NOTES AND MEMORANDA

PARETO AND THE DISTRIBUTION OF INCOME*

BY JOHN CREEDY

University of Reading, England

I. INTRODUCTION

Pareto's analysis of the distribution of income between individuals is often cited in favourable terms as a pioneering example of disinterested econometric research.[1] Indeed, Pareto's "Law"[2] became so widely accepted that a great deal of subsequent theoretical work has been devoted to the development of models of "genesis" which are capable of producing a Pareto distribution.[3]

Pareto's main conclusions were clearly (and repeatedly) stated. They are (a) that the distribution of income does not depend on chance or stochastic factors;[4] and (b) that there is considerable stability in the distribution, both over time and across countries.[5] These in turn led to the statement that any artificial attempts to alter the distribution would be futile. This latter view was also dominant in the contemporary writings of the Social Darwinists,[6] and although Pareto modified the extreme form of this statement in the later Manuel he still seemed to regard his "law" as fundamental.[7]

In a recent review of mathematical and statistical theories of income distribution Brown (1976) has examined Pareto's reasons for rejecting "le hasard"

*I am grateful to a referee for very helpful comments on an earlier draft of this paper.

[1]It is, however, worth noting Schumpeter's comment (1968) "... into everything that was not a theorem in the pure logic of economics the whole man and all the forces that conditioned him entered so unmistakably that it is more necessary than it usually is in an appraisal of scientific performance to convey an idea of that man and of those forces".

[2]The inappropriate use of the word "law" is interesting here. As Tawney, (1920) noted, to argue that a stable relationship implies that the distribution cannot be changed is to be, "... guilty of an innocent misunderstanding of the nature of economic laws". We also have "Gibrat's law of proportionate effect". Both Gibrat and Pareto were engineers before turning to economics.

[3]Well known examples are Champernowne (1953), Lydall (1968), Mandlebrot (1960); See Steindl (1965) for further cases. The Pareto distribution is often used for interpolation in official statistics (and even to stimulate more intensive search for tax evaders by Josiah Stamp).

[4]"La forme de la courbe... ne correspond nullement à la courbe si l'acquisition et la conservation de la richesse ne dépendaient que le hasard" (Pareto (1909)).

[5]"... la statistique nous apprend que la courbe ... varie très peu dans l'espace et dans le temps; des peuples différent et à des époqnes différent donnent des courbes très semblables. Il y a donc une stabilité remarquable dans la forme de cette courbe". (Pareto (1909)).

[6]Only a year after the publication of Pareto's first paper Spencer wrote (1897 p. 260), "What can be more extreme absurdity than that of proposing to improve social life by breaking the fundamental law of social life?" Pareto did *not* of course argue that changes in rank did not take place; Indeed much of his Traite was concerned with the circulation of elites. He also states in the Manuel (1909 p. 386) "La partie intérieure est... en perpétuel movement; tandis que certains individus montent les régions supérieures, d'autres en descendent".

[7]He states, for example (1909 p. 388), "... si la constitution sociale venait a changer d'une facon radicale ... il semble difficile qu'il ait plus de hiérarchie, et la forme de cette hiérarchie pourrait entre semblable à cette qui nous est donnee par les revenues des individus, mais elle ne correspondrait pas à des revenue en argent".

as a factor affecting the distribution. Pareto's view of "chance" seems to have been limited to a simple binomial process, and a consequent comparison with (the upper tail of) the Normal distribution.[8] It is, of course, true that any functional form can be made consistent with a number of stochastic models of genesis of income distribution.[9] It may well be thought that Pareto would have rejected "chance" factors even if his observed distribution had followed the Normal form. Indeed it can reasonably be argued that Pareto's attitude towards the genesis of the distribution was strongly influenced by the apparent stability which Pareto observed.[10] The "stability" conclusion derives from the parameter estimates of the now well known Pareto distribution. Brown (1976) provides a reminder that Pareto actually suggested a more general form of the "courbe des revenus", which was quickly rejected.

The main purpose of this note is, then, to examine in more detail Pareto's discussion of his more general equation and his reasons given for prefering the simple two parameter form. Since his views about income distribution were repeated so often in his economic and sociological work, an analysis of the "scientific status" of this work is of particular interest.

II. The General Pareto Distribution

The best known form of Pareto's "Law" is expressed by the equation:

$$\log N = \log K - \alpha \log x \tag{1}$$

Where x is the level of income, N is the number of people with incomes equal to or greater than x and K and α are parameters. The latter is, of course, independent of the units in which income is measured. In terms of the distribution function $F(x)$, equation (1) can be rewritten as:

$$\log \{1 - F(x)\} = \log A - \alpha \log x \tag{2}$$

(Note that the intercept changes because $1 - F(x)$ represents the *proportion*, while N is the absolute number, of incomes above or equal to x). It is convenient to write $\log A = \alpha \log x_0$, where x_0 is now the "threshold" income level.[11] Whence:

$$F(x) = 1 - (x_0/x)^{\alpha} \tag{3}$$

After presenting the log-linear form Pareto noted a certain amount of concavity and suggested (1896 p. 3) . . . "une seconde approximation du phénomène" . . . ,

[8]See the long footnote in the Cours (1897 p. 316–317).

[9]Pareto's argument was neatly dismissed by Gibrat (1931 p. 193), "Pareto éliminait le hasard sous le prétext que sa formule différait de la formule de Gauss. Démonstration sans valeur, car le 'hasard' peut donner naissance a des repartitions très différent de celle de la courbe en cloche. Les répartitions normales sont l'exception". And later in a footnote, "Il ne faut pas oublier que l'on peut toujours trouver un changement de variable tel que la courbe en cloche devienne une courbe quelconque donée à l'avance".

[10]In Pareto (1909 p. 47) he observed, "Il se peut qu'il n'y ait là qu'une coincidence fortuite, mais elle est assez singuliere pour attenir l'attention".

[11]The use of α as a measure of inequality depends, of course, on the assumption that x_0 remains constant. Pareto did not himself use α (see Pareto (1909) p. 320).

the four parameter form:

$$\log N = \log A - \alpha \log(x+\theta) - \beta x \tag{4}$$

Stating that, (1896, p. 3), "C'est probablement la form générale des courbe de distribution". Nevertheless Pareto only gives one set of estimates of the parameters of (4). For the Grand Duché d'Oldenbourg in 1890 he obtained:

$$\log N = 8.72204 - 1.465 \log(x+220) - 0.0000274x \tag{5}$$

but remarks (1896), "Pour d'autres pays, on a des valeurs de θ et β, encore plus petites et qui, en bien cas paraissent être d'un ordre grandeur inférieur à celui des erreurs d'observation". Pareto then rejects this functional form in favour of his initial log-linear form.[12] This seems rather unusual when it is realised that the values of β and θ are not invariant with respect to the units of measurement, and that β is likely to be very small since the units of x are large and are elsewhere transformed by taking logarithms.[13]

It is also of interest to note a point made by Pareto in (1906 p. 78),

> "Il ne faut pas poursuivre une précision illusoire, et calculer laborieusement un grand nombre de décimales, qui, a fond, ne signifient rien du tout."

And yet Pareto presented the only value of β in equation (5) to seven decimal places!

Pareto did seem to attach more importance to the parameter θ. The inclusion of θ has the convenient effect of making the curve cut the N axis rather than approaching it asymptotically—and has also been found useful in the completely different area of estimating quasi demand curves for recreational facilities.[14] Pareto later reports that after examining distributions of different countries and sources of income (1897 p. 310), "la constant θ est negative, quand il s'agit du produit du travail; elle est positive quand il s'agit de la répartition de la fortune; ell est nulle, ou généralement assex petite, quand il s'agit du revenu totaux".

Nevertheless no values of θ are reported, and Pareto does not consider how a distribution of total income with zero θ could be produced by some combination of separate distributions of "profits" and "wages". Furthermore such a combination is unlikely to give rise to even a "Pareto type" distribution. Consider, for example, the convolution of two Pareto variates x and y where:

$$x \text{ is } F_1(x|\alpha_1, x_0, \theta_1) \text{ and } y \text{ is } F_2(y|\alpha_2, y_0, \theta_2)$$

Then the distribution function of $z = x + y$ is given by:

$$F(z) = \int F_1(z-y)\, dF_2(y)\, dy \tag{6}$$

[12]He states (1896 p. 3) "Il sera donc prudent de ne faire usage qu'avec une grande réserve des résultats de ce cas extrème".

[13]In fact ordinary least squares estimates of $\log N = \log A - \alpha \log x - \beta x$ using Pareto's data gave only 2 out of 17 values of β which were not significantly different from zero. Furthermore 5 cases produced estimates of β which were larger than that obtained for Oldenbourg.

[14]Here the dependent variable is the visitor rate and the independent variable is the distance travelled. See Cheshire and Stabler (1974).

Where integration is over the range of y. Substitution into (6) gives

$$F(z) = 1 - x_0^{\alpha_1}\alpha_2 y_0^{\alpha_2}\int (z-y+\theta_1)^{-\alpha_1}(y+\theta_2)^{-(\alpha_2+1)}\,dy \tag{7}$$

and it is certainly not clear that (7) would produce an $F(z)$ of Pareto's general form.

The "general form" of Pareto's equation is, of course, non-linear. The estimation of the parameters would have presented severe numerical difficulties for Pareto, and it is rather surprising that although he claimed to have obtained many estimates no explanation was given of the method of estimation used[15] (and only one example was presented—that of Oldenbourg). It is therefore of interest to examine the available evidence using an appropriate method of estimation based on the method of maximum likelihood.

III. Estimation and Results

Consider the estimation of the parameters a, α and θ of:

$$n_i = \log N_i = a - \alpha \log (x_i + \theta) + u_i$$

from a sample with T income groups. On the assumption that the u_i are independently normally distributed as $N(0, \sigma^2)$, then the loglikelihood of the sample is given by:

$$\log L = -\frac{T}{2}\log 2\pi - \frac{T}{2}\log \sigma^2 - \frac{1}{2\sigma^2}\sum_{i=1}^{T} u_i^2 \tag{8}$$

writing $w = 1/\sigma^2$ and $y = \theta + x$ the first derivatives of (8) are

$$\delta \log L/\delta a = w\Sigma u \qquad \delta \log L/\delta \alpha = -w\Sigma u \log y$$

$$\delta \log L/\delta \theta = -w\alpha\Sigma(u/y) \qquad \delta \log L/\delta \sigma^2 = -\frac{T}{2}w\left(1 + \frac{w}{T}\Sigma u^2\right) \tag{9}$$

The set of non linear maximum likelihood equations (9) can be solved using Fisher's method of Scoring. The so-called scoring equations are[16]

$$\left[-E\left\{\frac{\delta^2 \log L}{\delta p_i\, \delta p_j}\right\}\right][{}^{j+1}p_i - {}^{j}p_i] = \left[\frac{\delta \log L}{\delta p_i}\right] \tag{10}$$

Where ${}^{j}p_i$ is the estimate of the ith parameter p_i in the jth interation. The vector of first derivatives $\delta \log L/\delta p_i$ is the vector of "efficient scores" while the matrix of negative expectations of second derivatives is the "Information matrix". The inverse of this matrix in the last iteration is the variance-covariance matrix of the parameter estimates.

[15] Pareto found the method of least squares too laborious, noting (1896) "... la méthode des moindres carrés ... conduit ... à de très longs calcus. On a donc tâché de lui substituer d'autres méthodes plus simples". An interesting discussion of estimatimation of Pareto's curve can be found in Bowley (1926).

[16] This method is the same as Newton's, except that the 2nd derivatives are replaced by their negative expectations.

For σ^2 then:

$$\frac{\delta^2 \log L}{\delta(\sigma^2)^2}=\frac{T}{2\sigma^4}-\frac{1}{\sigma^6}\Sigma u^2$$

and

$$-E\left(\frac{\delta^2 \log L}{\delta(\sigma^2)^2}\right)=\frac{T}{2}w^2$$

Using the fact that the mean is independent of the variance for the normal distribution the information matrix can be partitioned, and (10) then gives:

$$\frac{T}{2}w^2({}^{j+1}\sigma^2-{}^{j}\sigma^2)=-\frac{Tw}{2}\left(1+\frac{w}{T}\Sigma u^2\right)$$

whence

$$(11) \qquad {}^{j+1}\sigma^2=\frac{1}{T}\Sigma u^2$$

Which need only be applied when stable values of the other parameters have been obtained. The scoring equations for the parameters a, θ and α can then be obtained as:

$$\begin{bmatrix} T & -\Sigma \log y & -\alpha\Sigma(1/y) \\ & \Sigma(\log y)^2 & \Sigma(u+\alpha \log y)/y \\ & & \alpha\Sigma(\alpha-u)/y^2 \end{bmatrix}\begin{bmatrix} j+1_a-j_a \\ j+1_\alpha-j_\alpha \\ j+1_\theta-j_\theta \end{bmatrix}\begin{matrix} = \\ = \\ = \end{matrix}\begin{bmatrix} \Sigma u \\ -\Sigma u \log y \\ -\alpha\Sigma u/y \end{bmatrix}$$

All values of u and y are calculated using parameter estimates obtained in the jth iteration.

The above routine was applied to the data used by Pareto, and the results given in Table 1. It can be seen that most of the values of θ are highly significant—and certainly do not justify Pareto's remarks about the relative values of θ. (Also compare the value for oldenbourg with that given by Pareto). Furthermore only four of the values of θ are negative, again contradicting Pareto's remark that θ is negative for earnings from employment. The ordinary least squares estimates of α using the conventional form of equation (2) are also given in Table 1 for comparison.

IV. Conclusion

The above results clearly show that Pareto's statements concerning the general form of the upper tail of the income distribution are not supported by the evidence which he actually used. Furthermore some of his statements are, on closer examination, not consistent.

It is also rather suprising that Pareto should publish only one estimate of his more general equation, when he claimed to have obtained many others (his summary—in terms of the sign of θ for different sources of income—is not confirmed by the data, but could easily have been checked from the convexity or concavity of the double-log graphs). There is no explanation of how such

TABLE 1

MAXIMUM LIKELIHOOD ESTIMATES OF $N = A/(x+\theta)^{\alpha}$

Sample	Constant (log A)	α	θ	*O.L.S. est of* α	Pareto's α
	(standard errors are in parenthesis)				
1 Angleterre 1843	20.001	1.718	123.959	1.556	1.50
	(0.524)	(0.059)	(46.244)	(0.039)	
2 Angleterre 1879–80	19.123	1.355	−27.933	1.395	1.35
	(0.229)	(0.028)	(14.535)	(0.020)	
3 Villes Italiennes	22.300	1.566	346.36	1.434	1.45
	(0.353)	(0.036)	(94.508)	(0.028)	
4 Saxe 1880	23.576	1.632	84.45	1.589	1.58
	(0.063)	(0.007)	(11.081)	(0.014)	
5 Saxe 1886	22.854	1.515	0.543	1.514	1.51
	(0.065)	(0.007)	(10.944)	(0.005)	
6 Bale	17.789	1.226	−80.435	1.249	—
	(0.265)	(0.026)	(77.757)	(0.017)	
7 Oldenbourg 1890	27.918	2.376	1108.86	1.634	1.465
	(1.115)	(0.114)	(210.969)	(0.100)	
8 Crédit Foncier de	13.699	1.601	9.186	1.089	1.5773
France 1888	(0.127)	(0.023)	(0.611)	(0.087)	
9 Crédit Foncier de	14.445	1.894	6.773	1.359	1.8172
France 1895	(0.289)	(0.052)	(0.969)	(0.096)	
10 Austria 1891	31.756	2.972	388.75	2.697	2.7114
	(0.705)	(0.075)	(91.385)	(0.039)	
11 Wartemberg 1890	20.520	1.488	458.27	1.191	1.4186
	(0.218)	(0.023)	(41.971)	(0.058)	
12 Amburgo 1891	18.806	1.111	−102.005	1.146	1.1308
	(0.182)	(0.018)	(43.515)	(0.013)	
13 Schaumbourg Lippe	16.632	1.299	9.472	1.282	1.2983
1893	(0.723)	(0.094)	(51.087)	(0.031)	
14 Zurich 1891	27.444	2.389	263.077	2.103	—
	(0.907)	(0.101)	(93.940)	(0.049)	
15 Wartemberg 1890	37.685	3.474	1539.08	2.030	2.881
	(1.521)	(0.160)	(96.820)	(0.158)	
16 Brema	17.444	1.124	−179.871	1.196	1.1814
	(0.279)	(0.029)	(55.794)	(0.025)	
17 Sassonia Weimia	23.386	1.860	348.59	1.563	1.6305
Eisenach 1892	(0.395)	(0.044)	(53.588)	(0.037)	1.7342

Samples 1 to 7 are distributions given in "La courbe de la répartition de la Richesse" (1896). *Universite de Lausanne.* Recueil publié par la Faculté de Droit à l'occasion de l'exposition nationale Suisse. Genève 1896. Sample 8–17 from "Aggiunta allo studio della curva delle entrate". *Giornale degi Economisti,* Janvier 1897, pp. 15–26.

'estimates' were obtained—though at that date any iterative method would have been extremely laborious.

Pareto's initial interpretation of his 'results' has for some time been questioned; it now seems that there are further grounds for criticism.

REFERENCES

Bowley, A. L., 1926, *Elements of Statistics*: P. S. King & Son.

Brown, J. A. C., 1976, "The mathematical and statistical theory of income distribution," in *The Personal Distribution of Income* edited by A. B. Atkinson for the Royal Economic Society: Allen & Unwin.

Champernowne, D. G., 1953, "A model of income distribution," *Economic Journal*, 68.
Cheshire, P. C. and Stabler, M. J., 1976, "Joint consumption benefits in recreational site 'surplus': An empirical estimate". *Regional studies*, 10, No 3.
Gibrat, R., 1931, *Les Inegalites Economiques.* Paris: Libraire du Recueil Sirey.
Lydall, H., 1968, *The Structure of Earnings*: Oxford University Press.
Mandelbrot, B., 1960, "The Pareto-Levey Law and the distribution of income". *International economic Review*, 1.
Pareto, V., 1896, "La Courbe de la Repartition de la Richesse", *Universite de Lausanne.*
———, V., 1897, *Cours d'economie Politique.*
———, V., 1909, *Manuel d'economie Politique.* Geneve: Librairee Droz.
———, V., 1906, "Determination mathematique des resultats des experience". Reprinted in Vol 8 of *Deuvres Complete.*
Schumpeter, J., 1968, *Ten Great Economists.* Allen & Unwin.
Spencer, H., 1897, *Principles of Ethics*, Vol. II.
Steindl, J., 1965, *Random Processes and the Growth of Firms: A study of the Pareto Law.* Griffin: London.
Tawney, R., 1920, *Equality.* Allen and Unwin.