ON FITTING THE PARETO LAW TO INCOME DISTRIBUTIONS

Summary

The current method of fitting the Pareto distribution is shown to have a number of serious defects. In the first place, it is usually fitted to the entire range of income observed without first finding out the range where the Pareto law can give a good fit. Second, unweighted least squares gives wrong weightage to different points ; weighted least squares should be used, but even that is not perfectly satisfactory. Finally, the current method does not ensure that the expected frequencies add up to the total number of observations—a defect which can be easily remedied.

Maximum likelihood and other methods of estimation are proposed.

For studies on differences in concentration, it would be safer to calculate the concentration measures directly, than depending on the Pareto constant[1].

1. *Introduction.*

The standard method of fitting the Pareto curve to size-distributions of income may be outlined as follows :

Given a grouped size-distribution showing n_j incomes in the size-class $x_{j-1} - x_j$ (where $j=1, 2, \cdots, k$ and $x_k = \infty$), the first task is to find out for each x_j ($j=0, 1, \cdots, k-1$) the number $N(x_j)$ of incomes not less than x_j. Clearly $N(x_j) = n_{j+1} + \cdots + n_k$. Next one plots a graph showing $N(x_j)$ against x_j ($j=0, 1, \cdots, k-1$) on log-log scale. This graph should be linear if the Pareto law holds i.e., if $N(x) = \frac{A}{x^{\alpha}}$. One should find out from this graph how many points counting from that corresponding to x_{k-1} towards the left seem to fall approximately on a straight line. The Pareto law is fitted to these points by minimising $\Sigma \left[\log N(x_j) - \log A + \alpha \log x_j \right]^2$ where the sum is over the chosen set of points $[x_j, N(x_j)]$.

[1] Some of the observations made below were mentioned very briefly by Johnson [2].

2. *Defects of the Standard Method of Fitting.*

We first point to the defects of the common method of fitting by unweighted least squares :

(a) First of all, the Pareto law is very often fitted to the entire range of income for which the distribution of persons is available, even though the graph mentioned above shows considerable non-linearities near the lower income brackets. This is against the spirit of the Pareto law which is avowedly concerned with the "sufficiently high" income levels only. As a result, the Pareto law very often gives a remarkably poor fit to empirical data. Consider, for instance, the text-book example presented in Davis [1, p. 26] and reproduced in Lange [3], based on US data for 1918[2].

income in dollars	no. of incomes cumulated from above (000)	
	observed	fitted
500	35541	50722
1000	23010	15648
1500	10512	7864
2000	5290	4827
3000	2225	2426
5000	842	1020
10000	254	315
25000	62	66
50000	21	21
100000	7	6
200000	2	2

(b) Why is it, then, that Davis and others are satisfied with such fits ? The explanation is that the graph above-mentioned is deceptive and gives rise to complacence even where the observed and 'expected' frequencies seriously diverge. In virtue of the properties of logarithms, the graph shows the same vertical deviation when, for example, the observed and expected frequencies are 6 and 2 respectively as when the two frequencies are 600,000 and 200,000 respectively ; but the divergence is far more serious, from the statistical view-point, in the second case. This is why the extremely serious deviations near the lower income brackets are understated in the graph.

(c) This brings us to the more fundamental question of the rationale of using unweighted least squares for estimation. There seems

[2] Although goodness of fit tests are not to be taken seriously, can one ignore the fact that χ^2 runs to millions (whatever definition of χ^2 is used in this case, where observed and expected frequencies have different totals) ?

to be little logic behind the method. Like the graph, the method also gives equal weightage to deviations between 6 and 2, and between 600,000 and 200,000, say. In effect, the points for high incomes receive much larger weight. This is why the fit is generally good at high income ranges. Weighted least squares would, of course, be an improvement. It is easy to see that under assumptions of random sampling (which are subject to criticism)

$$V[\log N(x)] \sim [C.\ V.\ \text{of}\ N(x)]^2 = \frac{Q(x)}{E\{N(x)\}}$$

where $Q(x)$ =proportion of observations below x and the other symbols have usual meanings. We should therefore minimise[3]

$$\sum_{0}^{k-1} \frac{N(x_j)}{Q(x_j)} [\log N(x_j) - \log A + \alpha \log x_j]^2.$$

Even this is not perfect, since the different deviations are correlated. But this would give greater weightage to the higher points for lower income values and remove a great defect of the common method.

(d) Another serious defect of the common method is that the expected frequencies do not add up to the total of observed frequencies, viz., $N(x_0)$. This also can be easily avoided ; one can estimate α by some suitable method and obtain A from the relation

$$N(x_0) = \frac{A}{x_0^{\alpha}} \quad \cdots \quad \cdots \quad \cdots \quad \cdots \quad (1)$$

Indeed the Pareto law is a uniparameter law

$$\left.\begin{aligned} N(x) &= N(x_0) \left(\frac{x_0}{x}\right)^{\alpha} \\ \text{i e. } F(x) &= 1 - \left(\frac{x_0}{x}\right)^{\alpha} \end{aligned}\right\} \quad \cdots \quad \cdots \quad (2)$$

and there is little point in trying to estimate the parameter A in any other way.

---

[3] Here and subsequently we assume that the points selected for fitting the Pareto curve are numbered 0, 1, 2, ..., k-1.

3. *The Maximum Likelihood and other Methods.*

The method of estimating A given above may always be used. As regards α, two simple methods suggest themselves, viz., (i) equating observed mean $\bar{x}$ to the theoretical expression $\frac{\alpha x_0}{\alpha - 1}$, and (ii) equating L, the observed Gini-Lorenz concentration coefficient[4], to the theoretical expression $\frac{1}{2\alpha - 1}$. The latter is perhaps the best procedure (vide Section 4); but it may also be noted that assuming that the observations form a random sample from population defined by equation (2), the maximum likelihood estimate of α is obtained by solving

$$\sum_{1}^{k-1} n_j \frac{x_j^{\alpha} \log x_j - x_{j-1}^{\alpha} \log x_{j-1}}{x_j^{\alpha} - x_{j-1}^{\alpha}}$$

$$= \sum_{1}^{k-1} n_j \left(\log x_{j-1} + \log x_j \right) + n_k \log x_{k-1} - N(x_0). \log x_0$$

This solution can be done by standard methods; the r. h. s., it may be noted, is independent of α.[5]

To recapitulate, it is most important to find out from the graph (say) from what income upward the Pareto law really gives a good fit, and not to fit the law blindly over the whole (income-tax) range as done too often. For, while the Davis data cited in section 2 give α =1·69 by the current least-squares method, the estimate is 1·45 when we use equation (1) and $\bar{x} = \frac{\alpha x_0}{\alpha - 1}$, 1·82 when we use equation (1) and $L = \frac{1}{2\alpha - 1}$, and only 1·13 when equation (1) and the maximum likelihood method are used. For another data of Szeliski [1, p. 29] the corresponding estimates are : least squares—1·48, based on mean—1·69, based on Lorenz ratio—1·81 and maximum likelihood —1·66. Actually, the Pareto law cannot give a good fit over the whole

[4] Estimated by graphical methods or by numerical quadrature.

[5] Apart from deviations from random sampling, the income data are known to be defective; so the maximum likelihood method is not to be taken very seriously.

range and all these estimates are useless. Second, one must estimate A by equation (1). And third, one should give proper weightage to the different deviations $\log N(x) - \log A + \alpha \log x$; but if this is done, while the range of income is too wide for Pareto law to give a good fit, improved methods may give a Pareto line passing close to the higher points which would receive large weight but very far from the lower ones which would get small weights.

4. *Studies on Concentration.*

The Pareto curve has often been fitted in connection with studies on concentration of income : changes in the estimates of α have been the basis of observations on changes in concentration. The procedure is fraught with great risk unless the Pareto fit is very close, which is not the case in many applications.

Thus, Patel [4] used the common method for fitting Pareto curves to the distributions of all income-tax assessees in India (individuals plus Hindu undivided families) separately for several years. The range chosen was Rs. 5,000 per year and above. We give below the estimates of concentration coefficient obtained from estimated values of α and also those directly obtained by graphical methods by plotting the concentration curve. The large discrepancies cannot be due to the subjective element in the graphical method.

	1950-51	1955-56	1956-57	1957-58	1958-59	1959-60
1. From Pareto fit	0·449	0·396	0·377	0·366	0·348	0·340
2. Direct	0·477	0·437	0·436	0·424	0·416	0·415

Davis [1, pp. 34-35 and p. 429] also uses the estimates of α in the same way. The examples given above and the following two from Davis' book should suffice to illustrate the risk involved in drawing conclusions about changes in concentration from the estimates of α. For the distribution of income among personal income recipients in the US for the year 1918 (over the range $500 and above) shown in Davis [1, p. 26] the Pareto fit gives $L=0·418$, while direct graphical method gives $L=0·379$; again for the distribution of Szeliski for the year 1929 (over the range $1000 and above) shown in Davis [1, p. 29], the Pareto fit gives $L=0·510$, while direct graphical method gives $L=0·382$.

It should be clear that the estimate of Lorenz ratio from Pareto can err by 0·05 or even 0·1, which is too much for an index which normally ranges between 0·25 and 0·50. To estimate the share of top 10% or bottom 50% (say) from the Pareto curve is even more risky In general, for studies on changes in concentration it would be safer to estimate L directly than depending on the estimates of the Pareto constant α.

REFERENCES

1. Davis, H. T. The Theory of Econometrics, The Principia Press, Inc. Bloomington, Indiana, 1941.
2. Johnson, N. O. "The Pareto Law", The Review of Economic Statistics, Vol. 19, 1937, pp. 20-26.
3. Lange, O. Introduction to Econometrics, Pergamon Press, London, 1959.
4. Patel, R. C. "Recent Trends in Distribution of Personal Income", The Economic Weekly, Bombay, Vol. XV, No. 10, 1963, pp. 445-448.

N. BHATTACHARYYA & R. N. MUKHERJEE