ON DESIGNING A ONE-STAGE "BEHAVIORAL MODEL" TO EXPLAIN CITY SIZES

by

Trudy Ann Cameron
Department of Economics, UCLA

Discussion Paper #409
Department of Economics
University of California, Los Angeles
405 Hilgard Ave
Los Angeles, CA 90024

June 1986

On Designing A One-Stage "Behavioral Model" to Explain City Sizes

by

Trudy Ann Cameron
Department of Economics
University of California, Los Angeles

October, 1985
Revised April, 1986

ABSTRACT

A large number of papers have estimated the parameters of Pareto distributions for city sizes in different countries, but only one has attempted to explain the differing magnitudes of these parameters with a set of country-specific explanatory variables. It is reassuring that there has now been some research which advances beyond simple "curve-fitting" to explore systematically the "behavioral" determinants of city size. However, this newer research uses two-stage OLS methods which yield invalid second-stage standard errors (and consequently, questionable hypothesis tests).

In this paper, we propose an alternative, one-stage behavioral model which has the potential to generate more-useful results by being better able to uncover the uncontaminated systematic relationships between city size and its determinants. In general, these new models are non-linear in parameters, so that they require more-sophisticated econometric techniques. However, nonlinear optimization methods are steadily becoming more accessible to researchers; work need no longer be limited by OLS techniques.

On Designing A One-Stage "Behavioral Model" to Explain City Sizes

by

Trudy Ann Cameron

1. Introduction

A relatively simple concept in Urban Economics, the size distribution of cities, has received a considerable amount of attention in the empirical literature--sufficiently for it to have gained a place in Urban Economics textbooks (i.e. Mills and Hamilton, pp. 58-62). Increasingly complex methodological techniques have been employed in the pursuit of an intriguing statistical regularity. The "original" models have now been applied to data from new countries, and updated more-general models have also been tested. But if one criticism can be leveled at this work, it might be the following: the "traditional" empirical approach to this problem has limited the generality of the modeling process. Determination and refinement of the exact value of the shape parameter of a Pareto distribution of city sizes is a worthwhile endeavor. However, not much "behavioral modeling" has been undertaken. For instance, while we know what city sizes <u>are</u> (or have been) in different countries at different times, it would be very useful to uncover the factors which influence city sizes. With this knowledge, we could make useful predictions about anticipated trends in city-sizes, based on our expectations about the "explanatory" variables. Unfortunately, the two-stage estimation process which has typically been employed is too restrictive in its first stage and yields inaccurate standard errors in the second stage.

In this paper, we will first review the state-of-the-art of behavioral modeling for the distribution of city sizes. We will then propose a specification for a "behavioral model" and describe tractable, straightforward nonlinear estimation techniques.

We hope that those researchers who have undertaken empirical work in this area will be inspired to explore this alternative class of specifications. In some cases, it would seem relatively inexpensive to augment these data and then to reestimate using an explicit behavioral model like the one proposed here.

2. Extant Models for City Size Distributions

Beginning with the seminal work by Zipf (1949) which uncovered the so-called "rank-size" rule, a large number of investigators have explored the theoretical and empirical validity of the Pareto distribution (of which the "rank-size" rule is a special case) as a description of city sizes. (See Beckmann, 1958, Berry, 1961, Richardson, 1973, and Alperovich, 1982, among others.) Empirical efforts have focused on the Pareto absolute frequency (as opposed to relative frequencies) and have utilized the absolute "survival function". This statistical construct gives the absolute number of cities greater than or equal to a certain size, $G(x)$, as a function of that size, x. Algebraically, the hypothesized relationship is:

$$G(x) = A_0 x^{-a} \qquad (1)$$

or, $$\log G(x) = \log A_0 - a \log x$$

If $a = 1$, then $xG(x) = A$, a constant equal to the population of the largest city (since $G(x) = 1$ for the largest city and x is its size). Many investigators have treated the rank-size hypothesis as simply H_0: $a = 1$. More recently, Alperovich (1984) has suggested that the appropriate rank-size rule hypothesis is the joint hypothesis H_0: ($a = 1$ and A = size of largest city).

All of these studies, however, are primarily descriptive. The objective is the careful estimation of magnitudes of the individual parameters of the size distribution and hypothesis tests regarding their true values.

To our knowledge, the sole attempt thus far to formulate a behavioral model has been undertaken by Rosen and Resnick (1982). These investigators fit separate Pareto size distributions for separate countries in the first stage of their estimation. Subsequently, they attempt to explain cross-country variations in each of the two estimated parameters, A_0 and a, by regressing each parameter on a vector of characteristics for each country. They experiment with population, GNP, and total miles of railway as determinants of the Pareto parameters.

Without diminishing the fundamental importance of the Rosen and Resnick breakthrough into behavioral modeling, some shortcomings to their approach can be identified. Most importantly, the second stage of their two-stage estimation technique ignores the fact that the dependent variables are estimated quantities. Most crucial to the interpretation of Rosen and Resnick's results is the fact that the simple OLS standard error estimates generated at the second stage are incorrect. Intuitively, since the dependent variable is merely a point estimate of a random variable which itself has variance, ignoring the presence of this additional variance will bias the second stage standard error estimates. We cannot be certain a priori about the direction of this bias, but as a consequence, explanatory variables which appear to be statistically significantly different from zero at the second stage may in fact not be so, or vice versa. To remedy this problem, it would be preferable to estimate all parameters at once, in only one step.

3. <u>Alternative Modeling Strategies</u>

One presumption which hinders generalization of the literature on city size distributions is the persistent adherence to the Pareto distribution. While the theoretical developments borrowed from the industrial organization literature on the size distribution of firms suggests the Pareto distribution, which undoubtedly explains this choice by empirical investigators, there has been little experimentation with alternative distributions. It is important to note that the Pareto distribution was also at one time favored to describe the size distribution of incomes. However, the recent econometrics literature, specifically McDonald (1984), has considered an extremely general class of four-parameter density functions: generalized beta distributions of the first and second kind. He also cites recent work with a three-parameter special case of both of these distributions, the generalized gamma distribution. In particular, Esteban (1981) is reported to have shown that the generalized gamma distribution has similar "tail behavior" or includes as special or limiting cases, the lognormal, Weibull, simple gamma, exponential, normal and Pareto distributions.

While the Pareto distribution has frequently performed well in describing city size distributions, there is no edict requiring that only this distribution should be used. Indeed, the Pareto distribution has the decidedly inconvenient feature of being defined over the domain $1 < t < \infty$. Furthermore, the mean of this distribution goes to infinity as its shape parameter falls to one or below. (This is an important problem, because fitted values of this parameter below unity have frequently appeared for city size distributions. See Rosen and Resnick's results.) This idiosyncracy makes the Pareto distribution somewhat inhospitable for use in a non-normal regression model.

If we are willing, however, to switch to the more highly parameterized generalized gamma family of distributions, incorporation of alternative distributional assumptions into a regression model becomes quite feasible. The literature on "survival analysis" or "product life testing" (i.e. Lawless, 1982) contains extensive analyses of this distribution and its special cases in both a simple univariate distribution-fitting context, as well as in log-linear non-normal regression models. For regression situations where the dependent variable can take on only non-negative values, and where there is no censoring or truncation involved, Cameron and White (1985) provide an extensive description of the incorporation of the generalized gamma distribution into both linear and log-linear regression models, as well as an empirical application which illustrates that applied economists should be concerned not only about the functional form of their assumed relationship between the dependent and explanatory variables, but also about the distributional assumptions they adopt for the conditional distribution of the dependent variable. Limiting one's attention to the usual normality or log-normality assumptions may be inappropriate.

It is quite simple to summarize the process of generating a regression model where the dependent variable--here, city size--has a non-normal distribution. Rosen and Resnick limit the explanatory variables to country-level characteristics, but it is entirely conceivable that attributes of the city itself could be called upon to explain a larger- or smaller-than-anticipated size. City sizes t might be assumed to follow a generalized gamma (GG) distribution, but the mean of that distribution may well depend upon a vector of "explanatory" variables, x_{ijr}, (where i=city and j=country), r = 1,...,k.

The reader is referred to the Cameron and White paper for a full discussion of generalized gamma regression models, but the basic steps for the linear version of the model are as follows:

a.) Start with the formula for the unconditional density function;

b.) Let the mean of the distribution equal $x'\beta$;

c.) Solve the equation in (b.) for the scale parameter of the density in terms of $x'\beta$ and the other parameters of the distribution (the shape parameters);

d.) Substitute the expression from (c.) into the original density function (a.) wherever the previously constant scale parameter appears. The result is a density function, the mean of which varies with the values of a set of explanatory variables.

In Rosen and Resnick's two-stage Pareto formulation, a larger value for the "a" exponent implies a city size distribution with a higher proportion of small cities and fewer large cities. This would correspond in the generalized gamma model to a smaller value of the linear combination $x'\beta$. Here, however, a range of explanatory variables is called upon directly to account for variations in the scale parameter.

One shortcoming of city size data has been overlooked thus far, however. Population data are typically reported only for cities larger than a certain size. City sizes smaller than this level are unobserved. The full distribution of city sizes must be inferred from the shape of the upper tail of the distribution. Survival data analysts are familiar with the problem of unobservable portions of the distributions they are attempting to fit, but their problem is fundamentally different than the one which arises here. Survival time is observed exactly for products which fail (or for patients who die) within the time span of the study. For those products which don't fail (or patients who don't die), we know that their survival time is larger than the observation period, but we also know specifically how many elements of our sample fall into this category and their characteristics (specifically, the values of their associated "explanatory variables"). This is an example of a distribution which is "censored." For city size, however, the available data rarely include even a footnote about how many cities are smaller than the reporting threshold size. It may not even be objectively clear what constitutes a "city", if smaller settlements were recorded. We know nothing at all about these smaller cities. The observable distribution is therefore "truncated."

GG regression models are complicated but not impossible to specify and to estimate in the censored case. However, due to the fact that the GG distribution function does not have a closed form, it would be considerably more difficult and expensive to formulate and estimate a truncated GG regression model. However, the two-parameter Weibull distribution, one of its special cases, has not been explored and may well be appropriate (see Hastings and Peacock, 1974). The attractiveness of this distribution stems from its convenient closed-form cumulative distribution function.

4. A Truncated Weibull Regression Model to Explain City Sizes

A Weibull-distributed random variable t is defined over $0 < t < \infty$. Scale parameter b and shape parameter c are both constrained to be strictly positive. The probability density function is:

$$f(t) = \frac{c\ t^{c-1}}{b^c} \exp\left[-(t/b)^c\right] \qquad (2)$$

The associated cumulative distribution function is:

$$F(t) = 1 - \exp[-(t/b)^c] \tag{3}$$

For non-truncated data, the mean of the univariate Weibull distribution might be assumed to be a linear-in-parameters function of a vector of explanatory variables, x_{ijr}, $r = 1,\ldots,k$. (Subscripts will occasionally be suppressed.)

$$E(t) = b\,\Gamma[(c+1)/c] = x'\beta \tag{4}$$

$$\Rightarrow b = x'\beta \,/\, \Gamma(c^*) \qquad \text{where } c^* = (c+1)/c$$

Thus, for the regression model, we will have

$$f(t_{ij}|x_{ij}) = \frac{c\, t_{ij}{}^{c-1} \exp\,[-(t_{ij}/(x'\beta/\Gamma(c^*)))^c]}{(x'\beta/\Gamma(c^*))^c} \tag{5}$$

$$F(t_{ij}|x_{ij}) = 1 - \exp\,[\,-(t_{ij}/(x'\beta/\Gamma(c^*)))^c]$$

For a distribution truncated from below at t_j^*, the probability of observing $t = t_{ij}$, given x_{ij}, will be given by:

$$f^*(t_{ij}|x_{ij}) = \frac{c\, t_{ij}{}^{c-1}\ \exp[-(t_{ij}/(x'\beta/\Gamma(c^*)))^c]}{(x'\beta/\Gamma(c^*))^c\ \exp[-(t_j^*/(x'\beta/\Gamma(c^*)))^c]} \tag{6}$$

Which simplifies readily to:

$$f^*(t_{ij}|x_{ij}) = \frac{c\, t_{ij}{}^{c-1}}{(x'\beta/\Gamma(c^*))^c}\ \exp\left[\frac{(t_j^{*c} - t_{ij}{}^c\ )}{(x'\beta/\Gamma(c^*))^c}\right] \tag{7}$$

For each observation, then, one must have data on the size of the city, t_{ij}, the minimum size of city reported for the country in question, t_j^*, as well as a vector or country and perhaps city-specific explanatory variables, x_{ij}.

The log-likelihood function for a sample of N cities will be:

$$\log L = n(\log c + c \log\Gamma(c^*)) + c \sum \log(t_{ij}/x'\beta) \quad (8)$$
$$- \sum \log t_{ij} + \Gamma(c^*)^c \sum [(t^*_j/x'\beta)^c - (t_{ij}/x'\beta)^c]$$

where sums are over i from 1 to N, and j indexes particular countries.

Optimal values of the parameter vector β and the distribution shape parameter c can be found by maximizing this log-likelihood function, given the data. Nonlinear optimization computer programs such as GQOPT can search for the optimum using numerically computed derivatives, or analytic derivatives can be provided for more economical estimation. Analytic derivative formulas for the gradient vector and the Hessian matrix corresponding to this log-likelihood function are provided in the Appendix. When optimal values of the unknown parameters have been achieved, the asymptotic variance-covariance matrix for the parameters is given by the inverse of the Hessian matrix, evaluated at the optimal parameters.

The fitted model can be interpreted in much the same way as are ordinary least squares regression models. In OLS, we assume that the conditional distribution of the dependent variable is normal, with its mean determined by $x'\beta$. In the absence of heteroskedasticity, the variance of this conditional distribution is assumed to be constant. Here, we are merely assuming instead that the conditional distribution for the dependent variable is Weibull. We again let the expected value of the conditional distribution be $x'\beta$. We assume that the shape parameter, c, is constant. The β parameters can, as usual, be interpreted as the effect on mean city size of a unit change in the level of each explanatory variable.

5. Implications

We have noted some limitations of existing empirical work analyzing city size distributions. To overcome some of these, we have outlined a richer, more-general model which could conceivably be employed with existing data sets (possibly after they have been augmented with more information) to yield results which are potentially of more interest to policy-makers than many of the models which have been employed so far. Particularly promising is the possibility that the parametric "distribution-fitting" branch of the literature can soon be extended via non-normal regression models (such as the one proposed here) to accommodate the influence of a variety of explanatory variables upon the mean of the distribution. Once these techniques are developed, it will be possible to incorporate more of the implications of the economic theory behind city size distributions developed by researchers such as Beckmann (1958), Tinbergen (1967), Beckmann and McPhearson (1970), and more-recently, Alperovich (1982a, 1982b) and Taylor (1986).

Researchers up until now have endeavored to specify models which have some of the features of the "behavioral" model proposed here. However, the unqualified use of two-stage OLS estimation procedures yields unreliable standard error estimates. Our models continue in the spirit of these attempts, but "collapse" the models so that they may be estimated in one step and will yield acceptable asymptotic standard errors. We have also been careful to address the fundamental truncation of city size distributions in our model. Failure to treat this data truncation explicitly could easily result in biased parameter estimates.

We hope that the methodologies proposed here will be tested empirically. The results should add significantly to our understanding of the determinants of city sizes.

REFERENCES

1. G. Alperovich, Capital-labor substitution and city hierarchies, Ann. Reg. Sci., 16, 57-63 (1982a).

2. G. Alperovich, Scale economies and diseconomies in the determination of city size distribution, J. Urban Econ., 12, 202-213 (1982b).

3. G. Alperovich, The size distribution of cities: on the empirical validity of the rank-size rule, J. Urban Econ., 16, 232-239 (1984).

4. M.J. Beckmann, City hierarchies and the distribution of city size, Econ. Develop. and Cultural Change, 6, No.3, 243-248 (1958).

5. M.J. Beckmann and J.C. McPhearson, City size distribution in a central place hierarchy: An Alternative approach, J. Reg. Sci., 10, 25-33 (1970).

6. B.J.L. Berry, City size distribution and economic development, Econ. Develop. and Cultural Change, 9, No. 4, 573-587 (1961).

7. T.A. Cameron and K.J. White, Generalized gamma family regression models for long-distance telephone call durations, Discussion paper #363, Department of Economics, University of California, Los Angeles, January 1985.

8. J. Esteban, Income-share elasticity, density functions and the size distribution of income, mimeographed manuscript, University of Barcelona, 1981.

9. GQOPT: A Package for Numerical Optimization of Functions, distributed by Richard Quandt, Department of Economics, Princeton University, Princeton, N.J. 08544.

10. N.A.J. Hastings and J.B. Peacock, "Statistical Distributions," John Wiley and Sons, New York (1974).

11. J.F. Lawless, "Statistical Models and Methods for Lifetime Data," John Wiley and Sons, New York (1982).

12. J.B. McDonald, Some generalized functions for the size distribution of income, Econometrica, 52, 647-663 (1984).

13. E.S. Mills and B.W. Hamilton, Urban Economics, Third Edition, Scott, Foresman and Company, Glenview, Illinois. (1984)

14. H.W. Richardson, Theory of the distribution of city sizes: Review and prospects, Reg. Studies, 7, 239-251 (1973).

15. K.T. Rosen and M. Resnick, The size distribution of cities: an examination of the pareto law and primacy, J. Urban Econ., 8, 165-186 (1980).

16. C.A. Taylor, "Spatial Utility Equilibrium and City Size Distribution in a Central Place System," J. Urban Econ., 19, 1-22 (1986).

17. J. Tinbergen, The hierarchy model of the size distribution of centres, Papers Reg. Sci. Assoc., 20, 65-68 (1967).

18. G.K. Zipf, "Human Behavior and the Principle of Least Effort," Addison-Wesley, Cambridge, Mass. (1949)

APPENDIX I

GRADIENT AND HESSIAN ELEMENTS FOR A TRUNCATED WEIBULL REGRESSION MODEL OF CITY SIZES

The elements of the gradient vector corresponding to the log-likelihood function given in equation (8) in the body of this paper are given as follows. First, we adopt some simplifying notation (all sums are from 1 to N):

$$Z = z/x'\beta, \quad \text{for } z = x_{ijr},\ t_j^*,\ t_{ij}$$

$$\Gamma'(z) = \partial\, \Gamma(z)/\, \partial z$$

$$\Gamma''(z) = \partial^2\, \Gamma(z)/\, \partial z^2$$

$$\Psi(z) = \partial \log \Gamma(z)/\, \partial z$$

$$\Psi'(z) = \partial\, \Psi(z)/\, \partial z$$

The derivatives are then:

$$\frac{\partial \log L}{\partial \beta_r} = -\ c \sum X_{ijr} - c\, \Gamma(c^*) \sum \{ X_{ijr} [T_j^{*c} - T_{ij}^{\ c}] \}$$

$$\text{for } r = 1,\ldots,k$$

$$\frac{\partial \log L}{\partial c} = (n/c)\ [1 - \Psi(c^*)] + n \log \Gamma(c^*) + \sum \log T_{ij}$$

$$-\ [\ \Gamma'(c^*)/c^2\ ] \sum [\ T^*_j{}^c - T_{ij}{}^c\ ]$$

$$+\ \Gamma(c^*) \sum [\ T^*_j{}^c \log T^*_j - T_{ij}{}^c \log T_{ij}\ ]$$

The elements of the corresponding Hessian matrix are given by:

$$\frac{\partial^2 \log L}{\partial\beta_r \partial\beta_s} = c \sum X_{ijr} X_{ijs}$$

$$+\ c\ (1 + c)\ \Gamma(c^*) \sum \{ X_{ijr} X_{ijs} [T_j^{*c} - T_{ij}^{\ c}] \}$$

$$\text{for } r,s = 1,\ldots,k$$

$$\frac{\partial^2 \log L}{\partial \beta_r \partial_c} = -X_{ijr} + [\Gamma'(c*)/c - \Gamma(c*)] \sum \{ X_{ijr} [T_j*^c - T_{ij}^c]\}$$

$$-c\ \Gamma(c*) \sum \{ X_{ijr} [T_j*^c \log T_j* - T_{ij}^c \log T_{ij}] \}$$

$$\text{for} \quad r = 1,\ldots,k$$

$$\frac{\partial^2 \log L}{\partial c^2} = -(n/c^2)(1 - 2\ \Psi(c*) + \Psi'(c*)/c)$$

$$+ [(\Gamma''(c*) + 2\ \Gamma'(c*))/c^4] \sum [T_j*^c - T_{ij}^c]$$

$$- [2\Gamma'(c*)/c^2] \sum [T_j*^c \log T_j* - T_{ij}^c \log T_{ij}]$$

$$+ \Gamma(c*) \sum [T_j*^c (\log T_j*)^2 - T_{ij}^c (\log T_{ij})^2]$$