and so is the logdensity of the lognormal for large σ, which renders the tails indistinguishable. Both distributions are regularly varying,[3] i.e., they are heavy-tailed, and their tails have similar properties. Robert Daris and Lucio Torelli (2000) denote the Weibull factor of a distribution $F(x)$ as $W = x\, d/dx\, [\ln(-\ln F(x))]$,[4] and show that $W \sim 1/\ln x$ for the Pareto distribution and $W \sim 2/\ln x$ for the lognormal.

It is natural, then, that the upper tail of city sizes can be fit to a Pareto distribution. After all, Zipf's law has repeatedly been shown to hold for the largest cities,[5] and it establishes that one cannot reject that the tail of the distribution is Pareto. Gabaix and Yannis Ioannides (2004) propose an unbiased estimate for the regression of log rank on log size that corrects for the bias in the log-log representation. Using Monte Carlo simulations, they show that the OLS coefficient in the range that is usually considered for city size distributions is biased downward and the standard errors are underestimated by a large factor. Kwok Tong Soo (2005) makes an exhaustive comparison for 73 countries, and using OLS, he finds that Zipf's Law holds for 20 countries.

Given that the tail of a lognormal is indistinguishable from the Pareto under certain circumstances, the researcher who is interested in the tail properties of a size distribution can choose which one to use. The Pareto distribution is very appealing theoretically for its simple functional form. That makes the Pareto distribution a useful tool in analyzing size distributions of population or any other relevant economic variable. In the context of executive compensation, for example, Gabaix and Augustin Landier (2008) focus on the tail properties of the earnings distribution. Using the Pareto distribution allows for closed-form derivations which would be much harder with the lognormal distribution, a complicated three-parameter function (mean, variance, and truncation). The Pareto distribution is an extremely useful tool for analytical tractability, while fitting the data for the largest observations, and while continuing to be consistent with proportional growth.

Focusing on the tail per se can therefore be valuable. However, the method employed in Levy (2009) for refuting the lognormal in favor of the Pareto has a number of shortcomings, mainly due to the fact that it ignores part of all available data:

1. The *data are available* and are part of the distribution that is being estimated. Unlike the case of metropolitan areas, the data are not censored below a given size (100,000 in the case of metropolitan areas (MAs)). In those cases of censored data, one may estimate the properties of a tail, derived from a larger, unobserved distribution. Here, however, the data of the entire distribution are observed. Only half a percent of all observations are used.

[3] A positive function is regularly varying at ∞ of index α if

$$\lim_{x\to\infty} \frac{f(tx)}{f(x)} = t^{\alpha},\ t > 0.$$

[4] For the Weibull distribution, the Weibull factor is constant.

[5] See, among many others, George Zipf (1949) and Xavier Gabaix (1999).

This content downloaded from 88.18.104.120 on Sat, 27 Apr 2013 06:08:21 AM
All use subject to JSTOR Terms and Conditions

2. How should one interpret Levy's (2009) comparison of a given lognormal distribution (with $\mu = 7.28$ and $\sigma = 1.75$) to the Pareto? It is not a test of the lognormal against Pareto, but rather a test of significance of the lognormal with prespecified parameters against "anything other than that" (including another lognormal or a mixture of lognormals). The estimate of a subsample is bound to yield a better fit than the fit for the same subsample estimated on the entire sample. The subsample estimate is less constrained, as it does not have to fit the censored data. Using this logic, one could further partition the subsample of 150 observations and reject the fit of the Pareto proposed by Levy (2009) (with estimated coefficient $\hat{a} = 1.3542$) to generate a different estimate $\hat{a}$ with an even better fit for some $N < 150$.

3. With all the data available, and given that one nonetheless does not want to use all data, the question arises what the *appropriate truncation point* is. The choice of the truncation point becomes *endogenous* and can be chosen subjectively to favor one hypothesis over another. Should we repeat the exercise for all possible truncation points and stop where the deviation from the lognormal is biggest? This resembles pre-testing in standard regression analysis, and it is well known that pre-testing leads to biased estimates.

III. The Economics of Population Dynamics

The consistent finding is that population growth across cities is proportionate, and in the light of Gibrat's law, it is not surprising that the ensuing distribution is lognormal. Moreover, the lognormal tail and the Pareto tail are hard to distinguish, thus corroborating Zipf's law.

The relevance of understanding the properties of the size distribution derives from the light it sheds on the economic dynamics that drive population mobility. That allows us to evaluate how policy interventions like transportation subsidies or local taxes will affect prices, mobility, and ultimately welfare. The contribution of "Gibrat's Law for (All) Cities" (Eeckhout 2004) is that shocks to the TFP (total factor productivity) of local economies can induce population mobility that is consistent with Gibrat's law. The economic mechanism, driven by economic shocks translated through the general equilibrium price mechanism in the labor and housing markets induces population dynamics that are consistent with those empirically observed. Recent research confirms that those economic mechanisms are consistent with the data. For example, Morris Davis and François Ortalo-Magné (2007) find that the household expenditure share on housing is remarkably constant over time and across US metropolitan areas. Using an equilibrium model that builds on Eeckhout (2004), they estimate the expenditure share on housing at 0.24 and evaluate the quality of life and welfare across cities.

Such findings based on Gibrat's law are important for economic policy because we can readily measure the impact of policies on house prices, mobility, and welfare. Hopefully, this reply clarifies that the effort in further analyzing those economic mechanisms of population dynamics is well spent. The analysis of Gibrat's law for cities is built on solid ground and it is a robust starting point for further contributions to the understanding of the economic driving forces of population dynamics.

REFERENCES

Daris, Roberto, and Lucio Torelli. 2000. "Some Indices for Heavy-Tailed Distributions." Unpublished.
Davis, Morris A., and François Ortalo-Magné. 2007. "Household Expenditures, Wages, Rents." Unpublished.
Eeckhout, Jan. 2004. "Gibrat's Law for (All) Cities." *American Economic Review*, 94(5): 1429–51.
Gabaix, Xavier. 1999. "Zipf's Law for Cities: An Explanation." *Quarterly Journal of Economics*, 114(3): 739–67.

This content downloaded from 88.18.104.120 on Sat, 27 Apr 2013 06:08:21 AM
All use subject to JSTOR Terms and Conditions

Gabaix, Xavier, and Yannis M. Ioannides. 2004. "The Evolution of City Size Distributions." In *Handbook of Regional and Urban Economics Volume 4 Cities and Geography*, ed. J. V. Henderson and J.-F. Thisse, 2341–78. Amsterdam: Elsevier.

Gabaix, Xavier, and Augustin Landier. 2008. "Why Has CEO Pay Increased So Much?" *Quarterly Journal of Economics*, 123(1): 49–100.

Gibrat, Robert. 1931. *Les Inégalités économiques*. Paris: Sirey.

Goldstein, Michel L., Steven A. Morris, and Gary G. Yen. 2004. "Problems with fitting to the Power-law distribution." *The European Physical Journal B-Condensed Matter*, 41(2): 255–58.

Levy, Moshe. 2009. "Gibrat's Law for (All) Cities: Comment." *American Economic Review*, 99(4): 1672–75.

Mitzenmacher, Michael. 2003. "A Brief History of Generative Models for Power Law and Lognormal Distributions." *Internet Mathematics*, 1: 226–51.

Soo, Kwok Tong. 2005. "Zipf's Law for Cities: A Cross-country Investigation." *Regional Science and Urban Economics*, 35: 239–63.

Zipf, George K. 1949. Human Behavior and the Principle of Least Effort. Cambridge, MA: Addison-Wesley.

This content downloaded from 88.18.104.120 on Sat, 27 Apr 2013 06:08:21 AM
All use subject to JSTOR Terms and Conditions