The Beaman Distribution: A New Descriptive Model for the Size Distribution of Incomes

Fiona C. Maclachlan and John E. Reith*

A new two-parameter descriptive model for the size distribution of incomes is presented. The Beaman distribution, originally found to describe the relationship between prices and sales volumes of textiles and related products, is shown to fit closely U.S. Census data on individual incomes over the past twenty years. The two parameters of the model represent the minimum and median levels of income. An important advantage of the Beaman distribution over other two-parameter models is that negative incomes can be accounted for.

KEYWORDS: Income distribution, gamma distribution, Weibull distribution.
JEL: C160, D310

* *Manhattan College and E. I. Du Pont de Nemours & Co. (retired), respectively*. An accompanying *Mathematica*® notebook is available at *home.manhattan.edu/~fiona.maclachlan/beaman*.

I. Introduction

Since Pareto many attempts have been made to discover a mathematical relationship that describes precisely the skewed distribution of income. Pareto's distribution with its two parameters is parsimonious, but the fit to most data sets is poor. On the other extreme, the generalized beta with five adjustable parameters (McDonald & Xu 1995) can be made to fit income data almost perfectly. In selecting the optimal point of tradeoff between parsimony and fit, a judgment has to be made about the relative importance of the two qualities. We shall argue below that there are slight but systematic imperfections in the income data used to test the models, meaning that a better fit does not necessarily imply a better model. In considering the trade-off between parsimony and fit, therefore, the value of parsimony ought to be given greater weight than if the data were perfect.

McDonald (1984) identifies four two-parameter models in the modern literature. He ranks the four models by goodness of fit to family income data from 1970, 1975 and 1980 as follows: the gamma, Weibull, Fisk, and lognormal.[1] The gamma distribution, originally fitted to income data by Amorosa (1925), was introduced to English speaking readers by A. B. Z. Salem and T. D. Mount (1974). The Weibull distribution, used by Swedish engineer Wallobi Weibull to model metallurgical failure rates, was later found to have a wide variety of applications outside of engineering (Weibull 1951). Peter R.

[1] A more recent study (Bandourian et al. 2002) examines international data and finds the Weibull ranks ahead of the gamma.

Fisk (1961) proposed what he called the sech^2 distribution[2] for incomes within specific occupations. Finally, the lognormal distribution was introduced as a model of the size distribution of incomes by French economist, Robert Gibrat (1931).

In this paper we introduce an additional two-parameter model. Using 23 years of data on male and female incomes we show that it consistently provides a better fit to the data than the lognormal or Fisk, and is competitive with the gamma and Weibull. Furthermore, unlike any of the existing two-parameter distributions, the distribution presented here allows for the possibility of negative incomes. If capital losses are to be included in income measures, the possibility of negative incomes becomes an important consideration.

---

[2] Fisk (1961, p.175) gives the CDF of his distribution as $F(t) = \dfrac{(t/t_0)^{\alpha}}{1+(t/t_0)^{\alpha}}$.

II. The Distribution

The distribution has its origin in the textile market research group at duPont in the 1970's. With access to consumer spending data from a sample of roughly 10,000 households, the group was able to examine sales volume at different prices for various products. Plots of the data revealed a distribution that was stable through time and across products. Apparently following a lognormal distribution, a better fit was found with a distribution for which the CDF is given by:

$$F(x) = \frac{1}{\frac{(b-a)^3}{(x-a)^3} + 1}$$

The distribution was found through a trial and error process using monthly consumer spending data on over 100 end uses for a period of ten years. The duPont researchers found that setting the exponent just under three[3] lead to the best fit but decided to round it up to three for aesthetic and practical reasons. The two parameters have direct economic interpretation. One represents the minimum price (or income), while the other represents the median. Figure 1 illustrates the fitted distribution using data points for the income for males in 2002, and shows the position of the two parameters along the horizontal axis.

Reith (1986) first introduced the equation in an unpublished paper in which he presents a number of results based on his market research at duPont. He labeled the distribution after the late Ralph G. Beaman, the colleague in the market research group at duPont who contributed the most in its formulation. While it was not possible to recover

[3] Reith's best memory of the exponent is 2.988.

the underlying data from the product market studies, we discovered that the distribution provides a remarkably good fit to publicly available income data.

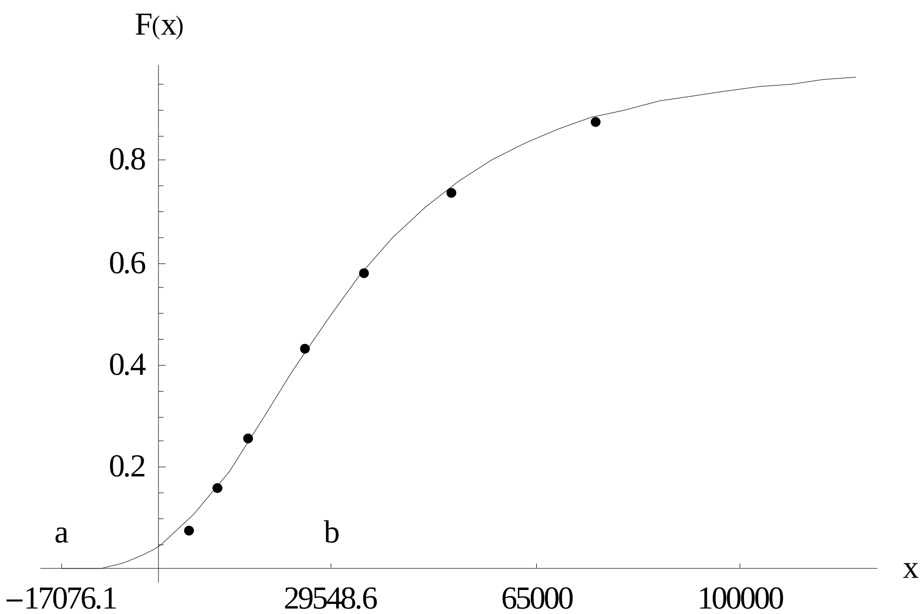

FIGURE 1.—Beaman CDF fitted to 2002 male incomes.

A notable difference between the results of the product market studies and our work on income distribution presented here is the large negative value that we find in estimating the minimum value parameter. Over the 23 years, the magnitude of the negative intercept is usually over one half the estimated median. For example, in 2002 the male median income is estimated at $29,549, while the minimum is estimated at negative $17,076. For product markets, estimates of the minimum price were rarely negative, and only slightly negative when the market prices were very low.

III. The Data

The data used to test the fit are drawn from the Statistical Abstract of the United States (1995-2004) and cover the 23 years, 1970, 1980-1993, 1995-2002. We consider individual incomes, labeled “income of people” by the Census, for males and females. Since many studies of income distribution are addressing the social justice issues of poverty and inequality, it is common to examine family income data. Our focus, however, is on the purely empirical question of the market determined distribution of claims to output across individuals.

Another result of the traditional focus on inequality and poverty is the way in which the Census measures income. Of particular importance to our study is the fact that government transfer payments are included, and capital gains are not. The Census Bureau (2003, p. 2) provides the following definition:

> Money income (MI) is collected for all people in the sample 15 years old and over. Money income includes earnings, unemployment compensation, workers’ compensation, social security, supplemental security income, public assistance, veterans’ payments, survivor benefits, income from estates, trusts, educational assistance, alimony, child support, assistance from outside the household, and other miscellaneous sources. It is income before deduction for taxes or other expenses and does not include lump-sum payments or capital gains.

If lower income individuals are more likely to receive a significant portion of their income in the form of government transfers, then one should expect the official data points on the left hand tail of the distribution to lie below the pure market determined values. For example, if the U.S. Census says that 8% of the population earns less than $5,000, then the percentage when transfers are not included would be larger than 8%.

Interestingly, the fitted Beaman distribution is more likely than not to cut above the minimum data point, as can be seen in Figure 1.[4]

The absence of capital gains from the income measures precludes the possibility of negative incomes and leads to an incomplete picture of income distribution. An underlying theory of income distribution that makes reference to the preference for risk (e.g., Friedman 1953) would ideally be tested with data that includes capital losses. As larger proportions of the population earn their income from invested savings, capital losses become an increasingly important consideration.

IV. Comparing the Fit

In order to compare the fit of the five two-parameters models, we find parameter estimates using the Levenberg-Marquardt algorithm in the *Mathematica 5.0* nonlinear regression package. Figures 2 and 3 show the sum of the squared residuals for the five distributions across the 23 years. We see that the Beaman distribution generates a sum of squared residuals that is competitive with those from the Weibull and gamma distributions. The fit of the Beaman distribution relative to the gamma and Weibull is better for males, than for females. The inclusion of government transfer payments, alimony and child support in income measures may be responsible for this discrepancy.

---

[4] A complete set of plots can be viewed at:

home.manhattan.edu/~fiona.maclachlan/beaman

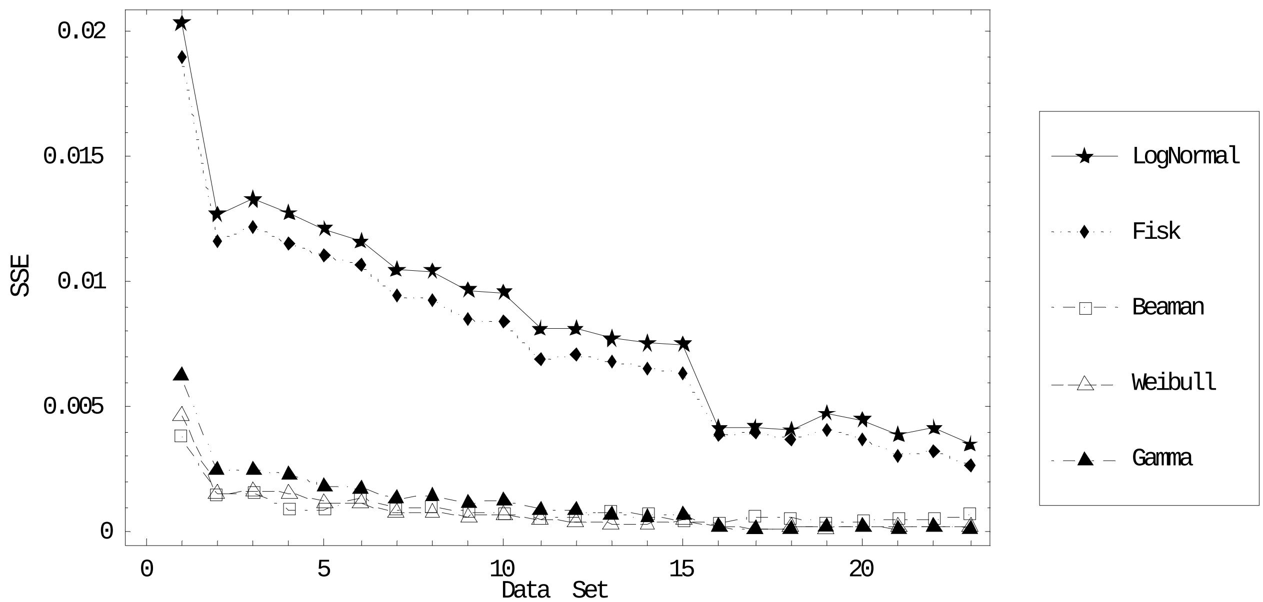

FIGURE 2.—Sum Squared Errors for Male Incomes.

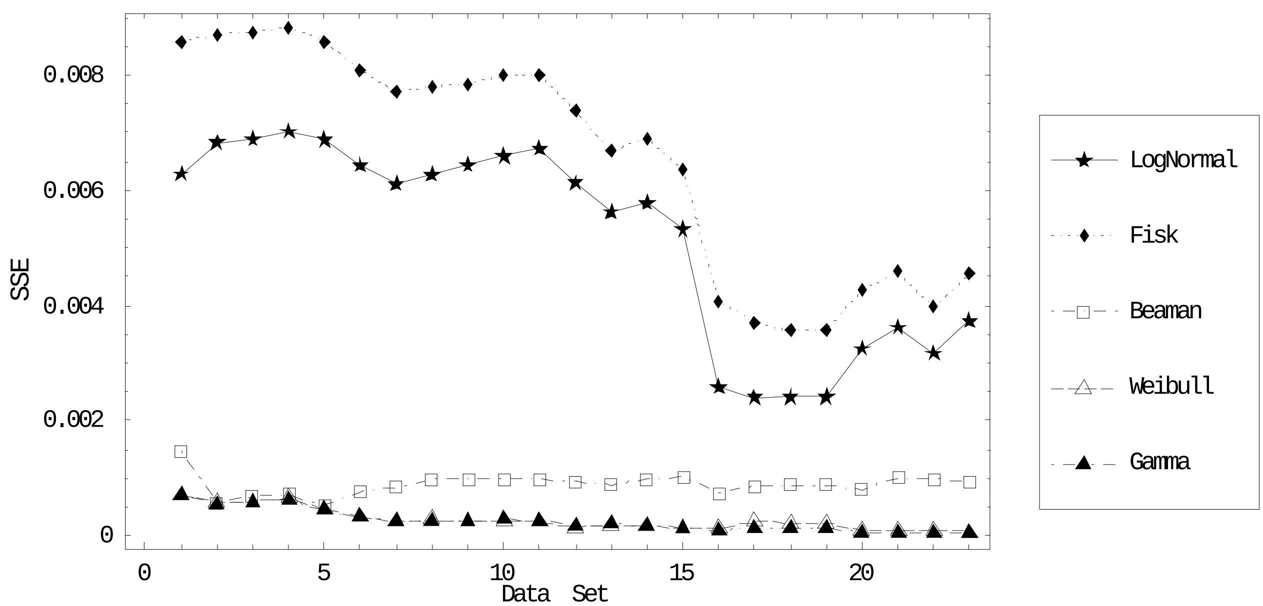

FIGURE 3.—Sum Squared Errors for Female Incomes.

V. Conclusion

In this paper, we address the question originally posed by Pareto. Is there a stable mathematical relationship that describes the distribution of income generated in a market economy? One difficulty in answering this question is that the data provided by the Census on individual incomes are not a pure reflection of market outcomes. The inclusion of transfer payments in income measures is helpful for researchers charting poverty and inequality but it adds confusion to the question of which statistical distribution reflects most accurately the results of economic forces. Similarly, the absence of capital gains from the measures creates an incomplete picture of an underlying economic reality in which a growing proportion of the population earns an income from invested savings.

The Beaman distribution is a parsimonious descriptive model of the size distribution of income that stands up well relative to existing models and has the added advantage that it can account for negative incomes. Furthermore, since the official measures of income with which we are testing the models include various transfer payments, the fit of the Beaman distribution might be even better than it appears.

REFERENCES

AMOROSA, L. (1925): “Ricerche Intorno alla Curva dei Redditi,” *Annali di Mathematica Pura ed Applicata,* Ser. 4-21, pp. 123-159.

BANDOURIAN, R., J.B. MCDONALD AND R. S. TURLEY (2002): “A Comparison of Parametric Models of Income Distribution Across Countries and Over Time” (June 2002), Luxembourg Income Study Working Paper No. 305.

FISK, P. R. (1961): “The Graduation of Income Distributions,” *Econometrica*, Vol. 29, No. 2 (Apr., 1961), pp. 171-185.

FRIEDMAN, M. (1953): “Choice, Chance and the Personal Distribution of Income,” *Journal of Political Economy*, Vol. 61, No. 4 (Aug., 1953), pp. 277-290.

GIBRAT, R. (1931): *Les Inégalités économiques,* Paris: Recueil Sirey.

MCDONALD, J. B. (1984): “Some Generalized Functions for the Size Distribution of Income,” *Econometrica*, Vol. 52, No. 3 (May, 1984), pp. 647-664.

______________ AND Y. J. XU (1995): “A Generalization of the Beta Distribution with Applications,” *Journal of Econometrics*, Vol. 66, No. , pp. 133-152.

REITH, J. E. (1986): "Some Rules of Market Behavior," mimeo.

SALEM, A. B. Z. AND T. D. MOUNT (1974): “A Convenient Descriptive Model of Income Distribution,” *Econometrica*, Vol. 42, No. 6 (Nov., 1974), pp.1115-1127.

US CENSUS BUREAU (2003): “Income in the United States: 2002,” *Current Population Reports*, P60-221, Washington D.C.: Government Printing Office.

----------------------- (1995-2004): *Statistical Abstracts of the United States*, Washington D.C.: Government Printing Office.

WEIBULL, W. (1951): “A Statistical Distribution Function of Wide Applicability,” *Journal of Applied Mechanics*, Sept. 1951, pp. 293-297.